# INSTITUTO DE COMPUTAÇÃO
## UNIVERSIDADE ESTADUAL DE CAMPINAS

**Image Retrieval based on discrete distributions of distinctive Color and Scale representative Image Regions (CSIR)**

*Jurandy Almeida*     *Anderson Rocha*
*Ricardo Torres*     *Siome Goldenstein*

Technical Report - IC-07-28 - Relatório Técnico

September - 2007 - Setembro

# Image Retrieval based on discrete distributions of distinctive Color and Scale representative Image Regions (CSIR)*

Jurandy Almeida, Anderson Rocha, Ricardo Torres, and Siome Goldenstein

September 15, 2007

## Abstract

Content-based image retrieval (CBIR) is a challenging task. Common techniques use only low-level features. However, these solutions can lead to the so-called 'semantic gap' problem: images with high feature similarities may be different in terms of user perception. In this paper, our objective is to retrieve images based on color cues which may present some affine transformations. For that, we present CSIR: a new method for comparing images based on discrete distributions of distinctive color and scale image regions. We validate the technique using images with a large range of viewpoints, partial occlusion, changes in illumination, and various domains.

## 1 Introduction

Content-based image retrieval (CBIR) is a challenging task. Common techniques use low-level features and explores local shape and intensity information for viewpoint and occlusion [1]; wavelets and autoregressive models [2]; surface reflection [3]; and Gabor filters [4]. Even fractal transformations can hold interesting resuls [5].

Some CBIR techniques use segmentation as a pre-processing stage. However, experience has demonstrated that segmentation is suited only for narrow domains due to its own difficulty [6, 7]. The most preferred descriptors for retrieval in broad-image domains use color and texture information [6, 8–10]. Some approaches rely on color histograms and color correlograms [11, 12], color coherence vectors [13], and border/interior pixel classification [14]. Sometimes, features such as shape and silhouette [15,16], and moment invariants [17,18] can reduce the 'semantic gap' problem: images with high feature similarities may be different in terms of user perception.

Recent developments have used middle- and high-level information to improve the low-level features. Li et al. [19] have performed architectonics building recognition using color, orientation, and spatial features of line segments. Raghavan et al. [20] have designed a similarity-preserving space transformation method of low-level image space into a high-level vector space to improve retrieval. Some researchers have used bag of features for

1

image categorization [21–23]. Others have used Bayesian approaches to unsupervised one-shot learning of object categories [24]. However, these approaches often require complex learning stages and can not be directly used for image retrieval tasks.

To address these problems, we present a new method to compare images using discrete distributions of distinctive **C**olor and **S**cale representative **I**mage **R**egions (CSIR). The key advantages of this method are: (1) it is robust to viewpoint, occlusion, and illumination changes; (2) it is invariant to image transformations such as rotation and translation; (3) it does not not require any learning stage; and (4) it uses an effective metric to compare images with different number of features. Hence, it does not require a fixed number of features for an image.

To support such statements, and to show that the method can be used in CBIR tasks, we validate the technique using images with a large range of viewpoints, partial occlusion, changes in illumination, and various domains.

## 2 Image descriptors

In this section, we present some low-level feature descriptors widely used in the literature. In Section 3, we compare these descriptors with our technique.

In general, we classify color image descriptors into three categories: (1) global-based; (2) partition-based; and (3) region-based.

1. **Global-based**. It comprises methods that globally describe the color distribution of images. Such methods do not take into account the color spacial distribution. These methods are both time and space computational efficient. GCH (c.f., Sec. 2.1), and BIC (c.f., Sec. 2.2), are examples of such techniques.

2. **Partition-based**. It comprises methods that spatially decompose the image into a fixed number of regions. It region is individually analyzed in order to capture the color spatial distribution. Such methods do not take into account image's visual cues. CCV (c.f., Sec. 2.3), and LCH (c.f., Sec. 2.4) are examples of such approaches.

3. **Region-based**. It comprises methods that use segmentation to decompose images according to visual cues. The number of obtained regions as well as shape, size, and location vary from image to image. The objective is not to find and separate objects in the image but to find similar group of pixels. CBC (c.f., Sec. 2.5) is an example of such techniques.

### 2.1 Global Color Histogram (GCH)

The simplest approach to encode the information present in an image is the Global Color Histogram (GCH) [11]. A GCH is a set of ordered values, one for each distinct color, representing the probability of a pixel being of that color. Uniform quantization and normalization are used to reduce the number of distinct colors and to avoid scaling bias [11]. The $L_1$ (City-block) or $L_2$ (Euclidean) are the most used metrics for histogram comparison.

Histograms are effective for retrieval if there is uniqueness in the color pattern present in the images we want to compare. However, GCH can be sensitive to changes in viewpoint, occlusion, and illumination [7].

## 2.2   Border/Interior Classification (BIC)

Stehling et al. [14] have presented the border/interior pixel classification (BIC), a compact approach to describe images. BIC relies on the RGB color-space uniformly quantized in $4 \times 4 \times 4 = 64$ colors. After the quantization, the image pixels are classified as *border* or *interior*. A pixel is classified as *interior* if its 4-neighbors (top, bottom, left, and right) have the same quantized color. Otherwise, it is classified as *border*.

After the image pixels are classified, two color histograms are computed: one for border pixels and another for interior pixels. The two histograms are stored as single histogram with 128 bins. BIC compares the histograms using the *dLog* distance function [14]

$$dLog(q,d) = \sum_{i=0}^{i<M} \|f(q[i]) - f(d[i])\|$$  (1)

$$f(x) = \begin{cases} 0, & \text{if } x = 0 \\ 1, & \text{if } 0 < x < 1 \\ \lceil \log_2 x \rceil + 1, & \text{otherwise} \end{cases}$$  (2)

where $q$ and $d$ are two histograms with $M$ bins each. The value $q[i]$ represents the $i^{th}$ bin of histogram $q$, and $d[i]$ represents the $i^{th}$ bin of histogram $d$.

## 2.3   Color Coherence Vectors (CCVs)

Zabih et al. [13] have presented an approach to compare images based on color coherence vectors. They define color's coherence as the degree to which pixels of that color are members of large similarly-colored regions. They refer to these significant regions as coherent regions. Coherent pixels are part of some sizable contiguous region, while incoherent pixels are not.

In order to compute the CCVs, first the method blurs and discretizes the image's color-space to eliminate small variations between neighboring pixels. Next, it finds the connected components in the image aiming to classify the pixels within a given color bucket as either coherent or incoherent.

CCV binary classification is based on a non-binary visual property of the images (the size of the connected components) and an empirical size threshold is needed. Hence the most of the useful information about the size of the connected components is lost in this reduction.

## 2.4   Local Color Histogram (LCH)

Tan et al. [25] have presented an approach based on local color histograms (LCH). This technique decomposes the image into equally-sized cells and individually describes each cell using a local color histogram.

The image contents are represented using a local color histogram matrix, one for each cell

$$h_{i,j,k} = \frac{a_{i,j,k}}{n} \tag{3}$$

where $n$ is the number of image's pixels, $a_{i,j}$ is a cell, and $k$ is a quantized color. The LCH distance between two images is the difference of corresponding cell histograms using $L_1$.

## 2.5  Color-based clustering (CBC)

Stehling et al. [26] have presented a region based approach to retrieve images named color-based clustering (CBC). This method decomposes the image into disjoint connected components. Each region presents a minimum size $s_{min}$ and a maximum color dissimilarity $d_{max}$. Each region is defined in terms of the average color in the Lab space ($L$,$a$,$b$), its normalized horizontal and vertical center ($h$,$v$), and its size in pixels normalized with respect to the image's size ($s$).

The $L_2$ Euclidean distance between two regions $a_i$ of image $A$ and $b_j$ of image $B$, is

$$\mathcal{D}(a_i, b_j) = \alpha \times L_2^{color}(a_i, b_j) + (1 - \alpha) \times L_2^{center}(a_i, b_j) \tag{4}$$

where $L_2^{color}(a_i, b_j)$ considers the $L, a, b$ color differences of images $A$ and $B$. $L_2^{color}(a_i, b_j)$ considers the center differences. The parameter $\alpha$ measures which sum component needs to be most valued. The distance $d(A, B)$ of two images is the weighted distance $\mathcal{D}(a_i, b_j) \; \forall a_i, b_j \; \in A, B$. We have used IRM [27] for such computation.

## 3  The CSIR framework

In presence of different viewpoint, occlusion, and illumination in broad-image domains, the direct use of color and texture descriptors can fail.

To address this problem, we present a new method for CBIR based on images' discrete distributions of distinctive local representative features and color properties: CSIR. It is a region-based technique to retrieve images using color visual cues which are robust to pose, orientation, and scale changes. Our framework is based on three key steps: (1) feature region detection; (2) description; and (3) comparison metric; as we illustrate in Algorithm 1.

Our CSIR approach is different from previous literature [28, 29], where the authors describe the images based on histograms of gradient orientation and do not codify color information of the images. In other related work, the authors transform the image to an invariant color-space [30] while here we merge low-level information and local representative features using the image color-space.

### 3.1  Feature region detection

In this stage, we are interested in patterns that can be repeatedly found amongst similar images independent of some affine transformations (e.g., pose, orientation, and scale).

To find such regions, we use a feature region detector or operator [31]. The detectors provide regions which later we use as support regions to compute color descriptors.

---

**Algoritm 1** The CSIR framework.

---

**Require:** Input image $I$;

  1: **Feature region detection:** search for local scale and rotation invariant feature regions $\mathcal{R}$.                                                    ▷ Sec. 3.1

  2: **Description:**                                                                      ▷ Sec. 3.2

        i     **Construct** a separate Gaussian pyramid $\mathcal{G}_R$, $\mathcal{G}_G$, and $\mathcal{G}_B$ for each color channel (R,G,B) of the image $I$.

        ii     **For each** feature region $r \in \mathcal{R}$

                •     **Extract** local scaled and oriented patches $\mathcal{P}$ from the Gaussian color pyramids $\mathcal{G}_R$, $\mathcal{G}_G$, and $\mathcal{G}_B$.

                •     **For each** patch $p \in \mathcal{P}$

                        –     **Calculate** a local low-level color descriptor (e.g., BIC, GCH).

  3: **Comparison metric:** use an appropriate metric to compare the images.     ▷ Sec. 3.3

---

Many different techniques for describing local image regions are available: the rotation invariant Harris points [31]; the rotation and scale invariant Harris-Laplace, Laplace-of-Gaussian [32], Difference-of-Gaussian [28,33], and Hessian-Laplace regions [28,34]. Furthermore, there are the affine transformations invariant Harris-Affine [35], and Hessian-Affine regions [35], among others.

Here, we use the Difference-of-Gaussian (DoG) operator. This idea was first proposed by Crowley and Parker [33]. We search for local extrema in the 3D scale-space representation of an image $I(x, y, \sigma)$ where $\sigma$ is the scale. In this approach, we create a pyramid representation of an image using difference-of-Gaussian filters. We detect a feature point if a local 3D extremum is present and if its absolute value is higher than a threshold.

To build the scale-space pyramid, we successively smooth and sample the input image with a Gaussian kernel. We obtain the DoG representation by subtracting two successive smoothed images. The local 3D extrema in the pyramid representation determine the localization and the scale of the feature regions.

We have used the DoG operator for a number of reasons. First, it is very efficient: we build all DoG levels by using only smoothing and sub-sampling operations. Second, DoG operator provides a close approximation to the scale-normalized Laplace of Gaussian (LoG) regions [32]. This approximation is interesting because Mikolajczyk [34] have showed that the maxima and minima of LoG operator produces the most stable image features compared to a range of other possible operators, such as gradient, Hessian, or Harris corner. However, LoG is more computational intensive than DoG.

We perform the feature region detection in the luminance channel and do not codify

color information. Our experiments have shown that two good choices are the $V$ channel for HSV color representation or $Y$ for YCbCr. In Figure 1, we present the result of this stage for an input image.



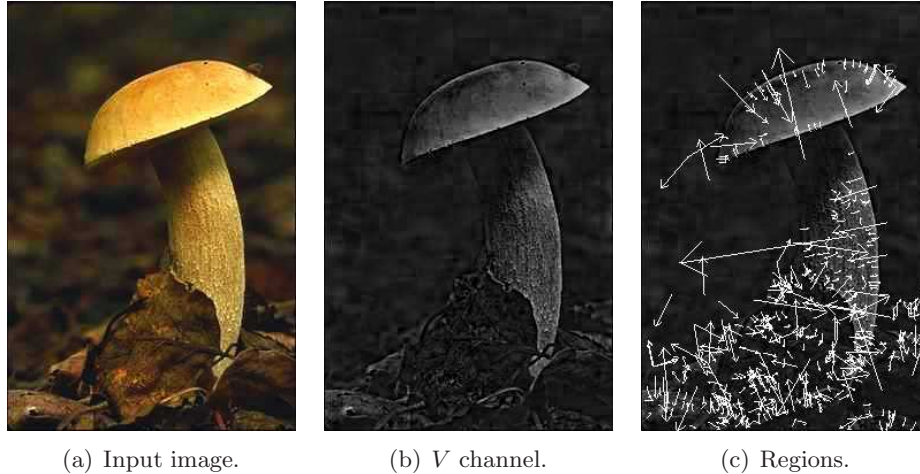(a) Input image.          (b) $V$ channel.          (c) Regions.

Figure 1: Feature regions detection.

Formally, let the scale space of an image be a function $L(x, y, \sigma)$, produced from the convolution of a Gaussian random variable $G(x, y, \sigma)$ with an input image $I(x, y)$

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y), \tag{5}$$

where '*' is the convolution operator in $x$ and $y$, and

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2 + y^2}{2\sigma^2}}. \tag{6}$$

As proposed by Lowe [28], to find good representative and invariant regions in scale-space, we can search for local extrema in the DoG function convolved with the image, $D(x, y, \sigma)$. We can compute this step using two successive scales separated by a constant multiplicative $k$. The constant factor $k$ is required for true scale invariance [28, 32].

$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \\ &= L(x, y, k\sigma) - L(x, y, \sigma). \end{aligned} \tag{7}$$

## 3.2   Description

In this stage, our objective is to describe each persistent region found in Stage 1. It is ideal that each region codifies color information that present repeatable patterns amongst similar images.

Stage 1 provides important information that enables us to retrieve similar images even when such images are slightly modified by some affine transformations. However, this approach still does not consider color information or even capture the color spatial distribution.

In Stage 2, we use the Stage 1 resulting regions to find similar image color cues capturing their distribution and location in the images.

When we use color descriptors, we represent images' color patterns. Given a query image, we are interested in finding similar images and we use repeatable color patterns to find the answers. CSIR approach introduces a new concept. Instead of using the color pattern analysis in the whole image (like previous approaches), we analyze color patterns in representative image's regions. We show, experimentally, that analyzing only these regions, instead of the whole image, it is possible to improve effectiveness in CBIR tasks.

In order to describe the representative image's regions, we construct a separate Gaussian pyramid $\mathcal{G}_R$, $\mathcal{G}_G$, $\mathcal{G}_B$ for each color channel (R,G,B) of the input image $I$. Figure 2 shows the RGB-composed resulting pyramid of an input image.
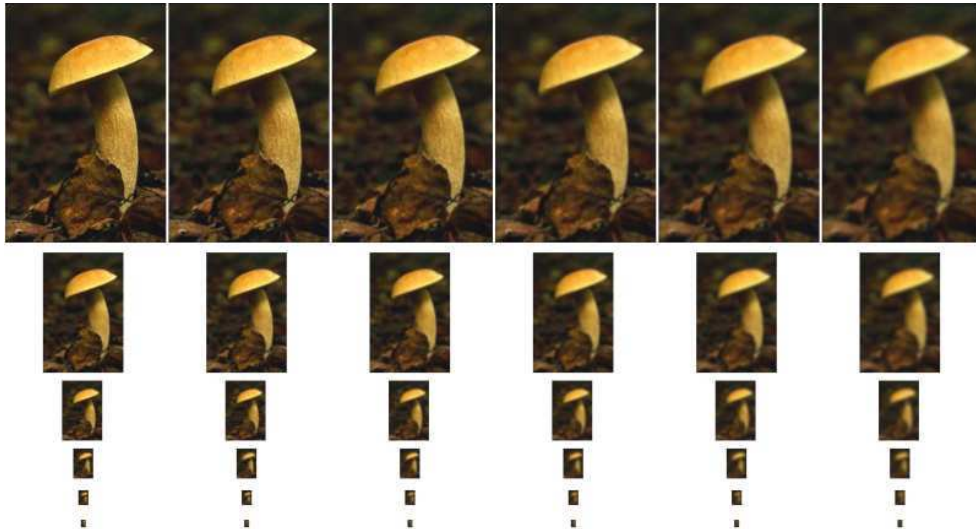


Figure 2: Six-octave RGB-composed resulting pyramid of the image in Figure 1(a). Each octave has 6 scales.

For each region of Stage 1, we extract a local scaled and oriented patch feature region from the Gaussian color pyramid. The patches capture the different illumination, viewpoint, and orientations of the image. Figure 3 shows the resulting patches for the input image in Figure 1(a). There are three patches for each scale, eight patches for each octave, from left to right, top to bottom.

Next, for each extracted patch, we calculate a local low-level image descriptor (e.g., BIC, GCH, CCV) that represents patches' local color information. Figure 4 shows the resulting features for three patches for each scale, eight patches for each octave, from left to right, top to bottom. In this case, we have used BIC to encode the low-level information on each patch. According to the BIC classification, white color represents *border* and black represents *interior*.
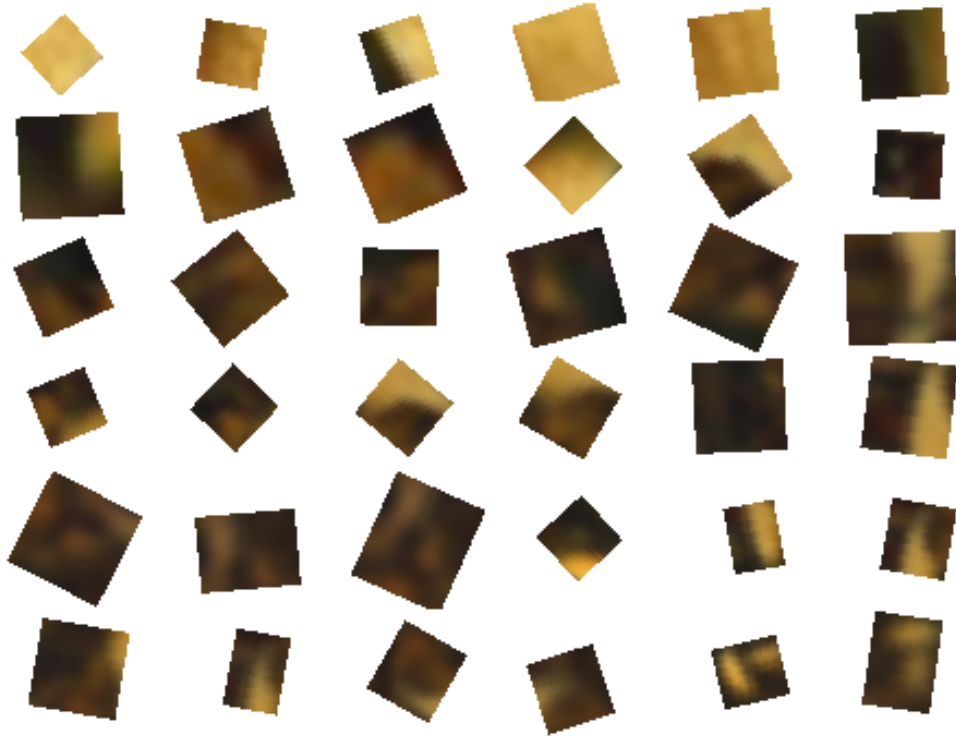
Figure 3: Some resulting patches of the image in Figure 1(a).

## 3.3   Comparison metric

CSIR method provides color and scale information regions that describe an image. The number of feature for each image is different. The more complex an image the more feature regions CSIR provides for it. Hence, we need to compare images with different number of features. For that, we model the image features as hyper points under an unknown distribution. Further, we use the Earth Mover's Distance (EMD) metric to evaluate dissimilarity between two multi-dimensional distributions (image features). The advantage is that EMD "lifts" this distance from individual features to full distributions [36].

Intuitively, given two distributions $\mathbf{B_p}$ and $\mathbf{B_q}$, we can view $\mathbf{B_p}$ as a mass of earth properly spread in space, and $\mathbf{B_q}$ as a collection of holes in that same space. Then, the EMD measures the least amount of work needed to fill the holes with earth. Here, a unit of work corresponds to transporting a unit of earth by a unit of ground distance.

EMD provides a way to compare images based on their discrete distributions of local features. Let $(\mathcal{X}, \mathcal{D})$ be a metric space, $\mathbf{B_p}, \mathbf{B_q} \subset \mathcal{X}$ be two equal-mass sets, and $\pi$ be a a matching between $\mathbf{B_p}$ and $\mathbf{B_q}$. The EMD is the minimum possible cost of $\pi$ and is defined as
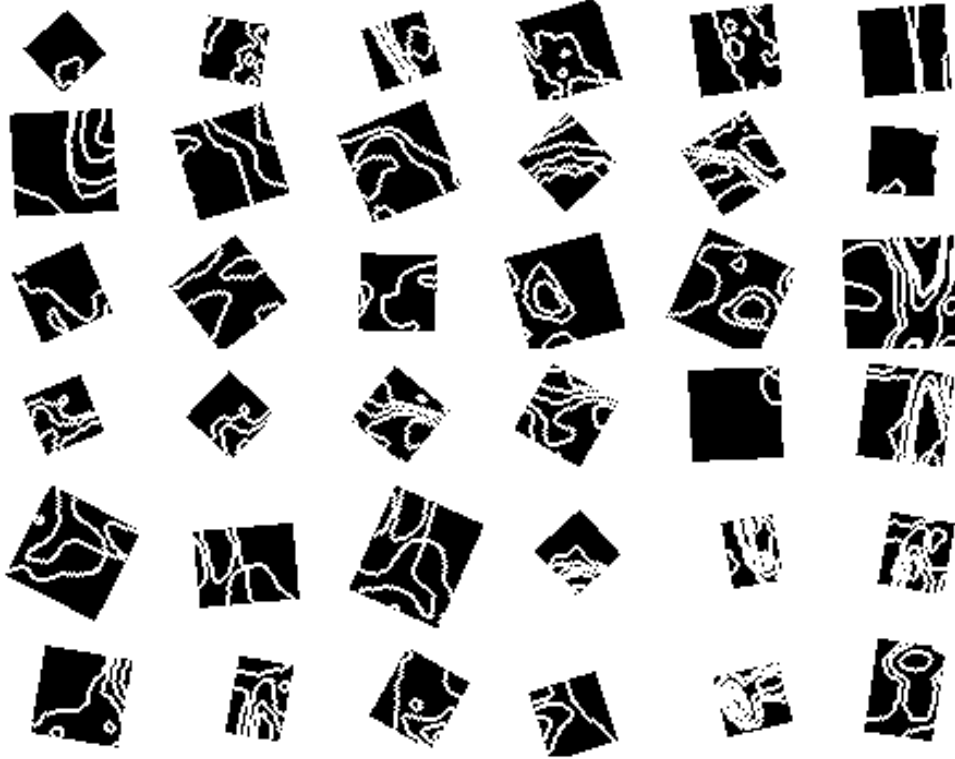
Figure 4: Some resulting low-level information patches of the image in Figure 1(a).

$$EMD(\mathbf{B_p}, \mathbf{B_q}) = \min_{\pi:\mathbf{B_p} \to \mathbf{B_q}} \sum_{s \in \mathbf{B_p}} \mathcal{D}(s, \pi(s)). \tag{8}$$

The computation of $\mathcal{D}$ is based on establishing the correspondence between two images' unordered features. However, the complexity of finding the optimal exact correspondence between two equal-mass sets is cubic in the number of features per set. Hence, we have used a low-distortion EMD embedding [37] to reduce the problem of correspondence between sets of local features to an $L_1$ distance.

## 4    Experiments

In this section, we compare our CSIR approach to the set of image descriptors described in Section 2. For each image, we compute the feature vector using the selected descriptors. We sort the resulting feature vectors using a proper comparison metric.

We show that our approach is more resilient to some affine transformations than previous approaches that use color information on the whole image. Further, we provide results that point out that CSIR is indeed suitable for CBIR tasks.

### 4.1    Methodology

In this work, we have used the *query-by-example* (QBE) paradigm [7]. In QBE, we give an image as a visual example to the system and we query for images that are similar to the given example. Clearly, the effectiveness of these systems is dependent on the properties of the example image.

In order to assess the system effectiveness, we have a database with reference models, a set of images that represent the queries, and a common metric to be used in the effectiveness retrieval assessment. Here, our reference models are equal to the query models, and we test all images in the database against the remaining images, one at a time.

To evaluate the descriptors we present in this paper, we have used two image databases described in the literature. To create a more realistic scenario, we have merged these two databases.

The first database is a selection of the Corel Photo Gallery and is the same as the reported in [14]. This database is highly heterogeneous and comprises images with different domains.

The second database is freely available[1] and comprises images with a common background and different viewpoint, occlusion, and illumination.

As our objective in this paper is to retrieve objects in a database, we have excluded images in both databases that do not represent an explicit object. The resulting combined database we use in the experiments comprises 1,320 broad-image domains spanned into 72 different object classes. Each class contains at least 5 images. We present some examples of the resulting database in Figure 5.

We use the *Precision × Recall* [6,7] metric to assess the retrieval effectiveness. *Precision* is the ratio of the number of relevant images retrieved to the total number of irrelevant and relevant images retrieved. *Recall* is the ratio of the number of relevant images retrieved to the total number of relevant images in the database.

Also, we have used some unique value measurements in the validation. The first one corresponds to the resulting precision when the number of retrieved images is enough to include all relevant images for a given query. This measurement is named *R-value* [38], hence $p_R$ stands for the precision in this point.

We also evaluate the measurements $p_{30}$, $r_{30}$, $p_{100}$, and $r_{100}$. These values are estimatives of the number of retrieved images which a common user would assess in a practical retrieval system [38].

---

[1] `http://www.mis.informatik.tu-darmstadt.de/Research/Projects/categorization/eth80-db.html`

Figure 5: Resulting database. *Boats*, *Rodeo* and *Car* classes.

Finally, we considered the average value for three ($3P$) and eleven ($11P$) points in the precision/recall curve [38]. We obtain the $3P$ value averaging the precision through three predefined recall points (usually 20%, 50% e 80%). We obtain the $11P$ value averaging 11 predefined recall points (usually 0%, 10%, ..., 90%, 100%).

## 4.2 Overall results

Figure 6 shows the results for seven image descriptors [11, 13, 14, 25, 26, 29]. Here, CSIR is represented by two curves: $CSIR_{BIC}$ and $CSIR_{GCH}$ that uses BIC and GCH respectively as the local low-level image descriptor.

Table 1 presents the average result for seven unique value measures we use in this paper: $3P$, $11P$, $p_{30}$, $r_{30}$, $p_{100}$, $r_{100}$ e $p_R$.

Table 1: Unique values measurements results.

|  | $3P$ | $11P$ | $p_{30}$ | $r_{30}$ | $p_{100}$ | $r_{100}$ | $p_R$ |
|---|---|---|---|---|---|---|---|
| **CSIR$_{BIC}$** | **.67** | **.58** | **.53** | **.66** | **.32** | **.90** | **.28** |
| BIC | .49 | .42 | .39 | .52 | .23 | .76 | .17 |
| **CSIR$_{GCH}$** | **.44** | **.38** | **.37** | **.49** | **.21** | **.70** | **.13** |
| CBC | .31 | .27 | .27 | .38 | .16 | .58 | .09 |
| GCH | .27 | .24 | .23 | .34 | .14 | .50 | .10 |
| LCH | .26 | .23 | .23 | .34 | .14 | .51 | .10 |
| CCV | .26 | .23 | .23 | .34 | .14 | .50 | .10 |

In fact, the use of representative regions to better represent color cues in the image does improve the retrieval effectiveness for broad-domain images under different illumination, occlusion, and focus conditions as we see in Figure 6 and Table 1. The greater the value
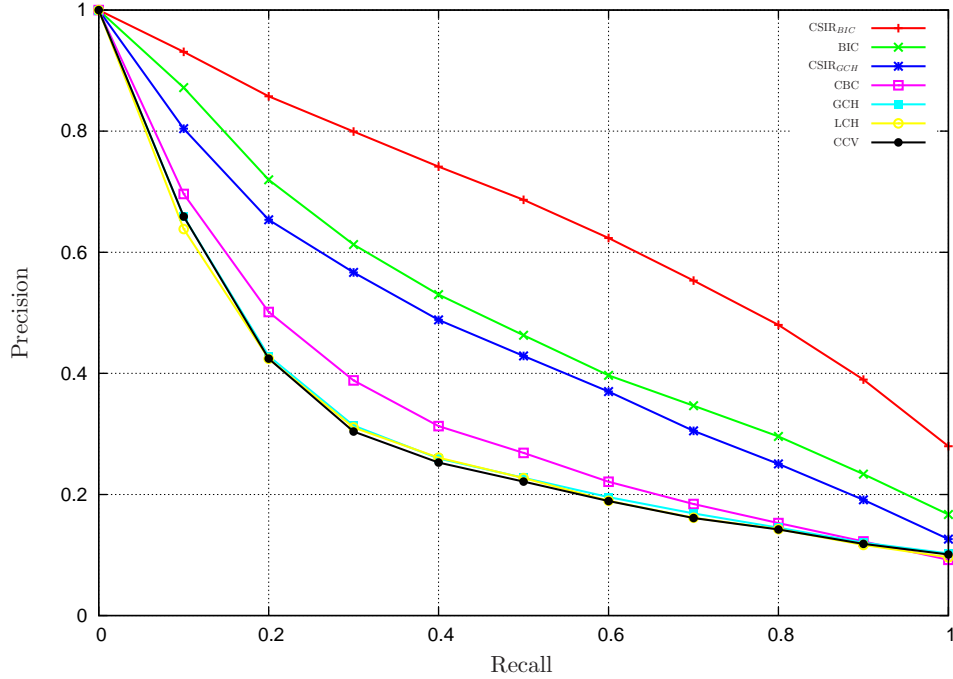
Figure 6: CSIR$_{BIC, GCH}$ *vs.* existing approaches.

the better the descriptor. For instance, the CSIR$_{BIC}$ is $\approx 37\%$ better than traditional BIC and CSIR$_{GCH}$ is $\approx 63\%$ better than GCH.

## 4.3   Visual examples

In this section, we show two resulting queries $Q_1$ and $Q_2$ for CSIR$_{BIC}$ and BIC. We show the query on top left and the resulting retrieved images sorted from left to right, top to bottom.

We show $Q_1$ and its top-11 results in Figures 7(a) and 7(b). The use of discrete distributions of distinctive color and scale image regions of CSIR$_{BIC}$ yields better results than the BIC global analysis. For instance, BIC retrieves the non-relevant image $R_9$.

We show $Q_2$ and its top-11 results in Figures 8(a) and 8(b). CSIR$_{BIC}$ captures the variations in viewpoint, partial occlusion, and illumination. Note that BIC retrieves the non-relevant images $R_2, R_4, R_5, R_7$ and $R_9$.

## 5   Conclusions

In this paper, we have presented CSIR: a new method for comparing images based on their discrete distributions of distinctive color and scale representative regions.

Our method is robust to viewpoint, occlusion, and illumination changes; it is invariant to image transformations such as rotation and translation; and it does not not require any

(a) $\text{CSIR}_{BIC}$ retrieves all images correctly.

(b) BIC retrieves the non-relevant image $R_9$.
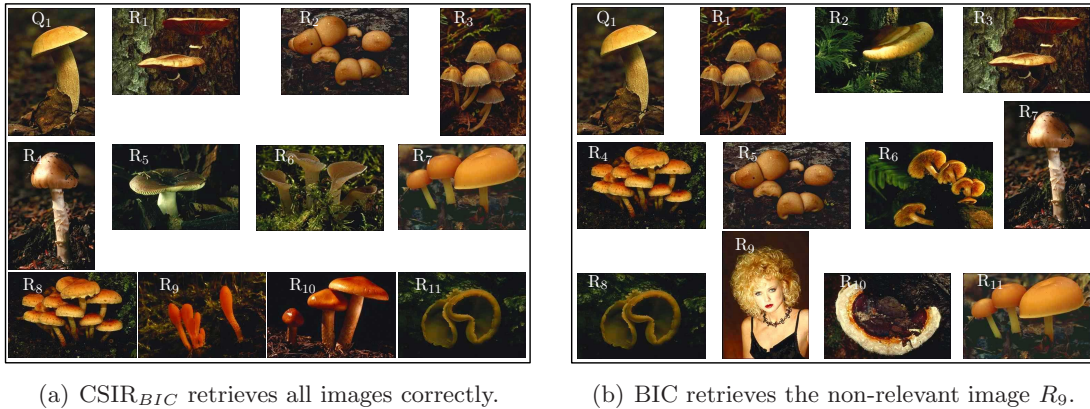
Figure 7: $Q_1$ top-11 results.

learning stage.

Our key contribution is that instead of using the color pattern analysis in the whole image (as previous approaches) we use distinctive color and scale representative patterns that can be repeatedly found amongst similar images independent of some affine transformations.

We have provided experiments showing that CSIR is suitable for CBIR tasks and that it provides good retrieval effectiveness.

Future work includes the evaluation of other feature region operators and low-level image descriptors to improve the image representation.

# 6    Acknowledgments

# References

[1] C. Shimid and R. Mohr, "Local grayvalue invariants for image retrieval," *TPAMI*, vol. 19, no. 5, pp. 530–535, 1997.

[2] P. M. Tardiff and A. Zaccarin, "Multiscale autoregressive image representation for texture segmentation," in *VIII Image Processing*, vol. 3026, 1997, pp. 327–337.

[3] E. Angelopoulou and L. B. Wolff, "Sign of gaussian curvature from curve orientation in photometric space," *TPAMI*, vol. 20, no. 10, pp. 1056–1066, 1998.

[4] X. Fu, Y. Li, R. Harrison, and S. Belkasim, "Content-based image retrieval using gabor-zernike features," in *ICPR*, 2006, pp. 417–420.
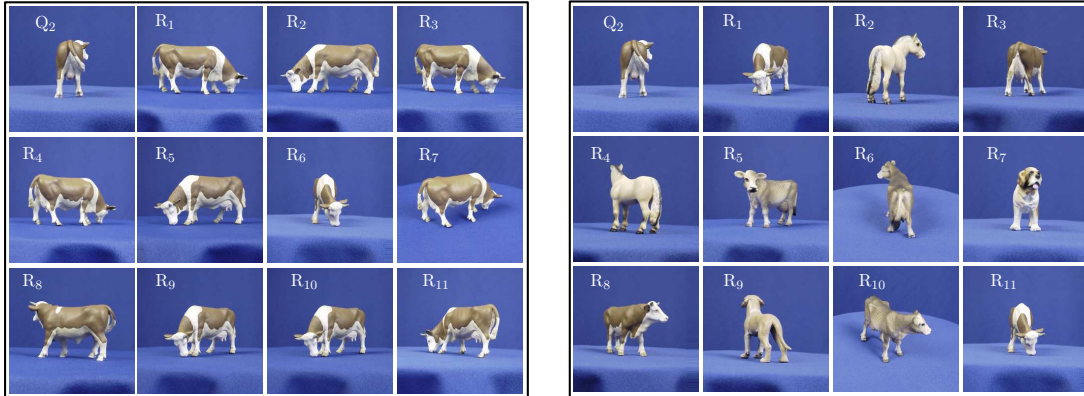
(a) CSIR$_{BIC}$ retrieves all images correctly.

(b) BIC retrieves the non-relevant images $R_2$, $R_4$, $R_5$, $R_7$, and $R_9$.

Figure 8: $Q_2$ top-11 results.

[5] L. M. K. et al., "Fast texture database retrieval using extended fractal features," in *VI Storage and Retrieval for image and video databases*, vol. 3312, 1998, pp. 162–173.

[6] A. W. M. Smeulders, M. Worring, S. Santine, A. Gupta, and R. Jain, "Content-based image retrieval at the end of early years," *TPAMI*, vol. 22, no. 12, pp. 1349–1380, 2000.

[7] R. S. Torres and A. X. Falcão, "Content-based image retrieval: Theory and applications," *Revista de Informática Teórica e Aplicada*, vol. 13, no. 2, pp. 161–185, 2006.

[8] T. Tan, "Rotation invariant texture features and their use in automatic identification," *TPAMI*, vol. 20, no. 7, pp. 751–756, 1998.

[9] D. A. Forsyth and M. M. Fleck, "Automatic detection of human nudes," *IJCV*, vol. 32, no. 1, pp. 63–77, 1999.

[10] M. Mirmehdi and M. Petrou, "Segmentation of color texture," *TPAMI*, vol. 22, no. 2, pp. 142–159, 2000.

[11] M. J. Swain and B. H. Ballard, "Color indexing," *IJCV*, vol. 7, no. 1, pp. 11–32, 1991.

[12] J. Huang, S. R. Kumar, M. Mitra, W.-J. Zhu, and R. Zabih, "Spatial color indexing and applications," *IJCV*, vol. 35, no. 3, pp. 245–268, 1999.

[13] G. Pass, R. Zabih, and J. Miller, "Comparing images using color coherence vectors," in *ACM Multimedia*, 1996, pp. 65–73.

[14] R. O. Stehling, M. A. Nascimento, and A. X. Falcão, "A compact and efficient image retrieval approach based on border/interior pixel classification," in *11th Intl. Conf. on Information and Knowledge Management*, 2002, pp. 102–109.

[15] N. Alajlan, M. S. Kamela, and G. Freeman, "Multi-object image retrieval based on shape and topology," *Signal Processing: Image Communication*, vol. 21, no. 10, pp. 904–918, 2006.

[16] B. Wang and J. A. Bangham, "Shape retrieval using matching pursuit decomposition," in *IEEE AVSS*, 2006, pp. 98–104.

[17] K. Jarrah, M. Kyan, I. Lee, and L. Guan, "Application of image visual characterization and soft feature selection in content-based image retrieval," in *Multimedia Content Analysis, Management, and Retrieval*, vol. 6073, 2006.

[18] C. S. Sastry, A. K. Pujari, and B. L. Deekshatulu, "A fourier-radial descriptor for invariant feature extraction," *Intl. Journal of Wavelets, Multiresolution and Information Processing*, vol. 4, no. 1, pp. 197–212, 2006.

[19] Y. Li and L. G. Shapiro, "Consistent line clusters for building recognition in cbir," in *16th ICPR*, vol. 3, 2002, pp. 30 952–30 957.

[20] B. Shah, V. Raghavan, P. Dhatric, and X. Zhao, "A cluster-based approach for efficient content-based image retrieval using a similarity-preserving space transformation method," *JASIST*, vol. 57, no. 12, pp. 1694–1707, 2006.

[21] M. Marszałek and C. Schmid, "Spatial Weighting for Bag-of-Features," in *CVPR*, 2006, pp. 2118–2125.

[22] J. Sivic and B. Russell and A. Efros and A. Zisserman and and W. Freeman, "Discovering objects and their location in images," in *ICCV*, 2005, pp. 370–377.

[23] K. Grauman and T. Darrell, "Efficient Image Matching with Distributions of Local Invariant Features," in *CVPR*, 2005, pp. 627–634.

[24] L. F. Fei, R. Fergus, and P. Perona, "One-shot learning of object categories," *TPAMI*, vol. 28, no. 4, pp. 594–611, 2006.

[25] H. Lu, B. C. Ooi, and K. Tan, "Efficient image retrieval by color contents," in *Intl. Conf. on Applications of Databases*, vol. 819, June 21–23 1994, pp. 95–108.

[26] R. O. Stehling, M. A. Nascimento, and A. X. Falcão, "An adaptive and efficient clustering-based approach for content-based image retrieval in image databases," in *IEEE Intl. Database Engineering and Applications Symposium*, July 2001, pp. 356–365.

[27] J. Li, J. Z. Wang, and G. Wiederhold, "IRM: Integrated region matching for image retrieval," in *ACM Intl. Conf. on Multimedia*, 2000, pp. 147–156.

[28] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *IJCV*, vol. 60, no. 2, pp. 91–110, 2004.

[29] Y. Ke and R. Sukthankar, "Pca-sift: A more distinctive representation for local image descriptors," in *CVPR*, vol. 02, 2004, pp. 506–513.

[30] A. E. Abdel-Hakim and A. A. Farag, "CSIFT: A SIFT descriptor with color invariant characteristics," *CVPR*, vol. 2, pp. 1978–1983, 2006.

[31] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *TPAMI*, vol. 27, no. 10, pp. 1615–1630, October 2005.

[32] T. Lindeberg, "Scale-space theory: a basic tool for analyzing structures at different scales," *Journal of Applied Statistics*, vol. 21, no. 1, pp. 225–270, 1994.

[33] J. Crowley and A. Parker, "A Representation for Shape Based on Peaks and Ridges in the Difference of Low-pass Transform," *TPAMI*, vol. 6, no. 2, pp. 156–170, 1984.

[34] K. Mikolajczyk and C. Schmid, "Scale & Affine Invariant Interest Point Detectors," *IJCV*, vol. 60, no. 1, pp. 63–86, 2004.

[35] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool, "A comparison of affine region detectors," *IJCV*, vol. 65, no. 1/2, pp. 43–72, 2005.

[36] Y. Rubner, C. Tomasi, and L. J. Guibas, "The earth mover's distance as a metric for image retrieval," *IJCV*, vol. 40, no. 2, pp. 99–121, 2000.

[37] P. Indyk and N. Thaper, "Fast image retrieval via embeddings." in *ICCV*, 2003.

[38] R. A. Baeza-Yates and B. Ribeiro-Neto, *Modern Information Retrieval.* Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1999.