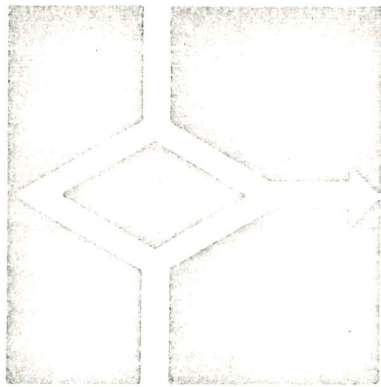


1961



CONGRESSO ANNUALE  
ANNUAL CONFERENCE

A.I.C.A.  
ASSOCIAZIONE  
ITALIANA  
PER IL CALCOLO  
AUTOMATICO

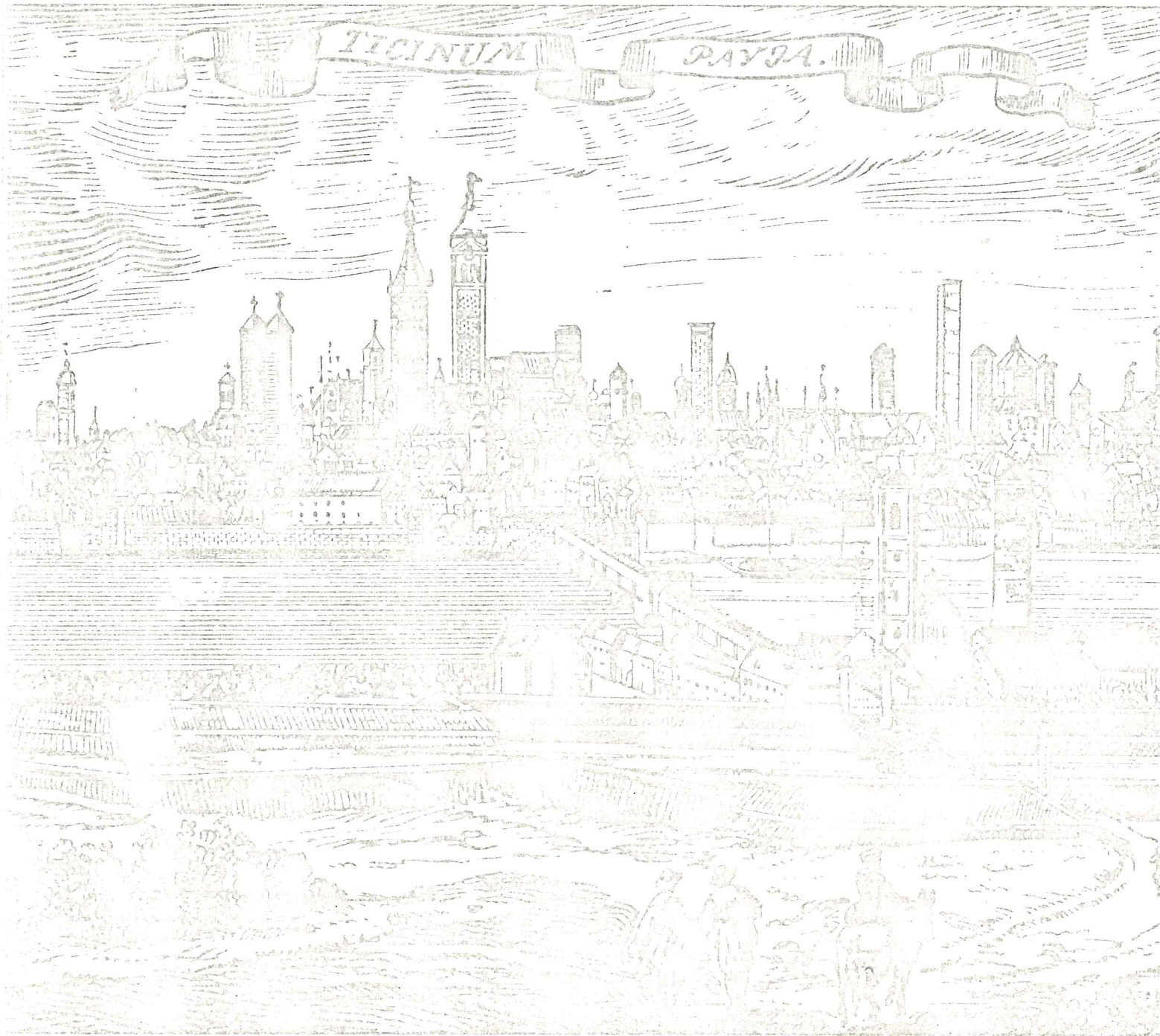
PAVIA 23-25 Settembre 1981

1981

IST. EL. INF.  
BIBLIOTECA  
Posiz. *RECETTIVO*

*LSI-19*

atti  
volume I



# SULLA DETERMINAZIONE AUTOMATICA DI CARIOTIPI

A. ANDRONICO(\*), P. L. CASALINI(\*\*), A. TONAZZINI(\*\*\*)

(\*) I.R.I.-C.N.R. Via S. Maria, 46 -- 56100 Pisa

(\*\*) I.S.I. -- Corso Italia, 40 -- 56100 Pisa

(\*\*\*) Istituto di Matematica, Via Buonarroti, 2 -- 56100 Pisa

*Sommario.* Si descrive un metodo per la determinazione automatica di cariotipi. Il lavoro è articolato in tre parti. Nella prima viene descritto il processo che porta all'acquisizione delle informazioni di carattere geometrico relative alle figure contenute nel fotogramma di una piastra metafase; dopo un breve cenno sulla digitalizzazione, che trasforma i dati analogici del fotogramma in una matrice numerica facilmente trattabile, si descrive la determinazione automatica del livello di taglio dell'immagine digitalizzata e il prelievo del contorno delle figure in essa contenute. Nella seconda parte viene descritta una procedura automatica per la classificazione dei singoli cromosomi in base a certi loro parametri geometrici caratteristici. Nella terza si descrive una procedura per l'accoppiamento dei cromosomi. Tale procedura, del tutto generale, si basa sulla costruzione di una «funzione di somiglianza» il cui studio porta alla suddivisione in classi di appartenenza delle forme contenute nell'immagine in esame.

## 1. INTRODUZIONE

Anche i metodi descritti nella letteratura più recente sul problema dell'analisi e del riconoscimento automatico di cromosomi allo scopo di ottenere il cariotipo non portano a soluzioni complete [10], [11], [12]. Alcuni autori hanno adottato tecniche che considerano fotogrammi ottenuti da vetrini bandeggiati, e anche in questi casi non si hanno ancora risultati di costruzione completa del cariotipo in modo automatico. Poiché il problema è ancora aperto per alcuni aspetti d'attuazione, e dato l'interesse sempre più attuale di utilizzare metodi automatici o semi-automatici nell'analisi della metafase (in cui è possibile riscontrare particolari configurazioni), in questo lavoro viene presentato lo stato attuale di una ricerca in corso sull'argomento.

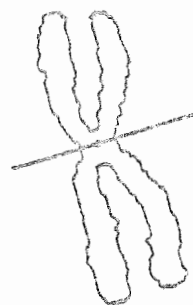
## 2. PROBLEMATICA AFFRONTATA

Nei fotogrammi ottenuti da vetrini contenenti cromosomi umani, questi appaiono come figure piane aventi una struttura geometrica ben definita, anche se non rappresentabile analiticamente.

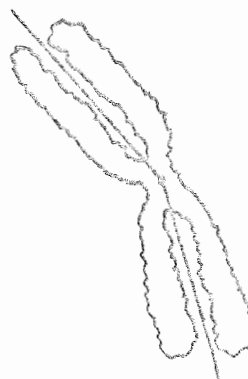
Brevemente, essi sono costituiti di due filamenti (*crromatidi*) uniti per una regione detta *centromero*. Sezionando un cromosoma ortogonalmente alla sua dimensione maggiore a livello centromerico, questo risulta suddiviso in due parti (*bracci*) (Fig. 1.).

Per ogni braccio è poi possibile definire un *asse di simmetria* (Fig. 2.).

Fino a pochi anni fa il metodo più comunemente usato in biologia per determinare il *cariotipo* (cioè l'ordinamento dei cromosomi in coppie di omologhi) era quello di effettuare un ingrandimento fotografico del vetrino, ritagliare i cromosomi ed accoppiare fra loro i più simili in base a osservazioni empirico-geometriche. Tale «somiglianza» veniva essenzialmente decisa in funzione della posizione del centromero e delle lunghezze totali dei due cromosomi in esame.



“Fig. 1. Struttura di un cromosoma”



“Fig. 2. Assi di un cromosoma”

Attualmente, mediante tecniche particolari (bandeggio) è possibile evidenziare su ogni cromosoma delle bande di colorazione diversa che permettono una più accurata determinazione del cariotipo [15].

Tuttavia il primo approccio di carattere geometrico rimane ancora valido e questo lavoro ne rappresenta un tentativo di automazione.

### 3. ACQUISIZIONE DELL'INFORMAZIONE ESSENZIALE

Il primo passo per l'analisi tramite computer di un vetrino consiste nella *digitalizzazione* del fotogramma da esso ricavato.

In generale un fotogramma (negativo o diapositivo) può essere considerato come un oggetto piano la cui luminosità (densità ottica) varia da punto a punto. Questo fatto può essere espresso matematicamente mediante una funzione  $f(x, y)$ , il cui valore rappresenta il livello di grigio che compare al punto del piano di coordinate  $(x, y)$  in un opportuno sistema di riferimento.

Mediante un processo di campionamento e una successiva quantizzazione (digitalizzazione), dall'insieme dei valori della  $f$  viene estratto un sottoinsieme finito di valori interi. Il risultato di una tale operazione è una matrice  $A$  bidimensionale il cui elemento  $a_{ij}$  rappresenta il livello di grigio medio corrispondente al rettangolo del fotogramma individuato dalle coordinate  $(i, j)$  di un suo ben determinato punto.

Per il presente lavoro tale processo viene effettuato fuori linea col sistema a flying-spot S.A.D.A.F. in funzione presso P.L.E.I. del C.N.R. di Pisa [1].

Ma per un'analisi di tipo geometrico delle forme che compongono un'immagine, l'informazione ottenuta con la digitalizzazione è ridondante in quanto contiene, oltre a informazioni morfologiche, anche informazioni fotometriche.

Il primo passo per l'estrazione dalla matrice della sola informazione geometrica consiste nello scegliere un livello di grigio  $\ell$ , ed effettuare un taglio (*thresholding*) sulla  $A$ , ottenendo una matrice  $A'$ , definita nel seguente modo:

$$a'_{ij} = \begin{cases} 0 & \text{se } a_{ij} < \ell, \\ 1 & \text{se } a_{ij} \geq \ell, \end{cases}$$

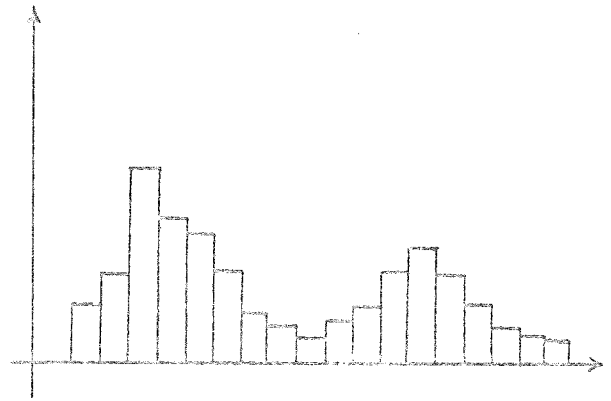
se si è digitalizzato un fotogramma negativo, altrimenti, se il fotogramma era diapositivo:

$$a'_{ij} = \begin{cases} 1 & \text{se } a_{ij} \leq \ell, \\ 0 & \text{se } a_{ij} > \ell, \end{cases}$$

La determinazione di  $\ell$  è fatta in modo automatico seguendo un procedimento euristico [8] che qui accenniamo brevemente.

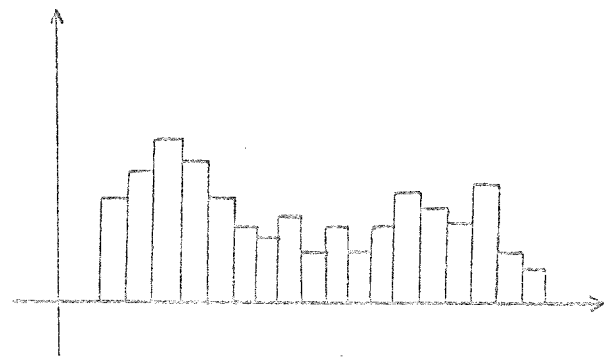
Se  $p$  sono i livelli di grigio ammessi, si può associare alla matrice  $A$  un vettore  $S$  (*vettore statistico*) di  $p$  elementi, il cui elemento  $k$ -esimo è uguale al numero degli elementi di  $A$  che hanno valore  $k$  ( $k=0, p-1$ ). Il metodo euristico nasce dallo studio dei minimi dell'istogramma del vettore statistico, cioè della rappresentazione cartesiana discreta, in cui sulle ascisse vengono riportati i livelli di grigio e sulle ordinate gli ele-

menti del vettore. Esso, in condizioni ideali, dovrebbe essere *bimodale* (Fig. 3.) con i due picchi in corrispondenza dei livelli di grigio del fondo e del soggetto rispettivamente: allora il livello di taglio sarebbe quello individuato dal minimo compreso fra tali picchi.



“Fig. 3. Iistogramma ideale”

Nella realtà i due picchi sono separati da più minimi alternati con massimi relativi (Fig. 4.) e questo complica notevolmente la determinazione di  $\ell$ , della quale si riportano qui le linee generali.



“Fig. 4. Iistogramma reale”

Considerando il vettore statistico  $S$  come la tabulazione di una funzione in  $p$  punti e definendo opportunamente, per terne di valori consecutivi, crescita, decrescenza, non crescita e non decrescenza, si determinano successioni di intervalli entro cui sono identificabili massimi e minimi relativi, che vengono poi raggruppati in due insiemi  $M$  e  $N$ . Siano  $L_M$  ed  $L_N$  rispettivamente il massimo e il minimo assoluti di  $M$ . Si sceglie come  $\ell$ , il minimo più vicino alla media dei minimi compresi fra  $L_M$  e  $L_N$ . Con tale metodo si riesce a determinare automaticamente  $\ell$ , per immagini il cui istogramma è, nel senso precisato in [8], “quasi bimodale”.

In questa categoria di immagini rientrano assai bene le piastrine metafasiche, a condizione di una buona ripresa fotografica.

Una volta effettuato il taglio, alla matrice  $A'$  ottenuta si applica un algoritmo di «inseguimento» [9] che rende, per ogni figura (determinata in base alla definizione di componente connessa di una matrice [16]), due vettori delle coordinate del suo contorno, nel sistema di riferimento che ha come

ascisse le colonne e come ordinate le righe della matrice.

Per vari motivi, fra cui una registrazione fotografica non buona, può succedere che le curve discrete così ottenute non rappresentino solo i contorni dei singoli cromosomi. Si può verificare infatti che alcuni fra i cromosomi più piccoli vengano perduti, altri risultino spezzati a livello centromerico e altri ancora uniti o sovrapposti, e questo senza possibilità di ottimizzazione (Fig. 5.).



“Fig. 5. Alterazioni provocate dal taglio su una particolare piastra metafasica digitalizzata”

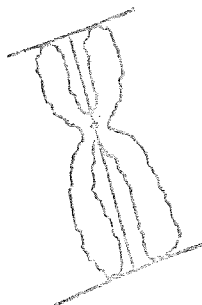
Queste curve devono essere eliminate dal successivo processo di classificazione in modo interattivo.

Viceversa le curve che rappresentano contorni di blasti (cioè nuclei di linfociti del sangue), o di piccole macchie di rumore possono essere eliminate automaticamente con metodi statistici basati sul conteggio dei punti di tali curve.

#### 4. LA PROCEDURA DI CLASSIFICAZIONE

Per ogni singolo cromosoma la classificazione viene fatta mediante il calcolo di 5 parametri essenziali e precisamente: area, perimetro, lunghezze dei due bracci, lunghezza totale. L'area e il perimetro vengono calcolati approssimando il cromosoma con il poligono avente per vertici i punti del suo contorno discreto.

Le lunghezze dei due bracci si calcolano come distanza fra il punto che, con buona approssimazione, rappresenta il centromero e le rette tangenti ai bracci (Fig. 6.).



“Fig. 6. Determinazione della lunghezza dei bracci”

L'algoritmo che ricerca il centromero prende come input una prima valutazione dei coefficienti angolari degli assi dei bracci del cromosoma e come output rende un'approssimazione delle coordinate del centromero e nuove valutazioni dei suddetti coefficienti. Tale algoritmo si presta quindi a un processo iterativo il quale, in modo del tutto euristico, si è verificato convergere velocemente, fornendo, in pochi passi d'iterazione, una soddisfacente valutazione delle coordinate del centromero.

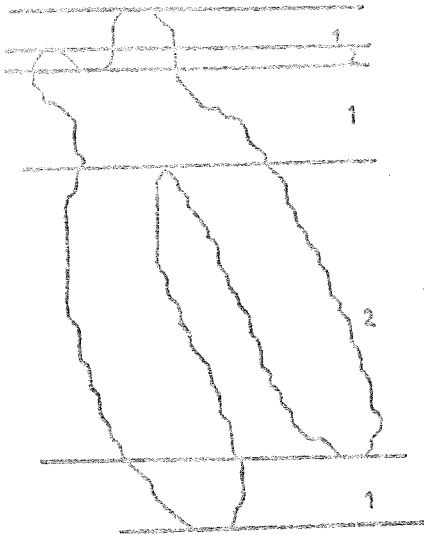
Le linee generali di tale algoritmo sono le seguenti. Dato uno dei due assi, si procede alla scansione del cromosoma con rette ad esso ortogonali. Nel caso che i due cromatidi siano ben separati lo scopo della scansione è quello di suddividere il cromosoma in zone caratterizzabili in base al numero dei punti in cui vengono intercettate dalle rette (Fig. 7.).

Viceversa se i due cromatidi appaiono quasi uniti lo scopo della scansione è quello di calcolare le ampiezze dei segmenti intercettati sul cromosoma stesso dalle rette (Fig. 8.).

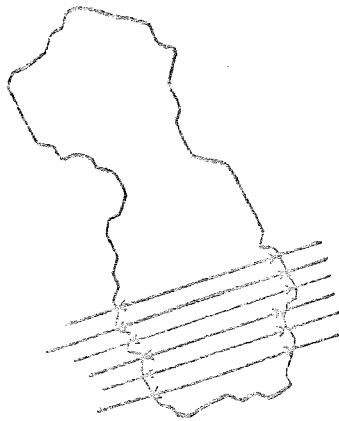
In entrambi i casi la zona centromerica è ben individuabile. Ripetendo il procedimento per l'altro asse e mediando i due risultati si ottiene una approssimazione del centromero.

Le rette tangenti ai bracci sono individuate ciascuna dai due punti del contorno del braccio in questione di massima distanza dal centromero, rispettivamente  $A'$  e  $B'$  in Fig. 9.

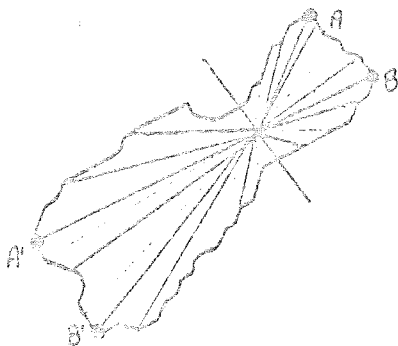
L'input del primo passo d'iterazione è costituito dal coefficiente angolare di una delle due diagonali del rettangolo circoscritto al cromosoma opportunamente scelta (Fig. 10).



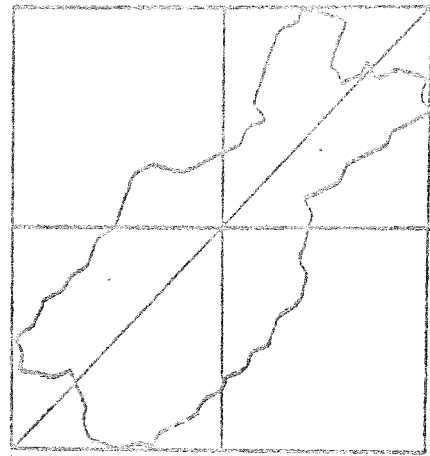
“Fig. 7. Suddivisione in zone di un cromosoma”



“Fig. 8. Calcolo delle ampiezze di un cromosoma”



“Fig. 9. Calcolo dei punti per cui passano le rette tangenti ai bracci”



“Fig. 10. Determinazione della prima valutazione dell'asse di un cromosoma”

Questo metodo dà buoni risultati già al secondo passo di iterazione anche per cromosomi ripiegati a livello centromerico.

### 5. LA PROCEDURA DI ACCOPPIAMENTO

Una volta classificati tutti i cromosomi di un intero fotogramma, a ciascuno di essi è associabile il pattern dei valori che i parametri significativi assumono nel cromosoma stesso.

Il metodo usato per effettuare gli accoppiamenti è del tutto generale e applicabile a molte e differenti categorie di immagini.

Siano  $A_i$ , per  $i=1, r$  tutti i cromosomi classificati con associati i pattern  $(a_i^{(1)}, \dots, a_i^{(n)})$  per  $i=1, r$ . Per ogni coppia di cromosomi  $A_i$  e  $A_j$  si costruisce il pattern  $(c_1, \dots, c_n)$  così definito:

$$c_k = \begin{cases} a_i^{(k)} / a_j^{(k)} & \text{se } a_i^{(k)} \leq a_j^{(k)} \\ a_j^{(k)} / a_i^{(k)} & \text{se } a_i^{(k)} > a_j^{(k)} \end{cases}$$

per  $k=1, n$  (in questo caso  $n=5$ )

$c_k$  può essere definito come il *grado di accoppiamento relativo al parametro k-esimo* fra  $A_i$  e  $A_j$ . Diremo *grado di accoppiamento assoluto* fra  $A_i$  e  $A_j$  il numero

$$R(i, j) = \frac{\sum_{k=1}^n \mu_k \cdot c_k}{\sum_{k=1}^n \mu_k}$$

dove i  $\mu_k$  sono pesi prefissati in base ad osservazioni statistiche.

Da numerose prove effettuate su campioni già costruiti, si è riscontrato che quando  $R(i, j) < 0,3$  i cromosomi  $A_i$  e  $A_j$  non sono mai omologhi. Si può quindi costruire una matrice  $M$  così definita:

$$M_{ij} = \begin{cases} R(i, j) & \text{se } i \neq j \text{ e } R(i, j) \geq 0,3 \\ 0 & \text{altrimenti} \end{cases}$$

La matrice  $M$  risulta essere quadrata, simmetrica e di ordine  $r$ .

Il metodo adottato per decidere le coppie di omologhi può essere così schematizzato. Data la riga  $i$ -esima della matrice si cerca il suo elemento di massimo valore. Sia esso  $M_{ij}$ , allora si dirà che i cromosomi  $A_i$  e  $A_j$  sono omologhi. Si procederà poi a porre a 0 le righe e le colonne della matrice di indici  $i$  e  $j$ . Se durante una tale scansione della matrice si incontrerà una riga di tutti zeri significherà che il cromosoma corrispondente o ha già trovato il suo omologo o non è accoppiabile.

Un procedimento di questo tipo è del tutto simile al Metodo Ungherese [17] applicato su una matrice di assegnazione  $M'$  costruita nel seguente modo:

$$M'_{ij} = [k M_{ij}]$$

dove  $k = 10^h$  con  $h$  massimo numero di cifre significative degli  $M_{ij}$ ,  $v_i$ ,  $v_j$ .

Negli esperimenti fatti con questo metodo si costruisce il cariotipo con una sicurezza compresa fra il 75% e l'85%.

A nostro giudizio, essendo il metodo di accoppiamento del tutto generale e fondato su presupposti matematici ben precisi (si veda la teoria dei Fuzzy Sets [13], [14]) la percentuale di errore deve essere imputata alla imprecisa valutazione dei parametri significativi, dovuta essenzialmente alla discretizzazione apportata dalla digitalizzazione del fotogramma, al taglio, nonché alla cattiva registrazione fotografica.

In condizioni ottimali il valore dei pesi associati a ogni grado di accoppiamento relativo dovrebbe essere unitario. Tuttavia, proprio per quanto detto sopra, si è riscontrato che certi parametri (quali ad esempio il perimetro e l'area) possono venire in qualche modo alterati (si pensi ad esempio ad un taglio che provoca l'unione dei due cromatidi). Se il fotogramma risulta investito in modo uniforme da tale alterazione della morfologia cromosomica, il metodo di accoppiamento, essendo basato sul rapporto fra i valori dei parametri, non ne risulterà influenzato. Viceversa, se solo una parte dei cromosomi sono rimasti alterati, è opportuno modificare i pesi associati ai vari gradi di accoppiamento o addirittura eliminare i parametri alterati dalla procedura stessa.

Il programma da noi messo a punto, tenendo conto dei casi che possono verificarsi, fornisce il cariotipo costruendo la matrice di assegnazione in differenti modi, scegliendo poi come soluzione ottimale il cariotipo che presenta il maggior numero di accoppiamenti.

Questo procedimento fa sì che la scelta dei  $\mu_i$  sia determinata, fissato quale o quali parametri vengono pesati in modo unitario, con criteri di proporzionalità.

## 6. ALCUNI RISULTATI

In Fig. 11. è riportato il fotogramma di una piastra metafase ottenuta in condizioni quasi ideali, in quanto risulta particolarmente nitido e privo di blasti o rumori di altra natura.

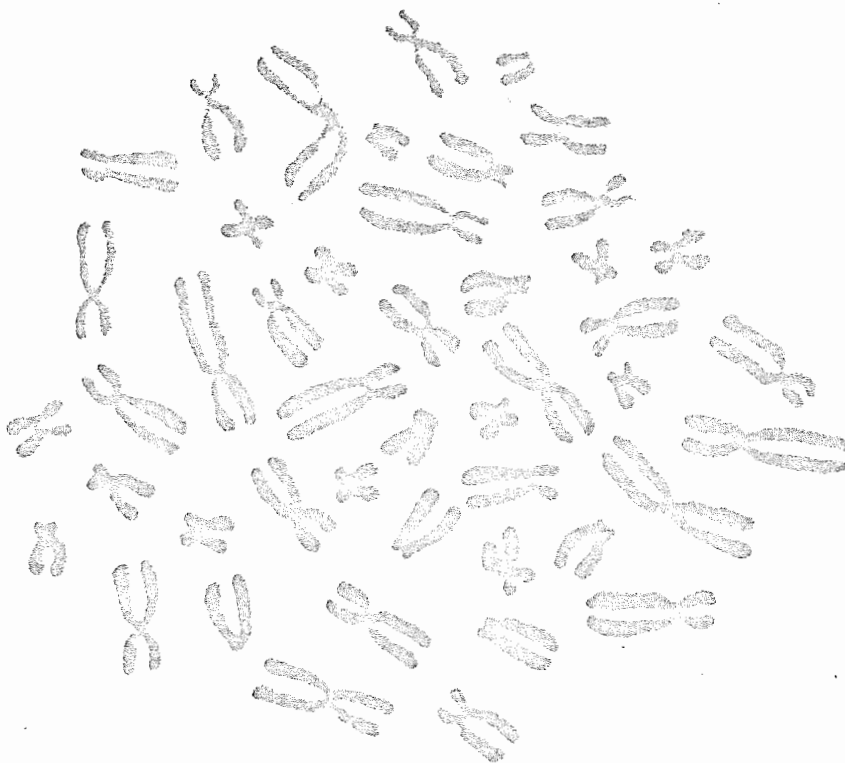


Fig. 11. Fotogramma di una piastra metafase

In Fig. 12. è riportato il risultato dei processi di digitalizzazione, taglio e inseguimento dei contorni effettuati sul fotogramma di Fig. 11., così come appare su terminale video. I numeri che compaiono accanto a ogni cromosoma indicano l'ordine con cui i contorni sono stati prelevati dalla matrice di digitalizzazione.

In Fig. 13. sono riportati alcuni esempi di coppie di omologhi, visualizzate su terminale video al termine del processo di accoppiamento.

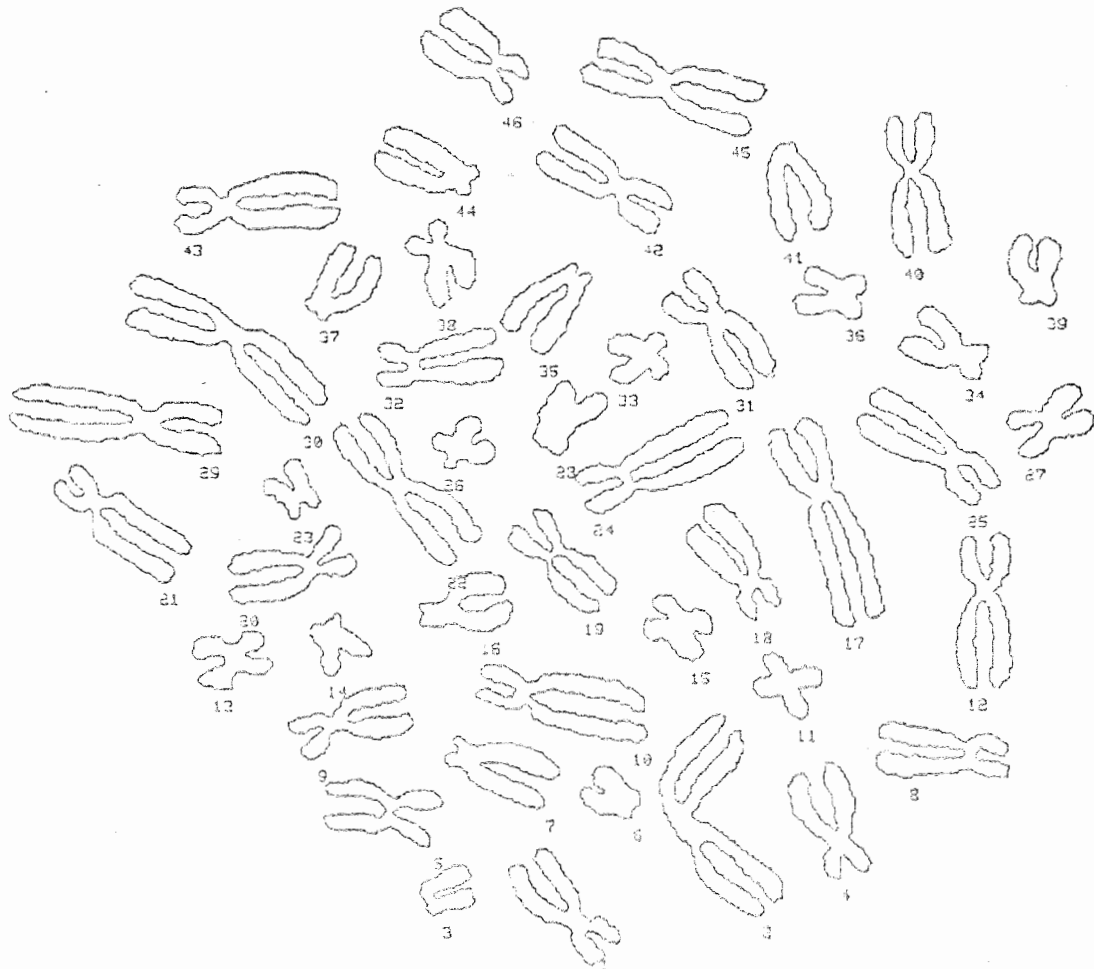
In Fig. 14. è riportato il cariotipo completo costruito sulla base degli accoppiamenti ottenuti automaticamente. Le frecce indicano gli errori, cioè gli scambi fra cromosomi di gruppi differenti, rispetto al cariotipo effettivo fornito dai biologi e riportato in Fig. 15. Gli scambi di cromosomi fra coppie di uno stesso gruppo non sono significativi in quanto, senza le tecniche del bandeggio, non è possibile determinare gli esatti accoppiamenti all'interno di ciascun gruppo.

In figura 16 è riportato il fotogramma di una piastra metafisica ottenuto in condizioni di ripresa fotografica e di rumore del vetro molto peggiori rispetto al fotogramma di Fig. 11. Si può notare la presenza, sulla destra, di un grande biasto.

In Fig. 17. è riportato il risultato del prelievo dell'informazione geometrica contenuta nel fotogramma di Fig. 16. Si noti l'assenza del contorno del biasto, eliminato automaticamente. Il taglio ha qui però provocato l'unione di alcuni cromosomi e la perdita di uno dei piccoli cromosomi.

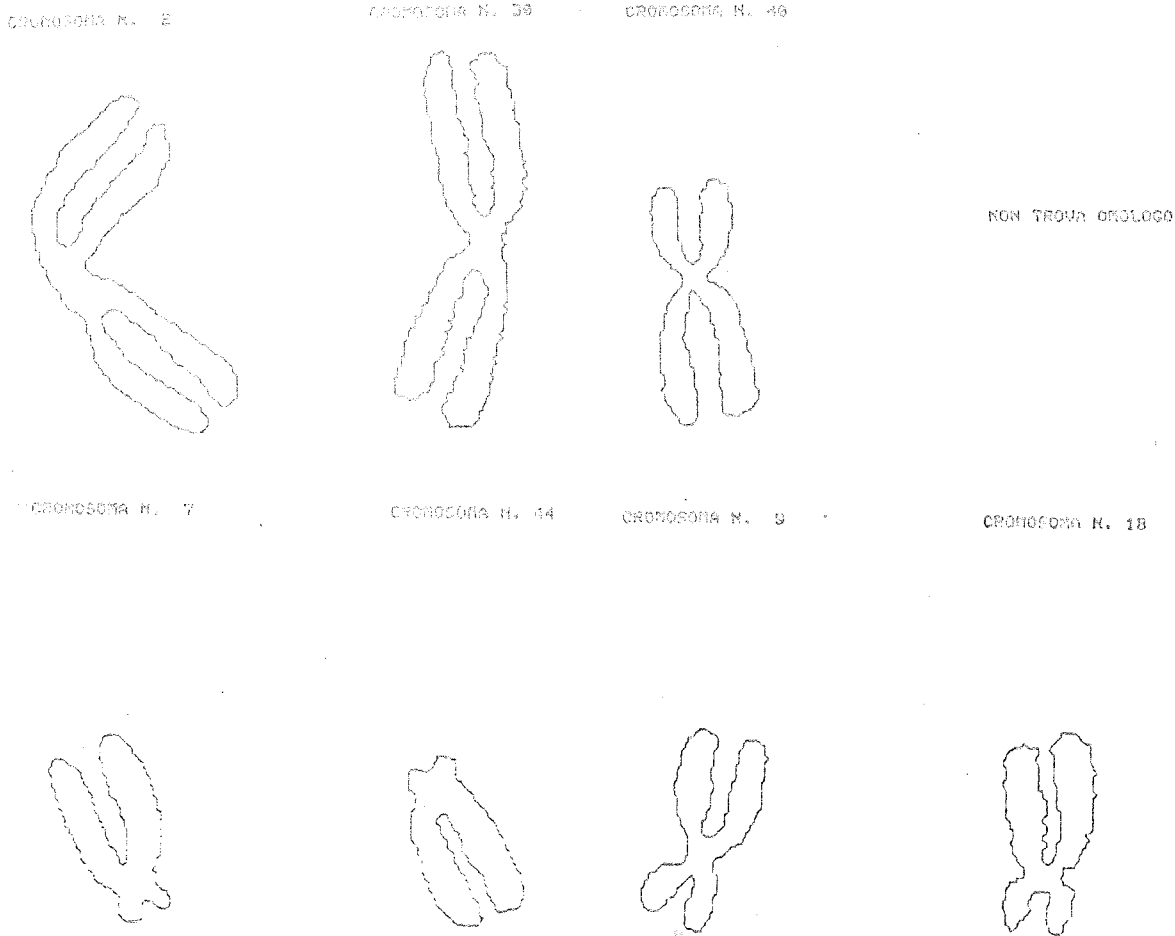
In Fig. 18. sono riportati alcuni esempi di accoppiamenti ottenuti dall'analisi della piastra metafisica di Fig. 16. Si noti come il taglio ha distorto il contorno del cromosoma n. 10.

Di questo esempio non si riporta il cariotipo completo in quanto non sarebbe significativo per l'impossibilità di classificare alcuni cromosomi.



"Fig. 12. Risultato del prelievo dell'informazione geometrica contenuta nel fotogramma di Fig. 11."

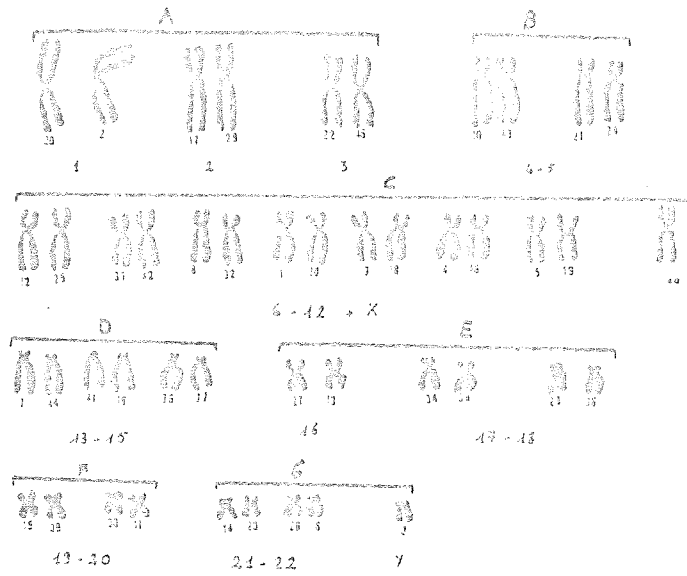
Sulla determinazione automatica di cariotipi



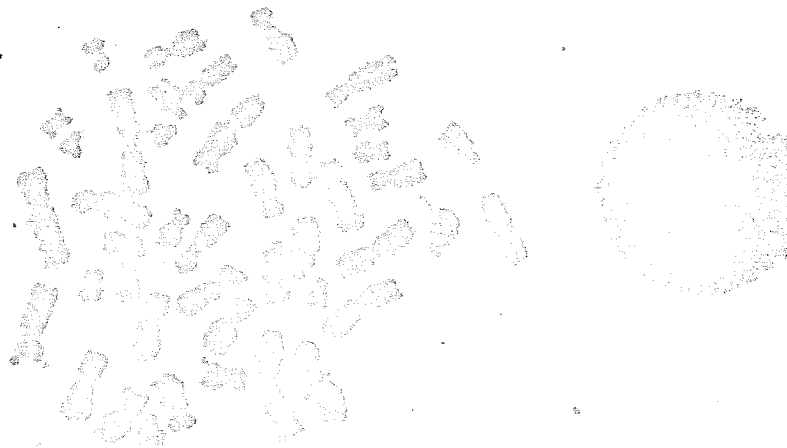
"Fig. 13. Alcuni accoppiamenti ottenuti dall'analisi della piastra metafase di Fig. 11."



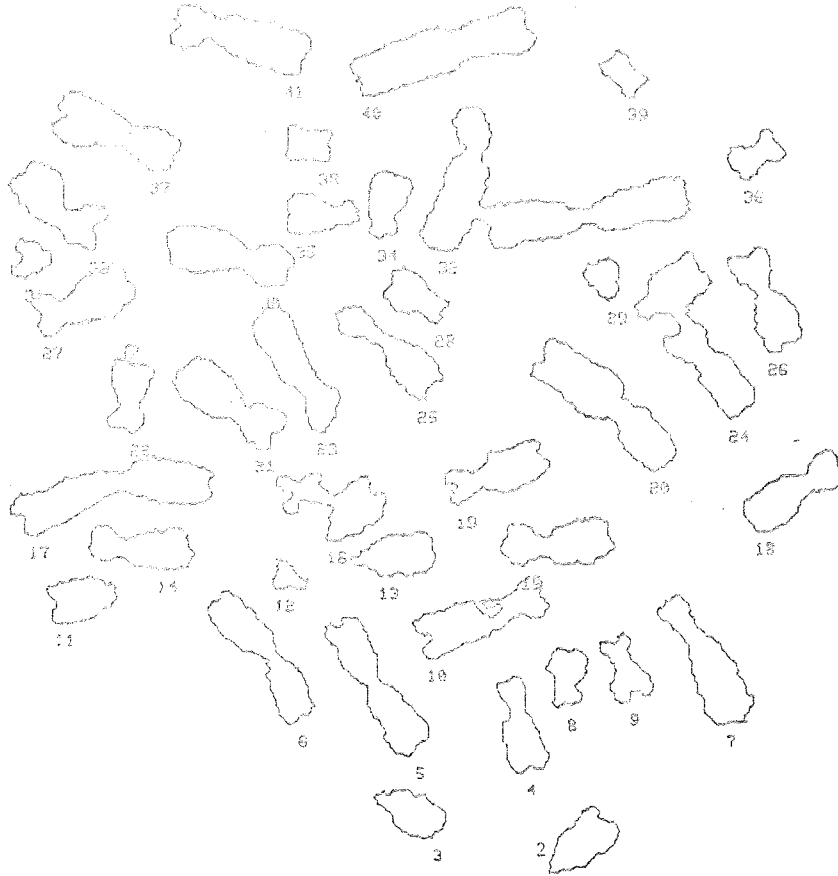
"Fig. 14. Cariotipo automatico della piastra metafase di Fig. 11."



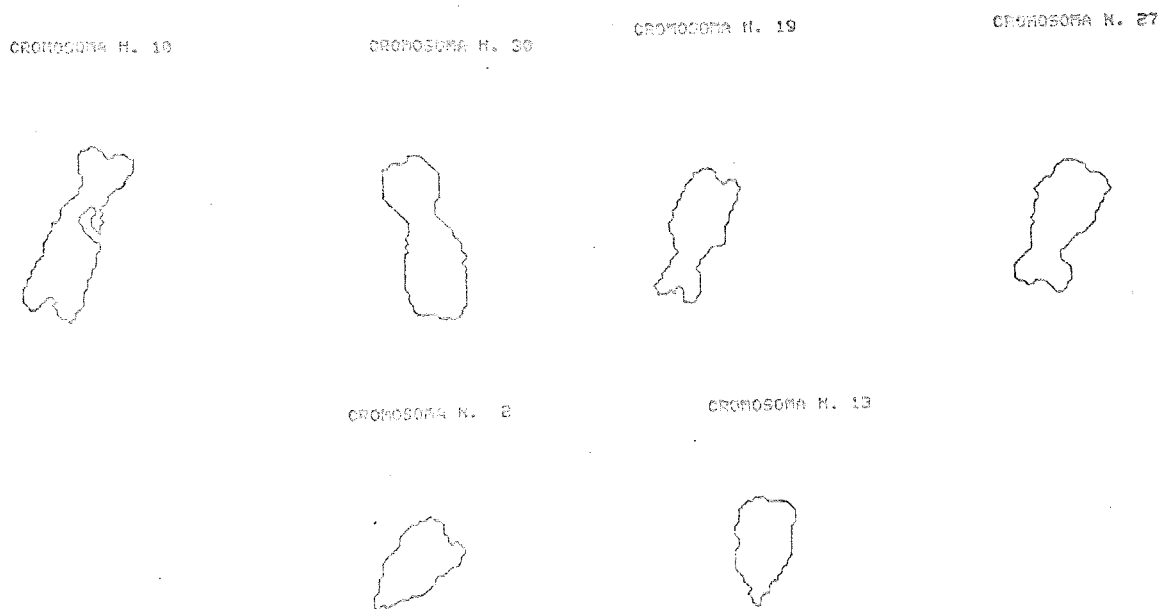
"Fig. 15. Cariotipo fornito dai biologi"



"Fig. 16. Fotogramma di una piastra cromosomica"



“Fig. 17. Visualizzazione su terminale video dell'informazione prelevata dal fotogramma di Fig. 16”



“Fig. 18. Coppie di omologhi carcioidi dall'analisi della piastra metafasica di Fig. 16”.

Si è voluto tuttavia, con esso, mostrare la necessità, comune e prioritaria nel campo dell'elaborazione e analisi di immagini, di effettuare riprese fotografiche il più possibile nitide e fedeli all'originale.

## 7. BIBLIOGRAFIA

- [1] L. Azzarelli, R. Ponicucci, Descrizione del sistema S.A.D.A.F. per la lettura e digitalizzazione di fotogrammi, Nota Tecnica C72-2, I.E.I. 1972.
- [2] H. C. Handrews, Computer techniques in image processing, (Academic Press, 1970).
- [3] A. Grasselli, Automatic Interpretation and classification of images, (Academic Press, 1969).
- [4] A. Rosenfeld, Picture processing by computer, (Academic Press, 1969).
- [5] G. Callus, N. Montanaro, G. A. Maccaro, Una applicazione del «pattern recognition» in biologia: tentativi di appaiamento e classificazione dei cromosomi umani con l'impiego di un elaboratore elettronico, *Atti del XIV Congresso Scientifico Internazionale per l'Elettronica* (1967).
- [6] D. Rucovitz, Automatic chromosome analysis, *Brit. Med. Bull.* 24,3,260 (1968).
- [7] F. Bernardi, P. L. Casalini, G. Marchetti, Un metodo di implementazione interattivo per l'analisi di cromosomi tramite computer, I.E.I. Nota Interna B78-36, (Dicembre 1978).
- [8] P. L. Casalini, L. Monaci, Un metodo per la determinazione automatica del livello di soglia per immagini di tipo cartografico, Nota Interna B79-27, I.E.I. (Dicembre 1979).
- [9] P. L. Casalini, R. Favilla, L. Monaci, Studio preliminare per l'automatizzazione dell'acquisizione di informazioni da immagini di tipo cartografico, I.E.I. Nota Interna B80-3, (Aprile 1980).
- [10] J. Piper, E. Craun, D. Rucovitz, H. Rutledge, Automation of chromosome analysis, *Signal Processing*, (July 1980) pp. 203-221.
- [11] G. A. Van Zee, F. C. A. Groen, A Contrast Enhancing Filter for Banded Chromosomes, *Computer and Biomedical Research* 11, pp. 325-335 (1978).
- [12] N. E. Morton, S. Yee, Estimation of Polymorphism in Quantitative Measurements on Homologous Pairs of Human Chromosomes, *Computer and Biomedical Research* 11, pp. 89-101 (1978).
- [13] L. A. Zadeh, Fuzzy Sets, *Inf. and control* 3, pp. 338-353 (1965).
- [14] R. Bellman, R. Kalaba, Zadeh L. A., Abstraction and Pattern Classification, *J. of Math. Anal. and Appl.* 13, pp. 1-7 (1966).
- [15] F. Nuzzo, Il cariotipo umano normale e patologico, Piccin Editore, (1975).
- [16] F. Bernardi, P. L. Casalini, G. Marchetti, Digitalizzazione e geometria dell'immagine, Nota Interna B79-13, I.E.I. (Maggio 1979).
- [17] J. Munkres, Algorithms for the assignment and transportation, *J. Soc. Industrial Applied Math.*, vol. 3 (marzo 1957), pag. 32-38.