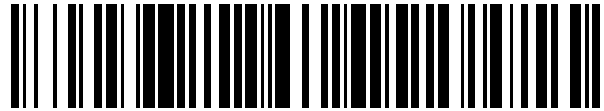


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 392 320**

51 Int. Cl.:

G06K 9/00 (2006.01)

G06K 9/62 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **09738413 .5**

96 Fecha de presentación: **15.04.2009**

97 Número de publicación de la solicitud: **2291794**

97 Fecha de publicación de la solicitud: **09.03.2011**

54 Título: **Tratamiento de imágenes**

30 Prioridad:

28.04.2008 EP 08155309

28.04.2008 GB 0807675

45 Fecha de publicación de la mención BOPI:

07.12.2012

45 Fecha de la publicación del folleto de la patente:

07.12.2012

73 Titular/es:

BAE SYSTEMS PLC (100.0%)

6 Carlton Gardens

London SW1Y 5AD, GB

72 Inventor/es:

WILLIS, CHRISTOPHER, JON

74 Agente/Representante:

CARPINTERO LÓPEZ, Mario

ES 2 392 320 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

DESCRIPCIÓN

Tratamiento de imágenes

Esta invención se refiere a tratamiento de imágenes y en particular a un método y a un dispositivo para clasificar el contenido en una imagen hiperespectral, por ejemplo mediante clasificación de regiones de uno o más píxeles que aparecen en una escena y la detección de anomalías.

La mayoría de métodos conocidos de tratamiento de imágenes hiperespectrales, basados en una teoría de probabilidad estadística bien conocida y aceptada, aplicada al tratamiento de reconocimiento estadístico de patrones, han sido desarrollados suponiendo que los datos espectrales siguen una distribución gaussiana multivariante correlacionada. Estos métodos conocidos incluyen una aproximación de detección tal como el método RX descrito en Reed, I. S., Yu, X., "Adaptive Multiple Band CFAR Detection of an Optical Pattern with Unknown Spectral Distribution", IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. 38, pp. 1760-1770, 1990, o un proceso de reconocimiento más clásico tal como el descrito en Fukunaga, K., "Introduction to Statistical Pattern Recognition", Second Edition, Academic Press, 1990.

Sin embargo, muchos de los resultados derivados usando aproximaciones basadas en el modelo estadístico gaussiano muestran tasas de falsas alarmas mucho mayores de lo que se podría esperar. Es decir, los estadísticos derivados usando los métodos de reconocimiento muestran colas más pesadas de lo esperado en sus distribuciones. Este efecto es apuntado, y pueden observarse ejemplos de él, en Manolakis, D., Shaw, G., "Detection Algorithms for Hyperspectral Imaging Applications", IEEE Signal Processing Magazine, pp. 29-43, Enero de 2002, y en Stein, D. W. J., Beaven, S. G., Hoff, L. E., Winter, E. M., Schaum, A. P., Stocker, A. D., "Anomaly Detection from Hyperspectral Imagery", IEEE Signal Processing Magazine, pp. 58-69, Enero de 2002.

El artículo anteriormente citado de Manolakis y Shaw muestra ejemplos de "colas pesadas" en la salida de un detector de anomalías basado en la distancia de Mahalanobis, en imágenes hiperespectrales, en sus figuras 14 y 15. Estas figuras también muestran los resultados de intentos de ajustar una mezcla de distribuciones F al estadístico de distancia de Mahalanobis, o alternativamente miembros parametrizados de la familia de distribuciones α -estables simétricas (SaS, del inglés "Symmetric α -Stable"). Manolakis & Shaw también apuntan que si los datos siguieran una distribución t de Student multivariante correlacionada, entonces el estadístico de distancia de Mahalanobis tendría una distribución F. En ambos casos, tanto en la mezcla de distribuciones F como en los casos derivados de SaS, se consigue un mejor ajuste que para la aproximación basada en la gaussiana, aunque no se discuten métodos para parametrizar estas distribuciones más exóticas.

El artículo anteriormente citado de Stein et al. también muestra un ejemplo de colas pesadas en el estadístico de distancia de Mahalanobis en su figura 4. Evidentemente, ha habido un intento de ajustar un modelo de mezcla al estadístico observado y de obtener un buen ajuste a partir de una mezcla de tres distribuciones gamma. Sin embargo, sigue sin estar claro cómo volver a ligar la distribución del estadístico de distancia de Mahalanobis observado a la distribución de los datos de imagen hiperespectral recogidos.

Sun J et al: "Robust Mixtures in the Presence of Measurement Errors" ACM International Conference Proceeding Series - Proceedings, Twenty-Fourth International Conference on Machine Learning, ICML 2007 Association for Computing Machinery US, vol. 227, 2007, páginas 847-854, da a conocer una aproximación basada en mezcla para modelado robusto de densidad y detección de valores raros (*outliers*) para datos multivariantes experimentales que incluyen información de error de medida. El modelo está diseñado para inferir medidas atípicas que no son debidas a errores, con el objetivo de recuperar objetos peculiares potencialmente interesantes.

Peel D et al: "Robust Mixture Modelling using the t-distribution", Statistics and Computing, vol. 10, 2000, páginas 339-348 da a conocer un modelado de datos mediante una mezcla de distribuciones t. Se describe el uso del algoritmo ECM para ajustar un modelo de mezcla de distribuciones t y se dan ejemplos de su uso en el contexto de agrupación (*clustering*) de datos multivariantes en la presencia de observaciones atípicas en la forma de ruido de fondo.

En un primer aspecto, la presente invención estriba en un método para clasificar regiones dentro de una escena representada en una imagen hiperespectral, en que las regiones dentro de la escena son clasificadas de acuerdo con su probabilidad de pertenencia a uno o más componentes en un modelo estadístico que tiene una verosimilitud de modelo de ser representativo del contenido de la escena, en que el método comprende los pasos de:

- (i) para cada elemento muestral de aprendizaje en un conjunto de datos de aprendizaje, asignar probabilidades de pertenencia actuales iniciales a cada elemento muestral de aprendizaje;
- (ii) asignar cada elemento muestral de aprendizaje a uno de los componentes de acuerdo con su probabilidad de pertenencia actual;
- (iii) para cada componente, determinar la probabilidad a priori de componente y otros estadísticos de componente, usando una medida de rareza determinada para cada elemento muestral de aprendizaje;

(iv) estimar una nueva probabilidad a posteriori de componente para cada elemento muestral de aprendizaje usando probabilidades condicionales de componente derivadas usando dicha medida de rareza y dichos otros estadísticos de componente; y

5 (v) repetir los pasos (ii) a (iv) para mejorar la verosimilitud de modelo, usando la nueva probabilidad a posteriori de componente para cada elemento muestral de aprendizaje procedente del paso (iv) como probabilidad de pertenencia actual para el respectivo elemento muestral de aprendizaje en el paso (ii), en que dicha medida de rareza es determinada usando un parámetro v de una distribución t de Student aplicada para pesar la contribución de los valores muestrales de aprendizaje a los estadísticos para cada componente.

10 De acuerdo con este primer aspecto de la presente invención, se ha encontrado que la inclusión de una medida de rareza, apropiadamente formulada, en el cálculo de estadísticos de componente proporciona un pesado más apropiado para la contribución de cada uno de los elementos muestrales de aprendizaje a esos estadísticos al generar el modelo. En particular, la medida de rareza tiende a evitar problemas en técnicas de modelado estadístico conocidas en las que aparecen distorsiones en los estadísticos de componente debido a la presencia de valores raros en los datos de aprendizaje.

15 Preferiblemente, el conjunto de datos de aprendizaje comprende datos que representan una o más regiones de la escena. Éste es el caso particularmente cuando una escena está siendo analizada sustancialmente en tiempo real y se está haciendo un intento de modelar la escena, usando algunos o todos los datos que representan la escena, y luego identificar ciertas regiones de la misma escena, no necesariamente regiones que fueron usadas en el modelado de la escena, para un análisis más detallado o diferente, por ejemplo usando sensores diferentes. De este modo, en una realización preferida de la presente invención de acuerdo con este primer aspecto, el método comprende además el paso de:

(vi) para una región dada en la escena, no representada en el conjunto de datos de aprendizaje, determinar la probabilidad de su pertenencia a uno o más componentes del modelo estadístico generado en los pasos (i) a (v) del método.

25 El conjunto de datos de aprendizaje puede comprender datos que representan todas las regiones de la escena, o puede comprender datos que representan una o más regiones de una escena diferente. Esta última opción permite generar un modelo estadístico de una escena con contenido similar y luego, algún tiempo después, que ese modelo sea usado para clasificar una o más regiones de una nueva escena. Este modo de operación puede ser particularmente útil en vuelo de avión sobre un área de paisaje similar, por ejemplo un desierto, donde un modelo estadístico construido usando datos capturados de una o más escenas previas puede usarse para analizar más rápidamente regiones de escenas nuevas sin tener que generar un modelo estadístico nuevo. Por supuesto, puede generarse una librería de modelos estadísticos para diferentes tipos de escena para uso posterior. El modelo más apropiado puede ser seleccionado para uso con una escena nueva según sea necesario.

30 En todos los casos, las escenas pueden ser clasificadas y/o modeladas al nivel de resolución de píxel. Así, puede generarse un modelo estadístico sobre la base de que cada elemento muestral de aprendizaje comprende datos que definen un píxel en una escena respectiva y cada uno de los componentes en el modelo estadístico representa uno o más píxeles en la escena. Cuando se clasifican regiones de una escena, cada una de las regiones puede comprender un píxel en la escena. Éste es el caso particularmente cuando se clasifican regiones en una escena hiperespectral.

40 La medida de rareza puede comprender una estimación del parámetro v determinada separadamente para cada componente en cada aplicación del paso (iii) del método, en combinación con un valor para el número de componentes en el modelo estadístico, para pesar la contribución de los valores muestrales de aprendizaje a los estadísticos para cada componente respectivo. Alternativamente, puede ser determinado un valor estimado común del parámetro v para todos los componentes, tal como se determina en cada aplicación del paso (iii) del método. En particular, la estimación del parámetro v puede comprender fijar el valor del parámetro v de modo que sea igual al número de componentes en el modelo estadístico en cada aplicación del paso (iii) del método.

45 Los análisis de Manolakis & Shaw y de Stein et al. anteriormente citados parecen haber sido llevados a cabo en gran medida a posteriori, observando, e intentando ajustar modelos al estadístico derivado. En contraste, las realizaciones preferidas de la presente invención intentan explotar la hipótesis de un modelo estadístico de cola pesada en la derivación del propio estadístico de prueba. En una realización preferida, la presente invención se basa en la derivación de una interpretación del contenido de una escena hiperespectral en términos de diferentes tipos de material de cubierta de suelo. Una distribución t de Student multivariante correlacionada es usada para caracterizar cada material. Esto resulta en la derivación de un modelo de mezcla de distribuciones t para toda la escena, que, a su vez, conduce a un modelo de mezcla de distribuciones F para la distancia de Mahalanobis. Esto último permite el desarrollo de un detector de anomalías (también conocidas como novedades o valores raros), con la distancia de Mahalanobis como estadístico de prueba.

En una realización preferida de la presente invención, el método comprende además el paso de identificar una región de la escena que tenga una probabilidad de pertenencia a uno o más componentes del modelo estadístico

que esté por debajo de un umbral predeterminado. De este modo, las regiones que no parecen corresponder con las representadas en el modelo estadístico – los así denominados valores raros – pueden ser fácilmente identificadas para análisis adicional.

5 En un segundo aspecto, la presente invención estriba en un dispositivo de tratamiento de datos programado para implementar el método de acuerdo con el primer aspecto de la presente invención.

En un tercer aspecto, la presente invención también se extiende a un programa de ordenador que, cuando es cargado en un ordenador y es ejecutado, hace que el ordenador implemente el método de acuerdo con el primer aspecto de la presente invención.

10 En un cuarto aspecto, la presente invención se extiende además a un producto de programa de ordenador, que comprende un medio legible por ordenador que lleva almacenado en él un medio de código de programa de ordenador que cuando es cargado en un ordenador, y es ejecutado, hace que el ordenador implemente el método de acuerdo con el primer aspecto de la presente invención.

Serán descritas ahora realizaciones preferidas de la presente invención, sólo a modo de ejemplo.

15 Procesos de interpretación tradicionales de escenas aplicados a imágenes hiperespectrales hacen uso típicamente de métodos del tipo de maximización de la esperanza (EM, del inglés "Expectation Maximisation") o de maximización EM estocástica (del inglés "Stochastic EM") para segmentar una escena en regiones de un tipo de material común. Estos métodos caen en la clase de métodos conocidos como "segregación (*unmixing*)". En estos procesos de interpretación de escenas se supone que los estadísticos de los materiales en la escena pueden ser representados por modelos estadísticos gaussianos multivariantes y los procesos de segregación EM o SEM calculan iterativamente los estadísticos de tales distribuciones (el paso de Esperanza) y luego llevan a cabo una reasignación o bien suave (para maximización EM) o probabilística (para maximización SEM) de elementos muestrales de aprendizaje a componentes de modelo de mezcla (el paso de Maximización) hasta la convergencia.

20 Es probable que todas las escenas, excepto las más uniformes, representadas en imágenes hiperespectrales contengan pequeñas regiones de píxeles con una firma espectral que difiere significativamente de la del grueso de los píxeles en la escena. Uno de los objetivos principales del uso del proceso de interpretación (segregación) de escenas (EM o SEM) para imágenes hiperespectrales es que el modelo estadístico resultante proporcione una buena representación de la mayoría de los píxeles en la escena, permitiendo con ello que las pequeñas regiones de material no común sean identificadas como valores raros (también denominados anomalías o píxeles/regiones novedosos). Tales procesos son útiles, por ejemplo, en detección de objetivos o en prospección de minerales raros, etc. Sin embargo, en el modo más común de aplicación de estas técnicas de interpretación de escenas, la escena usada para construir (formar por aprendizaje) el modelo estadístico es la misma en la que deben ser encontrados los píxeles raros. Esto puede ser un problema para los procesos de interpretación de escenas ya que los píxeles anómalos contaminan el modelo cuando está siendo derivado. Esto es, dichos píxeles pueden tirar de o estirar la distribución o las distribuciones que forman el modelo alejándolas de los vectores medios y matrices de covarianza que habrían sido derivadas si estos píxeles anómalos no hubieran estado presentes. Éste es un problema particular para distribuciones gaussianas en las que el ajuste por mínimos cuadrados de los datos confiere la influencia más grande sobre el modelo a los elementos muestrales más raros (los píxeles hiperespectrales anómalos en este caso).

25 El inventor de la presente invención se ha dado cuenta de que lo que se requiere es un modelo estadístico para un componente de material único que efectivamente reduzca la influencia de los píxeles más raros al tiempo que siga representando un buen modelo de los valores comunes, y un medio para derivar un modelo así a partir de una escena con muchos tipos de material dentro de un marco de segregación. El requisito anterior de un modelo estadístico de componente que tolere valores raros sin afectar significativamente la media derivada y los estadísticos de amplitud puede ser satisfecho mediante el uso de una así denominada distribución de cola pesada. Hay muchos tipos y variantes de distribuciones de cola pesada pero aquéllos que se sabe que han sido considerados incluyen miembros de la familia de distribuciones α -estables simétricas (S α S) (véase Manolakis & Shaw, 2002, citado anteriormente), y un estadístico derivado no especificado modelado mediante una mezcla de distribuciones gamma (véase Stein et al., 2002, citado anteriormente). Sin embargo, en ninguna de estas aproximaciones conocidas hay una motivación aparente de considerar el requisito del modelo, o una exposición de medios para generarlo.

30 Otro tipo de distribución de cola pesada es la distribución t de Student y una aproximación basada en esta distribución es el sujeto de la presente invención.

Habiendo establecido la necesidad de distribuciones de cola pesada en la representación de escenas hiperespectrales que contienen anomalías, se requiere un medio de parametrizar un modelo de mezcla compuesto por componentes de este tipo. Un conjunto de esquemas de maximización de la esperanza adecuados para parametrizar una mezcla estadística de distribuciones t de Student multivariantes es descrito en los párrafos posteriores.

55 Un método de interpretación de escenas hiperespectrales de acuerdo con una realización preferida de la presente invención comprende dos pasos principales:

- (1) Extracción de un modelo de mezcla usando un conjunto de datos de aprendizaje que representa una región de aprendizaje seleccionada de una escena dada. El modelo de mezcla comprende una segmentación de la escena de aprendizaje en componentes identificables, en que cada componente comprende un modelo estadístico que representa un material identificable diferenciado, por ejemplo, en la escena.
- 5 (2) Asignación de los píxeles de un área de prueba de la escena a los componentes del modelo de mezcla extraído.

Dependiendo de la aproximación particular escogida, la región de aprendizaje seleccionada para el paso (1) puede ser toda la escena y/o puede ser coincidente con parte o toda el área de prueba del paso (2). Alternativamente, la región de aprendizaje y el área de prueba pueden ser completamente distintas.

- 10 Preferiblemente, el proceso de extracción de modelo en el paso (1) usa una variante del algoritmo estocástico de maximización de la esperanza (SEM) para extraer un modelo de mezcla definiendo K componentes identificables a partir del conjunto de datos de aprendizaje. El algoritmo SEM se describe, por ejemplo, en Masson, P., Pieczynski, W., "SEM Algorithm and Unsupervised Statistical Segmentation of Satellite Images", IEEE Transactions on Geoscience and Remote Sensing, Vol. 31, n° 3, pp. 618-633, Mayo de 1993. El algoritmo SEM, en su forma estándar, se usa para extraer los parámetros – la media y la matriz de covarianza para cada uno de los K componentes identificables – de un modelo de mezcla gaussiano y es en sí mismo un desarrollo del algoritmo de maximización de la esperanza (EM). Para más información sobre el método EM, se hace referencia a un artículo de Dempster, M. M., Laird, N. M., Rubin, D. B., "Maximum Likelihood from Incomplete Data via the EM Algorithm", Journal of the Royal Statistical Society, Series B, Vol. 39, pp. 1-38, 1977.
- 15
- 20 La familia de algoritmos de maximización de la esperanza proceden calculando iterativamente la probabilidad a posteriori de cada elemento muestral de un conjunto de datos de aprendizaje y actualizando luego los estadísticos de la agrupación usando estas probabilidades a posteriori. Según avanza el algoritmo, la verosimilitud a posteriori de los elementos muestrales es incrementada iterativamente y el algoritmo termina cuando los cambios en la probabilidad a posteriori de un paso al siguiente caen por debajo de algún umbral predeterminado (establecido por el usuario).
- 25

El algoritmo SEM convencional es una variación de la aproximación EM en la cual se considera que cada elemento muestral de datos de aprendizaje pertenece a un único componente del modelo de mezcla, en vez de tener su probabilidad a posteriori distribuida entre los componentes. Esta aproximación permite adicionalmente la identificación de componentes redundantes de modelo que están "explicando" un número demasiado bajo de los elementos muestrales de datos de aprendizaje. Estos componentes de modelo pueden ser eliminados en caso apropiado, por ejemplo si el número de píxeles que representan un material dado en una escena es demasiado pequeño para modelado estadístico.

30

De acuerdo con una realización preferida de la presente invención, se ha diseñado una variación del algoritmo SEM convencional para uso en un proceso de interpretación de escenas hiperespectrales, que procede como sigue:

- 35 (i) Para cada elemento muestral de datos de aprendizaje, asignar una probabilidad de que pertenezca a cada uno de los K componentes iniciales. Si no está disponible ninguna información a priori relativa a probabilidades de pertenencia a componentes, estas probabilidades pueden tomarse de una distribución uniforme.
- 40 (ii) Asignar cada elemento muestral de aprendizaje a uno de los K componentes de acuerdo con la probabilidad de pertenencia actual del elemento muestral de aprendizaje. La pertenencia está denotada por la variable indicadora Z_{ik} , que, para el algoritmo SEM, toma un valor de uno para un único valor del índice k para cada uno de los elementos muestrales de datos de aprendizaje indexados por i ; todos los demás valores de Z_{ik} son cero.
- 45 (iii) Usando la pertenencia a componente actual, calcular la probabilidad a priori de componente y otros estadísticos de componente, es decir el vector medio y la matriz de covarianza para cada elemento muestral de datos de aprendizaje. Éstos son determinados usando el siguiente método.

El método requiere el cálculo de una pareja de variables auxiliares. La primera variable auxiliar es la distancia de Mahalanobis del i -ésimo valor muestral de aprendizaje x_i , actualmente asignado al centro del k -ésimo componente, es decir

50

$$(\Delta'_{ik})^2 = (x_i - \mu'_k)^T (\Sigma'_k)^{-1} (x_i - \mu'_k)$$

donde los estadísticos del componente (μ_k', Σ_k') son las estimaciones previas producidas por este algoritmo SEM. Para la primera iteración de este algoritmo SEM, se usan estimaciones de máxima verosimilitud para las estimaciones previas de los vectores de media condicional μ_i' y las matrices de covarianza Σ_i' del componente.

La segunda variable auxiliar es una medida de la falta de rareza de cada elemento muestral, dada por

5

$$u_{ik} = \frac{p + v_k'}{(\Delta_{ik}')^2 + v_k'}$$

10 donde el valor v_k' usado es la estimación previa del parámetro del k -ésimo componente y p es la dimensionalidad del conjunto de datos. Esta medida de falta de rareza pretende medir la contribución de los valores muestrales de aprendizaje x_i a la media y covarianza del k -ésimo componente. Obsérvese que, al ser el valor esperado de la distancia de Mahalanobis propiamente p , el parámetro u de un elemento muestral será cercano a 1, excepto para valores raros. Para la primera iteración de este algoritmo SEM, se usan estimaciones basadas en momentos para los parámetros v_i' condicionales del componente.

La probabilidad a priori del componente es calculada usando

15

$$p_k = \frac{\sum_{i=1}^n z_{ik}}{n}$$

20 Usando las variables auxiliares anteriormente descritas, los nuevos valores de la media y la covarianza para el k -ésimo componente son

$$\mu_k = \frac{\sum_{i=1}^n z_{ik} u_{ik} x_{ik}}{\sum_{i=1}^n z_{ik} u_{ik}}$$

y

25

$$\Sigma_k = \frac{\sum_{i=1}^n (z_{ik} u_{ik})(x_{ik} - \mu_{ik})(x_{ik} - \mu_{ik})^T}{\sum_{i=1}^n z_{ik} u_{ik}}$$

Se discuten posteriormente en detalle técnicas preferidas para la estimación del parámetro v_k o v .

30 (iv) Usar los estadísticos calculados para estimar la probabilidad a posteriori de componente de cada elemento muestral de aprendizaje x usando la probabilidad condicional de componente

35

$$p(x | \mu, \Sigma, v) = \frac{\Gamma\left(\frac{v+p}{2}\right)}{\Gamma\left(\frac{v}{2}\right) (v\pi)^{p/2} |\Sigma|^{1/2}} \left(1 + \frac{\Delta^2}{v}\right)^{-\frac{v+p}{2}},$$

donde Δ^2 es la distancia de Mahalanobis del elemento muestral de aprendizaje x usando las estimaciones actuales de los estadísticos de componente, y la probabilidad a priori de componente.

- (v) Asignar cada elemento muestral de aprendizaje a uno de los K componentes de acuerdo con su probabilidad (a posteriori de componente) de pertenencia actual.
- (vi) Si algún componente es demasiado pequeño, reducir K y redistribuir los elementos muestrales de aprendizaje actualmente asignados a los componentes eliminados aleatoriamente entre todos los demás componentes, y volver al paso (iii);
- (vii) Repetir a partir del paso (iii) hasta que el cambio en la verosimilitud total del modelo sea menor que algún umbral pre-establecido.

En el caso particular en que los conjuntos de datos de aprendizaje y de prueba son los mismos, las asignaciones de clase de los elementos muestrales de aprendizaje, realizadas en el paso de transferencia final (v) anterior, se usan como segmentación (z_{ik}) de los datos que representan la escena para visualización y pasos de tratamiento subsiguientes.

Serán descritos ahora en cierto detalle ejemplos preferidos de técnicas de evaluación del parámetro v , de acuerdo con realizaciones preferidas de la presente invención. Preferiblemente, pueden usarse cuatro métodos diferentes en dos situaciones. Éstos son:

- (A) basado en la estimación de momentos, usando específicamente la estimación de la kurtosis;
- (B) usando una modificación de un esquema esbozado en el artículo de Shoham et al., citado anteriormente;
- (C) usando un método de Monte Carlo de cadena de Markov (MCMC, del inglés "Markov-Chain Monte Carlo"); y
- (D) fijando el valor de modo que iguale al número de componentes considerados.

Los métodos (A)-(C) pueden usarse o bien para la situación en la que cada componente tiene un valor v separado o para el caso en que v es común a través de todos los componentes. El método (D) se usa sólo en la situación de v común.

El método (A) está basado en una estimación de la kurtosis. La kurtosis está definida en este caso por la fórmula:

$$\kappa = \frac{\mu_4}{\mu_2^2}$$

donde μ_i es el i -ésimo momento central (de modo que el denominador es aquí la varianza al cuadrado). La kurtosis se estima separadamente para cada componente, y para cada una de las p dimensiones de los datos (canales/bandas), individualmente, dentro del componente. Se toma el promedio de las respectivas estimaciones de kurtosis de banda para formar una kurtosis media para cada componente.

Para el caso de v separado los valores de kurtosis de componente son usados para extraer una estimación del parámetro v usando lo siguiente:

$$v_i = \frac{4\kappa_i - 6}{\kappa_i - 3}$$

donde v_i es la estimación de componente del parámetro v .

Para la estimación del parámetro v común se calcula el promedio con pesos a priori de los valores de kurtosis de componente usando la estimación actual de la probabilidad a priori de componente. La estimación de kurtosis común resultante se usa para calcular el valor v común como sigue:

$$v = \frac{4\kappa - 6}{\kappa - 3}$$

En ambos casos, tanto el de v común como el de v separado, el valor de parámetro calculado se usa en la siguiente iteración del esquema SEM. El estimador, basado en la kurtosis, del parámetro v es de programación muy directa y tiene poco coste computacional.

El método (B) es una modificación del esquema esbozado en el artículo de Shoham et al., citado anteriormente. En particular, el método (B) usa una modificación del estimador empírico descrito por Shoham et al. Las modificaciones propuestas en la presente invención se refieren al uso de la aproximación de Shoham et al. dentro de un esquema estocástico de maximización de la esperanza (con pertenencias duras a componente de los elementos muestrales), en vez de en el esquema de maximización de la esperanza (con pertenencias blandas a componente) descrito en su artículo, citado anteriormente.

El método (B) requiere el cálculo de una secuencia de variables auxiliares. La primera requerida es la distancia de Mahalanobis del elemento muestral (indexado por i) actualmente asignado al centro de componente (indexado por k), es decir,

10

$$\Delta_{ik}^2 = (x_i - \mu_k)^T \Sigma_k^{-1} (x_i - \mu_k)$$

donde los estadísticos de los componentes (μ_k, Σ_k) son las estimaciones actuales producidas por el esquema SEM.

La segunda variable auxiliar es una medida de la falta de rareza de cada elemento muestral, dada por

15

$$u_{ik} = \frac{p + v_k'}{\Delta_{ik}^2 + v_k'}$$

donde el valor v_k' usado es la estimación previa para el parámetro del k -ésimo componente y p es la dimensionalidad del conjunto de datos. Obsérvese que, como el valor esperado de la distancia de Mahalanobis es en sí misma p , entonces el parámetro u de un elemento muestral será cercano a 1, excepto para valores raros.

20

Para el caso de valor v separado, la tercera variable auxiliar está dada por

25

$$y_k = -\Psi\left(\frac{p + v_k'}{2}\right) - \frac{1}{n_k} \sum_{i=1}^{n_k} \left[\log\left(\frac{2}{\Delta_{ik}^2 + v_k'}\right) - u_{ik} \right]$$

donde n_k es el número de elementos muestrales asignados actualmente al k -ésimo componente del modelo de mezcla y $\Psi(\cdot)$ es la función digamma.

Finalmente, los valores del componente v son calculados usando la fórmula empírica de Shoham et al.

30

$$v_k = \frac{2}{y_k + \log y_k - 1} + 0.0416 \left(1 + \operatorname{erf} \left(0.6594 \log \left(\frac{2.1971}{y_k + \log y_k - 1} \right) \right) \right)$$

Para el caso de v común, las variables auxiliares están dadas por

35

$$u_{ik} = \frac{p + v'}{\Delta_{ik}^2 + v'}$$

donde v' es la estimación previa del parámetro v común, e

40

$$y = -\Psi\left(\frac{p + v'}{2}\right) - \frac{1}{n} \sum_{i=1}^n \left[\log\left(\frac{2}{\Delta_{ik}^2 + v'}\right) - u_{ik} \right].$$

Aquí el propio valor v está dado por

$$v = \frac{2}{y + \log y - 1} + 0.0416 \left(1 + \operatorname{erf} \left(0.6594 \log \left(\frac{2.1971}{y + \log y - 1} \right) \right) \right)$$

El esquema de Shoham et al. modificado del método (B) es un poco más complicado de programar, pero sigue siendo relativamente directo y tiene poco coste computacional.

La tercera aproximación, el método (C), usa un método de Monte Carlo de cadena de Markov (MCMC) para inferir el parámetro v usando estimaciones actuales de los otros parámetros. Como en el caso de los métodos (A) y (B), el método (C) puede ser aplicado o bien separadamente a cada componente o a través de todos los componentes resultando en una estimación de v común. El método MCMC proporciona un gran conjunto de herramientas de técnicas de inferencia estadística. El método (C) está basado en la propuesta de un nuevo valor potencial para el parámetro sobre el cual se está realizando la inferencia, seguida por o bien la aceptación o bien el rechazo de este valor propuesto dependiendo de cómo afecta a la verosimilitud de observar los datos suministrados. La probabilidad de aceptación en un esquema MCMC estándar está dada por

$$j = \min \left(1, \frac{p(\underline{x} | \theta') p(\theta') p(\theta | \theta')}{p(\underline{x} | \theta) p(\theta) p(\theta' | \theta)} \right).$$

Aquí, el primer término de cociente es el cociente de las verosimilitudes de observar los datos bajo el modelo propuesto con el conjunto de parámetros θ' y el modelo actual con el conjunto de parámetros θ , el segundo cociente es entre las probabilidades a priori de los dos conjuntos de parámetros y el cociente final es entre las distribuciones propuestas del conjunto de parámetros actual condicionado a los valores propuestos y viceversa. Para una introducción a estos métodos, véase por ejemplo Green, P. J., "A Primer on Markov chain Monte Carlo" en Complex Stochastic Systems, pp. 1-62, Barndorff-Nielsen, O. E., Cox, D. R. and Kluppelberg, C. (eds.), Chapman and Hall, London, (2001).

El esquema MCMC en el método (C) usa los valores actuales para los vectores de media de componentes μ_k , las matrices de covarianza Σ_k , y las distancias de Mahalanobis de los elementos muestrales Δ_{ik}^2 (calculadas como anteriormente). Los cocientes de verosimilitud dentro del esquema son calculados usando la función de densidad de probabilidad para variables aleatorias que siguen una distribución t de Student, es decir

$$p(x | \mu, \Sigma, v) = \frac{\Gamma\left(\frac{v+p}{2}\right)}{\Gamma\left(\frac{v}{2}\right) (v\pi)^{p/2} |\Sigma|^{1/2}} \left(1 + \frac{\Delta^2}{v}\right)^{-\frac{v+p}{2}}.$$

El esquema MCMC tiene un único parámetro para inferencia en cualquier momento. Para el caso de v separado, éste es el v_k de cada componente uno por uno, mientras que para el caso de v común todos los componentes son tratados conjuntamente.

Para las probabilidades a priori del parámetro, se escogen distribuciones previas no informativas de modo que el cociente a priori es siempre 1. Para la distribución propuesta, se selecciona la log-normal usando el valor actual como la media y con una varianza ajustable. Esto es un ejemplo de un camino aleatorio sobre el método de escala logarítmica que simplifica la evaluación del cociente de propuesta y restringe naturalmente la estimación de v a valores positivos según se requiere. La varianza de la distribución propuesta puede ajustarse para restringir el cociente de aceptación MCMC al rango deseado (típicamente 25%-45%).

La aproximación MCMC tiene muchos aspectos a su favor cuando se lleva a cabo inferencia estadística. No es el menor la producción de una muestra a posteriori del parámetro o parámetros de interés que puede ser usada luego en inferencia bayesiana. Sin embargo, para la aplicación aquí considerada, la aproximación MCMC es probablemente excesiva y, adicionalmente, tiene varias características que la hacen demasiado complicada. Éstas incluyen:

- Requisitos computacionales significativos en el cálculo del valor de verosimilitud del modelo para el parámetro propuesto;

- problemas al evaluar convergencia y mezcla, y;
- falta de un camino claro para seleccionar o extraer un valor de parámetro único a partir de la muestra a posteriori generada por el esquema MCMC.

5 Al usar varios de los métodos (A) – (C) anteriormente discutidos, se ha observado que los valores derivados para el parámetro v a menudo o bien convergían a valores cercanos al del número de componentes en el modelo de mezcla, o bien resultaban tener tal valor cuando eran calculados. Por lo tanto, se propone un método adicional (D) que comprende fijar el valor v igualándolo al número de componentes. Claramente éste el esquema más simple para establecer el valor del parámetro con el menor coste de programación y computacional.

10 Se han propuesto cuatro métodos, (A) a (D), para la evaluación del parámetro v en el método SEM para modelos de mezcla de distribuciones t de Student. Tres de los métodos pueden usarse sobre una base componente a componente y todos pueden ser usados dentro de un modelo basado en v común. Los mejores resultados han sido observados, para los datos hiperespectrales examinados, para los esquemas basados en v común. Más allá de ello se ha observado poca diferencia en rendimiento entre los diferentes esquemas de parametrización. En este sentido, el método (A) ha sido priorizado en aplicaciones típicas de la presente invención, al ser simple, rápido y

15 razonablemente adaptable a todas las situaciones observadas.

REIVINDICACIONES

1. Un método para clasificar regiones dentro de una escena representada en una imagen hiperespectral, en que las regiones dentro de la escena son clasificadas de acuerdo con su probabilidad de pertenencia a uno o más componentes en un modelo estadístico que tiene una verosimilitud de modelo de ser representativo del contenido de la escena, en que el método comprende los pasos de:
- (i) para cada elemento muestral de aprendizaje en un conjunto de datos de aprendizaje, asignar probabilidades de pertenencia actuales iniciales a cada elemento muestral de aprendizaje;
 - (ii) asignar cada elemento muestral de aprendizaje a uno de los componentes de acuerdo con su probabilidad de pertenencia actual;
 - (iii) para cada componente, determinar la probabilidad a priori de componente y otros estadísticos de componente, usando una medida de rareza determinada para cada elemento muestral de aprendizaje;
 - (iv) estimar una nueva probabilidad a posteriori de componente para cada elemento muestral de aprendizaje usando probabilidades condicionales de componente derivadas usando dicha medida de rareza y dichos otros estadísticos de componente; y
 - (v) repetir los pasos (ii) a (iv) para mejorar la verosimilitud de modelo, usando la nueva probabilidad a posteriori de componente para cada elemento muestral de aprendizaje procedente del paso (iv) como probabilidad de pertenencia actual para el respectivo elemento muestral de aprendizaje en el paso (ii); en que dicha medida de rareza es determinada usando un parámetro v de una distribución t de Student aplicada para pesar la contribución de los valores muestrales de aprendizaje a los estadísticos para cada componente.
2. El método según la reivindicación 1, en que el conjunto de datos de aprendizaje comprende datos que representan una o más regiones de la escena.
3. El método según la reivindicación 1 o la reivindicación 2, que comprende además el paso de:
- (vi) para una región dada en la escena, no representada en el conjunto de datos de aprendizaje, determinar la probabilidad de su pertenencia a uno o más componentes del modelo estadístico generado en los pasos (i) a (v) del método.
4. El método según la reivindicación 1, en que el conjunto de datos de aprendizaje comprende datos que representan todas las regiones de la escena.
5. El método según la reivindicación 1 o la reivindicación 3, en que el conjunto de datos de aprendizaje comprende datos que representan una o más regiones de una escena diferente.
6. El método según una cualquiera de las reivindicaciones precedentes, en que cada elemento muestral de aprendizaje comprende datos que definen un píxel en una escena respectiva.
7. El método según una cualquiera de las reivindicaciones precedentes, en que cada uno de los componentes en el modelo estadístico representa uno o más píxeles en la escena.
8. El método según una cualquiera de las reivindicaciones precedentes, en que cada una de dichas regiones comprende un píxel en la escena.
9. El método según cualquiera de las reivindicaciones 1 a 8, en que dicha medida de rareza comprende una estimación del parámetro v determinada separadamente para cada componente en cada aplicación del paso (iii) del método, en combinación con un valor para el número de componentes en el modelo estadístico, para pesar la contribución de los valores muestrales de aprendizaje a los estadísticos para cada componente respectivo.
10. El método según cualquiera de las reivindicaciones 1 a 9, en que dicha medida de rareza comprende una estimación de un valor común del parámetro v para todos los componentes, determinada en cada aplicación del paso (iii) del método, en combinación con un valor para el número de componentes en el modelo estadístico, para pesar la contribución de los valores muestrales de aprendizaje a los estadísticos para cada componente respectivo.
11. El método según la reivindicación 10, en que dicha estimación del parámetro v comprende fijar el valor del parámetro v de modo que sea proporcional al número de componentes en el modelo estadístico en cada aplicación del paso (iii) del método.
12. El método según una cualquiera de las reivindicaciones precedentes, que comprende además el paso de identificar una región de la escena que tenga una probabilidad de pertenencia a uno o más componentes del modelo estadístico que esté por debajo de un umbral predeterminado.

13. Un dispositivo de tratamiento de datos programado para implementar el método según una cualquiera de las reivindicaciones 1 a 12.
14. Un programa de ordenador, que cuando es cargado en un ordenador y es ejecutado, hace que el ordenador implemente el método según una cualquiera de las reivindicaciones 1 a 12.
- 5 15. Un producto de programa de ordenador, que comprende un medio legible por ordenador que lleva almacenado en él un medio de código de programa de ordenador que cuando es cargado en un ordenador, y es ejecutado, hace que el ordenador implemente el método según una cualquiera de las reivindicaciones 1 a 12.