STATISTICAL MODELLING FOR IMAGE RETRIEVAL
USING A BIOLOGICAL MODEL OF THE PERCEPTIVE COLOUR SPACE

Anne Guérin-Dugué(1), Christophe Biernacki(2), Jeanny Hérault(3)
(anne.guerin@inrialpes.fr, christophe.biernacki@math.univ-fcomte.fr, jeanny.herault@inpg.fr)

1. INTRODUCTION

The use of colour in computer vision has received growing attention for segmentation compression or indexing, see [8] for a review at this date. For image retrieval systems, colour information is one of the most important features. In this work, we propose a perceptive colour space based on a biological model of the retina. In this space, a model of Gaussian mixture appears very efficient for colour distribution. Two strategies are presented, retrieval by maximum likelihood on (i) the global chromatic distribution, (ii) the local spatio-chromatic distribution.

2. COLOUR REPRESENTATION

2.1. Perceptive Colour Model

In the retina of man and primates, the three kinds of photoreceptors are known to present a non-linear and adaptive transduction function. If is the image of the i-th colour component activating the photoreceptors of class i, the response of this class is a compressed image:

\[ c_i(x,y) = \frac{C_j(x,y)}{C_j(x,y) + h(x,y) * C_i(x,y)} \]

where the term \( h(x,y) * C_i(x,y) \) is a kind of local mean of the input image, obtained by the convolution with a normalised low-pass kernel: \( h(x,y) \).

The result is that (i) each photoreceptor produces an output image which is more compressed in highlights, and (ii) the resulting colour of each pixel is affected by the local surrounding colour of the scene, i.e. mainly the colour of illumination.

This second property is very interesting because it induces a kind of colour-constancy behaviour: due to the high-pass behaviour of the compression law, the colour variations are mainly due to the objects in the scene (high-pass signal) than to the slowly varying illumination (low-pass signal).

Then if we name \( R(x,y) \), \( G(x,y) \) and \( B(x,y) \) the "stimulation colour space", we can name the compressed signals \( r(x,y) \), \( g(x,y) \) and \( b(x,y) \) the "perceptive colour space" from (1). Then we can derive three more orthogonal signals: \( l_c = \frac{r(x,y) + g(x,y) + b(x,y)}{3} \), \( a_c = r(x,y) - g(x,y) \), \( b_c = b(x,y) - \frac{r(x,y) + g(x,y)}{2} \). These three signals define the new perceptive compressed colour space with the corresponding lumiance \( l_c \) and chrominance \( a_c, b_c \) components. In the following, this space is called "Compressed AB".

2.2. Classical Colour Spaces

Two classical colour spaces have been selected for comparison [12], (i) the "YChCr" (linear) space used for compressed images in the JPEG format, and (ii) the "LAB" empirical perceptive (non-linear) space defined by the CIE. Non-linear functions are introduced in relation to perceptive data [9], but not based on a biological model contrary to our "Compressed AB" space.

3. STATISTICAL MODELIZATION

3.1. Related work

From the initial study presented by Swain and Ballard [11], colour information becomes one of the main features
for image segmentation and indexing. The colour distribution of an image is a global characterisation with a loss of spatial information. Usually the empirical histogram models this colour distribution. It is advantageously simple and fast, but it is also sensitive to the bin size. Here, we use an alternative model by Gaussian mixture [2, 4, 6]. Compared to the histogram technique, the advantages are the reduced number of parameters, the regularisation effect on noisy data and the flexibility of the method providing more elaborated strategies. Conversely, it is more time consuming.

To recover the spatial organisation of the image, several strategies may apply. In [4], all the features (spatial co-ordinates (2), texture (3), colour (3)) are gathered for both modelling by a Gaussian mixture in 8 dimensions and segmentation. The integration of the spatial location in the features vector produces miss-classification in too large homogeneous areas which are sometimes arbitrary split into two pieces. In [6] to overcome this difficulty, the spatial information is integrated after an image segmentation into homogeneous colour region. The model of the spatio-chromatic distribution of each segmented area is then an unique Gaussian mode in 5 dimensions (colour (3), spatial co-ordinates (2)).

The study presented here has been inspired by [6]. The differences are in the implementation method, the choice of the colour space, the suppression of the luminance information and the constraints for the spatio-chromatic covariance matrix. All these choices will be justified.

3.2. Model Description

The model of the spatio-chromatic distribution is a Gaussian mixture according to a three-steps strategy, (i) global chromatic modelling, (ii) spatial classification, (iii) spatio-chromatic modelling. Let us notice \( X^{\text{c}} = \{ c_1(i), c_2(i); i=1\ldots n \} \), the 2D colour information of n pixels from the image (fig. 1b). The global colour distribution is estimated by a Gaussian mixture:

\[
  f(c) = \sum_{k=1}^{K} p_k N(c; \theta_k) \quad (2)
\]

with \( c=(c_1,c_2) \), \( p_k \) the weight \( (p_1+\ldots+p_K=1) \), \( N(\cdot; \theta_k) \) the Gaussian density and \( \theta_k \) its parameters (mean and covariance). The Expectation-Maximisation algorithm is used to determine the maximum likelihood parameters of the mixture with \( K \) Gaussians [5]. Following [3], the unknown number of mixture components can be chosen by optimising the ICL information criterion (Integrated Completed Likelihood):

\[
  \text{ICL} = \text{BIC} - 2.24 \ln(n) - 2.24 \ln(l_{\Theta}) \quad (3)
\]

with \( l(\hat{\Theta}) \) the maximum likelihood of the parameters, \( d_f \) the number of free parameters needed for a model with \( K \) Gaussians and \( E \) the entropy of the colour partition. ICL can be interpreted as a penalisation of the Bayesian Information Criterion (BIC) or equivalently of the Maximum Description Length criterion (MDL). ICL is more robust and parsimonious than BIC : the number \( K \) of selected modes is lower [3].

To model the spatio-chromatic distribution \( X^{\text{sc}} \), \( X^{\text{sc}} = \{ c_1(i), c_2(i), x(i), y(i); i=1\ldots n \} \), the intermediate step is a spatial classification of all the pixels in one of the \( K \) chromatic modes (fig. 1c). After running a connected-component algorithm, \( K_{\text{sc}} \) (superior or equal to \( K \)) spatially connected components are obtained. The last step is the Gaussian modelling in 4D \( (c_1, c_2, x, y) \) of each connected spatial component (fig. 1d). This approach [6] is suitable for spatial components with an ellipsoid-like shape. If not, the model of the spatial components can be a Gaussian mixture on the spatial dimensions (fig. 1e).

Let us remark in fig. 1d, how the blob corresponding to the sky is split into 2 smaller blobs in fig. 1e to better fit the spatial organisation of the image.

For each image, the index is this new vector of parameters \( \hat{\Theta}_{\text{sc}} \), characterising the spatially localised colour distribution of the image. Contrary to [6], the covariance model is constrained with null spatio-chromatic co-variances (4 values). These parameters provide a too fine information level : this is the colour gradation inside each connected spatial component.
3.3. Results

The database (4600 images) has been provided by a photographer. Three hundred images have been selected according to five selected contexts (city, forest, indoor scene, beach, desert). Figure 2 illustrates this approach with a city photo. The model of the empirical distribution (fig. 2.b) is a mixture of 3 Gaussian modes.

Comparing the three selected colour spaces, we have in average a more parsimonious model with our “Compressed AB” space. In this space, the distribution is more Gaussian and then the model by Gaussian mixture fits very well. Consequently, the spatio-chromatic modelling is also more parsimonious (table 1 and fig. 2).

4. CHARACTERISATION AND IMAGE INDEXING

4.1. Methodology for the global characterisation

The global colour distribution of each image \( I^{(l)} \) in the database is characterised by its parameters vector \( \hat{\Theta}^{(l)} \) for the \( K^{(l)} \) Gaussians in the mixture. Let us consider the query by the presentation of an image \( I \). To retrieve images relatively to this query, we compute the likelihood of the parameters vector \( \hat{\Theta}^{(l)} \) characterising the images \( I^{(l)} \) in the database, relatively to the image \( I \). The retrieved images are sorted by decreasing order of the likelihood. Let us assume the equi-probability of all the images in the database, then we have:

\[
P_r(I^{(l)} / I) \propto P_r(I / I^{(l)}) = \ell(\hat{\Theta}^{(l)} / X)
\]

\[
\ell(\hat{\Theta}^{(l)} / X) = \prod_{l=1}^{L} \sum_{k=1}^{K^{(l)}} p_k^{(l)} N(c(i);\Theta_k^{(l)}),
\]

with \( X \) the set of observations extracted from the image \( I \). Only 1% of all the pixels are selected at random from \( I \) to compute this likelihood. The same percentage has been used to run the “EM” algorithm for computing the index \( \hat{\Theta}^{(l)} \). Colour information characterises regions. Only relevant regions are then sampled and the smaller regions are statistically neglected with a such sampling ratio. Consequently processing time is then reduced.

The precision criterion assesses the efficiency of such a strategy. It is the ratio of relevant images over the number of retrieved images. An image is considered as relevant if it belongs to same category than the query image.

4.2. Methodology for the local characterisation

The methodology is the same as previously but the likelihood is computed with the spatio-chromatic index \( \hat{\Theta}_{SC}^{(l)} \) and the spatio-chromatic set of observations from \( I \) (\( X_{sc} \)). On the precision measures, we have noticed better results by constraining the spatio-chromatic co-variances with null values. In fact, a full spatio-chromatic co-variance matrix provides a too fine modelling level and then constraints the image matching too much.

5. EXPERIMENTAL RESULTS

We have compared for the three colour spaces, six strategies for image indexing and matching :

- Indexing by the empirical histogram of the global colour distribution and matching by the Chi2 distance, the Jeffrey divergence and the histogram intersection [10],
- Indexing by the Gaussian mixture on the global colour distribution and matching by the likelihood (method called “GD:GM”),
- Indexing by the Gaussian mixture on the local colour distribution into two cases, one Gaussian mode per spatially connected component (“GD:GM; LD:1G”) or \( \eta \) modes per component- selected by the ICL criteria- to model non ellipsoid-like shape (“GD:GM; LD:GM”). The matching is implemented by the likelihood.

The luminance information is not used here. Processing orientation features from luminance image is an efficient way to integrate luminance in a retrieval system [7]. This question is not addressed here while the main topic concerns colour distribution modelling.

The following table summarises the results on the average precision integrated up to the rank 10.

<table>
<thead>
<tr>
<th></th>
<th>“AB”</th>
<th>“Comp. AB”</th>
<th>“CbCr”</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \overline{K} )</td>
<td>4.25</td>
<td>3.60</td>
<td>4.84</td>
</tr>
<tr>
<td>( \overline{K}_{SC} )</td>
<td>5.70</td>
<td>5.49</td>
<td>7.20</td>
</tr>
</tbody>
</table>

Table 1 : Average value of the number of Gaussian mode for the global colour distribution.

Figure 2 : Histogram of the optimal value \( K \) selected by the ICL criterion on all the images for respectively (a) “AB”, (b) “Compressed AB” and (c) “CbCr” space.
Table 2 : Benchmark of the indexing and matching methods according to the colour space.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>“AB”</td>
<td>0.740</td>
<td>0.743</td>
<td>0.752</td>
<td>0.835</td>
<td>0.801</td>
<td>0.794</td>
</tr>
<tr>
<td>“Comp. AB”</td>
<td>0.782</td>
<td>0.777</td>
<td>0.798</td>
<td>0.843</td>
<td>0.844</td>
<td>0.840</td>
</tr>
<tr>
<td>“CbCr”</td>
<td>0.755</td>
<td>0.755</td>
<td>0.746</td>
<td>0.832</td>
<td>0.797</td>
<td>0.784</td>
</tr>
</tbody>
</table>

First of all, whatever the strategy, indexing in the “Compressed AB” space provides better results. In this space, we have noticed that the chromatic modes are well highlighted and separated from each others. This improves the discrimination between images, whatever the modelling methods. Afterwards, whatever the colour space, the description by a statistical model gives better results than a description by the empirical histogram.

Figure 3 illustrates the proposed strategies by an image retrieval. Let us notice the rank of the non-relevant image (desert) in the three requests, and remark the similitude of the spatial organisation of the first images in third request.

6. CONCLUSIONS

We have focused on a statistical image representation for image retrieval using a biological model of the human perceptive colour space. Two representation levels have been addressed (global and local). The colour space integrates the photoreceptors non-linearities and induces a kind of colour-constancy behaviour. Thus, the chrominance modes of the images are well enhanced allowing an efficient and parsimonious statistical modelling by a Gaussian mixture. With this global representation, retrieval by maximum likelihood overcomes classical techniques based on distances between histograms. Moreover, it allows a local level representation by a spatio-chromatic modelling. Like this, retrieval is more focused knowing the organisation of the chromatic blobs in the images. Furthermore, thanks to this statistical framework, other features can be easily combined.

7. REFERENCES


Acknowledgements

The authors would like to thank Gilles Celeux , Florence Forbes for fruitful discussions and Samuel Blanck for the software implementation. This research was partially funded by the project "ACTIV2" from the Rhône-Alpes region.