

QÜESTIÓ, vol. 26, 1-2, p. 61-85, 2002

DETECCIÓN DE RASGOS EN IMÁGENES BINARIAS MEDIANTE PROCESOS PUNTUALES ESPACIALES MARCADOS

J. MATEU*

G. LORENZO**

Universitat Jaume I*

En este trabajo consideramos el problema de la detección de rasgos bajo la presencia de ruido en imágenes que tras un cierto tratamiento se reducen a binarias, por la presencia de dos tipos de elementos. Podemos encontrar ejemplos de este problema en la detección de minas por medio de imágenes de avión o satélite, en la búsqueda de rasgos en imágenes microscópicas de células, o en la caracterización de fallas en zonas de terremotos.

En primer lugar revisamos algunos métodos de detección jerárquicos basados en modelos probabilísticos, en los que los rasgos proceden de distribuciones normales multivariantes y el ruido surge según un proceso de Poisson espacial. Posteriormente, presentamos una nueva solución al problema mediante el uso de procesos puntuales espaciales marcados. Definimos un proceso puntual marcado en el que a cada localización se le asigna un par de marcas: las distancias al K -ésimo vecino más cercano y una variable dicotómica diferenciadora del rasgo frente al ruido. Estas distancias se modelizan como una mixtura de distribuciones cuyos parámetros se determinan mediante el algoritmo EM.

Finalmente, la nueva metodología es evaluada y contrastada sobre simulaciones y casos reales.

Detection of features in binary images by means of spatial marked point processes

Palabras clave: Algoritmo EM, mixturas de distribuciones, proceso de Poisson, procesos puntuales espaciales

Clasificación AMS (MSC 2000): 62M30, 60G55

* Autor para la correspondencia. Email: mateu@mat.uji.es. Fax: 964.728429

** Email: valentin@mat.uji.es

* Departamento de Matemáticas, Universitat Jaume I. Campus Riu Sec, E-12071 Castellón, Spain.

– Recibido en diciembre de 2000.

– Aceptado en febrero de 2002.

1. INTRODUCCIÓN

El análisis de imágenes difusas en las que bajo un cierto ruido se desarrolla algún fenómeno de interés es necesario en muchos contextos. Un ejemplo es el problema de detectar campos de minas (rasgos en general) en una superficie sobre la base de una imagen tomada por un avión de reconocimiento (Dasgupta & Raftery, 1998; Kaufman & Rousseeuw, 1990). Después de su procesamiento, tal imagen se reduce a un conjunto de objetos, algunos de los cuales serán minas y otros ruido (como por ejemplo, rocas u objetos metálicos). Los objetos identificados son pequeños y pueden ser representados por puntos sin pérdida de información relevante. La tarea del investigador será determinar si hay o no minas y dónde están localizadas (Cox, 1975). En este sentido partimos de la base de que las imágenes a analizar han sido procesadas de forma que nuestras unidades de trabajo consisten en localizaciones puntuales en una región del plano. Este problema puede verse también como el problema de buscar subregiones de mayor densidad dentro de una región donde se desarrolla un proceso puntual (Byers & Raftery, 1996, 1998). Un ejemplo típico se muestra en la Figura 1.

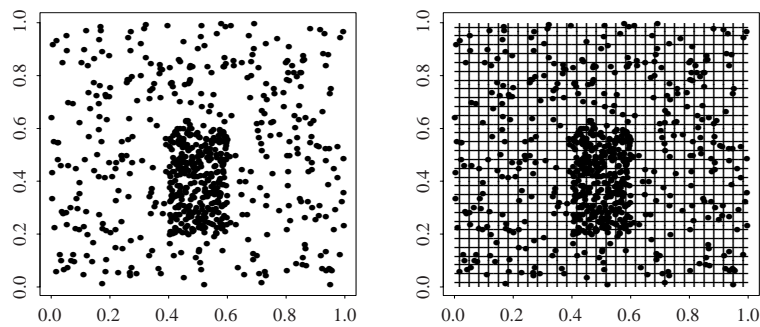


Figura 1. Datos simulados (ruido y rasgo) y retículo regular 30×30 sobre los datos.

Por lo dicho anteriormente, la teoría de procesos puntuales espaciales proporcionará el marco adecuado para la resolución de este problema. Un proceso puntual espacial es una colección de puntos distribuidos espacialmente en una región plana del espacio (Diggle, 1983). Los procesos puntuales constituyen una rama de la metodología estadística que es capaz de analizar dependencias espaciales entre las observaciones, que se suponen han sido generadas por algún mecanismo aleatorio y desconocido. Datos en la forma de localizaciones espaciales aparecen en muy diversos contextos en donde su análisis es esencial: localizaciones de árboles en bosques, centros de núcleos de células, epicentros de terremotos, localizaciones de ciudades, etc. Además, disponemos de un proceso, el proceso de Poisson, que sirve de marco con el que comparar cualquier tipo de estructura puntual. Este proceso se caracteriza por la ausencia de dependencia espa-

cial entre las observaciones y constituye el enlace entre la teoría de patrones espaciales y las técnicas usuales estadísticas.

En los últimos años, el problema de la detección de rasgos subyacentes en imágenes con ruido se ha centrado en el uso de métodos cluster basados en modelos probabilísticos (Banfield & Raftery, 1993; Dasgupta & Raftery, 1998; Fraley & Raftery, 1999) adaptando al caso de patrones puntuales espaciales algunas ideas surgidas en el contexto del análisis cluster. Aunque estos métodos funcionan adecuadamente en la práctica, muestran algunos inconvenientes: *a)* suelen ser matemáticamente bastante complicados; *b)* exigen suposiciones de Gaussianidad sobre los datos que no siempre son adecuadas; *c)* para su correcta aplicación necesitan la especificación de la forma y tamaño de los clusters.

Los inconvenientes antes mencionados motivan la propuesta de modelización que presentamos en este trabajo. Dado que las imágenes binarias pueden ser consideradas como un proceso puntual espacial bivalente, proponemos la definición de un proceso puntual marcado con dos tipos de marcas asociadas a cada posición espacial: una primera marca viene definida por las bien conocidas distancias al K -ésimo punto más cercano (Diggle, 1983; Collins, 1995) y una segunda marca que habrá que calcular y que identificará o clasificará la correspondiente localización espacial en rasgo o ruido. La novedad de esta propuesta no reside en el uso en sí de las distancias sino en la metodología que proponemos para la clasificación. Este método es totalmente original y lo más importante no hace ninguna suposición sobre la forma, número de rasgos o modelo probabilístico de los mismos.

El plan de este trabajo es el siguiente. En la sección 2 se definen los procesos puntuales espaciales. La sección 3 se centra en el método probabilístico, mientras que la nueva metodología propuesta viene en la sección 4. Finalmente, la sección 5 analiza datos simulados y reales.

2. PROCESOS PUNTUALES ESPACIALES

Un proceso puntual es un modelo estocástico dado por un conjunto de puntos $\{x_i\}_{i=1}^n$ en algún subconjunto $X \in \mathbb{R}^d$ (normalmente $d = 2$). Si las localizaciones contienen medidas o etiquetas, el proceso puntual será un proceso puntual marcado (Diggle, 1983; Penttinen, Stoyan & Henttonen, 1992; Cressie, 1993). Por ejemplo, sucesos pueden ser árboles en un bosque, ciudades en una región geográfica, o epicentros de terremotos. Las marcas correspondientes pueden ser, respectivamente, especies, tipos o diámetros de árboles, tamaños de ciudades o magnitudes de terremotos.

En general asumiremos que las localizaciones de los sucesos $\{x_i\}_{i=1}^n$ y sus marcas $\{Z(x_i)\}_{i=1}^n$ son realizaciones de algún proceso estocástico de la forma $\{Z(\mathbf{x}) : \mathbf{x} \in D\}$,

donde $\mathbf{x} = \{x_1, \dots, x_n\}$ y tanto $Z(\cdot)$ como D son aleatorios (Diggle, 1983; Cressie, 1993).

Un proceso puntual marcado se define de la siguiente forma. Sea (Ω, Λ, ρ) un espacio de medida, tal que $\rho(\Omega) < \infty$. Tomamos ahora $\Omega = X \times F$ con $X \subset \mathbb{R}^d$ y $F \subset \mathbb{R}$, $\Lambda = \chi \times \xi$, donde χ es la σ -álgebra de Borel de X y ξ son los conjuntos de Borel de F y, finalmente, $\rho = \nu \times \pi$, donde ν es una medida finita, no atómica en X (típicamente la medida Lebesgue) y π es una medida en F correspondiendo a la distribución de marcas. El espacio exponencial (Cressie, 1993) se denota por Ω_e y se equipa con la correspondiente σ -álgebra Λ_e . Un elemento $(\mathbf{x}, Z(\mathbf{x})) \in \Omega_e$ se toma como un patrón puntual espacial. Los procesos puntuales marcados se pueden también definir a través de medidas aleatorias (Cressie, 1993).

Dos conceptos importantes de los procesos puntuales son la estacionariedad y la isotropía. Un proceso puntual es *estacionario* si es estadísticamente invariante bajo traslaciones, es decir, $N = [\mathbf{x}, Z(\mathbf{x})]$ y $N_t = \{[\mathbf{x} + t; Z(\mathbf{x} + t)]\}$ tienen la misma distribución para todo t . Un proceso puntual es *isotrópico* si su distribución es invariante bajo rotaciones, es decir, N y $N_r = \{[r\mathbf{x}; Z(r\mathbf{x})]\}$ tienen la misma distribución para cada rotación r con respecto al origen. Si el proceso es a la vez estacionario e isotrópico se conoce como invariante frente al movimiento.

Un proceso puntual espacial de Poisson es aquel proceso en el que los puntos son independientes y localizados aleatoriamente en la región de estudio. Este proceso constituye el patrón básico de comparación (Cressie, 1993).

3. MÉTODOS BASADOS EN MODELOS PROBABILÍSTICOS

3.1. Introducción

Las técnicas de detección basadas en modelos probabilísticos se basan en la suposición de que los datos vienen generados por mixturas de distribuciones de probabilidad, distinguiendo, cuando sea posible, entre las distribuciones de los rasgos frente a las del ruido (McLachlan & Basford, 1988; Banfield & Raftery, 1993; Bensmail, Celeux, Raftery & Robert, 1994).

Estos métodos consideran a priori una información importante para el desarrollo del modelo puesto que determinará de un modo decisivo elementos como la cantidad, forma, tamaño y orientación de los clusters. Se trata de una colección de técnicas paramétricas en las que suponemos que los datos proceden de una distribución Normal Multivariante (NMV) o modelo Gaussiano. De los dos parámetros a considerar en dicho modelo (vector de medias y matriz de varianzas-covarianzas), tiene especial relevancia la descomposición espectral de la matriz de varianzas-covarianzas Σ . Esta matriz toma

diferentes expresiones dependiendo del cluster o grupo donde se defina. De esta forma, una reparametrización de esta matriz en el grupo o cluster k -ésimo toma la siguiente forma

$$(1) \quad \Sigma_k = \lambda_k D_k A_k D_k',$$

donde

1. λ_k denota el primer valor propio de la descomposición espectral de Σ_k , y controla el *volumen* de los rasgos (clusters),
2. D_k es la matriz de vectores propios de Σ_k , y controlará la *orientación* de los rasgos,
3. A_k es la matriz diagonal cuyos elementos son proporcionales a los valores propios, y controla la *forma* de los rasgos en el siguiente sentido: si $A_k = \text{diag} \{ \alpha_{1k}, \dots, \alpha_{pk} \}$ con $1 = \alpha_{1k} \geq \alpha_{2k} \geq \dots \geq \alpha_{pk} > 0$, entonces si los α_{jk} son de magnitud similar, el k -ésimo rasgo tenderá a ser esférico, mientras que si $\alpha_{2k} \ll 1$, estará concentrado sobre una línea; si $\alpha_{2k} \approx 1$ y $\alpha_{3k} \ll 1$, estará concentrado sobre un plano bidimensional en un espacio p -dimensional y así sucesivamente.

Las características de forma, volumen y orientación de las distribuciones son estimadas a partir de los datos y pueden variar o mantenerse fijas entre los rasgos. Así, por ejemplo, si $\Sigma_k = \sigma^2 I = \lambda I$ para todo k , siendo I la matriz identidad, el método lleva a detectar rasgos hiperesféricos con la misma varianza. Si $\Sigma_k = \Sigma$, para todo k , obtenemos el método de varianza constante (Wolfe, 1970; Fraley & Raftery, 1998).

3.2. Adaptación al caso de procesos puntuales espaciales

En este caso, el modelo considerado será un modelo Gaussiano para los rasgos y un modelo de Poisson para el ruido (Gower, 1967; Marriott, 1975; Banfield & Raftery, 1993). La detección de los rasgos se realiza en dos etapas:

1. En la primera etapa, mediante un modelo de cluster jerárquico se proporciona una primera estimación de los rasgos.

Para la primera etapa, partimos de la ya conocida descomposición espectral de la matriz Σ de varianzas-covarianzas, $\Sigma_k = \lambda_k D_k A_k D_k'$, pero adaptándola al caso que nos ocupa, es decir, con patrones puntuales espaciales (McLachlan & Basford, 1988). En este caso, la forma de los rasgos es la misma, pero su volumen y orientación son diferentes, es decir, $A_k = \text{diag} \{ 1, \alpha \}$, $k = \{ 1, \dots, G \}$ donde $\alpha < 1$ (ya que en un proceso puntual espacial hay dos dimensiones, $d = 2$) y G indica el número de rasgos a detectar. Además, $\alpha = \frac{\lambda_2}{\lambda_1}$, ($\lambda_1 > \lambda_2$). Cuando α es mucho más pequeño que uno, los rasgos que resultan tenderán a ser largos y estrechos, mientras que si se acerca a uno, los rasgos tenderán a ser de una forma más circular. Por esta razón α se llama el *parámetro forma*.

2. En la segunda etapa estos clusters-rasgos se refinan utilizando el algoritmo E. M. (Dempster *et al.*, 1977) como método para la obtención de la estimación máximo verosímil en presencia de datos faltantes. Supongamos una población con distribución una mezcla de densidades de la forma

$$(2) \quad f(X; \theta) = \sum_{k=1}^G \pi_k f_k(X; \theta)$$

donde:

(a) $f_k(X; \theta)$ se distribuye como una *NMV* (μ_k, Σ_k)

(b) $\sum_{k=1}^G \pi_k = 1$

Así, para n observaciones de esta mezcla de distribuciones, definimos un proceso puntual espacial marcado de la forma $pem_i = (x_i, z_i)$, donde $x_i = (x_{i1}, x_{i2})$ son las localizaciones espaciales originales y $z_i = (z_{i1}, \dots, z_{iG})$, con

$$z_{ik} = \begin{cases} 1 & \text{si la } i\text{-ésima observación está en el } k\text{-ésimo rasgo} \\ 0 & \text{en otro caso} \end{cases}$$

las marcas añadidas.

El vector de z_i sigue una distribución multinomial con parámetros $(1; \pi_1, \dots, \pi_G)$. Esto permite calcular la función de verosimilitud completa

$$(3) \quad L(pem; \theta) = \sum_{i=1}^n \sum_{k=1}^G z_{ik} \{ \log \pi_k + \log f_k(x_i; \theta) \}$$

En el primer paso del algoritmo EM (paso E) hay que calcular el estimador para las etiquetas que hemos puesto a cada observación $\hat{z}_{ik} = p_{ik} = E(z_{ik} | x_1, \dots, x_n; \theta)$, la cual es la probabilidad a posteriori de que x_i esté en el rasgo k -ésimo. En el paso siguiente (paso M) se obtienen los estimadores máximo verosímiles de θ y π . Para ello, se maximiza la versión estimada de (3) dada por

$$(4) \quad L^*(pem; \theta) = \sum_{i=1}^n \sum_{k=1}^G \hat{z}_{ik} \{ \log \pi_k + \log f_k(x_i; \theta) \}$$

El parámetro de forma α se estimará por máxima verosimilitud a partir del modelo de mezclas utilizado por el algoritmo EM. Supongamos que tenemos una mezcla de distribuciones Gaussianas, satisfaciendo $\Sigma_k = \lambda_k D_k A_k D_k'$, con $A_k = A = \text{diag}\{1, \alpha, \dots, \alpha\}$ ($k = 1, \dots, G$), y un ruido que se distribuye como un proceso de Poisson homogéneo.

Sea n_0 el número de puntos del ruido, d la dimensión de los datos (usualmente $d = 2$) y $\hat{\Sigma}_k$ la matriz de varianzas-covarianzas muestral para el rasgo k -ésimo y cuya descomposición espectral es $\hat{\Sigma}_k = L_k \Omega_k L_k'$. Si comenzamos a partir de la función de verosimilitud maximizada, asumiendo que α es conocida,

$$(5) \quad \begin{aligned} 2 \log L = & - (n - n_0) (d \log (2\pi) + d (1 - \log (d)) + \log (|A|)) \\ & - d \sum_{k=1}^G n_k \log \left(\frac{\text{tr}(\Omega_k A^{-1})}{n_k} \right) - 2n_0 \log (V) \end{aligned}$$

donde V es el volumen ocupado por los datos en R^d . Una buena definición de V viene dada por el volumen del rectángulo más pequeño cuyos lados son paralelos a los ejes coordenados, que contienen a los datos

$$(6) \quad V = \prod_{j=1}^d \left(\max_{i=1, \dots, n} \{x_{ij}\} - \min_{i=1, \dots, n} \{x_{ij}\} \right)$$

Finalmente, maximizando la probabilidad (5), con respecto a α , obtenemos

$$(7) \quad \sum_{k=1}^G \left[\frac{n_k}{n - n_0} \right] \frac{w_{k2} + \dots + w_{kd}}{\alpha w_{k1} + \dots + w_{kd}} = 1 - \frac{1}{d}$$

donde los w_{kj} representan los elementos de la matriz Ω_k .

La ecuación (7) es un polinomio en α de orden G con G raíces distintas como máximo. Sin embargo, sólo nos interesan las que se encuentran entre $[0, 1]$. En el caso particular en el que sólo hay un grupo ($G = 1$),

$$\hat{\alpha} = \frac{w_2 + \dots + w_d}{(d - 1) w_1}$$

Por lo general, (7) se resuelve mediante un sistema de búsqueda de posibles valores de α en entornos de retículos.

Finalmente, se puede determinar el número óptimo de rasgos representando cada elección como un modelo estadístico y comparándolos mediante el factor Bayes. Hay una serie de razones que lo avalan (contempla la distribución asintótica en modelos mixtos, considera modelos que pueden ser comparados por métodos frecuentistas, utiliza los p-valores) (Sokal & Michener, 1958). Además, gracias al algoritmo EM, podemos encontrar la función de verosimilitud maximizada de las mixturas y utilizar un método BIC (Bayesian Information Criterion) (Smith & Spiegelhalter, 1980; Kass & Raftery, 1995).

Como puede verse esta metodología, aunque se ha demostrado eficaz (Dasgupta & Raftery, 1998), no deja de tener un cierto fundamento matemáticamente complicado

con fuertes suposiciones sobre los modelos de probabilidad de los datos y no resulta sencilla llevarla a la práctica. Ciertamente es que con la ayuda del software MCLUST (Fraley & Raftery, 1999) la aplicación se hace más llevadera. Obsérvese que una colección de localizaciones espaciales en el plano equivale a disponer de dos variables sobre las que es susceptible la aplicación de las técnicas tradicionales de cluster jerárquico. Y esto es lo que se hace en una primera etapa. Posteriormente se refina el resultado obtenido mediante la reparametrización de la matriz de varianzas-covarianzas en términos del parámetro de forma α . Además, se aplica el algoritmo EM para mejorar la clasificación, la cual se lleva a cabo obteniendo los valores de la variable clasificadora z_{ik} .

4. MÉTODO DE DETECCIÓN BASADO EN DISTANCIAS AL K-ÉSIMO VECINO MÁS PRÓXIMO

En esta sección proponemos un nuevo método de detección de rasgos en imágenes con ruido basado en la construcción de procesos puntuales marcados donde las marcas provienen a partir del cálculo de cierto tipo de distancias entre los sucesos. Se trata de un método no paramétrico en que no hace falta establecer ninguna suposición sobre la Gaussianidad de los datos, forma o número de rasgos. En este sentido consideramos que se trata de un método de mayor aplicabilidad y sencillez.

La originalidad de esta propuesta no se basa en el uso de las distancias al K -ésimo vecino más próximo pues bien es sabido que ya fueron definidas por Diggle (1983) y posteriormente reanalizadas por Collins (1995), sino en el método de clasificación. La idea subyacente es formar un proceso puntual espacial marcado donde una de las marcas viene definida por este tipo de distancias (donde haremos uso de las propiedades que de ellas se derivan) y la otra marca actuará de datos faltantes a ser estimados para proveer la clasificación. Por tanto las dos secciones siguientes tratan estos dos puntos básicos.

4.1. Construcción del proceso puntual marcado

Asumiremos que el ruido está distribuido como un *proceso puntual de Poisson homogéneo* y el rasgo también está distribuido como un *proceso de Poisson restringido* a la imagen de definición del mismo, y que puede tener cualquier forma.

Construimos el siguiente proceso puntual marcado. Sea $N = \{x_i, y_i, d_i, \delta_i\}_{i=1}^n$ con (x_i, y_i) las localizaciones originales, d_i las distancias de cada localización a su K -ésimo vecino más cercano (Collins, 1995) y $\delta_i \in \{0, 1\}$. Además, $\delta_i = 1$, si el i -ésimo punto pertenece a un rasgo y $\delta_i = 0$ si se encuentra en el ruido.

Veamos algunas propiedades de este tipo de distancias. Como disponemos de un proceso de Poisson homogéneo, podemos encontrar la función de distribución de la distancia

al K -ésimo vecino más próximo, D_K , de un punto aleatorio en el proceso

$$(8) \quad P(D_K \geq d) = \sum_{k=0}^{K-1} \frac{e^{-\lambda\pi d^2} (-\lambda\pi d^2)^k}{k!} = 1 - F_{D_K}(d)$$

siendo λ la intensidad del proceso o número de puntos por unidad de área.

Esta expresión generaliza la obtenida por Diggle (1983) para distancias al primer vecino más próximo. Si D_K es mayor que d entonces debe ser uno de los $0, 1, \dots, K-1$ puntos en el círculo. Pero además, sabiendo que la función de densidad $f_{D_K}(d)$ viene dada por la derivada de F_{D_K} , podemos desarrollarla para obtener

$$\begin{aligned} f_{D_K}(d) &= \frac{\partial F_{D_K}(d)}{\partial d} = \frac{d \left(1 - \frac{e^{-\lambda\pi d^2} (-\lambda\pi d^2)^k}{k!} \right)}{dd} \\ &= - \sum_{k=0}^{K-1} \left[\frac{e^{-\lambda\pi d^2} (-2\lambda\pi d) (\lambda\pi d^2)^k}{k!} + \frac{e^{-\lambda\pi d^2} k (\lambda\pi d^2)^{k-1} (2\lambda\pi d)}{k!} \right] \\ &= - \sum_{k=0}^{K-1} \frac{e^{-\lambda\pi d^2} 2\lambda^k \pi^k d^{2k-1} (-\lambda\pi d^2 + k)}{k!} \\ &= e^{-\lambda\pi d^2} 2 \sum_{k=0}^{K-1} \frac{(\lambda\pi)^k d^{2k-1} (\lambda\pi d^2 - k)}{k!} \\ &= e^{-\lambda\pi d^2} 2 \left[\sum_{k=0}^{K-1} \frac{(\lambda\pi)^{k+1} d^{2k+1}}{k!} - \sum_{k=0}^{K-1} \frac{(\lambda\pi)^k d^{2k-1} k}{k!} \right] \\ &= e^{-\lambda\pi d^2} 2 \left[\sum_{k=0}^{K-1} \frac{(\lambda\pi d^2)^k \lambda\pi d}{k!} - \sum_{k=0}^{K-1} \frac{(\lambda\pi d^2)^k d^{-1} k}{k!} \right] \\ &= e^{-\lambda\pi d^2} 2 \left[\lambda\pi d \sum_{k=0}^{K-1} \frac{(\lambda\pi d^2)^k}{k!} - (\lambda\pi d^2) \sum_{k=0}^{K-1} \frac{(\lambda\pi d^2)^{k-1}}{(k-1)!} \right] \\ (9) \quad &= \frac{e^{-\lambda\pi d^2} 2 (\lambda\pi)^K d (d^2)^{K-1}}{(K-1)!} = \frac{e^{-\lambda\pi d^2} 2 (\lambda\pi)^K d^{2K-1}}{(K-1)!} \end{aligned}$$

La función de densidad anterior (9) corresponde a la función de densidad de una variable aleatoria con una distribución Gamma transformada, $Y \sim \Gamma(K, \lambda\pi)$ donde $Y = (D_K)^2$. Concretamente es un ejemplo de una *Distribución Gamma Generalizada* (Stacy, 1962).

Además, dada esta estructura, encontrar el estimador máximo verosímil de la razón λ es relativamente sencillo. Para el conjunto de observaciones de distancias d_i , la función de verosimilitud toma la forma

$$L(d_i; \lambda) = \prod_{i=1}^n f_{D_K}(d_i; \lambda) = \frac{e^{-\lambda \pi \sum_{i=1}^n d_i^2} [2(\lambda \pi)^K]^n \prod_{i=1}^n d_i^{2K-1}}{(K-1)!}$$

Tomando logaritmos y derivando obtenemos

$$\begin{aligned} \ln L(d_i; \lambda) &= -\lambda \pi \sum_{i=1}^n d_i^2 + n [\ln 2 + K [\ln \lambda + \ln \pi]] + \ln \left[\prod_{i=1}^n d_i^{2K-1} \right] \\ &\quad - \ln((K-1)!) \end{aligned}$$

$$\frac{d \ln L(d_i; \lambda)}{d \lambda} = -\pi \sum_{i=1}^n d_i^2 + \frac{nK}{\lambda},$$

de donde podemos obtener el estimador máximo verosímil para la intensidad λ , por medio de la expresión

$$(10) \quad \hat{\lambda} = \frac{nK}{\pi \sum_{i=1}^n d_i^2}.$$

4.2. Determinación de los parámetros de clasificación. Algoritmo EM

Teniendo en cuenta que el modelo más simple está formado por un rasgo y por un ruido (el resto de observaciones), ambos distribuidos como dos procesos de Poisson homogéneos espaciales, podemos decir que la distribución de D_K es una mezcla de las dos distribuciones Gamma Generalizadas anteriormente comentadas

$$(11) \quad D_K \simeq p \Gamma^{(1/2)}(K, \lambda_1 \pi) + (1-p) \Gamma^{(1/2)}(K, \lambda_2 \pi)$$

siendo $\Gamma^{(1/2)}(a, b)$ la distribución de la Gamma Generalizada dada en (9).

El objetivo ahora es la obtención de los valores de las marcas de clasificación δ_i , $\forall i \in \{1, \dots, n\}$ y también de los parámetros λ_1 , λ_2 y p . Para ello haremos uso del algoritmo EM (Dempster *et al.*, 1977; Celeux & Govaert, 1992; Hathaway, 1986). Los dos pasos de este algoritmo son:

1. **Paso E.** En este paso obtendremos el valor esperado de las variables δ_i , $E(\hat{\delta}_i(t+1))$, a partir de los valores estimados de $p(t)$, $\lambda_1(t)$, $\lambda_2(t)$ en la iteración anterior. Se

trata de encontrar la función $Q(p, \lambda_1, \lambda_2 | X, d) = E(\ln(L(p, \lambda_1, \lambda_2 | d_i, \delta_i)))$, que para la localización i -ésima y la intensidad λ_1 toma la forma

$$\begin{aligned} Q_i(p, \lambda_1, \lambda_2 | d_i, \delta_i) &= \frac{p(d_i, \delta_i | \lambda_1)}{p(d_i, \delta_i | \lambda_1)p(\lambda_1) + p(d_i, \delta_i | \lambda_2)p(\lambda_2)} \\ (12) \qquad \qquad \qquad &= \frac{\hat{p}(t) f_{D_k}(d_i; \hat{\lambda}_1(t))}{\hat{p}(t) f_{D_k}(d_i; \hat{\lambda}_1(t)) + (1 - \hat{p}(t)) f_{D_k}(d_i; \hat{\lambda}_2(t))} \end{aligned}$$

2. **Paso M.** En este paso se obtienen los estimadores para λ_1 y λ_2 que en la iteración posterior se utilizarán para realizar el paso E. La probabilidad p se puede estimar por

$$(13) \qquad \qquad \qquad \hat{p}(t+1) = \sum_{i=1}^n \frac{\hat{\delta}_i(t+1)}{n}.$$

Teniendo en cuenta que la función de verosimilitud viene dada por

$$L = \prod_{i=1}^n f_{D_K}(d_i; \lambda),$$

para λ_1 tenemos que

$$\begin{aligned} L &= \prod_{i=1, \delta_i=1}^n f_{D_K}(d_i; \lambda) = \prod_{i=1, \delta_i=1}^n \frac{e^{-\lambda_1 \pi d_i^2} 2 (\lambda_1 \pi)^K d_i^{2K-1}}{(K-1)!} \\ &= \frac{e^{\sum_{i=1, \delta_i=1}^n \lambda_1 \pi d_i^2} 2^{\sum_{i=1, \delta_i=1}^n \delta_i} (\lambda_1 \pi)^{\sum_{i=1, \delta_i=1}^n \lambda_1 \pi d_i^2} \prod_{i=1, \delta_i=1}^n d_i^{2K-1} \delta_i}{[(K-1)!]^{\sum_{i=1, \delta_i=1}^n \delta_i}}, \end{aligned}$$

con logaritmo

$$\begin{aligned} \ln L &= \sum_{i=1}^n \lambda_1 \pi d_i^2 \delta_i + \sum_{i=1}^n \delta_i (\ln 2) + k \left(\sum_{i=1}^n \delta_i \right) \ln (\lambda_1 \pi) \\ &\quad + \sum_{i=1}^n \ln (d_i^{2K-1} \delta_i) - \sum_{i=1}^n \delta_i \ln (K-1)! \end{aligned}$$

Finalmente, derivando obtenemos

$$\begin{aligned} \frac{\partial \ln L}{\partial \lambda_1} &= \frac{\partial}{\partial \lambda_1} \left[-\lambda_1 \pi \sum_{i=1}^n d_i^2 \delta_i + K \left(\sum_{i=1}^n \delta_i \right) [\ln \lambda_1 + \ln \pi] \right] \\ &= -\pi \sum_{i=1}^n d_i^2 \delta_i + K \left(\sum_{i=1}^n \delta_i \right) \frac{1}{\lambda_1}, \end{aligned}$$

de donde se obtiene el estimador para λ_1 (escrito ya en notación iterativa)

$$(14) \quad \hat{\lambda}_1(t+1) = \frac{K \sum_{i=1}^n \hat{\delta}_i(t+1)}{\pi \sum_{i=1}^n d_i^2 \hat{\delta}_i(t+1)}.$$

Analogamente, el estimador para λ_2 viene dado por

$$(15) \quad \hat{\lambda}_2(t+1) = \frac{K \sum_{i=1}^n (1 - \hat{\delta}_i(t+1))}{\pi \sum_{i=1}^n d_i^2 (1 - \hat{\delta}_i(t+1))}.$$

Finalmente, el método de clasificación consistirá en clasificar una observación como rasgo si su δ_i final se encuentra más cerca del 1 y como ruido en otro caso.

5. SIMULACIONES Y APLICACIONES

En esta sección ilustramos la metodología presentada en este trabajo sobre unos datos simulados y reales. El objetivo es mostrar que la propuesta de clasificación para la detección de rasgos frente a ruido basada en la construcción de procesos marcados con marcas definidas a partir de un cierto tipo de distancias funciona correctamente en la práctica. Por tanto constituirá una buena alternativa al uso de las técnicas de cluster jerárquico. Quede claro que, como bien mostramos en primer lugar en esta sección, las técnicas de cluster jerárquico funcionan correctamente en la práctica (ver Figura 2) a costa de la fuerte exigencia sobre la distribución de los datos y no se trata tanto de comparar ambas metodologías como de mostrar que nuestra alternativa produce al menos los mismos resultados bajo un menor costo de exigencia probabilística y también computacional.

El primer conjunto de datos que se utiliza (Figura 1) consta de 700 puntos simulados, 300 de los cuales son rasgo y corresponderían a un proceso aleatorio en la región $[0.4, 0.6] \times [0.2, 0.6]$, mientras que los 400 puntos restantes son ruido, también definidos como un proceso aleatorio, pero esta vez en toda la región $[0, 1] \times [0, 1]$. En esta misma figura, en la segunda gráfica aparece un retículo regular de 900 celdas (30×30). Esta malla colocada sobre los datos nos permitirá calcular posteriores distancias.

Antes de analizar el comportamiento de nuestra metodología, vamos a clasificar y separar el rasgo del ruido por medio de técnicas jerárquicas basadas en modelos probabilísticos. Para ello hacemos uso del software MCLUST (Fraley & Raftery, 1999). Los resultados de una primera aplicación de este software vienen representados en la Figura 2. Observamos que los puntos representados por círculos se encuentran aglutinados en la zona donde se encuentra el rasgo; el ruido viene representado por las localizaciones en forma de cuadrados. En general, la clasificación obtenida por este método se ajusta

adecuadamente a los datos simulados. Además, un análisis de los valores BIC confirmó que el número óptimo de rasgos a detectar era 1. Tal y como hemos comentado en la sección 3, este procedimiento funciona bien pero tiene fuertes exigencias sobre la distribución de probabilidad de los datos y su uso es complicado en la práctica.

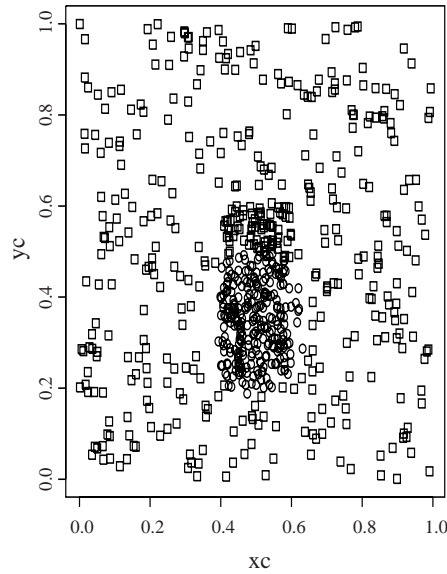


Figura 2. Método jerárquico probabilístico sobre los datos simulados.

En segundo lugar, y como motivación para el uso de nuestra metodología, aplicamos sobre los datos simulados diferentes tipos de distancias, las definidas al K -ésimo vecino más cercano (D_K) y otras definidas desde un retículo regular impuesto sobre los datos (ver Figura 1) al K -ésimo vecino de entre las observaciones reales más cercano (DG_K) (Diggle, 1983; Collins, 1995). En este caso nuestro único objetivo es ver si estas distancias pueden detectar de forma intuitiva diferencias entre rasgo y ruido.

En la Figura 3 se muestran histogramas de frecuencias absolutas para las distancias D_K para distintos valores de K . Se observa una clara bimodalidad al aumentar el valor de K (es decir, al aumentar el número del vecino más próximo que se toma de referencia para calcular la distancia). Las distancias alrededor de la primera moda (distancias más pequeñas), corresponden a los valores de las distancias de los puntos del rasgo, mientras que la segunda moda corresponde a los valores de las distancias de los puntos del ruido. De esta forma las distancias la K -ésimo vecino más cercano pueden ser utilizadas para discriminar entre rasgo y ruido.

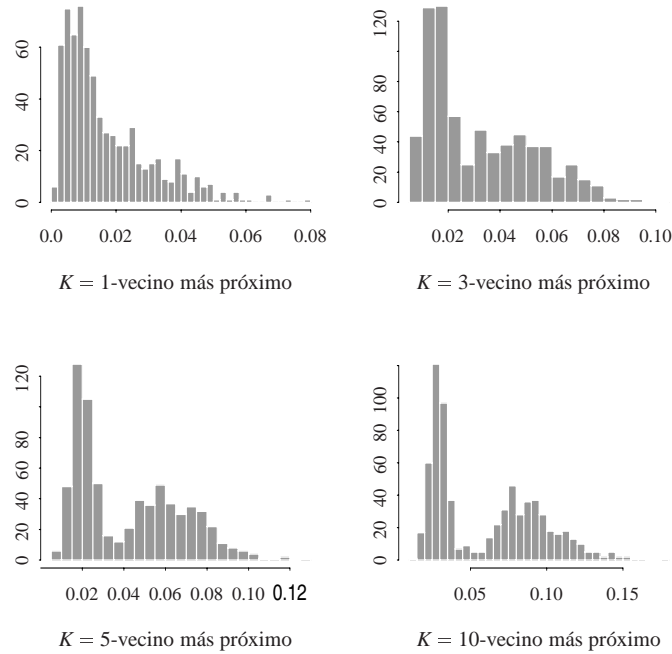


Figura 3. Histogramas para las distancias D_K con $K = 1, 3, 5, 10$ para los datos simulados.

Si evaluamos la distancia DG_K desde cada punto de un retículo o malla regular al K -ésimo punto más cercano de entre los originales encontramos de nuevo claras diferencias entre las distancias de punto de retículo a localización de ruido o rasgo (ver Tablas 1 y 2 para diferentes valores de K).

Tabla 1. Medidas descriptivas de las distancias DG_K para el rasgo de los datos simulados.

DG_K (Rasgo)	Min	Q_{25}	Mediana	Media	Q_{75}	Max
$K = 1$	0.001071	0.005459	0.008011	0.008237	0.010858	0.01823
$K = 3$	0.005813	0.01134	0.01338	0.01434	0.01675	0.02709
$K = 5$	0.008987	0.01613	0.01923	0.01908	0.02138	0.03088
$K = 10$	0.01917	0.02605	0.0281	0.02824	0.03102	0.04037

Tabla 2. Medidas descriptivas de las distancias DG_K para el ruido de los datos simulados.

DG_K (Ruido)	Min	Q_{25}	Mediana	Media	Q_{75}	Max
$K = 1$	0.0008557	0.01575	0.0234	0.02528	0.03301	0.07655
$K = 3$	0.008291	0.03625	0.04668	0.04744	0.05698	0.104
$K = 5$	0.02069	0.05135	0.06182	0.06286	0.07294	0.1329
$K = 10$	0.03391	0.07633	0.08909	0.0903	0.1019	0.1731

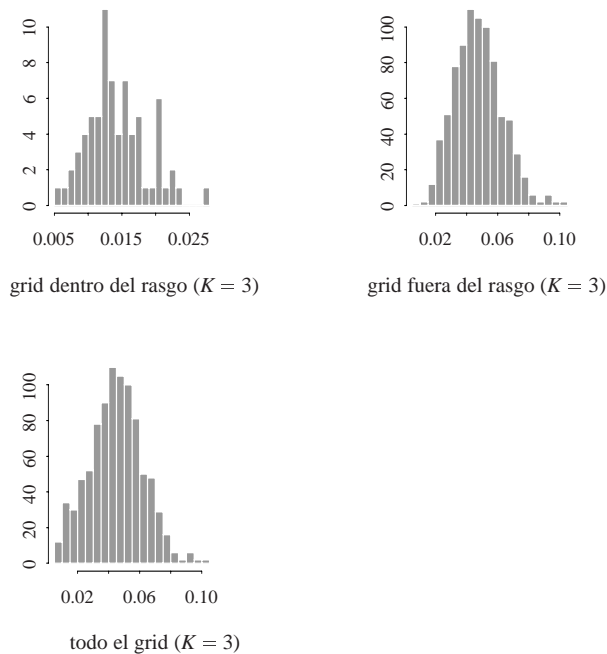


Figura 4. Histogramas para las distancias DG_K con $K = 3$ para los datos simulados.

La Figura 4 representa histogramas referentes a las distancias DG_K con $K = 3$. Notar que el tercer histograma es el resumen conjunto de los datos anteriores y para el caso $K = 3$ no se aprecia tanto la bimodalidad (que representaría las diferencias entre rasgo y ruido). Sin embargo, analizándolos por separado si vemos que el primer histograma (rasgo) muestra distancias 0.005 a 0.025, mientras que en el segundo el rango de distancias es de 0.02 a 0.1. En la Figura 5 repetimos el proceso anterior pero para $K = 10$, y en este caso, en la tercera gráfica sí aparece una bimodalidad clara. Por tanto, mayores valores de K proporcionan diferencias más claras entre rasgo y ruido.

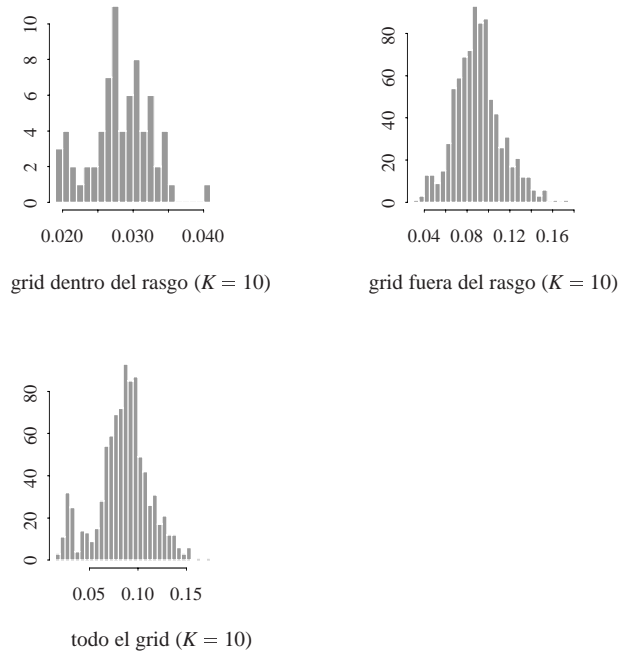


Figura 5. Histogramas para las distancias DG_K con $K = 10$ para los datos simulados.

Visto pues que el uso de cierto tipo de distancias puede discriminar entre rasgo y ruido, analizamos finalmente nuestros datos simulados con nuestra propuesta basada en la definición de un proceso puntual espacial marcado con unas marcas basadas en las distancias al K -ésimo vecino más cercano y otras marcas utilizadas como criterio de clasificación y obtenidas por el algoritmo EM. Para ello utilizamos diferentes valores de K , $K = 5, 10, 15$, y los resultados de clasificación se muestran en las Figuras 6, 7 y 8. Es claro que un aumento del valor de K va asociado con una mejor clasificación. En cada gráfica se presentan los histogramas de las distancias D_K , así como un mapa de probabilidad de pertenencia al rasgo. Vemos que los resultados demuestran la utilidad de nuestra metodología en el objetivo de la detección de rasgos.

Finalmente, para acabar con nuestro estudio de detección de rasgos, analizamos unos datos reales consistentes en la detección de zonas cancerígenas (zonas de una mayor acumulación de cuerpos) en una célula de un tejido biológico. El objetivo consiste en detectar y separar los dos rasgos del ruido existente en la imagen (ver Figura 9). Para ello hemos aplicado de nuevo nuestra técnica. Los resultados de clasificación para $K = 5, 10, 15$ se muestran en las Figuras 9, 10 y 11. De nuevo, es claro que un aumento del valor de K va asociado con una mejor clasificación.

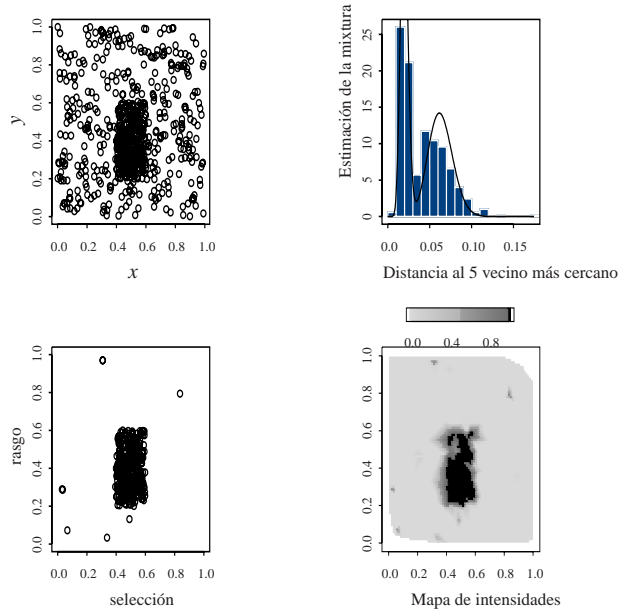


Figura 6. Detección de un rasgo con distancias al 5º vecino más próximo.

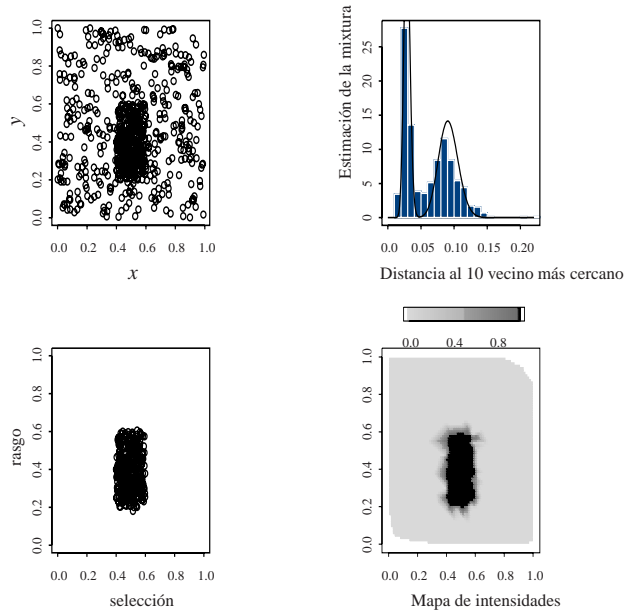


Figura 7. Detección de un rasgo con distancias al 10º vecino más próximo.

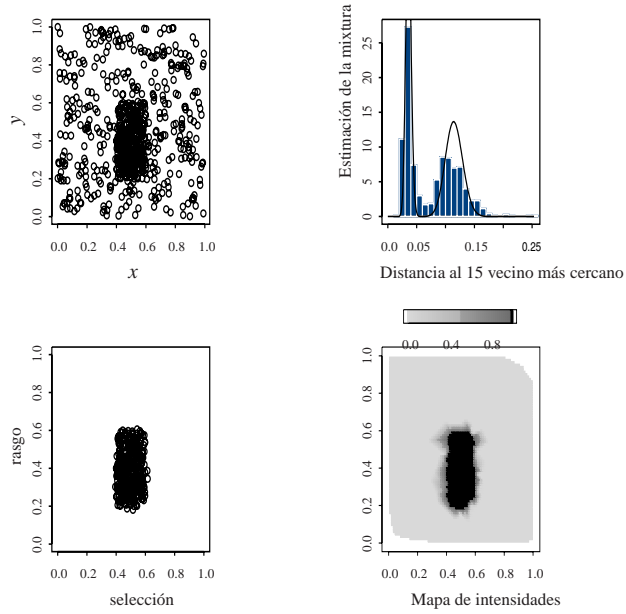


Figura 8. Detección de un rasgo con distancias al 15º vecino más próximo.

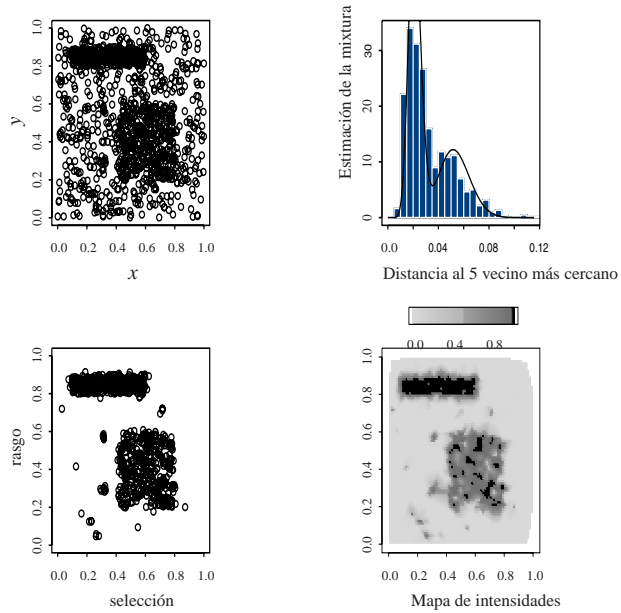


Figura 9. Detección de dos rasgos en células enfermas con distancias al 5º vecino más próximo.

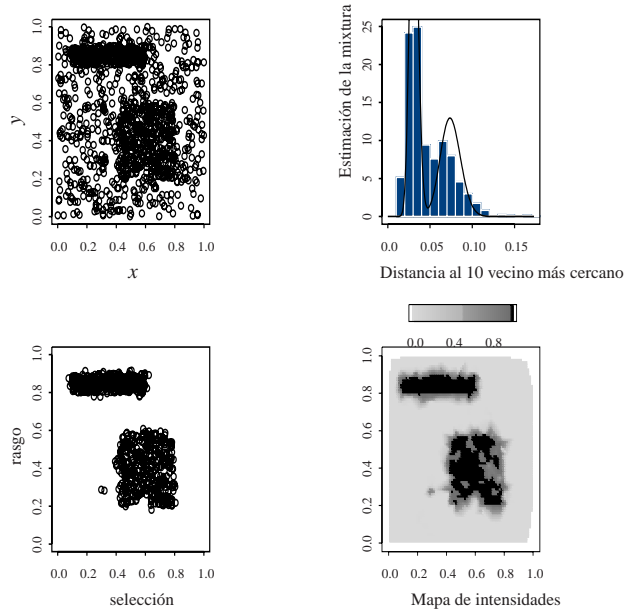


Figura 10. Detección de dos rasgos en células enfermas con distancias al 10º vecino más próximo.

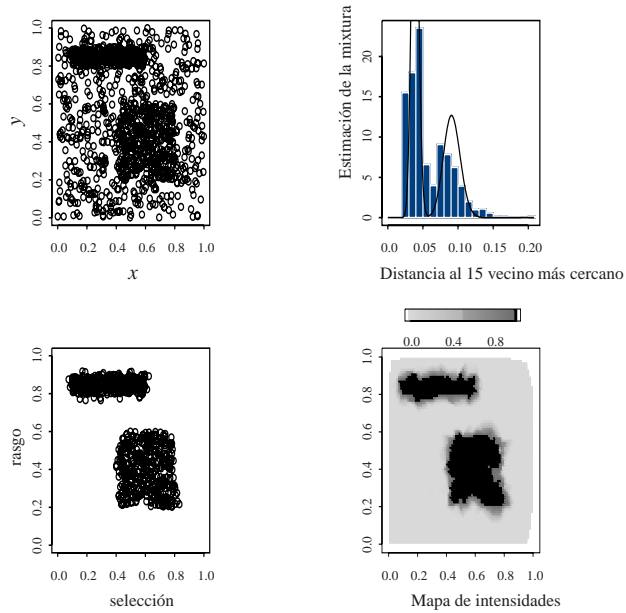


Figura 11. Detección de dos rasgos en células enfermas con distancias al 15º vecino más próximo.

6. CONCLUSIONES Y DISCUSIÓN

En este trabajo se han presentado dos métodos complementarios sobre detección de rasgos subyacentes a ruido en imágenes usualmente digitalizadas. El primero de ellos, conocido y presentado en los últimos años, consiste en una adaptación al caso espacial de técnicas de cluster jerárquico. El segundo explota las propiedades de cierto tipo de distancias entre sucesos en el contexto de los procesos espaciales marcados para obtener una clasificación a través del algoritmo EM. Ambos métodos discriminan correctamente entre rasgo y ruido, si bien el segundo de ellos es más sencillo de llevar a la práctica y más robusto en el sentido de que no hace ningún tipo de suposición sobre los modelos probabilísticos de los datos. Además, el método de distancias clasifica, en general para valores altos de K , mejor que los modelos jerárquicos.

Por otra parte, las expresiones utilizadas en el algoritmo EM se restringen a sólo dos intensidades diferentes (λ_1, λ_2). Una buena generalización de este método sería el incorporar la posibilidad de detección de todo un rango de posibles intensidades. En este caso la distribución de las distancias D_K sería la siguiente:

$$(16) \quad D_K \sim \sum_{j=1}^s p_j \Gamma\left(\frac{1}{2}\right) (K, \lambda_j \pi),$$

donde

$$\sum_{j=1}^s p_j = 1$$

En este caso, el paso E del algoritmo EM vendría dado por

$$(17) \quad E\left(\widehat{\delta}_{ij}(t+1)\right) = \frac{\widehat{p}_j(t) f_{D_s}\left(d_i; \widehat{\lambda}_j(t)\right)}{\sum_{j=1}^s \widehat{p}_j(t) f_{D_s}\left(d_i; \widehat{\lambda}_j(t)\right)}, \quad \sum_{j=1}^s \widehat{p}_j = 1 \quad i = 1, \dots, n.$$

Y el paso M por

$$(18) \quad \widehat{\lambda}_j(t+1) = \frac{s \sum_{i=1}^n \widehat{\delta}_{ij}(t+1)}{\pi \sum_{i=1}^n d_i^2 \widehat{\delta}_{ij}(t+1)} \quad j = 1, \dots, s$$

con

$$(19) \quad \widehat{p}_j(t+1) = \frac{\sum_{i=1}^n \widehat{\delta}_{ij}(t+1)}{n},$$

siendo

$$\sum_{j=1}^s \widehat{\delta}_{ij}(t+1) = 1.$$

Actualmente los autores están trabajando sobre el uso de funciones LISA, funciones basadas en densidades producto de procesos puntuales, para proporcionar nuevos métodos de clasificación.

AGRADECIMIENTOS

Los autores quieren agradecer a los referees las sugerencias recibidas sobre una versión anterior del trabajo.

REFERENCIAS

- Banfield, J. D. & Raftery, A. E. (1993). «Model-based Gaussian and non-Gaussian clustering». *Biometrics*, 49, 803-849.
- Bensmail, H., Celeux, G., Raftery, A. E. & Robert, C. P. (1997). «Inference in model-based cluster analysis». *Statistics and Computing*, 7, 1-10.
- Byers, S. & Raftery, A. E. (1997). *Bayesian estimation and segmentation of spatial point processes using Voronoi tilings*, Technical Report, University of Washington, 32.
- Byers, S. & Raftery, A. E. (1998). «Nearest neighbor clutter removal for estimating features in spatial point processes». *Journal of the American Statistical Association*, 93, 577-584.
- Celeux, G. & Govaert, G. (1992). «A classification EM algorithm for clustering and two stochastic versions». *Computational Statistics and Data Analysis*, 14, 315-332.
- Collins, L. B. (1995). *Inter-event distance methods for the statistical analysis of spatial point processes*. Ph. D. Dissertation, Department of Statistics, University of Chicago, Chicago, Illinois.
- Cox, D. R. (1975). «Partial likelihood». *Biometrika*, 62, 269-276.
- Cressie, N. A. C. (1993). *Statistics for spatial data*, Wiley-Interscience.
- Dasgupta, A. & Raftery, A. E. (1998). «Detecting features in spatial point processes with clutter via model-based clustering». *Journal of the American Statistical Association*, 93, 294-302.
- Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). «Maximum likelihood for incomplete data via the EM algorithm (with discussion)». *Journal of the Royal Statistical Society*, B, 39, 1-38.
- Diggle, P. J. (1983). *Statistical analysis of spatial point patterns*, Academic Press.
- Fraley, C. & Raftery, A. E. (1998). «How many clusters? Which clustering method? - Answers via model-based cluster analysis». *The Computer Journal*, 41, 578-588.
- Fraley, C. & Raftery, A. E. (1999). «MCLUST: software for model-based cluster and discriminant analysis». *Technical report*, 342, Department of Statistics, University of Washington.

- Gower, J. C. (1967). «A comparison of some methods of cluster analysis». *Biometrics*, 23, 623-637.
- Hathaway, R. J. (1986). «Another interpretation of the EM algorithm for mixture distributions». *Journal of Statistics & Probability Letters*, 4, 53-56.
- Kass, R. E. & Raftery, A. E. (1995). «Bayes Factors». *Journal of the American Statistical Association*, 90, 773-795.
- Kaufman, L. & Rousseeuw, P. J. (1990). *Finding groups in data. An introduction to cluster analysis*. John Wiley & Sons.
- Marriott, F. H. C. (1975). «Separating mixtures of normal distributions». *Biometrics*, 31, 767-769.
- McLachlan, G. J. & Basford, K. E. (1988). *Mixture models: inference and applications to clustering*. Marcel Dekker.
- Penttinen, A. K., Stoyan, D. & Henttonen, H. M. (1992). «Marked point processes in forest statistics». *Forest Science*, 38, 806-824.
- Smith, A. & Spiegelhalter, D. (1980). «Bayes factors and choice criteria for linear models». *Journal of the Royal Statistical Society, A*, 42, 213-220.
- Stacy, E. W. (1962). «A generalization of the gamma distribution». *Annals of Mathematical Statistics*, 33, 1187-1192.
- Wolfe, J. H. (1970). «Pattern clustering by multivariate mixture analysis». *Multivariate Behavioral Research*, 5, 329-350.

ENGLISH SUMMARY

DETECTION OF FEATURES IN BINARY IMAGES BY MEANS OF SPATIAL MARKED POINT PROCESSES

J. MATEU*

G. LORENZO**

Universitat Jaume I*

In this paper the features detection in binary images under the presence of clutter is considered. Practical examples of this problem can be found in the mine detection through satellite images, in the detection of features in biological cell images or in the detection of faults in earthquake problems. We first revise detection methods based on probabilistic models, in which features are normally distributed, and the clutter comes from a spatial Poisson process. Then, we show a new solution to the problem in hand by means of spatial marked point processes. The marked processes are defined in terms of two variables: distances to the k th-nearest neighbour and a Bernoulli variable to define feature and clutter. The distances are modelled through a mixture of probability distributions whose parameters are obtained by the EM algorithm. Finally, the new methodology is evaluated and tested over simulations and real-case studies.

Keywords: EM algorithm, mixtures, Poisson process, spatial point processes

AMS Classification (MSC 2000): 62M30, 60G55

* Autor para la correspondencia. Email: mateu@mat.uji.es. Fax: 964.728429

** Email: valentin@mat.uji.es

* Departamento de Matemáticas, Universitat Jaume I. Campus Riu Sec, E-12071 Castellón, Spain.

– Received December 2000.

– Accepted February 2002.

The analysis of noisy images in which features are present under substantive clutter is a key tool in many different contexts. After being processed, the image reduces to a set of objects (points), some of them will identify features and others will stand for clutter. So the final image consists of a collection of planar points located in a region of interest. Consequently, the theory of spatial point processes can handle this kind of problem. A spatial point process is a set of points spatially distributed over the plane that show certain types of interaction. A basic and benchmark process is the Poisson process, for which the points are randomly distributed.

In the last years, the problem of feature detection in presence of clutter has focused on probability-based cluster methods. However, these methods are rather complicated, they need the data to be Normally distributed and need the shape and size parameters be specified. Consequently, in this paper we propose a different approach based on spatial marked processes. Each location is marked with two variables, the K -th nearest-neighbour distances and a binary variable identifying the type of point (feature or clutter). This method puts no hypothesis in the probability distribution of the data and has no size or shape parameters to define.

Methods based on probabilistic models

These techniques are parametric and are based on the hypothesis that the data comes from a mixture of probability distributions. They consider a priori information determining parameters such as shape, size and orientation of the clusters. If the data is Gaussian, then the variance-covariance matrix can be decomposed using the spectral decomposition. This decomposition defines the basic tool to proceed as gives light to elements that characterize the cluster.

Method based on K th nearest-neighbour distances

Now, we assume that both, features and clutter come from a Poisson spatial process and define the following spatial process. Let $N = \{x_i, y_i, d_i, \delta_i\}_{i=1}^n$ with (x_i, y_i) the original spatial locations, d_i the K th nearest-neighbour distances and $\delta_i \in \{0, 1\}$. Moreover, $\delta_i = 1$, if the i th point belongs to a feature and $\delta_i = 0$ if it is clutter. Now, the probability distribution function of the K th nearest-neighbour distance, D_K , is given by

$$(1) \quad P(D_K \geq d) = \sum_{k=0}^{K-1} \frac{e^{-\lambda\pi d^2} (-\lambda\pi d^2)^k}{k!} = 1 - F_{D_K}(d)$$

where λ is the intensity of the process.

The corresponding probability density function, $f_{D_K}(d)$, corresponds to a Generalized Gamma distribution where $Y = (D_K)^2$ and $Y \sim \Gamma(K, \lambda\pi)$. Then, the likelihood function takes the form

$$L(d_i; \lambda) = \prod_{i=1}^n f_{D_K}(d_i; \lambda) = \frac{e^{-\lambda\pi\sum_{i=1}^n d_i^2} [2(\lambda\pi)^K]^n \prod_{i=1}^n d_i^{2K-1}}{(K-1)!}$$

So, our basic starting point is the following:

$$(2) \quad D_K \simeq p\Gamma^{(1/2)}(K, \lambda_1\pi) + (1-p)\Gamma^{(1/2)}(K, \lambda_2\pi)$$

and the aim is to estimate the classification marks $\delta_i, \forall i \in \{1, \dots, n\}$ together with the intensity parameters λ_1, λ_2 and p . For this we use the EM algorithm.