An Effective Approach Towards Content-Based Image Retrieval

Rokia Missaoui, M. Sarifuddin, and Jean Vaillancourt

Département d'informatique et d'ingénierie,
Université du Québec en Outaouais
C.P. 1250, Succ. B, Gatineau (Qc), Canada, J8X 3X7
{rokia.missaoui, sarifuddin, jean.vaillancourt}@uqo.ca

Abstract. This paper describes a content-based approach to improve image retrieval effectiveness. First, we define two new measures for computing similarity among images based on color histograms, namely the dissimilitude distance $DS*$ and the similarity distance E. The latter is incorporated into the exponentiation part of the Gibbs distribution and into the generalized Dirichlet mixture, while the former is compared to five similarity measures: L_1, L_2 (Euclidean distance), E as well as Gibbs and Dirichlet distributions integrating the similarity measure E. Then, in order to overcome the limitations (and inappropriateness) of some previous information retrieval measures in evaluating the efficiency of an image retrieval process, three variants of a new effectiveness measure are proposed and experimented on an image collection for different similarity distances.

1 Introduction

Content-based image retrieval (CBIR) has emerged as an important area in computer vision, multimedia computing and databases. In order to make image databases easier to explore, we have developed a three-step process and a prototype for image mining and retrieval (see [5]). The main objectives of such a work are: (i) the development of an image feature extraction module, (ii) the design of a data mining tool dedicated to image clustering, classification and association rule generation, and (iii) the design of an image retrieval module which allows the identification of images that are similar to a given image query. In this paper, we limit ourselves to the third objective by describing the mechanisms put together to improve image retrieval effectiveness.

The rest of the paper is organized as follows. Section 2 presents a brief background on image color representation. Section 3 presents a new similarity distance E for image retrieval as well as its integration into two separate distributions, namely Gibbs distribution (more precisely, its exponentiation part) and generalized Dirichlet mixture [1]. A new retrieval measure is defined in Section 4 while details about the experimentation of our solution to improving image retrieval effectiveness is provided in Section 5.

2 Color Spaces

Most image retrieval systems follow the paradigm of representing images using a set of features, such as color, texture, shape and layout. Among these features, color is the most

P. Enser et al. (Eds.): CIVR 2004, LNCS 3115, pp. 335–343, 2004.

frequently used visual property in content-based image retrieval because it is relatively robust, and invariant with respect to image size and orientation.

It is known that the RGB space is not perceptually uniform in the sense that color differences captured by the Euclidean distance, for example, in the three-dimensional RGB space do not correspond to color differences as perceived by humans. The CIE (*Commission Internationale de l'Eclairage*) has then defined two perceptually uniform or approximatively-uniform color spaces $L^*a^*b^*$ and $L^*u^*v^*$. Further, the $L^*C^*H^*$ (Lightness, Chroma, and Hue) and $L^*t^*\theta^*$ (t = Chroma and θ^* = Hue) color spaces have been defined as derivatives of $L^*u^*v^*$ and $L^*a^*b^*$ [4].

In our work we use 3-D $L^*a^*b^*$ and $L^*C^*H^*$ color spaces to represent and extract color properties of images. Any color in $L^*a^*b^*$ space is represented in a cubic coordinate system of axes L^*, a^*, and b^*. The mapping from $L^*a^*b^*$ to $L^*C^*H^*$ can be expressed in terms of polar coordinates with the perceived lightness and the psychometric correlates of chroma and hue angle using the following formula:

$$C^*_{ab} = \sqrt{a^{*2} + b^{*2}} \text{ and } H^*_{ab} = tan^{-1}\left(\frac{b^*}{a^*}\right) \tag{1}$$

To get a good precision with a reasonably fair execution time, we apply Wand's quantization method [8] to 3-D color histogram $L^*C^*H^*$. We divide the hue angle H^* is divided into 17 colors ($k = 0, 2, \cdots, 16$), and each color is then split into 12 chroma ($n = 0, 2, \cdots, 11$) and 15 lightness values ($m = 0, 2, \cdots, 14$). For white and black images, color is split into 15 different lightness values.

Finally, each histogram is divided into 3075[1] color bins and represented by a vector $\mathbf{V} = (V_{0,0,0}, V_{0,0,1}, \cdots, V_{k,m,n}, \cdots, V_{16,14,11})$ where $V_{k,m,n}$ represents a color bin and stands for the percentage of pixels having the color, lightness and chroma in the quantization intervals $k \triangle h$, $m \triangle l$ and $n \triangle c$.

3 Similarity Color Distance

Color-based similarity analysis can be conducted using either color vectors or color histograms and bins. Moreover, it can be conducted using either similitude, dissimilitude or both of them.

In case of similitude analysis, a simple metric distance L_q such as L_1 (city-block, $q = 1$) or L_2 (Euclidean distance, $q = 2$) can be used. However, L_1 and L_2 are not appropriate for the identification of dissimilitude.

$$L_q = \left(\sum_c \left|V^X(c) - V^Y(c)\right|^q\right)^{(1/q)} \tag{2}$$

As opposed to L_1 and L_2 distances which compute the difference between two histograms w.r.t. to color c, a metric called histogram intersection [7] is defined as the common proportion of color c in two histograms.

In this section we define a new similarity distance which takes into account both the dissimilitude and the similitude of two images X and Y with respect to a set of colors. We

[1] $= (17 \times 15 \times 12) + 15$.

first present a dissimilitude distance named DS^*, and then a similarity distance called E. Both distances are semi-metric ones. Finally, we integrate E into two distributions, namely Gibbs random field model and generalized Dirichlet mixture.

3.1 Dissimilitude Distance

We define the dissimilitude distance DS_c^* between two images X and Y with respect to color c as equal to $D_c + L_c$ where D_c is an indicator of a potential absence of color c in one of the images, and L_c is the difference between the proportions of color c in images X and Y. The former helps discard some images that do not share the same set of colors with the query image.

$$L_c = \left|(V^X(c) - V^Y(c))\right| \text{ and } D_c = \begin{cases} L_c \; if \;\; L_c = max(V^X(c), V^Y(c)) \\ 0 \qquad\qquad otherwise. \end{cases} \tag{3}$$

where $V^X(c)$ and $V^Y(c)$ are the proportions of color c in histograms related to images X and Y respectively. It turns out that the dissimilarity distance D_c is equivalent to L_c (e.g., L_1) when color c is present in both histograms.

Based on the $L^*C^*H^*$ color histogram, the color c corresponds to a color bin at coordinates k, m, and n. Using equation 3, the dissimilitude distance of a color bin and a color histogram can be expressed respectively by:

$$DS^*_{k,m,n} = \begin{cases} 2 \times L_{k,m,n} \quad if \;\; L_{k,m,n} = max(V^X_{k,m,n}, V^Y_{k,m,n}) \\ \quad L_{k,m,n} \qquad\qquad otherwise. \end{cases} \tag{4}$$

$$DS^* = \Big(\sum_{k=0}^{16}\sum_{m=0}^{14}\sum_{n=0}^{11}(DS^*_{k,m,n})^q\Big)^{1/q}. \tag{5}$$

Figures 1, 2 and 3 in Section 5 show the image retrieval output when L_1, L_2 and DS^* distances are used respectively. The leftmost top image on each figure represents the *image query* while the *retrieved images* are displayed by the system in a relevance ranking sequence (from top to bottom and left to right). One can see that DS^* leads to a smaller number of false alarms (i.e., number of irrelevant images included in the answer set) than L_1 and L_2.

3.2 Similarity Distance

Similarity between two color histograms (related to images X and Y) with respect to color c can be expressed by R_c as a ratio between the dissimilitude DS_c^* and the similitude S_c as follows:

$$R_c = \frac{DS_c^*}{S_c}, \text{ where } S_c = min\big(V^X(c), V^Y(c)\big) \tag{6}$$

In order to better highlight similitude and dissimilitude between two images w.r.t. to color c, we propose a new similarity distance as a combination of the two components

$DS^*_{k,m,n}$ and $S_{k,m,n}$. The following two formulae express the similarity distance for color bins and color histograms respectively:

$$E_{k,m,n} = \begin{cases} DS^*_{k,m,n}\big(1 + log(R_{k,m,n})\big) & \text{if } S_{k,m,n} > DS^*_{k,m,n} \\ DS^*_{k,m,n} & \text{otherwise} \end{cases} \tag{7}$$

where $S_{k,m,n}$ is given by Equation 11

$$E = \Big(\sum_{k=0}^{16}\sum_{m=0}^{14}\sum_{n=0}^{11}\left(E_{k,m,n}\right)^q\Big)^{1/q} \tag{8}$$

Figure 4 shows that E leads to a better retrieval output than L_1, L_2 and DS^*.

To further improve the effectiveness of the image retrieval procedure, we propose to incorporate the empirically superior distance measure E into two models that have been shown to possess powerful properties in image retrieval system in the past: the Gibbs random fields and Dirichlet mixture.

3.3 Similarity Distance and Gibbs Distribution

Gibbs distributions and Gibbs random fields are very popular in Statistical Physics and have been successfully used in image processing such as image enhancement, texture analysis, and image comparison [6].

A Gibbs random field (GRF) can be thought of as a random coloring of points on a lattice. It is therefore convenient to represent it mathematically as a family F of random variables taking values in a set $\mathcal{S}$ and parameterized through each possible color configuration f on the lattice (which in our case is the two-dimensional support for the images). Usually, as is the case here, such a distribution is defined as follows :

$$P(f) = \frac{exp - \frac{U(f)}{T}}{\sum_{f \in F} exp - \frac{U(f)}{T}} \tag{9}$$

where the denominator is just a normalizing constant, T (a measure of the entropy of the distribution, usually referred to as the temperature) is set to value 1 for simplicity and $U(f)$ is the so-called energy function, which in our case will take the form of a sum over neighboring configurations to f as defined by a prescribed neighborhood system $\mathcal{N}$.

$$U(f) = \sum_{C \in \mathcal{C}_{\mathcal{N}}}\Big(\sum_{c \in \mathcal{C}_C(f)} V_C(c)\Big) \tag{10}$$

where $\mathcal{C}_{\mathcal{N}}$ is the set of clique types generated by the neighborhood system $\mathcal{N}, \mathcal{C}_C(f)$ is the set of instances of the clique type C in the lattice f, and $V_C(.)$ is the potential function associated with clique type C [3].

Using a similar reasoning as before, we can define the similitude measure between a color bin at position k, m, n in the histogram of images X and Y as follows:

$$S_{k,m,n} = max\Big(min\big(V^X_{k,m,n}, V^Y_{k,j,i}\big), \big(min\big(V^X_{k,j,i}, V^Y_{k,m,n}\big)\Big)^{i=(n-1,1,n+1)}_{j=(m-1,1,m+1)} \tag{11}$$

Since each histogram is quantized into k colors and each color is split according to n chroma values and m lightness values, the similitude of a color bin $V^X(k, m, n)$ in the histogram of the query image can be computed with respect to the neighborhood ($\mathcal{N} = 2$) of chroma and lightness of the color bin $V^Y(k, m, n)$ related to a target image.

While dissimilitude is computed using Equation 4, the similarity distance E of color k and lightness m for any value of the chroma n (in the two histograms) is calculated using the following equation:

$$E_{k,m} = U(f) = \begin{cases} e^{-DS^*_{k,m}\left(1+log\left(\frac{DS^*_{k,m}}{S_{k,m}}\right)\right)} & \text{if } S_{k,m} > DS^*_{k,m} > 0 \\ e^{-DS^*_{k,m}} & \text{otherwise} \end{cases} \quad (12)$$

$$U(X,Y) = \sum_{n=0}^{16} \sum_{m=0}^{14} E_{k,m} \quad (13)$$

where $S_{k,m}$ and $DS^*_{k,m}$ are given by:

$$S_{k,m} = \left(min\left(\sum_{n=0}^{11} V^X_{k,m,n}, \sum_{n=0}^{11} V^Y_{k,m,n}\right)^q + \sum_{n=0}^{11} \left(S_{k,m,n}\right)^q\right)^{(1/q)} \quad (14)$$

$$DS^*_{k,m} = \left(\left|\sum_{n=0}^{11} V^X_{k,m,n} - \sum_{n=0}^{11} V^Y_{k,m,n}\right|^q + \sum_{n=0}^{11} \left(DS^*_{k,m,n}\right)^q\right)^{(1/q)} \quad (15)$$

The motivation behind using an average value of chroma in Equation 12 is due to the fact that the variation of chroma does not lead to an abrupt change to a color perception while a lightness variation does.

Equation 12 is used to calculate the similarity between two histograms of images X and Y for a given color k and an identified lightness m, while Equation 13 is used to compute the distance between those histograms.

Using the distribution described by Equation 9 and taking Formula 13 as the energy function together with a default value of 1 for T, we define the probability that an image Y_j $(j = 1, 2, ..., J)$ in the database be similar to a given query image X as:

$$P(X, Y_j) = \frac{exp\left(-\sum_{k=0}^{16} \sum_{m=0}^{11} E^j_{k,m}\right)}{\sum_{j=1}^{J} exp\left(-E^j\right)} \quad (16)$$

where E^j is the similarity distance between the query image X and image Y_j of the database.

Figure 5 illustrates image retrieval using similarity distance as an integration of E into the exponentiation part of Gibbs distribution.

3.4 Similarity Distance and Dirichlet Distribution

Let $\boldsymbol{V} = (V_1, ..., V_I)$ be a vector of positive random color variables and V_i the maximal probability value of the i^{th} color in the two histograms of X and Y (i.e., $V_i = max(V^X_{k,m}, V^Y_{k,m})$ or $V_i = max(V^X_k, V^Y_k)$) with $\sum_{i=1}^{l} V_i < A$ and $0 < V_i < 1$.

Based on the generalized Dirichlet mixture [1], the joint density function is given by:

$$p(V_1, ..., V_I) = \frac{\Gamma\big(|\boldsymbol{\alpha}|\big)}{A^{|\boldsymbol{\alpha}|}\prod_{i=1}^{I}\Gamma\big(\alpha_i\big)}\prod_{i=1}^{I} V_i^{(\alpha_i-1)} \tag{17}$$

Vector $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_I)$ can be perceived as a similarity distance for events governed by V_i, and hence α_i can be instantiated to $E_{k,m}$ (see Equation 12) for $V_i = max(V^X_{k,m}, V^Y_{k,m})$, or to $E_k = \sum_m E_{k,m}$ for $V_i = max(V^X_k, V^Y_k)$, where $V^X_k = \sum_m \sum_n V^X_{k,m,n}$ and $V^Y_k = \sum_m \sum_n V^Y_{k,m,n}$. The parameter A is a constant.

Figure 6 shows that the integration of the similarity distance E into the Dirichlet distribution leads to the best retrieval effectiveness among the six alternatives considered.

4 Image Retrieval Effectiveness

The performance of an image retrieval system may be analyzed according to its accuracy and its efficiency. While the latter is estimated based on execution time and storage requirements, the former corresponds to system effectiveness in retrieving the images that are the most closely similar to the image query.

Indicators such as false alarm, false dismissal, precision and recall are commonly used for retrieval effectiveness computation. However, they do not really reflect the accuracy of the image retrieval system because the ranking of each displayed image is generally not taken into account. The normalized recall measure partially overcomes this limitation.

Faloutsos et *al.* [2] have defined a measure for evaluating the effectiveness of QBIC system. For each image query, the average rank (AVRR) of all relevant retrieved images is computed as well as the ideal average rank of relevant images (IAVRR). The formula assumes that the system returns all the P relevant images which, in the ideal case (IAVRR), occupy the first P positions. This effectiveness measure obviously takes into account the ranking of relevant images. However, it ignores the deviation between the ideal ranking and the actual ranking of a relevant image. For example, if the system returns images in a completely inverse order of the ideal ranking, the following formula returns a perfect effectiveness value (= 1).

$$\text{Eff} = \frac{AVRR}{IAVRR}, \quad \text{where} \quad IAVRR = \sum_{i=1}^{P} \frac{i}{P} \quad \text{and} \quad AVRR = \sum_{i=1}^{P} \frac{r_i}{P} \tag{18}$$

where P is the total number of relevant images, $i = (1, 2, \cdots, P)$ is similarity image ranking by human expert judgement and r_i corresponds to system image ranking (in a decreasing relevance order).

In this section we propose a new effectiveness measure which overcomes the limitations indicated so far. Let P be the total number of relevant images in the image database, R the total number of retrieved images ($R \geq P$) and P_R the accuracy ratio defined either by $P_R = \frac{P}{R}$ or $\frac{1}{1+log(\frac{R}{P})}$, where $0 \leq P_R \leq 1$. We define the (actual) average rank as $AVRR = \sum_{i=1}^{P} \frac{i}{P} + \sum_{i=1}^{P} \frac{|i-r_i|}{P}$ while the ideal average rank $IAVRR$ is kept unchanged [2].

Table 1. Displayed images using L1, L2, DS*, E, E+Exp and E+Dirichlet distances. False alarms are in **bold**.

Distance	Relevant image ranking (P=10)										R
Expert	1	2	3	4	5	6	7	8	9	10	10
L_1	1	2	5	4	3	6	**31**	**11**	**27**	**12**	31
L_2	1	2	**28**	2	10	**20**	**31**	5	**19**	**23**	31
DS^*	1	2	4	6	3	5	**26**	**13**	**23**	**12**	26
E	1	2	4	6	3	5	**24**	**14**	**20**	10	24
$E + Exp$	1	2	4	3	5	7	**12**	10	**17**	**11**	17
$E + Dirichlet$	1	2	4	3	6	5	7	8	**13**	10	13

Table 2. Retrieval effectiveness computation for L1, L2, DS* (q = 1), E, E+Exp and E+Dirichlet distances using seven measures.

Effectiveness method	Effectiveness retrieval of six different metric distances						
	Expert	L_1	L_2	DS^*	E	$E + Exp$	$E + Dirichlet$
Faloutsos	1.0	2.04	2.96	1.89	1.76	1.38	1.09
Kendall	1.0	0.689	0.533	0.667	0.667	0.778	0.867
Salton	1.0	0.995	0.991	0.996	0.997	0.998	0.999
Parkaew	1.0	0.883	0.678	0.865	0.863	0.908	0.927
Eff_{ord}	1.0	0.512	0.364	0.545	0.579	0.743	0.873
Eff_{sys} (a)	1.0	0.167	0.116	0.209	0.241	0.437	0.672
Eff_{sys} (b)	1.0	0.368	0.242	0.385	0.419	0.604	0.781

In the following we define two variants of our effectiveness measure. The first one, called Eff_{ord}, exploits ranking in a more accurate way than in [2] while the second, called Eff_{sys}, improves the former by taking into account the number R of retrieved images needed to display the P relevant ones. The last one can be split into two distinct variants depending on the value given to P_R (see above). The second variant is more appropriate when $R \gg P$ while the first one performs better when R is a small multiple of P.

$$\text{Eff}_{ord} = \frac{\sum_{i=1}^{P} i}{\sum_{i=1}^{P} i + \sum_{i=1}^{P} |i - r_i|} \tag{19}$$

$$\text{Eff}_{sys} = \frac{P}{R} \frac{\sum_{i=1}^{P} i}{\sum_{i=1}^{P} i + \sum_{i=1}^{P} |i - r_i|} \tag{20a}$$

$$\text{Eff}_{sys} = \frac{1}{1 + log(\frac{R}{P})} \frac{\sum_{i=1}^{P} i}{\sum_{i=1}^{P} i + \sum_{i=1}^{P} |i - r_i|} \tag{20b}$$

Our preliminary experiments show that for a given recall, the highest (respectively the lowest) precision occurs for $E + Dirichlet$ (respectively $L2$).

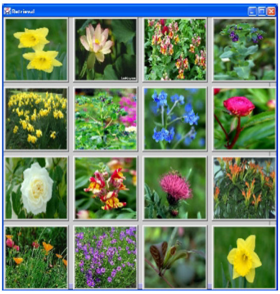

Fig. 1. Image retrieval using L2 distance.

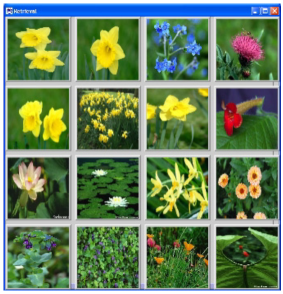

Fig. 2. Image retrieval using L1 distance.

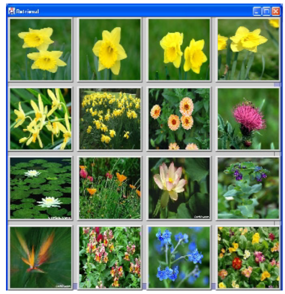

Fig. 3. Image retrieval using DS* distance (q=1).

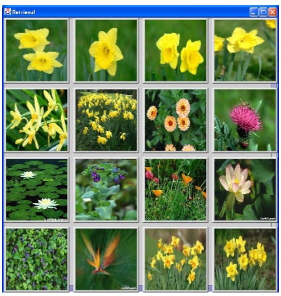

Fig. 4. Image retrieval using E distance (q=1).

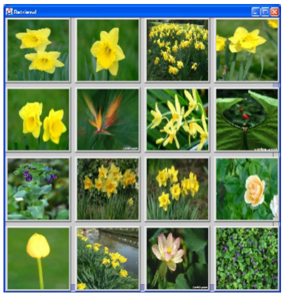

Fig. 5. Image retrieval using E+Exp distance.

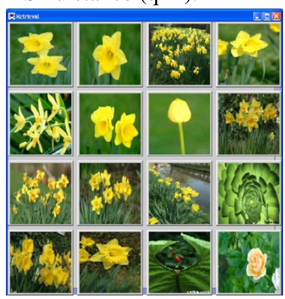

Fig. 6. Image retrieval using E+Dirichlet distance.

5 Empirical Analysis

We have conducted the empirical analysis into two steps: (i) image retrieval using each one of the six distances on a collection of 1069 images, and (ii) effectiveness computation based on the expert's ranking of similar images and using seven effectiveness retrieval measures. For the first step, ten image queries were addressed to the database by four users (students and faculty members) and average execution time was computed. For each similarity measure, the system retrieves the R images needed to display the P (set to 10) relevant images. Some irrelevant images appear in the answer set (false alarms) while some relevant images will be missed (false dismissals).

While the two steps aim at analyzing the retrieval effectiveness of each one of the distances, the second step helps identify the behavior of the newly proposed effectiveness measures, namely Eff_{ord} and Eff_{sys}.

Figures 1 through 6 show the images ranked by the system (from top to bottom and from left to right) when an image query (leftmost top image) is submitted. They clearly show that image retrieval effectiveness is the highest when E is integrated into the generalized Dirichlet mixture and the lowest when the Euclidean distance is used. Indeed, the number of false alarms and false dismissals is the smallest for $E+Dirichlet$ followed by $E+Gibbs$ (exponentiation), followed by E, followed by $DS*$, and so on. The worst similarity ranking is provided by the Euclidean distance.

Table 1 confirms our preceding observations about the performance of $E+Dirichlet$ against the effectiveness of the other similarity measures. The findings remain true in Table 2, except for Faloutsos's measure.

6 Conclusion

In this paper, we have defined two distances: the dissimilitude distance $DS*$ and the similarity distance E and proposed three variants of a new retrieval effectiveness measure. When incorporated into the Gibbs random field and particularly to the generalized Dirichlet mixture, the distance E appears to be a good similarity measure. Empirical analysis of six similarity measures is conducted on color histograms of an image database and shows that retrieval effectiveness is the highest for $E + Dirichlet$ and the lowest for the Euclidean distance.

Our current activities concern the design of new algorithms for color layout extraction (including spatial relationships identification) and image segmentation in order to get a more discriminating power in image retrieval and hence increase our system retrieval effectiveness.

Acknowledgments. We are grateful to VRQ and Canadian Heritage for their financial support.

References

1. B. Bouguila, D. Ziou, and J. Vaillancourt. Maximum likelihood estimation of the generalized dirichlet mixture. *Tech. Rep., Dep. CS, Université de Sherbrooke*, 2002.
2. C. Faloutsos, W. Equitz, M. Flickner, W. Niblack, D. Petkovic, and R. Barber. Efficient and effective querying by image content. *IBM Research Center*, 1994.
3. S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Trans. Pattern Anal. Much. Intell.*, 6, no. 6:721–741, November 1984.
4. B. Hill, Th. Roger, and F.W. Vorhagen. Comparative analysis of the quantization of color spaces on the basis of the cielab color-difference formula. *ACM Trans. on Graphics*, 16:109–154, April 1997.
5. R. Missaoui, M. Sarifuddin, J. Vaillancourt, Y. Hamouda, and H. Laggoune. A framework for image mining and retrieval. In *Proceedings of the SPIE Visual Communications and Image Processing, VCIP-03, Lugano, Switzerland*, pages 430–438, 2003.
6. B. Rémillard and C. Beaudoin. Statistical comparison of images using gibbs random fields. In *Proceedings of Vision Interface'99*, pages 612–617, 1999.
7. M. Swain and D. Baladar. Color indexing. *Computer Vision*, 7, No. 1:11–32, 1991.
8. M.P. Wand. *Data-based choice of histogram bin width*, volume Working paper series of *Australian graduate school of management*. University of New South Wales, Australia, 1996.