

Classificazione di cellule epiteliali HEP-2 mediante l'utilizzo dei tensori di Fisher

Lorenzo Cioni

lore.cioni@gmail.com

27 Agosto 2015

Sommario

Analizzare e classificare le cellule epiteliali di tipo 2 (HEp-2) mediante l'utilizzo della tecnica della immunofluorescenza indiretta è uno standard per rilevare malattie al tessuto connettivo umano, come ad esempio l'Artrite Reumatoide. Purtroppo questo metodo è molto costoso in termini di tempo e di lavoro impiegato e particolarmente soggettivo.

Questo elaborato ha come finalità quella di implementare un metodo per la classificazione di questo tipo di cellule basato sull'utilizzo del descrittore di covarianza e dei tensori di Fisher per l'estrazione di *features* dalle immagini.

Indice

1	Introduzione	1
2	Teoria	2
2.1	Il filtro di Gabor	2
2.2	Descrittore di Covarianza	4
2.3	Gaussian Mixture Model	4
2.4	Fisher Tensors	4
3	Dataset	4
4	Risultati	6
5	Implementazione	6
5.1	Esecuzione	6
6	Conclusioni	6

1 Introduzione

Una delle procedure standard per il rilevamento di malattie al tessuto connettivo umano, come ad esempio l'Artrite Reumatoide o il Lupus, è l'utilizzo di Immuno-

fluorescenza Indiretta sulle cellule epiteliali di tipo 2, altrimenti conosciute come HEp-2.

Questo tipo di analisi ha due principali svantaggi: è molto soggettiva e richiede un gran numero di ore lavorative. Si è così pensato ad un metodo per automatizzare il processo per ottenere risultati migliori sia sotto il profilo medico che dal punto di vista di tempo impiegato.

Il metodo proposto e implementato è tratto da un articolo pubblicato in occasione del contest di *Pattern Recognition 2014*¹ [1]. Per la classificazione delle cellule si procede inizialmente all'estrazione di un adeguato numero di *features* attraverso l'utilizzo del *Descrittore di Covarianza* [2], vengono poi utilizzati i *Tensori di Fisher* che codificano informazioni aggiuntive rispetto alla distribuzione delle *features* ed infine le cellule vengono classificate tramite un SVM multiclasse.

I test per la valutazione della bontà del metodo sono stati effettuati sul dataset della competizione².

2 Teoria

Vengono ora presentate le tecniche utilizzate all'interno del programma: l'estrazione di *features* dall'immagine mediante l'utilizzo dei *filtri di Gabor*, il Descrittore di Covarianza (*Covariance Descriptor*), il Modello Mistura di Gaussiane (*Gaussian Mixture Model, GMM*) e i Tensori di Fisher (*Fisher Tensors*).

2.1 Il filtro di Gabor

I filtri di Gabor³ sono filtri passa-banda usati nell'analisi di immagini principalmente per l'estrazione di *features* e l'analisi basata sulla tessitura.

La risposta finita all'impulso di questi filtri è calcolata come prodotto di uno sviluppo Gaussiano con oscillazione complessa. Estendendo queste funzioni a due dimensioni è possibile creare filtri sensibili all'orientazione⁴ e sotto certe condizioni è possibile approssimare linearmente la fase.

Sia (x, y) un punto dell'immagine. L'equazione per il filtro di Gabor 2D è la seguente:

$$G(x, y) = e^{(-\frac{(x')^2 + \gamma^2 (y')^2}{2\sigma^2})} \cos(2\pi \frac{x'}{\lambda})$$

con

$$x' = x \cos \theta + y \sin \theta$$

$$y' = -x \sin \theta + y \cos \theta$$

I parametri del filtro sono:

¹ICPR Contest 2014 - HEp-2 Cells Classification Contest

²HEp-2 Dataset

³Gabor, D.: *Theory of communication*. In J. IEE, vol. 93, pp. 429-457, Londra, 1946.

⁴Daugman, J. G.: *Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters* J. Optical Society of America A, vol. 2, no. 7, pp. 1160-1169, July 1985.

- θ : orientazione del filtro, espressa in gradi.
- λ : lunghezza d'onda del fattore coseno, espressa in pixels.
- γ : specifica l'ellitticità del supporto della funzione di Gabor.
- σ : è il parametro che regola l'involuppo Gaussiano.

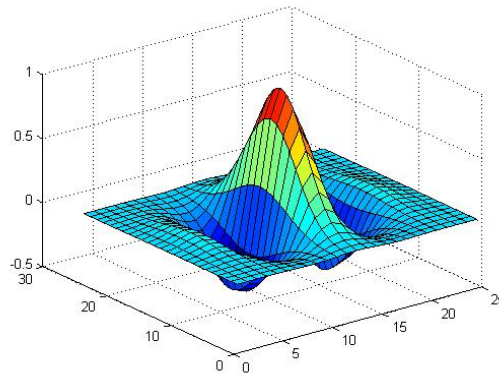


Figura 1: *Esempio di filtro di Gabor 2D*

Al fine di estrarre utili *features* da un'immagine è utile utilizzare un set di filtri di Gabor con parametri diversi, ad esempio la scala e l'orientazione.

Nel programma viene utilizzato un set di filtri di Gabor con 4 diverse angolazioni e 3 diverse scale per un totale di 12 filtri:

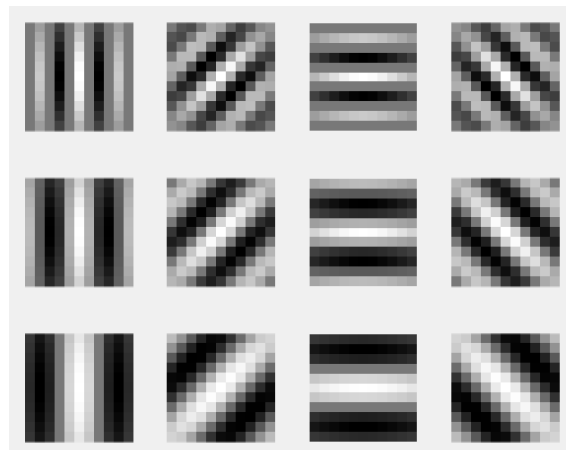


Figura 2: *Set di filtri di Gabor con 4 angolazioni e 3 scale*

Per ciascun pixel dell'immagine viene dunque calcolato il vettore delle *features* ottenute mediante la convoluzione dell'immagine con i filtri di Gabor (dividendo la parte reale dalla parte immaginaria).

2.2 Descrittore di Covarianza

Il Descrittore di Covarianza (Covariance Descriptor⁵, CovD) viene utilizzato in letteratura per descrivere e caratterizzare regioni di immagini.

Data un'immagine con associato, a ciascun pixel, un vettore di *features*, il Descrittore di Covarianza codifica informazioni che riguardano le *varianze* di queste ultime all'interno di una regione, la loro correlazione e la loro distribuzione nello spazio. Questo metodo è molto utilizzato in particolare modo quando si ha bisogno di rimanere invarianti rispetto all'illuminazione o alla rotazione.

Data un'immagine I con associato, per ciascun pixel (x, y) , il vettore di features di dimensione d , si considera F come l'immagine di features a d livelli. Viene dunque considerata una regione di F tale che $R \subset F$ e consideriamo $\{f_i\}_{i=1,\dots,n} \in R$ la i -esima features, con n numero di pixels della regione.

A questo punto la regione R viene rappresentata da una matrice di covarianza di dimensione dXd dei *features points*:

$$C = \frac{1}{n-1} \sum_{k=1}^n (f_k - \mu)(f_k - \mu)^T$$

con μ la media dei *features points*.

Poichè la matrice C è definita positiva e simmetrica può essere considerata un *tensore*.

2.3 Gaussian Mixture Model

Un Modello Mistura di Gaussiane (*Gaussian Mixture Model*) è una funzione parametrica di distribuzione di probabilità rappresentata come la somma pesata di Gaussiane. A Gaussian Mixture Model (GMM) is a parametric probability density function represented as a weighted sum of Gaussian component densities. GMMs are commonly used as a parametric model of the probability distribution of continuous measurements or features in a biometric system, such as vocal-tract related spectral features in a speaker recognition system. GMM parameters are estimated from training data using the iterative Expectation-Maximization (EM) algorithm or Maximum A Posteriori (MAP) estimation from a well-trained prior model.

2.4 Fisher Tensors

3 Dataset

Per la valutazione dell'efficienza dell'algoritmo implementato è stato utilizzato una parte del dataset dell'*ICPR HEp-2 Cell Classification Contest*.

Il dataset contiene al suo interno 149 immagini che rappresentano le scansioni di vetrini di laboratorio. All'interno di queste immagini è possibile distinguere 6 diversi pattern: *punteggiato*, *nucleolare*, *citoplasmico*, *omogeneo*, *granulare*, *negativo*. Alcune immagini del dataset erano però classificate come *altro*, rendendo difficile una

⁵Oncel Tuzel, Fatih Porikli, Peter Meer, *Region Covariance: A Fast Descriptor for Detection and Classification* Mitsubishi Electric Research Laboratories, Inc., 2006.

loro classificazione: queste sono state dunque rimosse fino ad ottenere un totale di 137 immagini.

Le immagini sono state acquisite tramite un microscopio a fluorescenza (40 ingrandimenti) in combinazione con una lampada di mercurio vaporizzato a 50W e una fotocamera digitale SLIM con risoluzione 1920 x 1110.

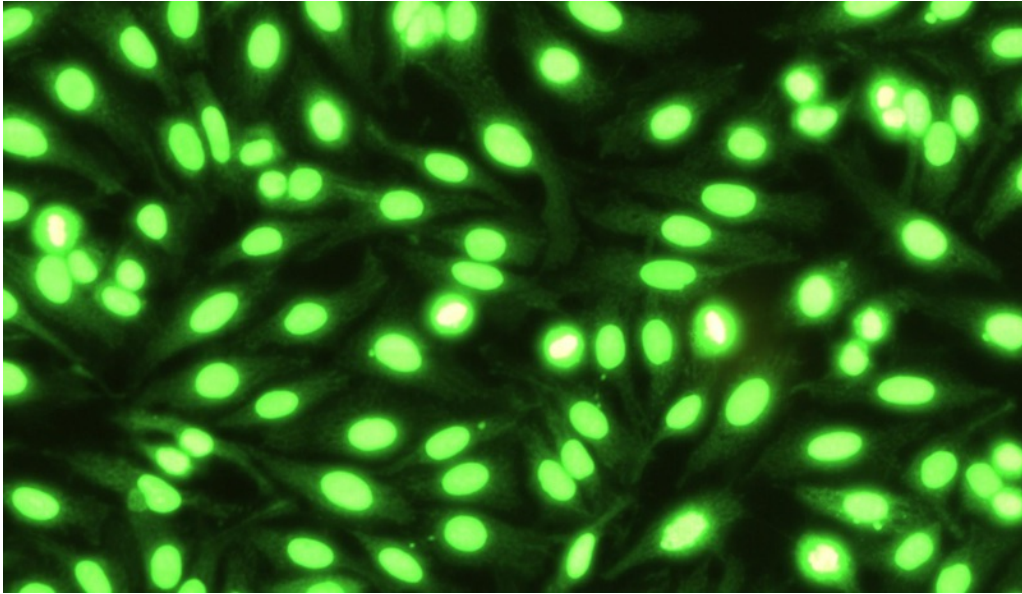


Figura 3: *Esempio di immagine del dataset*

Ciascuna immagine è stata poi annotata da un medico specialista ed associata ad una delle classi di pattern sopra esposte. Le annotazioni sono state inserite in una tabella così da poterne ricavare un *training set*.

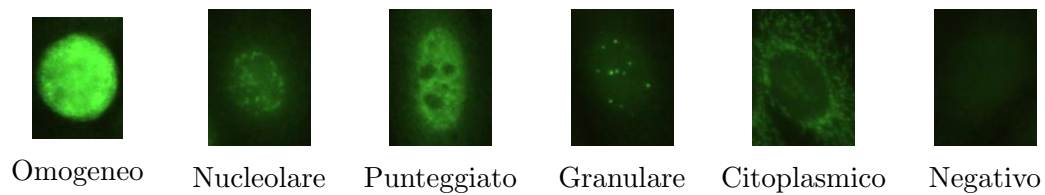


Figura 4: Esempi di cellule classificate nel dataset

4 Risultati

5 Implementazione

5.1 Esecuzione

6 Conclusioni

Riferimenti bibliografici

- [1] Masoud Faraki, Mehrtash T. Harandi, Arnold Wiliem, Brian C. Lovell, *Fisher tensors for classifying human epithelial cells*. Pattern Recognition, Volume 47, 2014, pp. 2348 - 2359.
- [2] Oncel Tuzel, Fatih Porikli, Peter Meer, *Region Covariance: A Fast Descriptor for Detection and Classification*. Mitsubishi Electric Research Laboratories, Inc., 2006.
- [3] Gabor, D.: *Theory of communication* In J. IEE, vol. 93, pp. 429-457, London, 1946.