Available at
www.ComputerScienceWeb.com
POWERED BY SCIENCE DIRECT®

Pattern Recognition Letters 24 (2003) 227–236

Pattern Recognition Letters

www.elsevier.com/locate/patrec

Dissimilarity computation through low rank corrections

Dorin Comaniciu [a,*], Peter Meer [b], David Tyler [c]

[a] Imaging and Visualization Department, Siemens Corporate Research, Inc., 755 College Road East, Princeton, NJ 08540, USA
[b] Department of Electrical and Computer Engineering, Rutgers University, 94 Brett Road, Piscataway, NJ 08854-8058, USA
[c] Department of Statistics, Rutgers University, Piscataway, NJ 08854, USA

Received 24 September 2001; received in revised form 15 May 2002

Abstract

Most of the energy of a multivariate feature is often contained in a low dimensional subspace. We exploit this property for the efficient computation of a dissimilarity measure between features using an approximation of the Bhattacharyya distance. We show that for normally distributed features the Bhattacharyya distance is a particular case of the Jensen–Shannon divergence, and thus evaluation of this distance is equivalent to a statistical test about the similarity of the two populations. The accuracy of the proposed approximation is tested for the task of texture retrieval.

Keywords: Dissimilarity metric; Bhattacharyya distance; Low rank correction; Statistical homogeneity; Decision theoretic image retrieval

1. Introduction

Visual features such as texture or color are often defined at the output of a window operator or pixel-wise computations. From the ensemble of outputs, statistical information about the variation of the feature across a region (or the entire image) can be obtained, most often described by a mean vector and a covariance matrix.

The identification of visual features that provide sufficient discrimination between the image classes in a database has recently received a lot of attention, see (Antani et al., 1998; Comaniciu et al., 2000; Cox et al., 1998; Flickner et al., 1995; Ma and Manjunath, 1997; Ortega et al., 1997; Pentland et al., 1996; Smith and Chang, 1998). However, much less work has been devoted to finding compact feature representations and efficient computation methods for the dissimilarity measure used to infer the ranking of the features (Kelly et al., 1996; Santini and Jain, 1999; Vasconcelos and Lippman, 1999).

In this paper we focus on the properties of the Bhattacharyya distance (Fukunaga, 1990, p. 99) which allow a compact feature representation and an efficient dissimilarity computation. Our work is motivated by the observation that most of the energy of a multivariate feature is frequently

* Corresponding author. Tel.: +1-609-734-3643; fax: +1-609-734-6565.

E-mail addresses: comanici@scr.siemens.com (D. Comaniciu), meer@caip.rutgers.edu (P. Meer), dtyler@stat.rutgers.edu (D. Tyler).

0167-8655/03/$ - see front matter © 2002 Elsevier Science B.V. All rights reserved.
PII: S0167-8655(02)00214-3

contained in a low dimensional subspace. We show that when the features are normally distributed the Bhattacharyya distance is equal to a specialized version of the Jensen–Shannon divergence (Lin, 1991). As a result, the computation of the dissimilarity measure is equivalent to a statistical test for the similarity of the two populations.

As an application we investigate the task of texture retrieval based on two commonly used representations: the multiresolution simultaneous autoregressive (MRSAR) model (Mao and Jain, 1992) and Gabor features (Manjunath and Ma, 1996). For two standard texture databases using the approximated Bhattacharyya distance consistently yielded a retrieval performance comparable to that using the exact distance, and always better than that using the traditional Mahalanobis distance. The proposed approximate computation decreased the arithmetic complexity and the logical database size.

The paper is organized as follows. Section 2 presents the statistical motivation for using the Bhattacharyya distance as dissimilarity measure. In Section 3 the arithmetic complexity of the exact Bhattacharyya distance and the size of the associated logical data are discussed. The approximate computation of the Bhattacharyya distance is described in Section 4. Section 5 presents an experimental evaluation of the proposed dissimilarity measure.

2. Motivation for Bhattacharyya distance

In its general form, the Bhattacharyya distance between two arbitrary distributions $\{p_i(\boldsymbol{x})\}_{i=1,2}$ is defined as (Kailath, 1967)

$$d_{\mathrm{B}}^2 = -\log \int \sqrt{p_1(\boldsymbol{x})p_2(\boldsymbol{x})}\,\mathrm{d}\boldsymbol{x}. \tag{1}$$

Although not a metric (it does not obey the triangle inequality), the distance (1) is popular in classification problems since it is closely related to the Bayes error. Geometric interpretation of the Bhattacharyya distance, its relation to the Fisher measure of information, the statistical properties of the sample estimates, and explicit forms for various distributions are given in (Djouadi et al., 1990; Kailath, 1967).

Our motivation for the use of the Bhattacharyya distance is given by its relationship to the Jensen–Shannon divergence based statistical test of the homogeneity between two distributions (Lin, 1991)

$$\begin{aligned}\mathrm{JS}(p_1,p_2) = &\int p_1(\boldsymbol{x}) \log \frac{p_1(\boldsymbol{x})}{(p_1(\boldsymbol{x})+p_2(\boldsymbol{x}))/2}\,\mathrm{d}\boldsymbol{x}\\ &+ \int p_2(\boldsymbol{x}) \log \frac{p_2(\boldsymbol{x})}{(p_1(\boldsymbol{x})+p_2(\boldsymbol{x}))/2}\,\mathrm{d}\boldsymbol{x}.\end{aligned} \tag{2}$$

We show in Appendix A that if the two distributions are normal, Jensen–Shannon divergence reduces to the Bhattacharyya distance (up to a constant). Thus, we formulate the similarity problem as a goodness-of-fit test between the empirical distributions p_1 and p_2 and the homogeneous model $(p_1+p_2)/2$.

In a comparative study of similarity measures Puzicha et al. (1997) found the Jensen–Shannon divergence superior to the Cramer–von Mises and Kolmogorov–Smirnov statistical tests. The Jensen–Shannon divergence has also been used by Ojala et al. (1996) for texture classification. More recently, Vasconcelos and Lippman (2000) presented a discussion on similarity measures. They show that most of the similarity functions are related to the Bayesian criterion.

Other measures such as the Fisher linear discriminant function yield useful results only when the two distributions have different means (Fukunaga, 1990, p. 132), whereas the Kullback divergence (Cover and Thomas, 1991, p. 18) provides in various instances lower performance than the Bhattacharyya distance (Kailath, 1967). The Bhattacharyya distance is a particular case of the Chernoff distance (Fukunaga, 1990, p. 97). While the latter in general provides a better bound for the Bayesian error, it is more difficult to evaluate. The Chernoff and Bhattacharyya bounds have been used recently in (Konishi et al., 1999) to analyze the performance of edge detectors.

The expression (1) is defined for arbitrary distributions, however, we will assume in the sequel that the distribution of the feature of interest is unimodal, characterized by its mean vector $\boldsymbol{\mu} \in R^p$ and covariance matrix $\boldsymbol{C} \in R^{p\times p}$. This is equivalent to describing each image as homogeneous relative to the feature under consideration. When the feature distribution is a multivariate normal, the mean vector and covariance matrix uniquely define it, otherwise they provide an incomplete but most often satisfactory representation. In (Xu et al., 2000) we have discussed the multimodal case, corresponding to nonhomogeneous images.

3. Exact distance computation

For two p-dimensional *normal* distributions characterized by $\boldsymbol{\mu}_\mathrm{q}$, $\boldsymbol{C}_\mathrm{q}$ and $\boldsymbol{\mu}_\mathrm{d}$, $\boldsymbol{C}_\mathrm{d}$ (subscripts q and d were chosen for the intended application, denoting query and database), the Bhattacharyya distance (1) becomes

$$d_\mathrm{B}^2 = \frac{1}{4}(\boldsymbol{\mu}_\mathrm{q} - \boldsymbol{\mu}_\mathrm{d})^\mathrm{T}(\boldsymbol{C}_\mathrm{q} + \boldsymbol{C}_\mathrm{d})^{-1}(\boldsymbol{\mu}_\mathrm{q} - \boldsymbol{\mu}_\mathrm{d}) + \frac{1}{2}\log\frac{\left|\frac{\boldsymbol{C}_\mathrm{q}+\boldsymbol{C}_\mathrm{d}}{2}\right|}{\sqrt{|\boldsymbol{C}_\mathrm{q}||\boldsymbol{C}_\mathrm{d}|}} \tag{3}$$

where $|\cdot|$ is the determinant and T is the transpose operator. The first term in (3) gives the class separability due to the mean-difference, while the second term gives class separability due to the covariance-difference.

The efficient way to compute (3) is through Cholesky factorization (Golub and VanLoan, 1996, p. 143). Since $\boldsymbol{C}_\mathrm{q}$ and $\boldsymbol{C}_\mathrm{d}$ are symmetric and positive definite, their sum has the same property. The case of singular matrices (which is less frequent in practice) can be always avoided by reducing the dimension of the feature space, and will not be considered in the sequel. Hence, we can write

$$\boldsymbol{C}_\mathrm{q} + \boldsymbol{C}_\mathrm{d} = \boldsymbol{L}\boldsymbol{L}^\mathrm{T}, \tag{4}$$

where $\boldsymbol{L} \in R^{p\times p}$ is lower triangular with positive diagonal entries. Computing the inverse as

$$(\boldsymbol{C}_\mathrm{q} + \boldsymbol{C}_\mathrm{d})^{-1} = (\boldsymbol{L}^{-1})^\mathrm{T}\boldsymbol{L}^{-1} \tag{5}$$

and introducing (4) and (5) into (3), yields

$$d_\mathrm{B}^2 = \frac{1}{4}\|\boldsymbol{L}^{-1}(\boldsymbol{\mu}_\mathrm{q} - \boldsymbol{\mu}_\mathrm{d})\|_2^2 + \frac{1}{2}\log\frac{\prod_{i=1}^{p} l_{ii}^2}{2^p\sqrt{|\boldsymbol{C}_\mathrm{q}||\boldsymbol{C}_\mathrm{d}|}}, \tag{6}$$

where $\|\cdot\|_2$ is the Euclidean norm, $\{l_{ii}\}_{i=1,\dots,p}$ are the diagonal elements of $\boldsymbol{L}$, and $(\prod_{i=1}^{p} l_{ii}^2)/2^p$ is the determinant of $(\boldsymbol{C}_\mathrm{q} + \boldsymbol{C}_\mathrm{d})/2$.

Since the determinant for the query $|\boldsymbol{C}_\mathrm{q}|$ is computed only once and the determinant for a database entry $|\boldsymbol{C}_\mathrm{d}|$ is computed off-line, the online computation of each Bhattacharyya distance according to expression (6) requires about

- $p^2/2$ flops to obtain the sum $\boldsymbol{C}_\mathrm{q} + \boldsymbol{C}_\mathrm{d}$;
- $(p^3/3) + (p^2/2)$ flops for the Cholesky factorization (Golub and VanLoan, 1996, p. 144);
- p^2 flops to compute the vector $\boldsymbol{L}^{-1}(\boldsymbol{\mu}_\mathrm{q} - \boldsymbol{\mu}_\mathrm{d})$ through backward substitution.

Hence, the computation of the dissimilarity measure (6) involves a total of $(p^3/3) + 2p^2$ flops. Each entry in the logical database should contain the mean vector $\boldsymbol{\mu}_\mathrm{d}$, the covariance matrix $\boldsymbol{C}_\mathrm{d}$, and the value of the determinant $|\boldsymbol{C}_\mathrm{d}|$, which means about $p^2/2$ floating point numbers, due to the symmetry of $\boldsymbol{C}_\mathrm{d}$.

4. Approximate distance computation

The Bhattacharyya distance (3) is symmetric in $\boldsymbol{C}_\mathrm{q}$ and $\boldsymbol{C}_\mathrm{d}$, but the search in the database involves repeated computation of (3) with the same $\boldsymbol{C}_\mathrm{q}$ and different $\boldsymbol{C}_\mathrm{d}$s. To reduce the logical database size, the $\boldsymbol{C}_\mathrm{d}$s should be stored as efficiently as possible. Often the feature vectors belong to a subspace and the complexity of the retrieval process can be decreased if the *effective* dimension of the feature is taken into account.

The underlying structure of a multivariate feature is revealed by principal component analysis, e.g., singular value decomposition (SVD) of its covariance matrix. Let SVD of the matrices $\boldsymbol{C}_\mathrm{q}$ and $\boldsymbol{C}_\mathrm{d}$ be

$$\boldsymbol{C}_\mathrm{q} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^\mathrm{T} \quad \text{and} \quad \boldsymbol{C}_\mathrm{d} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^\mathrm{T}, \tag{7}$$

where $\boldsymbol{U} \in R^{p\times p}$ and $\boldsymbol{V} \in R^{p\times p}$ are orthogonal matrices, $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1, \dots, \sigma_p) \in R^{p\times p}$ and $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \dots, \lambda_p) \in R^{p\times p}$ are diagonal, the singular values (SVs) $\{\sigma_i\}_{i=1,\dots,p}$ and $\{\lambda_i\}_{i=1,\dots,p}$ being strictly positive and in a decreasing order.

When the effective dimensionality of the feature in the database is r, the last $p-r$ SVs of $\boldsymbol{C}_\mathrm{d}$ are small, we can make a first approximation by assuming that they are all equal to a value λ computed as their geometric mean,

$$\lambda_i \approx \lambda = \left(\prod_{j=r+1}^{p} \lambda_j\right)^{\frac{1}{p-r}} \qquad i = r+1,\dots,p. \tag{8}$$

It is easy to see that (8) preserves the determinant of $\boldsymbol{C}_\mathrm{d}$ (which is equal to $\prod_{i=1}^{p} \lambda_i$). The geometric interpretation is that the hyperellipsoid E_d defined by $E_\mathrm{d} = \{\boldsymbol{C}_\mathrm{d}\boldsymbol{x} : \|\boldsymbol{x}\|_2 = 1\}$, is replaced by another hyperellipsoid with the same volume, but having the values of the last $p-r$ semi-axes equal to λ. The approximation (8) is expected to introduce only small changes in the matrix $\boldsymbol{C}_\mathrm{d}$. For example, when $r=1$, a 3D ellipsoid having $\boldsymbol{\Lambda} = \mathrm{diag}(8,2,1)$ (Fig. 1(a)) becomes the ellipsoid having $\boldsymbol{\Lambda} = \mathrm{diag}(8, 1.41, 1.41)$ in Fig. 1(b).

Denoting by $\boldsymbol{I}_p$ the identity matrix of order p, by $\{\boldsymbol{v}_i\}_{i=1,\dots,p}$ the columns of $\boldsymbol{V}$, and using the orthogonality of $\boldsymbol{V}$ we have $\sum_{i=1}^{p} \boldsymbol{v}_i\boldsymbol{v}_i^\mathrm{T} = \boldsymbol{I}_p$. Then

$$\begin{aligned} C_\mathrm{d} &\approx \sum_{i=1}^{r} \lambda_i \boldsymbol{v}_i\boldsymbol{v}_i^\mathrm{T} + \lambda \sum_{i=r+1}^{p} \boldsymbol{v}_i\boldsymbol{v}_i^\mathrm{T} \\ &= \lambda\boldsymbol{I}_p + \sum_{i=1}^{r} (\lambda_i - \lambda)\boldsymbol{v}_i\boldsymbol{v}_i^\mathrm{T} = \lambda\boldsymbol{I}_p + \boldsymbol{V}_r\boldsymbol{\Psi}\boldsymbol{V}_r^\mathrm{T} \\ &= \lambda\boldsymbol{I}_p + \boldsymbol{W}_r\boldsymbol{W}_r^\mathrm{T}, \end{aligned} \tag{9}$$

where $\boldsymbol{\Psi} = \mathrm{diag}(\psi_1,\dots,\psi_r) \in R^{r\times r}$ has positive entries $\psi_i = \lambda_i - \lambda$ for $i = 1,\dots,r$, $\boldsymbol{V}_r = [\boldsymbol{v}_1,\dots,\boldsymbol{v}_r] \in R^{p\times r}$, and

$$\boldsymbol{W}_r = \boldsymbol{V}_r\boldsymbol{\Psi}^{\frac{1}{2}} \in R^{p\times r}. \tag{10}$$

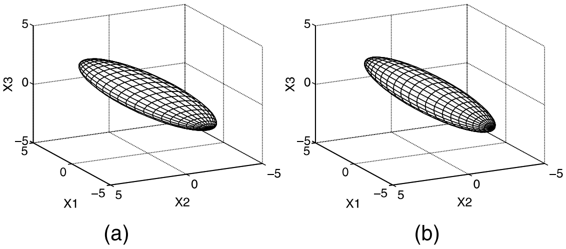

Fig. 1. The employed approximation changes the shape of the original ellipsoid (a) to that of the ellipsoid (b). The volume remains the same.

Using (9) and the orthogonality of $\boldsymbol{U}$ we can write the sum of the two covariance matrices as

$$\begin{aligned} \boldsymbol{C}_\mathrm{q} + \boldsymbol{C}_\mathrm{d} &\approx \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^\mathrm{T} + \lambda\boldsymbol{I}_p + \boldsymbol{W}_r\boldsymbol{W}_r^\mathrm{T} \\ &= \boldsymbol{U}\boldsymbol{\Gamma}\boldsymbol{U}^\mathrm{T} + \boldsymbol{W}_r\boldsymbol{W}_r^\mathrm{T} = \boldsymbol{C}_\gamma + \boldsymbol{W}_r\boldsymbol{W}_r^\mathrm{T}, \end{aligned} \tag{11}$$

where $\boldsymbol{\Gamma} = \mathrm{diag}(\gamma_1,\dots,\gamma_p) \in R^{p\times p}$ with $\gamma_i = \sigma_i + \lambda$ for $i = 1,\dots,p$.

The relation (11) shows that $\boldsymbol{C}_\mathrm{q} + \boldsymbol{C}_\mathrm{d}$ can be approximated as a sum of a full rank matrix $\boldsymbol{C}_\gamma = \boldsymbol{U}\boldsymbol{\Gamma}\boldsymbol{U}^\mathrm{T}$ and a rank r correction $\boldsymbol{W}_r\boldsymbol{W}_r^\mathrm{T}$. The following two sections will exploit this decomposition to approximate the two terms of the Bhattacharyya distance (3).

4.1. First Bhattacharyya term

The first term of (3) requires the inverse $(\boldsymbol{C}_\mathrm{q} + \boldsymbol{C}_\mathrm{d})^{-1}$. A rank r correction to a matrix results in a rank r correction of the inverse expressed by the Sherman–Morrison–Woodbury formula (Golub and VanLoan, 1996, p. 50)

$$\begin{aligned} (\boldsymbol{C}_\gamma + \boldsymbol{W}_r\boldsymbol{W}_r^\mathrm{T})^{-1} &= \boldsymbol{C}_\gamma^{-1} - \boldsymbol{C}_\gamma^{-1}\boldsymbol{W}_r \\ &\quad \times (\boldsymbol{I}_r + \boldsymbol{W}_r^\mathrm{T}\boldsymbol{C}_\gamma^{-1}\boldsymbol{W}_r)^{-1}\boldsymbol{W}_r^\mathrm{T}\boldsymbol{C}_\gamma^{-1} \end{aligned} \tag{12}$$

with $\boldsymbol{I}_r$ being the identity matrix of order r. Since $\boldsymbol{C}_\gamma^{-1} = \boldsymbol{U}\boldsymbol{\Gamma}^{-1}\boldsymbol{U}^\mathrm{T}$, the right side of the relation (12) becomes

$$\boldsymbol{U}\boldsymbol{\Gamma}^{-\frac{1}{2}}[\boldsymbol{I}_p - \boldsymbol{Z}_\gamma(\boldsymbol{I}_r + \boldsymbol{Z}_\gamma^\mathrm{T}\boldsymbol{Z}_\gamma)^{-1}\boldsymbol{Z}_\gamma^\mathrm{T}]\boldsymbol{\Gamma}^{-\frac{1}{2}}\boldsymbol{U}^\mathrm{T}, \tag{13}$$

where $\boldsymbol{Z}_\gamma \in R^{p\times r}$ is given by

$$\boldsymbol{Z}_\gamma = \boldsymbol{\Gamma}^{-\frac{1}{2}}\boldsymbol{U}^\mathrm{T}\boldsymbol{W}_r = \boldsymbol{\Gamma}^{-\frac{1}{2}}\boldsymbol{U}^\mathrm{T}\boldsymbol{V}_r\boldsymbol{\Psi}^{\frac{1}{2}} = \boldsymbol{\Gamma}^{-\frac{1}{2}}\boldsymbol{Z}\boldsymbol{\Psi}^{\frac{1}{2}}, \tag{14}$$

where

$$\boldsymbol{Z} = \boldsymbol{U}^\mathrm{T}\boldsymbol{V}_r \in R^{p\times r}. \tag{15}$$

Finally, defining the vectors $\boldsymbol{\xi} = \boldsymbol{\Gamma}^{-\frac{1}{2}}\boldsymbol{U}^\mathrm{T}(\boldsymbol{\mu}_\mathrm{q} - \boldsymbol{\mu}_\mathrm{d})$ and $\boldsymbol{v} = \boldsymbol{Z}_\gamma^\mathrm{T}\boldsymbol{\xi}$, the first term of the Bhattacharyya distance (3) is approximated by

$$d_{\mathrm{B1}}^2 \approx \tfrac{1}{4}[\boldsymbol{\xi}^\mathrm{T}\boldsymbol{\xi} - \boldsymbol{v}^\mathrm{T}(\boldsymbol{I}_r + \boldsymbol{Z}_\gamma^\mathrm{T}\boldsymbol{Z}_\gamma)^{-1}\boldsymbol{v}]. \tag{16}$$

The matrix inversion in (16) is solved through Cholesky factorization. Since $(\boldsymbol{I}_r + \boldsymbol{Z}_\gamma^\mathrm{T}\boldsymbol{Z}_\gamma) \in R^{r\times r}$, the amount of computation is negligible for $r \ll p$.

Hence, the online computation of (16) requires about

- $2rp^2$ flops to obtain the matrix $\boldsymbol{Z}$;
- $2p^2$ flops to derive the vector ξ.

All the other operations, such as the derivation of $\boldsymbol{Z}_\gamma$ or $\boldsymbol{v}$ involve a number of flops of order rp which again are negligible. In conclusion, for $r \ll p$ the approximate computation of the first term d_{B1}^2 of the Bhattacharyya distance requires about $2rp^2 + 2p^2$ flops.

4.2. Second Bhattacharyya term

To obtain the second term of the Bhattacharyya distance (3) we need the SVs of $\boldsymbol{C}_q + \boldsymbol{C}_d$. By using (10), (11), and (15) we have

$$\begin{aligned}\boldsymbol{C}_q + \boldsymbol{C}_d &\approx \boldsymbol{U}\boldsymbol{\Gamma}\boldsymbol{U}^T + \boldsymbol{V}_r\boldsymbol{\Psi}\boldsymbol{V}_r^T = \boldsymbol{U}(\boldsymbol{\Gamma} + \boldsymbol{Z}\boldsymbol{\Psi}\boldsymbol{Z}^T)\boldsymbol{U}^T \\ &= \boldsymbol{U}\left(\boldsymbol{\Gamma} + \sum_{i=1}^{r}\psi_i \boldsymbol{z}_i \boldsymbol{z}_i^T\right)\boldsymbol{U}^T, \end{aligned} \tag{17}$$

where $\boldsymbol{z}_i$ are the columns of $\boldsymbol{Z}$ with $\|\boldsymbol{z}_i\|_2 = 1$, for $i = 1, \dots, r$. From the last relation it results that the SVs of $\boldsymbol{C}_q + \boldsymbol{C}_d$ are approximately given by the SVs of the symmetric positive definite matrix $\boldsymbol{B} \in R^{p\times p}$

$$\boldsymbol{B} = \boldsymbol{\Gamma} + \sum_{i=1}^{r}\psi_i \boldsymbol{z}_i \boldsymbol{z}_i^T, \tag{18}$$

whose expression contains r rank one updating steps. It was shown (Golub, 1973; Golub and VanLoan, 1996, p. 443), that the p SVs of a rank one updated matrix $\boldsymbol{\Gamma} + \psi \boldsymbol{z}\boldsymbol{z}^T$, with $\boldsymbol{\Gamma} = \mathrm{diag}(\gamma_1, \dots, \gamma_p) \in R^{p\times p}$ and $\boldsymbol{z} = (z_1, \dots, z_p)^T \in R^p$, are the zeros of the secular equation

$$f(\zeta) \equiv 1 + \psi \sum_{j=1}^{p} \frac{z_j^2}{\gamma_j - \zeta} = 0. \tag{19}$$

The secular equation (19) appears in the context of symmetric eigenvalue problems (Dongarra and Sorensen, 1987) and can be solved through a fast, quadratically convergent procedure, as described in (Barlow, 1993; Bunch et al., 1978).

Based on the above formulation, we solve r secular equations to find the set of SVs $\{\beta_{ij}\}_{i=1,\dots,r,\, j=1,\dots,p}$ of the matrices

$$\boldsymbol{B}_i = \widehat{\boldsymbol{B}}_{i-1} + \psi_i \boldsymbol{z}_i \boldsymbol{z}_i^T, \quad i = 1, \dots, r, \tag{20}$$

where $\widehat{\boldsymbol{B}}_0 = \boldsymbol{\Gamma}$ and $\widehat{\boldsymbol{B}}_{i-1} = \mathrm{diag}(\beta_{i-1,1}, \dots, \beta_{i-1,p})$ for $i > 1$. Finally, the SVs β_j of $\boldsymbol{B}$ are approximated as

$$\beta_j \approx \beta_{rj}, \quad j = 1, \dots, p, \tag{21}$$

which means that we compute r updates of the SVs, but neglect the updating of the singular vectors.

The second term of Bhattacharyya distance (3) is then given by

$$d_{B2}^2 \approx \frac{1}{2}\log\frac{\prod_{j=1}^{p}\beta_j}{2^p\sqrt{|\boldsymbol{C}_q||\boldsymbol{C}_d|}}. \tag{22}$$

The number of flops necessary to find all sets $\{\beta_{ij}\}$ is not significant, having the order of rp (Bunch et al., 1978).

Therefore, in the case of $r \ll p$ the arithmetic complexity of the approximate computation of the complete Bhattacharyya distance remains $2rp^2 + 2p^2$ flops. Each entry in the logical database should contain the mean vector $\boldsymbol{\mu}_d$, the matrix $\boldsymbol{V}_r$, the SVs $\lambda_1, \dots, \lambda_r$, and the value of the determinant $|\boldsymbol{C}_d|$, which means about $(r+1)p$ floating point numbers. Compare this with $(p^3/3) + 2p^2$ flops and $p^2/2$ floating point numbers required by the full computation.

5. An experimental validation

We compared the performance of the approximate Bhattacharyya distance similarity measure against the exact Bhattacharyya distance and the traditionally employed Mahalanobis distance, in a texture retrieval task. Two standard texture databases, VisTex Texture Database (1995) and Brodatz (1966) were employed and to generate the image classes the technique described in (Picard et al., 1993) was used. The VisTex database contained 1188 subimages representing 132 equally populated classes. The Brodatz database contained 1008 subimages corresponding to 112 equally populated classes. Typical images are shown in Fig. 2.

Two different texture representations were used in the experiments. Similar to Picard et al. (1993) we used 15-dimensional feature vectors to define a MRSAR model (Mao and Jain, 1992). The mean and covariance of these vectors were computed for each query image and database entry. The second representation employed was with Gabor features similar to those described in (Manjunath and Ma, 1996). However, we also took into account the cross-correlation between the filtered images. Thus, each input image was processed at four scales and six orientations to obtain 24 filtered images that defined a 24-dimensional feature space. The mean vector and the covariance matrix of the vectors in the space were then computed. Fig. 3 shows that indeed, most of the energy of both MRSAR (Fig. 3(a) and (b) for VisTex and Brodatz, respectively) and Gabor (Fig. 3(c) for VisTex) features is contained in a low dimensional space.

The Mahalanobis distance is a widely used dissimilarity measure. In fact it can be obtained from the Bhattacharyya distance by taking $\boldsymbol{C}_{\mathrm{q}} = \boldsymbol{C}_{\mathrm{d}} = \boldsymbol{C}$,

$$d_{\mathrm{M}}^2 = (\boldsymbol{\mu}_{\mathrm{q}} - \boldsymbol{\mu}_{\mathrm{d}})^{\mathrm{T}} \boldsymbol{C}^{-1} (\boldsymbol{\mu}_{\mathrm{q}} - \boldsymbol{\mu}_{\mathrm{d}}). \tag{23}$$

To quantitatively assess the retrieval performance over an entire database, the average recognition rate (ARR) showing the average percentage of images from the class of the query (maximum eight images) being among the first N retrievals, was used (see Picard et al. (1993) for details). Fig. 4 shows the retrieval performance when MRSAR features are employed for retrievals from the VisTex database. In Fig. 4(a) the dissimilarity measure is the complete Bhattacharyya distance and the Mahalanobis distance defined in

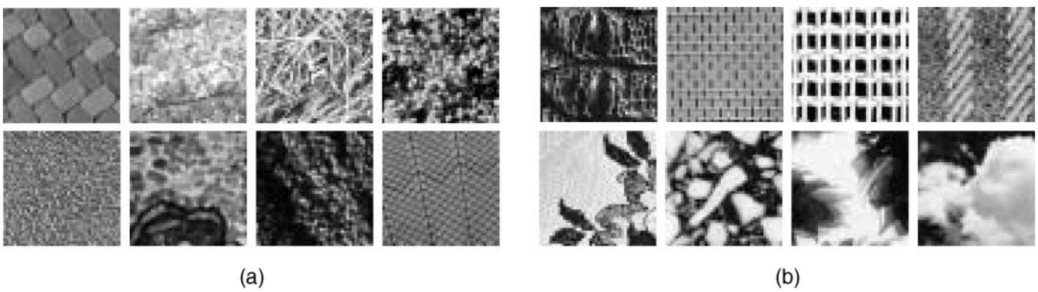

Fig. 2. Examples of 128 × 128 images from: (a) VisTex database; (b) Brodatz database.

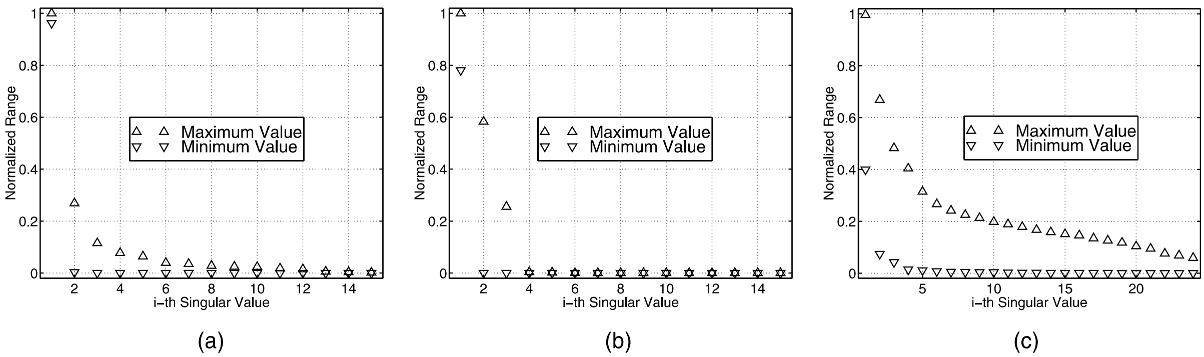

Fig. 3. The normalized range of the SVs of texture features for different databases. (a) 1188 vectors of SVs for the MRSAR features from VisTex, (b) 1008 vectors of SVs for the MRSAR features from Brodatz, (c) 1188 vectors of SVs for the Gabor features from VisTex.

two ways, the common matrix being taken as that of the query image $\boldsymbol{C} = \boldsymbol{C}_{\mathrm{q}}$ or that of the database entry $\boldsymbol{C} = \boldsymbol{C}_{\mathrm{d}}$. We provided both sets of results since there is no theoretical rationale which of the matrices $\boldsymbol{C}_{\mathrm{q}}$ or $\boldsymbol{C}_{\mathrm{d}}$ should be employed, the performance depending on the specific database.

The retrieval performance for the same features and database, when the approximated $(r = 1, 2)$ Bhattacharyya distance is used, is shown in Fig. 4(b) and remains superior to that of Mahalanobis distance. For $r \geqslant 3$ the performance for the approximated and full cases coincide. Retrieval examples are shown in Fig. 5 for $r = 2$. The results for the Brodatz database, shown in Fig. 6(a) and (b), are similar. Note the higher values of ARR due to the more homogeneous nature of this database. It is interesting to observe that the retrieval performance for $r = 2$ can be better than in the case of using the complete Bhattacharyya distance. The reason is that the covariance matrix estimation is affected by errors when the matrix dimension is large. Therefore, the approximation of the covariance matrix in a lower dimensional space might be beneficial for the retrieval process. We, however, cannot claim that this happens consistently.

For a third retrieval experiment we used the Gabor features to describe the VisTex database (Fig. 7). The number of dimensions for the approximation had to be increased to $r = 4$ since the Gabor features exhibit a lesser compaction property (Fig. 3(c)). Observe that the selection of the value of r for a given database should be performed experimentally, since r depends on the compactness of the features employed.

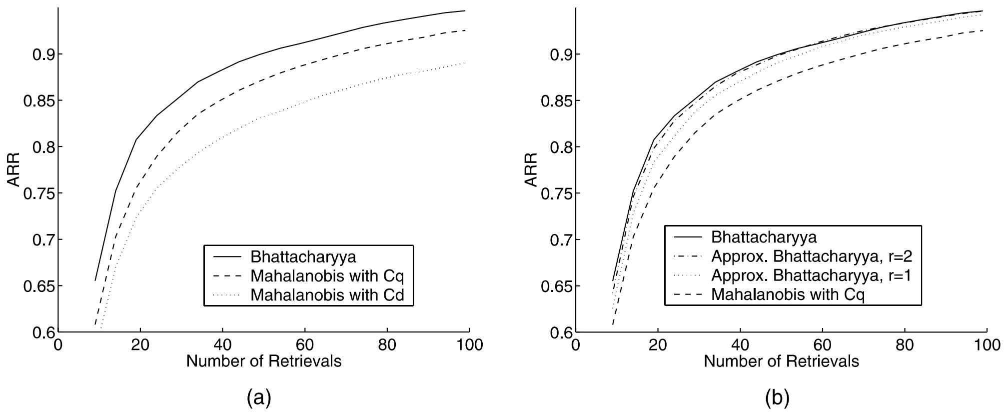

Fig. 4. Retrieval performance of MRSAR features describing the VisTex database, expressed as the ARR function of the number of retrievals. (a) Results obtained by using the complete Bhattacharyya and differently defined Mahalanobis distances, (b) results obtained by approximating the Bhattacharyya distance.

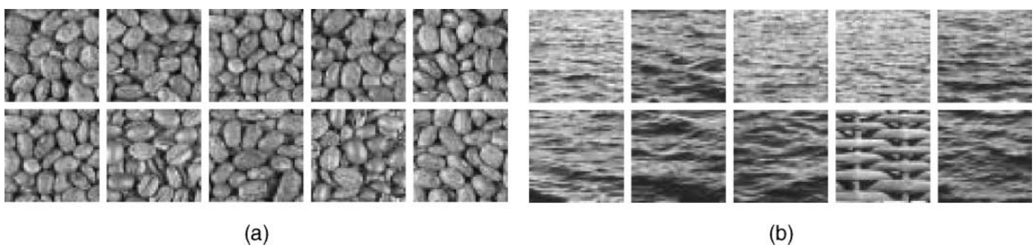

Fig. 5. Two retrieval examples (ordered from left–right, top–down) obtained with MRSAR features describing the VisTex database and approximated Bhattacharyya distance with $r = 2$. The first image in each case is the query image. (a) *Beans:* the sixth and eighth retrievals are from the *Coffee* class. (b) *Water:* the eighth retrieval is from the *Fabric* class.

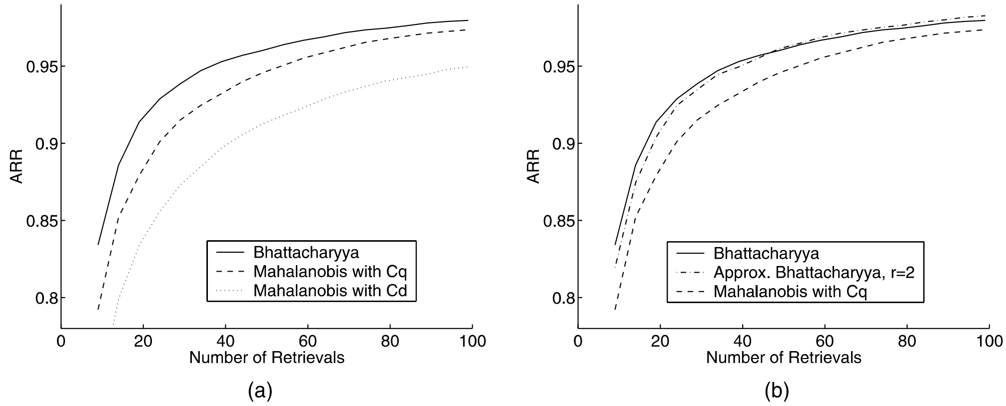

Fig. 6. Retrieval performance of MRSAR features describing the Brodatz database, expressed as the ARR function of the number of retrievals. (a) Results obtained by using the complete Bhattacharyya and differently defined Mahalanobis distances, (b) results obtained by approximating the Bhattacharyya distance.

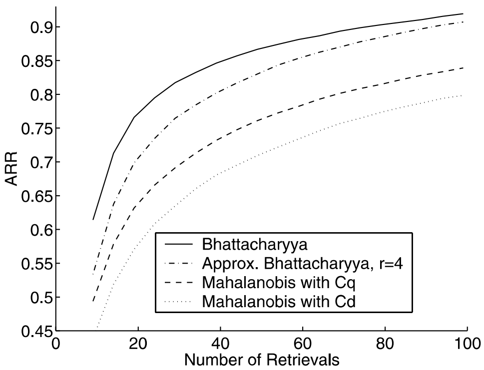

Fig. 7. Retrieval performance of Gabor features describing the VisTex database, expressed as the ARR function of the number of retrievals. Results obtained by using the complete Bhattacharyya, approximated Bhattacharyya with $r = 4$, and differently defined Mahalanobis distances.

6. Conclusion

We have presented a technique for the efficient computation of the dissimilarity between features, based on the approximation of the Bhattacharyya distance. The proposed methodology has a statistical motivation and is appropriate for multivariate features with unimodal distributions. We validated the theory for the task of texture retrieval by employing standard texture databases.

Different extensions of the method to the multimodal case are possible. Some of them were investigated in (Xu et al., 2000) The underlying distribution of the vectors in the feature space can be modeled with a finite mixture of multivariate normal distributions whose expression is used in the computation of the general Bhattacharyya distance (1). In this case, the dissimilarity metric results in a weighted sum of terms similar to (3). Other approach would involve a nonparametric strategy (Comaniciu and Meer, 1999) that computes the general Bhattacharyya distance (1) directly from the data samples.

Acknowledgements

Peter Meer and David Tyler were supported by the National Science Foundation under the grant IRI-99-87695. A preliminary version appeared in the IEEE Workshop on Content-Based Access of Image and Video Libraries, Fort Collins, Colorado, 1999.

Appendix A. Specialized Jensen–Shannon divergence for multivariate normal distributions

Let us assume that $\{p_i(\boldsymbol{x})\}_{i=1,2}$ are two p-dimensional normal distributions

$$p_i(\boldsymbol{x}) = \frac{1}{|2\pi \boldsymbol{C}_i|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_i)^{\mathrm{T}} \boldsymbol{C}_i^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i)\right). \quad \text{(A.1)}$$

The Jensen–Shannon divergence (2) can be specialized to the normal case as

$$\mathrm{JS}(p_1, p_2) = \int p_1(\boldsymbol{x}) \log \frac{p_1(\boldsymbol{x})}{r(\boldsymbol{x})}\, \mathrm{d}\boldsymbol{x} + \int p_2(\boldsymbol{x}) \log \frac{p_2(\boldsymbol{x})}{r(\boldsymbol{x})}\, \mathrm{d}\boldsymbol{x}, \tag{A.2}$$

where $r(\boldsymbol{x})$ is the most likely normal source for the homogeneous model $(p_1 + p_2)/2$, having the mean $\boldsymbol{\mu} = (\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)/2$ and covariance $\boldsymbol{C} = (\boldsymbol{C}_1 + \boldsymbol{C}_2)/2$ (El-Yaniv et al., 1997).

Using (A.1) in (A.2) together with the identity $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{x} = \mathrm{tr}\, \boldsymbol{C}^{-1}\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}$, we first obtain

$$\log \frac{p_i(\boldsymbol{x})}{r(\boldsymbol{x})} = \frac{1}{2} \log \frac{|\boldsymbol{C}|}{|\boldsymbol{C}_i|} - \frac{1}{2} \mathrm{tr}\, \boldsymbol{C}_i^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_i)(\boldsymbol{x} - \boldsymbol{\mu}_i)^{\mathrm{T}} + \frac{1}{2} \mathrm{tr}\, \boldsymbol{C}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})^{\mathrm{T}} \tag{A.3}$$

for $i = 1, 2$, where tr denotes the trace of a matrix. Performing the integration yields

$$\int p_i(\boldsymbol{x}) \log \frac{p_i(\boldsymbol{x})}{r(\boldsymbol{x})}\, \mathrm{d}\boldsymbol{x} = \frac{1}{2} \log \frac{|\boldsymbol{C}|}{|\boldsymbol{C}_i|} + \frac{1}{2} \mathrm{tr}\, \boldsymbol{C}_i\boldsymbol{C}^{-1} - \frac{p}{2} + \frac{1}{2} \mathrm{tr}\, \boldsymbol{C}^{-1}(\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^{\mathrm{T}}. \tag{A.4}$$

Summing (A.4) for $i = 1, 2$ and substituting $\boldsymbol{\mu} = (\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)/2$ and $\boldsymbol{C} = (\boldsymbol{C}_1 + \boldsymbol{C}_2)/2$ we obtain

$$\begin{aligned}\mathrm{JS}(p_1, p_2) &= \log \frac{\left|\frac{\boldsymbol{C}_1 + \boldsymbol{C}_2}{2}\right|}{\sqrt{|\boldsymbol{C}_1||\boldsymbol{C}_2|}} + \frac{1}{2} \mathrm{tr}(\boldsymbol{C}_1 + \boldsymbol{C}_2)\left(\frac{\boldsymbol{C}_1 + \boldsymbol{C}_2}{2}\right)^{-1} \\ &\quad - p + \frac{1}{8} \mathrm{tr}\, \boldsymbol{C}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^{\mathrm{T}} + \frac{1}{8} \mathrm{tr}\, \boldsymbol{C}^{-1}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^{\mathrm{T}} \\ &= \log \frac{\left|\frac{\boldsymbol{C}_1 + \boldsymbol{C}_2}{2}\right|}{\sqrt{|\boldsymbol{C}_1||\boldsymbol{C}_2|}} + \frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^{\mathrm{T}} \times (\boldsymbol{C}_1 + \boldsymbol{C}_2)^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2),\end{aligned} \tag{A.5}$$

an expression equal (up to a constant scale) to that of the Bhattacharyya distance (3).

References

Antani, S., Kasturi, R., Jain, R., 1998. Pattern recognition methods in image and video databases: past, present and future. In: Advances in Pattern Recognition Lecture Notes in Computer Science, Vol. 1451. Springer, Berlin, pp. 31–53.

Barlow, J., 1993. Error analysis of update methods for the symmetric eigenvalue problem. SIAM J. Matrix Anal. Appl. 14, 598–618.

Brodatz, P., 1966. Textures: A Photographic Album for Artists and Designers. Dover, New York.

Bunch, J., Nielsen, C., Sorensen, D., 1978. Rank-one modification of the symmetric eigenproblem. Numerische Matematik 31, 31–48.

Comaniciu, D., Meer, P., 1999. Distribution free decomposition of multivariate data. Pattern Anal. Appl. 2, 22–30.

Comaniciu, D., Meer, P., Foran, D., 2000. Image guided decision support system for pathology. Machine Vision Appl. 11 (4), 213–224.

Cover, T., Thomas, J., 1991. Elements of Information Theory. John Wiley and Sons, New York.

Cox, I., Miller, M., Minka, T., Yianilos, P., 1998. An optimized interaction strategy for bayesian relevance feedback. In: Proceedings IEEE Conference on Computer Vision and Pattern Recognition, Santa Barbara, pp. 553–558.

Djouadi, A., Snorrason, O., Garber, F., 1990. The quality of training-sample estimates of the Bhattacharyya coefficient. IEEE Trans. Pattern Anal. Machine Intell. 12, 92–97.

Dongarra, J., Sorensen, D., 1987. A fully parallel algorithm for the symmetric eigenvalue problem. SIAM J. Sci. Comput. 8, 139–154.

El-Yaniv, R., Fine, S., Tishby, N., 1997. Agnostic classification of markovian sequences. In: Advances in Neural Information Processing Systems, NIPS-97, Vol. 10, pp. 465–471.

Flickner, M. et al., 1995. Query by image and video content: the qbic system. Computer 9 (9), 23–31.

Fukunaga, K., 1990. Introduction to Statistical Pattern Recognition, second ed. Academic Press.

Golub, G., 1973. Some modified matrix eigenvalue problems. SIAM Rev. 15, 318–334.

Golub, G.H., VanLoan, C., 1996. Matrix Computations, third ed. John Hopkins U. Press.

Kailath, T., 1967. The divergence and Bhattacharyya distance measures in signal selection. IEEE Trans. Commun. Tech. 15, 52–60.

Kelly, P., Cannon, M., Barros, J., 1996. Efficiency issues related to probability density function comparison. SPIE Storage and Retrieval for Still Image and Video Databases IV Vol. 2670, 42–49.

Konishi, S., Yuille, A., Coughlan, J., Zhu, S., 1999. Fundamental bounds on edge detection: an information theoretic evaluation of different edge cues. In: Proceedings IEEE Conference on Computer Vision and Pattern Recognition, Fort Collins, pp. 573–579.

Lin, J., 1991. Divergence measures based on the Shannon entropy. IEEE Trans. Inf. Theory 37, 145–151.

Ma, W., Manjunath, B., 1997. Netra: a toolbox for navigating large image databases. In: Proceedings IEEE International Conference on Image Processing, Santa Barbara, Vol. 1, pp. 568–571.

Manjunath, B., Ma, W., 1996. Texture features for browsing and retrieval of image data. IEEE Trans. Pattern Anal. Machine Intell. 18, 837–842.

Mao, J., Jain, A., 1992. Texture classification and segmentation using multiresolution simultaneous autoregressive models. Pattern Recognit. 25, 173–188.

Ojala, T., Pietikainen, M., Harwood, D., 1996. A comparative study of texture measures with classification based on feature distributions. Pattern Recognit. 29, 51–59.

Ortega, M. et al., 1997. Supporting similarity queries in MARS. In: Proceedings ACM Conference on Multimedia, Seattle, pp. 403–413.

Pentland, A., Picard, R., Sclaroff, S., 1996. Photobook: content-based manipulation of image databases. Int. J. Comput. Vision 18, 233–254.

Picard, R., Kabir, T., Liu, F., 1993. Real-time recognition with the entire Brodatz texture database. In: Proceedings IEEE Conference on Computer Vision and Pattern Recognition, New York, pp. 638–639.

Puzicha, J., Hofmann, T., Buhmann, J., 1997. Non-parametric similarity measures for unsupervised texture segmentation and image retrieval. In: Proceedings IEEE Conference on Computer Vision and Pattern Recognition, San Juan, Puerto Rico, pp. 267–272.

Santini, S., Jain, R., 1999. Similarity measures. IEEE Trans. Pattern Anal. Machine Intell. 21, 871–883.

Smith, J., Chang, S.-F., 1998. Visually searching the web for content. IEEE Multimedia Mag. 4 (3), 12–20.

Vasconcelos, N., Lippman, A., 1999. Probabilistic retrieval: new insights and experimental results. In: Proceedings of IEEE Workshop on Content-Based Access on Image and Video Libraries, Fort Collins, Colorado, pp. 62–66.

Vasconcelos, N., Lippman, A., 2000. A unifying view of image similarity. In: Proceedings International Conference on Pattern Recognition, Barcelona, Spain, Vol. I, pp. 38–41.

VisTex Texture Database, 1995. Maintained by the Vision and Modeling group at the MIT Media Lab. Available at http://www.white.media.mit.edu/vismod/imagery/VisionTexture/vistex.html.

Xu, K., Georgescu, B., Comaniciu, D., Meer, P., 2000. Performance analysis in content-based retrival with textures. In: Proceedings International Conference on Pattern Recognition, Barcelona, Spain, Vol. IV, pp. 275–278.