Perceptual Color Descriptor Based on a Spatial Distribution Model: PROXIMITY HISTOGRAMS

Moncef Gabbouj, Murat Birinci and Serkan Kiranyaz
Institute of Signal Processing
Tampere University of Technology
Tampere, FINLAND
Moncef.gabbouj@tut.fi

***Abstract*— Color is the main source of information particularly for content-analysis and retrieval. Most of the color descriptors, however, show severe limitations and drawbacks due to their incapability of modelling the human color perception. Moreover, they cannot characterize all the properties of the color composition in a visual scenery. In this paper we present a perceptual color feature, which describes all major properties of prominent colors both in spatial and color domain. In accordance with the well-known *Gestalt* law, we adopt a top-down approach in order to model (see) the whole color composition before its parts and in this way we can avoid the problems of pixel-based approaches. In color domain the dominant colors are extracted along with their global properties and quad-tree decomposition partitions the image so as to characterize the spatial color distribution (SCD). The proposed color model distils the histogram of inter-color distances. Combination of the extracted global and spatial properties forms the final descriptor, which is neither biased nor become noisy from the presence of such color elements that cannot be perceived in both spatial and color domains. Finally a penalty-trio model fuses all color properties in a similarity distance computation during retrieval. Experimental results approve the superiority of the proposed technique against well-known global and spatial descriptors.**

Keywords-component; Spatial Color Descriptor, Content-based Retrieval, Perceptual Color Model.

I. Introduction

Modeling the true color composition of an image can turn out to be a powerful feature for the purpose of content-based image retrieval (CBIR), if extracted in a perceptually oriented way and semantically intact. Moreover, color structure in a visual scenery is robust to noise, image degradations, changes in size, resolution and orientation. Eventually most of the existing CBIR systems use various color descriptors in order to retrieve relevant images (or generally speaking the visual multimedia material); however their retrieval performance is usually limited especially on large databases due to lack of the discrimination power of such color descriptors. One of the main reasons for this is because most of them are designed based on some heuristics or naïve rules that are not formed with respect to what humans or more specifically human visual system (HVS) finds "relevant" in color similarity. The word "relevance" is described as "the ability (as of an information retrieval system) to retrieve material that satisfies the needs of the user". Therefore, it is of decisive importance that human color perception is respected whilst modeling and describing any color composition of an image.

Accordingly, the study of human color perception and similarity measurement within the color domain become crucial and there is a wealth of research performed in this field. For example in [2], Broek et. al. focused on the utilization of color categorization (or called as *focal* colors) for CBIR purposes and introduce a new color matching method, which takes human cognitive capabilities into account. They have exploited the fact that humans tend to think and perceive colors only in 11 basic categories. In [14], Mojsilovic et. al. performed series of psychophysical experiments analyzing how humans perceive and measure similarity within the domain of color patterns. Their experiments concluded five perceptual criteria (called as "basic color vocabulary") important for comparing the color patterns as well as a set of rules (called as "basic color grammar") governing the use of these criteria in similarity judgment. One observation worth mentioning here is that human eye cannot perceive a large number of colors at the same time, nor able to distinguish similar (close) colors well. Based on this, they show that at the coarsest level of judgment, HVS primarily uses *dominant colors* (i.e. the few colors prominent in the scenery) to judge similarity. Henceforth, the two rules are particularly related for modeling the similarity metrics of the human's color perception. The first one indicates that the two color patterns that have similar dominant colors (DCs) are perceived as similar. The second rule states that two multicolored patterns are perceived as similar if they possess the same (dominant) color distributions regardless of their content, directionality, placement or repetitions of a structural element. In short humans focus on a few DCs and their (spatial) distributions while judging the color similarity between images and our ability to extract such a global color view out of a visual scenery, no matter it is a digital image or a natural 3D view, is indeed amazing. But, it is obvious that humans can neither see individual pixels, nor perceive even a tiny fracture of such a massive amount of color levels, which are present in nowadays digital imaging technology and thus it is crucial to perform certain steps in order to extract the true "perceivable" elements (the true DCs and their global distributions). In other words the un-perceivable elements (as we call them *outliers*), which does not have significant contribution or weight over the present color structure, in both color and spatial (pixel) domain, should be suppressed or removed. Recall that according to two color perception rules presented in [15], two images that are perceived as similar in terms of color composition have a similar DC

This work was supported by the Academy of Finland, project No. 213462 (Finnish Centre of Excellence Program (2006 - 2011)

978-1-4244-3757-3/09/$25.00 ©2009 IEEE

properties; however the color properties of their outliers might be entirely different and hence this can affect (degrade, bias or shift) any similarity measurement if not handled accordingly.

In order to remove the outliers and to secure the global (perceptual) color properties, we adopt a top-down approach both in DC extraction and modeling their global spatial distribution. This approach is in fact phased from the well-known *Gestalt* rule of perception, [26]: "Humans see the whole before its parts", therefore, the method strives to extract what is the (next) global element both in color and spatial domain, which are nothing but the DCs and their spatial distribution within the image. In order to achieve such a (global) spatial representation within an image, starting from the entire image, quad-tree decomposition is applied to the current (parent) block only if it cannot host the majority of a particular DC, otherwise it is kept intact (non-decomposed) representing a single, homogeneous DC presence in it. So this approach tries to capture the "whole" before going through "its parts" and whenever the whole body can be perceived with a single DC, it is kept "as is". Hence the outliers (few alien pixels, which cannot anyway be perceived in a big block of a particular DC) can be suppressed from the spatial distribution and furthermore, the resultant (block-wise) partitioned scheme can be efficiently used for a global modeling and due description of the spatial distribution. After the image is (quad-tree) decomposed, we then represent this global spatial distribution via inter-proximity statistics of the DCs, i.e. the so-called color proximity histograms. Finally a penalty-trio model uses both global and spatial color properties and performs an efficient similarity metric.

The proposed method is fully automatic (i.e. without any supervision, feedback or training involved). Forming the whole process as a Feature eXtraction (*FeX*) module into MUVIS framework [10], allows us to test the mutual performance in the context of multimedia indexing and retrieval.

II. RELATED WORK

There is a wealth of research done and still going on in developing content-based multimedia indexing and retrieval systems such as MUVIS [10], QBIC [7], PicHunter [4], VisualSEEk [22], Virage [25], Image-Rover [20], VideoQ [3], etc. In such frameworks, database primitives are mapped into some high dimensional feature domain, which may consist of several types of descriptors such as visual, aural, etc. From the latitude of low-level descriptors, careful selection of some sets to be used for a particular application may capture the semantics of the database items in a content-based multimedia retrieval (CBMR) system. In this paper, we shall restrict the focus on CBIR domain, which employ only *color* as the descriptor for image retrieval.

In one of the earlier works of color descriptors, Kato et. al. [9] used the color of every corresponding pixel in two images for comparison and the number of corresponding pixels having the same color determines the similarity between them. Recall the HVS fact mentioned earlier about humans inability to see individual pixels or to perceive large amount of color levels and hence this approach did not provide robust solutions, i.e. slight changes in camera position, orientation, noise or lightning conditions may cause significant degradations in the similarity computation. Swain and Ballard [23] proposed the first color histogram, which solves this sensitivity problem. In their work color histograms are extracted and histogram intersection method is utilized for comparing two images (i.e. the query and next image in the database). Since this method is quite simple to implement and gives reasonable results especially in small to medium size databases, several other histogram-based approaches emerged, such as [4], [7], [10], [16], [20], [21] and [24]. The primary feature of such histogram-based color descriptors (be it in RGB, CIE-Lab, CIE-Luv, or HSV) is that they cluster the pixels into fixed color bins, which are quantizing the entire color space using a pre-defined color palette. yet their performance is still quite limited and usually degrades drastically in large databases.

In order to address the drawbacks of such a static quantization scheme, various DC descriptors, e.g. [1], [5], [6], [13] and [15], have been developed using dynamic quantization with respect to image color content. DCs, if extracted properly according to the aforementioned color perception rules, can indeed represent the prominent colors in any image. They have a global representation, which is compact and accurate and they are also computationally efficient since only few colors that are usually present in a natural image and perceivable by a human eye are described. In the current work we implement a top-down DC extraction scheme, similar to one in [5] where the method is entirely designed with respect to HVS color perceptual rules.

One of the most promising approach among all SCD descriptors is the color correlogram [8], which is a table where the k^{th} entry for the color histogram bin pair (i, j) specifies the probability of finding a pixel of color-bin j at a distance k within a maximum range d, from a pixel of color-bin i in an image I with dimensions $W \, x \, H$. Accordingly, Ma and Zhang [11], conducted a comprehensive performance evaluation among several global/spatial color descriptors for CBIR and reported that correlogram achieves the best retrieval performance among the others, such as color histograms, Color Coherence Vector (CCV) [17], color moments, etc. However, the computation complexity is a critical factor for the feasibility of correlogram. The naïve algorithm to compute correlogram takes $O(WHd^2)$, which is a massive computation. The fast algorithm can reduce this to $O(WHd)$; however, requiring $O(WHdm)$ memory space (in bytes) to store them. Apart from such severe feasibility problems, correlogram has several limitations and drawbacks. The first and the foremost is its pixel-based structure, which characterizes the color proximities in pixel level. Such a high-resolution description not only makes it too complicated and infeasible to perform, it also becomes meaningless with respect to HVS color perception rules simply because individual pixels do not mean much for human eye. Furthermore, since correlogram is a pixel level descriptor working over RGB color histogram, the *outliers*, both in color and spatial domains will cause the aforementioned feasibility problems on computational (memory and speed) cost and storage, which makes correlogram inapplicable in many cases or significantly degraded and limited so as to make it feasible again. Finally using the probability alone, makes the descriptor insensitive to the dominance of the color or its area (weight) in the image. This is basically due to the normalization by the amount (weight or area) of color and such an important perceptual cue is lacking in correlogram's description.

978-1-4244-3757-3/09/$25.00 ©2009 IEEE

III. PERCEPTUAL COLOR DESCRIPTOR IN CBIR: PROXIMITY HISTOGRAMS

Under the light of earlier discussion, the proposed color descriptor is designed to address the drawbacks and problems of the color descriptors, particularly the *Correlogram*. In order to achieve this, it is mainly motivated from human color perception rules and therefore, global and spatial color properties are extracted and described in a way HVS perceives them.

A. Formation of the Color Descriptor

As the main outline of the proposed scheme shown in Figure 1, we use a DC extraction algorithm similar to one in [5] where the method is entirely designed with respect to HVS color perceptual rules and configurable with few thresholds, T_S (color similarity), T_A (minimum area), ε_D (minimum distortion) and $N_{DC}^{\max}$ (maximum number of DCs). As the first step, the true number of DCs present in the image (i.e. $1 \le N_{DC} \le N_{DC}^{\max}$) are extracted in CIE-Luv color domain and back-projected to the image for further analysis involving extraction of the spatial properties (SCD) of DCs. Let C_i represents the ith DC class (cluster) with the following members: c_i is the color value (centroid), and w_i is the weight (unit normalized area) Due to the DC thresholds set beforehand, $w_i > T_A, \left|c_i - c_j\right| > T_S \quad for\ 1 \le i, j \le N_{DC}$.

During the back-projection phase, the DC, which has the closest centroid value to a particular pixel color, will be assigned to that pixel. As a natural consequence of this process, spatial outliers, i.e. isolated pixel(s), which are not populated enough to be perceivable, can emerge (e.g. see the example in Figure 1) and should thus be eliminated. Due to the perceptual approach based on the Gestalt rule, "Humans see the whole before its parts", a top down approach such as quad-tree decomposition can process the "whole" first, meaning the largest blocks possible, which can be described (and perceived) by a single DC, before going into its "parts". Due to its top-down structure, it is not degraded from the aforementioned problems that pixel-based approaches usually do and starting from the entire image where DCs are already back-projected it extracts the largest "rectangular" blocks, which makes further analysis easier and more efficient than the arbitrary shape regions. Two parameters are used to configure the quad-tree: T_W, which is the minimum weight (dominance) within the current block required from a DC not to go down for further partition and $D_{QT}^{\max}$, which is the depth limit indicating the maximum amount of partition (decomposition) allowed.

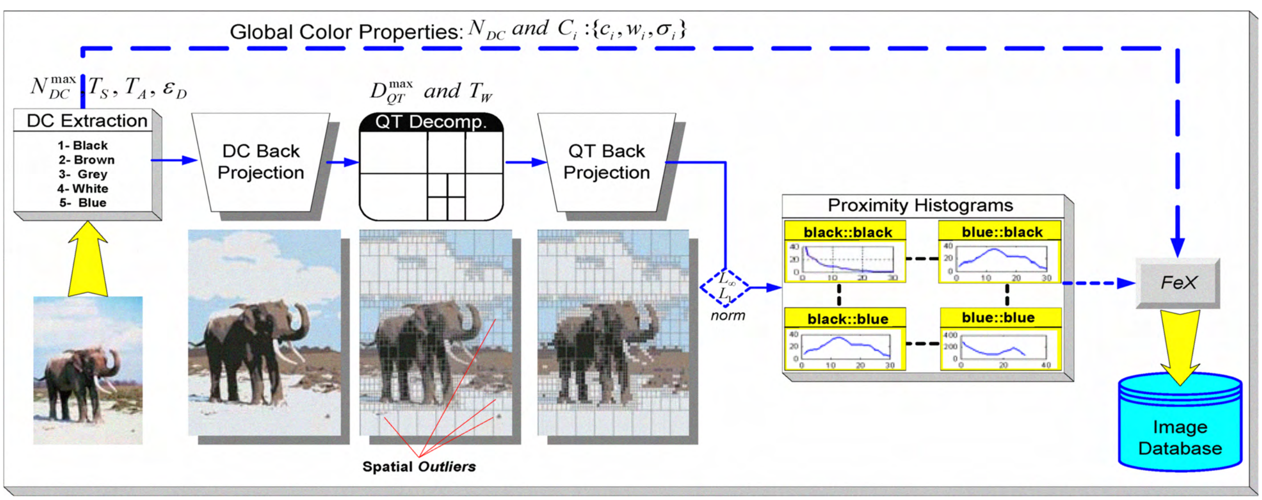

Figure 1: Overview of the proposed color descriptor formation.

T_W can be set in accordance with T_A, i.e. $T_W \cong 1 - T_A$. Therefore, for the typical T_A setting (between 2-5%), T_W can be conveniently set as $T_W \ge 95\%$. Since $D_{QT}^{\max}$ determines when to stop the partitioning abruptly, it should not be set too low not to cause inhomogeneous (mixed) blocks and on the other hand, extensive experimental results suggest that $D_{QT}^{\max} > 6$ is not required even for the most complex scenes since the results are almost identical to one with $D_{QT}^{\max} = 6$. Therefore, the typical range is $4 \le D_{QT}^{\max} \le 6$. Let B^p corresponds to pth partition of the block B where p=0 is the entire block and $1 \le p \le 4$ represents the pth quadrant of the block. The 4 quadrants can be obtained simply by applying equal partitioning to the parent block or via any other partitioning scheme, which is optimized to yield most homogenous blocks possible. For simplicity we use the former case and accordingly a generic QT algorithm, *QuadTree*, can be expressed as follows:

> ***QuadTree*** (*parent*, *depth*)
> - Let $w_{\max}$ be the weight of the DC, which has the maximum coverage in *parent* block.
> - If ($w_{\max} > T_W$) then **Return.**
> - Let $B^0 = parent$
> - For $\forall p \in [1,..,4]$ do:
>   - ***QuadTree*** (B^p , *depth*)
> - **Return**.

The QT decomposition of a (back-projected) image *I* can then be initiated by calling ***QuadTree*** (*I*, 0) and once the process is over, each QT block carries the following data: its depth $D \le D_{QT}^{\max}$, where the partitioning is stopped, its location in the image and the major DC, which has the highest weight in the block (i.e. $w_{\max} > T_W$) and perhaps some other DCs, which are eventually some spatial outliers due to their minority. In order to remove those spatial outliers, a QT back-projection of the major DC into its host block is sufficient. The final scheme where outliers in both color and spatial domains are removed and the (major) DCs are assigned (back-projected) to their blocks can be conveniently used for further (SCD) analysis to extract spatial color features. Note that QT blocks can vary in size depending on the depth, yet even the smallest (highest depth) block is large enough to be perceivable and carry a homogenous DC. So instead of performing pixel-level analysis such as in correlogram, the uniform grid of blocks in the highest depth ($D = D_{QT}^{\max}$) can be used for characterizing the global SCD and extracting the spatial features in an efficient way. Once the QT back-projection of major DCs into their host blocks are completed, all QT blocks hosting a single (major) DC with a certain depth ($D \le D_{QT}^{\max}$) are further partitioned into the blocks in highest depth (i.e. $D = D_{QT}^{\max}$) so as to achieve a proximity histogram in the highest block-wise resolution. Therefore, in such a uniform block-grid, the image I will have NxN blocks where $N = 2^{D_{QT}^{\max}}$ and each of which hosts a single DC. Accordingly the problem of computing inter-DC proximities turns out to be block distances and hence the block indices in each dimension (i.e. $\forall x, y \in [1, N]$) can directly be used for distance (proximity) calculation. Since the number of blocks does not change with respect to image dimension(s), the resolution invariance is, therefore, achieved (e.g. the same image in different resolutions will have identical proximity histograms as opposed to significantly varying *Correlograms* due to its pixel-based computations). As shown in Figure 2, we used L_∞ norm for block-distance calculations. Let $b_1 = (x_1, y_1)$ and $b_2 = (x_2, y_2)$ be two blocks, the distance in L_∞ norm, $L_\infty : \|b_1 - b_2\| = \max(|x_1 - x_2|, |y_1 - y_2|)$. Using the block indices, the block distances become integer numbers and note that for a full range histogram, the maximum (distance) range will be $[1, L]$ where L is $N-1$ in L_∞. A block-wise proximity histogram for a DC pair c_i and c_j stores in its k^{th} bin the number of blocks hosting c_j (i.e. $\forall b_j | I(b_j) = c_j$, equivalent to amount of color c_j in I) from all blocks hosting c_i (i.e. $\forall b_i | I(b_i) = c_i$, equivalent to amount of color c_i in I) in a distance k. So such a histogram clearly indicates how close or far two DCs and their spatial distribution with respect to each other. Yet the histogram bins should be normalized by the total number of blocks, which can be found k blocks away from the source block b_i hosting the DC c_i because this number will significantly vary with respect to the distance (k) and the position of source block (b_i). Therefore, the k^{th} bin of the normalized proximity histogram, $\Phi_{c_i}^{c_j}(k)$, between the DC pair c_i and c_j can be expressed as,

$$\Phi_{c_i}^{c_j}(k) = \sum_{b_i} \sum_{b_j} \Delta(b_i, b_j, k) \qquad \textit{where}$$

$$\Delta(b_i, b_j, k) = \begin{cases} N(b_i, k)^{-1} & \textit{if } b_i \in I(c_i), b_j \in I(c_j), \|b_i - b_j\| = k \\ 0 & \textit{else} \end{cases} \quad (1)$$

k = 1

3	5	5	5	5	5	5	3
5	8	8	8	8	8	8	5
5	8	8	8	8	8	8	5
5	8	8	8	8	8	8	5
5	8	8	8	8	8	8	5
5	8	8	8	8	8	8	5
5	8	8	8	8	8	8	5
3	5	5	5	5	5	5	3

k = 2

5	6	9	9	9	9	6	5
6	7	11	11	11	11	7	6
9	11	16	16	16	16	11	9
9	11	16	16	16	16	11	9
9	11	16	16	16	16	11	9
9	11	16	16	16	16	11	9
6	7	11	11	11	11	7	6
5	6	9	9	9	9	6	5

k = 7

15	8	8	8	8	8	8	15
8	0	0	0	0	0	0	8
8	0	0	0	0	0	0	8
8	0	0	0	0	0	0	8
8	0	0	0	0	0	0	8
8	0	0	0	0	0	0	8
8	0	0	0	0	0	0	8
15	8	8	8	8	8	8	15

Figure 2: $N(b_i, k)$ templates in 8x8 block grid ($D_{QT}^{\max} = 3$) for 3 range values in L_∞ norm.

Note that the normalization factor, $N(b_i, k)$, by the total number of neighbor blocks in distance k, is independent from the DC distribution and hence it can be only computed once and used for all images in the database. Figure 2 presents $N(b_i, k)$ templates computed for all blocks ($\forall b_i \in I$), both norms and some range values. In the figure for illustration purposes N is kept as 8 ($D_{QT}^{\max} = 3$) and note that normalization cannot be applied for those blocks where $N(b_i, k) = 0$ since the range (k) is out of image boundaries and hence $\Phi_{c_i}^{c_j}(k) = 0 \; for \; \forall c_i$. Due to space limitations computational complexity analysis is skipped in this paper.

B. Retrieval via Penalty-Trio Model

In a retrieval operation in MUVIS, a particular feature of the query image, Q, is used for (dis-) similarity measurement with the same feature of a database image, I. Repeating this process for all images in the database, D, and ranking the images according to their similarity distances yield the retrieval result. The proposed color descriptors of Q and I contain both global and spatial color properties. Let C_i^Q and C_j^I represent the i^{th} and j^{th} ($i \le N_{DC}^Q$, $j \le N_{DC}^I$) DC classes where N_{DC}^Q and N_{DC}^I are number of DCs in Q and I respectively. Along with these global properties, the proposed SCD descriptors of Q and I contain proximity histogram ($\Phi_{c_i}^{c_j}(k)$). Henceforth for the similarity distance computation over the proposed color descriptor, both global and spatial color properties are used within a penalty-trio model, which basically penalizes the following properties:

- P_ϕ: The amount of different (mismatching) DCs
- The differences of the matching DCs in:
  - P_G: Global color properties
  - P_{SCD}: SCD properties

So the penalty-trio over all color properties can be expresses as,

$$P_\Sigma(Q, I) = P_\phi(Q, I) + (\alpha P_G(Q, I) + (1 - \alpha) P_{SCD}(Q, I)) \quad (2)$$

where $P_\Sigma \le 1$ is the (unit) normalized total penalty, which corresponds to (total) color similarity distance and $0 < \alpha < 1$ is the weighting factor between global and spatial color properties. Therefore, the proposed penalty-trio model computes a complete distance measure from *all* color properties. Color (DC) matching is the key factor here. We form two sets: matching (S^M) and mismatching (S^ϕ) DC classes from C^Q and C^I by assigning each DC, $c_i \in C_i$, in one set, which cannot match any DC, $c_j \in C_i$, in the other (i.e. $\left|c_i - c_j\right| > T_S \quad for\ \forall i,j$) into S^ϕ and the rest (with at least one match) into S^M. Note that $S^M + S^\phi = C^Q + C^I$ and using the DCs in S^ϕ, P_ϕ can directly be expressed as,

$$P_\phi(Q,I) = \frac{\sum (w_i \ \left| C_i \in S^\phi \right.)}{2} \le 1 \tag{3}$$

The dissimilarity (penalty, P_ϕ) increases proportional with the total amount (weight) of DCs mismatching. In an extreme case where there are no colors matching, $S^M = \{\phi\} \Rightarrow \ P_\Sigma = P_\phi = 1$ makes sense since color-wise two images are nothing in common and hence entirely dissimilar. In another extreme case where all DCs are matching, so $S^\phi = \{\phi\} \Rightarrow P_\phi = 0$ and color (dis-)similarity will only emerge from global (P_G) and spatial (P_{SCD}) color properties of the (matching) DCs.

Typically P_ϕ contributes a certain color distance as a natural consequence of mismatching colors between *Q* and *I*, yet the rest of the distance will occur from the cumulated difference from the properties of matching colors (due to global or spatial properties). Assume without loss of generality that i^{th} DC class in set $C_i^Q : \{c_i^Q, w_i^Q, \sigma_i^Q\} \in S_Q^M$ matches to i^{th} DC in set $C_i^I : \{c_i^I, w_i^I, \sigma_i^I\} \in S_I^M$ (i.e. via sorting one set with respect to other). So the penalties for global and SCD properties can be expressed as,

$$P_G(Q,I) = \beta \sum_{i=1}^{N_M} \left| w_i^Q - w_i^I \right| + (1-\beta) \frac{\sqrt{\sum_{i=1}^{N_M} \left(c_i^Q - c_i^I\right)^2}}{T_d N_M} \le 1$$

$$P_{SCD}(Q,I) = \frac{\sum_{i=1}^{N_M} \sum_{j=1}^{N_M} \sum_{k=1}^{L} \Delta\left(\frac{\Phi_{c_i^Q}^{c_j^Q}(k)}{\max(w_i^Q, w_j^Q)} - \frac{\Phi_{c_i^I}^{c_j^I}(k)}{\max(w_i^I, w_j^I)} \right)}{N_M^2 L} \le 1 \tag{4}$$

$$where \qquad \Delta(x-y) = \left\{ \begin{array}{ll} 0 & if \ x = y = 0 \\ \dfrac{|x-y|}{(x+y)} & else \end{array} \right\}$$

where $0 < \beta < 1$, similar to α, is the weighting factor between the two global color properties: DC weights and centroids. Δ is the normalized difference operator, which emphasizes the difference from zero – nonzero pairs (e.g. =1). This is a common consequence when the DC pairs' area is relatively small but their SCD is quite different. It also suppresses the bias from similar SCDs of two DCs with large weights. Note that P_{SCD} computation should be independent from the effect of DC weights since this is already taken into consideration within P_G computation. As a result the combination of P_G and P_{SCD} represent the amount of dissimilarity occur in the color properties and the unit normalization allows the combination in a configurable way with weights α, β, which can favour one color property to another. With the combination of P_ϕ, the penalty trio models a complete similarity distance between two color compositions..

IV. EXPERIMENTAL RESULTS

The experiments are performed to evaluate the proposed color descriptor efficiency with respect to HVS perceptive criteria (subjective test) and to compare retrieval (via QBE) performances within image databases indexed by the proposed and competing (Correlogram and MPEG-7 DCD [13], [18]) *FeX* modules. In the experiments performed in this section, we used **Corel_10K** Image Database, which contains 10,000 images from *Corel* database bearing diverse contents such as *wild life, city, buses, horses, mountains, beach, food, African natives*, etc. All experiments are carried out on a Pentium-5 1.8 GHz computer with 1024 MB memory. The following parameters are used for all the experiments performed throughout this section: $N_{DC}^{\max} = 6, T_A = 2\%, T_S = 15$ for DC extraction, $T_W = 96\%, D_{QT}^{\max} = 6$ for QT decomposition and $T_C^{\min} = 45,\ T_C^{\min} = T_S,\ \alpha = \beta = 0.5$ for penalty-trio model. For *Correlogram*, we set RGB color histogram quantization as 3x3x3 bins with *d=10.*

TABLE **1: ANMRR scores of the proposed and the competing descriptors on Corel_10K database.**

Methods	ANMRR
Dominant Color	0,458
Auto Correlogram	0,381
Correlogram	0,357
Prox. Histogram	**0.263**

TABLE 1 presents *Average Normalized Modified Retrieval Rank* (ANMRR) scores as the retrieval performance criteria for the proposed and two competing techniques. The query dataset is prepared a priori by regarding a certain degree of color-content coherency, that is, the content similarity can mostly be perceived by color similarity; however a unique, one-to-one correspondence between content and color similarities can never be guaranteed in such natural images due to the presence of other visual cues, such as texture, shape, etc. Nevertheless, according to ANMRR scores presented in the table, **Corel_10K** database the proposed descriptor achieves superior retrieval performance than the competing methods, i.e. *Correlogram*, *Auto-Correlogram* and *MPEG-7 DCD* combined with the quadratic distance computation. Moreover, we observed that in the majority of the queries (78%), the proposed method outperforms (auto-) Correlogram whereas the figure is even higher (92%) with MPEG-7 DCD.

V. CONCLUSIONS

The proposed color descriptor based on proximity color histograms utilizes the perceptual properties of the color composition in a visual scenery in order to maximize the description power. In other words, the so-called *outliers*, which are the un-perceivable color elements, are discarded for description efficiency. Basically a top-down approach is performed during the extraction of the global and spatial color properties. In this way severe problems and critical limitations of traditional pixel based methods are effectively avoided and in spatial domain only the perceived (visible) color components can be truly extracted using QT decomposition. During the retrieval phase, one-to-many color matching is performed in order to apply the penalty-trio model over matching (and possibly fused) DC sets. This greatly reduces the faulty mismatches and erroneous similarity distance computations. Penalty-trio computes the normalized differences in both spatial and global color properties and combines all so as to yield a complete comparison between two color compositions.

The proposed color descriptor has a major advantage of being applicable to any database size and image resolution. Thus it does not suffer from the infeasibility problems and severe limitations of Correlogram. Accordingly, it achieves a significant performance gain on ANMRR scores; however, this is still below our expectations due to two reasons: first and the foremost Correlogram has the advantage of describing the texture in color images thanks to its pixel-level analysis via co-occurrence probabilities. Yet the major reason is that color alone does not provide complete information for content-based retrieval over general, broad-context image databases. It has been shown that color properties correlate with the true content only in a certain extend, but cannot be used as a single cue to characterize the entire content [19]. The proposed descriptor is configurable, efficient, and applicable even to large image databases.

REFERENCES

[1] G. P. Babu, B. M. Mehtre, and M. S. Kankanhalli, "Color Indexing for Efficient Image Retrieval", *Multimedia Tools and Applications*, vol. 1, pp. 327-348, Nov. 1995.

[2] E. L. van den Broek, P. M. F. Kisters, and L. G. Vuurpijl, "The utilization of human color categorization for content-based image retrieval", *in Proc. of Human Vision and Electronic Imaging IX*, pp. 351-362, San José, CA (SPIE, 5292), 2004.

[3] S.F. Chang, W. Chen, J. Meng, H. Sundaram and D. Zhong, "VideoQ: An Automated Content Based Video Search System Using Visual Cues", *In Proc. of ACM Multimedia*, Seattle, 1997.

[4] I. J. Cox, M. L. Miller, S. O. Omohundro, O. N. Yianilos, "PicHunter: Bayesian Relevance Feedback for Image Retrieval", *In Proc. of Int'l Conference on Pattern Recognition*, pp. 361-369, 1996.

[5] Y. Deng, C. Kenney, M. S. Moore, and B. S. Manjunath, "Peer Group Filtering and Perceptual Color Image Quantization", In *Proc. of IEEE Int. Symposium on Circuits and Systems, ISCAS*, vol. 4, pp. 21-24, 1999.

[6] J. Fauqueur and N. Boujemaa, "Region-Based Image Retrieval: Fast Coarse Segmentation and Fine Color Description", in *Proc. of IEEE Int. Conf. on Image Processing (ICIP'2002)*, Rochester, USA, Sep. 2002.

[7] J. Hafner, H. S. Sawhney, W. Esquitz, M. Flickner, W. Niblack, "Efficient Color Histogram Indexing for Quadratic Form Distance Functions", *IEEE Trans. Pattern Analysis and Machine Int.*, vol. 17, pp. 729-736, 1995.

[8] J. Huang; S.R. Kumar, M. Mitra, W.-J. Zhu, R. Zabih, "Image indexing using color correlograms", *in Proc. of Computer Vision and Pattern Recognition*, pp.762-768, 17-19 Jun. 1997.

[9] T. Kato, T. Kurita, and H. Shimogaki, "Intelligent visual interaction with image database systems—toward the multimedia personal interface," *Journal of Information Processing,* vol. 14, No. 2, pp. 134–143, 1991.

[10] MUVIS. [Online] http://muvis.cs.tut.fi/

[11] W. Y. Ma and H. J. Zhang, "Benchmarking of Image Features for Content-based Retrieval", *in Proc. Conf. Signals, Systems and Computers*, pp. 253-257, 1998.

[12] J. MacQueen, "Some Methods for Classification and Analysis of Multivariate Observations", *In Proc. of Fifth Berkeley Symp. On Math. Stat. and Prob*., pp. 281-296, 1967.

[13] B. S. Manjunath, J.-R. Ohm, V. V. Vasudevan, and A. Yamada, "Color and Texture Descriptors", *IEEE Trans. On Circuits and Systems for Video Technology*, vol. 11, pp. 703-715, Jun. 2001.

[14] A. Mojsilovic, J. Kovacevic, J. Hu, R. J. Safranek, K. Ganapathy, "Matching and Retrieval based on the Vocabulary and Grammar of Color Patterns", *in IEEE Trans. on Image Processing*, vol. 9, no. 1, pp. 38-54, Jan. 2000.

[15] A. Mojsilovic, J. Hu and E. Soljanin, "Extraction of Perceptually Important Colors and Similarity Measurement for Image Matching, Retrieval and Analysis", *IEEE Trans. On Image Proc.*, vol. 11, pp. 1238-1248, Nov. 2002.

[16] V. Ogle and M. Stonebraker. Chabot, "Retrieval from a relational database of images", *IEEE Computer*, 28(9), pp. 40–48, 1995.

[17] G. Pass, R. Zabih, and J. Miler, "Comparing Images Using Color Coherence Vectors", *in Proc. of the ACM Multimedia'96*, pp. 65-72, Boston, Nov. 1996.

[18] L.-M. Po and K.-M. Wong, "A New Palette Histogram Similarity Measure for MPEG-7 Dominant Color Descriptor", *In Proc. Int. Conf. on Image Proc., ICIP 2004*, pp. 1533-1536, 2004.

[19] B. Rogowitz, T. Frese, J. Smith, C. A. Bouman, and E. Kalin, "Perceptual Image Similarity Experiments",In *Proc. of SPIE*, vol. 3299, pp.576-590, 1997.

[20] S. Sclaroff, L. Taycher, M. La Cascia, "Image-Rover: A Content-Based Image Browser for the World Wide Web", *In Proc. of IEEE Workshop on Content-based Access Image and Video Libraries*, Puerto Rico, pp. 2-9, Jun. 1997.

[21] J.R. Smith and S. F. Chang, "Single color extraction and image query", *In Proc. of ICIP*, vol. 3, pp. 528-531, Oct. 1995.

[22] J.R. Smith and S. F. Chang, "VisualSEEk: a fully automated content-based image query system", *In Proc. of ACM Multimedia*, Boston, pp. 87-98, Nov. 1996.

[23] M.J. Swain and D.H. Ballard, "Color indexing", *International Journal of Computer Vision*, Vol. 7, No. 1, pp. 11–32, 1991.

[24] A. Vailaya, A. Jain, H. J. Zhang, "On images vs. landscapes", *Pattern Recognition*, vol. 31, pp. 1921-1935, 1998.

[25] Virage. [Online] www.virage.com

[26] M. Wertheimer, "Laws of organization in perceptual forms", Partial translation in W.B. Ellis, editor, "A Sourcebook of Gestalt Psychology", pages 71-88, NY, Harcourt, Brace and Company, 1938.

978-1-4244-3757-3/09/$25.00 ©2009 IEEE