

Reconocimiento de Patrones

Proyecto final

Detección de pólipos en Colonoscopia Virtual

Marcelo Fiori

IIE - Facultad de Ingeniería - UdelaR

Marzo/Abril - 2010

Agenda

- 1 **Introducción**
- 2 **El problema**
- 3 **Características seleccionadas**
- 4 **Disparidad entre clases**
- 5 **Clasificación y Resultados**
- 6 **Trabajo posterior**

Agenda

- 1 **Introducción**
- 2 El problema
- 3 Características seleccionadas
- 4 Disparidad entre clases
- 5 Clasificación y Resultados
- 6 Trabajo posterior

Cáncer colorectal

- Una de las mayores causas de muerte por cáncer en el mundo.
- Detección temprana de pólipos permite alcanzar tasas del 90 % de curabilidad.
- La técnica habitual es la colonoscopia óptica
 - elevada performance
 - invasiva y extremadamente cara (imposible utilización en campañas de *screening*)

Colonoscopia Virtual (VC)

- Reconstrucción 3D del colon a partir de cortes de tomografía computada (no invasiva y relativamente barata)
- Para poder utilizarla en *screening* es necesario que los procedimientos sean automáticos
- Actualmente la cantidad de falsos positivos y falsos negativos producida por éstos métodos está muy por encima de los máximos aceptados en la práctica médica

Agenda

- 1 Introducción
- 2 El problema**
- 3 Características seleccionadas
- 4 Disparidad entre clases
- 5 Clasificación y Resultados
- 6 Trabajo posterior

Objetivo

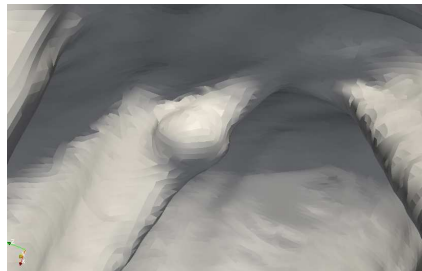
Detectar pólipos en VC

- Particularmente los menores a $10mm$ de diámetro.
- Tipos de lesiones:
 - Pediculado (tallo de implantación)
 - Sésil (base amplia de implantación)
 - Plano

Objetivo

Detectar pólipos en VC

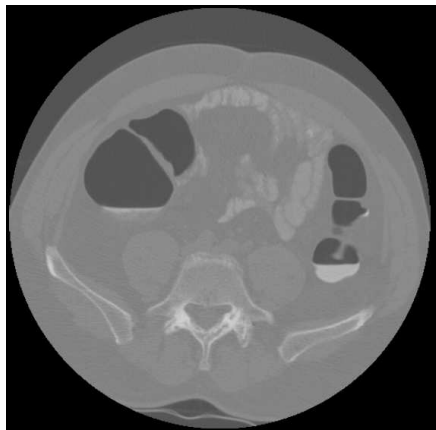
- Particularmente los menores a $10mm$ de diámetro.
- Tipos de lesiones:
 - Pediculado (tallo de implantación)
 - Sésil (base amplia de implantación)
 - Plano



Datos

Imágenes CT

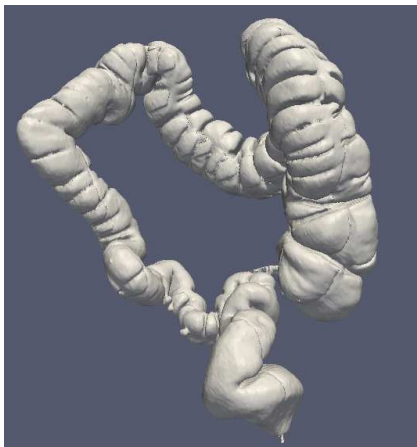
- Cortes de 512×512 pixels. Resolución x, y : $0,625\text{mm}/\text{pixel}$
- Separación entre cortes (resolución z): 1mm



Segmentación

Segmentación

Reconstrucción de superficie del colon a partir de cortes CT



Datos

Casos disponibles

- 10 casos segmentados
- 17 pólipos en total
 - 1 flat polyp
 - tamaños entre $4mm$ y $13mm$.

Agenda

- 1 Introducción
- 2 El problema
- 3 Características seleccionadas**
- 4 Disparidad entre clases
- 5 Clasificación y Resultados
- 6 Trabajo posterior

Características Geométricas

Curvatura

- Buscamos zonas de geometría particular

Shape Index y Curvedness

- Introducidos por Koenderink en 1990
- Transformación: $(\kappa_1, \kappa_2) \rightarrow (S, C)$

Características Geométricas

Curvatura

- Buscamos zonas de geometría particular

Shape Index y Curvedness

- Introducidos por Koenderink en 1990
- Transformación: $(\kappa_1, \kappa_2) \rightarrow (S, C)$

Características Geométricas

Curvatura

- Buscamos zonas de geometría particular

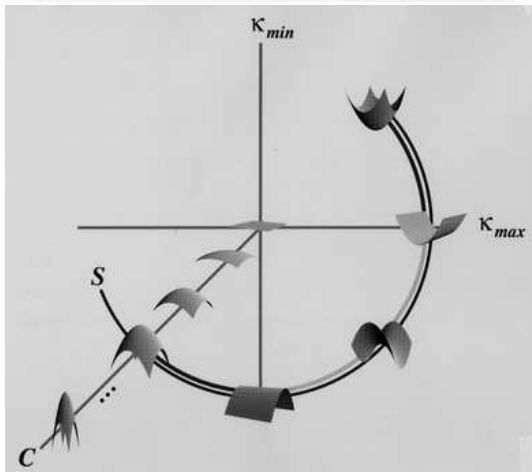
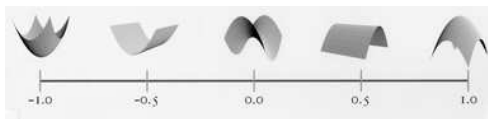
Shape Index y Curvedness

- Introducidos por Koenderink en 1990
- Transformación: $(\kappa_1, \kappa_2) \rightarrow (S, C)$

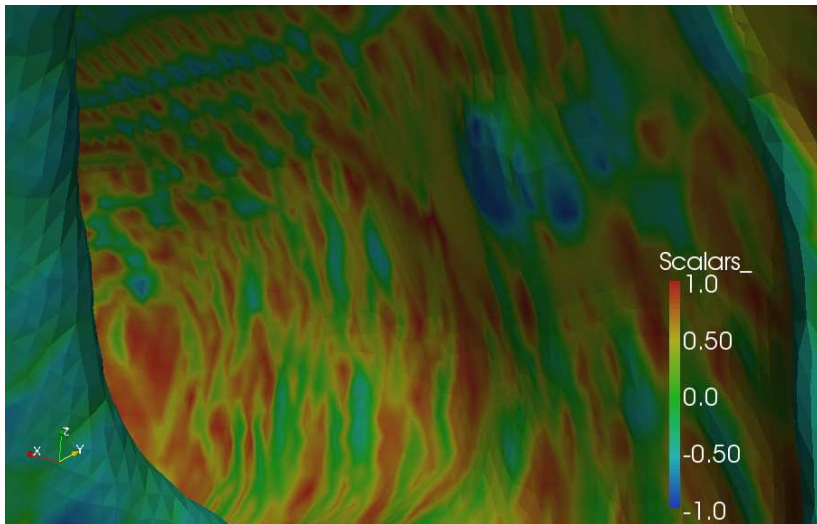
$$S = -\frac{2}{\pi} \arctan \frac{\kappa_{max} + \kappa_{min}}{\kappa_{max} - \kappa_{min}}$$

$$R = \sqrt{\frac{\kappa_{max}^2 + \kappa_{min}^2}{2}} \quad C = \frac{2}{\pi} \ln R$$

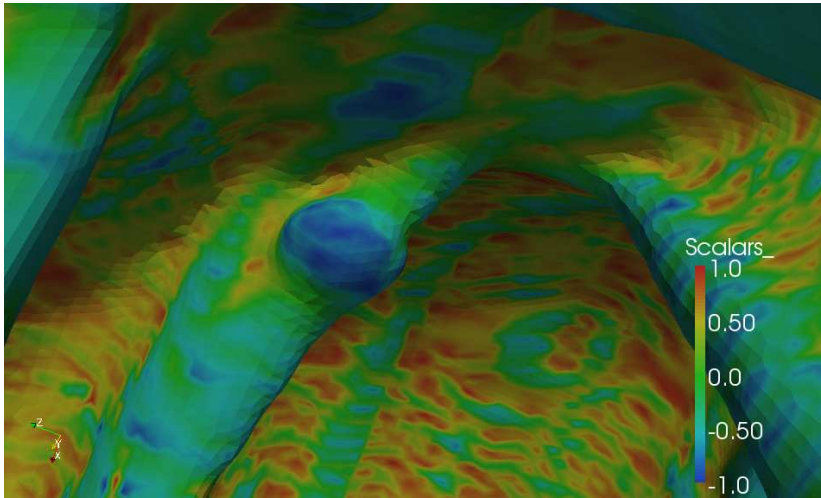
Shape Index y Curvedness



Shape Index y Curvedness



Shape Index y Curvedness



Detección primaria

Consideraciones locales

- Zonas con valores de Shape Index cercanos a -1 son consideradas como candidatas (detección primaria)
- Para cada zona Z_1 se considera un anillo Z_2
- Se calculan histogramas de Shape Index en Z_1 y Z_2 y se mide la distancia (L_1 y KL)



Características geométricas

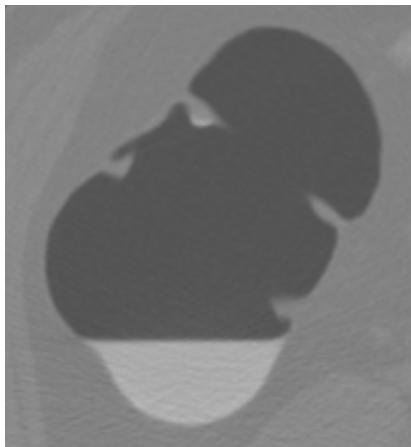
Cuatro características:

- distancia L_1 entre histogramas de Shape Index
- distancia KL entre histogramas de Shape Index
- factor de forma $F = \frac{4\pi \cdot Area}{Perimetro^2}$
- área de la zona candidata

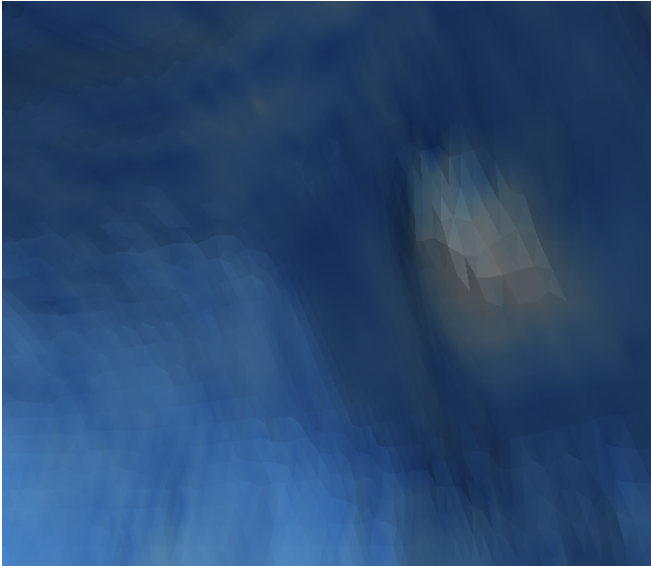
Características de nivel de gris

Absorvancia

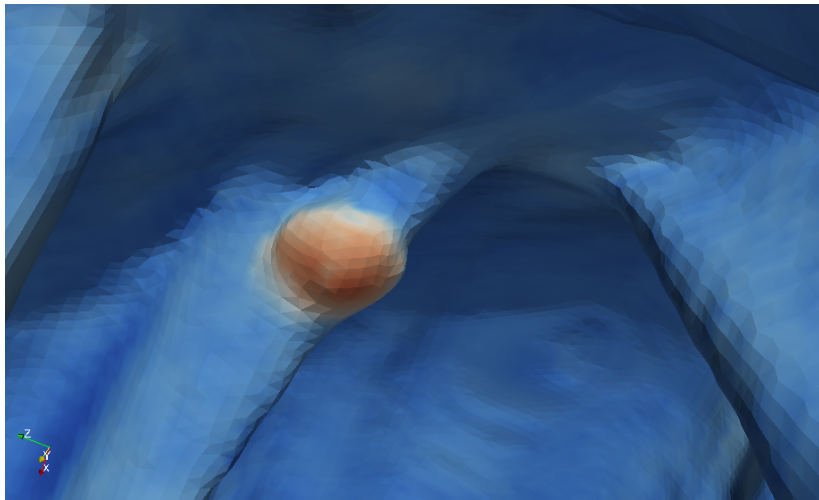
- Distinta actividad celular en pólipos



Nivel de gris

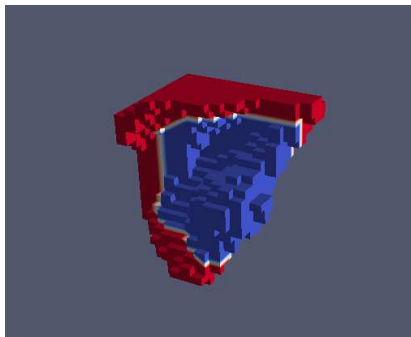


Nivel de gris



Características de nivel de gris

- Los valores de gris se encuentran en las imágenes CT, no en la superficie reconstruida.
- Para cada zona Z_1 se considera un volumen V_1 que la contenga, así como a los vóxeles de tejido cercanos.
- Se determina un volumen V_2 alrededor de V_1 para medir diferencias.



Características de nivel de gris

Características utilizadas

- Promedio de nivel de gris
- Características de Haralick
 - entropía
 - energía
 - máxima probabilidad
 - contraste
 - SumMean

Global y local

- Cada característica se calcula para V_1 y V_2 .
- Se consideran los valores relativos (diferencias entre V_1 y V_2) y absolutos (calculados solo en V_1)

Doce características:

Seis absolutas y seis relativas.

Agenda

- 1 Introducción
- 2 El problema
- 3 Características seleccionadas
- 4 Disparidad entre clases**
- 5 Clasificación y Resultados
- 6 Trabajo posterior

Disparidad entre clases

Desbalance:

- Luego de la detección primaria, se obtienen 1300 zonas candidatas, entre las cuales se encuentran los 17 pólipos.

Dos posibles soluciones:

- SMOTE (Synthetic Minority Over-sampling TEchnique)
- Muestreo de la clase predominante

Evaluación de SMOTE

Nuevos patrones

- Dos clases: A y B
- 5 características $\mathcal{N}(0, 1)$ (para ambas clases)

Dos grupos:

- Grupo 1: 100 instancias de A y 2000 de B
- Grupo 2: 20 instancias de A y 2000 de B

Teóricamente, cualquier clasificador rondaría el 50 % de clasificación correcta

Procedimiento

Para cada grupo:

- se generaron nuevos patrones de A
- se muestrearon patrones de B para equiparar
- se clasificó con *NaiveBayes* (con CV)

Sorteos:

- proporción de SMOTE de
 - 0 a 2000 % en grupo 1
 - 0 a 10000 % en grupo 2
- para cada proporción, se repitió el experimento 1000 veces.

Evaluación de SMOTE - Resultados

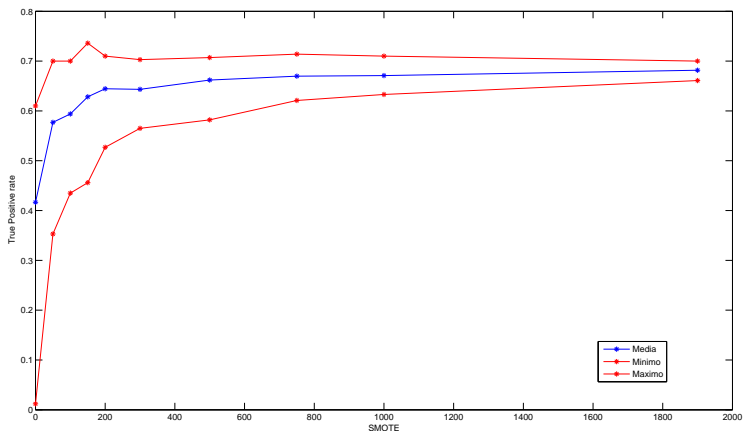


Figura: Grupo 1: TP rate vs. SMOTE prop.

Evaluación de SMOTE - Resultados

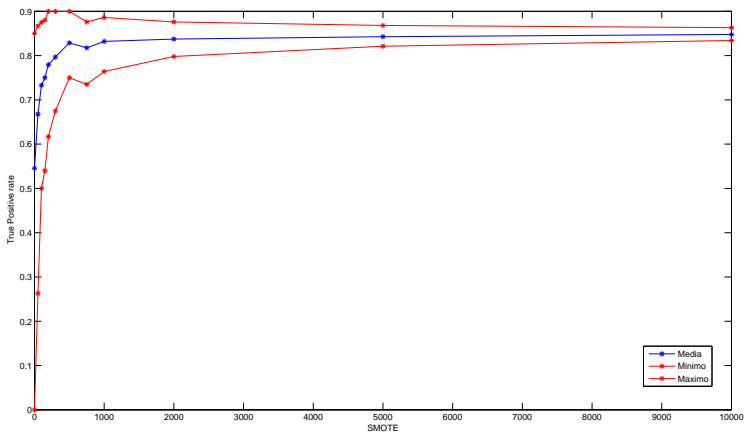


Figura: Grupo 2: TP rate vs. SMOTE prop.

Evaluación de SMOTE - Resultados

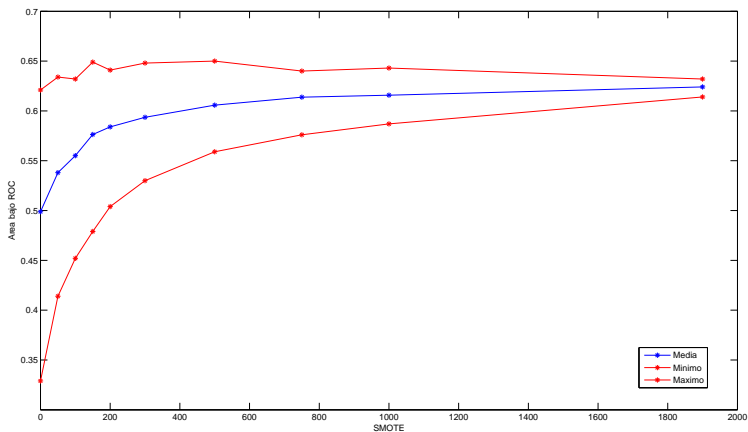


Figura: Grupo 1: AUC vs. SMOTE prop.

Evaluación de SMOTE - Resultados

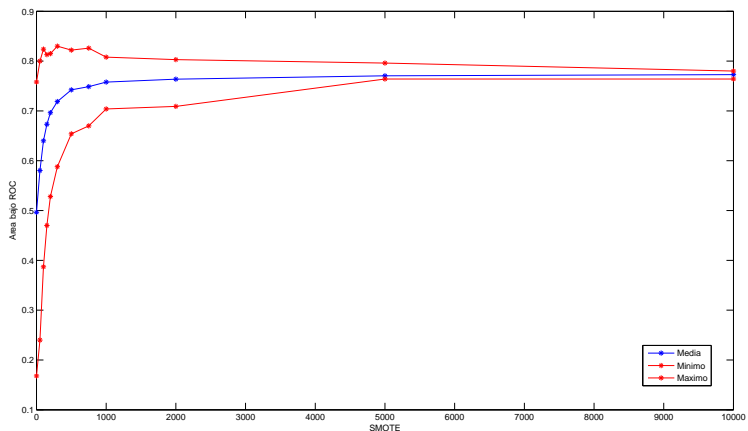


Figura: Grupo 2: AUC vs. SMOTE prop.

Evaluación de SMOTE

Observaciones

- A medida que aumenta la prop. de SMOTE la performance del clasificador aumenta
- Esto es más pronunciado cuanto menos patrones originales haya de la clase minoritaria

Disparidad entre clases

Forma de evaluación

- Vistos los problemas de SMOTE, se optó por evaluar mediante muestreo de la clase predominante.

Agenda

- 1 Introducción
- 2 El problema
- 3 Características seleccionadas
- 4 Disparidad entre clases
- 5 Clasificación y Resultados**
- 6 Trabajo posterior

Clasificación

Clasificadores

- *Naive Bayes*
- *AdaBoost* con árboles *C4.5*
- *SVM* de una clase

Clasificación

Clasificadores

- No se lograron buenos resultados con *SVM* de una clase.
- Las performances de *Naive Bayes* y *AdaBoost* fueron comparables.

Evaluación

- 1000 sorteos con semillas distintas.
- 10-CrossValidation utilizando *Naive Bayes*.
- Promediado de TP rate y AUC sobre los 1000 sorteos.
- Clasificación utilizando distintos subconjuntos de características

Resultados

	Geom	Gris(abs)	Gris(rel)	Geom + Gris(abs)	Geom + Gris(rel)	Todas
TP	0.895	0.739	0.951	0.912	0.975	0.956
FP	0.136	0.531	0.203	0.168	0.118	0.139
AUC	0.941	0.613	0.917	0.932	0.964	0.955

Cuadro: Resultados de clasificación utilizando *NaiveBayes*

Agenda

- 1 Introducción
- 2 El problema
- 3 Características seleccionadas
- 4 Disparidad entre clases
- 5 Clasificación y Resultados
- 6 Trabajo posterior**

Trabajo posterior

Trabajo realizado luego de la elaboración del informe:

- Entrenamiento mediante *Cost-sensitive learning* por el desbalance de clases (en lugar de muestreo de clase predominante).
- Mejoras en detalles de cálculos y definiciones.
- Resultados de 100% de sensibilidad con 3,2 FP por caso.