



POLITECNICO
MILANO 1863

Esercizio 1 Homework 2 - Laboratorio di elaborazione di bioimmagini

Autore: Francesco Benedetto 10663326

Indice

1	Introduzione	
2	Metodo	
2.1	Caricamento e analisi preliminare dell'immagine	1
2.2	Segmentazione preliminare	2
2.3	Etichettatura e analisi delle bounding box	2
2.4	Bounding box "problematiche"	3
2.5	Seconda segmentazione e seconda etichettatura	3
2.6	Classificazione dei nuclei e salvataggio metriche	3
3	Risultati	4
4	Conclusioni	5

1 Introduzione

Questo progetto si focalizza sulla segmentazione di nuclei cellulari e sulla loro classificazione in due categorie

principali: sani e malati.

Gli obiettivi principali sono:

- Segmentare i nuclei cellulari distinguendoli dal background e dal citoplasma.
- Identificare e separare nuclei uniti.
- Estrarre proprietà geometriche e di intensità per classificare i nuclei.
- Visualizzare e verificare i risultati della classificazione.

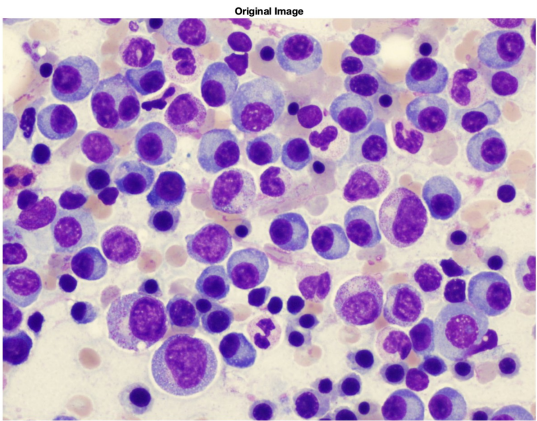


Figura 1: Immagine iniziale

2 Metodo

2.1 Caricamento e analisi preliminare dell'immagine

Essendo un'immagine a colori, si è cercato di capire quale spazio di colori (compresa l'immagine convertita in scala di grigi) sia più adatto per la segmentazione. La scelta si è basata sull'istogramma di questi ultimi, scegliendo quello con le valli più marcate, che permettano meglio di segmentare i nuclei tramite thresholding. Per la segmentazione iniziale si sarebbe potuto

usare anche un Kmeans ($k=3$ 1.Background 2.Citoplasma 3.Nuclei). Nonostante ciò si è preferito usare il thresholding nello spazio di colori HSV, in quanto ha poi permesso di trovare più facilmente il valore di intensità del viola scuro (utile alla classificazione).

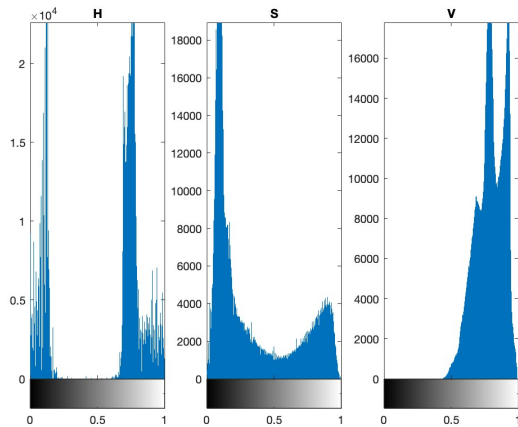


Figura 2: Istogrammi spazio di colore HSV

2.2 Segmentazione preliminare

Si è deciso di segmentare i nuclei e non la cellula completa (nucleo + citoplasma) in quanto la segmentazione di questi ultimi fa sì che ci siano meno sovrapposizioni. L'analisi successiva, non dovrebbe essere influenzata da questa scelta, in quanto la classificazione può essere basata su proprietà geometriche e statistiche dei nuclei. Dopo aver osservato gli istogrammi si è applicato un thresholding manuale a due canali dello spazio HSV:

- Il canale H è stato scelto per eliminare esclusivamente il background, mentre il canale S per eliminare il citoplasma:
 - per valori H minori di 0.523 si ha solo background, altrimenti nuclei più citoplasma.
 - per valori S maggiori 0.6 si hanno solo i nuclei, per valori minori citoplasma.

Con questa combinazione di soglie si è applicata la prima segmentazione, ottenendo la maschera binaria dei nuclei, i cui valori di intensità vengono invertiti in modo tale da avere il background nero e gli oggetti (i nuclei)

bianchi. Successivamente è stata effettuata anche una prima pulizia:

- Rimozione del rumore: `bwareaopen` è stata utilizzata per rimuovere oggetti piccoli (area minore di 50 pixel).
- Operazioni morfologiche: `imopen` con un disco di raggio 5 per separare nuclei adiacenti.

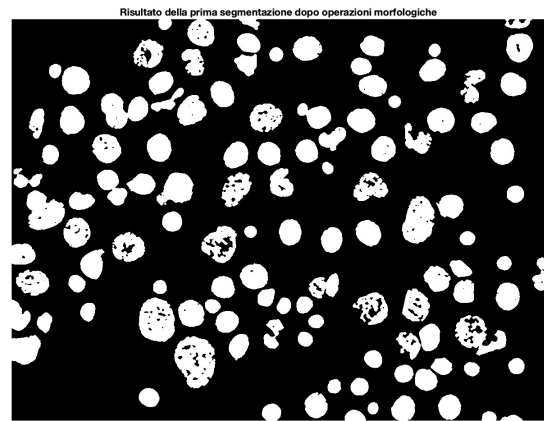


Figura 3: Risultato della prima segmentazione dopo operazioni morfologiche

2.3 Etichettatura e analisi delle bounding box

Dopo la segmentazione dei nuclei, sono stati etichettati gli oggetti e visualizzate le loro bounding box per verificare se ci sono alcune di queste con nuclei uniti. È stato notato che ci sono alcune bounding box "problematiche" che contengono più nuclei, contando quindi come un unico oggetto. Per affrontare questo problema, è stato deciso di scrivere un secondo codice ('Es1_Analisi_BB_Problematiche.m') per non interferire nel flusso di quello principale.

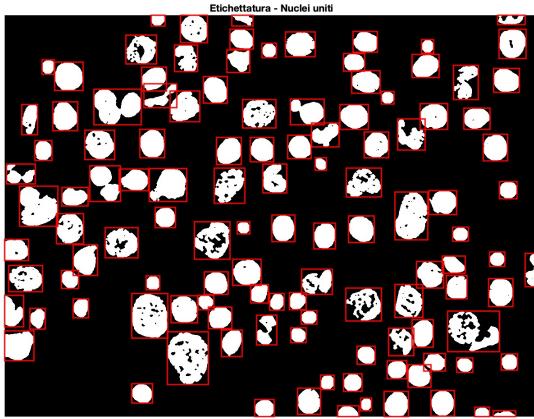


Figura 4: Le bounding box dovrebbero delimitare un singolo oggetto

2.4 Bounding box "problematiche"

Nello script dedicato esclusivamente all'analisi delle bounding box anomale, si è cercato di dare una soluzione ai seguenti limiti:

- più di un nucleo per bounding box
- sovrapposizioni
- analisi più dettagliata di nuclei con forme atipiche

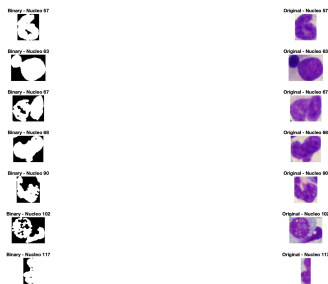


Figura 5: Esempi di boundingbox "problematiche"

Gli oggetti composti da due o tre nuclei (solo circa 15 sul totale di 130) sono stati segmentati manualmente. Tra questi ci sono due bounding box in cui ci sono parti di nucleo che si sovrappongono. Essendo quindi pochissimi casi, si è deciso di perdere quella poca informazione. In alternativa si sarebbe potuto applicare la Hough Transform cercando circonferenze, ma solo per casi in cui le cellule avessero forma circolare. Per le

altre si sarebbero approssimate erroneamente le altre forme come cerchi. Alla fine di questa sezione è stato salvato il risultato per poterlo utilizzare nel codice principale.

2.5 Seconda segmentazione e seconda etichettatura

A seguito dell'analisi dettagliata, tutti i nuclei prima uniti, sono stati separati correttamente. Questi ultimi da qui in poi sono stati considerati come oggetti distinti (ognuno con la propria bounding box. Come operazione morfologica finale alcuni nuclei che presentavano dei buchi all'interno sono stati riempiti mediante `imfill` (facendo così si sono recuperati pixel che avrebbero influenzato il calcolo dell'area). Gli oggetti sono stati a questo punto nuovamente etichettati (questa volta definitivamente) e per ognuno di loro sono state calcolate diverse proprietà (successivamente descritte).

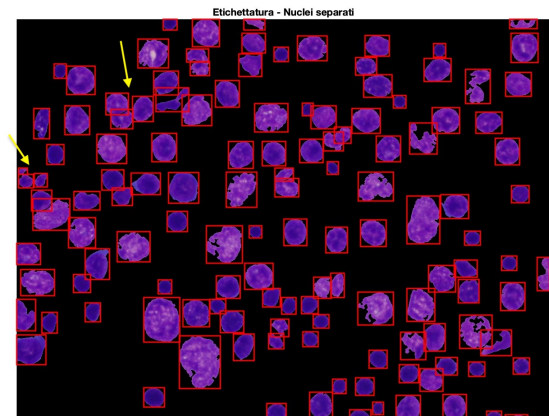


Figura 6: Le frecce gialle indicano due esempi di bounding box che sono state modificate nel codice secondario

2.6 Classificazione dei nuclei e salvataggio metriche

Le classi dei nuclei su cui è stata basata la classificazione sono 2: **SANI** e **MALATI**. All'interno della classe **MALATI**, si considerano anche i nuclei di cui la classificazione è ambigua. Questo permette di avere una *recall* più alta per avere il minor numero possibile di *falsi negativi*. La classificazione si basa su 4 proprietà:

area, eccentricità, deviazione standard dell'intensità e intensità media del canale V.

- **Area:** nuclei più grandi (area maggiore) possono essere sintomo di malattia.
- **Eccentricità:** nuclei poco circolari (eccentricità alta) possono essere sintomo di malattia (il nucleo inizia a cambiare forma).
- **Deviazione standard dell'intensità della scala di grigi:** nuclei con texture poco uniforme (std più alta) possono essere sintomo di malattia.
- **Intensità media del canale V:** nuclei con colore che tende al rosa possono essere sintomo di malattia.

N.B. Questo approccio tende ad automatizzare la classificazione. Per selezionare precisamente le cellule da classificare manualmente, si potrebbe usare `bwselect`.

I parametri migliori trovati si basano sulla classificazione di questa immagine specifica. Per generalizzare il codice, sarebbe opportuno diminuire la soglia dell'eccentricità circa a 0.70 e modificare le soglie. In questo caso la soglia dell'eccentricità è stata aumentata per classificare meglio i nuclei ai bordi dell'immagine (essendo tagliati sono meno circolari). Per essere considerato malato, basta che il nucleo superi almeno una delle seguenti soglie:

- **Area Threshold** = 2050; Soglia di area
- **Eccentricity Threshold** = 0.9; Soglia di eccentricità
- **Std Threshold** = 0.3; Soglia di STD
- **V Threshold** = 0.63; Soglia di intensità del canale V (< 0.63 è viola scuro, > è rosa)

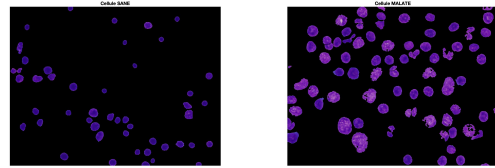


Figura 7: A sinistra le segmentazione (con colorazione originale) dei nuclei SANE a destra dei nuclei MALATI

3 Risultati

Di seguito i risultati della segmentazione e del calcolo delle metriche per classe:

Classe	SANE	MALATI
Numero di nuclei	48	82
Area media	1098.33 ± 428.21	2386.89 ± 1167.86
Eccentricità media	0.48 ± 0.19	0.59 ± 0.17
Deviazione std	0.07 ± 0.01	0.07 ± 0.01
Intensità media V	0.57 ± 0.03	0.66 ± 0.03

Tabella 1: Proprietà media e deviazione standard per ciascuna classe di nuclei.

La deviazione standard è simile per le due classi (molto bassa), dato che sono stati utilizzati solo i nuclei (texture più uniforme rispetto a nucleo + citoplasma)

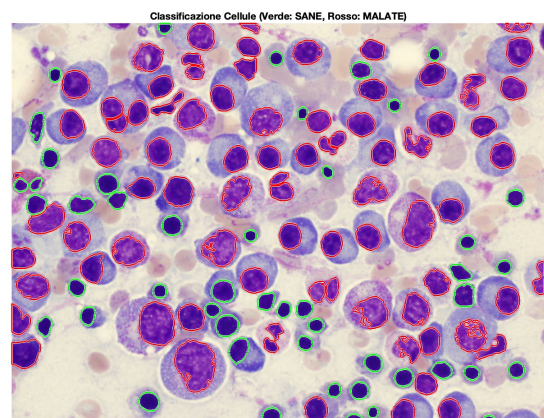


Figura 8: Risultato della classificazione: in verde nuclei SANE in rosso nuclei MALATI

4 Conclusioni

Il processo di segmentazione ha avuto successo nell'identificare i nuclei e nel separare quelli adiacenti. Le proprietà geometriche e di intensità estratte hanno permesso di classificare correttamente i nuclei in due categorie: sani e malati. I risultati mostrano una buona discriminazione tra le due classi, con nuclei malati che mostrano un'area maggiore, una maggiore eccentricità e una più alta deviazione standard dell'intensità.