

Estimación de la Complejidad Local en Imágenes de Texturas

*Agustin Mailing^{†‡}, Tomás Crivelli[†] y Bruno Cernuschi-Frías^{†‡}

[†]Facultad de Ingeniería, Universidad de Buenos Aires, Argentina

[‡]IAM-CONICET, Buenos Aires, Argentina

Resumen Cuando se utilizan modelos de diferentes órdenes para el modelado de datos, normalmente, es el modelo más complejo quien ajustará mejor pues presenta mayor flexibilidad. No obstante, en ocasiones los datos pueden ser descriptos de igual manera por modelos de distinto orden. En este caso decimos que la complejidad de los datos (de acuerdo a los modelos de referencia) está dada por el modelo más sencillo que lo describe. En este trabajo se presenta un método para determinar esta complejidad localmente en una imagen (de texturas) de manera sencilla dentro de un esquema de modelos encajados. Si bien la estructura de los modelos utilizados es simple, permitiendo su clara y fácil implementación, el método es lo suficientemente flexible y general para ser utilizado con estructuras y modelos muy complejos.

Key words: Complejidad, Texturas, Procesos Aleatorios, Procesamiento de imágenes

1. Introducción

En la gran mayoría, por no decir prácticamente todos los algoritmos y métodos que involucran imágenes está presente en forma explícita o no, la complejidad de los datos que el tipo de imágenes transporta. Esto es la *Información*. Si se pretende comprimir una imagen su complejidad es quien determinará cuál es la tasa de compresión máxima que se obtendrá sin pérdida o cuánta información se perderá (en algunos casos para siempre) según la compresión pretendida.

En algunos casos la información que transporta una imagen es determinística: por ejemplo en una típica foto; y en otros la información esta dada por la estadística de los datos: una imagen del movimiento de gente en una calle concurrida o un cine, o una imagen satelital de un cielo nublado. En este trabajo, es esta información estadística la que permite determinar qué regiones (de aquí el término *local*) contienen más información respecto del modelo seleccionado. Esto permitiría por ejemplo determinar qué región es la que contiene mayor información para ser procesada, enviada y/o almacenada.

* Trabajo realizado gracias al apoyo de la Universidad de Buenos Aires, CONICET, Argentina, y la Agencia Nacional de Promoción Científica y Tecnológica, PICT No. 209-2006, Argentina

Uno de los problemas centrales en el análisis de la información es la elección del orden de un modelo que se ajuste a los datos observados. En el contexto de un enfoque paramétrico, esto se relaciona al número de parámetros del modelo. Supongamos que los datos corresponden a una imagen, la cual se quiere clasificar entre diversas clases, cada una de las cuales está descrita por un modelo paramétrico de diferente orden. En general, el modelo de mayor orden se ajustará mejor a los datos con un menor error, independientemente de la clase a la que pertenecen.

Consideremos el siguiente ejemplo en una dimensión. Queremos decidir si un conjunto de datos con ruido fue generado por un polinomio de primer orden o un polinomio de segundo orden. Ajustamos, mediante una simple regresión, los dos parámetros del primer modelo (una recta) y los tres parámetros del segundo (una parábola), y decidimos por aquel que da un menor error cuadrático. Se puede ver que esta regla de decisión siempre elegirá el modelo cuadrático (el modelo de mayor orden) ya que en el peor de los casos ajusta con igual error que la recta. El resultado sería un error de clasificación sistemático.

El problema mencionado se relaciona al concepto de minimizar la complejidad del modelo [1], esto es, encontrar el modelo más simple que describa correctamente las observaciones. Uno de los enfoques existentes es el principio *Minimum Description Length (MDL)* [1], el cual considera los modelos como códigos capaces de comprimir la información y propone penalizar la longitud de cada código, dando lugar a una estrategia de selección del modelo. Otro de los enfoques es el Criterio de Información de Akaike, el cual, a pesar de ser ampliamente utilizado, resulta inconsistente [2].

En este contexto se propone un clasificador Bayesiano óptimo que tiene en cuenta la complejidad de los modelos involucrados, en la formulación de la regla de decisión entre un conjunto de clases. Las ideas presentadas en este trabajo extienden los resultados de [3] y [2], al caso de distribuciones de Gibbs para variables aleatorias no independientes, y en particular, el análisis de un modelo de campo aleatorio de Markov Gaussiano [4]. En nuestro enfoque puramente Bayesiano, la matriz de información de Fisher [5] juega un papel fundamental. Vale mencionar que de hecho existe una relación estrecha entre el principio MDL y la inferencia Bayesiana [6].

Se utiliza aquí la regla de decisión óptima, obtenida en el trabajo previo [7], que incorpora un término de penalización para el número de parámetros usados para describir un conjunto de datos. Este se obtiene directamente del clasificador Bayesiano clásico y aparece naturalmente al comparar modelos de clase de diferente tamaño. A diferencia de [7], se utiliza aquí un esquema de conjuntos encajados, lo que permite utilizar la invarianza del modelo respecto de la escala y obtener así a distintos niveles la complejidad estadística local más conveniente para modelar los datos.

2. Complejidad Bayesiana

Sean $\omega_1, \omega_2, \dots, \omega_c$ un conjunto de c clases. Sea $\mathbf{X} = \{x_p\}_{p:1..n}$ una imagen de n pixels. La regla de deción Bayesiana, para probabilidades a priori de clase iguales, es [8]:

$$\mathbf{X} \in \omega_k \quad \text{si} \quad p(\mathbf{X} | \omega_k) \geq p(\mathbf{X} | \omega_l) \quad \forall l \neq k. \quad (1)$$

Ahora asumamos que para cada clase ω_k tenemos un modelo definido por un vector de parámetros $\boldsymbol{\alpha}^{(k)} = \{\alpha_1^{(k)}, \alpha_2^{(k)}, \dots, \alpha_{d_k}^{(k)}\}$ de dimensión d_k . Luego, dentro de la teoría Bayesiana podemos escribir la verosimilitud

$$p(\mathbf{X} | \omega_k) = \int_{\boldsymbol{\alpha}^{(k)}} p(\mathbf{X} | \boldsymbol{\alpha}^{(k)}, \omega_k) p(\boldsymbol{\alpha}^{(k)} | \omega_k) d\boldsymbol{\alpha}^{(k)}. \quad (2)$$

Sea $\hat{\boldsymbol{\alpha}}^{(k)}$ el estimado de máxima verosimilitud (ML) de $\boldsymbol{\alpha}^{(k)}$ para la clase ω_k usando n puntos. A continuación desarrollamos la log-verosimilitud

$$L_k = \log p(\mathbf{X} | \boldsymbol{\alpha}^{(k)}, \omega_k) \equiv \log p(\mathbf{X} | \boldsymbol{\alpha}^{(k)}) \quad (3)$$

en una expansión en serie de Taylor alrededor de $\hat{\boldsymbol{\alpha}}^{(k)}$:

$$\begin{aligned} \log p(\mathbf{X} | \boldsymbol{\alpha}^{(k)}) &= \log p(\mathbf{X} | \hat{\boldsymbol{\alpha}}^{(k)}) \\ &\quad - \frac{1}{2} (\boldsymbol{\alpha}^{(k)} - \hat{\boldsymbol{\alpha}}^{(k)})^T \mathbf{H}(\hat{\boldsymbol{\alpha}}^{(k)}) (\boldsymbol{\alpha}^{(k)} - \hat{\boldsymbol{\alpha}}^{(k)}) + \dots, \end{aligned} \quad (4)$$

donde $\mathbf{H}(\hat{\boldsymbol{\alpha}}^{(k)}) = \{H_{ij}\}$ es el negativo del Hessiano de dimensión $d_k \times d_k$ con respecto a $\boldsymbol{\alpha}^{(k)}$,

$$H_{ij} = - \frac{\partial^2}{\partial \alpha_i^{(k)} \partial \alpha_j^{(k)}} \log p(\mathbf{X} | \boldsymbol{\alpha}^{(k)}) |_{\hat{\boldsymbol{\alpha}}^{(k)}}. \quad (5)$$

El término de primer orden desaparece ya que el gradiente de L_k evaluado en el estimador ML es nulo por definición. Los términos de orden superior son descartados para un tamaño de imagen n suficientemente grande, y en consecuencia, $p(\mathbf{X} | \boldsymbol{\alpha}^{(k)})$ resulta asintóticamente Gaussiano, siempre que el modelo esté correctamente especificado y $\mathbf{H}(\cdot)$ sea asintóticamente positiva definida [2]. Luego,

$$p(\mathbf{X} | \boldsymbol{\alpha}^{(k)}) \approx p(\mathbf{X} | \hat{\boldsymbol{\alpha}}^{(k)}) e^{-\frac{1}{2} (\boldsymbol{\alpha}^{(k)} - \hat{\boldsymbol{\alpha}}^{(k)})^T \mathbf{H}(\hat{\boldsymbol{\alpha}}^{(k)}) (\boldsymbol{\alpha}^{(k)} - \hat{\boldsymbol{\alpha}}^{(k)})} \quad (6)$$

Por otro lado, ya que con $n \rightarrow \infty$ el estimador ML converge al vector de parámetros verdadero (ver [3]), la verosimilitud se concentra alrededor de $\hat{\boldsymbol{\alpha}}^{(k)}$ y

$$p(\mathbf{X} | \boldsymbol{\alpha}^{(k)}) \approx p(\mathbf{X} | \hat{\boldsymbol{\alpha}}^{(k)}) (2\pi)^{d_k/2} \det \{\mathbf{H}\}^{-1/2} \delta(\boldsymbol{\alpha}^{(k)} - \hat{\boldsymbol{\alpha}}^{(k)}), \quad (7)$$

donde δ es la delta de Dirac. Reemplazando en (2) e integrando,

$$L_k \approx \log p(\mathbf{X} | \hat{\boldsymbol{\alpha}}^{(k)}) - \frac{d_k}{2} \log n / 2\pi - \frac{1}{2} \log \det \frac{\mathbf{H}}{n}. \quad (8)$$

donde hemos usado la identidad $\det \mathbf{H} = \det \left\{ \frac{n\mathbf{H}}{n} \right\} = n^{d_k} \det \left\{ \frac{\mathbf{H}}{n} \right\}$ y $\log p(\hat{\boldsymbol{\alpha}}^{(k)} | \omega_k)$ se desprecia con respecto a los demás términos para $n \rightarrow \infty$.

Se observa que además de la función de log-verosimilitud, aparecen términos adicionales que penalizan la dimensión d_k del modelo, en este caso, el número de parámetros.

2.1. Caso particular: distribuciones de Gibbs

Ahora consideramos el caso correspondiente a las distribuciones de Gibbs o, en forma equivalente, a los Campos Aleatorios de Markov (MRF) [9,4], donde $p(\mathbf{X}; \boldsymbol{\alpha}) = \exp Q(\mathbf{X}; \boldsymbol{\alpha}) / Z(\boldsymbol{\alpha})$. Primero suponemos que $Q(\cdot)$ es una función lineal de los parámetros y esto implica que $\frac{\partial^2}{\partial \alpha_i \partial \alpha_j} Q(\cdot) = 0$. Esta restricción al modelo se cumple en la mayoría de las situaciones de interés práctico. Por ejemplo, el MRF Gaussiano, los auto-modelos [9] y, en general, modelos de Gibbs definidos por funciones potenciales de la forma $V(\cdot; \alpha) = \alpha f(\cdot)$. Omitiendo el índice de clase k por simplicidad, tenemos

$$H_{ij} = -\frac{\partial^2}{\partial \alpha_i \partial \alpha_j} \log p(\mathbf{X} | \boldsymbol{\alpha})|_{\hat{\boldsymbol{\alpha}}} = \frac{\partial^2}{\partial \alpha_i \partial \alpha_j} \log Z(\boldsymbol{\alpha})|_{\hat{\boldsymbol{\alpha}}}. \quad (9)$$

Se puede observar que \mathbf{H} se aproxima a la matriz de información de Fisher [5], cuyo elemento ij th es $-E \left[\frac{\partial^2}{\partial \alpha_i \partial \alpha_j} \log p(\mathbf{X} | \boldsymbol{\alpha}) \right] \Big|_{\boldsymbol{\alpha}^{real}}$, dado que $\hat{\boldsymbol{\alpha}} \rightarrow \boldsymbol{\alpha}^{real}$. No es casualidad que esta cantidad tan relacionada con la teoría de la información y la entropía relativa aparezca naturalmente en el clasificador Bayesiano propuesto como un término de complejidad, haciendo evidente la conexión con otros enfoques como el principio MDL [1]. Finalmente se puede mostrar que,

$$H_{ij} = E \left[\frac{\partial}{\partial \alpha_i} Q(\mathbf{X}; \boldsymbol{\alpha}) \frac{\partial}{\partial \alpha_j} Q(\mathbf{X}; \boldsymbol{\alpha}) \right] - E \left[\frac{\partial}{\partial \alpha_i} Q(\mathbf{X}; \boldsymbol{\alpha}) \right] E \left[\frac{\partial}{\partial \alpha_j} Q(\mathbf{X}; \boldsymbol{\alpha}) \right] \Big|_{\hat{\boldsymbol{\alpha}}} \quad (10)$$

A continuación aplicamos este resultado con el objeto de obtener la log-verosimilitud (8), para un modelo de campo aleatorio de Markov Gaussiano. Consideremos entonces un modelo de Gibbs Gaussiano definido por:

$$Q(\mathbf{X}; \boldsymbol{\alpha}) = \sum_p ax_p - bx_p^2 + \sum_{\{p,q\}} h_{p,q}x_px_q, \quad (11)$$

donde los pares $\{p, q\}$ son puntos vecinos de la imagen. Aquí vamos a considerar vecindades de 8 puntos y vamos a considerar un MRF estacionario con cuatro direcciones de interacción. Luego, $h_{p,q} \in \{h_{vert}, h_{horiz}, h_{diag}, h_{anti-diag}\}$ y tenemos $d = 6$ parámetros.

Al evaluar $\log \det \left\{ \frac{\mathbf{H}}{n} \right\}$ observamos su comportamiento asintótico con $n \rightarrow \infty$. Los elementos de la matriz \mathbf{H} pueden computarse usando la ecuación (10). En general, cada elemento resulta de orden $O(n)$, por ejemplo $H_{aa} = n \sum_q \sigma_{1q}$, $H_{ab} = 2n\mu \sum_q \sigma_{1q}$, $H_{ah_{horiz}} = n\mu \sum_{\{p,q\}} \sigma_{1p} + \sigma_{1q}$, etc. σ_{pq} es el coeficiente de correlación entre x_p y x_q , $\mu = E[x_p] \forall p$, y $\{p, q\}$ son pares de vecinos horizontales en este caso. Cada suma recorre los coeficientes de correlación, lo cual converge a una constante independiente de n , dado que para una distribución de Gibbs Gaussiana la correlación entre dos puntos de la imagen decae exponencialmente con la distancia entre ambos. Luego $\log \det \left\{ \frac{\mathbf{H}}{n} \right\}$ es de orden $O(1)$ haciéndose despreciable comparado con $(d_k/2) \log n / 2\pi$. En futuras investigaciones se hará un análisis exhaustivo del error cometido en esta aproximación. De hecho, en [2] se remarca que no siempre es posible despreciar el término en cuestión. Finalmente de (8) obtenemos:

$$L_k \approx \log p(\mathbf{X} | \hat{\boldsymbol{\alpha}}_k) - (d_k/2) \log n / 2\pi. \quad (12)$$

Con la ecuación (12) podemos computar L_k para el caso de una distribución de Gibbs Gaussiana. Incluso la misma aproximación es válida para distintos tamaños de vecindad, mientras esta sea pequeña comparada con el tamaño de la imagen. Se puede observar también que para modelos con el mismo número de parámetros, el término $(d_k/2) \log n / 2\pi$ no tiene ningún efecto, y el clasificador propuesto resulta en el clásico test de verosimilitud [8]. De aquí en más llamamos a este término, *término de complejidad*.

3. Determinación de la complejidad-local en texturas Gaussianas

El término de complejidad penaliza el número de parámetros utilizados para describir el conjunto de datos. Supongamos entonces que tenemos una ventana W con n puntos y la particionamos en N_k sub-imágenes disjuntas $\{A_1^{(k)}, \dots, A_m^{(k)}, \dots, A_{N_k}^{(k)}\}$ según alguna de las configuraciones k del conjunto de configuraciones elegido. Cada sub-imagen $A_m^{(k)}$ contiene $n_m^{(k)}$ puntos (con $n = \sum_m n_m^{(k)}$) y es modelada con una distribución de Gibbs Gaussiana con d parámetros $\boldsymbol{\alpha}_m^{(k)}$.

El objetivo es definir un clasificador Bayesiano que pueda determinar cual es la configuración de ventanas más probable. Para computar la log-verosimilitud para cada clase, cada sub-imagen dentro de una configuración es considerada independiente de las otras y en consecuencia podemos escribir

$$L_k = \sum_{m=1}^{N_k} \{ \log p(A_m^{(k)} | \hat{\alpha}_m^{(k)}) - (d/2) \log n_m^{(k)} / 2\pi \} \quad (13)$$

Aún cuando cada modelo Gaussiano ajustado a cada sub-imagen es del mismo orden d , el número diferente de particiones N_k y su diferente tamaño $n_m^{(k)}$ es lo que difiere entre clases y lo que hace al término de complejidad cumplir un papel importante. Cada configuración queda descripta por un conjunto de $N_k d$ parámetros, los cuales son estimados con un número diferente de puntos. En definitiva, la regla de decisión tiene en cuenta la complejidad de la configuración para una ventana, lo cual esta directamente relacionado a la forma de la partición. Con este esquema, el clasificador es capaz de decidir si dentro de un ventana aparecen una o varias texturas.

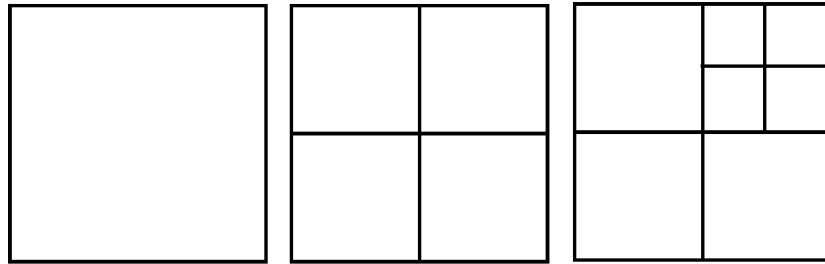


Figura 1. Esquema de los modelos utilizados para determinar la complejidad de la imagen. Hacia la derecha mayor complejidad (1,4 y 7).

Como las sub-imágenes son tomadas independientes, el proceso puede aplicarse del mismo modo a cada sub-imagen. De este modo el proceso puede aplicarse recursivamente mientras la cantidad de puntos sea suficiente. Se construye así un conjunto de configuraciones (modelos) que se encuentran encajados (ver esquema en Fig. 1). Desde el punto de vista de la implementación resulta de interés que cada la decisión se hace sobre cada sub-imagen por lo que no es necesario estimar la complejidad de cada modelo y tomar la decisión sobre todas las configuraciones posibles.

Así, dado un modelo estadístico (en este caso un MRF) y un número de particiones K (4 en Fig. 1), la complejidad de cada sub-imagen puede tomarse como: 1 si mediante (13) se determina que toda la sub-imagen puede modelarse con una sola clase (d parámetros), o K si es mejor particionarla ($d \times K$ parámetros).

En este último caso, cada una de las K sub-imágenes pasa por el mismo proceso resultando un método recursivo.

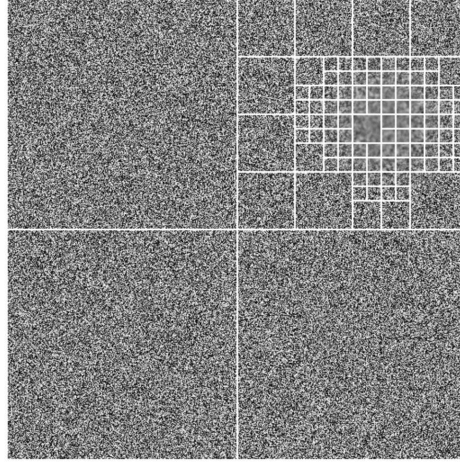


Figura 2. En la textura se introduce una *mancha* pequeña.

3.1. Aplicación a texturas Gaussianas

Utilizando el modelo MRF Gaussiano completo desarrollado en 2.1 con $d = 6$ y utilizando la aproximación aplicada en la ecuación (13). Los parámetros son estimados con un estimador de pseudo-máxima-verosimilitud [9] dado que el estimador ML clásico no es aplicable debido a la compleja forma de la función de partición $Z(\alpha_m^{(k)}) = \int_{\Lambda_m^{(k)}} \exp Q(\Lambda_m^{(k)}; \alpha_m^{(k)}) d\alpha_m^{(k)}$. Para calcular $\log p(\Lambda_m^{(k)} | \hat{\alpha}_m^{(k)}) = Q(\Lambda_m^{(k)}; \hat{\alpha}_m^{(k)}) - \log Z(\hat{\alpha}_m^{(k)})$ el valor de $\log Z(\hat{\alpha}_m^{(k)})$ puede aproximarse por simulación numérica [10] o bien por integración de la expansión en serie de $\exp Q(\Lambda_m^{(k)}; \hat{\alpha}_m^{(k)})$. Aquí utilizamos este último método. En la figura 4 se muestra el resultado de la aplicación de este método.

4. Conclusiones

En el trabajo se muestra una aplicación donde la complejidad de los datos (imágenes de texturas) respecto de una estructura de modelos estadísticos es utilizada para construir un mapa de complejidad de la imagen. Para esto se utiliza una penalización sobre la dimensión del modelo *no impuesta artificialmente*. A partir de esta base se utiliza un conjunto de modelos encajados que permite una

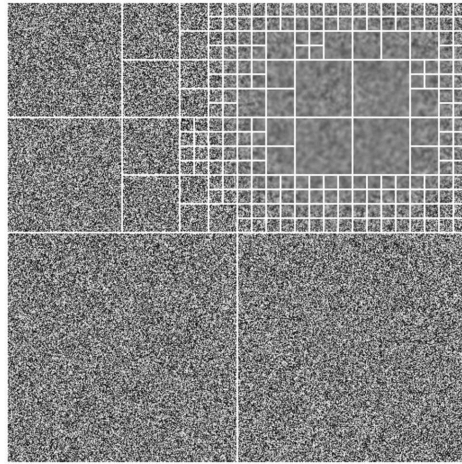


Figura 3. Cuando la *mancha* es más grande pero la estadística es igual la complejidad se concentra en los bordes.

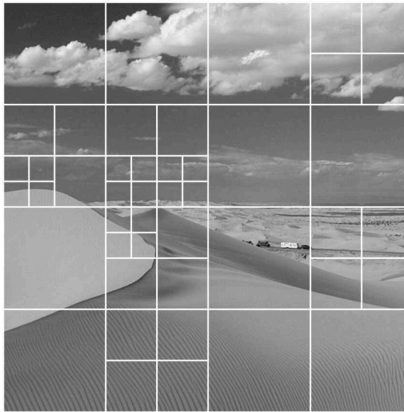


Figura 4. Imagen de mayor complejidad estadística usando el método de 3.1

implementación sencilla y reduce el costo computacional: puesto que la decisión sobre el modelo se realiza en cada etapa(escala).

Por otro lado, como puede observarse en la comparación entre las figuras 2, 3 y 4 la complejidad se concentra donde la estadística varía con lo cual se podría asociar esta complejidad a la detección de una *zona de borde estadístico* entre las distintas texturas. Esto asumiendo por su puesto, que el modelo es capaz de modelar los datos y que no nos encontramos en un caso de *sub-modelado*.

Teniendo en cuenta los resultados obtenidos favorables a partir de un modelo estadístico y modelo de partición muy sencillos (aplicados a datos estadísticamente sencillos también) parece evidente su extensión a modelos más complejos y por tanto de mayor aplicación.

Referencias

1. Rissanen, J.: Information and Complexity in Statistical Modeling. Springer Publishing Company, Incorporated (2007)
2. Kashyap, R.: Optimal choice of AR and MA parts in autoregressive moving average models. IEEE Trans. on Pattern Analysis and Machine Intelligence **4**(2) (Mar 1982) 99–104
3. Bolle, R., Cooper, D.: Bayesian recognition of local 3D shape by approximating image intensity functions with quadric polynomials. IEEE Trans. on Pattern Analysis and Machine Intelligence **6**(4) (Jul 1984) 418–429
4. Kindermann, R., Snell, J.: Markov random fields and their applications. American Mathematical Society, Providence, Rhode Island (1980)
5. Cover, T., Thomas, J.: Elements of Information Theory. Wiley-Interscience (2006)
6. Rissanen, J.: Fisher information and stochastic complexity. IEEE Trans. on Information Theory **42**(1) (Jan 1996) 40–47
7. Crivelli, T., Mailing, A., Cernuschi-Frias, B.: Complexity-based border detection for textured images. In: Acoustics, Speech and Signal Processing, IEEE International Conference on. (2010)
8. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification (2nd Edition). Wiley-Interscience (2000)
9. Besag, J.: Spatial interaction and the statistical analysis of lattice systems. Journal of the Royal Statistical Society. Series B **36** (1974) 192–236
10. Potamianos, G., Goutsias, J.: Stochastic approximation algorithms for partition function estimation of gibbs random fields. IEEE Trans. on Information Theory **43**(6) (Nov 1997) 1948–1965
11. Paget, R., Longstaff, I.: Texture synthesis via a noncausal nonparametric multiscale markov random field. Image Processing, IEEE Transactions on **7**(6) (Jun 1998) 925–931
12. Zhu, S.C., Yuille, A.: Region competition: unifying snakes, region growing, and bayes/mdl for multiband image segmentation. Pattern Analysis and Machine Intelligence, IEEE Transactions on **18**(9) (Sep 1996) 884–900
13. Crivelli, T., Cernuschi, B., Bouthemy, P., Yao, J.: Mixed-state Markov random fields for motion texture modeling and segmentation. In: Proc. of IEEE International Conference on Image Processing, ICIP'06. (Atlanta, 2006) 1857–1860