

Evaluation of Distances Between Color Image Segmentations

Jaume Vergés-Llahí and Alberto Sanfeliu

Institut de Robòtica i Informàtica Industrial
Technological Park of Barcelona, U Building
{jverges,asanfeliu}@iri.upc.es

Abstract. We illustrate the problem of comparing images by means of their color segmentations. A group of seven distances are proposed within the frame of the Integrated Region Matching distance and the employ of Multivariate Gaussian Distributions (MGD) for the color description of image regions. The performance of these distances is examined in tasks such as image retrieval and object recognition using the two segmentation algorithms in [1] and [2]. The best overall results are obtained for both tasks using the graph-partition approach along with the Fréchet distance, outperforming other distances in comparing MGDs.

Keywords: color segmentation, image retrieval, object identification.

1 Introduction

The aim of this paper consists in comparing images on the base of their color segmentation. The necessity for this arises from well-known tasks such as object recognition, image indexing and retrieval, as well as others related to mobile robotics. Content-based comparison of image segmentations can be viewed as an object recognition problem. Nevertheless, our situation is slightly opener than that of classical object recognition since segmentation of real images is most of the time imperfect. We need a more flexible way to compare segmented images on the basis of their content whether segmentation is perfect or not.

Consequently, our concern in this paper is to study different distances between images based on their color segmentation which can cope with imperfect segmentations. Such a measure would be helpful both in finding images in a database, identifying objects and comparing image by their content. Our aims heed the set of techniques known as *Content-Based Image Retrieval* (CBIR) since those methods develop measures among images based on their content to effectively indexing and searching in large-scale image databases. More precisely, CBIR is the set of techniques for retrieving semantically-relevant images from an image database on automatically-derived image features.

The reasons for using this framework are mainly two. First, these techniques are not focused on finding an exact identification, rather more flexible retrieval schemes are usually undergone. Second, they tend to cope with imperfect or inexact segmentations. Therefore, this paper proposes some distances between regions in the context of a CBIR distance between segmentations to evaluate their performance in tasks such as object identification and image retrieval.

2 Related Work

Because of the lack of space, we only refer to some of the best-known CBIR techniques. CBIR for general purpose databases is still a great challenging problem because of the size of such databases, the difficulties in both understanding images and formulating a query, and the properly evaluation of results. A common problem for all CBIR systems is to extract a *signature* from images based on their pixel values and to define an effective rule for comparing images.

The signature, whose components are called *features*, serves as the image representation. The important reason for using signatures, besides the significant compression of the image representation, is that of improving the correlation between image pixels and image semantics, i.e., understanding the image content by means of its pixel values. Most existing general-purpose CBIR systems roughly fall into the next three categories, depending on the image signatures

- *Histograms*: IBM QBIC [3], MIT Photobook [4].
- *Color layouts*: VIRAGE [5], Columbia VisualSEEK and WebSEEK [6], Stanford WBIIS [7], and WALRUS [8].
- *Region-based systems*: NeTra [9], Blobworld [10], SIMPLicity [11], SNL [12].

After extracting signatures, the next step is to determine a *comparison rule*, which includes a *querying scheme* – global or partial search – and the definition of a *similarity measure* between two images. Despite numerous types of signatures can be employed, such as shape and texture, our concern here is just those based on the color information of image pixels.

3 Distance Between Segmented Images

Intuitively, a region-based distance between two segmented images can be defined as the total amount of differences between corresponding regions. It is clear that only regions which are likely the same in both segmentations must be taken into account. Otherwise, the measure would be biased. In addition, any distance based on image segmentation should be tolerant to inaccuracies. To define this kind of measures, an attempt to match regions must be carried out somehow, despite it could be error prone and time consuming. Nevertheless, matching has to be softened by allowing one region of an image to be matched to several regions in another one since segmentations are not perfect.

Furthermore, a similarity measure among regions is equivalent to a distance between sets of points in a feature space. Every point in the space corresponds to a 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. In this paper, we first describe the region descriptor applied throughout the work to account for color distributions. Afterwards, a number of different approaches to measure the similarity between these descriptors are proposed. Finally, we describe how to group all these measures into only one distance between two segmented images.

3.1 Multivariate Gaussian Distributions

We describe region colors as a *Multivariate Gaussian Distribution* (MGD) of probability for its simplicity and compactness, along with its good mathematical properties. Besides, it is a natural way of introducing *Gaussian Mixture Models* (GMM) as a method of representing segmented images [2]. Hence, let the color of a region be a random variable $X \sim N_d(\mu, \Sigma)$ distributed as

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

where $\mathbf{x} \in \mathbb{R}^d$ is a sample point, $\Theta = \{\mu, \Sigma\}$ are the mean and covariance matrix of P , respectively, and d is the space dimension¹. We note as \mathcal{X} the distribution of the random variable X . So, for each region \mathcal{R}_i in image \mathcal{I} there will be a distribution \mathcal{X}_i whose parameters are $\{\bar{\mathbf{x}}_i, \Sigma_i, \omega_i\}$, being ω_i the region weight, and $\bar{\mathbf{x}}_i$ and Σ_i are the sample mean and covariance matrix, respectively.

3.2 Distances Between Image Regions

Now, we discuss seven diverse distances between two regions \mathcal{R}_x and \mathcal{R}_y , expressed in terms of their distributions \mathcal{X} and \mathcal{Y} . Nevertheless, this is not the discriminant case where a distance $\mathcal{D}(\mathbf{x}, \mathcal{Y})$ between a point \mathbf{x} and a distribution \mathcal{Y} is computed to know whether \mathbf{x} belongs to \mathcal{Y} . Rather, we need to estimate the distance between two whole MGDs, i.e., $\mathcal{D}(\mathcal{X}, \mathcal{Y})$, corresponding to random variables $X \sim N_d(\mu_x, \Sigma_x)$ and $Y \sim N_d(\mu_y, \Sigma_y)$.

Euclidean Distance. It consists only in computing the Euclidean distance between the two means, $\bar{\mathbf{x}}$ and $\bar{\mathbf{y}}$, each of them representing a distribution center

$$\mathcal{D}^2(\mathcal{X}, \mathcal{Y}) = \|\bar{\mathbf{x}} - \bar{\mathbf{y}}\|^2 \quad (2)$$

Its simplicity is both a pro and a con, since it neither appraises the shape of the distributions nor its relative size, conveyed in the covariance matrices.

Mahalanobis Distance. Shape and orientation could be interesting when trying to compare distributions, for example, whenever two distributions are centered at the same point. Mahalanobis distance introduces these subtleties into account by using as a metric the inverse of the covariance matrix of the particular distribution. This way, the squared distance between a sample point \mathbf{y} and a distribution \mathcal{X} is computed as $\mathcal{D}^2(\mathbf{y}, \mathcal{X}) = (\mathbf{y} - \bar{\mathbf{x}}) \Sigma_x^{-1}(\mathbf{y} - \bar{\mathbf{x}})^t$.

Since these distances are bilinear, it is true that the mean distance to a distribution of a set of points $\{\mathbf{y}_i\}_{i=1, \dots, n_y}$ equals to the distance between the sample mean $\bar{\mathbf{y}}$ and that distribution, that is, $\frac{1}{n_y} \sum_{i=1}^{n_y} \mathcal{D}^2(\mathbf{y}_i, \mathcal{X}) = \mathcal{D}^2(\bar{\mathbf{y}}, \mathcal{X})$. Parameters $\{\bar{\mathbf{x}}, \Sigma_x\}$ belonging to \mathcal{X} were estimated using the sample $\{\mathbf{x}_j\}_{j=1, \dots, n_x}$.

¹ Usually 3 in a color coordinates such as *RGB*, *HSI* or *Lab*.

Therefore, the reverse distance $\mathcal{D}^2(\bar{\mathbf{x}}, \mathcal{Y})$ between $\{\mathbf{x}_i\}$ points and the distribution \mathcal{Y} , whose parameters are computed using the sample $\{\mathbf{y}_i\}$, can also be taken into account. Hence, it seems natural to define the total distance between the two distributions \mathcal{X} and \mathcal{Y} as the mean of the two previous distances, that is, $\mathcal{D}^2(\mathcal{X}, \mathcal{Y}) = \frac{1}{2}(\mathcal{D}^2(\bar{\mathbf{x}}, \mathcal{Y}) + \mathcal{D}^2(\bar{\mathbf{y}}, \mathcal{X}))$, which is equivalent to the expression

$$\mathcal{D}^2(\mathcal{X}, \mathcal{Y}) = (\bar{\mathbf{x}} - \bar{\mathbf{y}}) \left[\frac{1}{2} (\Sigma_x^{-1} + \Sigma_y^{-1}) \right] (\bar{\mathbf{x}} - \bar{\mathbf{y}})^t \quad (3)$$

Fréchet Distance. Another way to compute a distance between two MGDs is the Fréchet distance [13]. Fréchet distance between two random variables X and Y is defined by $\min_{X, Y} E\{\|X - Y\|^2\}$. This is a special case of the Monge–Kantorovich mass transference problem. Dowson and Landau [13] solved this problem for the case of \mathcal{X} and \mathcal{Y} being *elliptically symmetric*, which is the condition of the MGD. Hence, the distance between \mathcal{X} and \mathcal{Y} can be written as

$$\mathcal{D}^2(\mathcal{X}, \mathcal{Y}) = \|\bar{\mathbf{x}} - \bar{\mathbf{y}}\|^2 + tr \left[\Sigma_x + \Sigma_y - 2(\Sigma_x \Sigma_y)^{1/2} \right] \quad (4)$$

Fréchet distance is composed of two terms, namely, an Euclidean distance among means and a distance on the space of covariance matrices. Additionally, it is a closed-form solution to the Earth’s Mover Distance (EMD) in the situation of two equally weighted Gaussian² and a natural distance for the Gaussian region representation.

Fröbenius Distance. Alike Fréchet distance, this one computes the distance between two MGD by addition of two partial distances, one among means (Euclidean) and another among covariances, which is defined between matrices based on the norm of the difference matrix computed from the covariances Σ_x and Σ_y , calculated as if they were vectors (componentwise), i.e.,

$$\mathcal{D}^2(\mathcal{X}, \mathcal{Y}) = \|\bar{\mathbf{x}} - \bar{\mathbf{y}}\|^2 + \|\Sigma_x - \Sigma_y\|^2 \quad (5)$$

Bhattacharyya Distance. Bhattacharyya’s affinity kernel [14] was extensively used as a similarity measure in tasks such as object tracking [15]. It is defined as $K(\mathcal{X}, \mathcal{Y}) = \int_{\Omega} \sqrt{P_x(\mathbf{v})P_y(\mathbf{v})} d\mathbf{v}$, where P_x and P_y are PDFs of the random variables X and Y , respectively, and $\mathbf{v} \in \Omega \subset \mathbb{R}^d$. This is a divergence-type measure interpretable as a (normalized) correlation between PDFs [15].

A closed form in the case of MGDs for the above kernel is suggested in [16] as $K(\mathcal{X}, \mathcal{Y}) = k \cdot \exp(-\frac{1}{2}\mathcal{D}^2(\mathcal{X}, \mathcal{Y}))$, where $k = |\Sigma_z|^{1/2}/(|\Sigma_x|^{1/4}|\Sigma_y|^{1/4})$ and, by analogy, $\mathcal{D}^2(\mathcal{X}, \mathcal{Y})$ is

$$\mathcal{D}^2(\mathcal{X}, \mathcal{Y}) = \frac{1}{2} (\bar{\mathbf{x}} \Sigma_x^{-1} \bar{\mathbf{x}}^t + \bar{\mathbf{y}} \Sigma_y^{-1} \bar{\mathbf{y}}^t - 2\bar{\mathbf{z}} \Sigma_z^{-1} \bar{\mathbf{z}}^t) \quad (6)$$

with the additional definitions of matrix $\Sigma_z = [\frac{1}{2}(\Sigma_x^{-1} + \Sigma_y^{-1})]$ and vector $\bar{\mathbf{z}} = \frac{1}{2}(\Sigma_x^{-1}\bar{\mathbf{x}} + \Sigma_y^{-1}\bar{\mathbf{y}})$.

² This assumption can be totally assumed whenever segmented images are employed.

Kullback–Leibler Distance. The Kullback–Leibler (KL) divergence is a measure of the alikeness between two PDFs based on information theoretic motivations [17], defined as $KL(\mathcal{X}, \mathcal{Y}) = \int_{\Omega} P_x(\mathbf{v}) \log(P_x(\mathbf{v})/P_y(\mathbf{v})) d\mathbf{v}$. If both P_x and P_y are MGDs, it turns into $KL(\mathcal{X}, \mathcal{Y}) = \text{tr}(\Sigma_y^{-1}\Sigma_x) + \log|\Sigma_y| - \log|\Sigma_x| - d$.

Since KL divergence is not symmetric in general, it must be symmetrized before defining a proper distance as follows $KLS(\mathcal{X}, \mathcal{Y}) = \frac{1}{2}(KL(\mathcal{X}, \mathcal{Y}) + KL(\mathcal{Y}, \mathcal{X}))$. Consequently, the KLS distance thus obtained is given as

$$KLS(\mathcal{X}, \mathcal{Y}) = \frac{1}{2} (\text{tr}(\Sigma_y^{-1}\Sigma_x) + \text{tr}(\Sigma_x^{-1}\Sigma_y) - 2d) \quad (7)$$

As for the cases of Fröbenius and Fréchet, Eq. (7) only represents a metric in the covariance space, so the distance is $\mathcal{D}^2(\mathcal{X}, \mathcal{Y}) = \|\bar{\mathbf{x}} - \bar{\mathbf{y}}\|^2 + KLS(\mathcal{X}, \mathcal{Y})$.

Jensen–Shannon Distance. KL divergence has a number of numerical difficulties when covariances are close to singularity. A variant to overcome such a problem is the *Jensen–Shannon divergence* (JSD), defined as $JSD(\mathcal{X}, \mathcal{Y}) = \frac{1}{2}(KL(\mathcal{X}, \frac{\mathcal{X}+\mathcal{Y}}{2}) + KL(\mathcal{Y}, \frac{\mathcal{X}+\mathcal{Y}}{2}))$, which in the MGD case changes into

$$JSD(\mathcal{X}, \mathcal{Y}) = \frac{1}{2} \left(\text{tr} \left(2(\Sigma_x + \Sigma_y)^{-1} \Sigma_x \right) + \text{tr} \left(2(\Sigma_x + \Sigma_y)^{-1} \Sigma_y \right) \right) \quad (8)$$

Again, Eq. (8) is just a distance between covariances and must be completed to get a distance between distributions as $\mathcal{D}^2(\mathcal{X}, \mathcal{Y}) = \|\bar{\mathbf{x}} - \bar{\mathbf{y}}\|^2 + JSD(\mathcal{X}, \mathcal{Y})$.

3.3 IRM Similarity Measure

Integrated Region Matching (IRM) measures the overall similarity between images by integrating distances among regions of two images. An advantage of the overall similarity measure is its robustness against poor segmentations. Precisely, a region-to-region match is obtained when regions are significantly similar to each other in terms of the extracted signature \mathcal{X}_i , i.e., the most similar regions are matched first. Then, the whole distance is computed as a weighted sum of distances between region pairs as follows

$$IRM(\mathcal{X}, \mathcal{Y}) = \sum_{i=1}^n \sum_{j=1}^m s_{ij} d_{ij} \quad (9)$$

where $d_{ij} = \mathcal{D}(\mathcal{X}_i, \mathcal{X}_j)$ and $s_{ij} \geq 0$ is the level of significance between two regions, which indicates the relevance of the matching for determining the whole distance between the two images. It is required that the most similar regions get the highest priority, so the IRM algorithm in [11] attempts to assign as much significance as possible to region pairs with the least distance d_{ij} .

Additionally, the selection of the region weights w_i must be faced. These values are related to the levels of significance as $\sum_j s_{ij} = w_i$ and $\sum_i s_{ij} = w'_j$. Despite that choice can be done *uniformly* for all regions, we prefer the *area percentage scheme*, where w_i is the ratio between region area and image area, since more salient objects in an image tend to occupy larger areas, besides of being less sensitive to inaccurate segmentations.

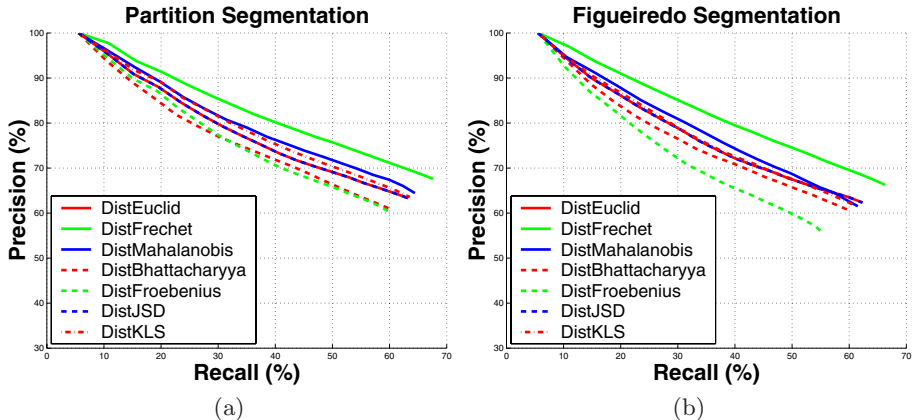


Fig. 2. Precision vs. Recall plots corresponding to each segmentation algorithm: (a) graph-partition and (b) Figueiredo’s EM.

Matching experiment consists in evaluating every distance as a way to perform object identification. The former database was divided into two disjoint subsets containing 9 images per object, that is, 459 images per set in total. One set is the *testing* set while the other is the *sample* set. Then, the experiment 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.

Table 1. Object recognition results per segmentation method.

Distance	Figueiredo [%]	Partition [%]
Euclidean	89.11 (4)	91.29 (3)
Fréchet	94.34 (1)	95.21 (1)
Mahalanobis	89.54 (2)	91.50 (2)
Bhattacharyya	86.71 (5)	88.02 (6)
Fröbenius	84.53 (6)	89.32 (5)
Jensen–Shannon	89.11 (4)	91.29 (3)
KL Symmetric	89.32 (3)	91.07 (4)

Results corresponding to object identification are exhibited in Table 1 performed using Lab color coordinates after the two previously mentioned color image segmentation algorithms, namely, graph-partition and Figueiredo’s EM. In regard to the recognition rates, the best overall results were obtained using our segmentation and the Fréchet distance, and was as high as 95.21% of correct object identifications, outperforming Mahalanobis distance. Euclidean distance obtains medium positions, similarly to JSD and KLS, whereas the worst ones are both Fröbenius and Bhattacharyya.

5 Conclusions

This paper illustrated the problem of comparing images by means of their color segmentations. A group of seven distances were proposed within the frame of the IRM distance and the employ of Multivariate Gaussian Distributions (MGD) for the color description of image regions. The performance of these distances was examined in tasks such as image retrieval and object recognition using the two segmentation algorithms in [1] and [2]. The best overall results were obtained for both tasks using the graph-partition approach along with the Fréchet distance, outperforming other distances in comparing MGDs.

References

1. Vergés-Llahí, J., Climent, J., Sanfeliu, A.: Colour image segmentation solving hard-constraints on graph-partitioning greedy algorithm. In: Proc. 15th International Conference on Pattern Recognition , ICPR'00. Volume 3. (2000) 629–632
2. Figueiredo, M., Jain, A.: Unsupervised learning of finite mixture models. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **24** (2002) 381–396
3. Flickner, M., Sawhney, H., Niblack, W., Ashley, J., Dim, B., Huang, Q., Gorkani, M., Hafner, J., Lee, D., Petkovic, D., Steele, D., Yanker, P.: Query by image and video content: The QBIC system. *Computer* **28** (1995) 23–32
4. Pentland, A., Picard, R., Sclaroff, S.: Photobook: Tools for content-based manipulation of image databases. *Proc. SPIE* **2185** (1994) 34–47
5. Gupta, A., Jain, R.: Visual information retrieval. *Comm. ACM* **40** (1997) 69–79
6. Smith, J., Chang, S.: VisualSEEK: A fully automated content-based query system. In: Proc. ACM Multimedia. (1996) 87–98
7. Wang, J., Wiederhold, G., Firschein, O., Sha, X.: Content-based image indexing and searching using Daubechies' wavelets. *Int'l Digital Libraries* **1** (1998) 311–328
8. Natsev, A., Rastogi, R., Shim, K.: WALRUS: A similarity retrieval algorithm for images databases. *IEEE Trans. on Knowledge and Data Eng.* **16** (2004) 301–316
9. Ma, W., Manjunath, B.: NeTra: A toolbox for navigating large image database. In: Proc. IEEE Int'l Conf. Image Processing. (1997) 568–571
10. Carson, C., Belongie, S., Greenspan, H., Malik, J.: Blobworld: Image segmentation using expectation-maximization and its application to image querying. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **24** (2002) 1026–1038
11. Wang, J., Li, J., Wiederhold, G.: SIMPLiCity: Semantics-sensitive integrated matching for picture libraries. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **23** (2001) 947–963
12. Nascimento, M., Sridhar, V., Li, X.: Effective and efficient region-based image retrieval. *Journal of Visual Languages and Computing* **14** (2003) 151–179
13. Dowson, D., Landau, B.: The Fréchet distance between multivariate normal distributions. *Journal of Multivariate Analysis* **12** (1982) 450–455
14. Bhattacharyya, A.: On a measure of diverg. between two statistical populations defined by their probability distrib. *Bull. Calcutta Math. Soc.* **35** (1943) 99–110
15. Comaniciu, D., Ramesh, V., Meer, P.: Kernel-based object tracking. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **25** (2003) 564–577
16. Kondor, R., Jebara, T.: A kernel between sets of vectors. In: Proc. Int. Conf. on Machine Learning, ICML 2003. (2003)
17. Kullback, S.: *Information Theory and Statistics*. Dover, New York (1968)