

Normalized colour segmentation for human appearance description *

Robert Benavente, Gemma Sánchez, Ramon Baldrich, Maria Vanrell, Josep Lladós
Computer Vision Center, Dept. Informàtica.
Universitat Autònoma de Barcelona,
08193 Bellaterra (Barcelona), Spain
{robert,gemma,ramon,maria,josep}@cvc.uab.es

Abstract

In this paper we present a colour segmentation method based on a normalized colour naming algorithm which removes the effects of the varying conditions due to changes in scene illuminant. Images labelled with the colour name and intensity of small regions are further processed by a region growing step providing a sound segmentation. The method has been tested on a large set of images we get from a surveillance system, whose goal is the automatic retrieval of people from an image database using their appearance description. It is given in terms of placement of the colour regions in clothes. Finally, a quantitative measurement to evaluate the performance of the algorithm has been defined.

Keywords: Colour naming, colour normalization, region growing, surveillance, performance evaluation.

1 Introduction

Segmentation consists in partitioning the image into some non-intersecting regions according to a certain homogeneity criterion. A broad set of segmentation methods can be found in the literature. Often they are domain-dependent and, up to now, there is no general method able to segment any image. A complete survey of segmentation methods can be found in [11]. Criteria to segment images can be given in terms of discontinuity or similarity of regions. Thus, we have edge-based [10] or region-based methods [7], respectively. Region-based methods intend to map the properties of scene surfaces to homogeneity criteria of image pixels. The two visual cues explaining surface properties of a scene are colour and texture. In this work we present a region-based method dealing with colour properties of scene surfaces.

A common colour property of a surface is its colour

name. It is the adjective that natural language uses to describe the physical effect of the light reflected by a surface. Colour constancy ability of the human visual system usually makes the mapping between a surface and its colour name to be unique. Therefore, any colour naming method claiming uniqueness will have to consider the colour lighting conditions of a scene and its variability. The method we present in this work tries to solve these two problems by introducing a colour normalization jointly with a learning step. It avoids to recover lighting conditions of the scene in order to be subtracted from the image being processed, which is what colour constancy techniques usually do [5, 9].

Color naming is a key issue in the application we are working on. The application purpose is to retrieve personal data of people based on queries on its appearance. The final goal is people identification while they are inside a building with highly restricted access. Each database entry consists of three fields namely, personal data, an image acquired by the surveillance camera while the person is in front of the reception desk and finally, its appearance description. Last item is automatically introduced by the vision system that we are developing. Since retrieval queries will be made by the security staff, the appearance descriptions have to be stored in terms of "dark hair", "red tie", "dark blue shirt".

In order to automatically construct the above descriptions, colour segmentation will be a basic step. Once the foreground, i.e. people silhouette, is discriminated from the scene background, the subsequent processing and interpretation will be mainly based on colour regions we will extract. Since the colour name will be an essential feature for appearance description we have developed a colour segmentation method based on colour names. The method has three main steps: a colour normalization of the whole image, an image labelling using colour names and a segmentation refinement by a region-growing algorithm. Next section is devoted to introduce this method in more detail and, afterwards, we show how it performs on a large set of images of the application explained above. Finally, we discuss these results and the future work.

* Work partially supported by projects 2FD97-1800 and TAP98-0618 of Spanish CICYT, by UAB FIPD grant and by Casinos de Catalunya.

2 Segmentation based on colour names

The colour segmentation method we propose in this section assumes that any meaningful region can be described by its colour name. The presence of any texture in the region would complicate it, therefore we are just directed to label regions with a more or less homogeneous colour.

The main problem that any colour-based vision task has to face up is that colour perception is related to the illuminant spectral composition. Thus, the same surface may present very different appearances under different illuminants. The human visual system has an adaptive mechanism that allows avoiding the variations of the scene light and assign constant colour names to surfaces. Since common cameras do not have this ability, dealing with the illumination variability is one of the main problems in computer vision. In order to solve the problem of colour variability, many techniques have been developed during the last decade. These computational approaches try to introduce the chromatic adaptation ability of the human visual system. A good comparison between some of these techniques can be found in [6].

Our colour naming method will deal with the colour variability problem by applying the colour normalization presented by Finlayson et al. in [4]. This method is fairly simple because it only works on colour pixels of the image and it does not need any other calibration procedure. In the definition of this normalization, two important assumptions are made: a linear response to intensity changes and a Von Kries model for the chromatic adaptation. These two assumptions define a linear diagonal model which is usually assumed in computer vision [3]. The proposed normalization has two steps. The first one is an *intensity normalization*, that is, a transformation $R(I)$ to chromaticity coordinates defined as:

$$R(I)_{i,j} = \frac{I_{i,j}}{\sum_{k=1}^3 I_{i,k}} \quad (1)$$

where I is the image to normalize, and the subscripts i, j are referred to the i -th pixel of the j -th channel of I . The second step is a *channel normalization*, $C(I)$, that avoids the effect of the illuminant colour. It is expressed as:

$$C(I)_{i,j} = \frac{P/3 \cdot I_{i,j}}{\sum_{k=1}^P I_{k,j}} \quad (2)$$

where P is the number of pixels in I . Colour normalization is the result of the iterative application of R and C on the image. Then, the *normalized image* will be defined as:

$$N(I) = C(R(C(R(\dots(C(R(I))\dots)))))) \quad (3)$$

The above iterative process will finish when transformations imply very small changes on the image. In our application seven iterations usually suffice

Given the three-channel representation of a pixel, (n_r, n_g, n_b) , of the image $N(I)$, it can be assured that $n_r + n_g + n_b = 1$, that is, all the normalized coordinates are on the plane, $R + G + B = 1$, of the RGB colour space. Therefore, normalized colour information can be considered on a two-dimensional space without loss of information. Hence, we define

$$\mathcal{D}(I_i) : (n_r, n_g, n_b)_i \longrightarrow (n_u, n_v)_i \quad (4)$$

where (n_u, n_v) are the normalized chromaticity coordinates and subscript i denotes the i -th pixel of I . These normalized chromaticity coordinates represent the projection of the colour information on the plane $R + G + B = 1$ after the effects of the light source on the scene have been removed. For the rest of this paper, we will refer to them as the *Normalized Chromaticity Coordinates (NCC)*, and the two-dimensional space will be referred as the *Normalized Chromaticity Diagram (NCD)*.

Given that these normalized coordinates only depend on the image information, they can not be considered as a standard illuminant-independent space. The suitability of this normalization for colour naming has been widely analyzed in [1], where it has been shown that the only constraint it requires is to have a common context in all the images to be labelled. It is clear that this constraint is fulfilled by the application we have introduced in section 1, where all images share the same scene background.

The set of names we will consider in our method is the same as the one used by Berlin and Kay in [2]. It is formed by: grey, blue, green, yellow, orange, red, pink and purple. To extend this list to a wider set of colours we will add an intensity descriptor, that is, light, normal or dark. The combination between certain colour and an intensity descriptor will allow to distinguish other common colour names. Thus, light grey is white, dark grey is black, and dark orange is brown. Therefore, our colour naming algorithm will involve two processes, one for colour description and a second for intensity description. In both cases we will need two phases: learning and naming.

Learning.

The learning process has two goals: first to arise a tessellation of the *NCD*, where every region will represent all the normalized coordinates of a colour label; second, to define two thresholds for every colour that will represent intensity descriptors.

To achieve these goals, a learning set LS of images containing regions of homogeneous colour, will be used. The colour name and the intensity descriptor of each region in the images of LS are known and will be denoted by l_h and l_i , respectively. The learning process performs in this way:

1. For every image, $I \in LS$ we obtain their normalized coordinates by computing $\mathcal{D}(N(I))$.

2. For every region, r_t , from an image $I \in LS$, we calculate the average of the NCC computed in step 1 of the pixels in the region. These average point, (n_u^t, n_v^t) will be associated to the region colour label, l_h^t .
3. The convex hull of the average points with the same label will be the basis to tessellate the NCD .
4. For every region r_t , from an image $I \in LS$, we compute the region intensity as

$$i^t = \frac{1/Q \cdot \sum_{k=0}^Q i_k^t}{1/P \cdot \sum_{k=0}^P i_k} \quad (5)$$

where i_k is the length of the colour vector of the pixel k of an image, I , i.e. $i_k = (I_{k,1}^2 + I_{k,2}^2 + I_{k,3}^2)^{\frac{1}{2}}$, Q is the number of pixels in region r_t and P is the number of pixels of image I . Global image intensity has to be considered in order to avoid dependency on changes in lighting conditions.

5. For all the regions with the same l_h , we get the values allowing to separate i^t in three intensity degrees, light, normal and dark, given by l_i of regions. These values will be the intensity thresholds denoted as: $(\tau_{l_h}^1, \tau_{l_h}^2)$.

Naming.

Naming process of a region, r_t of an image I is a simple mapping between image regions and labels. This transformation will be denoted as NM , and is given by

$$NM : r_t \longrightarrow (l_h^t, l_i^t, n_u^t, n_v^t) \quad (6)$$

where the point (n_u^t, n_v^t) is the average of the region, r_t on the image $\mathcal{D}(N(I))$. l_h^t is the label corresponding to this point within the NCD , and, l_i^t is assigned according to the corresponding values of τ^1 and τ^2 for this colour.

This algorithm has been tested on a large set of images of the same application, where images were labelled by human operators being at the reception desk. It clearly improves previous tests using the RGB space or its chromaticity coordinates. Normalized coordinates allow an improvement from 50% of success to 80% on homogeneous regions. See [1] for a further study.

Once the colour naming algorithm has been introduced, a colour segmentation strategy is directly derived by applying the NM mapping on the regions define by every pixel neighbourhood of an image. This strategy presents three kind of problems when we apply it in our application. First of all, as the images have noise, an oversegmentation appears with very small regions we are not interested in. The noise is due to the acquisition device and the uncontrolled environmental conditions, such as light that causes specular reflection which are not taken into account in the normalization algorithm. Secondly, the segmentation precision we

need is lower than the one we have. There are some details, such as buttons and little patterns due to the clothes fabrics, that also result in an oversegmented image. The third problem is the division of areas with a hue which lies near the border between two different regions of the NCD . As the division of the different hues in the chromaticity diagram is fixed in the learning step of the algorithm, the annotation of ambiguous colours such as greenish-blue, might result in the splitting of the area into two different regions with a different label each one.

To solve these problems, in a refinement step of the segmentation, the labelled regions are clustered according to area, chromaticity and intensity criteria. Thus, following the idea given in [8], regions are organized in a graph structure and they are clustered in terms of a pyramidal graph contraction procedure. It consists in a bottom-up approach in which, at every iteration k , a more reduced set of regions is built from the set of regions of level $k - 1$. It is performed by merging neighbouring regions according to a distance measure define in terms of chromatic information. Regions are represented by an *attributed region adjacency graph* structure $G(V, E, \nu)$ where each vertex $r_i \in V$ is a region, and the set of edges E represent neighbouring relationships between regions. The graph is labelled by vertices, i.e. a labelling function $\nu(r_i) = (|r_i|, n_u^i, n_v^i)$ associates to each region r_i its area, and the average of the NCC according to eq. 6.

Let $G_k(V_k, E_k, \nu_k)$ be a graph of level k . Given two regions $r_i, r_j \in V_k$ such that there exists an edge $e \in E_k$ joining r_i and r_j , the *normalized chromaticity distance* is define as follows:

$$ncd(r_i, r_j) = [(n_u^i - n_u^j)^2 + (n_v^i - n_v^j)^2]^{\frac{1}{2}} \quad (7)$$

If r_i and r_j are not neighbouring regions or they have different intensity, it is define as $ncd(r_i, r_j) = \infty$.

The naming process define in eq. 6 results in an initial set of regions $V_0 = \{r_i\}$ labelled according to the mapping NM . Thus, the initial graph $G_0(V_0, E_0, \nu_0)$ is constructed from the above set of regions. However, before start the region growing process, small regions, which are likely to be produced by noise, are removed from this graph. Thus, each region $r_i \in V_0$ such that $|r_i| < T_a$ is absorbed by its neighbouring region r_j with minimum distance $ncd(r_i, r_j)$.

At iteration $k + 1$, a new graph $G_{k+1}(V_{k+1}, E_{k+1}, \nu_{k+1})$ is constructed from $G_k(V_k, E_k, \nu_k)$ by merging neighbouring regions in terms of the distance ncd . To do this, a weight w_i is assigned to each graph vertex $r_i \in V_k$:

$$w_i = \frac{\sum_{r_j \in n(r_i)} ncd(r_i, r_j)}{|r_i|} \quad (8)$$

where $n(r_i)$ is the set of neighbouring regions of r_i . The weight w_i conveys two concepts: the average distance between the region r_i and its neighbours, and its size. Those

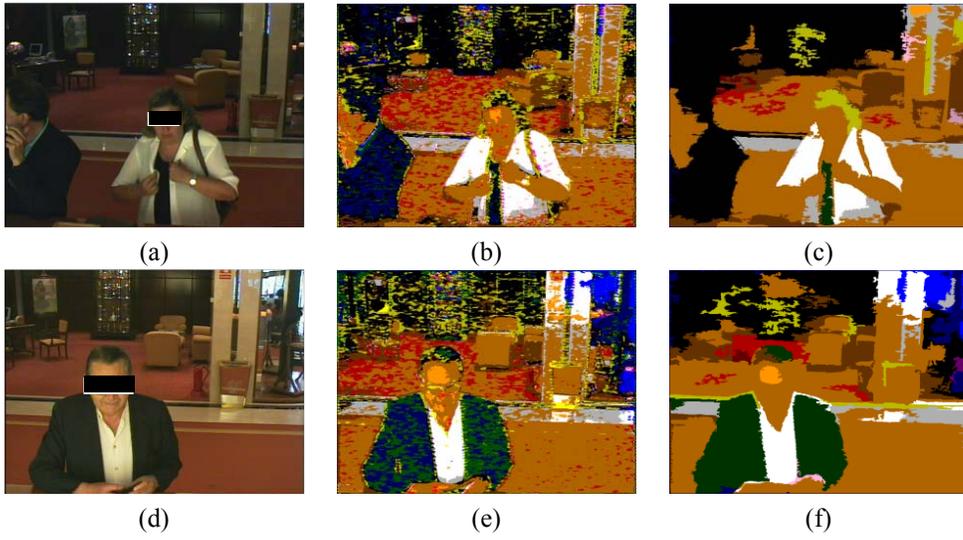


Figure 1. Segmentation results. (a)(d) original images, (b)(e) segmentation after normalized colour naming, (c)(f) segmentation after region growing refinement

vertices with minimum weight among their local neighbours are selected as survivors and initialize the set of vertices V_{k+1} of level $k + 1$. Afterwards, each non-surviving vertex is merged with its neighbouring surviving vertex with minimum ncd . If the distance is greater than a threshold T_{ncd} , it is also selected as a survivor. Informally, this graph contraction procedure can be explained as follows: the greatest regions similar to their corresponding neighbours in terms of ncd act as seems of a new set of regions V_{k+1} , then the remainder regions of V_k are merged with closest regions under the threshold T_{ncd} . Once the new graph G_{k+1} is constructed, the function ν_{k+1} is computed. The algorithm stops when $G_k = G_{k+1}$ and G_k represents the segmentation result.

3 Performance evaluation

Segmentation is often evaluated visually or in terms of the performance of subsequent domain-dependent processes. However, some efforts have been made to design suitable sets of ground truthed data and evaluation mechanisms able to automatically validate the performance of the algorithm. Zhang [12] distinguished two categories of analytical methods for image segmentation evaluation: the goodness methods and the discrepancy methods. In the first category, some quantitative measures such as inter and intra region homogeneity are extracted from the segmented image and used as parameters of an evaluation function. The second category uses a reference of expected segmentation which is compared with the actually segmented image to

give a discrepancy measure between them. The latter category is better for objectively assess segmentation algorithms. In this work, we have formulated a discrepancy method. For each input image, a synthetic image containing the expected regions has been generated. Then, a distance between the segmented image and the synthetic one has been defined. Thus, Let r_e and r_s be a region of the synthetic image and a region obtained by the segmentation algorithm, respectively. The distance $d(r_e, r_s)$ between these two regions is define as a weighted sum of the colour distance d_c and the overlapping distance d_o :

$$d(r_e, r_s) = w_c d_c(r_e, r_s) + w_o d_o(r_e, r_s) \quad (9)$$

To facilitate the comparison between different images, the above distances are normalized between 0 and 1.

The colour distance is formulated as the Euclidean difference between the expected colour of the region and the actual colour once the region has been labelled, i.e. according to eq. 7, it can be define as $d_c(r_e, r_s) = ncd(r_e, r_s)$. The overlapping distance is define as follows:

$$d_o(r_e, r_s) = 1 - \min\left(\frac{|r_e \cap r_s|}{|r_e|}, \frac{|r_e \cap r_s|}{|r_s|}\right), \quad (10)$$

where $|r|$ denotes the area of the region r and $r_e \cap r_s$ denotes the intersecting region between r_e and r_s .

Consequently, the distance between an expected region r_e and the set $S = \{r_s^i\}$ obtained by the segmentation can be define as follows:

$$D(r_e, S) = \min_i d(r_e, r_s^i). \quad (11)$$

Thus, given two sets of regions $E = \{r_e^i\}$ and $S = \{r_s^j\}$, the evaluation function that measures the dissimilarity, i.e. the accuracy of the segmentation, is defined as:

$$F(E, S) = \frac{\sum_{i=1}^{card(E)} D(r_e^i, S)}{card(E)}. \quad (12)$$

Figure 1 shows some particular results. For the two cases presented, the original image, the naming result and the final segmentation are displayed. In the annotation results (Figs. 1(b) and 1(e)) we can observe the oversegmentation produced by noisy input images, solved by the region growing step. On the other hand, the problem of annotation in uncertainty regions of *NCD* due to ambiguous colours, as described in Section 2, can be assessed in the second result. Here, the jacket could be labelled as "greenish-blue", and it results in regions labelled as "green" and other ones as "blue" in Fig. 1(e). The region growing step merges these sub-regions and a final uniform region is obtained.

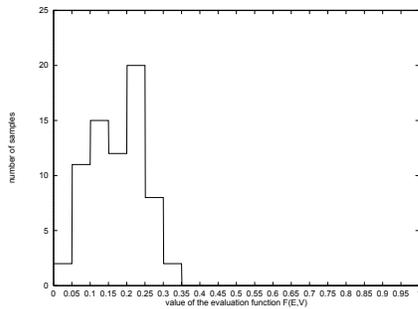


Figure 2. Dissimilarity measurements histogram among a set of 70 samples.

The experimentation set consisted of 70 images acquired by the surveillance application described in Section 1. A quantitative study is reported in Fig. 2. This graphic plots the density of dissimilarity values computed from the evaluation set. We can observe that in all samples, the dissimilarity value is less than 0.4, and a very high percentage is around 0.2. The subsequent step in our application consists in an interpretation of region structure, matching it with a set of structural patterns that describe known clothing configurations. The experimentation confirms that the image is likely to be understood as long as region structure and colour are basically preserved. Thus, the obtained dissimilarity values are good enough in our application.

4 Conclusions

In this paper we have proposed an algorithm for colour image segmentation based on normalized colour features towards giving symbolic descriptors. A colour normalization

of the input image followed by an image labelling using colour names has been used to do a coarse segmentation. Afterwards, a refinement step based on a region-growing process has been performed to solve oversegmentations due to noise and annotations of ambiguous colours. The segmentation algorithm has been used as a low level step in a surveillance application consisting in people appearance description. The performance of the algorithm has been quantitatively measured with a comprehensive set of test images.

Once the results show the suitability of the algorithm for the application where it is used, some efforts have to be done to improve the learning step developing an automatic procedure able to infer the right *NCD* tessellation and the intensity thresholds. Secondly, domain-dependent knowledge could guide the region growing, that is, since segmentation is applied to people images, simple models about clothing and physiognomic features could contribute to overcome erroneous segmentations due to noise, ambiguous colours, etc. Currently, we are working with images acquired in the actual environment of the surveillance application. Our segmentation algorithm is being thoroughly tested with a prototype of that application.

References

- [1] R. Benavente. Dealing with colour variability: application to a colour naming task. Technical Report 32, Computer Vision Center, 1999.
- [2] B. Berlin and P. Kay. *Basic Color Terms: Their Universality and Evolution*. University of California Press, Berkeley, 1969.
- [3] G. Finlayson, M. Drew, and B. Funt. Diagonal transforms suffice for color constancy. In *Proceedings of the 4th ICCV '93*, pages 164–171, 1993.
- [4] G. Finlayson, B. Schiele, and J. Crowley. Comprehensive colour image normalization. In *Proceedings of the 5th ECCV '98*, pages 475–490, 1998.
- [5] D. Forsyth. A novel algorithm for color constancy. *International Journal on Computer Vision*, 5(1):5–36, 1990.
- [6] B. Funt, K. Barnard, and L. Martin. Is machine colour constancy good enough? In *Proceedings of the 5th ECCV '98*, pages 445–459, 1998.
- [7] T. Gevers. *Color Image Invariant Segmentation and Retrieval*. PhD thesis, University of Amsterdam, 1996.
- [8] S. W. Lam and H. H. Ip. Structural texture segmentation using irregular pyramid. *Pattern Recognition Letters*, pages 691–698, July 1994.
- [9] L. Maloney and B. Wandell. Colour constancy: a method for recovering surface spectral reflectance. *Journal of the Optical Society of America*, 3(1):29–33, 1986.
- [10] R. Nevatia. A color edge detector. In *Proceedings of the IJCPR-III*, pages 829–832, 1976.
- [11] N. Pal and S. Pal. A review on image segmentation techniques. *Pattern Recognition*, 26(9):1277–1294, 1993.
- [12] Y. Zhang. A survey of evaluation methods for image segmentation. *Pattern Recognition*, 29(8):1335–1346, 1996.