ACTA STEREOL 1992; 11/2: 191-197
ORIGINAL SCIENTIFIC PAPER

STEPWISE DISCRIMINANT ANALYSIS : AN AID FOR COLOR IMAGE SEGMENTATION

Catherine Souchier, Martine Ffrench, Régine Catallo,
Paul André Bryon,

Cytologie analytique, Université Claude Bernard,
8 avenue Rockefeller, 69373 LYON Cedex 08, France

ABSTRACT

Color image analysis allows new applications. A procedure is presented for color image segmentation. It implies measurement of pixels associated with different image colors and a stepwise discrimant analysis. It is illustrated on immunoenzymatic labelings.

Key words : color, image analysis, immunoenzymatic labeling.

INTRODUCTION

Nowadays, recent developments in image analysis allow true color analysis of histological sections, cellular smears or imprints. True color analysis provides more information about specimens, and increases the possible applications of image analysis. It allows objects with different color values, but matching grey tones to be detected. For example, pink and pale blue cytoplasms on a Papanicolaou specimen can be separated, as yellow-brown and red-purple markers on double immunoenzymatical labeled (PAP/APAAP) specimens. With monochrome images, it would be difficult, and even impossible without color filters.
Color trichromatic theory implies that any color can be matched by a vectorially additive combination of three primaries - red, green and blue -. In image analysis, a color image is automatically split into red, green and blue spectral bands, and stored into three separated numerical "grey-tones" images (Levine MD, 1985, Ballard DH, 1982). Grey operations and transformations can be performed on all of the three images.
Monochrome segmentation algorithm, as thresholding - the simplest one -, can be applied on any of the three images and on every possible combination of images for better color discrimination. For instance, the red, green and blue images (Charpin, 1988) or the hue, luminance or saturation images (Cohen, 1988), which can be easily obtained from true color images in most commercial image analyzers, were used in the estrogen receptors application. Moreover, many other images resulting from linear (addition, subtraction, division, multiplication) and non linear (minimum or maximum) grey image operations, or from image transformations with filters, such as a median filter or a morphological one, can be employed. For instance, Mac Auley et al (1989) used the linear

combinations obtained by principal component analysis. The choice of the most valid image(s) for the best image segmentation is thus expressed. We worked with a simple procedure for color image segmentation. It requires image instructions and data processing software currently implemented and used in image and data analysis, and consists of two main steps 1) acquisition of pixels color values associated with different image colored zones and 2) stepwise discriminant data analysis. The procedure is presented and results are illustrated on double (PAP/APAAP) immunoenzymatic labeled human bone marrow aspirates of patients presenting myeloma. In this application, the black and white image is difficult to be analysed (fig 1).

MATERIAL AND METHODS

a) Biological specimens

Bone marrow aspirates were incubated for one hour with Bromodeoxyuridine (BrdU). Cell smears were performed by cytocentrifugation and were methanol fixed. A double immunoenzymatic labeling of intracytoplasmic Ig by peroxidase-antiperoxidase (PAP) technique and BrdU by alkaline phosphatase-anti-alkaline phosphatase (APAAP) technique after DNA denaturation with formamide was realized (Ffrench M, 1990). Cytoplasms of plasma cells were brown (diaminobenzidine (DAB) reaction product) and S phase nuclei were red (fast red reaction product). Unlabeled nuclei were blue (haematoxylin counterstaining) (fig 1).

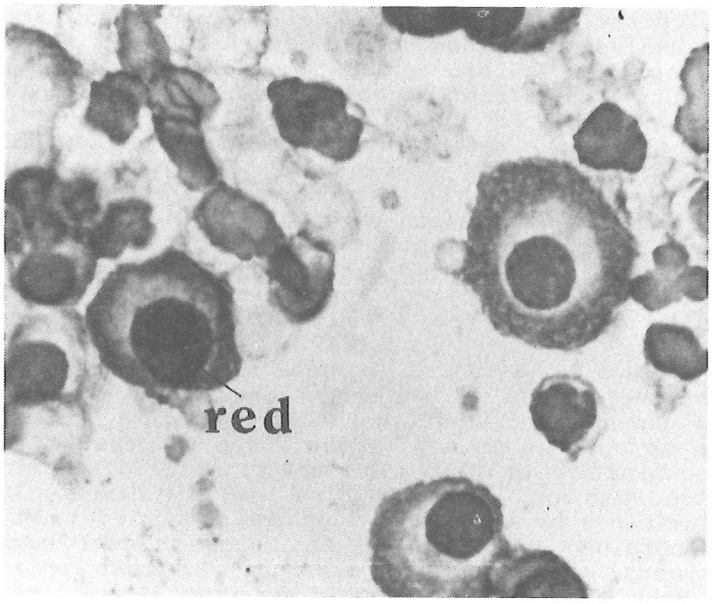

Fig.1. Black/white image of double immunoenzymatic labeled bone marrow aspirates. The recognition of proliferative nuclei (red) is difficult.

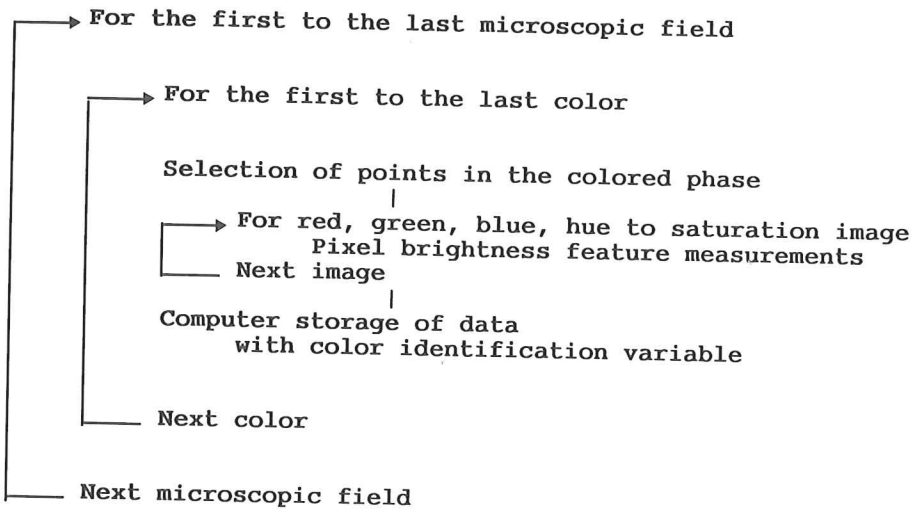

Fig. 2. Flow chart of image processing program

b) Image analysis

b1. Procedure

A simple procedure was elaborated in order to find the most suitable image(s) for image segmentation. It implies a preliminary study which consists of two main steps.

1) Color pixel values associated with different image colored zones were acquired. Points were selected with a cursor, and color data (red, green, blue, hue and saturation values) were acquired and stored in a data file with a color zone identification variable (fig 2). Each individual piece of data corresponds to one pixel only (table 1). On average, two hundred pixels per color and per case were acquired.

2) Data were analyzed to find the variable(s) or the combination(s) of variables which emphasize the differences between image colors (image zones), and a stepwise discriminant analysis was performed.

a) With only two colors, one discriminant factor - initial image or weighted linear combination image - was thus obtained and should be used for image segmentation in the image analysis routine. Image analysis system limits - integer values, 0 and 255 extreme values, 8 or 16 bits arithmetic operations - were however to be considered. Discriminant analysis was thus repeated with one (or several) more variable(s) : discriminant factor estimations which could be obtained with the image analysis system. As an improvement in approximation requires more arithmetic grey operations and thus more acquisition times, real efficiency of different possibilities were studied.

b) With more than two colors, a correct classification rate could require more than one discriminant factor. The classification functions could be used, but they necessitate constant mean values and thus, more reproducibility from slice to slice. Implementation of discrimant factors could also be performed, but segmentation parameters which imply different thresholdings and boolean operations were considered as too difficult to control with our image analysis instructions package. The solution used consisted of colors grouping. Contrast between one color or one color group with the rest of the image was analyzed as in the preceding part (a).

Table 1. Data set example obtained on double labeled plasma cells. Background were notified as white, plasma cells as brown, S phase nuclei as red, and other cellular components as blue.

Pixel	COLOR ZONE	COLOR VALUES RED	GREEN	BLUE	OF HUE	PIXELS SATUR
1	White	227	250	231	142	19
2	White	218	242	227	134	23
i	White	r_i	g_i	b_i	h_i	s_i
200	White	248	245	248	28	2
201	Red	135	69	115	10	58
202	Red	94	9	45	253	75
i	Red	r_i	g_i	b_i	h_i	s_i
400	Red	171	89	149	10	73
401	Brown	178	134	83	212	84
402	Brown	193	144	101	216	78
i	Brown	r_i	g_i	b_i	h_i	s_i
600	Brown	160	107	77	220	74
601	Blue	195	205	207	100	10
602	Blue	201	206	211	85	11
i	Blue	r_i	g_i	b_i	h_i	s_i
800	Blue	183	188	210	72	26

b2. Instrumentation

The study was performed on an image analysis system (Samba 2002, TITN) connected by a tri-CCD camera (SONY) with a Zeiss microscope (Axioplan). Plasma cells observation was done with a 40X oil immersion objective (NA=1.4). Ten different smears were analyzed, and 200 pixels by color (white, brown, red, and blue) were measured (table 1). Routines were developed in extended Pascal, which allows interfacing with image software. Data processing was realized on a 286 PC compatible computer (Victor V286) using BMDP statistical software (Dixon et al, 1988).

RESULTS

Similar discriminant analysis results were found for the 10 different smears, and mean values were retained (table 2). In all cases, the weight of the blue image was most important for the plasma cells selection (brown pixels), and the weight of the green image for the S phase nuclei (red pixels).

Table 2. Discriminant functions for the selection of brown pixels of plasma cells or for the selection of red pixels of S phase nuclei

	DISCRIMINANT FUNCTIONS		
	a) SELECTION OF BROWN PIXELS		b) SELECTION OF RED PIXELS
1	Blue - 0.22*Red - 0.54*Green	1	Green - 0.38*Red - 0.53*Blue
2	Blue - 0.32*Red - 0.51*Green	2	Green - 0.27*Red - 0.63*Blue
3	Blue - 0.33*Red - 0.54*Green	3	Green - 0.59*Red - 0.43*Blue
4	Blue - 0.24*Red - 0.49*Green	4	Green - 0.66*Red - 0.42*Blue
5	Blue - 0.52*Red - 0.48*Green	5	Green - 0.30*Red - 0.55*Blue
6	Blue - 0.20*Red - 0.62*Green	6	Green - 0.61*Red - 0.30*Blue
7	Blue - 0.10*Red - 0.61*Green	7	Green - 0.63*Red - 0.42*Blue
8	Blue - 0.38*Red - 0.58*Green	8	Green - 0.54*Red - 0.52*Blue
9	Blue - 0.59*Red - 0.49*Green	9	Green - 0.19*Red - 0.59*Blue
10	Blue - 0.40*Red - 0.46*Green	10	Green - 0.65*Red - 0.45*Blue
MEAN	Blue - 0.33*Red - 0.53*Green	MEAN	Green - 0.48*Red - 0.48*Blue

The image corresponding to the linear combination (Blue - (Red/3 + Green/2) + 128) was defined as the most valid for contrast improvement between the labeled cytoplasms of plasma cells (different levels of brown) and the rest of the image (table 3-a). The expression was defined from the mean value (table 2-a) to be performed by image analysis operations. Integer values of Red/3 and Green/2 were taken, and a constant (128) was added for inclusion of all values between 0 to 255. The image leads to nearly 96% correct classification, and significantly improves the results that could be obtained with the best initial image, i.e. the blue one, (improvement of +25%) or with the hue image (improvement of + 13%) (table 3-a).

The image corresponding to the linear combination (Green - (Red/2 + Blue/2) + 128) was defined as the most valid for contrast improvement between proliferative nuclei (different levels of purple) and the rest of the image (table 3-b). It leads to nearly 99% correct classification, and improves the results that could be obtained with the green image (improvement of +10%). The classification rate was significantly lower with the hue image (73%) (table 3-b).

Table 3. Correct classification percent of pixels with different levels of brown or red obtained on three different images

	CORRECT CLASSIFICATION PERCENTS a) OF BROWN PIXELS				b) OF RED PIXELS		
	B-R/3-G/2	BLUE	HUE		G-R/2-B/2	GREEN	HUE
1	95	78	78	1	98	86	73
2	98	79	86	2	99	85	73
3	96	69	75	3	98	91	69
4	97	61	84	4	98	84	73
5	94	73	89	5	98	91	71
6	93	63	81	6	99	98	72
7	96	63	84	7	99	90	71
8	99	86	79	8	98	78	75
9	93	72	73	9	99	92	63
10	98	70	98	10	100	95	90
MEAN	96	71	83	MEAN	99	89	73

DISCUSSION

Similar results were obtained from one smear to another (table 2), and satisfactory discrimations were obtained with the mean values (table 3). Differences were corrected by different threshold limits. Identification of image colors was only needed and the results of data analysis could be used in all applications implying the same double immunoenzymatic procedure, without any further preliminary study. It is an important point since this study can be only performed interactively because of color identification of some pixels. Compared to the principal component decomposition used by MacAulay (1989), it presents the only disadvantage of our procedure ; its main advantage being that discriminant analysis optimizes differences between groups (the image colors in our case), and thus, better improves the contrast of one colour.

Three other points should also be stressed to demonstrate the procedure interest.

The first is that the most valid image before thresholding is found in a reproducible and objective way. It is difficult with only images display to define the best choice, and where the preliminary study implied by our procedure is concerned, it is also time consuming. It should also be noted that differences between colors, as well as variabilities between pixels belonging to a same zone, are taken into account. It did happen that an image (the hue image for example) was retained while the image itself was poorly contrasted.

Secondly, the procedure is not limited to the choice between fixed possibilities but also makes it possible to find the most valid linear combination. This case is well illustrated in our example in which two linear combinations were retained.

Thirdly, the procedure is simple to use whatever the image analysis system. It is only necessary to draw points in the colored zones and to get the grey level of each pixel in the different colored images. The present study was done on one image analysis system but the procedure can be easily reproduced on other image analysis systems. For example, we have implemented it more recently on a Quantimet 570 image analysis system (Leica, Great Britain) without any difficulty. With these two systems, no specific image analysis instructions were written. We have mainly used an editor (cursor or mouse) and the instruction : "integrated grey level limited to a binary mask". Both are currently available on an image analysis system. Moreover, good results could be obtained if only mean grey values in small frames could be measured. In this case, for a satisfactory approximation, pixels inside the frame should have nearly the same color. For data analysis, stepwise discriminant analysis was used. It is also largely developed, and is present in much statistical data analysis software, such as wellknown BMDP or SAS statistical package. If necessary, successive one way analysis of variances could be performed to identify the best image, and some linear combinations could be even tested.

Color identification is an important step in total image colored components segmentation, and this systematic procedure was for us a useful and easy tool.

ACKNOWLEDGEMENT

This work was supported by the ARC under grant 6851.

REFERENCES

BALLARD DH, BROWN CM. Color: Computer vision, Jersey: Prentice-Hall, 1982:31-34.

CHARPIN C, MARTIN PM, DEVICTOR B, LAVAUT MN, HABIB MC, ANDRAC L, TOGA M. Multiparametric study (samba 200) of oestrogen receptor in immunocytochemical assay in 400 human breast carcinomas : analysis of estrogen receptor distribution heterogeneity in tissues and correlations with dextran coated charcoal assays and morphological data. Cancer Research 1988;48:1578-1586.

COHEN O, BRUGAL G, SEIGNEURIN D, DEMONGEOT J. Image cytometry of oestrogen receptors in breast carcinomas. Cytometry 1988;9:579-587.

FFRENCH M, MAGAUD JP, DUHAUT P, BRYON PA, VIALA JJ. Determination of plasma cell labelling index with bromodeoxyuridine using a double immunoenzymatic technique. Stain Technology 1990;65:1:31-36.

LEVINE MD. Color: Vision in man and machine, New York: McGraw-Hill book company, 1985:313-370.

MACAULAY C, TEZCAN H, PALCIC B. Adaptative color basis transformation : an aid in image segmentation. Anal Quant Cytol Histol 1989;11:1:53-58.

Received: 1992-05-11
Accepted: 1992-11-03