Chapter 7

Comparing Images by Content

7.1 Introduction

Following the ideas sketched elsewhere in this thesis, our aim in this Chapter is that of comparing images on the basis of the content obtained from their region segmentation. The necessity for this arises not only from well–established tasks such as object recognition, image indexing and retrieval, but also from problems related to the color constancy of images, since we are interested in objectively measuring the effects of such methods on the segmentation of color images.

Comparing image segmentations can be seen as an object recognition problem. Nevertheless, our situation here is slightly wider and bounds are less defined than these of classical object recognition, since these techniques usually rely on a clean segmentation of objects from the rest of the image or very often are designed for fixed and well–known geometric objects such as machinery parts. Neither of those constraints habitually hold in the case of natural images, where shape, size, and color of objects are quite variable and segmentations are most of the times imperfect.

Our impression is that a more flexible framework is needed in the present case. We need a way to compare segmented images despite the segmentation process is perfect or not. Furthermore, it is important to make clear that some of the tasks that use segmentation, such as finding an object in an environment, matching a given object or scene to a set of previously learnt objects or scenes, come together with the fact that a given image must be searched in a probably quite large database, which should be done as fast and efficiently as possible.

Consequently, our main concern in this Chapter will be to study different ways to define measures of similarity between images based on their color segmentation. These measures must cope with imperfect segmentations since imaging conditions vary more than anything in mobile robotics. Such measures would be extremely helpful both in finding images in a database, identifying objects and comparing image segmentations.

As a mean to attack this problem our interest heeds the set of techniques known as *Content–Based Image Retrieval* (CBIR), since those techniques develop measures among images based on their content to effectively indexing and

searching in large–scale image databases. More precisely, the CBIR approach consists of a set of methods for retrieving semantically–relevant images from an image database on automatically–derived image features.

The reasons we wield for using this framework are mainly the followings. First, these techniques are not essentially focused on finding an exact identification, but rather more flexible retrieval schemes are usually undertaken. Second, they tend to cope with imperfect or inexact segmentations. Inexactness is one of the major flaws in segmentation and can ruin other posterior procedures depending on segmented images. Furthermore, similarity measures employed in those approaches are also a good tool to objectively compare segmentations in the sense that two similar images should generate similar segmentations whose closeness could be numerically grasped by way of these similarity measures.

It is in this last sense that comparing segmentations gains meaning as a way to evaluate the effects of color constancy upon segmentations. The similarity within a set of images from the same scene which has undergone color constancy should be greater than that of the original ones, and so should be their corresponding segmentations. Therefore, before endeavouring the last task mentioned in this Section we first need a tool to compare image segmentations in order to find out whether or not our claim is true.

7.2 Outline of the Chapter

This Chapter is mainly devoted to the comparison of image segmentations and their application to the evaluation of color constancy upon segmentations. To do so, in Section 7.3 we describe a number of related approaches to study some of the possible solutions to this problem. Additionally, in Section 7.4 some usual types of region signatures are rehearsed. Consequently, in Section 7.5 we propose the use of multivariate Gaussian distributions to describe the color of image regions. Section 7.6 is an important one since all the distances between regions that are going to be employed in the next sections are one by one described there. After that, the question of defining a complete distance between two segmentations is endeavored in Section 7.7. Section 7.8 takes into account the experiments undertaken and the results obtained to show the performance of the previously defined list of distances and their application to comparing segmentations after applying a color constancy algorithm. Finally, some of the conclusions and achievements of the Chapter are summarized in Section 7.9.

7.3 Related Work

Some of the best known CBIR techniques are briefly summarized in this Section to illustrate the most relevant measures of similarity between images and their most interesting components before suggesting our own approach. The CBIR issue for image databases of general purpose is still a great and challenging problem because of the large size of such databases, the difficulty in understanding images, the difficulty in formulating a query, and the question of evaluating results properly. The common problems for most CBIR systems are to extract a *signature* for representing every image by its pixel values and to define a rule for comparing images in terms of these signatures.

To our knowledge, the important reason for using signatures, besides the significant compression of the image representation, is that of improving correlation between image pixels and image semantics, i.e., understanding image content only by means of its pixel values. Existing general–purpose CBIR systems roughly fall into three categories depending on the approach applied to extract image signatures, namely, histograms, color layouts, and region–based systems.

After extracting signatures, the step to take into account is the determination of a *comparison rule*, which includes a *querying scheme* and the definition of a *similarity measure* between two images. For most image retrieval systems, a query is basically specified by an image to which others should be matched. This is referred to as *global search* since similarity dwells in the overall properties of images. By contrast, there are also systems with *partial search*, which retrieve images based on a particular region in the image.

7.3.1 Histogram Signature

The best known image database system is IBM's QBIC [FSN+95], which allows an operator to specify various properties of a desired image, such as color distribution, texture or shape. The system then displays a selection of potential matches to those criteria, sorted by a score of the suitability of the match. Region segmentation is largely manual, despite newer versions contain simple automated segmentation facilities. The user is also required to balance different region attributes in order to arrive to a combined similarity measure.

MIT's Photobook [PPS94] incorporates more sophisticated representations of texture and also a higher degree of automatic segmentation. The above system uses histograms to globally represent image color. The major drawback of global histogram representation is that it lacks information about how color is spatially distributed and discards useful information such as object position, texture, and shape, which must be separately considered. Moreover, color histogram search is sensitive to intensity variation, color distortion and cropping[1].

7.3.2 Color Layout Signature

Other similar approaches are VIRAGE System [GJ97], and VisualSEEK and WebSEEK [SC96b] at Columbia. These systems attempt to reduce the influence of intensity variations and color distortions by employing spatial relations and *color layouts* in addition to elementary color, texture, and shape features.

The color layout approach was introduced to overcome the drawback of histogram search. In simple color layout indexing, images are partitioned into blocks and the average color of each block is stored, being thus a low resolution representation of the original image. Stanford's WBIIS [WWFS98] uses significant Daubechies's wavelet coefficients instead of averaging, which characterize color variances over the spatial extent of an image.

By adjusting block sizes or levels of the wavelet transform, coarseness in a color layout representation can be tuned. At a proper resolution, color layout representation naturally retains shape, location, and texture information. Nevertheless, as for the case of pixel representation, despite information is still there,

[1] Phenomenon produced by sensor saturation.

the retrieval system can not directly perceive it. Color layout search is sensitive to shifting, cropping, scaling, and rotation because images are described by a set of local properties [WLW01].

The approach taken by recent WALRUS [NRS04] attempts to reduce shifting and scaling sensitivity for the color layout, as well as that due to translation of objects within an image. In WALRUS, similarity between a pair of images is defined to be the fraction of the area in the two images covered by the matching regions from these images.

In order to extract such regions, this approach considers sliding windows of varying sizes and then clusters them based on the proximity to their signatures. Signatures are based on the first few coefficients of the Haar's wavelet transform. The lack of scale, translation, and orientation invariance of wavelets is a drawback that is handled in WALRUS using the separate windowing approach.

7.3.3 Region–based Signature

As an alternative, region–based retrieval systems attempt to overcome the deficiencies of color layout search by representing images at the level of objects. A region–based retrieval system applies a segmentation algorithm to decompose an image into regions allegedly corresponding to objects, if this decomposition were ideal. Region information can be used not only for locating relevant images, but also for classifying and retrieving purposes. Next, some of the most interesting systems are considered.

NeTra system [MM97a] performs retrieval by requiring the user to select one or more preextracted query regions and attributes, and aggregating the results using a weighted average of the different query features. Such a query system provides more control to the user but less usability, since segmented regions do not necessarily correspond to semantic objects that users may understand. Furthermore, segmentation often splits one object into several regions with none of them being representative for the whole object.

Blobworld approach [CBGM02] transforms raw pixel data into a small set of image regions that are internally coherent in color and texture. This representation is created by clustering pixels in a joint color–texture–position feature space. The segmentation algorithm, which is fully automatic, is based on the *Expectation Maximization* (EM) technique. A simple description of each region's color and texture feature is stored.

Color distribution is represented by a histogram of pixels within the region while texture is stored as the average texture *contrast* and *anisotropy*. In that system, the user composes a query by selecting some of the blobs extracted from the query image and matching them to those in the database, and finally specifying the relative importance for each blob's feature vector. The final query score is calculated using some fuzzy operators to put together the results from each blob.

SIMPLIcity [WLW01] is an image retrieval system which uses semantics classification methods, a wavelet–based approach for feature extraction, and the *Integrated Region Matching* (IRM) similarity measure based upon the segmentation of images. As in other region–based retrieval systems, an image is represented by a set of regions, roughly corresponding to objects, which are characterized by color, texture, shape, and position.

The system also classifies images into two great semantic categories, namely, *textured* and *nontextured* images, which enhances image retrieval by narrowing down the search range in the database. This approach also pays attention to the central problem of developing a similarity measure that combines information from all the extracted regions from the images being compared.

Greenspan et al. propose in [GGR01] an interesting probabilistic image matching scheme where image is continuously represented and the similarity measure and the distance computation are also defined in such continuous domain. Each image is first represented as a *Gaussian Mixture Model* (GMM) and images are compared and matched via a probabilistic measure of similarity between distributions which is based on the Kullback–Leibler divergence.

BlobEMD framework [GDR04] shows a novel method to compute the similarity between two sets of blobs representing image regions. As the previous approach, this one computes the blobs by a GMM and uses the *Earth Mover's Distance* (EMD) [RGT97] to compute the similarity among images.

The EMD similarity measure was developed for comparing histograms and has some common points with the IRM as will be shown lately in this Chapter. This allows similarity measures that are more insensitive to flaws in the segmentation process. Additionally, image segmentation can also be guided by a reference image allowing a context–based adaption of the image representation.

The SNL technique described in [NSL03] is a typical CBIR scheme. Once each image gets segmented, some features, such as color, shape, size, and position, are extracted. Afterwards, the IRM similarity measure [WLW01] is applied to compare image regions. The important point of this work is that the nonmetric nature of IRM is stated, preventing the use of metric access structures and filtering techniques based on the *triangle inequality*.

To overcome this issue, they purpose the alternative SNL* approach using the MICRoM distance [SNF02], which is a true metric, to compare segmented images in conjunction with a filtering step to substantially reduce the number of compared images. Since this method is computationally expensive, the authors propose the variant SNL^{+}, where the metric distance in SNL* is replaced by the original nonmetric IRM, while keeping the filtering step. It is shown that the retrieval speed is improved at the expense of little loss in the retrieval performance.

A quite different approach is the one in [CW02] that defines a *Unified Featured Matching* (UFM) measure which is based on fuzzy logic to retrieve images. In that system an image is represented by a set of segmented regions, where each of them is characterized by a fuzzy feature (fuzzy set) encompassing color, texture, and shape properties.

The similarity between two images is then defined as the overall similarity between two families of fuzzy features and quantified by the UFM measure, which integrates the properties of all the regions. First, a fuzzy similarity measure for two regions is introduced, which is then extended to construct a similarity vector including region–level similarities for all the regions. Finally, the UFM measure maps a similarity vector onto a scalar quantity, which accounts for the over–all image–to–image similarity.

7.4 Types of Region Signatures

Let us now consider the problem of defining region signatures and how to describe an image in terms of its constituent parts. As explained in Section 6.1 of Chapter 6, any image $\mathcal{I}$ can be seen as the union of several nonoverlapping components or regions $\mathcal{R}_i$ such as $\mathcal{I} = \cup_i \mathcal{R}_i$. The process of segmentation translates images into regions of similar features, such as color or texture.

Once those regions have been extracted from the image, they must be modeled someway. This consists in representing each region by a *primitive* encompassing its characteristics in terms of some physical quality such as color, shape, texture, or position. This structure must be flexible enough to describe a particular region while distinguish it from the rest, as well as wide enough to embrace a whole class of similar regions and capable to differentiate them from those of another class. Furthermore, features must be fast and easy to be computed and updated, besides needing little room to be stored.

As previously mentioned, a great number of different features have been proposed so far. Basically, they are all combinations of features such as color, shape, and texture extracted from each region. Some also take the region position into account. Hereafter we summarize some of the most extended region signatures gathered into the three main groups of shape, texture, and color.

- **Shape :** Boundary–based descriptors define the properties of the boundary ($2D$ closed curve) or the exterior of regions. Boundary–based techniques mainly use the outline of a region to compute its shape. Fourier description is one of the well–known methods belonging to this category where dominant Fourier coefficients of the transformed boundary are used as the shape descriptor [CB98].

  Other descriptors of this kind are shape ratios [NSL03] and normalized inertia moments [CW02]. The disadvantage of most of the shape–based retrieval systems is that boundary–based techniques are just applicable to databases with images containing sketches of only single objects.

- **Texture :** It is a powerful region descriptor that may be very helpful in the retrieval process. Texture on its own does not have the capability of finding similar images, but it can be used to preclassify textured images from nontextured ones and then be combined with other visual attributes like color to make the retrieval more effective. A popular representation of texture features is the co-occurrence matrix proposed by Haralick et al. in [HSD73].

  Other classical texture descriptors are the multi–orientation filter bank [MP90] and the second–moment matrix [GL96]. Yet, in the context of content–based retrieval task some additional texture descriptors have been introduced. For example in [CBGM02], polarity, anisotropy, and normalized texture contrast of a region around a pixel are taken into account as texture features.

  In [CW02, WLW01], texture is represented by the square root of the second order moment of wavelet coefficients in the high frequency bands, since moments of wavelet coefficients in various frequency bands have proven effective for discerning texture. Additionally, in WALRUS [NRS04] the

Haar's wavelet transform is computed in a set of sliding windows of varying size within the window capturing both shape and texture efficiently.

- **Color :** This is the most commonly used feature in region–based image retrieval and is also our main concern in this Thesis. In fact, as far as our knowledge arrives, all reviewed approaches try to take some advantage of color one way or the other, i.e., no approach avoids color as a feature. Essentially, the differences among those approaches lay in the color space employed and how color variation within a region is coped with.

  Several options can be chosen for the first problem. For example, *RGB* coordinates or one of its numerous linear transformations. Also, some nonlinear transformations have been proposed, such as HSI, Luv or Lab coordinates, as discussed in Section 6.4.6. A wider overview on color spaces can be found in [WS82, SK94].

  Geometrically speaking, applying a linear transformation onto *RGB* coordinates does not alter the distance between two points in the new space if the new metric is properly taken into account. Rather, nonlinear transformations try to express colors in a more natural way closer to the human perception of color differences. Some of these coordinates have been intentionally defined in a way such that a Euclidean distance is directly used to get similar results to those of a human observer [Fai97].

  The second question is to decide how to consistently express the color of a whole region. The easiest way is by way of a feature vector usually computed as the mean among the pixels in the region. Another popular representation is the color histogram of the region defined as a n–dimensional vector, where each i^{th} component represents the fraction of pixels in the region having the color assigned to the i^{th} bin, and n is the total number of bins in the histogram.

  Alternatively, if the segmentation is doing well and each region only encompasses pixels of similar color, it is clear that these colors will have some variation around the *mean* value. In order to cope with this source of variation, which has to do with surface and light changes, some approaches deal with Gaussian distributions, where a covariance matrix has to be considered besides the mean vector.

This way, any of the regions $\mathcal{R}_i$ an image is split into is described by a signature $\mathbf{\Theta}_i$ consisting in some of the region features previously described. Moreover, the importance of a region may be different depending on facts such as its relative area or position within the image. That is why a *significance value* ω_i must also be append to each region's signature. Consequently, any image $\mathcal{I}$ gets fully described by the set $\{(\mathbf{\Theta}_i, \omega_i),\ i = 1, \ldots, n\}$, where n is the number of components in the image. Our signature will be the *color* of the region that will be modeled in the next Section as a multivariate Gaussian distribution.

By only taking advantage of color, we want to study how far this class of feature can reach without the help of other signatures such as texture and shape. Furthermore, in our opinion, features of the kind of region position may lack of interest if objects are to be located anywhere in an image, whereas shape may present some unsolvable difficulties unless spatial information is provided since its heavy dependence on the point of view.

7.5 Multivariate Gaussian Distributions

Now, the problem consists in representing the color nature of a certain region in a coherent, sound, and useful way. At least three are the usual approaches to solve this problem. First, we could represent a whole region by only one color, by instance, the mean, the median, or any other statistic. In order to account for the color variation within a region $\mathcal{R}_i$, this single feature (vector) could be spanned either as a fuzzy membership function or as a probability distribution function. A third possibility is that of using nonparametric descriptors such as color histograms.

Specifically, we decided to describe color as a *Multivariate Gaussian Distribution* (MGD). This kind of descriptor has many advantages over others, such as simplicity and compactness of the representation, as well as good mathematical properties of the density functions. Besides, it is a natural way of introducing *Gaussian Mixture Models* (GMM) as a natural method for representing segmented images.

Hence, let the color of a certain region be a random variable $X \sim N_d(\xi, \Sigma)$ distributed after the following density function

$$P(\mathbf{x}|\mathbf{\Theta}) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} exp\left(-\frac{1}{2}(\mathbf{x}-\xi)\Sigma^{-1}(\mathbf{x}-\xi)^t\right) \tag{7.1}$$

where $\mathbf{x} \in \mathbb{R}^d$ is a sample point, the signature $\mathbf{\Theta} = \{\xi, \Sigma\}$ consists of the distribution parameters and d is the space dimension, usually 3 in a color coordinates such as RGB, HSI or Lab. Furthermore, ξ and Σ are the mean vector and the covariance matrix of the distribution P, respectively. We note as $\mathcal{X}$ the distribution of the random variable X.

So, each region $\mathcal{R}_i \in \mathcal{I}$ will be described by a distribution $\mathcal{X}_i$ with parameters $\{\overline{\mathbf{x}}_i, \mathbf{\Sigma}_i, \omega_i\}$, where ω_i corresponds to the region significance, and $\overline{\mathbf{x}}_i$ and $\mathbf{\Sigma}_i$ are respectively estimations for ξ_i and Σ_i using the colors of pixels in the i^{th} region.

7.6 Distances between Image Regions

As the natural previous step before any consideration about a distance between segmentations, in this Section we discuss different ways of defining a measure of similarity only between two single regions $\mathcal{R}_i$ and $\mathcal{R}_j$. Once the Gaussian distribution of colors for a region is established as the region descriptor, the idea is to propose a list of distances between such color descriptors. The kind of distances we are looking for must be capable of comparing two distributions expressed in terms of a mean vector and a covariance matrix.

It has to be clear than this is not like the discriminant case where the relative distance $\mathcal{D}(\mathbf{x}, \mathcal{Y})$ of a certain measurement $\mathbf{x}$ to a distribution $\mathcal{Y}$ is computed in order to know whether or not this measurement belongs to that distribution. Here, we need a way to estimate the distance between two Gaussian distributions, i.e., something expressed as $\mathcal{D}(\mathcal{X}, \mathcal{Y})$, corresponding to the random variables $X \sim N_d(\xi_x, \Sigma_x)$ and $Y \sim N_d(\xi_y, \Sigma_y)$.

7.6.1 Euclidean Distance

This is the first and easiest attempt to compute the distance between two Gaussian distributions. It only consists in computing the Euclidean distance between the two mean vectors, $\overline{\mathbf{x}}$ and $\overline{\mathbf{y}}$, that represent the centers of the distributions. So,

$$\mathcal{D}^2(\mathcal{X},\mathcal{Y}) = \|\overline{\mathbf{x}}-\overline{\mathbf{y}}\|^2 = (\overline{\mathbf{x}}-\overline{\mathbf{y}})\,\mathbf{I}_d\,(\overline{\mathbf{x}}-\overline{\mathbf{y}})^t \tag{7.2}$$

where $\mathbf{I}_d$ is the identity matrix of dimension d representing the metric of the distance.

Its simplicity is at the same time a pro and a con, since this distance neither appraises the shape of the two distributions nor its relative size, information conveyed in the covariance matrices. Euclidean distance will be the mark to beat by the next distances.

7.6.2 Mahalanobis Distance

Graphically, Euclidean distances generate a geometrical locus around a center with all points at a fixed distance, which is a circle in a 2D space. This means that any direction around that center has the same worth. Such a thing totally discards shape and orientation of distributions, which could be interesting when comparing them. For example, for two distributions located at the same point, the only appreciable differences would be their shape and relative alignment.

A way to introduce these subtleties into account is by using the Mahalanobis distance. This distance is a very useful way of determining the similarity of a set of measures from a sample of unknown parameters to a set of values from a collection of samples of known parameters. The locus generated in that case is an ellipse – in a 2D space – and it is equivalent to a Euclidean distance where the metric matrix is defined as the inverse of the covariance matrix of the known distribution.

This way, the squared distance between a sample point $\mathbf{y}$ and a distribution $\mathcal{X}$ with parameters $\{\overline{\mathbf{x}},\boldsymbol{\Sigma}_x\}$ is

$$\mathcal{D}^2(\mathbf{y},\mathcal{X}) = (\mathbf{y}-\overline{\mathbf{x}})\,\boldsymbol{\Sigma}_x^{-1}\,(\mathbf{y}-\overline{\mathbf{x}})^t \tag{7.3}$$

Since this kind of distances are bilinear, it is true that the mean distance to a given distribution of a set of samples $\{\mathbf{y}_i\}_{i=1,\dots,n_y}$ equals to the distance between the sample mean $\overline{\mathbf{y}}$ and that distribution, that is

$$\frac{1}{n_y}\sum_{i=1}^{n_y}\mathcal{D}^2(\mathbf{y}_i,\mathcal{X}) = \mathcal{D}^2\left(\frac{1}{n_y}\sum_{i=1}^{n_y}\mathbf{y}_i,\mathcal{X}\right) = \mathcal{D}^2(\overline{\mathbf{y}},\mathcal{X}) \tag{7.4}$$

Obviously, parameters $\{\overline{\mathbf{x}},\boldsymbol{\Sigma}_x\}$ have been estimated from another sample $\{\mathbf{x}_j\}_{j=1,\dots,n_x}$. Therefore, it is also possible to determine the reverse distance between this sample and a distribution $\mathcal{Y}$, namely, $\mathcal{D}^2(\overline{\mathbf{x}},\mathcal{Y})$, whose parameters are esteemed from the sample $\{\mathbf{y}_i\}_{i=1,\dots,n_y}$.

Hence, it seems natural to define the total distance between the two distributions $\mathcal{X}$ and $\mathcal{Y}$ as the mean of these two distances

$$\mathcal{D}^2(\mathcal{X},\mathcal{Y}) = \frac{1}{2}\left(\mathcal{D}^2(\overline{\mathbf{x}},\mathcal{Y}) + \mathcal{D}^2(\overline{\mathbf{y}},\mathcal{X})\right) \tag{7.5}$$

which is equivalent to

$$\mathcal{D}^2(\mathcal{X},\mathcal{Y}) = (\overline{\mathbf{x}} - \overline{\mathbf{y}}) \left[\frac{1}{2} \left(\Sigma_x^{-1} + \Sigma_y^{-1} \right) \right] (\overline{\mathbf{x}} - \overline{\mathbf{y}})^t \tag{7.6}$$

This distance is a Euclidean distance between two mean vectors where the metric matrix happens to be the average of the inverses of the covariance matrices corresponding to the measured samples.

7.6.3 Fréchet Distance

Another way to compute the distance between two multivariate Gaussian distributions is the Fréchet distance [HN99, DL82]. For two random variables X and Y, the Fréchet distance between distributions $\mathcal{X}$ and $\mathcal{Y}$ is defined by

$$\mathcal{D}^2(\mathcal{X},\mathcal{Y}) = \min_{X,Y} E\left\{ \| X - Y \|^2 \right\} \tag{7.7}$$

The Fréchet distance is a special case of the Monge–Kantorovich mass transference problem which is the basis to the *Earth Mover's Distance* (EMD) [GDR04] described later. Dowson and Landau [DL82] solved this problem for the case $\mathcal{X}$ and $\mathcal{Y}$ were elliptically symmetrical, which is the circumstance of Gaussian distributions. In that situation, the distance between $\mathcal{X}$ and $\mathcal{Y}$ can be written as

$$\mathcal{D}^2(\mathcal{X},\mathcal{Y}) = \|\overline{\mathbf{x}} - \overline{\mathbf{y}}\|^2 + tr\left[\mathbf{\Sigma}_x + \mathbf{\Sigma}_y - 2\left(\mathbf{\Sigma}_x \mathbf{\Sigma}_y\right)^{1/2} \right] \tag{7.8}$$

where $\overline{\mathbf{x}}$ and $\overline{\mathbf{y}}$ are the sample mean vectors and $\mathbf{\Sigma}_x$ and $\mathbf{\Sigma}_y$ are the covariance matrices of the distributions $\mathcal{X}$ and $\mathcal{Y}$, respectively.

As can be appreciated in Eq. (7.8), the Fréchet distance is composed of two different terms. the first one, on the right–hand side of Eq. (7.8), defines an Euclidean distance among all mean vectors, while the second term defines a distance on the space of all covariance matrices. Additionally, it is a closed–form solution of EMD in the case of two equal weight distributions and a natural distance for the MGD representation of image regions. Note that whenever two blobs have different weights ω, this distance is not strictly valid, but no alternative is available yet. Fortunately, as the distance employed later is based on the application of a previous region matching stage, the assumption of similar weights can be completely assumed.

7.6.4 Hotelling's T^2 Test

We now introduce a similarity measure between two Gaussian distributions based on the *Hotelling's T^2 test*, which is the multivariate generalization of the *Student's t–test* of significantly different means. In the univariate case, if

$$X_1, \ldots, X_n \;\sim\; N_1(\xi, \sigma^2) \tag{7.9}$$

with ξ and σ unknown, the hypothesis

$$\left.\begin{array}{l} H_0 : \xi = \xi_0 \\ H_1 : \xi \neq \xi_0 \end{array}\right\} \tag{7.10}$$

is tested using the statistics defined as

$$T = \frac{\overline{X} - \xi_0}{\sqrt{\hat{\sigma}^2/n}} \;\sim\; t_{n-1} \,|H_0 \tag{7.11}$$

where $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(\overline{X} - X_i\right)^2$ are mean and variance, respectively. Therefore, at level of significance α, the hypothesis H_0 can be rejected in favor of H_1 if

$$|t| = \sqrt{n}\frac{|\overline{x} - \xi_0|}{s} > t_{n-1,\alpha/2} \tag{7.12}$$

where $\overline{x}$ and s are the sample mean and variance values, respectively.

Equivalently, the squared version of this expression can be written as

$$t^2 = n\,(\overline{x} - \xi_0)\,s^{-1}\,(\overline{x} - \xi_0) > t^2_{n-1,\alpha/2} = F_{1,n-1,1-\alpha} \tag{7.13}$$

A natural multivariate generalization for $X_1, \ldots, X_n \;\sim\; N_d(\xi, \Sigma)$ with dimension $d \geq 1$ is the following Hotelling's T^2 statistic

$$T^2 = n\left(\overline{X} - \xi_0\right)\hat{\Sigma}^{-1}\left(\overline{X} - \xi_0\right)^t \;\sim\; \frac{(n-1)\,d}{n-d}\,F_{d,n-d}\,|H_0 \tag{7.14}$$

where distribution mean and covariance are, respectively, $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $\hat{\Sigma} = \frac{1}{n-1}\sum_{i=1}^{n}\left(\overline{X} - X_i\right)^t\left(\overline{X} - X_i\right)$. Consequently, the two sided Hotelling's T^2 test ($H_1 \colon \xi \neq \xi_0$) rejects H_0 in favor of H_1 at a significant level α if

$$t^2 = n\,(\overline{\mathbf{x}} - \xi_0)\,\boldsymbol{\Sigma}^{-1}\,(\overline{\mathbf{x}} - \xi_0)^t > \frac{(n-1)\,d}{n-d}\,F_{d,n-d,1-\alpha} \tag{7.15}$$

If the problem is geometrically considered, the latter statistic can be seen as the expression of the squared statistical distance between the sample mean $\overline{\mathbf{x}}$ and the distribution mean ξ_0. Note the relation there exists between this statistical distance and the Mahalanobis distance above.

Consider then the case of two independent identically distributed (i.i.d.) samples $\{\mathbf{x}_j\}_{j=1,\ldots,n_x}$ and $\{\mathbf{y}_i\}_{i=1,\ldots,n_y}$, where $\{\overline{\mathbf{x}}, \boldsymbol{\Sigma}_x\}$ and $\{\overline{\mathbf{y}}, \boldsymbol{\Sigma}_y\}$ are their corresponding sample mean vector and covariance matrix estimations, respectively. Hence, the Hotelling distance between two distributions $\mathcal{X}$ and $\mathcal{Y}$ is the following

$$\mathcal{D}^2(\mathcal{X}, \mathcal{Y}) = \left(\frac{n_x\,n_y}{n_x + n_y}\right)(\overline{\mathbf{x}} - \overline{\mathbf{y}})\,\boldsymbol{\Sigma}_{xy}^{-1}\,(\overline{\mathbf{x}} - \overline{\mathbf{y}})^t \;\sim\; \frac{(\nu-1)\,d}{(\nu-d)}\,F_{d,\nu-d} \tag{7.16}$$

where $\nu = n_x + n_y - 1$ and

$$\boldsymbol{\Sigma}_{xy} = \frac{(n_x - 1)\,\boldsymbol{\Sigma}_x + (n_y - 1)\,\boldsymbol{\Sigma}_y}{(n_x + n_y - 2)} \tag{7.17}$$

Comparing with the Mahalanobis distance, $\boldsymbol{\Sigma}_{xy}$ is the mean of matrices $\boldsymbol{\Sigma}_x$ and $\boldsymbol{\Sigma}_y$ weighted by the degrees of freedom for each sample, i.e., n_x and n_y.

7.6.5 α–Trimmed Hotelling's $\tilde{T}^2$ Test

Hotelling's T^2 test for the multivariate problem of examining the significance of the mean vector is known to be optimal under the assumption of multivariate normality. However, when the underlying normality assumption is not completely fulfilled, a more robust procedure is likely to be needed. This may be quite the case in segmentation, where region features, specially color, can not be assured to satisfy the normality assumption [SM01].

Here, we consider trimming multivariate samples by their coordinates. The vector of trimmed means so obtained is used to construct and develop the $\tilde{T}^2$ test as a robust trimmed analogous version of the Hotelling's T^2. An advantage of the coordinate–wise trimming is that whenever an observation is judged to be an *outlier* with regard to some components it is not completely eliminated from the sample and the information carried on the rest of components, which are still normal, is not lost after all. Let us now examine how to compute the trimmed versions of mean vectors and covariance matrices.

α–Trimmed Sample Mean Let $\{\mathbf{x}_i\}_{i=1,\ldots,n}$ be a random sample from a d–variate population with mean ξ and covariance matrix Σ. If only the j^{th} component is taken from those points at a time, the set $\{x_{ij}\}_{i=1,\ldots,n}$ is therefore generated, $\forall j = 1,\ldots,d$. Before computing the trimmed sample mean $\tilde{\mathbf{x}} = (\tilde{x}_1,\ldots,\tilde{x}_d)$, this set must be sorted out obtaining the coordinate–wise ordered set $\{x'_{ij}\}_{i=1,\ldots,n}$, where $x'_{1j} \leq \ldots \leq x'_{nj}$, $\forall j = 1,\ldots,d$.

Then, the j^{th} component of the trimmed sample mean, $\tilde{x}_j$, is computed as the mean value of a subset of elements taken at both sides from the middle element in the component–wise ordered set as follows

$$\tilde{x}_j = \frac{1}{n - 2\lfloor \alpha n \rfloor} \sum_{i=\lfloor \alpha n \rfloor + 1}^{n-\lfloor \alpha n \rfloor} x'_{ij} \ , \ \forall j = 1,\ldots,d \tag{7.18}$$

The number of these elements can be selected by tuning the *trimming parameter* $\alpha \in [0, 1/2]$. If $\alpha = 0$, all the elements in the sorted set $\{x'_{ij}\}$ are gathered and the trimmed mean $\tilde{\mathbf{x}}$ equals the sample mean $\overline{\mathbf{x}}$. If $\alpha = 1/2$, just the value in the middle of the ordering is collected and the trimmed mean $\tilde{\mathbf{x}}$ is then the *median* of the set.

α–Trimmed Covariance Matrix Similarly to the trimmed sample mean, in order to compute the trimmed covariance matrix, the sample points $\{\mathbf{x}_i\}$ must be coordinate–wise Winsorized, which means arranging them as follows

$$x^*_{ij} = \begin{cases} x'_{\lfloor \alpha n \rfloor + 1, j} & \text{if } x_{ij} < x'_{\lfloor \alpha n \rfloor + 1, j} \\ x_{ij} & \text{if } x_{ij} \in [x'_{\lfloor \alpha n \rfloor + 1, j}, x'_{n - \lfloor \alpha n \rfloor, j}] \\ x'_{n - \lfloor \alpha n \rfloor, j} & \text{if } x_{ij} > x'_{n - \lfloor \alpha n \rfloor, j} \end{cases} \tag{7.19}$$

where $i = 1,\ldots,n$ and $j = 1,\ldots,d$. This consists in truncating x_{ij} values outside the intervals limited by the lower bound $x'_{\lfloor \alpha n \rfloor + 1, j}$ and the upper bound $x'_{n - \lfloor \alpha n \rfloor, j}$. These elements are at the position $\lfloor \alpha n \rfloor + 1$ and $n - \lfloor \alpha n \rfloor$ of the ordering $\{x'_{ij}\}$, respectively.

Clearly, the elements of the trimmed covariance matrix $\tilde{\mathbf{\Sigma}}$ of the above Winsorized sample $\{x^*_{ij}\}$ is given by the expression

$$(\tilde{\mathbf{\Sigma}})_{ij} = \frac{1}{h-1}\sum_{k=1}^{n}(x^*_{ki} - \tilde{x}_i)(x^*_{kj} - \tilde{x}_j) \ , \ \forall i,j = 1,\ldots,d \tag{7.20}$$

where h is the number of elements within the interval $[x'_{\lfloor \alpha n \rfloor + 1,j}, x'_{n-\lfloor \alpha n \rfloor,j}]$ and equals $n - 2\lfloor \alpha n \rfloor$.

Thus, the trimmed Hotelling's $\tilde{T}^2$ statistic can be analogously stated as

$$\tilde{T}^2 = h\,(\tilde{\mathbf{x}} - \xi)\,\tilde{\mathbf{\Sigma}}^{-1}\,(\tilde{\mathbf{x}} - \xi)^t \tag{7.21}$$

According to Eq. (7.16), the two–sample version of $\tilde{T}^2$ is used to define the distance between two distributions $\mathcal{X}$ and $\mathcal{Y}$ as

$$\mathcal{D}^2(\mathcal{X},\mathcal{Y}) = \left(\frac{h_x\,h_y}{h_x + h_y}\right)(\tilde{\mathbf{x}} - \tilde{\mathbf{y}})\,\tilde{\mathbf{\Sigma}}^{-1}_{xy}\,(\tilde{\mathbf{x}} - \tilde{\mathbf{y}})^t \tag{7.22}$$

where

$$\tilde{\mathbf{\Sigma}}_{xy} = \frac{(h_x - 1)\,\tilde{\mathbf{\Sigma}}_x + (h_y - 1)\,\tilde{\mathbf{\Sigma}}_y}{(h_x + h_y - 2)} \tag{7.23}$$

7.6.6 Fröbenius Distance

Similarly to the Fréchet distance, the Fröbenius distance [GvL96] computes the total distance between two Gaussian distributions by adding two partial distances, one among mean vectors and another among covariance matrices, being the first one a simple Euclidean distance between the two mean vectors $\overline{\mathbf{x}}$ and $\overline{\mathbf{y}}$, whereas the other is a measure of similarity between matrices based on the norm of the difference matrix calculated from the covariance matrices $\mathbf{\Sigma}_x$ and $\mathbf{\Sigma}_y$ as if matrices were taken as vectors

$$\mathcal{D}^2(\mathcal{X},\mathcal{Y}) = \|\overline{\mathbf{x}} - \overline{\mathbf{y}}\|^2 + \sqrt{\sum_{i,j=1}^{d}\Big((\mathbf{\Sigma}_x)_{ij} - (\mathbf{\Sigma}_y)_{ij}\Big)^2} \tag{7.24}$$

7.6.7 Bhattacharyya Distance

Another way to produce a distance to measure the similarity between two distributions is the application of the Bhattacharyya's affinity kernel originally introduced in [Bha43], which has also been extensively used as a similarity measure in tasks such as object tracking in [CRM03]. Its definition is

$$K(\mathcal{X},\mathcal{Y}) = \int_{\Omega}\sqrt{P_x(\mathbf{v})P_y(\mathbf{v})}\,d\mathbf{v} \tag{7.25}$$

where P_x and P_y are the probability density functions of the random variable X and Y, respectively, and $\mathbf{v} \in \Omega \subset \mathbb{R}^d$. Additionally, Eq. (7.25) is a divergence–type measure that can be interpreted as the (normalized) correlation between the two probability density functions P_x and P_y [CRM03].

If our attention is restricted to the specific case of multivariate Gaussian distributions as in Eq. (7.1), a closed form for the above kernel is suggested in [KJ03] in the following terms

$$K(\mathcal{X},\mathcal{Y}) = \frac{|\mathbf{\Sigma}_z^{-1}|^{1/2}}{|\mathbf{\Sigma}_x|^{1/4}|\mathbf{\Sigma}_y|^{1/4}}\, exp\left(-\frac{1}{2}\mathcal{D}^2\left(\mathcal{X},\mathcal{Y}\right)\right) \tag{7.26}$$

where, by analogy, $\mathcal{D}^2\left(\mathcal{X},\mathcal{Y}\right)$ can be thought as a distance between two Gaussian distributions with parameters $\{\overline{\mathbf{x}},\mathbf{\Sigma}_x\}$ and $\{\overline{\mathbf{y}},\mathbf{\Sigma}_y\}$, that is,

$$\mathcal{D}^2(\mathcal{X},\mathcal{Y}) = \frac{1}{2}\left(\overline{\mathbf{x}}\mathbf{\Sigma}_x^{-1}\overline{\mathbf{x}}^t + \overline{\mathbf{y}}\mathbf{\Sigma}_y^{-1}\overline{\mathbf{y}}^t - 2\,\overline{\mathbf{z}}\mathbf{\Sigma}_z^{-1}\overline{\mathbf{z}}^t\right) \tag{7.27}$$

with the additional definitions of matrix $\mathbf{\Sigma}_z$ and vector $\overline{\mathbf{z}}$ as follows

$$\left.\begin{array}{rcl} \mathbf{\Sigma}_z & = & \frac{1}{2}\left(\mathbf{\Sigma}_x^{-1} + \mathbf{\Sigma}_y^{-1}\right) \\ \overline{\mathbf{z}} & = & \frac{1}{2}\left(\overline{\mathbf{x}}\mathbf{\Sigma}_x^{-1} + \overline{\mathbf{y}}\mathbf{\Sigma}_y^{-1}\right) \end{array}\right\} \tag{7.28}$$

7.6.8 Kullback–Leibler Distance

The Kullback–Leibler (KL) divergence is a measure of the distance between two probability density functions based on information theoretic motivations [Kul68]. KL divergence can be considered as kind of distance between two probability densities, though it is not a proper distance measure because of its lack of symmetry.

Let X and Y be two random variables with densities P_x and P_y, respectively. KL divergence from distributions $\mathcal{X}$ and $\mathcal{Y}$ is defined by

$$KL(\mathcal{X},\mathcal{Y}) = \int_\Omega P_x(\mathbf{v})\, log\frac{P_x(\mathbf{v})}{P_y(\mathbf{v})}\, d\mathbf{v} \tag{7.29}$$

If both P_x and P_y are two Gaussian distributions in a d–dimensional space, then the KL divergence turns into the following expression

$$KL(\mathcal{X},\mathcal{Y}) = tr\left(\mathbf{\Sigma}_y^{-1}\mathbf{\Sigma}_x\right) + log|\mathbf{\Sigma}_y| - log|\mathbf{\Sigma}_x| - d \tag{7.30}$$

Since KL divergence is not symmetrical[2] it must be symmetrized in order to define an appropriate distance. In general, the symmetrical version of the KL divergence, KLS, is obtained as

$$KLS(\mathcal{X},\mathcal{Y}) = \frac{1}{2}\left(KL(\mathcal{X},\mathcal{Y}) + KL(\mathcal{Y},\mathcal{X})\right) \tag{7.31}$$

Consequently, after applying Eq. (7.30) into Eq. (7.31), the KLS distance thus attained is given as

$$KLS(\mathcal{X},\mathcal{Y}) = \frac{1}{2}\left(tr\left(\mathbf{\Sigma}_y^{-1}\mathbf{\Sigma}_x\right) + tr\left(\mathbf{\Sigma}_x^{-1}\mathbf{\Sigma}_y\right) - 2d\right) \tag{7.32}$$

where d is the space dimension and $tr(\circ)$ is the trace of a matrix.

As in the cases of Fröbenius and Fréchet distances, Eq. (7.32) only represents a metric in the space of covariance matrices. So, in order to take the distance between the sample means into account, we define the following complete expression

$$\mathcal{D}^2\left(\mathcal{X},\mathcal{Y}\right) = \|\overline{\mathbf{x}} - \overline{\mathbf{y}}\|^2 + KLS(\mathcal{X},\mathcal{Y}) \tag{7.33}$$

[2] $KL(\mathcal{X},\mathcal{Y}) \neq KL(\mathcal{Y},\mathcal{X})$, in general.

7.6.9 Jensen–Shannon Distance

The above KL divergence has a number of numeric problems. First, if P_x is not *absolutely continuous* with respect to P_y, i.e., the support[3] of P_x is not a subset of the support of P_y, which results in a numerical instability when denominators in Eq. (7.32) are close to zero, i.e., whenever covariances are close to be singular.

A variant distance to overcome such problems is the *Jensen–Shannon distance* (JSD) defined in function of the KL divergence as

$$JSD(\mathcal{X},\mathcal{Y}) = \frac{1}{2}\left(KL(\mathcal{X},\frac{\mathcal{X}+\mathcal{Y}}{2}) + KL(\mathcal{Y},\frac{\mathcal{X}+\mathcal{Y}}{2})\right) \tag{7.34}$$

In the Gaussian case, Eq. (7.34) changes into the following one after applying Eq. (7.30) to it

$$JSD(\mathcal{X},\mathcal{Y}) = \frac{1}{2}\left(tr\left(2\left(\mathbf{\Sigma}_x+\mathbf{\Sigma}_y\right)^{-1}\mathbf{\Sigma}_x\right) + tr\left(2\left(\mathbf{\Sigma}_x+\mathbf{\Sigma}_y\right)^{-1}\mathbf{\Sigma}_y\right)\right) \tag{7.35}$$

Again, Eq. (7.35) is just a distance among covariance matrices, which must be completed as follows to create the whole distance among Gaussian distributions

$$\mathcal{D}^2\left(\mathcal{X},\mathcal{Y}\right) = \|\overline{\mathbf{x}}-\overline{\mathbf{y}}\|^2 + JSD(\mathcal{X},\mathcal{Y}) \tag{7.36}$$

7.7 Similarity Measure Between Images

Intuitively, a region–based similarity measure between two images can be defined as the total amount of the differences between pairs of corresponding regions. It is clear that only the regions which are likely to be the same in both images must be taken into consideration. Otherwise, the measure would be biased by those regions not being in fact the same. But how this matching has to be undertaken without an *explicit* region matching that could be error–prone and time consuming is a problem to be solved yet.

Additionally, a similarity measure based on the region segmentation of images should be tolerant to inaccurate segmentations. To define such a measure, an attempt to match regions must be carried out somehow after all. Being aware that the segmentation process is not perfect, the matching requirements have to be relaxed by allowing any region in the one image to be matched to several regions in the another image.

Mathematically, a similarity measure is also equivalent to a distance between sets of points in a feature space. In such representation, every point in the space corresponds to the feature vector or descriptor of a region. Although the distance between two points in a feature space can be chosen from a variety, it is not obvious how to define a distance between two groups or distributions of points. Moreover, this class of distance should be sufficiently consistent with the human concept of semantic *closeness* between two images.

In this Section we consider two similarity measures that tries to cope with the problems aforementioned, namely, *Earth Mover's Distance* (EMD) and IRM similarity. Both distances are computed as a combination of distances among regions in two image segmentations. The point that differentiates them is about

[3]The support of a function f is the set $\{\mathbf{x}\in\mathbb{R}^d | f(\mathbf{x})\neq 0\}$.

how these values are combined and the particular method that has been followed to compute the set of weights pondering the combination. *Grosso modo*, IRM is a close and useful simplification of EMD and will be the one used throughout the rest of the Chapter as a mean to evaluate distances between segmentations.

7.7.1 Earth Mover's Distance

EMD naturally extends the notion of distance between single elements to distance between sets of elements or distributions. Intuitively, given two distributions represented by sets of weighted features, one can be thought as a mass of earth properly spread in the feature space and the other as a collection of holes in the same space. EMD measures the least amount of work needed to fill the holes with earth. A unit of work corresponds to transporting a unit of earth by a unit of *ground distance*, which is a distance in the feature space.

Computing EMD is based on a solution to the transportation problem. This is a bipartite network flow problem which can be formalized as the following linear programming problem. Let $\mathcal{I} = \{(\mathbf{\Theta}_i, w_i)\}_{i=1,\ldots,n}$ be the first set of n regions, where $\mathbf{\Theta}_i$ is the descriptor representing the distribution $\mathcal{X}_i$ and w_i is the weight of the i^{th} region $\mathcal{R}_i$. Analogous definitions can be applied to the second set of n' regions $\mathcal{I}' = \{(\mathbf{\Theta}'_j, w'_j)\}_{j=1,\ldots,n'}$. $\mathcal{I}$ and $\mathcal{I}'$ represent two segmentations.

We can define $d_{ij} = \mathcal{D}(\mathcal{X}_i, \mathcal{X}'_j)$ as the ground distance between the color distributions of regions $\mathcal{R}_i \in \mathcal{I}$ and $\mathcal{R}'_j \in \mathcal{I}'$. Then, we need to find a set of optimal admissible flows f_{ij} from $\mathcal{I}$ to $\mathcal{I}'$ that minimize the overall *cost*, that is, $\sum_{i=1}^{n}\sum_{j=1}^{n'} f_{ij}d_{ij}$, being subjected to the following constraints

$$\begin{cases} f_{ij} \geq 0 & \forall i = 1,\ldots,n,\ \forall j = 1,\ldots,n' \\ \sum_{i=1}^{n} f_{ij} \leq w'_j & \forall j = 1,\ldots,n' \\ \sum_{j=1}^{n'} f_{ij} \leq w_i & \forall i = 1,\ldots,n \end{cases} \tag{7.37}$$

The first constraint allows flows from $\mathcal{I}$ to $\mathcal{I}'$ and not *vice versa*. The second constraint forces the total flow arriving in any region in the set $\mathcal{I}'$ not to exceed its capacity. The third one implies that the total flow going out from any region in the set $\mathcal{I}$ can not exceed its associated weight. Finally, the feasibility condition states that the total flow must not exceed the total supply, i.e.,

$$\sum_{i=1}^{n}\sum_{j=1}^{n'} f_{ij} = min\left\{\sum_{i=1}^{n} w_i, \sum_{j=1}^{n'} w'_j\right\} \tag{7.38}$$

Once the transportation problem is solved and the optimal flow f_{ij} is found, the earth mover's distance is defined as

$$EMD(\mathcal{X}, \mathcal{Y}) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n'} f_{ij}d_{ij}}{\sum_{i=1}^{n}\sum_{j=1}^{n'} f_{ij}} \tag{7.39}$$

The advantages of EMD are diverse. First, EMD applies to signatures, such as histograms or distributions. Second, the costs of moving earth reflect the notion of nearness properly without the quantization problems of most current measures. Third, EMD allows partial matches in a natural way. Fourth, if the ground distance is a metric and the total weights of two signatures are equal,

EMD is also a true metric. The next similarity measure, rather than being a rough simplification, is inspired from the EMD distance and follows the same spirit.

7.7.2 IRM Similarity Measure

The *Integrated Region Matching* (IRM) measure quantifies the overall similarity between two images by integrating the properties of all the regions in those images. The key advantage of this overall similarity is its robustness against poor segmentations. Hence, a region–to–region match is obtained whenever the components are significantly similar to each other in terms of their extracted features. The principle of matching in IRM is that the most similar region pair is matched first. Once all regions get matched one another, the complete distance estimation is computed as a weighted sum of the similarity between any pair of regions whose weights have been determined in the previous matching phase.

A matching between two regions $\mathcal{R}_i \in \mathcal{I}$ and $\mathcal{R}'_j \in \mathcal{I}'$ is assigned with a significance level $s_{ij} \geq 0$, which indicates the importance of that particular matching. Analogously to the EMD distance, the matching between two images can also be represented by a bipartite weighted graph where every vertex corresponds to a component one of these images. If two vertices are connected, the two regions are linked by an edge whose weight is the corresponding significance level. Otherwise, the significance level is zero or both elements belong to the same image.

The problem of defining a distance between sets of features is that of finding the proper significance levels so that a good similarity measure is obtained as an aftermath. Consequently, some properties on the significance levels must be stated beforehand. The first property to be satisfied is the fulfillment of significance, which ensures that all the regions participate in the similarity measure. This is sketched as follows

$$\begin{cases} s_{ij} \geq 0 & \forall i = 1, \ldots, n, \; \forall j = 1, \ldots, n' \\ \sum_{i=1}^{n} s_{ij} = w'_j & \forall j = 1, \ldots, n' \\ \sum_{j=1}^{n'} s_{ij} = w_i & \forall i = 1, \ldots, n \end{cases} \tag{7.40}$$

For normalization reasons, $\sum_{i=1}^{n} \sum_{j=1}^{n'} s_{ij} = \sum_{i=1}^{n} w_i = \sum_{j=1}^{n'} w'_j = 1$. Furthermore, it is also required that the most similar regions get the highest priority. The IRM algorithm attempts to fulfill those properties by assigning as much significance as possible to the region link with the least distance. The precise algorithm to do so will not be reproduced here and the reader can find it in [WLW01].

Finally, the IRM distance between two images is equal to the summation of all the weighted edge lengths in the bipartite graph, i.e.,

$$IRM(\mathcal{X}, \mathcal{Y}) = \sum_{i=1}^{n} \sum_{j=1}^{n'} s_{ij} \, d_{ij} \tag{7.41}$$

7.7.3 Region Weights Selection

In addition to the problem sketched above of defining a distance between segmented images, we can not forget the issue of choosing how the weights are to

be assigned to each region $\{w_i\}_{i=1,\ldots,n}$. A value for w_i is chosen to reflect the significance of the i^{th} component in the image. If every region were equally important – *uniform scheme* – then $w_i = 1/n$, where n is the total number of regions in $\mathcal{I}$.

Another selection strategy that can be followed instead is the one called *area percentage scheme*. In this case, weight w_i is the percentage of the whole image only covered by region $\mathcal{R}_i$ and its interest is based on the fact that important objects in an scene tend to occupy larger areas in the image.

The latter method is less sensitive to inaccurate segmentations than the uniform scheme. If an object is partitioned into several regions, the uniform scheme improperly raises their weights, whereas the percentage scheme balance their importance. On the other hand, if objects are merged into one region, the percentage scheme assigns a relatively higher weight to the whole region.

7.8 Experiments and Results

After suggesting a quite numerous list of measures objectively comparing images by means of the differences between their segmentations, now we describe the two sets of experiments carried out to test the preferences of such measures as well as to comment their corresponding results. First of all, we want to show how the previous group of distances performs in tasks such as image retrieval or object identification. For this purpose, we take advantage once more of the two segmentation algorithms introduced in the foregoing Chapter, the one proposed for us in this Thesis and Figueiredo's EM [FJ02].

In the second part of this Section we suggest using some of these distances to judge the performance of the color constancy algorithm introduced in Chapter 5. The idea is that a reduction in the color variation of images can also be measured as a consequent reduction in the distances between the corresponding segmented images. Hence, any identification system based on the appearance of objects will tend to do better whenever a color constancy stage is undergone, as it is claimed in [FBM98].

7.8.1 Performance in Retrieval and Identification

Image Database Description

First of all, we start at describing the image database used in the present set of experiments. These images belong to the celebrated *Columbia Object Image Library (COIL)*[4] database which consists of color images of 100 objects, which were placed on a motorized turntable against a black background. The turntable was rotated through 360 degrees to vary object pose with respect to a fixed color camera and images were taken at intervals of 5 degrees. This corresponds to 72 poses per object and a total amount of $7,200$ color images. More information about the database acquisition can be found in [NNM96].

In our experiments we used $N_s = 18$ images per object – corresponding to intervals of 20 degrees since 5 degrees provided too similar images – as well as just $N_o = 51$ of those objects. This way, we built up a database of $N_d = N_s \cdot N_o = 918$ color images, which is definitively a sufficient amount of images

[4] `http://www1.cs.columbia.edu/CAVE/research/softlib/coil-100.html`

Figure 7.1: The COIL database.

to test the performance of the previous distances. In Fig. 7.1 we can appreciate the front view of these objects.

It can be argued about the suitability of such a database to endeavour a testing set of experiments involving image retrieval. Usually, those experiments employ images similar to those in the Internet in a number several times bigger than we do. Nevertheless, the reader should not forget that our scope is mobile robotics, where it is not unlikely to find objects such as those pictured in the COIL database, after removing the background, though we are not going into further details on that issue.

As said, two different segmentation algorithms were employed in order to draw some conclusions out of the results about the influence that each type of technique has on the performance of those distances. First, in Fig. 7.2 we show the segmentation results of the objects' front views by applying the Figueiredo's EM algorithm [FJ02]. Similarly, in Fig. 7.3 analogous results of segmentations produced by our graph–partitioning algorithm [VLCS00] are pictured.

Furthermore, some more examples of objects are provided in Fig. 7.4 in order to present how objects change in shape and appearance through the diversity of poses they exhibit if the viewing angle is modified. In those Figures three sets of images are provided for each object. Pictures in the upper row show the actual object, the row in the middle corresponds to the Figueiredo's segmentation and the lower row to those segmented with our algorithm.

Experiment Details

The goal of these experiments is to establish the efficacy for each distance presented above in the tasks of image retrieval and object identification as a way to benchmark them for a posterior use in sight. To that purpose, we carry out two kinds of tests, namely, a retrieval experiment and a matching experiment.

Image retrieval is accomplished by emulating the response of a system to a certain image query in the following way. The response of the system consists of a set of images sorted by their increasing distances to a given image used as (global) query. Thus, for each type of distance from the lot defined before and for each image in the database, we generate one single response by computing the distance between such image to all other images in the database and, after sorting them by their resulting distances, we select the set of N_r images with the least distance to the query, where $N_r < N_s$. These images represent the set obtained from the retrieval system as a result to a global query. N_r is a variable value spanning from 1 to the number of images per category N_s.

Matching experiment consists in evaluating every distance as a way to perform object identification. If any of these distances were selective enough to decide whether or not a query corresponds to a given object, it would be very useful in such kind of tasks. Therefore, in order to estimate its suitability, the database is divided into two disjoint subsets containing 9 images per object, i.e., 459 images per set in total. One set is called the *testing* set, while the other is the *sample* set. Then, the test computes the total amount of correct identifications carried out by taking images from the testing set and finding the closest image in the sample set. A correct matching occurs whenever the closest image recovered from the sample set belongs to the same object as the image from the testing set.

Figure 7.2: Figueiredo's EM segmentation.

Figure 7.3: Graph–partitioning segmentation.

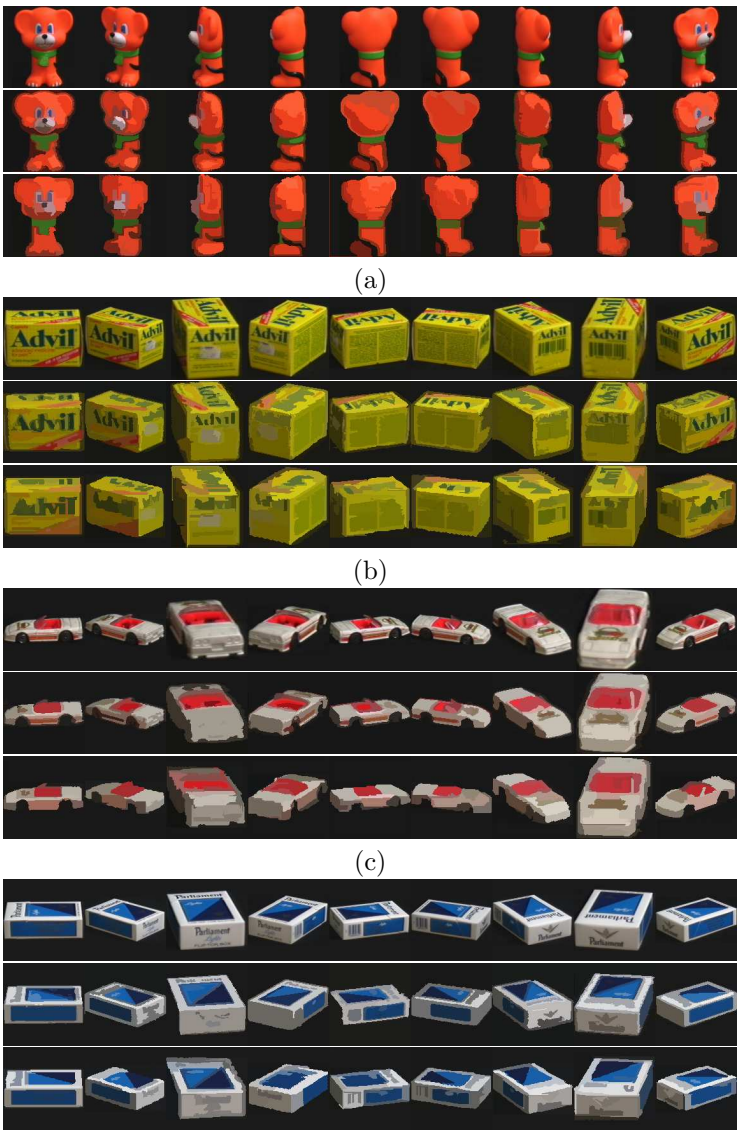


Figure 7.4: Segmentations of some objects – (a) Bear, (b) Advil, (c) Car and (d) Box – viewed through different poses. Upper Row: Original object images. Middle Row: Figueiredo's segmentation images. Lower Row: Images using our segmentation.

Measures of Performance

A widespread manner to objectively evaluate the performance of a retrieval system is to compute the percentage of relevant images retrieved. Ideally, a system should retrieve all the relevant images first, keeping as small as possible the number of nonrelevant images retrieved before those which are relevant. A retrieved image is considered relevant *if and only if* it is in the same category as that of the query. In our case, a category corresponds to the set of images belonging to an given object. Therefore, an image will be *relevant* if it belongs to the set of images of the queried object, while a *nonrelevant* image is that not belonging to them.

At least, two general measures can be defined to that goal, those of *recall* and *precision* [NSL03, CW02]. Recall is the percentage of the total number of relevant images retrieved and is defined as

$$\text{Recall} = \frac{\text{N}^o \text{ of Relevant Images Retrieved}}{\text{Total N}^o \text{ of Relevant Images}} \tag{7.42}$$

where the total number of relevant images stands for the number of images per category, i.e., per set of images. Precision refers to the capability of the system to retrieve only relevant images

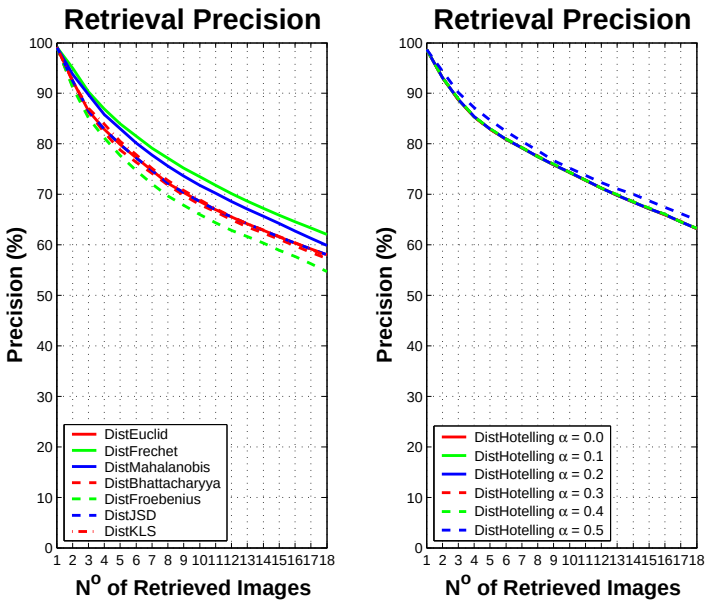
$$\text{Precision} = \frac{\text{N}^o \text{ of Relevant Images Retrieved}}{\text{Total N}^o \text{ of Images Retrieved}} \tag{7.43}$$

These two measures were calculated per category (object) and for an increasing number of retrieved images ($N_r = 1 \div N_s$). Additionally, a global measure was generated extending the average computation over the whole set of objects. Recall and precision measures per category are built up as the average of the results after taking as a query each image in that category. We did not consider any method speeding up the computation of a single query, rather than the symmetrical property of distances, which saved us as much as a half of the image–to–image distance computations.

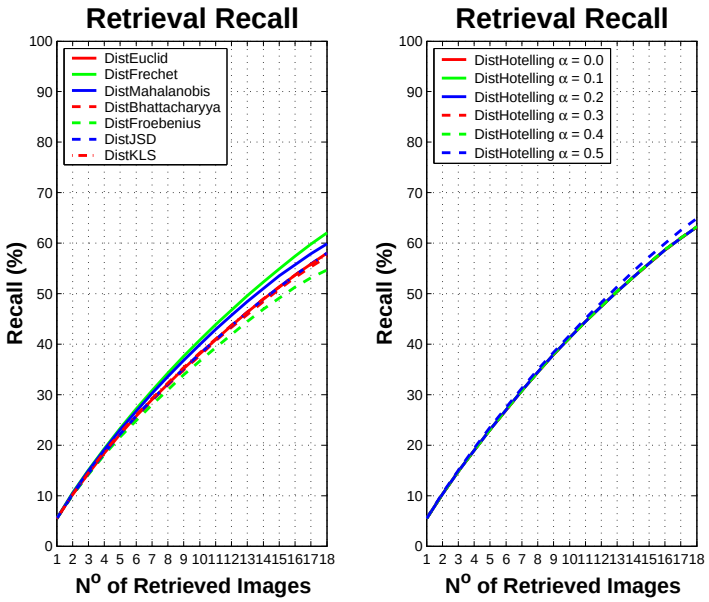
Precision and recall results as a function of the size of the retrieved set, N_r, are depicted in Fig. 7.5 to Fig. 7.9. Due to the great amount of distances to plot for each experiment, we decided to split these results into two graphics on behalf of clarity. For each graphic, the leftmost plot encompasses the following distances: Euclidean, Fréchet, Mahalanobis, Bhattacharyya, Fröbenius, Jensen–Shannon (JSD), and symmetrical KL distance (KLS). The rightmost plot endorses the results for the Hotelling distance with the trimming parameter $\alpha = 0.0, 0.1, 0.2, 0.3, 0.4$ and 0.5.

Specifically, Fig. 7.5, Fig. 7.6, and Fig. 7.8 correspond to results obtained employing images which were segmented using the Figueiredo's EM algorithm, while Fig. 7.7 and Fig. 7.9, to those segmented by our graph–partitioning algorithm. In the Figueiredo's segmentations we produced two kinds of results, depending on the color space employed in the computation of distances, namely, RGB or Lab. Fig. 7.5 was generated in RGB coordinates, while Fig. 7.6 in Lab. This choice does not affect the image segmentation process, which was separately done in RGB coordinates.

Since the results employing our segmentation algorithm were generated posteriorly to those by Figueiredo's and we realized that distances in Lab coordinates significantly improved the results attained for both precision and recall,

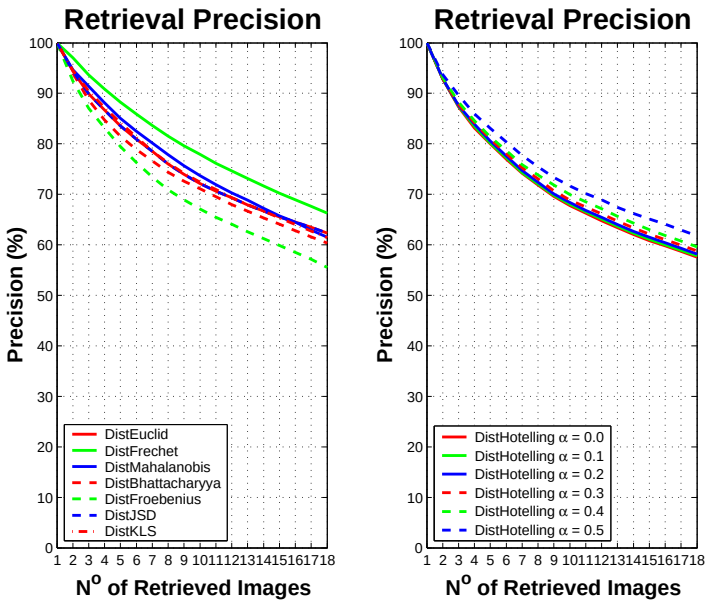


(a)

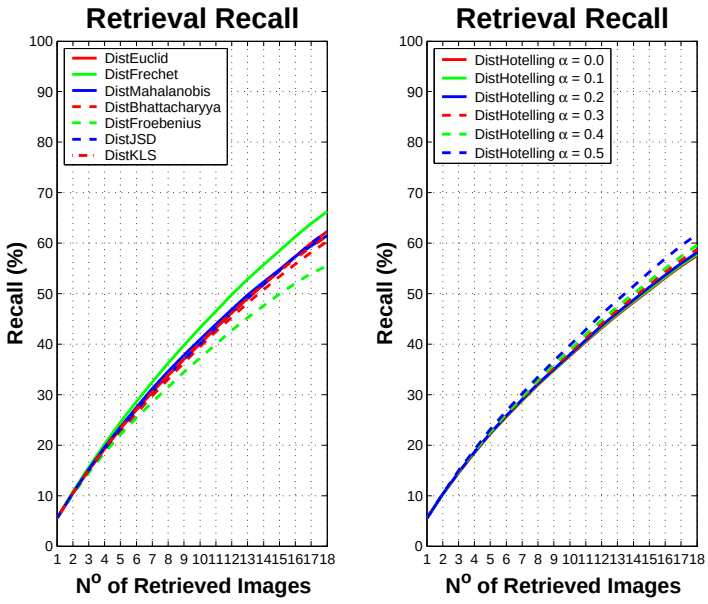


(b)

Figure 7.5: Results using Figueiredo's EM segmentation and RGB color coordinates: (a) retrieval precision plot and (b) retrieval recall plot.

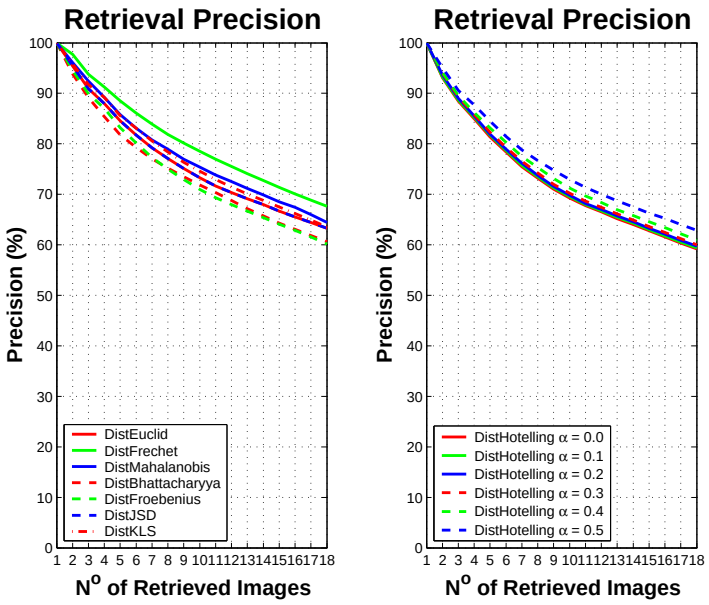


(a)

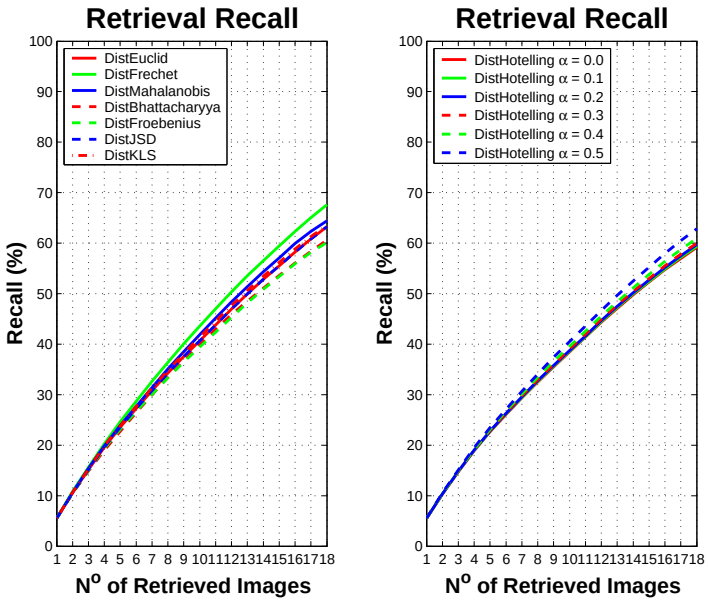


(b)

Figure 7.6: Results using Figueiredo's EM segmentation and Lab color coordinates: (a) retrieval precision plot and (b) retrieval recall plot.

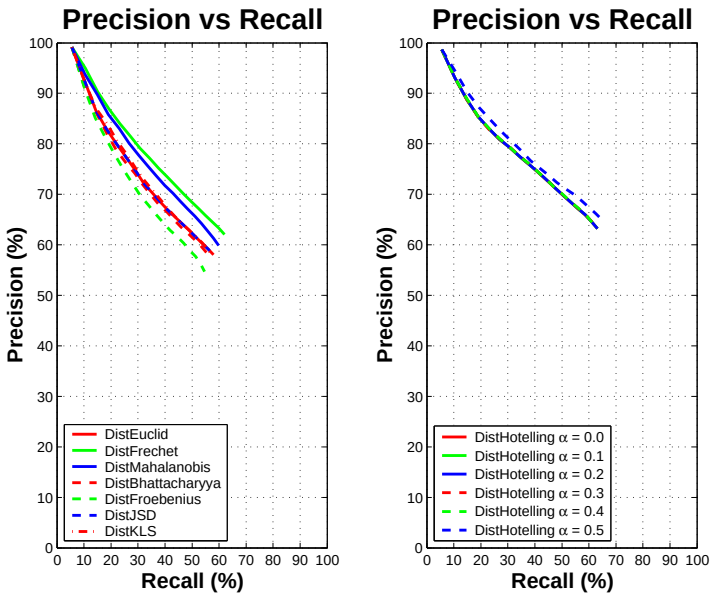


(a)

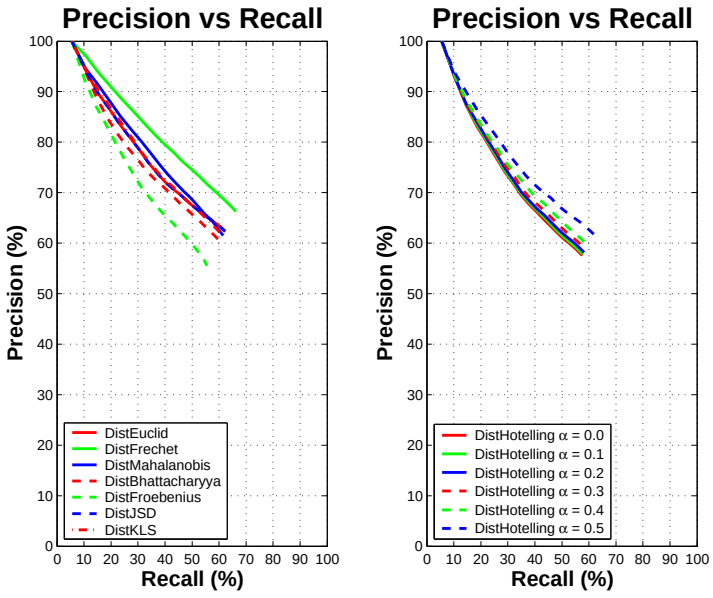


(b)

Figure 7.7: Results using graph–partitioning segmentation and Lab color coordinates: (a) retrieval precision plot and (b) retrieval recall plot.



(a)



(b)

Figure 7.8: Figueiredo's RGB and Lab precision vs. recall plots: (a) RGB and (b) Lab.

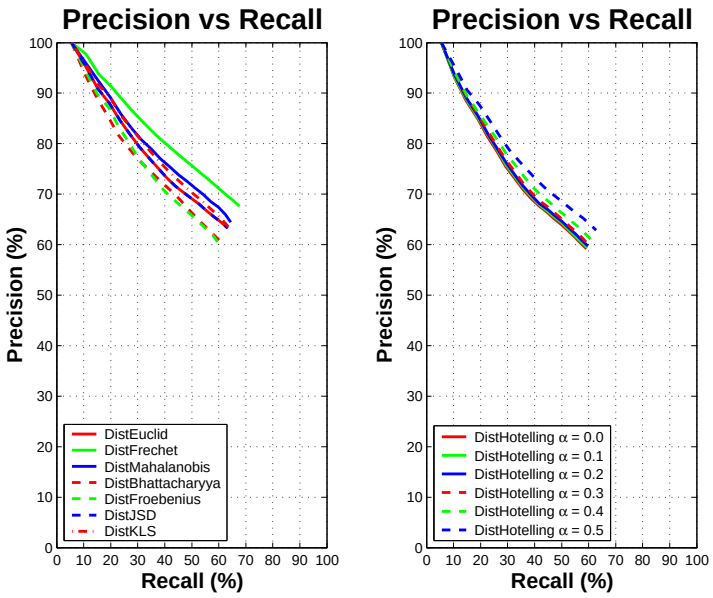


Figure 7.9: Precision vs. Recall plot corresponding to the graph–partitioning segmentation.

we decided to carry out the tests corresponding to the graph–partitioning segmentation only in Lab coordinates. Results obtained this way are depicted in Fig. 7.7 as explained before.

Another way of comprehensively showing results about precision and recall is by means of a *Precision vs. Recall* plot. We did this in Fig. 7.8, where results by Figueiredo's in both RGB and Lab coordinates are depicted, and also in Fig. 7.9, where those by our segmentation in Lab coordinates are analogously presented. This kind of plot shows how precision decays when the fraction of relevant images in a query response is pushed up. The slower the fall, the better the efficiency of the retrieval system.

It must be stated that in the worst case, namely, that of extracting as many images as the number of images in a set, $N_r = 18$ in our case, both precision and recall coincide in value. This is the situation exhibited in Table 7.1, which quantitatively shows the *best* and the *worst* values from the Precision vs. Recall plots as well as the particular distances that generated them.

Cases	**Figueiredo** [RGB]	**Figueiredo** [Lab]	**Partition** [Lab]
Best	64.91 (Hot. $\alpha = 0.5$)	66.23 (Fréchet)	67.60 (Fréchet)
Worst	54.70 (Fröbenius)	55.48 (Fröbenius)	59.16 (Hot. $\alpha = 0.0$)

Table 7.1: Best and worst results for the Precision vs. Recall plots per method.

We now change our attention to the issue of object recognition, whose results have been portrayed in Table 7.2. Values in this Table show the identification rates for each kind of distance and each segmentation algorithm. As before, Figueiredo's EM segmentation was carried out in both RGB and Lab spaces, while just in Lab coordinates for the graph–partitioning algorithm.

Distance	**Figueiredo** [RGB]	**Figueiredo** [Lab]	**Partition** [Lab]
Euclidian	85.19	89.11	**91.29** (3)
Fréchet	**91.29** (1)	**94.34** (1)	**95.21** (1)
Mahalanobis	86.71	**89.54** (2)	**91.50** (2)
Bhattacharyya	82.35	86.71	88.02
Fröbenius	82.79	84.53	89.32
Jensen–Shannon	85.19	89.11	**91.29** (3)
Symmetrical KL	84.97	**89.32** (3)	91.07
Hotelling ($\alpha = 0.0$)	**87.36** (3)	84.75	85.62
Hotelling ($\alpha = 0.1$)	**87.36** (3)	85.40	86.27
Hotelling ($\alpha = 0.2$)	**87.36** (3)	85.62	87.58
Hotelling ($\alpha = 0.3$)	**87.36** (3)	85.84	87.80
Hotelling ($\alpha = 0.4$)	**87.36** (3)	85.84	88.45
Hotelling ($\alpha = 0.5$)	**90.20** (2)	89.11	89.98

Table 7.2: Object recognition results per method.

Furthermore, on behalf of clarity, the three best results for each column have been marked in boldface with a number in parentheses, which stands for its order with regard to the degree of recognition. This way, it can be seen at once the distances giving better results for each kind of approach.

Result Discussion

First of all, let us consider precision and recall results as a function of the distance employed and the type of segmentation applied to images. By looking at the precision plots, we can realize that the results obtained in Lab coordinates are fairly better than those in RGB coordinates. This trend can also be observed in both Fig. 7.5(a) and Fig. 7.6(a), where the best curves in the former tend to decrease faster than those in the latter. The only difference in such behavior is found in the case of the Hotelling distance, where it is the other way round, i.e., results in RGB are better than those in Lab.

Furthermore, despite Lab plots have a wider range of variation, the ordering that appears after sorting the list of similarity measures by their performance is the same for both RGB and Lab coordinates. In Fig. 7.5(a) and Fig. 7.6(a) the best distance is that of Fréchet. The second one is Mahalanobis, which is in RGB close to that of Fréchet, but in Lab has a slightly worse performance, closer to the third group of distances, namely, Euclidean, KLS, and JSD. Bhattacharyya comes right after those, and the worst of all is Fröbenius, both in RGB and Lab.

An additional factor to take into account is that Hotelling's results depend on the trimming parameter $\alpha \in [0, 1/2]$. At least, theoretically, the idea behind this parameter is that it can be adjusted to get better results than strictly those

of the mean ($\alpha = 0$) or the median ($\alpha = 1/2$) somewhere in between those two values. However, results show that the precision curve increases with α. So, the best performance is achieved by the median and the worst by the mean. Unfortunately, in our experiments it seems that no better outcome is in the middle of those values as expected.

Nevertheless, the Hotelling distance attains a better performance than the Fréchet distance in the Figueiredo RGB segmentation, Fig. 7.5(a), and just as good as Mahalanobis in Lab space, Fig. 7.6(a). Analogous results can be appreciated examining Fig. 7.5(b) and Fig. 7.6(b), which correspond to the recall measure.

Regarding the results attained by our segmentation algorithm, displayed in Fig. 7.7(a), we find that the performance curves are sorted exactly the same way for both precision and recall plots, as they did in the former case of Figueiredo's segmentation. That is, first, Fréchet; then, Mahalanobis, Euclidean, KLS, Jensen–Shannon, Bhattacharyya, and Fröbenius at the end, but now Bhattacharyya performs as poorly as Fröbenius. On the other hand, the Hotelling distance presents a slightly worst performance, similar to that of Fröbenius in the worst case ($\alpha = 0.0$), and to that of Mahalanobis in the best case ($\alpha = 0.5$), which is almost the same result obtained in Fig. 7.6(a).

From these values and after having considered the plots in Fig. 7.5 to Fig. 7.7, some conclusions can be drawn about the performance of the distances between image segmentations presented so far. First, it is apparently clear that Lab coordinates work slightly better than RGB. Second, segmentations obtained with the graph–partitioning algorithm achieve higher values than those of Figueiredo's EM, probably because images got less oversegmented in the former case. Finally, it seems that the Fréchet distance outperforms the others, with the sole exception of the Hotelling distance with $\alpha = 0.5$ (median).

The worst distances ostensively are Fröbenius and Hotelling, with a trimming parameter close to $\alpha = 0.0$ (mean). A fact about the trimming parameter that comes out from the results is that in our experiment it was unable to find better outputs between the mean and the median, probably because the number of samples was relatively small. Yet, it provides a good mechanism to blend sample points even if only the median ($\alpha = 1/2$) is employed, since the selection is done component–wise, independently from the number of dimensions.

If now we have a look to the *Precision vs. Recall* plots in Fig. 7.8 and Fig. 7.9, we can perceive the same behaviour as before in attention to the set of distances, being even clearer which distances are better (or worse) than others. In particular, the best curve is the one in Fig. 7.9 attained by the Fréchet distance and Lab coordinates. These trend can also be appreciated in Table 7.1.

In regard to the recognition rates for each distance and segmentation scheme, almost the previous behaviour is again reflected also in Table 7.2. In the first column, i.e., Figueiredo's RGB segmentation, the best result is achieved by Fréchet and Hotelling ($\alpha = 0.5$), while Fröbenius and Bhattacharyya are now the worst. In the second column, which stands for the Figueiredo's segmentation in Lab, the best distances are Fréchet and Mahalanobis, whereas the worst are Fröbenius and Hotelling ($\alpha = 0.0$). Finally, in the third case, our segmentation algorithm in Lab, the best ones are again Fréchet and Mahalanobis, while Hotelling ($\alpha = 0.0$) keeps being the worst scored. The best overall result was obtained using the graph–partitioning segmentation in Lab coordinates along with the Fréchet distance, and was as high as 95.21% of correct identifications.

7.8.2 Effects of Color Constancy on Color Segmentation

This Section is where all the work done so far must help us to analyze the evolution of the differences between segmentations of images that have followed a color constancy (CC) stage if we compare them with the same lot which have not. Thus, the principal aim here is to objectively state two facts.

First, color constancy reduces the dissimilarity within the set of images belonging to the same scene taken under changing light conditions, as it is the kind found in mobile robotics. Secondly, we should also consider the effects of CC on the distance between images from different categories and whether a CC process would help to distinguish better these scenes one another.

In order to attain such goals, our experiments consist in computing the global differences between all the images coming from a given scene and also the distance between any pair of different sets. These values conform what we call the *intraclass* and the *interclass* mean distances, respectively. In the next lines our work will be the computation of such measures to show whether the CC phase is really working or not, and to what extent.

Image Database Description

Four sets of images were employed, each corresponding to a scene in a human–made environment, e.g., an office. All the images in each set pictured nearly the same view of the scene and the illumination was changed as much as it could be feasible in a way to generate rather variable sets. No care was taken to maintain the exact pose for these images and changes in light were obtained by turning the lights on and off as well as opening and shutting the window shutters. These illumination changes correspond to what a mobile robot would face in a usual walk around at different hours of the day.

To entirely show these sets we split them into two separate groups shown in Fig. 7.10 and Fig. 7.11. Three of these sets are in Fig. 7.10. It can be appreciated that neither the view of the scenes nor the light conditions are exactly the same in all the images. Unlike those sets, having 3 images each, the fourth one in Fig. 7.11 has 9 images and a special care was taken to greatly change their illumination even in a greater degree than in the previous sets.

More precisely, the latter set of images is shown in the first row of Fig. 7.11, where the effect of the light change in the global color of the scene is self–evident. Besides these images, Fig. 7.11 encompasses further interesting results. In the first column, the same images are again presented, but rather this time as the canonical images that have been employed by the rest of the set in the CC stage. The resulting images after the CC algorithm described in Chapter 5 are row–wise displayed. Each row corresponds to the set if the image in the first column is taken as the canonic one. Therefore, it can be perceived that the global color for each row closely resembles to the canonic one.

Despite the great number of images portrayed in Fig. 7.11, we only took just one row for our experiments –particularly, the last one – corresponding to one single canonic illumination. Since images in all the sets belong to the same office, we also employed the same canonic image for the 3 other sets. It is also feasible to use a different canonic image for each scene. *In summa*, the total number of images used in our experience was $N_d = 36$, split in two groups of $N_s = 18$, consisting of original images and color corrected images, respectively.

Figure 7.10: Sets with light variation (I): Sets 1, 2, and 3.

Figure 7.11: Sets with light variation (II): Set 4.

Experiment Details

In the next lines we describe the set of experiments carried out to study the effect of a CC stage in the segmentation of several sets of color images that have been captured from an autonomous robot. We want to know whether color constancy also reduces the distance between image segmentations and how this reduction comparatively affects the images that belong to the same scene or to those coming from different scenes.

In order to attain such a goal, we first compute a distance measure between pairs of original images, which have not followed any CC process. Afterwards, the same computations are applied to the same images after undergoing, this time, a CC stage. These calculation are simply done by way of applying a similarity measure between segmentations that come from the two groups of $N_s = 18$ images thus generated, images that belong to the scenes in Fig. 7.10 and Fig. 7.11.

The resulting distance existing between any two of the previous sets is calculated as the mean distance through all pairs of images in these sets. Hence, we must distinguish two kinds of such distances depending on the set where images originate, namely, the *intraclass* and the *interclass* distances. Intraclass distances account for the average distance among images within the same category, whereas in the case of interclass distances, images involved belong to different sets.

For computing similarities among segmentations we use two kinds of distances in the experiment, namely, the Euclidean distance defined in Section 7.6.1 and the Fréchet distance in Section 7.6.3. The first one is used for its simplicity, while the second has been stated in the previous Section to be the best in terms of performance in tasks of retrieval and recognition. Color descriptors are assumed to be in Lab coordinates for both types of distances.

Results corresponding to computations on the original images, where no CC has been applied yet, are summarized in Table 7.3 and 7.5 for the Euclidean and the Fréchet distances, respectively. Analogous results using the two types of distances but now corresponding to images that have suffered the CC process are placed in Tables 7.4 and 7.6, respectively.

In all these tables, boldfaced figures in the diagonal represent the intraclass distances within each image set, while those out of the diagonal are the interclass distance between pairs of image sets. Additionally, the average[5] through all the interclass distances for each set is given in the last row as a tool to know the mean distance from a certain set to the rest of sets considered. In principle, intraclass distances should be, at least, smaller than interclass distances if we intend to distinguish scenes one another.

Furthermore, in Table 7.7 to Table 7.10 we summarize the results already shown in Table 7.3 to Table 7.6 that specifically correspond to intraclass and interclass distances. These values are displayed forming two rows in each Table, whose columns belong to the sets of images considered in Fig. 7.10 and Fig. 7.11. In addition, there exists an extra column, labeled as *Row Mean*, which furnishes the mean values row–wise computed, i.e., the average through all the sets. That mean is also weighted by the number of elements per set.

The last row in those Tables indicates the *percent difference* between intraclass and interclass distances in respect to the intraclass distance, and is

[5] An average *weighted* by the number of images in the sets.

	Set 1	**Set 2**	**Set 3**	**Set 4**
Set 1	**7.462**	11.620	12.690	14.135
Set 2	11.620	**10.741**	13.811	16.253
Set 3	12.690	13.811	**8.804**	15.615
Set 4	14.135	16.253	15.615	**12.780**
InterC.	**13.343**	**14.838**	**14.669**	**15.334**

Table 7.3: Intraclass (diagonal) and interclass (last row) Euclidean distances without color constancy.

	Set 1	**Set 2**	**Set 3**	**Set 4**
Set 1	**6.101**	8.873	9.839	8.167
Set 2	8.873	**6.486**	7.212	8.924
Set 3	9.839	7.212	**3.106**	6.750
Set 4	8.167	8.924	6.750	**4.180**
InterC.	**8.643**	**8.571**	**7.460**	**7.947**

Table 7.4: Intraclass (diagonal) and interclass (last row) Euclidean distances with color constancy.

	Set 1	**Set 2**	**Set 3**	**Set 4**
Set 1	**10.629**	15.203	18.881	17.578
Set 2	15.203	**13.480**	17.342	19.764
Set 3	18.881	17.342	**11.468**	18.172
Set 4	17.578	19.764	18.172	**13.664**
InterC.	**17.364**	**17.367**	**18.148**	**18.505**

Table 7.5: Intraclass (diagonal) and interclass (last row) Fréchet distances without color constancy.

	Set 1	**Set 2**	**Set 3**	**Set 4**
Set 1	**9.844**	13.627	15.931	13.643
Set 2	13.627	**11.200**	14.061	14.791
Set 3	15.931	14.061	**8.631**	11.984
Set 4	13.643	14.791	11.984	**6.903**
InterC.	**14.097**	**14.412**	**13.189**	**13.473**

Table 7.6: Intraclass (diagonal) and interclass (last row) Fréchet distances with color constancy.

calculated as

$$\%Diff = \frac{IntreC - IntraC}{IntraC} \times 100 \tag{7.44}$$

showing how different the interclass distance is from the intraclass distance for a given image set. This index has been separately computed for each type of segmentation distance with values obtained before (Table 7.7 and Table 7.9) and after (Table 7.8 and Table 7.10) the color constancy step.

After applying color constancy to those images, it would be logic to get more homogeneous sets, whereas keeping their relative differences more or less constant. This should be reflected in a generalized increase of the $\%Diff$ index, i.e., color constancy should reduce the intraclass distance more than the interclass distance. That happening would prove that color constancy makes segmentations from the same set look more similar than those from different sets, which is helpful in identification tasks.

On the other hand, both the intraclass distance and the interclass distance can get independently affected by a color constancy step. To numerically grasp such effects we portray in Table 7.11 to Table 7.14 the results corresponding to the *percent reduction* of such distances due to color constancy, i.e.,

$$\%Red = \frac{Dist_{NoCC} - Dist_{CC}}{Dist_{NoCC}} \times 100 \tag{7.45}$$

where $Dist \in \{IntraC., InterC.\}$, and $NoCC$ and CC stand for distance values, respectively, *before* and *after* color constancy. These values come from Table7.3 to Table7.6 and have been rewritten in Table 7.11 to Table 7.14 for clarity. Such kind of measure is computed for the Euclidean and the Fréchet cases, and row–wise means are computed again weighting by the number of images per set.

In brief, $\%Diff$ helps to understand how color constancy makes image segmentations belonging to the same category look more similar than those from different sets. On the other hand, $\%Red$ illustrates the magnitude of the dissimilarity reduction among image segmentations that has been favoured by color constancy. In order to understand better those results, we have rehearsed them in Table 7.15 to Table 7.18, showing this time only the mean values for each image set of the two previous indices as well as these of the row–wise means.

Result Discussion

The aim in the past Section was to demonstrate that a CC step should improve the similarity among images and, by extension, among segmentations. We wanted to show how, by means of a content–based distance, this fact could be numerically stated. Moreover, it would be interesting to know if the CC could also improve the performance of such distances as a mean to recognize scenes from one single view taken under uncontrolled illumination conditions, despite this is not our interest here. Therefore, two indexes were defined in order to study the behavior of the differences between image segmentations as an aftermath of color constancy.

Let us first consider the evolution of the relative difference between the distances among images within a given set and among those belonging to distinct sets, i.e., the difference between the intraclass and the interclass distances. As said, these results can be succinctly viewed in Table 7.15 and Table 7.16. From such results we can claim that the mean $\%Diff$ value through all the sets presents

Distance	Set 1	Set 2	Set 3	Set 4	Row Mean
IntraC.	7.46	10.74	8.80	12.78	**10.89**
InterC.	13.34	14.84	14.67	15.33	**14.81**
% Diff.	**78.81 %**	**38.14 %**	**66.62 %**	**19.98 %**	**40.59 %**

Table 7.7: Percent difference between intraclass and interclass distances (Euclidean case without CC).

Distance	Set 1	Set 2	Set 3	Set 4	Row Mean
IntraC.	6.10	6.49	3.11	4.18	**4.71**
InterC.	8.64	8.57	7.46	7.95	**8.09**
% Diff.	**41.66 %**	**32.15 %**	**140.19 %**	**90.12 %**	**80.73 %**

Table 7.8: Percent difference between intraclass and interclass distances (Euclidean case with CC).

Distance	Set 1	Set 2	Set 3	Set 4	Row Mean
IntraC.	10.63	13.48	11.47	13.66	**12.76**
InterC.	17.36	17.37	18.15	18.51	**18.23**
% Diff.	**63.36 %**	**36.26 %**	**58.25 %**	**35.43 %**	**44.03 %**

Table 7.9: Percent difference between intraclass and interclass distances (Fréchet case without CC).

Distance	Set 1	Set 2	Set 3	Set 4	Row Mean
IntraC.	9.84	11.20	8.63	6.90	**8.40**
InterC.	14.10	14.41	13.19	13.47	**13.69**
% Diff.	**43.21 %**	**28.68 %**	**52.81 %**	**95.18 %**	**68.37 %**

Table 7.10: Percent difference between intraclass and interclass distances (Fréchet case with CC).

IntraC.	**Set 1**	**Set 2**	**Set 3**	**Set 4**	**Row Mean**
NoCC	7.46	10.74	8.80	12.78	**10.89**
CC	6.10	6.49	3.11	4.18	**4.71**
% Red	**18.24 %**	**39.61 %**	**64.72 %**	**67.29 %**	**54.08 %**

Table 7.11: Percent reduction due to CC in intraclass distance (Euclidean case).

InterC.	**Set 1**	**Set 2**	**Set 3**	**Set 4**	**Row Mean**
No CC	13.34	14.84	14.67	15.33	**14.81**
CC	8.64	8.57	7.46	7.95	**8.09**
% Red	**35.23 %**	**42.23 %**	**49.14 %**	**48.17 %**	**45.19 %**

Table 7.12: Percent reduction due to CC in interclass distance (Euclidean case).

IntraC.	**Set 1**	**Set 2**	**Set 3**	**Set 4**	**Row Mean**
No CC	10.63	13.48	11.47	13.66	**12.76**
CC	9.84	11.20	8.63	6.90	**8.40**
IntraC.	**7.39 %**	**16.91 %**	**24.74 %**	**49.48 %**	**32.91 %**

Table 7.13: Percent reduction due to CC in intraclass distance (Fréchet case).

InterC.	**Set 1**	**Set 2**	**Set 3**	**Set 4**	**Row Mean**
No CC	17.36	17.37	18.15	18.51	**18.23**
CC	14.10	14.41	13.19	13.47	**13.69**
% Red.	**18.81 %**	**21.53 %**	**27.33 %**	**27.19 %**	**24.87 %**

Table 7.14: Percent reduction due to CC in interclass distance (Fréchet case).

a leap from 40.59% to 80.73% in the Euclidean case, and also an analogous increase from 44.03% to 68.37% in the Fréchet case. Both results are rather significant and should help to differentiate better scenes one another.

Obviously, since not all the sets encompass the same color variation, which is clearly greater in the 4^{th} set, the biggest one, the amount of the increase in $\%Diff$ is not the same for all the sets either, and even for some there have been a slight decrease. In the Euclidean case, the greatest increase appears in the 4^{th} set, followed by that of the 3^{rd} set. For the 2^{nd} set, $\%Diff$ is almost the same before and after color constancy, while it has clearly decreased for the 1^{st} set. Results in the Fréchet case are very similarly to those just mentioned, unless for the 3^{rd} set, which is slightly inferior after the color constancy step. Nevertheless, the total $\%Diff$ variation is undoubtedly positive no matter the segmentation distance taken into account.

We now consider the effect of color constancy in the intraclass and the interclass distances independently taken in Table 7.17 and Table 7.18. If the mean values are considered for all the sets of images, it is pretty clear that there exists a reduction of those distances due to color constancy. In the Euclidean case, the decrease in the intraclass distance is 54.08% while that of the interclass distance is 45.19%. In the Fréchet case, they are 32.91% and 24.87%, respectively.

We can appreciate from these results two distinct facts. First, there has been a considerable general reduction in the dissimilarity between image segmentations as a consequence of the color constancy process. Secondly, this dissimilarity reduction is greater in the intraclass distances than in the interclass distances, which straightforwardly relates this improvement with the increase in the percent difference between these two types of distances previously described.

Again, some differences appear among the sets in terms of $\%Red$ and the same pattern is followed as for the case of $\%Diff$. Thus, sets having greater color variations also present greater dissimilarity reductions for both the Euclidean and the Fréchet distances. Above all them, there is the 4^{th} set. This set is rather significant since it is the one spanning the widest color variation and encompassing the largest number of images. On the contrary, in sets with smaller color variations (1^{rt} and 2^{nd}), $\%Red$ is pretty much the same, or even superior, for both the intraclass and the interclass distances.

In short, from the results exposed so far, we claim that a color constancy stage appreciably reduces the dissimilarity among color image segmentations, as already seen in Chapter 5 for their appearance. Thus, differences amid images in a particular set are reduced and, in a lesser degree, between images not belonging to the same sets. Moreover, the greater the color variation in a set, the more significant the reduction. Such results are basic in order to discern better from different scenes, thus achieving more selective segmentation similarity measures due to color constancy.

7.9 Conclusions

This Chapter illustrates the problem of comparing images by means of the content obtained from their color segmentations. A substantial group of measures of similarity is proposed within the frame of the IRM distance and the employ of multivariate Gaussian distributions for the color description of image regions. The performance of each and every one of these distances is examined

% Diff.	Set 1	Set 2	Set 3	Set 4	Row Mean
No CC	78.81 %	38.14 %	66.62 %	19.98 %	**40.59 %**
After CC	41.66 %	32.15 %	140.19 %	90.12 %	**80.73 %**

Table 7.15: Percent difference between intraclass and interclass distances before and after CC (Euclidean case).

% Diff.	Set 1	Set 2	Set 3	Set 4	Row Mean
No CC	63.36 %	36.26 %	58.25 %	35.43 %	**44.03 %**
After CC	43.21 %	28.68 %	52.81 %	95.18 %	**68.37 %**

Table 7.16: Percent difference between intraclass and interclass distances before and after CC (Fréchet case).

% Red.	Set 1	Set 2	Set 3	Set 4	Row Mean
IntraC.	18.24 %	39.61 %	64.72 %	67.29 %	**54.08 %**
InterC.	35.23 %	42.23 %	49.14 %	48.17 %	**45.19 %**

Table 7.17: Percent reduction due to CC in intraclass and interclass distances (Euclidean case).

% Red.	Set 1	Set 2	Set 3	Set 4	Row Mean
IntraC.	7.39 %	16.91 %	24.74 %	49.48 %	**32.91 %**
InterC.	18.81 %	21.53 %	27.33 %	27.19 %	**24.87 %**

Table 7.18: Percent reduction due to CC in intraclass and interclass distances (Fréchet case).

in tasks such as images retrieval and object recognition. The best overall results are obtained by using the graph–partitioning algorithm described in Chapter 6 along with Lab color space and Fréchet distance, being as high as 95.21% of correct object identifications, which outperforms the other approaches assayed in previous Sections.

Two of these distances (Fréchet and Euclidean) are selected afterwards in order to study the effects of the color constancy algorithm in Chapter 5 in the segmentation of images from an autonomous robot, focusing on how this process reduces the mean distances defined amid images within a category compared to those involving images in distinct sets in two different situations, namely, before and after carrying out the color constancy step. Thus, we want to answer the question whether or not the reduction in color dissimilarity may help these sets of images to be more easily distinguished each other by way of their segmentations.

Results exhibit that, if colors in those sets are originally variable enough, the reduction in the mean distance within a category is greater than that between different categories, which also means that the relative distance between image sets has neatly grown therefore. Otherwise, the reduction is more balanced, since there is no big differences to reduce. This way, the applied distance measures based on segmentations are more selective at distinguishing images from dissimilar scenes as a consequence of the color constancy step performed on them.

In brief, this Chapter shows that color constancy reduces the amount of color variation within sets of images picturing the same scene by producing more similar segmentations. As claimed in Chapter 3, this is necessary if a identification process based on the segmentation of color images must be independent from the conditions of illumination. Moreover, as a by–product of the aforementioned goal, a significant number of content–based similarity measures are also studied in tasks that involve either object recognition or image retrieval. Those measures are thus more selective at distinguishing images from unlike scenes due to color constancy.