ADAPTIVE CLUSTERING BY ANALYSIS OF THE CONNECTEDNESS DEGREE. APPLICATION TO SKIN SEGMENTATION.

Michèle Gouiffès

Institut d'Electronique Fondamentale CNRS UMR 8622, Université Paris-Sud 11, France

ABSTRACT

The paper presents a spatio-color classification in a chrominance-luminance space related to the dichromatic model. The unsupervised adaptive clustering is performed after the *color connectedness degrees* (CCD) of a color interval which embeds jointly 1) colorimetric information; 2) the probability that a given color is connected (in the image) to a set of similar colors. The chrominance CCDs are analyzed first while the luminance CCDs are studied only when necessary. Eventually, the method depends mainly on one parameter : the quantization step λ. The method is evaluated quantitatively in terms of quality and compactness (number of finals colors) on the Kodak image database. As an example, this generic technique is applied to skin detection.

1. INTRODUCTION

Many multimedia interactivity applications require some specific processings such as color clustering and/or segmentation: skin detection [1], retrieval or indexing, tracking [2]. Starting generally from the color distribution in the histogram, the clustering consists in determining the k most representative classes of pixels.

The performances of the classification depend highly on the choices made 1) on the colorspace 2) on the type of histogram. Color histograms have the advantage to be robust to small changes in viewpoint or appearance. Unfortunately, since they do not preserve topological (connectedness) or even any spatial information, they remain usually not informative or discriminant enough. Several elegant methods incorporate both spatial and colorimetric information into a same data representation [3–5]. Finally, the most classical way to improve the color data is to use 2^{nd} order statistical features as in the precursor work of [6]. Particularly, the Color Connectedness Degrees [7] was recently exploited in a segmentation procedure in order to secure a one-to-one relationship between color clusters and regions [8]. The procedure starts from a fixed number k of classes and determines the cubic color intervals in the RGB space with highest compactness degree and color homogeneity, therefore more likely to span real regions in the image.

Our clustering procedure relies on the CDD as in [8]. Our contributions lie both in the color representation and in the classification strategy:

1) A colorspace is designed to obey the assumptions of the dichromatic model which separates the chrominance of lambertian objects from their luminance, avoiding the problems of ill-definition arising in HSV space for example.

2) Chrominance and luminance are classified separately, which allows simpler data structures and reduced executing times. The chrominance classification is privileged. The luminance classification is performed in a second stage only when necessary.

3) The shape of the color clusters is adaptive.

4) The clustering is unsupervised, *i.e.* the number of classes k is not assumed to be known *a priori*.

The paper is organized as follows. Section 2 explains the prerequisites which this paper relies on, namely the colorspace and the definition of the connectedness degree of a color interval. The novel color classification scheme is the subject of Section 3. Section 4 asserts the relevance of the method by showing some results on real images. Finally, Section 5 proposes an application to skin detection.

2. PREREQUISITES

Color representation. In the present work which is focused on the processing of natural images, specular highlights are assumed to be insignificant. According to the dichromatic model [9] for lambertian objects, the colors of most natural images locate roughly along a finite number of straight lines in the RGB space, *i.e* along each body reflection vector, going from the RGB origin to the intrinsic color $\boldsymbol{c}_b$[1]. In the present work, the correspondence between the RGB coordinates $\boldsymbol{c}$ and the dichromatic model is made easier as colors are converted into a spherical frame (ρ, θ, ϕ), where ρ expresses the distance from the RGB origin $\boldsymbol{O} = (0,0,0)$ to the color $\boldsymbol{c}$ (then $\rho \simeq$ intensity). (θ, ϕ) express the chrominance, independantly from the luminance. Starting from this color representation called TPR, the procedure uses the Color Connectedness Degree (CCD) to deduce adaptively the main color classes.

The color connectedness degree. Let be a trichromatic image with components $\boldsymbol{c} = (c^1, c^2, c^3)$ and note $\boldsymbol{c}_i = (c^1{}_i, c^2{}_i, c^3{}_i)$

[1] For non-lambertian objects, specular reflection can also occur, producing specular vectors going from the body color to the illuminant color.

the color components of a pixel i at location $\boldsymbol{p}_i$. Assume that the chrominance is conveyed by components $(c^1{}_i, c^2{}_i)$ and the luminance is defined by the third one. A chrominance interval of size λ^2, the origin of which is the color $\boldsymbol{c}_i$, is defined as the color bin $I_i{}^\lambda = [c_i^1, c_i^1+\lambda][c_i^2, c_i^2+\lambda]$. The first order probability $\mathcal{P}_1(I_i^\lambda)$ is the probability that a pixel of color $\boldsymbol{c}_a$ belongs to I_i^λ. It is computed as the sum of all the first order probabilities $\mathcal{P}_1$ of the components $\boldsymbol{c} \in I_i^\lambda$.

Now, we define the co-occurrence probability of two colors $\mathcal{P}_{cc}(\boldsymbol{c}_a, \boldsymbol{c}_b)$ as the sum of all the occurrence probabilities of $\boldsymbol{c}_b$ in a 8-connected neighborhood $\mathcal{N}$ of a color $\boldsymbol{c}_a$. Then, the second order probability $\mathcal{P}_2(I_i{}^\lambda)$ of the color interval $I_i{}^\lambda$ is computed as the sum of the co-occurrence probabilities of all color couples $(\boldsymbol{c}_a, \boldsymbol{c}_b) \in I_i{}^\lambda$.Therefore, the connectedness degree of a color interval $\mathcal{D}(I_i{}^\lambda)$ is given as [7]:

$$\mathcal{D}(I_i{}^\lambda) = \frac{\mathcal{P}_2(I_i{}^\lambda)}{\mathcal{P}_1(I_i{}^\lambda)} = \frac{\sum\limits_{\boldsymbol{c}_a \in I_i{}^\lambda} \sum\limits_{\boldsymbol{c}_b \in I_i{}^\lambda} \mathcal{P}_{cc}(\boldsymbol{c}_a, \boldsymbol{c}_b)}{\sum_{\boldsymbol{c}_a \in I_i^\lambda} \mathcal{P}_1(\boldsymbol{c}_a)} \tag{1}$$

The CCD $\mathcal{D}(I_i{}^\lambda)$ is maximum when $I_i{}^\lambda$ corresponds to one only connected component in the image. The more regions, the lower the value of $\mathcal{D}$. Then, contrary to the correlogram or to the color histogram, a small (but perhaps salient) homogeneous region can get a high $\mathcal{D}$ value despite a small occurence probability $\mathcal{P}_1$.

Let us detail the procedure based on the analysis of the CCD in the TPR colorspace.

3. CLASSIFICATION METHOD

First of all, the RGB image is converted to TPR, and the saturation image S is computed. Then, an uniform quantization is performed on the given dynamics 2^M (M coding bits, 8 here), on N levels (intervals of size $\lambda = 2^M/N$). Then, chrominance and luminance are analyzed separately and the relative results are merged.

3.1. Chrominance analysis

The chrominance clustering is achieved following four main stages:

1) Computation of the color and spatial statistics of each color bin $\boldsymbol{b}_k$ of length λ ($N \times N$ bins): $\mathcal{P}_1(\boldsymbol{b}_k)$, $\mathcal{P}_2(\boldsymbol{b}_k)$, $\mathcal{D}(\boldsymbol{b}_k)$ and the coordinates $x_c(\boldsymbol{b}_k)$ and $y_c(\boldsymbol{b}_k)$ of the spatial centroids $\boldsymbol{p}_c$ in the image $\boldsymbol{c}$, of all the pixels with attributed color $\boldsymbol{b}_k$

2) Detection of the local maxima w.r.t. $\mathcal{D}_\lambda$. A local maximum $\mathcal{D}_m$ is detected at a bin $(\boldsymbol{b}_m)$ when the two following criteria $\mathcal{C}_1$ and $\mathcal{C}_2$ are met:

$$\mathcal{C}_1 = \mathcal{D}(\boldsymbol{b}_m) > \mathcal{D}(\boldsymbol{b}_v)\ \forall \boldsymbol{b}_v \in \mathcal{N}(\boldsymbol{b}_m) \tag{2}$$
$$\mathcal{C}_2 = \mathcal{D}(\boldsymbol{b}_m) > \mathcal{T}_1 \tag{3}$$

$\boldsymbol{b}_v$ is a color bin belonging to the 8-connected neighborhood of $\boldsymbol{b}_m$ (noted $\mathcal{N}(\boldsymbol{b}_m)$) in the $\mathcal{D}$ space. $\mathcal{T}_1$ is a threshold $\mathcal{T}_1 = 0.01 \times N_{pix}/(N \times N)$ designed to avoid insignificant local maxima. It is low enough not to be critical. N_{pix} is the number of pixels in the image. Additionally, when two local maxima are considered to be too close to one another (separated by only one chrominance bin), the lower value is removed.

3) Definition of the colorimetric subsets, *i.e* the sets of color bins. From the local maxima, the procedure consists in growing each color set (*i.e.* the union of color bins) in order to exhibit those of highest connectedness degree. Each bin in the (θ, ϕ) space is assigned to one of the classes of which the seed is a local maximum, according to a criterion jointly based on: the colorimetric distance in the (θ, ϕ) space; the centroid localizations in the image; the connectedness degree modification after assignment. A bin $\boldsymbol{b}_i$ is added to the mode which maximizes a similarity measure involving three different informations:

1. its chromaticity distance (euclidean distance) to the local maxima $\boldsymbol{b}_m$: $d^c(\boldsymbol{b}_i, \boldsymbol{b}_m)$
2. the spatial distance (euclidean distance) of the centroid $\boldsymbol{p}_c(\boldsymbol{b}_i)$ and the centroid of the local maximum $\boldsymbol{p}_c(\boldsymbol{b}_m)$: $d^s(\boldsymbol{b}_i, \boldsymbol{b}_m)$
3. the connectedness degree of the set of colors included in the union of the two bins $\boldsymbol{b}_i$ and $\boldsymbol{b}_m$: $d^{\mathcal{D}}(\boldsymbol{b}_i, \boldsymbol{b}_m) = \mathcal{D}(\boldsymbol{b}_i \bigcup \boldsymbol{b}_m)$

Finally, the similarity measure is written as:

$$\mathcal{S}(\boldsymbol{b}_i, \boldsymbol{b}_m) = \frac{d^c_{max}}{max(d^c(\boldsymbol{b}_i, \boldsymbol{b}_m), \varepsilon)} + \frac{d^s_{max}}{max(d^s(\boldsymbol{b}_i, \boldsymbol{b}_m), \varepsilon)} + \frac{d^{\mathcal{D}}(\boldsymbol{b}_i \bigcup \boldsymbol{b}_m)}{d^{\mathcal{D}}_{max}} \tag{4}$$

Note that each distance is normalized to the range $[0,1]$ in dividing it by its maximum value. ε denotes the lowest possible value of the criterion, it is used to avoid 0 at denominator.

4) Filtering of the chrominance classes, in order to eliminate classification defects. In a few words, a bin is re-affected when most of its neighbor pixels have a different class.

As an example, Fig.1(a) shows an example of an image to be processed[2]. Then, Fig.1 displays respectively: (b) the associate space of the CCD values ($N \times N$ structure); (c) the local maxima; (d) the color subsets. Finally, once each pixel has been assigned a *chrominance subset*, the chrominance clustering leads to the result of Fig.1(e).

3.2. Luminance (or ρ) analysis

The assumption is made, as in [10], that luminance information is useful only when colors are not saturated enough. Therefore when the mean saturation of a *chrominance subset* ranges below a given threshold, the class is declared unsaturated and may be divided. Our threshold is set to 20% as in [11] so that unsaturated and low saturated color subsets may be divided. Then, such a candidate color subset is subdivided only when its population is scattered enough in the

[2] Berkeley database: http://www.eecs.berkeley.edu/Research/Projects/CS /vision/grouping/segbench/

histogram. To that purpose, the energy is chosen as a scattering criterion, it is maximum for a perfectly homogeneous region. Therefore, when the energy $E(\rho)$ on the unsaturated color set is under a given threshold (0.8), then it is subdivided into *gray subsets* (*i.e.* ρ classes) following the same steps as above except for: 1) the procedure is simpler since the data is monodimensional; 2) only pixels defined as low saturated ($S < 20\%$) are processed.

Fig.1(f) shows the final chrominance/luminance clustering with $N = 64$. Note that not all the color subsets have been subdivided, namely the saturated color sets ($S > 20\%$) and/or the homogeneous color sets (with high energy).

For improved results, a Markov classification[3] can be preferred to the *Nearest Cluster Classification*. The result is shown on Fig.1(g).

4. QUALITATIVE ANALYSIS

The first experiments are based on the 24 images from the Kodak dataset [4] which is a representative and varied set of color images, composed of natural outdoor scenes, people, manufactured object, buildings. We propose to formulate the *goodness* of the classification as a trade-off between a large reduction of colors and regions in the image (compactness) with a satisfactory preservation of the original information (objective and subjective quality for example). Starting from that assertion, the four following comparison criteria are used: the *PSNR* (Peak Signal to Noise Ratio) computed between the *RGB* final classification result and the original image[5]; the *CIE*76 distance computed in the perceptually uniform colorspace *CIELAB*; the number of classes N_c and regions (8-connectedness) N_r. At first, the colorspace is TPR, the number of bins is $N = 64$, the colors are affected to the classes according to (4) (criterion noted $D+C+S$, for **D**egree, **C**olorimetric and **S**patial distance). The table 1 collects the results of each experiment (value of the mean criterion computed on the Kodak dataset) for 3 different studies (noted A, B, C) aiming at evaluating the quality and compactness of the proposed method. These results provide information about:

(A) the impact of N (quantization). Finally, $N = 64$ (1^{st} entry, table 1) of is a fair parameter, which provides a satisfying trade-off between quality and compactness. In addition, using a larger N means wasting much more memory and time.

(B) the impact of the colorspace. The proposed clustering technique can be used with any other chrominance/luminance colorspace. Note that not all colorspaces can reveal the color clusters so easily as the TPR colorspace. For comparison purposes, the classical *HSV* and the Otha independent components $I_1I_2I_3$ are tested. With HSV, the number of classes is very high for a lower *PSNR* and higher *CIE*76 (compared to 1^{st} entry). This colorspace is known to be sensitive to noise (ill-definition of hue at low saturation). Obviously, the use of $I_1I_2I_3$ leads to under-segmentation (low number of classes and regions but large reduction of *PSNR*). Finally, the quality is better with TPR and for a reasonable number of classes.

(C) the impact of the similarity criterion. We compare the use of the criterion $D+C+S$ for the clusters creation (1^{st} entry and eq.(4)) with the use of the mixture of color and spatial constraints ($C+S$). The mean *PSNR* for the criterion $C+S$, *i.e* without use of the CCD, is not significantly lower in average, however the number of classes is significantly higher. Indeed, the CCD contributes to enlarge color subsets. It favors relevant and large color subsets, naturally isolating the irrelevant color subsets related to sparse pixels in the image.

[3] This stage is not explained here for concision purposes, since it is not the core of the paper.

[4] http://r0k.us/graphics/kodak/

[5] Note that the *PSNR* values are low due to the huge reduction of the number of colors.

Table 1: Results of the evaluation criteria. (A) Impact of N: criterion $C+S+D$, TPR colorspace. (B) Impact of the colorspace, criterion $C+S+D$, $N = 64$. (C) Impact of the similarity criterion, TPR colorspace, $N = 64$.

		PSNR	*CIE*76	N_c	N_r
(A)	$N = 64$	20,19	10,48	22,38	3550,08
(A)	$N = 32$	18,02	14,31	5,83	801,17
(A)	$N = 96$	20,68	9,43	51,79	7789,58
(B)	*HSV*	19,59	10,47	92,83	13754,21
(B)	$I_1I_2I_3$	18,69	13,67	16,13	2423,13
(C)	C+S	18,94	12,74	17,88	2689,54

5. APPLICATION TO SKIN DETECTION

The detection uses the proposed classification in conjunction with the contribution of [1] which proposes simple way to select *skin* colors in an image, by multi-thresholding in the *RGB* colorspace. For each color cluster extracted previously, the number of *skin* pixels n_{skin} is counted. When the ratio R of n_{skin} on the number of pixels in the cluster is high enough ($R > 0.5$), the corresponding cluster is considered as a potential *skin* cluster. The relative pixels are then segmented into regions [12]. The selection of *skin* regions is then based on a simple size criterion. The regions the size of which is lower than 5% of the maximum size (size of the largest resulting region) are removed. Indeed, it is assumed that people occurring in the same scene appear with a comparable scale in the image.

6. CONCLUSION

A novel adaptive color classification procedure was designed upon the connectedness degree of color subsets. The color representation stems from the dichromatic model, in order to better stress *a priori* the different body colors in the image. The analysis is completed separately on chrominance and then on luminance when necessary. The most sensitive

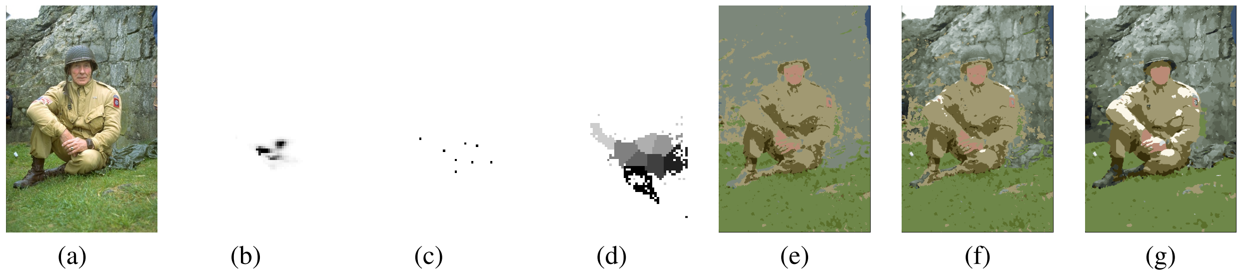

Fig. 1: (a) Initial image, (b) its $\mathcal{D}$ space. (c) Detection of local maxima. (d) Color subsets. (e) Chrominance clustering in (θ, ϕ). (f) Final clustering. (g) Markov clustering. The parameters are $N = 64$,association criterion $D + C + S$, TPR colorspace.

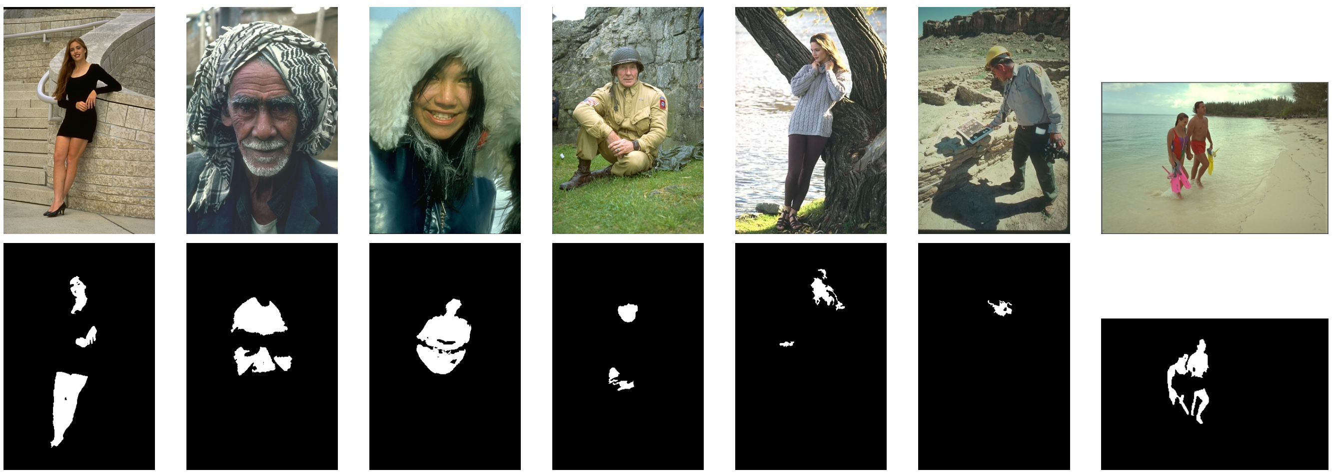

Fig. 2: Examples of skin detection with $N = 64$ in TPR, with use of CCD. Berkeley database.

parameter of this procedure turns out to be the quantization step λ. However $\lambda = 4$ leads to satisfactory results for most images. The method has been successfully applied to skin detection, but it is generic enough to be used for other problematics adressed by interactivity and multimedia applications. Now, the method has to be accelerated (SIMD, OpenMP) in order to enable a real-time execution (25 fps) for interactivity purposes.

7. REFERENCES

[1] J. Kovac, P. Peer, and F. Solina, "Human skin colour clustering for face detection," in *EUROCON 2003, International Conference on Computer as a Tool*, 2003.

[2] Li Peihua, "A clustering-based color model and integral images for fast object tracking," *Signal Processing: Image Communication*, vol. 21, no. 8, pp. 676–687, September 2006.

[3] S.T. Birchfield and S. Rangarajan, "Spatiograms versus histograms for region-based tracking.," in *CVPR*, 2005, pp. 1158–1163.

[4] N. Wada, S. Kaneko, and T. Takeguchi, "Using color reach histogram for object search in colour and/or depth scene," *Pattern recognition*, vol. 39, no. 5, pp. 881–888, 2006.

[5] H. Nezamabadi-Pour and E. Kabir, "Image retrieval using histograms of uni-color and bi-color blocks and directional changes in intensity gradient," *Pattern Recognition Letters*, vol. 25, no. 14, pp. 1547–1557, 2004.

[6] R.M. Haralick, K. Shanmugam, and Its'hak Dinstein, "Textural features for image classification," *IEEE Trans. on Systems, Man & Cybernetics*, vol. 3, no. 6, pp. 610–621, November 1973.

[7] M. Fontaine, L. Macaire, and J-G. Postaire, "Unsupervised segmentation based on connectivity analysis," in *ICPR*, Spain, 2000, vol. 1, pp. 600–603.

[8] L. Macaire, N. Vandenbroucke, and J-G. Postaire, "Color image segmentation by analysis of subset connectedness and color homogeneity properties," *CVIU*, vol. 102, no. 2006, pp. 105–116, March 2006.

[9] S.A. Shafer, "Using color to separate reflection components," *Color Research and Applications*, vol. 10(4), pp. 210–218, 1985.

[10] T. Carron and P. Lambert, "Color edge detector using jointly hue, saturation and intensity," in *IEEE ICIP*, November 1994, vol. 3, pp. 977–981.

[11] S. Sural, G. Qian, and S. Pramanik, "Segmentation and histogram generation using the hsv color space for image retrieval," in *in IEEE ICIP*, 2002, pp. 589–592.

[12] L. Lacassagne and B. Zavidovique, "Light speed labeling: efficient connected component labeling on risc architectures," *Journal of Real-Time Image Processing*, pp. 1–19, 2009.