



On AIR s.r.l.

Tecniche di riconoscimento statistico

Applicazioni alla lettura automatica di testi
(OCR)

Parte 5 – Tecniche OCR

Ennio Ottaviani

On AIR srl

ennio.ottaviani@onairweb.com

<http://onairweb.com/corsoPR>

A.A. 2008-2009

Introduzione

- L'obiettivo delle tecniche OCR è classificare dei pattern 2D (espressi come immagini digitali) in simboli alfanumerici, ottenendo una versione digitale (ASCII) di una immagine contenente testo
- Le tecniche utilizzate combinano aspetti di tipo statistico con aspetti di elaborazione delle immagini
- Applicazioni tipiche:
 - Data entry di archivi cartacei
 - Smistamento della corrispondenza
 - Lettura di moduli
 - Riconoscimento targhe



Il processo OCR

- Un qualsiasi processo OCR viene di norma suddiviso in 4 passi base che si ritrovano in tutte le applicazioni
 - Localizzazione del testo
 - Segmentazione dei caratteri
 - Classificazione dei caratteri
 - Analisi contestuale
- I primi due sono strettamente connessi alla problematica applicativa e coinvolgono metodi di elaborazione delle immagini
- La classificazione dei caratteri è un problema classico di PR
- L'analisi contestuale impiega metodi di ragionamento simbolico



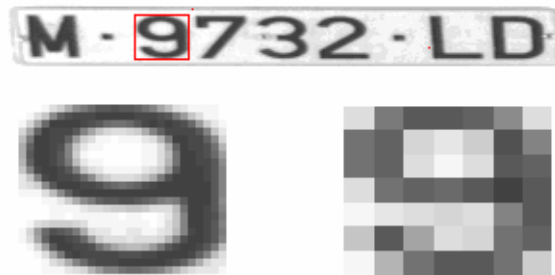
Features per OCR

- Dal punto di vista del PR, le applicazioni OCR richiedono innanzitutto una definizione delle *features* da utilizzare
- Le features ideali devono rispondere a diversi requisiti:
 - Computabilità
 - Discriminatività
 - Invarianza
- Il concetto di invarianza assume particolare importanza nelle applicazioni OCR dove il testo da leggere può essere deformato:
 - geometricamente (ruotato, scalato)
 - pittoricamente (chiaro, scuro)



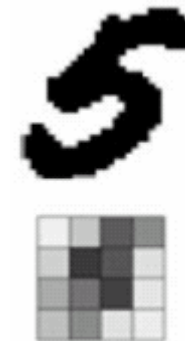
Zonatura

- Il metodo della zonatura prevede che l'immagine 2D a livelli di grigio del carattere da riconoscere venga tassellata in $M \times N$ regioni disgiunte rettangolari, e di ciascuna venga misurato il grigio medio



- Il metodo è adatto per immagini binarie, mentre nel caso generico occorre una normalizzazione dei livelli di grigio prima della zonatura

- La scelta di M, N è critica
 - valori piccoli → poco discriminanti
 - valori grandi → computazione onerosa



Momenti

- Uno schema per generare *feature* invarianti per trasformazioni assegnate di una immagine $f(x,y)$ si basa sui momenti

$$m_{pq} = \sum_{x=0}^{W-1} \sum_{y=0}^{H-1} x^p y^q f(x, y)$$

- L'immagine viene interpretata come una distribuzione di probabilità
- Se le coordinate (x,y) sono riferite al baricentro $(m_{10}/m_{00}, m_{01}/m_{00})$, si parla di momenti centrali μ_{pq} (invarianti per traslazione)
- I momenti centrali del secondo ordine misurano le varianze di x,y . I momenti di ordine superiore (normalizzati), sono utili per la classificazione



Momenti invarianti

- Descrittori di forma

$$\Theta = \frac{1}{2} \arctan \left(\frac{2\mu_{11}}{\mu_{20} - \mu_{02}} \right) \quad \lambda_i = \frac{\mu_{20} + \mu_{02}}{2} \pm \frac{\sqrt{4\mu_{11}^2 + (\mu_{20} - \mu_{02})^2}}{2} \quad e = \sqrt{1 - \frac{\lambda_2}{\lambda_1}}$$

- Invarianti di scala

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^\gamma} \quad \gamma = \lceil (p+q)/2 \rceil + 1$$

- Invarianti per rotazione (Hu)

$$\begin{aligned} h_1 &= n_{20} + n_{02}, \\ h_2 &= (n_{20} - n_{02})^2 + 4n_{11}^2, \\ h_3 &= (n_{30} - 3n_{12})^2 + (3n_{21} - n_{03})^2, \\ h_4 &= (n_{30} + n_{12})^2 + (n_{21} + n_{03})^2, \\ &\dots \end{aligned}$$



Altri insiemi di feature per OCR

- Non esiste un insieme di features che sia il migliore in qualunque applicazione, e la scelta è spesso frutto di tentativi
- Schemi di *feature extraction* molto diffusi sono:
 - coefficienti di Fourier 2D
 - codifiche del contorno
 - mappe di punti salienti
- La prassi suggerisce di calcolare molte più features di quanto realmente necessario, e di selezionare in seguito le migliori
 - *feature selection* (selezione di sottoinsiemi)
 - *feature generation* (combinazioni lineari ottimali)



Segmentazione caratteri

- Il processo di segmentazione è responsabile della estrazione dei singoli caratteri dall'immagine
- Vengono estratti di norma anche oggetti 2D che non sono caratteri, che il classificatore dovrà scartare
- Per le immagini binarie, la segmentazione deve semplicemente analizzare la connessione tra pixel adiacenti
- Spesso è utile la morfologia matematica per regolarizzare le forme e correggere difetti (es. caratteri fusi o frammentati)



Esempio

- Anche su immagini di buona qualità, gli errori sono inevitabili

PEGASYSTEMS INC
101 MAIN STREET
CAMBRIDGE MA 02142-1590

PEGASYSTEMS INC
101 MAIN STREET
CAMBRIDGE MA 02142-1590

- Un buon segmentatore deve usare strategie correttive, legate alla conoscenza della struttura del testo da leggere



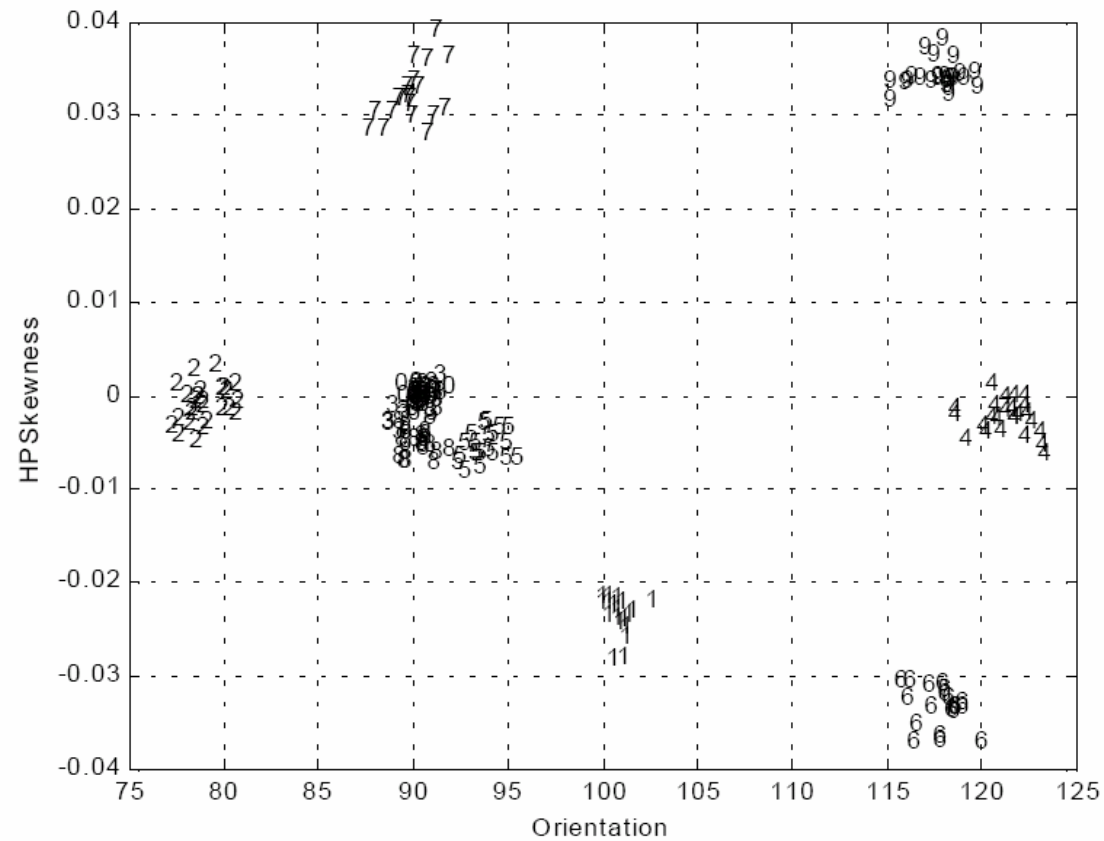
Classificazione caratteri

- In un sistema OCR, se la segmentazione ha operato correttamente, la classificazione di solito non presenta problemi, almeno per applicazioni con testi basati su simbologie standardizzate
- Anche poche features riescono già a ottenere discriminazioni
- La classificazione si complica quando la distribuzione delle classi contiene anche una variabilità indotta dagli stadi precedenti, che complica la definizione delle superfici di separazione
- A volte il training iniziale non è sufficiente per rappresentare tutte le possibili variabilità. E' importante quindi la capacità di generalizzare



Esempio

- Esempio di OCR a font fisso usando solo due *features* (orientazione e skewness)



OCR Postale

- Lo smistamento della corrispondenza usa da anni sistemi OCR per leggere indirizzi
- Le lettura del dattiloscritto o del manoscritto stampatello sono oggi considerate quasi “banali”
- Ancora non disponibili invece OCR efficaci per il corsivo
- La localizzazione dell’indirizzo (*Address Block Location, ABL*) è il vero problema delle applicazioni postali, a causa di:
 - ambiguità (indirizzi multipli)
 - complessità dei pezzi postali (es. riviste)
 - diverse tipologie di scrittura dell’indirizzo





On AIR s.r.l.

Tecniche di riconoscimento statistico

Applicazioni alla lettura automatica di testi
(OCR)

Parte 6 – Lettura targhe automobilistiche

Ennio Ottaviani

On AIR srl

ennio.ottaviani@onairweb.com

<http://onairweb.com/corsoPR>

A.A. 2008-2009

Aree applicative

- La lettura delle targhe (*License Plate Recognition*, LPR) è oggi un componente fondamentale per diverse applicazioni
 - Controllo di accessi (parcheggi, ZTL, ...)
 - Rilevamento di violazioni (passaggio col rosso, eccesso di velocità,...)
 - Recupero dei veicoli rubati
 - Analisi dei tempi di percorrenza
 - Esazione del pedaggio
 - Controlli di frontiera
 -
- Le targhe rappresentano un caso classico di OCR in partenza semplice (mono-font) ma reso complesso dalla variabilità indotta dall'ambiente



Localizzazione della targa

- La presenza di un targa viene identificata utilizzando appositi filtri che identificano la presenza di transizioni di grigio (gradienti)
- Di norma queste analisi vengono condotte lungo le righe dell'immagine, e confermate su righe adiacenti



Segmentazione della targa

- L'obiettivo della segmentazione è estrarre tutti i caratteri appartenenti alla targa. Sono di norma tollerati dei caratteri extra (spuri), ma è più complesso recuperare caratteri persi
- La segmentazione deve compensare le inevitabili variazioni pittoriche e geometriche dovute alle condizioni di ripresa
- Ci si riconduce quasi sempre ad una immagine binaria facile da trattare con tecniche classiche di elaborazione delle immagini



GE B56879



GE B56879



Analisi contestuale

- L'analisi contestuale comprende un insieme di metodologie per il trattamento di dati simbolici, finalizzato alla identificazione di termini appartenenti ad un lessico o comunque regolati da una grammatica
- Dati di ingresso
 - Lista di simboli candidati (caratteri veri e falsi) con significato univoco oppure multiplo (con costi relativi)
 - Grammatica espressa come insieme di regole che definiscono le concatenazioni di simboli ammissibili (sintassi)
 - Eventuali criteri geometrici aggiuntivi (posizioni)
- L'obiettivo finale è produrre una stringa di testo valida




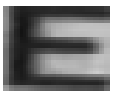
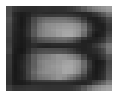
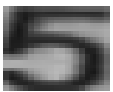
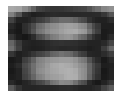
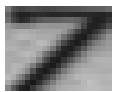

Analisi contestuale per le targhe

- Trattando testi di lunghezza limitata (e spesso nota a priori), si opera spesso per enumerazione diretta delle possibili concatenazioni e successiva valutazione
- Nei casi più complessi si usa un approccio markoviano (ogni simbolo viene scelto sulla base del precedente, cercando la sequenza ottimale). I sistemi di questo tipo sono detti *parsers*
- Attenzione: l'analisi contestuale permette di correggere errori di lettura e di saltare caratteri falsi ma non di recuperare caratteri persi!



Esempio

- Si ricerca l'allineamento ottimo di una sequenza di 8 simboli compatibile con la grammatica delle targhe italiane
- La funzione di costo cumula una quantità positiva legata alla distanza di Mahalanobis

							
6 42 G 60	E 35	B 33 8 41	5 21 S 66	6 35	B 23 8 25	7 15 Z 33	9 23 8 38

- Senza analisi contestuale la lettura a costo minimo sarebbe 6EB56B79 (costo 227). Con le regole della grammatica si ha invece GEB56879 (costo 247)

