

# Metodi di apprendimento non supervisionato per l'analisi di mappe di estinzione in astrofisica



## **Relatori**

Prof. Francesco Isgrò  
Dr. Massimo Brescia

## **Candidato**

Francesco Esposito  
N97000183

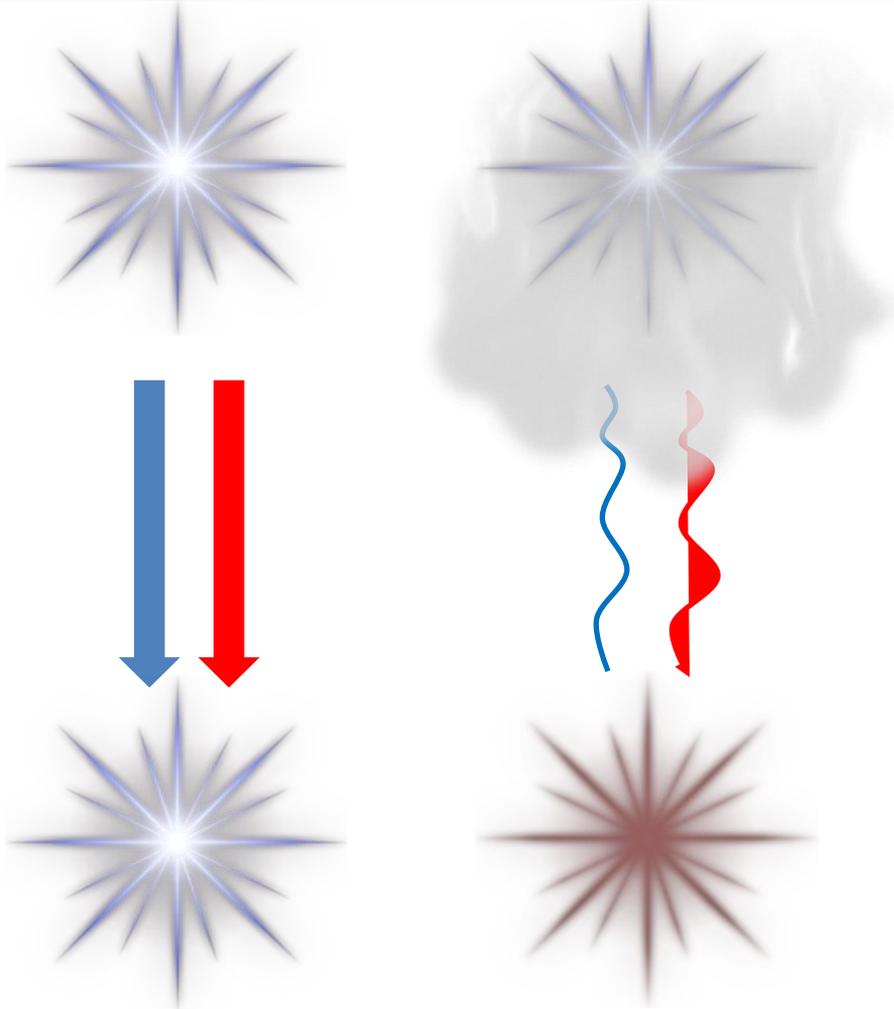
## **Correlatore**

Prof.ssa Paola Festa

---

ANNO ACCADEMICO 2014/2015

# Estinzione galattica



L'estinzione è una sorta di cataratta naturale dovuta all'assorbimento e scattering del flusso fotonico nei gas e polveri delle regioni di formazione stellare (nebulose) posti tra osservatore e sorgente.

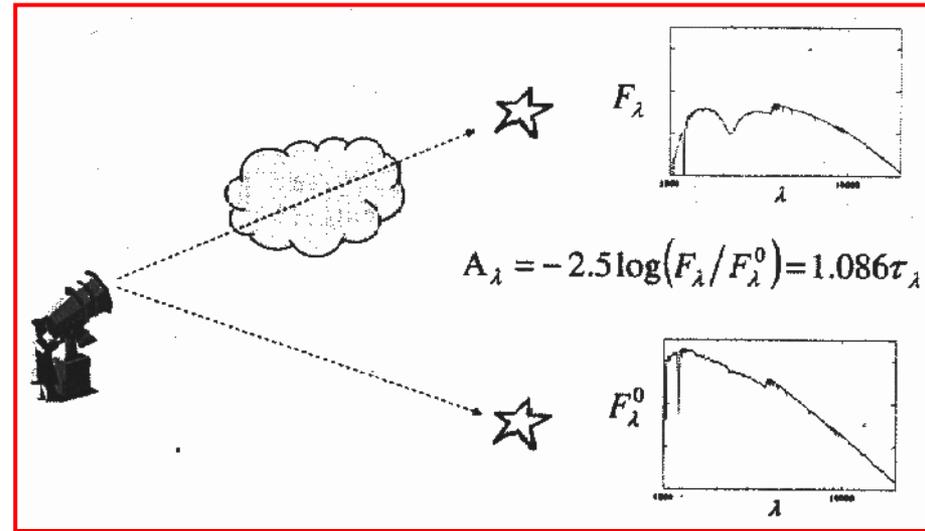
Un altro effetto è un maggiore assorbimento della luce blu, che rende l'oggetto più rosso.

La quantità di estinzione è estremamente variabile, dipendendo dalla natura delle zone di formazione stellare.



# Metodo tradizionale di calcolo

La curva di estinzione di una stella si misura attraverso il confronto tra il suo spettro e quello osservato per una stella simile di cui sia noto il flusso senza estinzione. È anche possibile utilizzare un modello teorico di spettro per una regione con una certa densità molecolare.

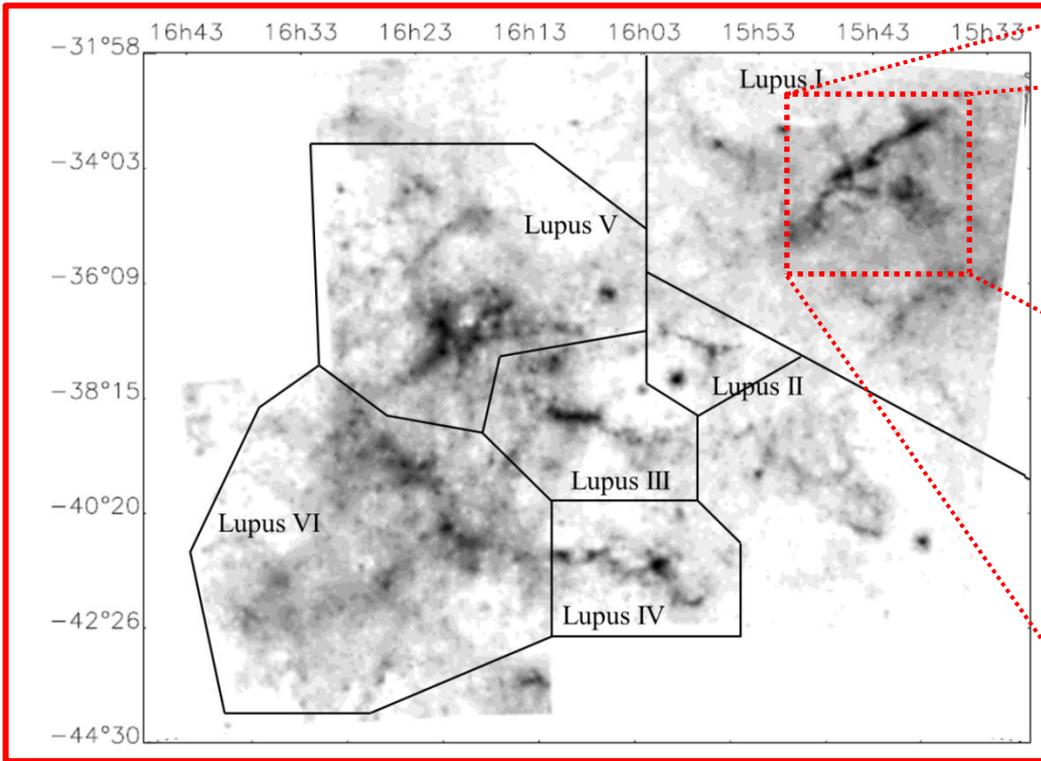


Nelle immagini astronomiche la relazione tra l'estinzione totale  $A_V$ , misurata in magnitudini, e la cosiddetta *column density* (colonna di densità) degli atomi di idrogeno  $N_H$ , misurata in  $\text{cm}^{-2}$ , mostra l'interazione tra il gas e le polveri nel mezzo interstellare. Da studi di regioni stellari effettuati con la spettroscopia ultravioletta e ai raggi X, risulta nota la relazione  $\frac{N_H}{A_V}$

$$\left\{ \begin{array}{l} \left[ \frac{N_H}{A_V} \right]_A = 1.37 \times 10^{21} \\ \left[ \frac{N_H}{A_V} \right]_C = 1.87 \times 10^{21} \end{array} \right. \begin{array}{l} \longrightarrow \\ \longrightarrow \end{array} \begin{array}{l} \text{Alcal\`a 2016, INAF-OAC (com. priv.)} \\ \text{Cambr\`esy 1999, A\&A 345} \end{array}$$

$[cm^{-2} mag^{-1}]$

# Nebulosa di Lupus



Mappa d'estinzione di Lupus osservata nella banda B (coordinate J2000; Cambrèsy 1999)

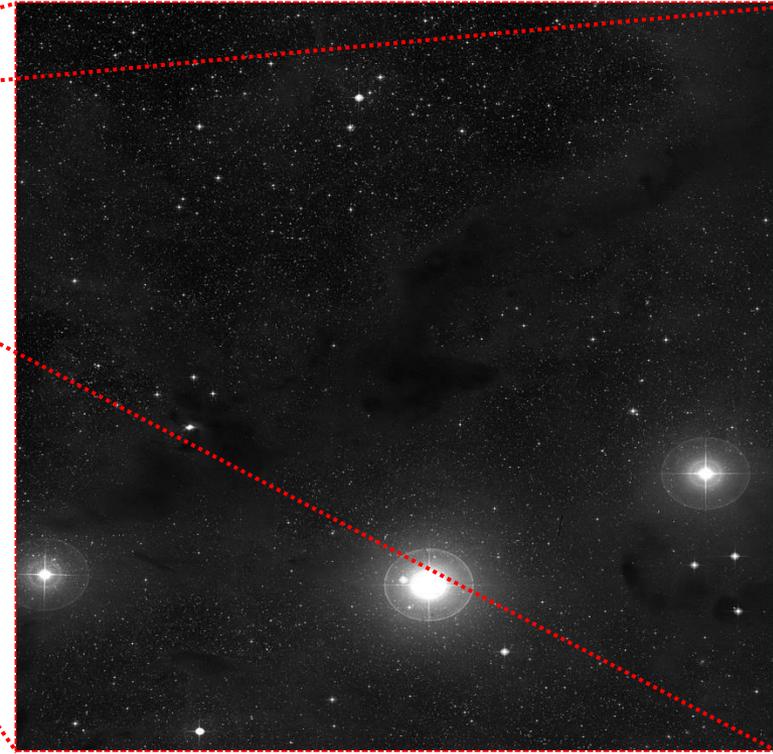


Immagine osservata da USNO-PMM di Lupus I nella banda B (coordinate J2000, 5x5 gradi DEC, 7000 x 7000 pixel di risoluzione)

La misurazione è stata effettuata attraverso un metodo di conteggio di stelle usando una griglia adattiva formata da una decomposizione basata su filtri wavelet applicata ad immagini ottiche ad alta risoluzione (Cambrèsy 1999)

# Thresholding e rapporto massa/estinzione

Il metodo descritto porta alla costruzione di mappe di estinzione sulle quali definire degli iso-contorni, è fondamentale per i calcoli di massa

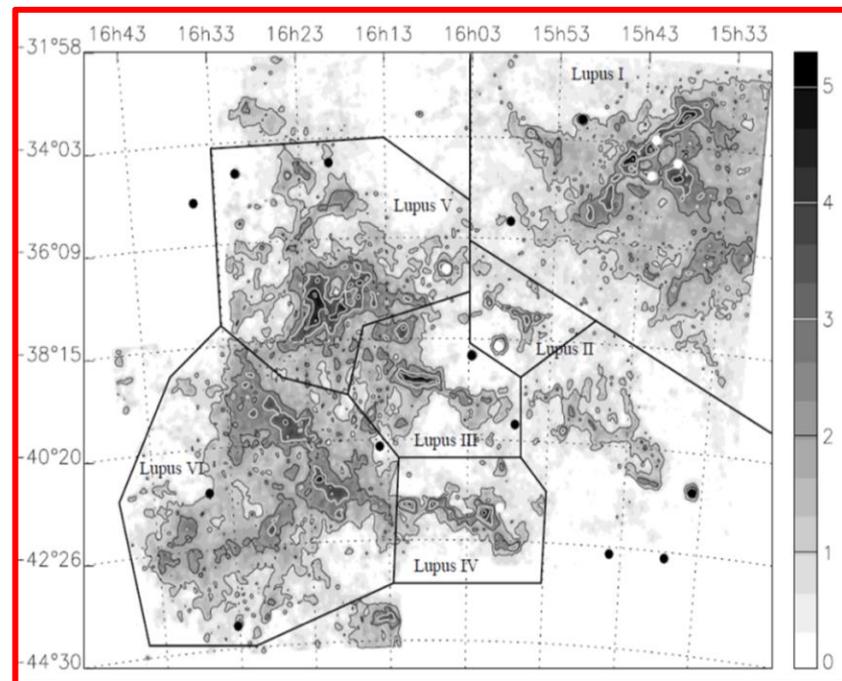
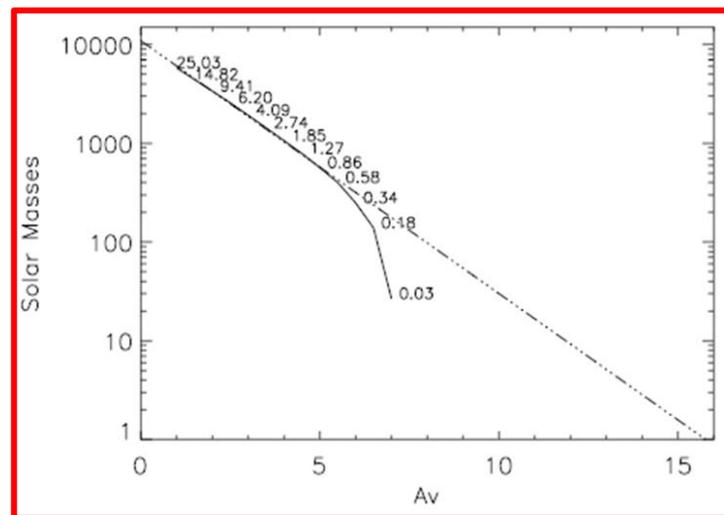
$$\log M = \log M_{tot} + \text{slope} \times A_V$$

$$M(A_V) = I_{pix}^2 \frac{N_H}{A_V} \mu \frac{m_H}{M_\odot} \sum_i A_V(i)$$

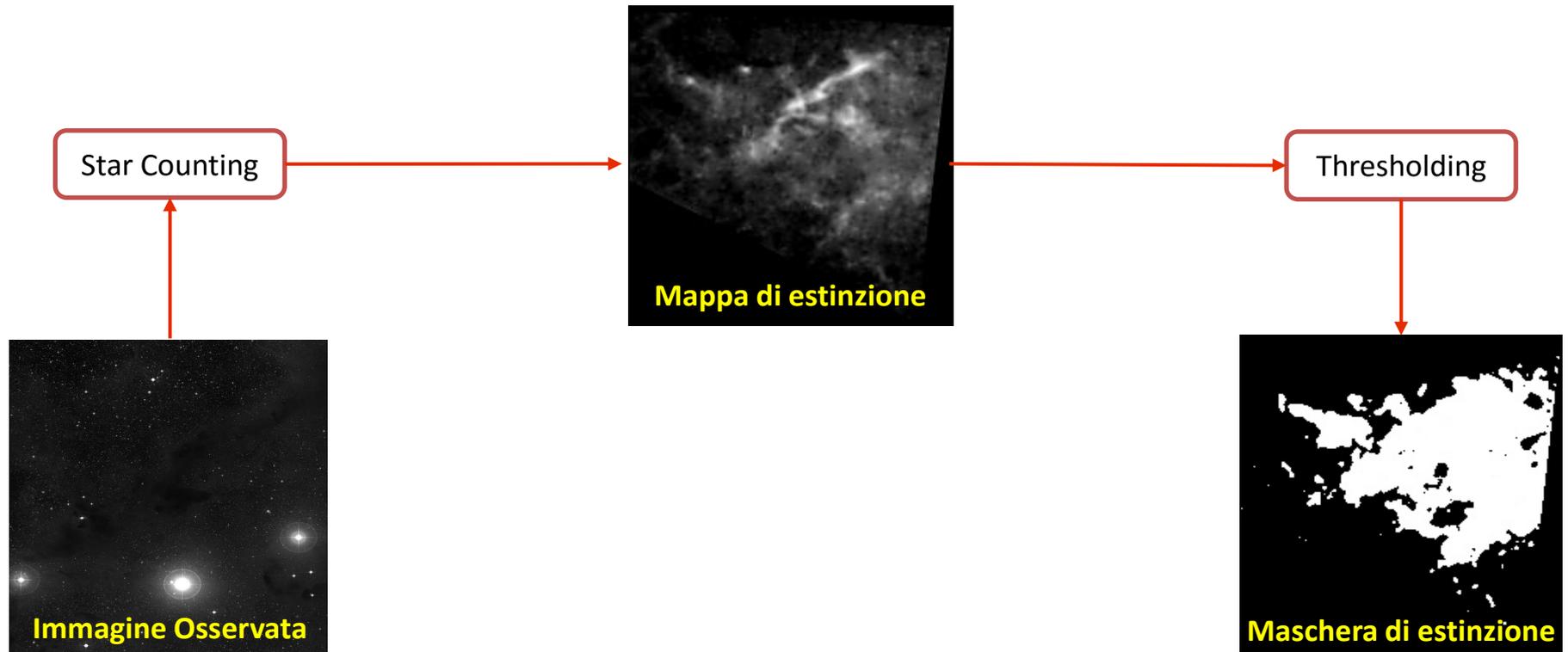
$$\frac{m_H}{M_\odot} = \frac{1.66}{1.98} \times 10^{-57} [M_\odot]$$

$$\begin{cases} [\mu]_A = 2.02 \\ [\mu]_C = 2.4 \end{cases}$$

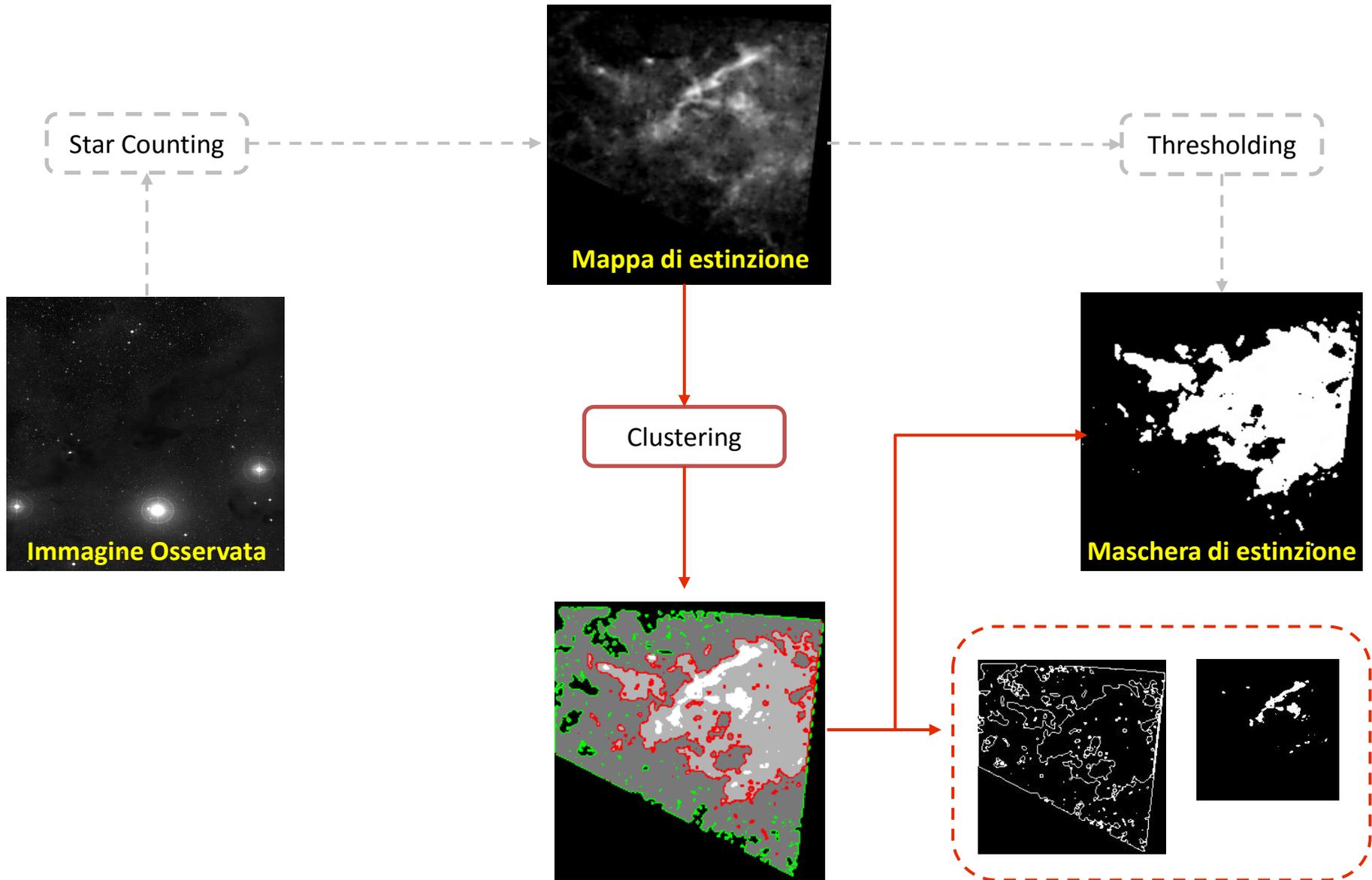
Regione	Distanza (pc)	$A_V^m$	$A_V^e$	$M_{tot}$	Slope
Lupus I	150	5.3	7.1	$10^4$	-0.56
Lupus II	150	3.8	5.7	80	-0.33
Lupus III	150	4.9	7.6	1150	-0.40
Lupus IV	150	5.3	7.0	630	-0.40
Lupus V	150	5.2	10.6	2500	-0.32
Lupus VI	150	4.8	7.0	$10^4$	-0.57



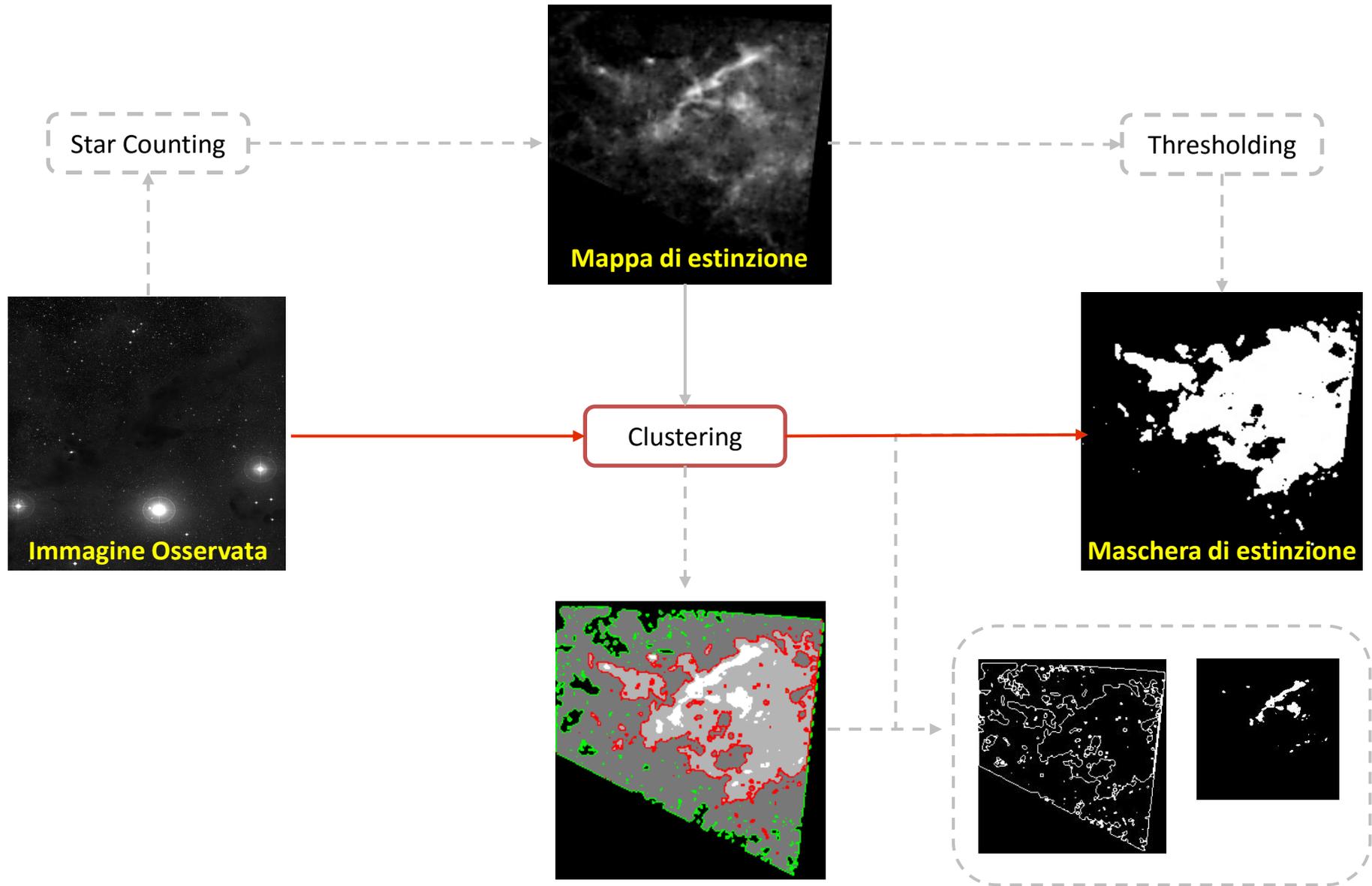
# Identificazione della regione di estinzione



# Identificazione della regione di estinzione



# Identificazione della regione di estinzione



# Problemi

Osservazioni  
astronomiche

→ {  
Immagini/spettri a basso SN Ratio  
Proiezione 2D, no profondità → foreground/background noise  
No verità assoluta (modelli/simulazioni) → **unsupervised**

Estinzione

→ {  
«effetto cataratta» non eliminabile  
Conteggio statistico affetto da errore (soprattutto distanza regione)  
Thresholding arbitrario del flusso  
Contorni regione non facilmente definibili

**Obiettivi**

Thresholding automatico (**edge detection**)  
Segmentazione dell'estinzione (**clustering**)

**Metodo**

Unsupervised Machine Learning

**Procedura**

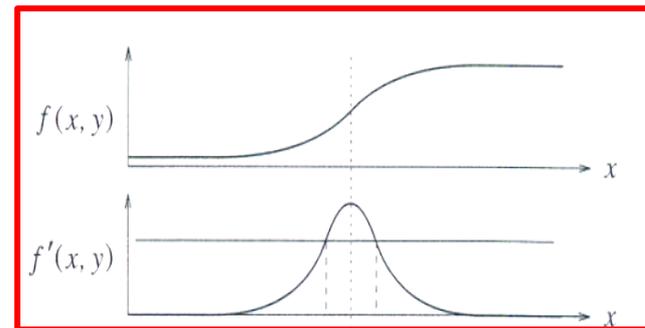
Immagini sintetiche  
Validazione tecnologica

Mappe di estinzione  
Validazione scientifica

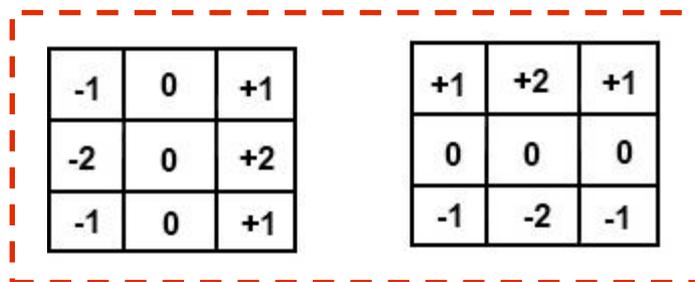
Immagini osservate  
RoI per calcolo estinzione

# Edge Detection

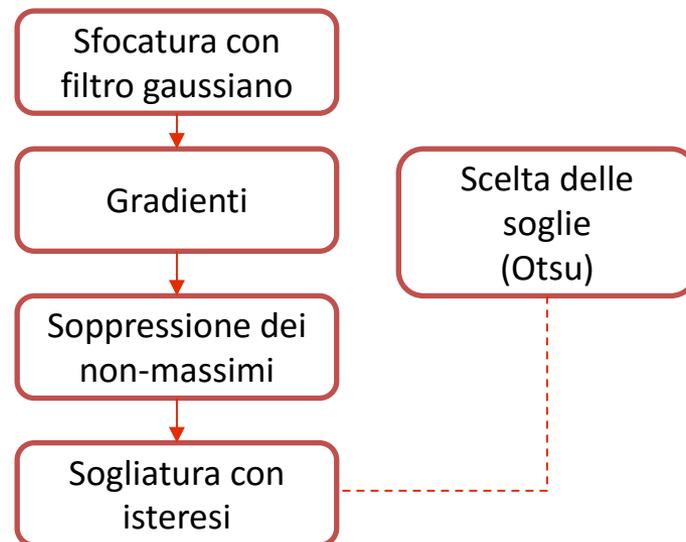
Gli edge si configurano come bruschi cambiamenti del livello di grigio identificabili come picchi della derivata prima. Su questo concetto basano il loro comportamento i più classici metodi di edge detection:



## Operatori di Sobel



## Metodo di Canny



# Fuzzy Edge Detection

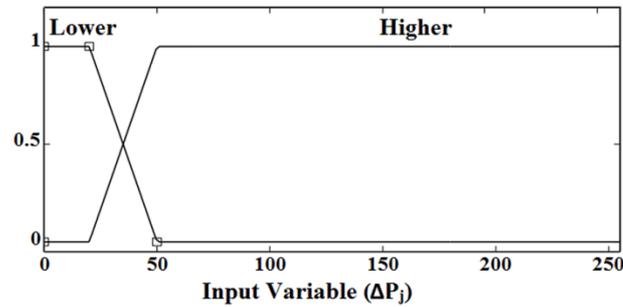
Kernel convolution



$P_1$	$P_2$	$P_3$	$\Delta P_1$	$\Delta P_2$	$\Delta P_3$
$P_4$	$P$	$P_5$	$\Delta P_4$	$P$	$\Delta P_5$
$P_6$	$P_7$	$P_8$	$\Delta P_6$	$\Delta P_7$	$\Delta P_8$

$$\Delta P_j = |P_j - P|$$

Fuzzification

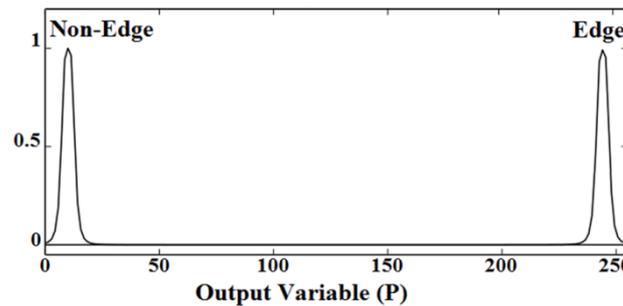


Fuzzy Inference System

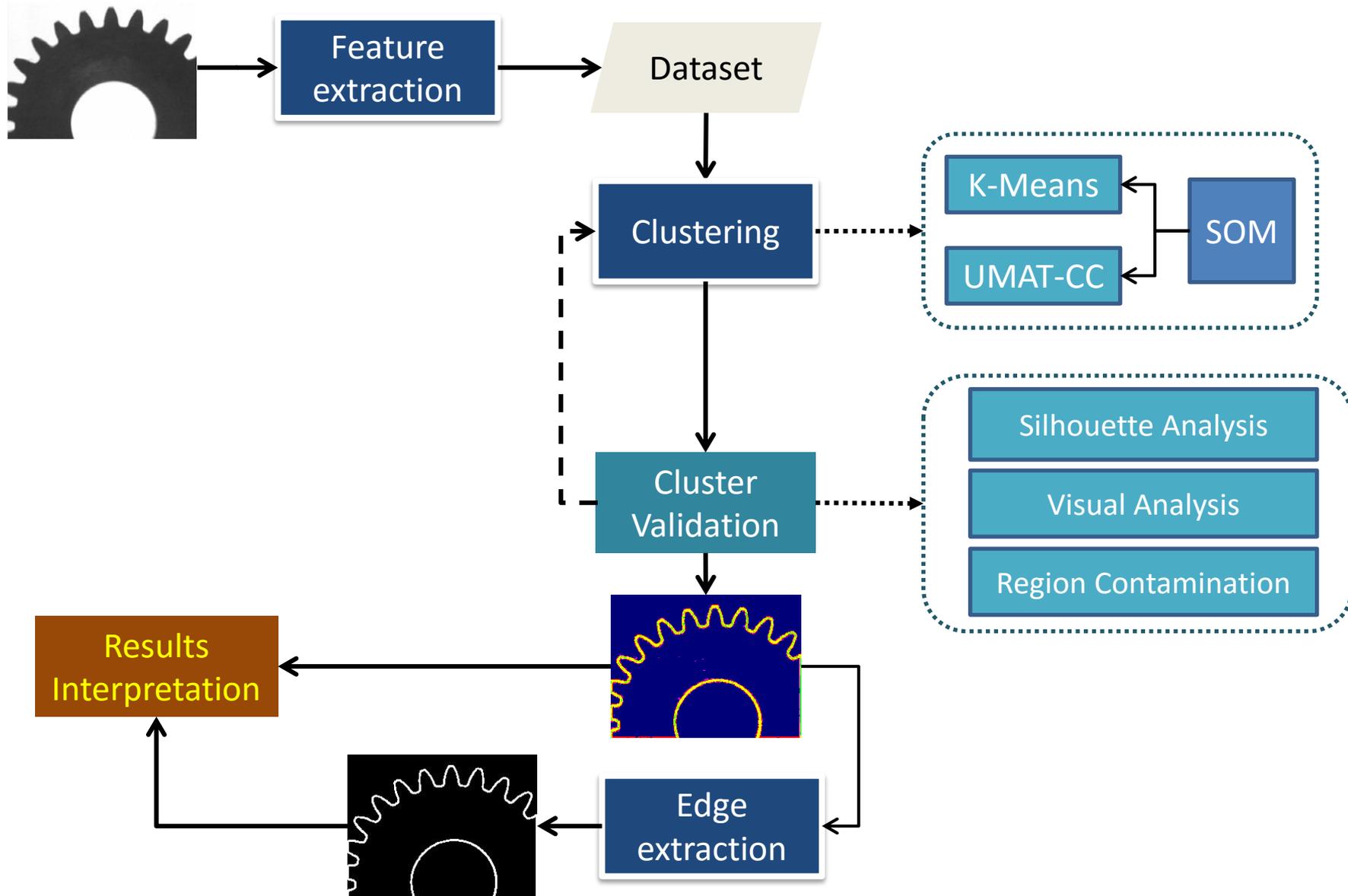


Rules	Input Variables								Output Variable
	$\Delta P_1$	$\Delta P_2$	$\Delta P_3$	$\Delta P_4$	$\Delta P_5$	$\Delta P_6$	$\Delta P_7$	$\Delta P_8$	$P$
1	Higher	Higher	None	None	None	None	None	Lower	Edge
2	Higher	None	None	High	None	None	None	Lower	Edge
3	None	Higher	Higher	None	None	None	None	Lower	Edge
4	None	None	None	Higher	None	Higher	None	Lower	Edge
5	Higher	Higher	None	None	None	None	Lower	None	Edge
6	Higher	None	None	Higher	None	None	Lower	None	Edge

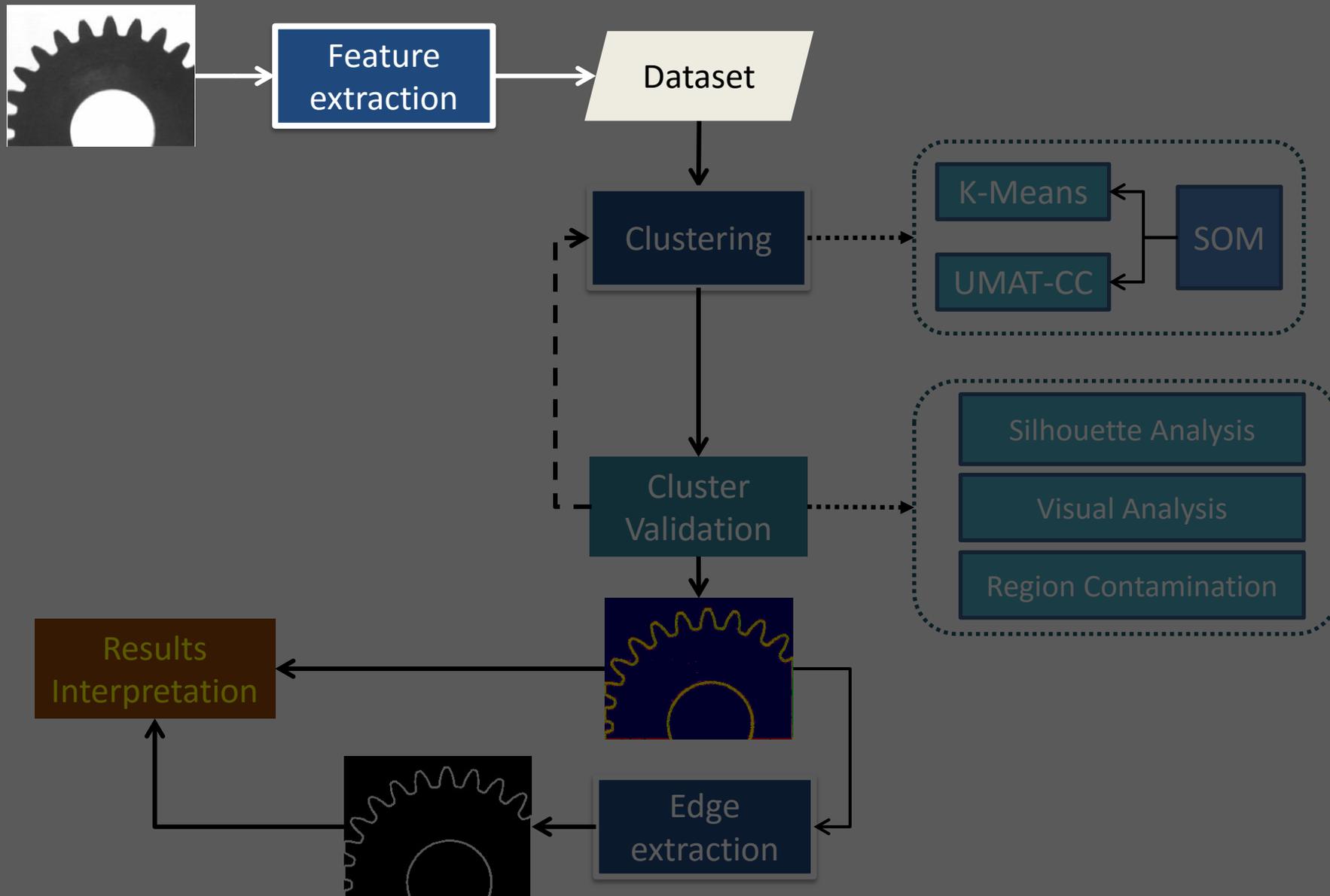
Defuzzification



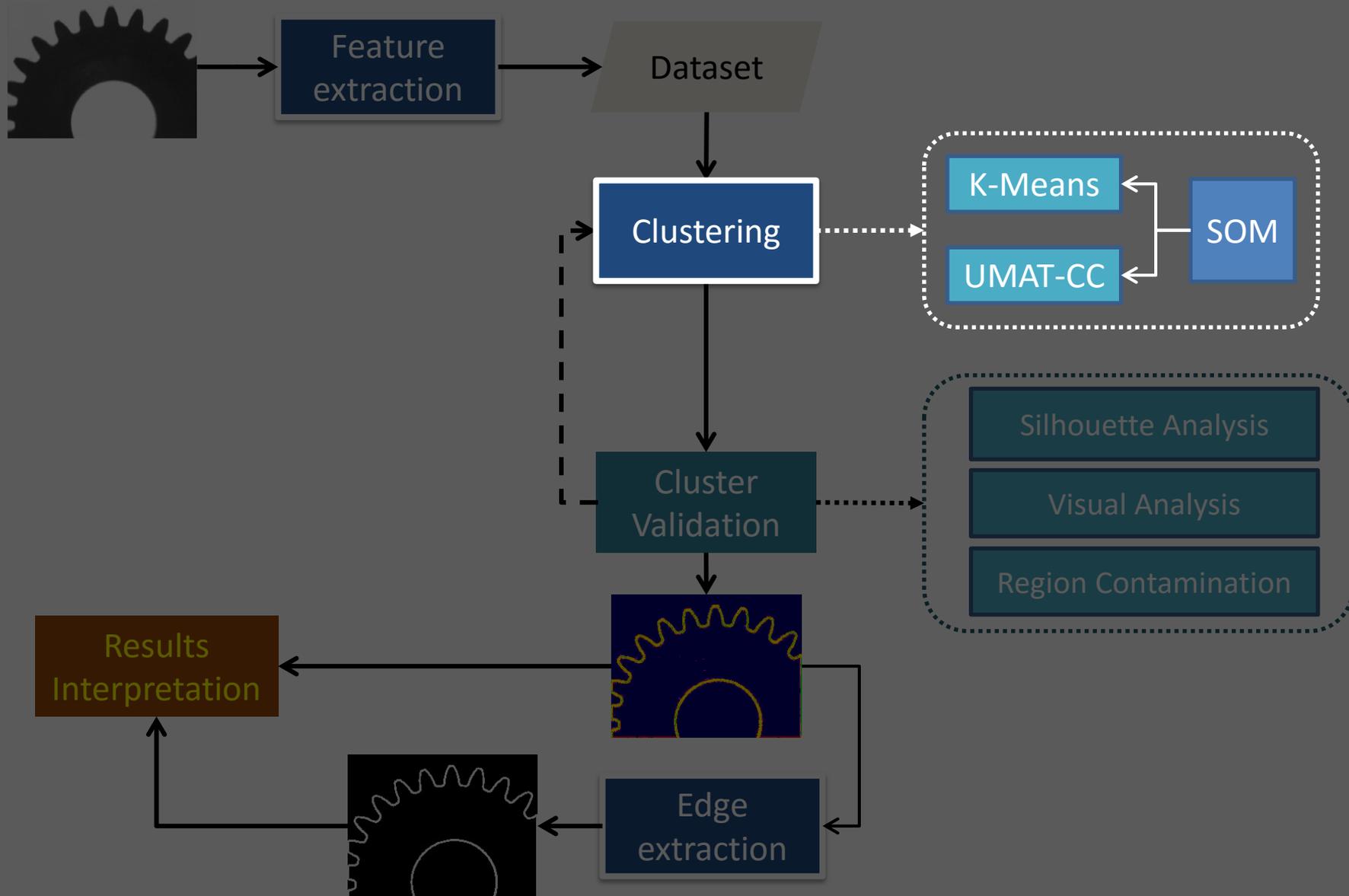
# Metodo Proposto



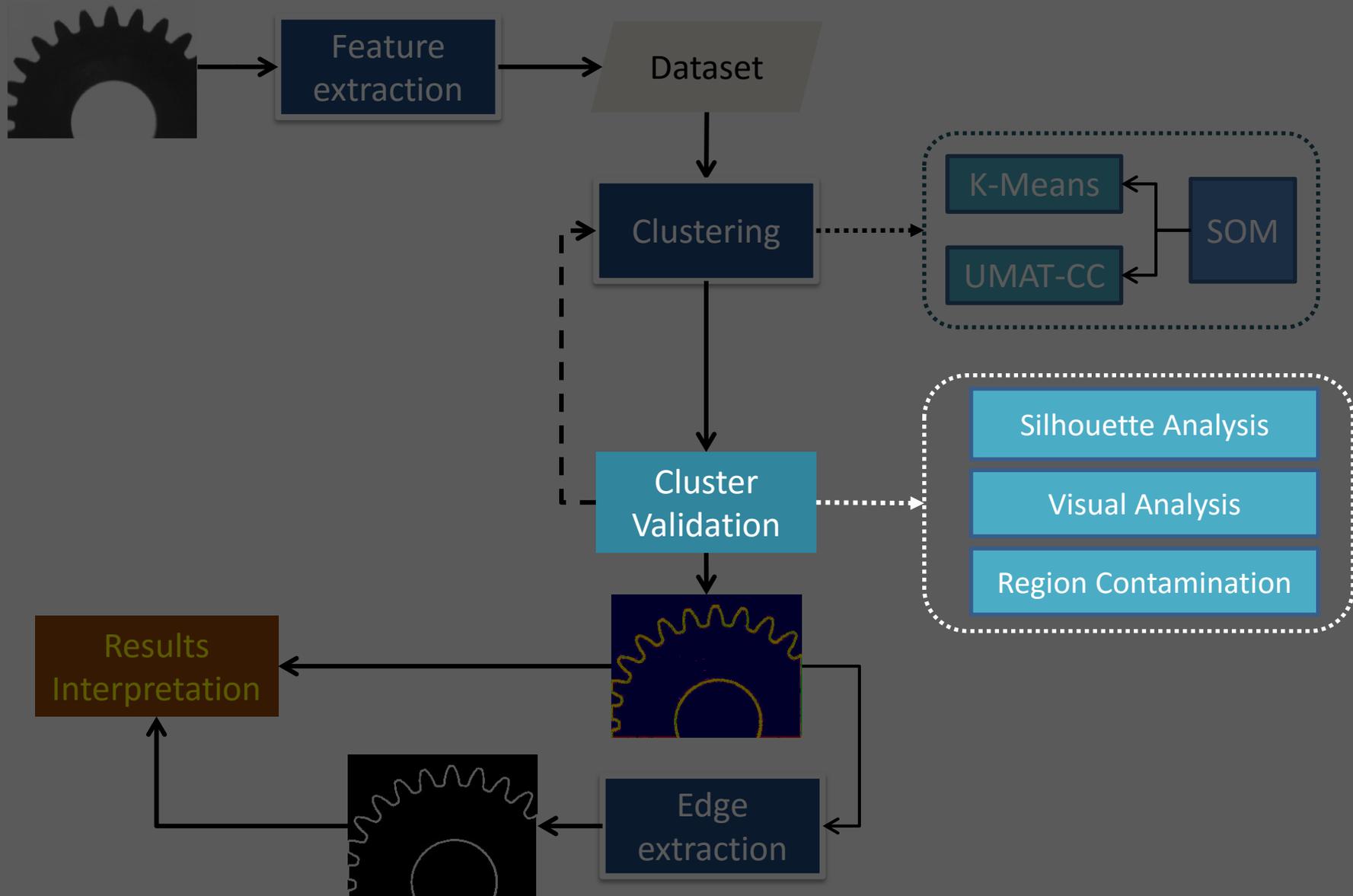
# Metodo Proposto



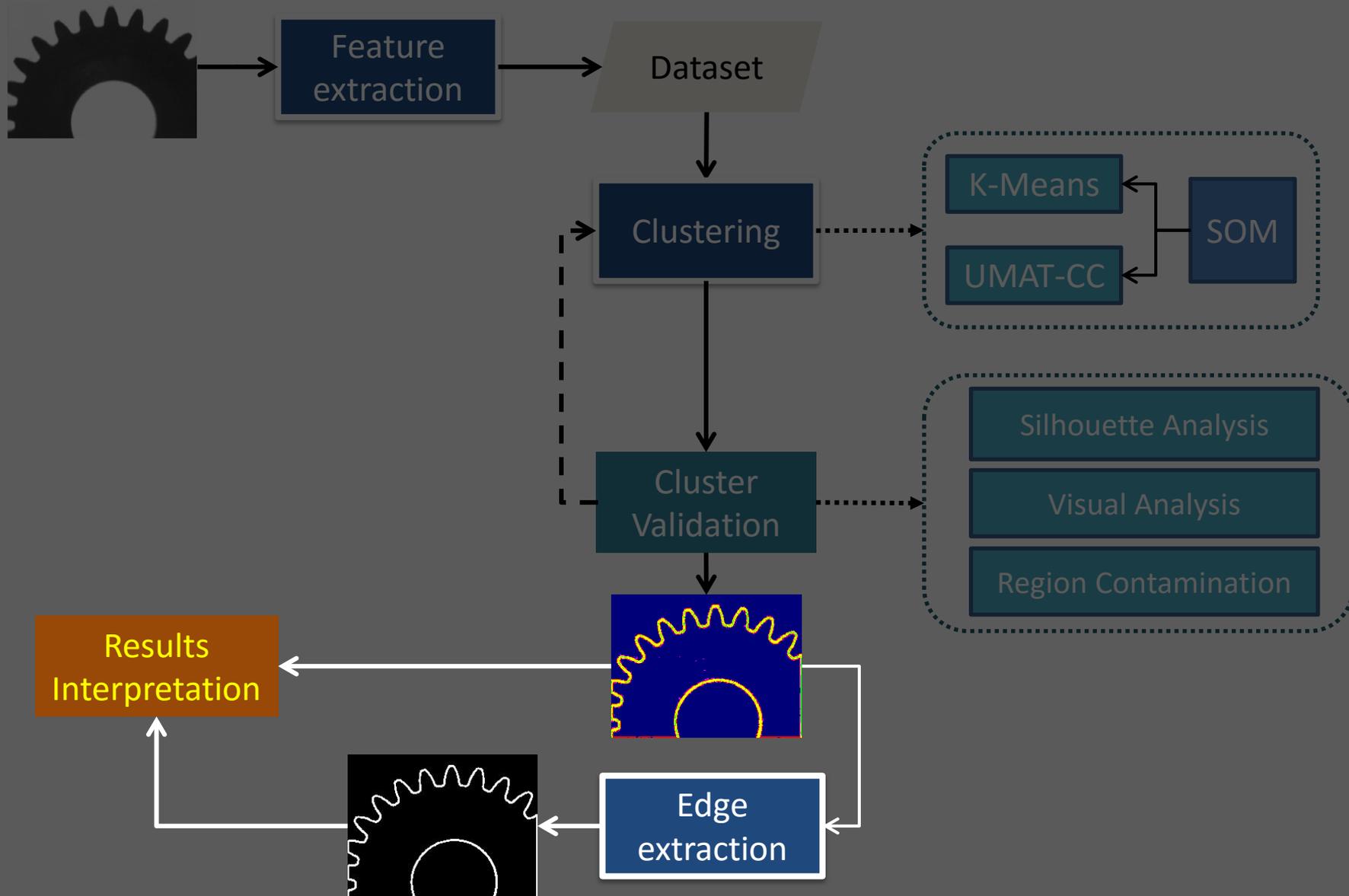
# Metodo Proposto



# Metodo Proposto

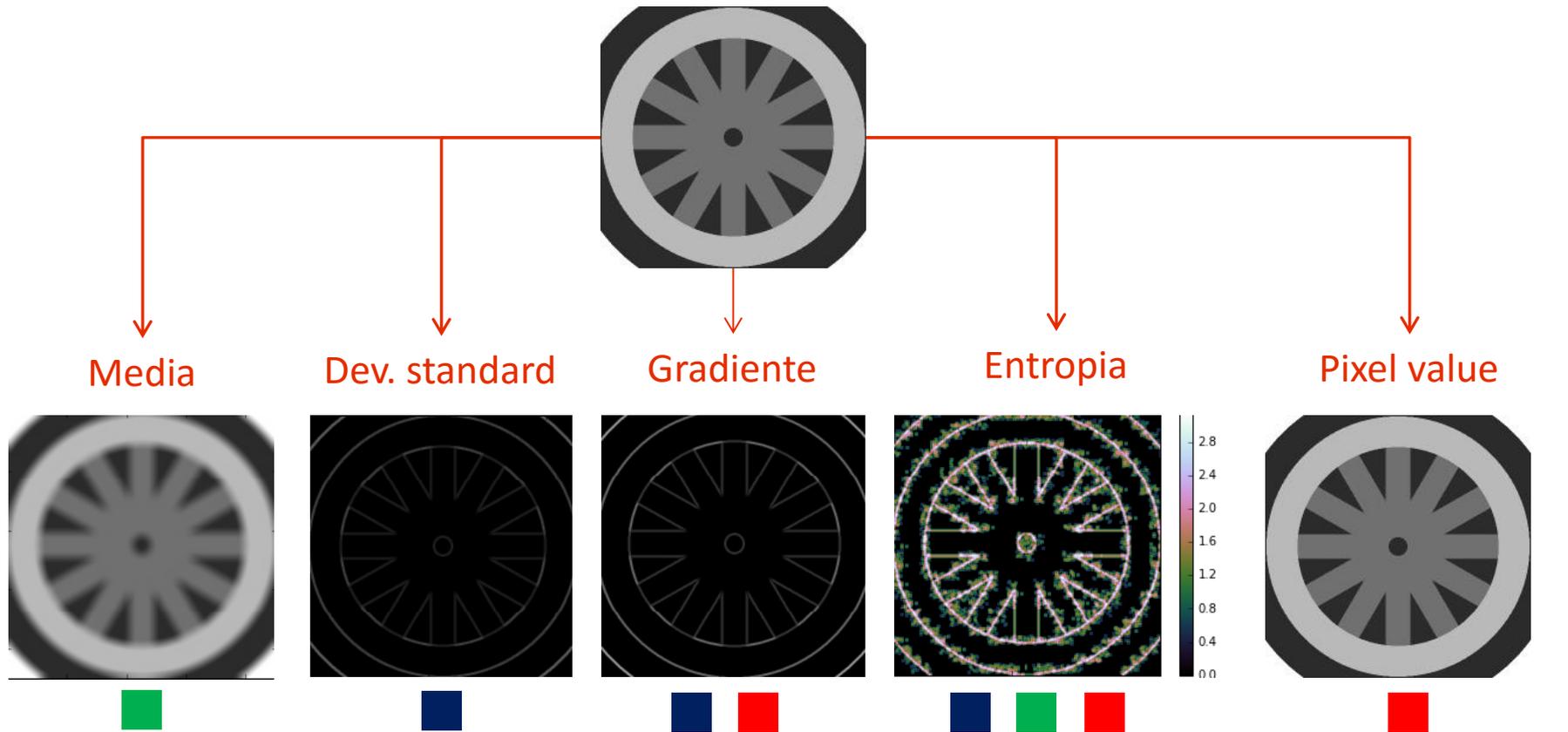


# Metodo Proposto



# Spazio dei parametri

Tramite il processo di feature extraction otteniamo, per ogni pixel, un N-vettore che lo identifica nello spazio dei parametri

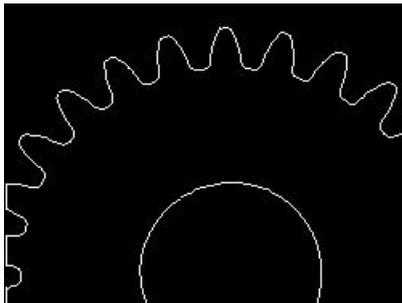


- Edge detection
- Segmentazione immagini osservate
- Segmentazione mappe d'estinzione

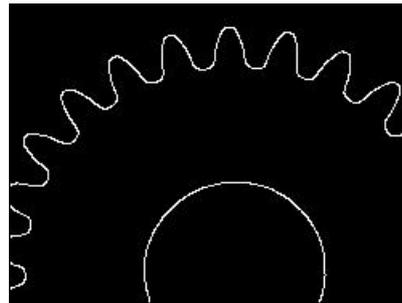
# Edge detection su immagini sintetiche



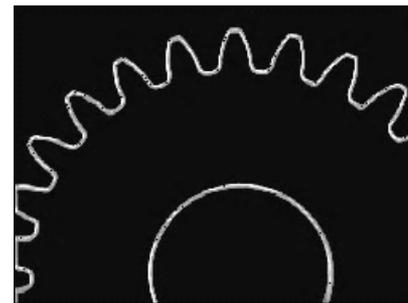
Sobel



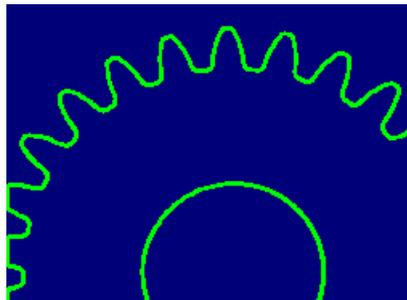
Canny



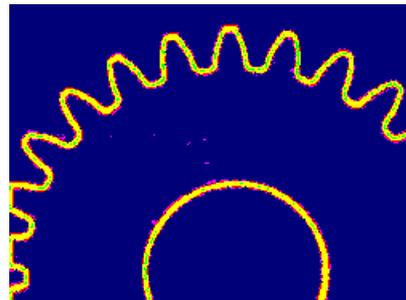
Fuzzy



Clustering

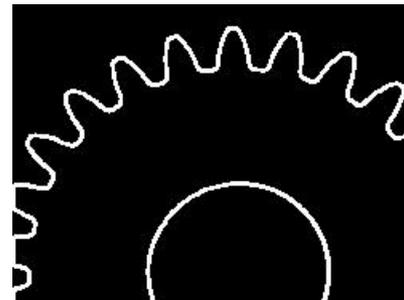


SOM + K-Means

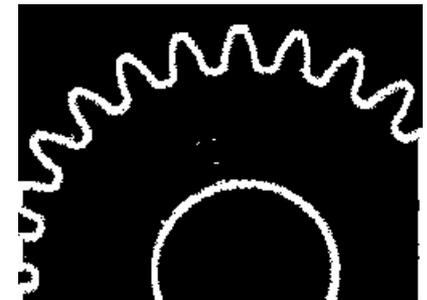


SOM + UMAT-CC

Edge Detection



SOM + K-Means



SOM + UMAT-CC

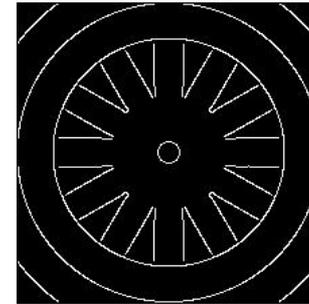
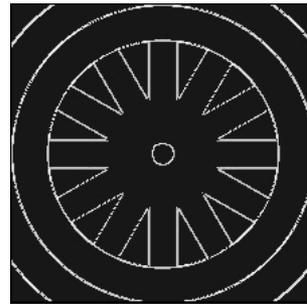
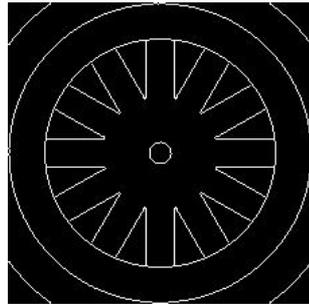
# Edge detection su immagini sintetiche



Sobel

Canny

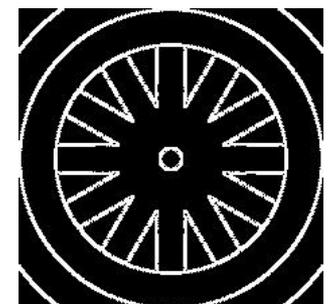
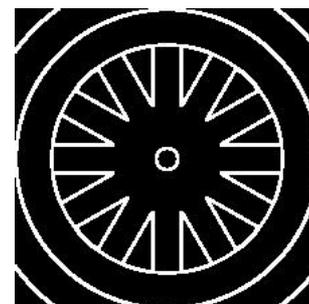
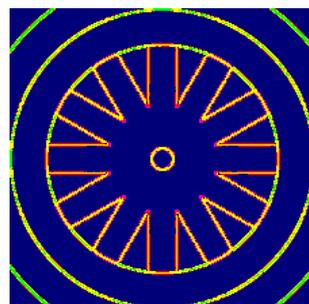
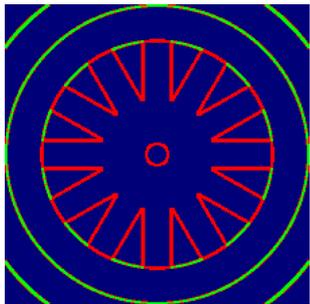
Fuzzy



Clustering



Edge Detection



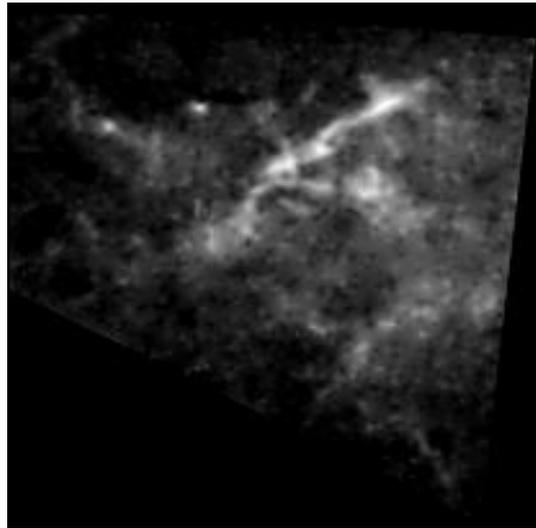
SOM + K-Means

SOM + UMAT-CC

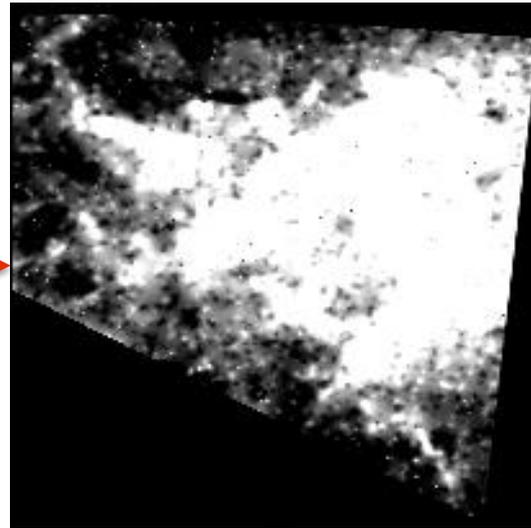
SOM + K-Means

SOM + UMAT-CC

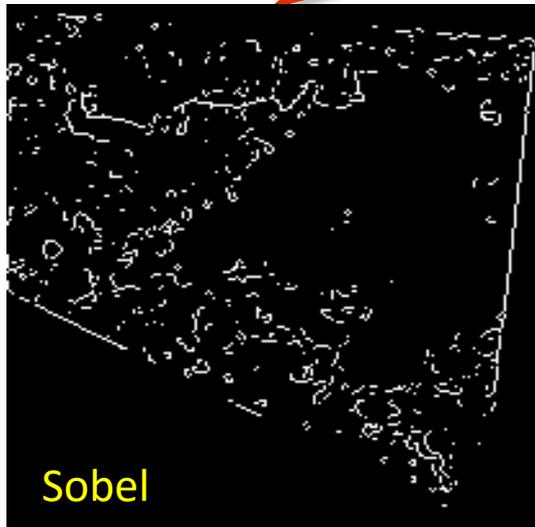
# Edge detection su mappe di estinzione



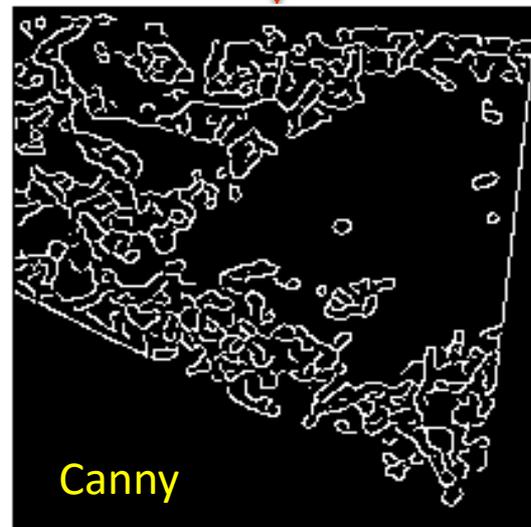
Aumento  
del  
contrasto



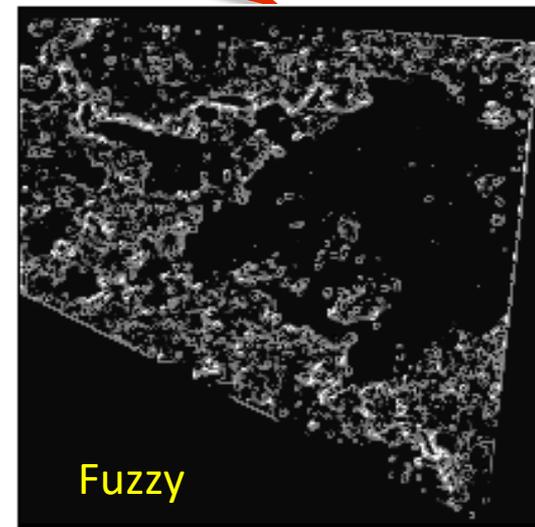
**Lupus I**



**Sobel**

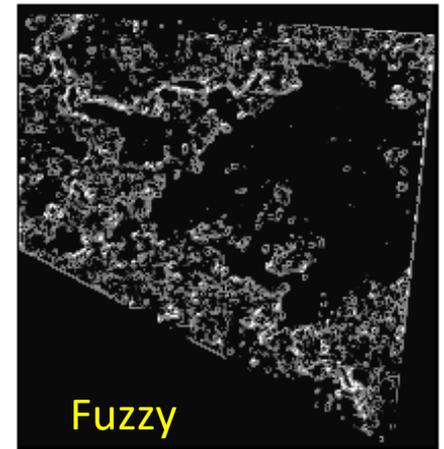
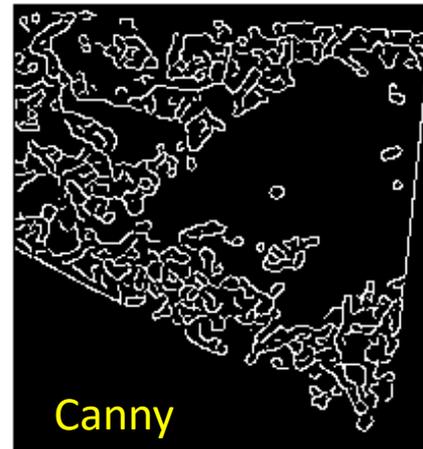
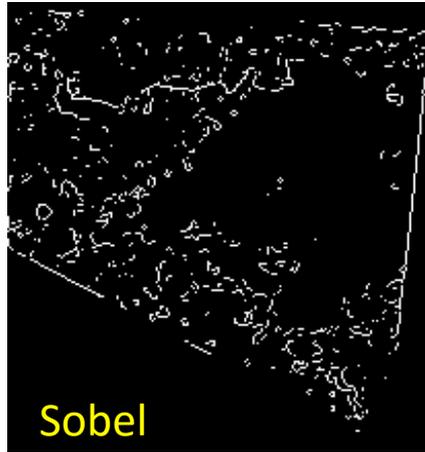
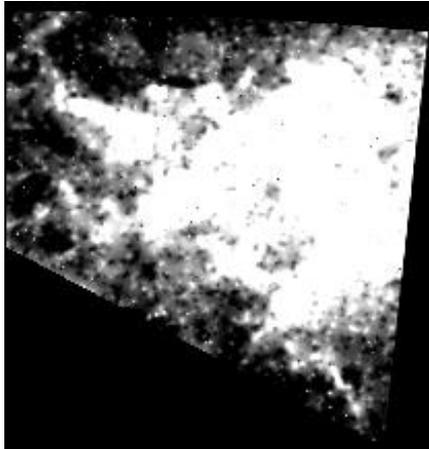


**Canny**

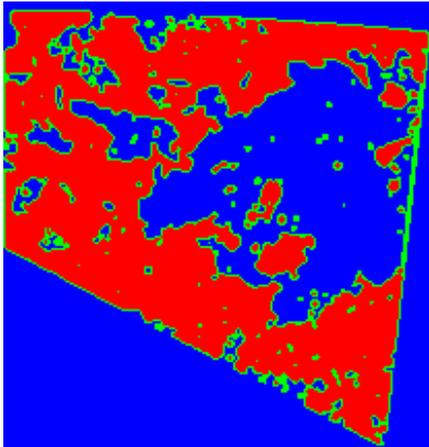


**Fuzzy**

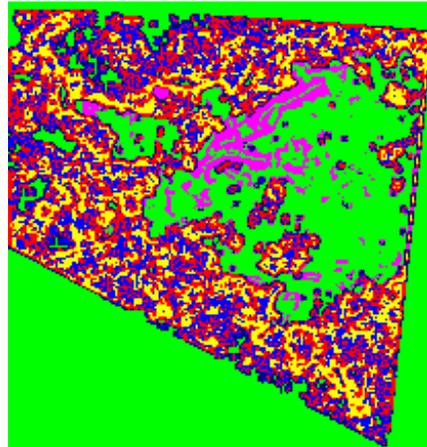
# Edge detection su mappe di estinzione



Clustering



SOM + K-Means

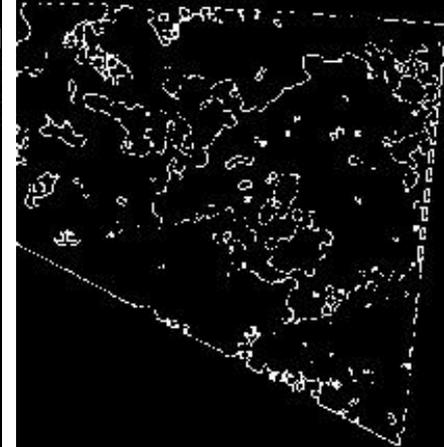


SOM + UMAT-CC

Edge Detection



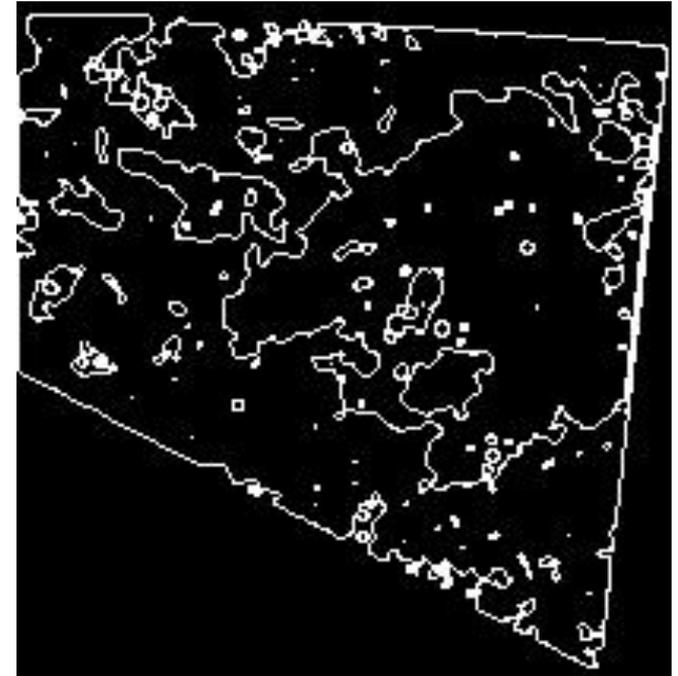
SOM + K-Means



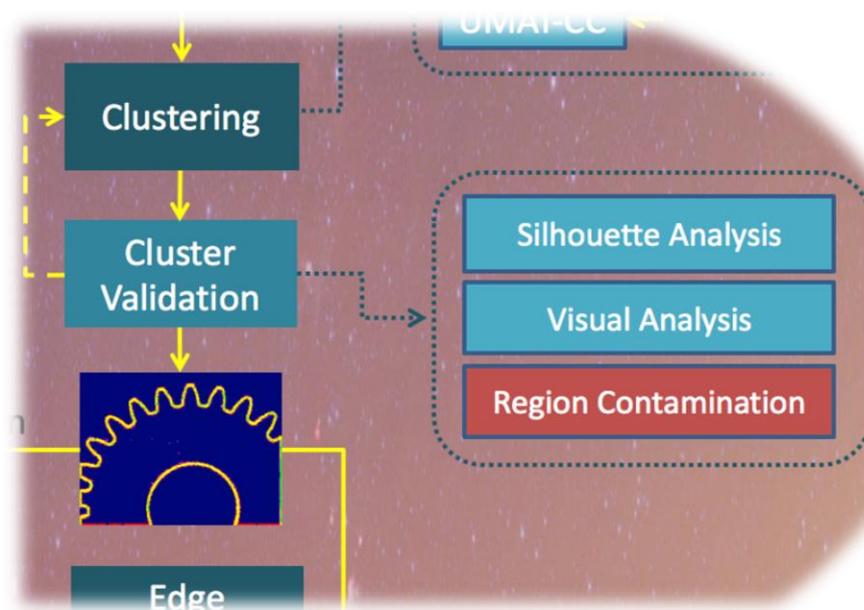
SOM + UMAT-CC

# Segmentazione su mappe di estinzione

Metodo in grado di circoscrivere delle regioni ma non di valutare la quantità di estinzione presente all'interno



Uso del valore del pixel (estinzione) come feature



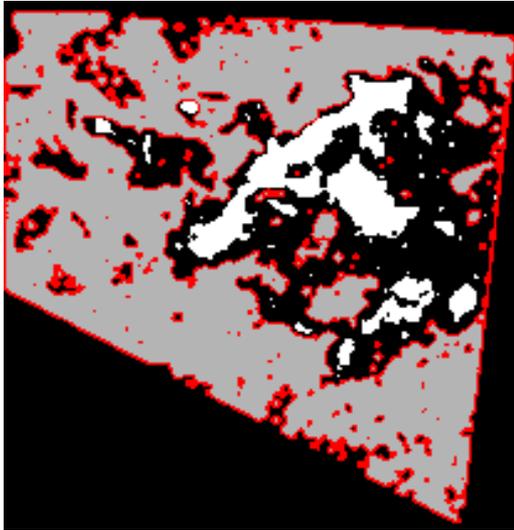
Metodo alternativo di validazione del clustering basato sul Region Contamination Index (RCI)

# Segmentazione su mappe di estinzione

Edge

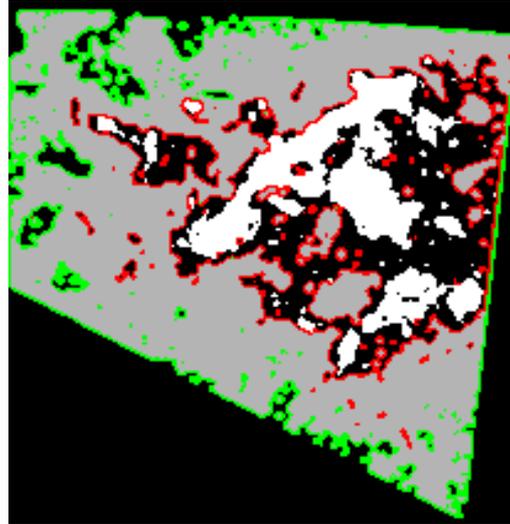
H    L

Scala di grigio in ordine decrescente secondo il valore medio dell'estinzione all'interno del cluster (H -> L)



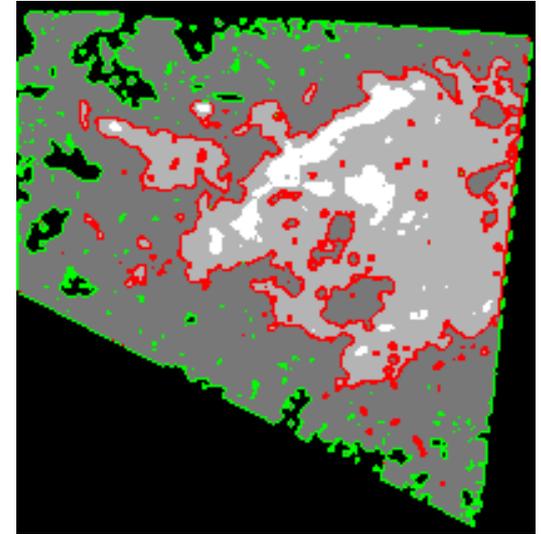
RCI 0.57

4



0.81

5



0

6

Region Contamination Index

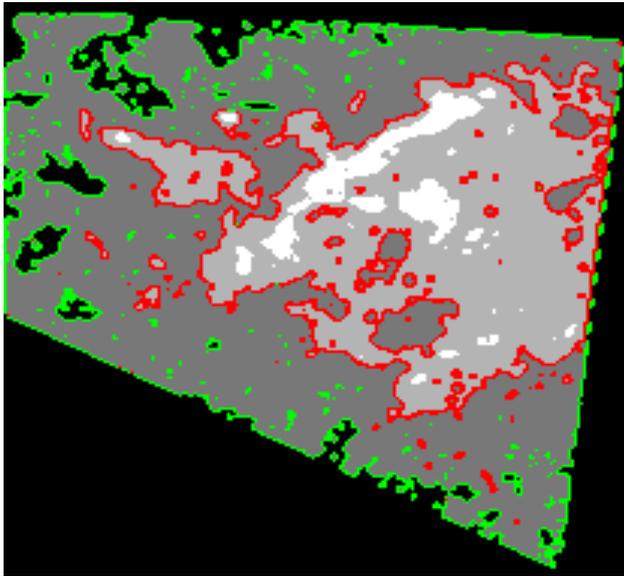
$$RCI = 1 - \frac{|P_H - P_L|}{P_H + P_L}$$

$P_H$  = N° di pixel ad alta estinzione (bianchi)

$P_L$  = N° di pixel a bassa estinzione (neri)

# Segmentazione su mappe di estinzione

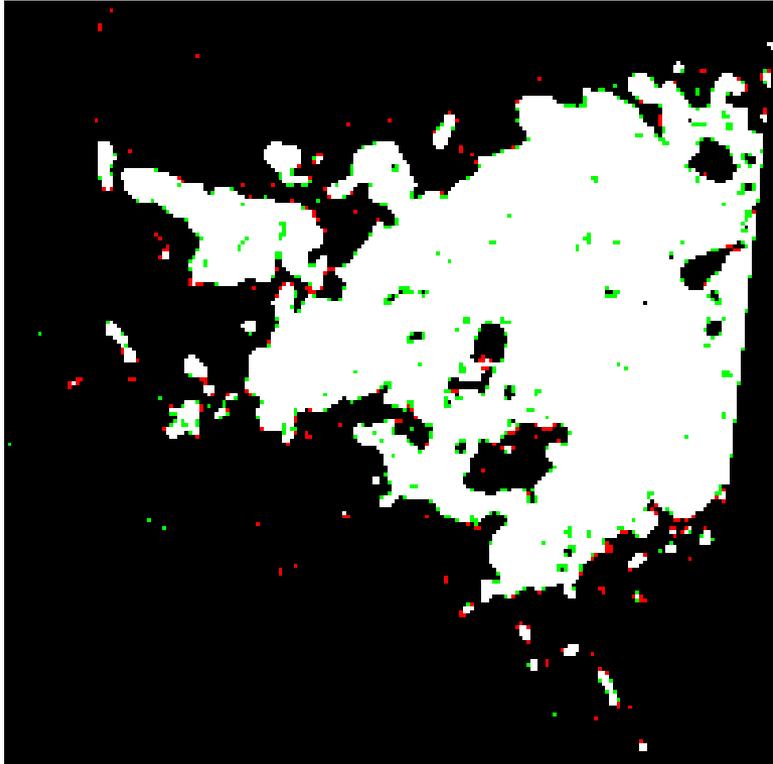
Dal risultato del clustering è possibile estrarre delle maschere binarie identificative delle regioni di estinzione.



E' possibile inoltre estrarre indicazioni aggiuntive: non esiste vincolo che limiti la coesistenza di differenti gradazioni di grigio, che non siano bianco e nero, all'interno della medesima regione



# Confronto tra regioni di estinzione



Thresholding a **0.98  $A_v$**

**Rapporto Massa**  
Clustering / Thresholding

$M_{Cambresy}$

$\sim +1.09\%$

$M_{Alcalà}$

$\sim +1.10\%$

**N° di pixel rispetto alla maschera di thresholding**

	In comune	11891
	Aggiunti	499
	Rimossi	221

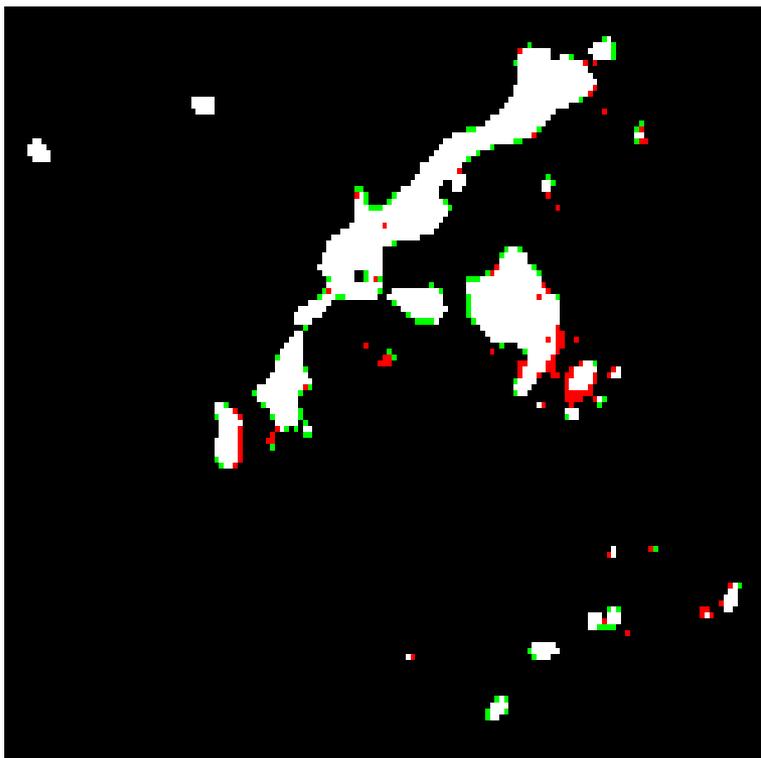
**Variazione numero pixel nella maschera**

+2.30%

**Estinzione media aggiunta**

$\sim 0.89$

# Confronto tra regioni di estinzione



Thresholding a  $2.5 A_v$

**Rapporto Massa**

Clustering / Thresholding

$M_{\text{Cambrèsy}}$

$\sim +0.32\%$

$M_{\text{Alcalà}}$

$\sim +0.32\%$

**N° di pixel rispetto alla maschera di thresholding**

	In comune	1151
	Aggiunti	110
	Rimossi	100

**Variazione numero pixel nella maschera**

$+0.80\%$

**Estinzione media aggiunta**

$\sim 2.44$

# Estrazione mappa di estinzione

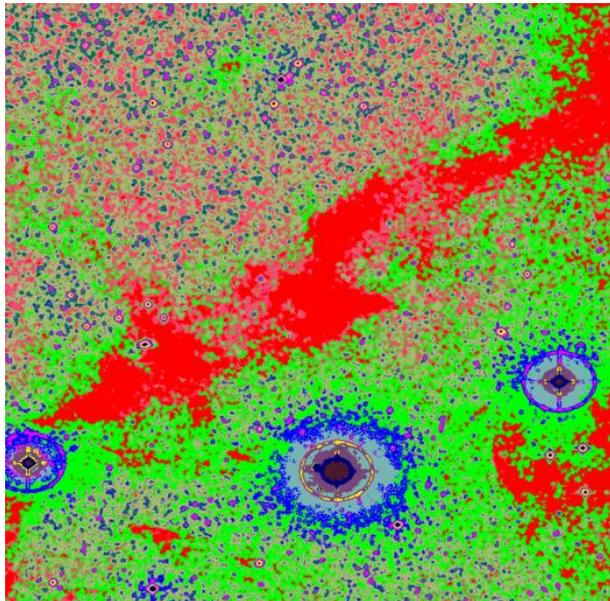


Partendo dall'immagine osservata è possibile, tramite il metodo di clustering proposto, estrarre una maschera che identifichi una RoI per il calcolo dell'estinzione

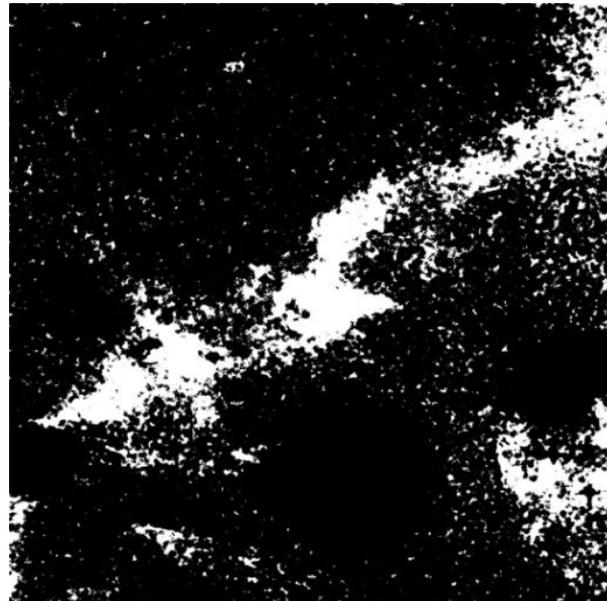
## Clustering vs thresholding

Clustering sulla mappa d'estinzione		Clustering sull'immagine osservata		threshold
$M_C$	$M_A$	$M_C$	$M_A$	$A_V$
$\sim +0.32\%$	$\sim +0.32\%$	$\sim -0.81\%$	$\sim -0.81\%$	2.5

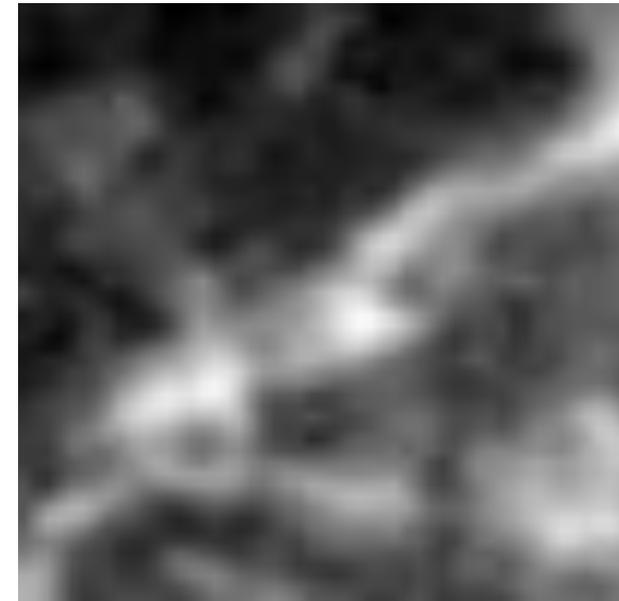
SOM + K-Means



RoI



Mappa di estinzione



# Conclusioni e sviluppi futuri

## Multi-Disciplinarietà

- Astrofisica
- Machine Learning
- Image Processing

## Complessità

Immagini ad alta risoluzione, basso SN Ratio, foreground/background noise, no ground truth assoluta

## Impatto Scientifico

Metodo con dimostrata validità confrontabile con i metodi tradizionali, con guadagno in termini di di automatizzazione e rispetto al quale rivela vantaggi esclusivi: maggior apporto informativo dei risultati con capacità di circoscrivere e delimitare le sotto-regioni a densità di estinzione variabile e risultato non assimilabile ad una sogliatura del livello di estinzione

## Sviluppi Futuri

Verifica più esaustiva dei parametri astrofisici derivati (ad es. la correlazione massa/luminosità) e approfondimento della rivelazione diretta dell'estinzione raffinando la scelta dello spazio dei parametri.

# Analisi greedy della Silhouette

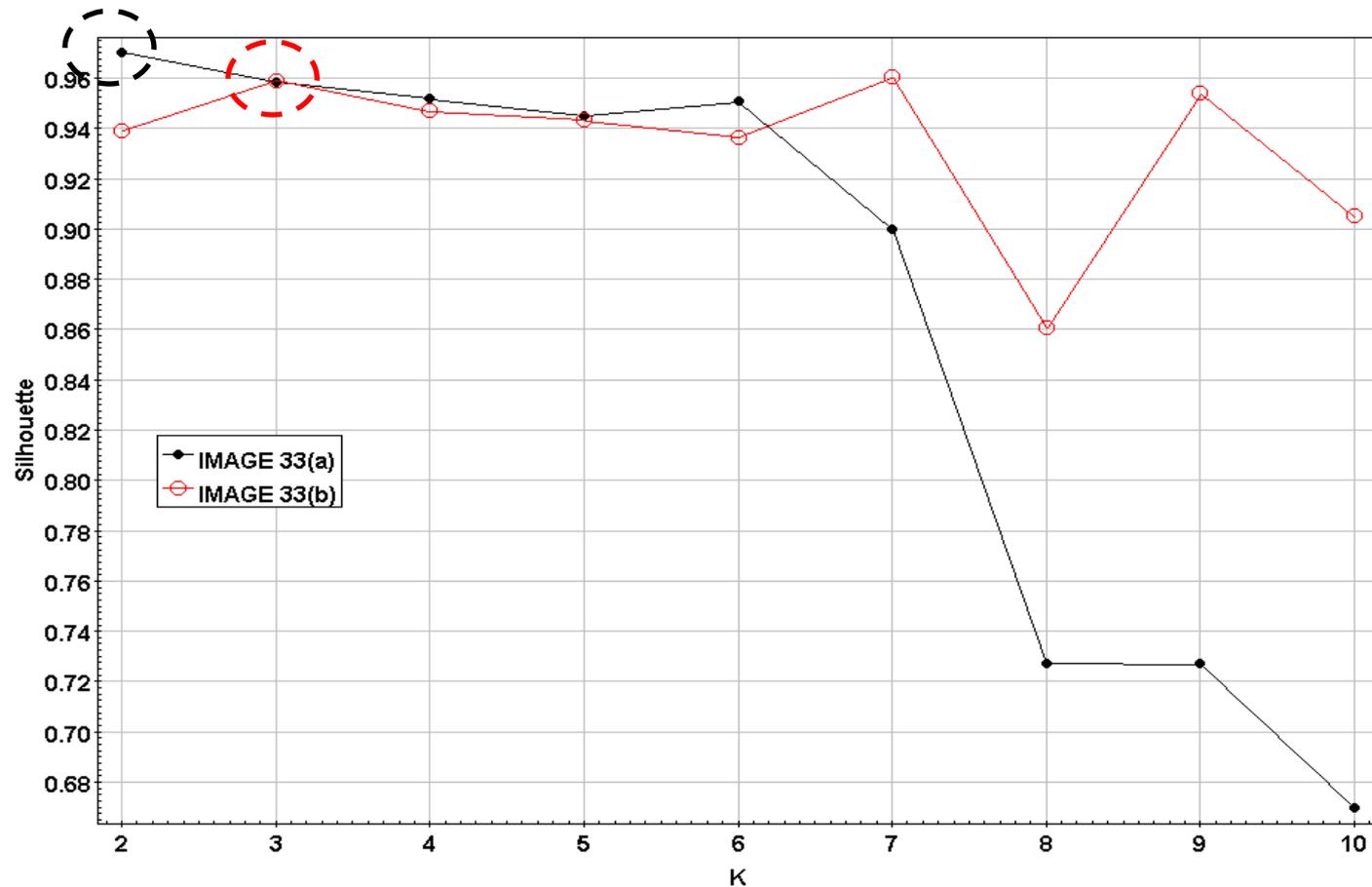
Tra i criteri più noti per la valutazione di un clustering figura l'analisi della Silhouette. E' definito criterio interno in quanto calcolato sulla base dei dati in input, valutando il rapporto tra la distribuzione *intra-cluster* e le distanze *inter-cluster*. La Silhouette utilizza le distanze medie tra punti e risulta fra i più accurati, sebbene ad un maggior costo computazionale [Petrovic 2006].



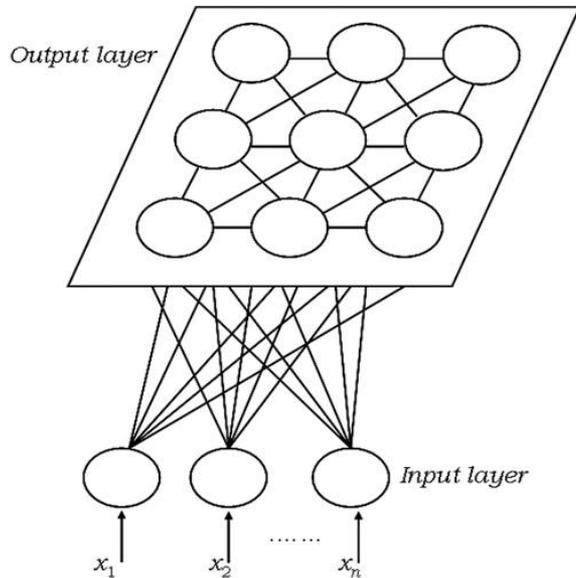
(a)



(b)

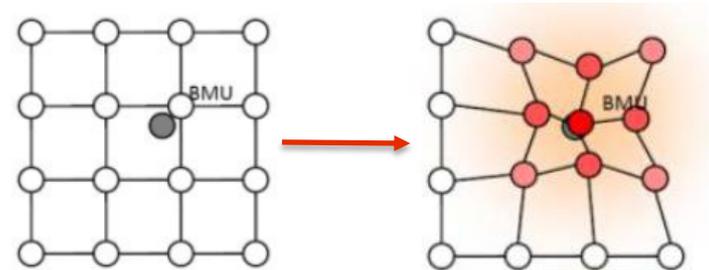


# Self Organizing Map

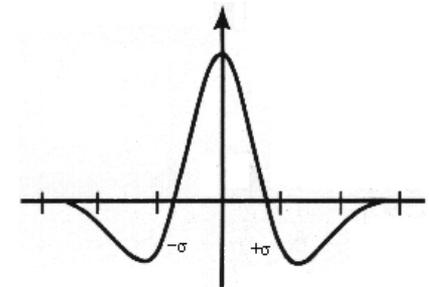


Ogni neurone dello strato riceve uno stimolo pari alla sommatoria degli input moltiplicati per il rispettivo peso sinaptico

I pesi del neurone vincente sono aggiornati, e quindi avvicinati, nello spazio dei parametri al vettore in input

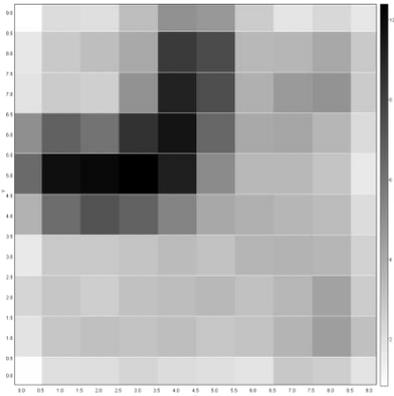


L'apprendimento è esteso ai nodi vicini secondo una funzione a cappello messicano che favorisce il formarsi di "bolle di attivazione" che identificano input simili

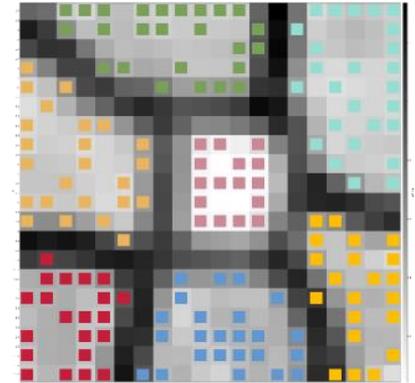


# U-Matrix A Componenti Connesse

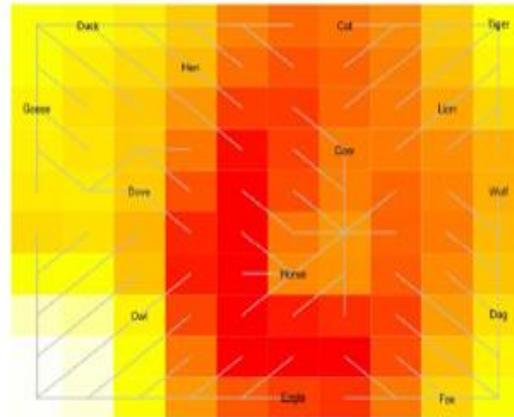
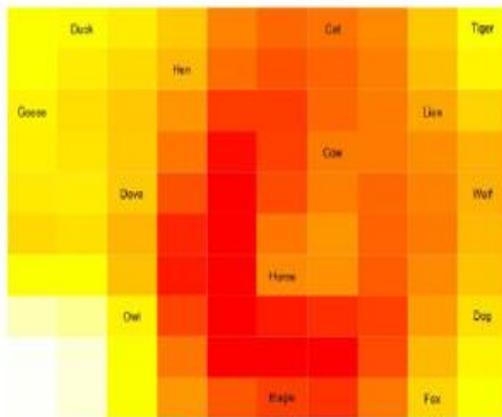
## U-Matrix



Strumento standard di visualizzazione di una SOM

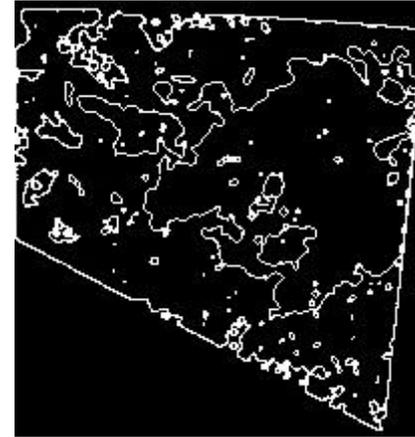
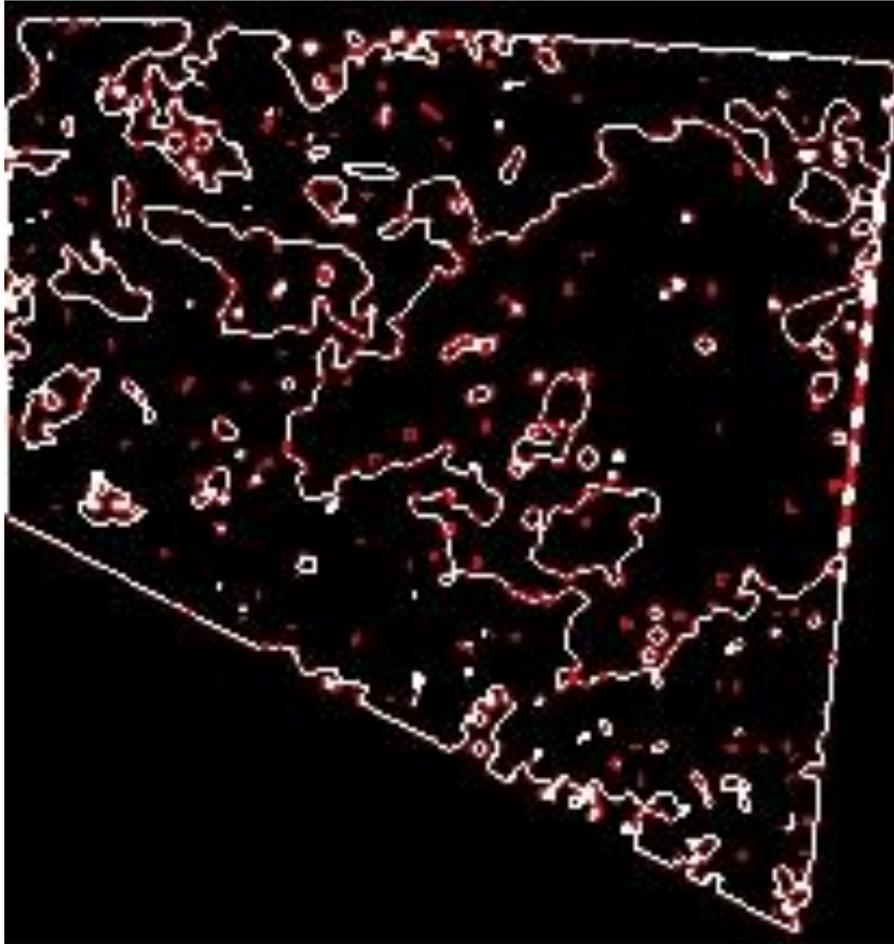


## Componenti Connesse

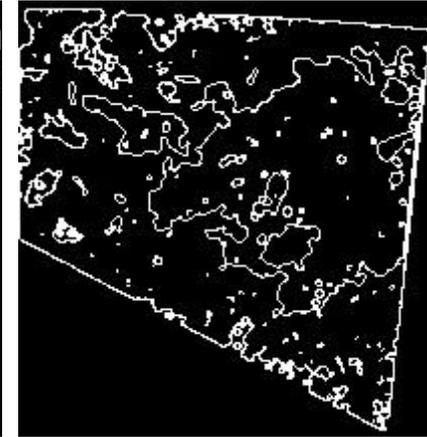


Per ogni nodo viene esaminato il valore, mostrato su heat map. Se il nodo esaminato non è quello con valore minimo allora ci si potrà muovere lungo un percorso attraversando di volta in volta il nodo con valore più basso e connettendo i nodi attraversati

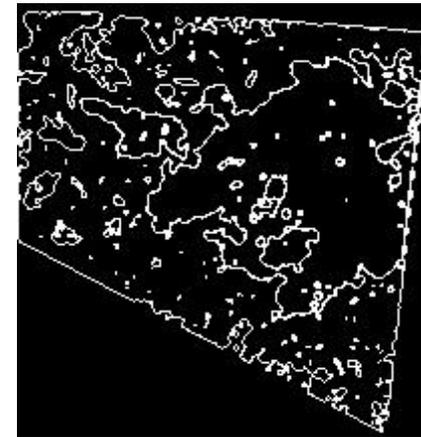
# Stabilità degli Edge



K = 4



K = 5



K = 6

L'aggiunta del valore dell'estinzione al dataset e l'aumentare del numero di cluster individuati non influenzano eccessivamente la stabilità degli edge individuati

- Pixel di edge non in comune
- Pixel in comune