

Classificazioni di immagini mediante algoritmi a reti neurali

Giuseppe ALONGE (*), Enrico BORGOGNO MONDINO (*), Giulio MAIONE (**),
Giuseppe TROPIANO (*), Daniele UGOLINI (*)

(*) Polo Nazionale Bioelettronica PST Elba, v. A. Moro 17 - Marciana Marina (LI) ,0565/99352

(**) Alenia Aerospazio - Divisione Spazio, v. A. Moro 17 - Marciana Marina (LI) ,0565/99352

Riassunto

L'*Information Technology Center* è la sezione del Polo Nazionale Bioelettronica destinata alla definizione di metodologie ed alla loro realizzazione mediante efficienti algoritmi diretti a servire realtà aziendali ad elevato contenuto tecnologico.

I progetti per ora in atto riguardano la definizione e l'applicazione di algoritmi a reti neurali per:

- a) il riconoscimento automatico della frutta dal punto di vista qualitativo
- b) la classificazione *supervised* e *unsupervised* di immagini satellitari pancromatiche e multispettrali
- c) lo studio multitemporale di immagini satellitari mirate al monitoraggio di alcuni aspetti ambientali (urbanizzazione, coste, agricoltura).

Abstract

The *Information Technology Center*, a section of Polo Nazionale Bioelettronica, is devoted to the development and the implementation of advanced and innovative algorithms, whose potential end-users are firms with a high technological profile.

The following three projects based on neural network application are under development:

- a) video-grading machine for fresh fruit
- b) supervised and unsupervised classification of panchromatic and multispectral remote sensing images
- c) multitemporal remote sensing image classification for environmental monitoring (urban, coastal, agriculture).

1. Selezione automatica di mele mediante ispezione visiva

Gran parte della lavorazione della frutta nelle centrali ortofrutticole è altamente automatizzata. Macchinari per l'automazione dei processi di calibrazione, di selezione per colore, di ceratura e di confezionamento sono ormai ampiamente utilizzati.

L'unica operazione che non è ancora stata automatizzata è la selezione per qualità, cioè la catalogazione del prodotto in base alle caratteristiche del suo aspetto (presenza o assenza di difetti sulla buccia).

I motivi per cui la selezione per qualità è ancora svolta manualmente o in modo semiautomatico è dovuto alla grande difficoltà che comporta rendere automatico questo processo. Tuttavia la selezione per qualità ha un ruolo molto importante in quanto permette di conferire caratteristiche di omogeneità

tipicamente industriali, secondo le richieste dell'utente finale, consumatore o industria, ad un prodotto che è naturalmente disomogeneo. L'omogeneità richiesta è relativa ad una serie di caratteristiche che si vuole siano comprese entro certi limiti o al di sopra di certi livelli.

Le mele sono classificate in tre categorie (I, II e III) in base alla presenza o assenza di difetti sulla superficie. L'aspetto più importante di questo processo è quello di distinguere il picciolo e il calice dai vari tipi di difetti (macchie, ammaccature, rugginosità).

Poiché in un'immagine digitalizzata il picciolo e il calice sono molto simili alle macchie, una semplice variazione locale tra pixel della stessa mela non permette di ottenere una corretta classificazione.

Inoltre i frutti sono presentati nella camera di visione in modo arbitrario, quindi non è possibile conoscere a priori la posizione del picciolo e del calice.

1.1 Descrizione del prototipo per acquisire le immagini

Il prototipo è composto da un nastro trasportatore, da una camera di visione e da un meccanismo di rotazione assiale delle mele.

La camera di visione è costituita da un box all'interno del quale sono presenti lampade ad alta frequenza e una telecamera CCD a colori. Per ottenere una maggiore diffusione della luce sono posti, di fronte alle lampade, dei pannelli di plexiglass.

Per ogni frutto sono acquisite in sequenza quattro differenti viste.

1.2 Descrizione algoritmi

Lo sviluppo di un sistema per l'automazione del processo di selezione in tempo reale per la frutta deve essere dotato di robustezza, velocità di calcolo (2/3 frutti al secondo) e capacità di selezione paragonabile a quella di un operatore umano. Quindi le soluzioni algoritmiche devono essere tali da non richiedere una grande quantità di calcolo, ma allo stesso tempo devono garantire la capacità di raggiungere gli obiettivi prefissati.

L'algoritmo di *grading* che si è definito si basa sull'utilizzo di reti neurali. Le tecniche di *image-processing* che comportano un elevato tempo di calcolo sono invocate solo quando necessarie. Infatti il primo passo consiste in una analisi globale molto veloce effettuata su tutte le viste del frutto, e solo se questo passo lo richiede si eseguono una o più analisi locali che comportano un maggior costo computazionale.

Poiché una grande percentuale dei frutti è di buona qualità, questa soluzione permette di ridurre i tempi di calcolo.

Le maggiori difficoltà nascono dal fatto che l'oggetto in esame è un prodotto della natura, per cui c'è una grande variabilità con cui i difetti si possono presentare sia nella forma sia nella tipologia. In generale risulta difficile riuscire a descrivere in termini matematici le caratteristiche intrinseche dei prodotti su cui si deve basare la selezione, ed in base a queste derivare una qualche funzione discriminante.

Diversamente è semplice costruire un database con un gran numero di "campioni" con i quali addestrare le reti neurali.

1.3 Definizione dell'architettura software

L'algoritmo è strutturato in tre diversi passi nei quali si prendono delle decisioni sulla qualità del frutto (v. fig. 1).

Nel primo passo vengono esaminate tutte le viste del frutto e solo se in una particolare immagine sono rilevate delle alterazioni sulla buccia, si continua l'ispezione. In questo primo passo la decisione è determinata da una rete neurale che classifica le *feature* estratte dall'istogramma.

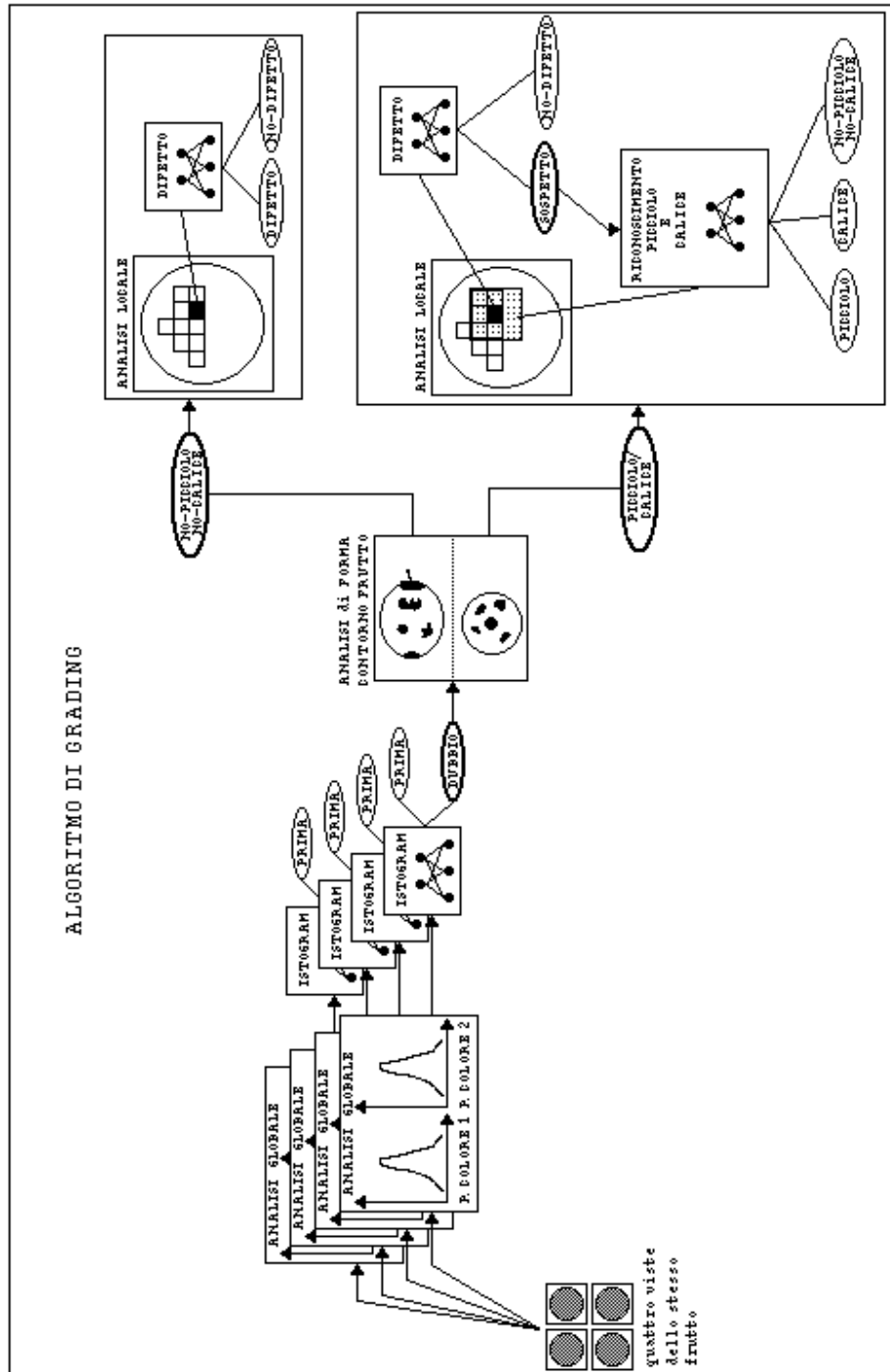


Figura 1. - Schema dell' algoritmo di grading

Il secondo passo consiste in un'analisi di forma del contorno della mela atta a stabilire la presenza del picciolo o del calice nella vista in esame. Questo passo viene eseguito solo nel caso in cui il precedente abbia stabilito la presenza di alterazioni sulla buccia.

Il terzo passo è composto da due azioni la cui scelta è determinata dal risultato della fase precedente:

1. no-picciolo, no-calice, nella vista in esame non è presente ne il picciolo ne il calice;
2. picciolo/calice, nella vista in esame è presente il picciolo o il calice.

Nel primo caso l'azione consiste nell'utilizzo di un operatore che permette di esaminare di volta in volta solo una porzione ridotta della superficie. Ogni regione è classificata come difetto o non-difetto usando un seconda rete neurale.

Nel secondo caso prima di classificare una regione come difetto si deve determinare se nella zona che si sta analizzando è presente il picciolo o il calice. Questa decisione è determinata da una terza rete neurale.

Una soluzione di questo tipo sembra essere più adatta per la frutta di forma sferica che ha una colorazione omogenea distribuita su tutta la superficie come ad esempio le arance e alcune varietà di mele ("Golden Delicious", "Red Delicious" e altre).

2. Classificazione di immagini telerilevate mediante algoritmi a reti neurali

Attualmente al PNB sono in fase avanzata di sviluppo due classificatori neurali: uno di tipo supervisionato (rete LVQ) ed uno di tipo non supervisionato (Neural Gas).

Riteniamo che l'approccio alla classificazione neurale sia più auspicabile delle metodologie tradizionali in virtù del fatto che, a differenza di queste, non richiede ipotesi preliminari circa la distribuzione di frequenza dei Numeri Indice (*DN*) dell'immagine.

La rete neurale impara da sé, adattandosi ad ogni tipo di distribuzione secondo le dinamiche sviluppate nella fase di *training*.

L'attenzione viene posta in questa sede sul classificatore unsupervised sviluppato al PNB, la cui originalità, nel campo del telerilevamento, sembra a tutt'oggi persistere e i cui risultati, pur richiedendo un maggior tempo di elaborazione, appaiono migliori delle tecniche tradizionali.

Resta inteso che dei classificatori non assistiti storici conserva pregi e soprattutto limiti legati all'interpretazione del risultato, la quale non può prescindere da un'attenta analisi da parte dell'operatore.

2.1 Classificazione unsupervised: rete Neural Gas

L'idea base è quella di applicare una metodologia a reti neurali con dimensionalità non costrittiva ma estremamente adattabile e capace di modellizzare al meglio il set di dati di addestramento (*training set*). Questo significa che la topologia dei nodi (o neuroni) della rete è assolutamente libera e non limitata all'interno di strutture fisse o semi-fisse (come reti SOM). La libertà di movimento dei nodi all'interno del training set conferisce alla rete un comportamento molto simile a quello di molecole gassose, da cui il nome.

2.1.1 Algoritmo

L'algoritmo (Fritzke, 1997) prevede due fasi:

- una fase di *training* nella quale vengono stimati i pesi dei nodi della rete
- una fase di classificazione in cui la rete addestrata sul *training-set* viene applicata alla scena da classificare

La rete, come tutte le reti neurali, è definita una volta che siano stati definiti:

- il numero dei nodi della rete, quindi dei *cluster* che si intende generare
- la dimensione del vettore di ingresso, cioè il numero delle bande
- la funzione per la stima dei pesi dei nodi
- la dinamica di competizione tra nodi al fine di ottenere l'assegnamento degli ingressi

2.1.2 Training

- Viene generato un *training set*, ottenuto campionando l'immagine con procedimento casuale.
- Ogni pixel campionato sulle n bande dell'immagine multispettrale costituisce il vettore di *input* ε per la rete (cioè per ogni nodo).
- Viene definita la matrice dei pesi $A = (w_1, w_2, \dots, w_N)$ dove $w_i \in \mathbb{R}^n$ è il vettore-peso (di dimensione pari al numero di bande considerate) associato al generico nodo i -esimo ed N è il numero dei *cluster*. Poichè il processo di addestramento è iterativo è richiesta una inizializzazione di tali vettori, che avviene definendone le componenti (associate alle diverse bande dell'immagine) come media aritmetica per banda dei DN dei *pixel* appartenenti al *training set*.
- Viene definito un parametro temporale t inizializzato a $t = 0$, il quale partecipa alla funzione per la correzione iterativa dei pesi. Il numero massimo di iterazioni (t_{max}) è pari alla dimensione del campione casuale utilizzato per il *training* (la dimensione del *training set* è a discrezione dell'operatore)
- Ad ogni step del procedimento iterativo di stima dei pesi:
 1. I vettori-peso vengono ordinati in base alla loro distanza euclidea dal vettore di *input* ε , viene definita una sequenza di indici (i_0, i_1, \dots, i_{N-1}) tale per cui w_{i_0} è il vettore-peso più vicino a ε e w_{i_k} (con $k = 0, 1, \dots, N-1$) è il vettore-peso per cui esistono k vettori w_j con $\|\varepsilon - w_j\| < \|\varepsilon - w_k\|$. Si assume $k_i(\varepsilon, A)$ come il numero k associato con il vettore-peso w_i .
 2. Viene calcolata la correzione vettoriale da applicare ad ogni vettore dell'ordinamento secondo la seguente regola di apprendimento:

$$\Delta w_i = \varepsilon(t) \cdot h_\lambda(k_i(\varepsilon, A)) \cdot (\varepsilon - w_i) \quad [1]$$

con le seguenti dipendenze temporali:

$$\lambda(t) = \lambda_i (\lambda_f / \lambda_i)^{t/t_{max}} \quad [2]$$

$$\varepsilon(t) = \varepsilon_i (\varepsilon_f / \varepsilon_i)^{t/t_{max}} \quad [3]$$

$$h_\lambda(k) = e^{-k/\lambda(t)} \quad [4]$$

I valori $(\lambda_i, \varepsilon_i)$ e $(\lambda_f, \varepsilon_f)$ sono definiti empiricamente. Nel caso specifico si sono utilizzati i seguenti valori (Martinetz et al., 1993): $\lambda_i = 10$, $\lambda_f = 0.01$, $\varepsilon_i = 0.5$, $\varepsilon_f = 0.005$

3. I vettori-peso vengono quindi corretti secondo il Δw_i corrispondente ($w_i = w_i + \Delta w_i$)
4. Viene incrementato il parametro temporale : $t = t + 1$
5. Se $t < t_{max}$ viene avviata l'iterazione successiva.

2.1.3 Classificazione vera e propria

Una volta stimati i pesi per ogni nodo della rete (che assume funzione di centroide per i diversi *cluster* che si intende generare) viene avviata la procedura di classificazione sull'intera immagine. Per ogni pixel dell'immagine viene calcolata la distanza euclidea tra il vettore che lo rappresenta ed il vettore-peso associato ad ogni nodo (risultato della fase di *training*). Il nodo cui compete la minore distanza euclidea risulta essere il vincente ed alla classe da questo rappresentata viene associato il *pixel* in esame.

2.2 Diretrici di sviluppo

L'affinamento della tecnica è tuttora in atto e mira essenzialmente a definire:

- una procedura ragionevole per l'individuazione automatica del miglior numero di classi (*cluster*) e quindi di nodi da adottare,
- criteri validi per il controllo dei parametri (dimensione del campione, convergenza dell'algoritmo),
- procedure di autogenerazione o di generazione controllata di sotto-reti all'interno di cluster giudicati troppo dispersivi e che si ritiene possano a loro volta essere classificati,
- metodologie per individuare zone critiche di classificazione (punti di frontiera) soprattutto per immagini ad elevato numero di bande.

3. Change Detection

Parte delle risorse del PNB sono impegnate nello sviluppo di metodologie per l'analisi multitemporale di immagini satellitari. In particolare è in fase di sviluppo un algoritmo che sintetizza procedimenti tradizionali con accorgimenti innovativi per l'estrazione di tematiche legate alla variabilità della scena. Questo prevede:

- 1) Una fase di georeferenziazione *image-to-image* mirata ad ottenere un'accurata corrispondenza tra pixel omologhi delle due immagini che partecipano allo studio.
- 2) Una fase di correzione radiometrica mirata a normalizzare le immagini rispetto alle condizioni di illuminazione mediante la definizione di una regressione lineare (o di ordine superiore) tra le immagini e generata su una porzione di queste ritenuta radiometricamente invariante su entrambe (centri abitati o zone d'acqua possibilmente). Nell'ipotesi che la regressione stabilisca la relazione tra le mutate condizioni di illuminazione, essa viene applicata ad una delle due immagini scelta come variabile indipendente ottenendone una nuova che rappresenta la stessa scena riportata a condizioni di illuminazione analoghe a quelle dell'immagine dipendente.
- 3) Una fase di *image differencing* o di *image rationing* tra la nuova immagine e l'immagine dipendente (che è restata immutata) intesa ad estrarre le differenze reali.
- 4) Una fase di visualizzazione del risultato intesa a consentirne un'immediata sovrapposizione sulle immagini di partenza.

E' in fase di studio un'estensione/modifica dell'algoritmo, che per ora prevede il processamento di immagini pancromatiche o monobanda, a immagini multispettrali.

Bibliografia

- Fritzke, B. (1997) - *Some Competitive Learning Methods* - Draft - Systems Biophysics - Institute for Neural Computation - Ruhr-Universitat Bochum, 20-21
- Recce M., Taylor J., Plebe A., Tropiano G. (1997) - *Video Grading of Oranges in Real Time* - Artificial Intelligence Review - Special Issue on AI Applications in Biology and Agriculture
- Recce M., Taylor J., Plebe A., Tropiano G. (1996) - *High Speed Vision-Based Quality Grading of Oranges* - 1996 International Workshop on Neural Networks for Identification, Control, Robotics, and Signal/Image Processing, Agosto 21-23, 1996, Venezia, 136-144