Color-metric Pattern-card Matching for Viewpoint Invariant Image Retrieval

T. Gevers and A.W.M. Smeulders
Department of WINS, University of Amsterdam
Kruislaan 403, 1098 SJ Amsterdam, The Netherlands
email: gevers@wins.uva.nl

Abstract

In this paper, viewpoint independent image retrieval by color-metric pattern-card matching is presented.

First, a photometric color invariant is proposed measuring a local color property of a pixel and its neighboring pixels while discounting the disturbing influences of shading, shadows and highlights. Color-metric pattern-cards are constructed on the basis of the photometric color invariant indicating whether a particular discrete photometric color invariant value is present in an image. To express similarity between color-metric pattern-cards, similarity functions are proposed and evaluated on a database of 500 images taken from 2-D and 3-D colored man-made objects in real world 3-D scenes.

The experimental results show that high image retrieval accuracy is achieved by two distinct similarity functions depending on the presence of object clutter in the scene. Furthermore, image retrieval by color-metric pattern-card matching is to a large degree robust to partial occlusion and a change in viewing position. Good run-time performance of the pattern-card matching process is achieved allowing for fast image retrieval by example image.

1 Introduction

Image retrieval by image example has been proposed recently, where a query image is given by the user on input, for example [3], [8], [9]. For such a case, to actually do image retrieval, the problem is to identify a query image as (a part of) target images in the image database. A typical application example is the problem of retrieving images containing an instance of a particular object. To that end, the query is specified by an example picture image taken from the object at hand.

Swain and Ballard [10] proposed a simple and very effective scheme to retrieve images entirely on the basis of color. Matching is expressed by histogram intersection. If the color distributions of query and target image are more or less the same, the matching rate is high. Swain's scheme has been extended by [2] to become illumination independent by indexing on an illumination-invariant set of color descriptors. These color-based schemes fail, however, when images are heavily contaminated by shadows, shading and highlighting cues.

In this paper, a photometric color invariant is proposed on which the color-metric histogram will be based. The photometric color invariant is defined as a function measuring some *local color* property at each image location while discounting the disturbing influences of shading, shadowing and highlighting cues. To make the histogram independent of the size of the object in view, we use a pattern-card (i.e. a thresholded color-metric histogram) indicating whether a particular discrete color invariant value is present in an image. Then, the problem of image retrieval is reduced to the problem to what extent pattern-card A derived from the query image is similar to pattern-card B^{I_k} constructed for each image I_k in the image database.

In this contribution, the aim is to arrive at fast and accurate pattern-card similarity functions $\mathcal{D}(A, B^{I_k})$ returning a numerical measure of similarity between A and B^{I_k}. To be useful for the purpose of image retrieval, pattern-card matching should satisfy the following criteria: 1. High discrimination power to allow for accurate image retrieval. 2. Robust to shadows, shading and specularities (highlights). 3. Robust to a change in viewpoint of the camera. 4. Robust to object size, occlusion and clutter in the scene. 5. Stable under substantial sensing and measurement error. 6. Easy to compute to allow for fast image retrieval.

The paper is organized as follows. In Section 2, photometric color invariant pattern-card matching is proposed for the purpose of image retrieval. In Section 3, various similarity functions are presented and their performance is evaluated in Section 4.

2 Image Retrieval by Photometric Color Invariant Pattern-Card Matching

In this section, a photometric color invariant is discussed first on which the color-metric pattern-card will be based. Then, image retrieval by color-metric pattern-card matching is discussed.

2.1 Viewpoint independent color model

In this paper, the reflection from composite materials is approximated by the sum of a body reflection component L_C and a surface reflection component S_C. This is the Torrance-Sparrow (T-S) reflection model:

$$M_C = L_C + S_C \tag{1}$$

where $C \in \{R, G, B\}$.

In this paper, body (diffuse) reflection L_C is assumed Lambertian [7]:

$$L_C = I_C k_C (\vec{n} \cdot \vec{s}) \tag{2}$$

where I_C is the intensity of a point light source and k_C is the diffuse reflection coefficient of the surface. The surface normal is denoted by $\vec{n}$ and illumination direction by $\vec{s}$. Lambertian reflection models dull, matte surfaces.

We use the modified Torrance-Sparrow specular reflection term [5]:

$$S_C = (\frac{FPG}{\vec{n} \cdot \vec{v}}) I_C \tag{3}$$

where $\vec{v}$ is a unit vector in the direction of the viewer and F is the Fresnel coefficient describing the percentage of the light reflected at the interface. Because the Fresnel coefficient is weakly dependent on incident angle and wavelength we will assume that it is a constant for a given material. P models the surface roughness and G is the geometric attenuation factor. Under the given conditions, the color of highlights is not related to the color of the surface on which they appear, but only on the color of the light source. In this paper, we assume that the illumination color is white (i.e. $I_R = I_G = I_B$).

Generally, hue is thought of as the angle between the diagonal grey-axis and a color point in the RGB-color space. The angle between the diagonal grey-axis and the color point in RGB is mathematically specified by the standard well-known transformation from the R, B and G values:

$$H(R, G, B) = \arctan(\frac{\sqrt{3}(G - B)}{(R - G) + (R - B)}) \tag{4}$$

By filling in $L_C + S_C$ in the hue equation, we can see that all possible pixels of the same shiny homogeneously colored surface have to be of the same hue:

$$H(L_R + S_R, L_G + S_G, L_B + S_B) = \arctan(\frac{\sqrt{3}(k_G - k_B)}{(k_R - k_G) + (k_R - k_B)}) \tag{5}$$

where illumination and geometric components are factored out resulting in an expression of (normalized) surface albedos proving the following lemma:

Lemma 1 *Assuming T-S reflection and white illumination, H is independent of the viewpoint, surface orientation, illumination direction, illumination intensity, and highlights.*

Therefore, the hue H is taken as a local quantitative, viewpoint-independent color feature. In the following, $H(x, y)$ denotes the hue value at image coordinates (x, y) ranging from $[0, 2\pi)$ computed by equation 4 from the original R, G and B values provided by the camera.

2.2 Photometric color invariant pattern-card construction

A simple photometric color invariant can be defined as a function which measures the hue at image coordinate (x, y) [4]. However, to incorporate some local spatial color information into the pattern-card, we consider the pair of hue values at either side of a significant hue edge.

To cope with the wrap-around nature of hue, the hue edge detector proposed in [4] is used to obtain local maximum hue gradient values $\mathcal{M}$, which will be zero except near locations where two regions of different hue (color) meet. Then, for each local maximum, two neighboring points are computed based on the direction of the gradient to determine the hue value on the both sides of the edge:

$$\vec{p}(\vec{x}) = (H(\vec{x} - \Delta\vec{n}_g), H(\vec{x} + \Delta\vec{n}_g)) \tag{6}$$

only computed for $\vec{x} \in \mathcal{M}$ positive that is only at the two sides of a maximum. Furthermore, $\vec{n}_g$ is the normal to the intensity gradient at $\vec{x}$ and Δ is a preset value (e.g $\Delta = 3$ pixels).

A 2D hue-hue histogram is constructed in the standard way on two hue axes by counting the number of times a hue-hue edge pair is present in an image:

$$\mathcal{T}(i, j) \hat{=} \eta(\vec{p}(\vec{x}) == (i, j)) \tag{7}$$

computed for $\vec{x} \in \mathcal{M}$ where η indicates the number of times $\vec{p}(\vec{x})$ equals the value of the indices (i, j). In the following, index (i, j) is called a hixel (i.e. histogram element).

Because the length of the hue-hue edge depends on the scale of the recorded object (e.g. distance object-camera), we define the photometric color invariant pattern-card, as a thresholded 2-D (hue-hue) histogram, by:

$$\mathcal{C} = \{(i, j) \in \mathcal{T} : \mathcal{T}(i, j) > t_b\} \tag{8}$$

indicating the set of hixels representing hue-hue edges to be present in the image regardless of the length of the hue-hue edge. t_b is a predefined threshold based on the noise level.

In the following, hue-hue edge pattern-cards are used for image retrieval.

2.3 Viewpoint independent image retrieval

Let the image database consist of a set $\{I_k\}_{k=1}^{N_b}$ of N_b color images. For each I_k, $\mathcal{C}^{I_k}$ is created to represent the presence of specific hue-hue edges in I_k. $\mathcal{C}^{\mathcal{Q}}$ is created in a similar way from the query image.

Then image retrieval is as follows. For each $\mathcal{C}^{I_k}$, $k = 1, ..., N_b$, similarity function $\mathcal{D}$ compares $\mathcal{C}^{I_k}$ with $\mathcal{C}^{\mathcal{Q}}$ to return a numerical measure $\mathcal{D}(\mathcal{C}^{\mathcal{Q}}, \mathcal{C}^{I_k})$ of similarity. Then, images are ordered according to $\mathcal{D}$ and displayed for viewing.

In the next section, several alternatives for similarity function $\mathcal{D}$ are defined.

3 Similarity Function $\mathcal{D}()$

Let the set minus operator be given by $\mathcal{C}^{\mathcal{Q}} \setminus \mathcal{C}^{I_k} = \{(i,j) \in \mathcal{C}^{\mathcal{Q}} : (i,j) \notin \mathcal{C}^{I_k}\}$. Similarity functions for object recognition can be characterized by two types of errors: false positives $\mathcal{C}^{\mathcal{Q}} \setminus \mathcal{C}^{I_k}$ and false negatives $\mathcal{C}^{I_k} \setminus \mathcal{C}^{\mathcal{Q}}$.

Similarity functions based on statistical and geometric attributes are presented in Section 3.1 and 3.2 respectively.

3.1 Statistic-based similarity functions

The first similarity function is defined on the absolute difference between the number of hixels in $\mathcal{C}^{\mathcal{Q}}$ and $\mathcal{C}^{I_k}$:

$$\mathcal{D}_m(\mathcal{C}^{\mathcal{Q}}, \mathcal{C}^{I_k}) = |\eta(\mathcal{C}^{\mathcal{Q}}) - \eta(\mathcal{C}^{I_k})| \tag{9}$$

where η denotes the number of hixels.

A similarity function expressing the number of corresponding hixels (logical AND) is given by:

$$\mathcal{D}_{and}(\mathcal{C}^{\mathcal{Q}}, \mathcal{C}^{I_k}) = \eta(\mathcal{C}^{\mathcal{Q}} \cap \mathcal{C}^{I_k}) \tag{10}$$

Similarity function $\mathcal{D}_{and}$ is sensitive to false positives but not to false negatives. Note that $\mathcal{D}_{and}$ can be interpreted as the binary version of histogram intersection proposed by Swain and Ballard [10].

The overall misclassification error (logical XOR) is given by:

$$\mathcal{D}_{xor}(\mathcal{C}^{\mathcal{Q}}, \mathcal{C}^{I_k}) = \eta((\mathcal{C}^{\mathcal{Q}} \setminus \mathcal{C}^{I_k}) \cup (\mathcal{C}^{I_k} \setminus \mathcal{C}^{\mathcal{Q}})) \tag{11}$$

$\mathcal{D}_{xor}$ is sensitive to both false positives and false negatives. An important property of $\mathcal{D}_{xor}$ is that it is a metric [1].

In the next section, similarity functions are discussed which take into account the distance of each misclassified hixel to the nearest correct set.

3.2 Geometry-based similarity functions

Because a hixel represents a hue-hue edge, the distance between hue values is to be defined first. Unlike intensity, hue is defined on a ring ranging $[0, 2\pi)$ instead of a linear interval. We define the difference between two hue values h_1 and h_2 as follows:

$$\mathcal{E}(h_1, h_2) = \sqrt{(\cos h_1 - \cos h_2)^2 + (\sin h_1 - \sin h_2)^2} \tag{12}$$

yielding a difference $d(h_1, h_2) \in [0, 2]$ between h_1 and h_2. The difference is a distance because it satisfies the metric criteria.

Then, the Euclidean distance between two hue-hue edges (h_{1_1}, h_{1_2}) and (h_{2_3}, h_{2_4}) is defined by:

$$\rho((h_{1_1}, h_{1_2}), (h_{2_3}, h_{2_4})) = \sqrt{\mathcal{E}(h_{1_1}, h_{2_3})^2 + \mathcal{E}(h_{1_2}, h_{2_4})^2} \tag{13}$$

where $h_{1_1}, h_{1_2}, h_{2_3}, h_{2_4} \in [0, 2\pi)$.

Let $d(\vec{a}, \mathcal{C}^{I_k})$ denote the shortest distance from $\vec{a} \in \mathcal{C}^{\mathcal{Q}}$ to $\mathcal{C}^{I_k}$:

$$d(\vec{a}, \mathcal{C}^{I_k}) = \min_{\vec{b} \in \mathcal{C}^{I_k}} \{\rho(\vec{a}, \vec{b}) : \vec{a} \in \mathcal{C}^{\mathcal{Q}}\} \tag{14}$$

Then, the geometry-based similarity functions used to compute similarity are as follows.

The first one is the root mean square similarity function mathematically specified as:

$$\mathcal{D}_{rms}(\mathcal{C}^{\mathcal{Q}}, \mathcal{C}^{I_k}) = \sqrt{\frac{1}{\eta(\mathcal{C}^{\mathcal{Q}})} \sum_{\vec{a} \in \mathcal{C}^{\mathcal{Q}}} d(\vec{a}, \mathcal{C}^{I_k})^2} \tag{15}$$

The second one is the maximum hixel difference:

$$\mathcal{D}_{max}(\mathcal{C}^{\mathcal{Q}}, \mathcal{C}^{I_k}) = \max_{\vec{a} \in \mathcal{C}^{\mathcal{Q}}} \{d(\vec{a}, \mathcal{C}^{I_k})\} \tag{16}$$

Furthermore, the Hausdorff distance is defined by:

$$\mathcal{D}_{haus}(\mathcal{C}^{\mathcal{Q}}, \mathcal{C}^{I_k}) = \max\{\max_{\vec{a} \in \mathcal{C}^{\mathcal{Q}}} d(\vec{a}, \mathcal{C}^{I_k}), \max_{\vec{b} \in \mathcal{C}^{I_k}} d(\vec{b}, \mathcal{C}^{\mathcal{Q}})\} \tag{17}$$

An important property of $\mathcal{D}_{haus}()$ is that it is a metric [1]. Except for $\mathcal{D}_{haus}()$, the above defined similarity functions are all sensitive to false positives and insensitive to false negatives.

In the next section, the various similarity functions are evaluated in practice.

4 Experiments

To evaluate the various similarity functions, the criteria 1-6 of Section 1 will be addressed.

The data sets on which the experiments will be conducted are described in Section 4.1. The error measures and pattern-card formation are given in 4.2 and 4.3 respectively.

4.1 Datasets

The database consists of $N_1 = 500$ images of 2-D and 3-D domestic objects, tools, toys, food cans, art artifacts etc., all taken from two households. Objects were recorded in isolation (one per image) with the aid of the SONY XC-003P CCD color camera (3 chips) and the Matrox Magic

Color frame grabber. The digitization was done in 8 bits per color. Objects were recorded against a white cardboard background. Two light sources of average day-light color are used to illuminate the objects in the scene. Objects were recorded at a pace of a few shots a minute. There was no attempt to individually control focus or illumination. Images show a considerable amount of noise, shadows, shading and highlights. As a result, recordings are best characterized as snap shot quality, a good representation of views from everyday life as it appears in home video, the news, and consumer photography in general.

A second, independent set (the query set) of recordings was made of randomly chosen objects already in the database. These objects, $N_2 = 70$ in number, were recorded again (one per image) with a new, arbitrary position and orientation with respect to the camera (some recorded upside down, some rotated, some at different scale).

In the experiments, all pixels in a color image are discarded with a local saturation smaller then 5 percent of the total range (this number was empirically determined by visual inspection); otherwise calculation of hue becomes unstable [6]. Consequently, the white cardboard background as well as the grey, white, dark or nearly colorless parts of objects as recorded in the color image will not be considered in the matching process.

4.2 Error measures

For a measure of match quality, let rank r^{Q_i} denote the position of the correct match for query image Q_i, $i = 1, ..., N_2$, in the ordered list of N_1 match values. The rank r^{Q_i} ranges from $r = 1$ from a perfect match to $r = N_1$ for the worst possible match.

Then, for one experiment, the average ranking percentile is defined by:

$$\overline{r} = \left(\frac{1}{N_2}\sum_{i=1}^{N_2}\frac{N_1 - r^{Q_i}}{N_1 - 1}\right)100\% \qquad (18)$$

The cumulative percentile of query images producing a rank smaller or equal to j is defined as:

$$\mathcal{X}(j) = \left(\frac{1}{N_2}\sum_{k=1}^{j}\eta(r^{Q_i} == k)\right)100\% \qquad (19)$$

where η reads as the number of query images having rank k.

4.3 Pattern-card formation

We determined the appropriate bin size for our application empirically. This has been achieved by varying the same number of bins on both hue axes over $q \in \{2, 4, 8, 16, 32, 64, 128, 256\}$ and chose the smallest q for which the number of bins is kept small for computational efficiency and large for retrieval accuracy. The results showed (not presented here) that the number of bins was of little influence on the retrieval accuracy when the number of bins ranges from $q = 32$ to $q = 256$. Therefore, the pattern-card bin size used during histogram formation is $q = 32$.

Also an appropriate value for t_b is to be determined during the construction of the pattern-card $\mathcal{C}$ as defined in Section 2.2, where we considered the total accumulation for a particular hixel $(i, j) \in \mathcal{T}$ to be due to noise, when the total accumulation is below t_b. Noise will introduce false positive and false negative errors affecting the performance of the similarity functions. Because noise is application dependent, we determined the appropriate value of t_b for our application by varying t_b with $q = 32$ over $t_b \in \{1, 2, 3, 4, 6, 8\}$ and have chosen $t_b = 3$ which produced the highest discriminative power averaged over all similarity functions.

For each query and target image, hue-hue edge pattern-card $\mathcal{C}^{\mathcal{Q}}$ and $\mathcal{C}^{I_k}$ have been created for $q = 32$ and $t_b = 3$.

4.4 Discriminative power

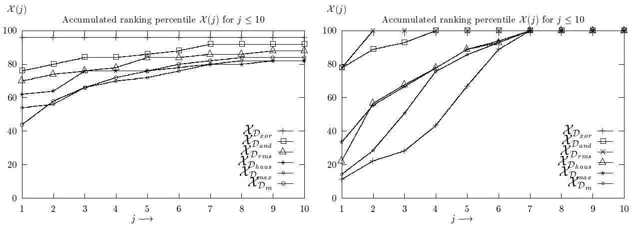

Figure 1. *a: The accumulated ranking percentile $\mathcal{X}$ without object clutter. b: The accumulated ranking percentile $\mathcal{X}$ with object clutter.*

Figure 1.a shows $\mathcal{X}$ for $j \leq 10$ differentiated for the various similarity functions, with $N_1 = 500$ and $N_2 = 70$.

The results show that for $\mathcal{D}_{xor}$, 98% of the images have rank 1. In other words, with the probability of 98 perfect matches out of 100, very high retrieval accuracy is achieved. This is due to the fact that $\mathcal{D}_{xor}$ is sensitive to both false positive and false negative error types, and relatively insensitive to outliers.

Furthermore, high retrieval accuracy is given by $\mathcal{D}_{and}$ and $\mathcal{D}_{rms}$, for which 92% and 89% of the correct matches are within the first 10 images. All other similarity functions produce slightly worse retrieval accuracy. $\mathcal{D}_{max}$ gives worst results.

The matching process has been executed very fast resulting in an average user time of 0.79 seconds on a SPARC 10. More details on the run-time complexity can be found in [5].

4.5 Stability to occlusion and a change in viewpoint

To test the effect of occlusion, 10 objects, already in the database of 500 recordings, were randomly selected and in total 40 images were generated by blanking out $o \in \{50, 65, 80, 90\}$ percent of the total object area. The average ranking percentile $\overline{r}$ with $N_1 = 500$ and $N_2 = 10$ is shown in Figure 2.a.

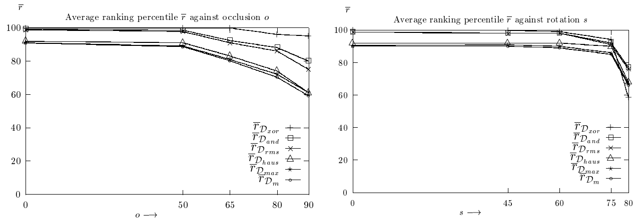

Figure 2. *a: The discriminative power $\overline{r}$ plotted against blanking out o. b: The discriminative power $\overline{r}$ plotted against change in viewpoint s.*

From the results we see that the effect of occlusion (i.e. blanking out object area) is largely the same for all similarity functions: namely a gradual decrease in retrieval accuracy beyond 50% blanking.

To test the effect of change in viewpoint, the 10 objects were put orthographically in front of the camera and in total 40 recordings were made by varying the angle between the camera for $s = \{45, 60, 75, 80\}$ degrees with respect to the object's surface normal. The average ranking percentile with $N_1 = 500$ and $N_2 = 10$ is shown in Figure 2.b.

Looking at the results, the rate of decrease in retrieval accuracy is almost negligible for $s < 75^o$. This means that hue-based pattern-card matching is highly robust to a change in viewpoint up to 75^o.

4.6 Discriminative power in the presence of object clutter

To test the effect of object clutter, a small and preliminary experiment has been conducted. 30 images have been recorded from cluttered scenes. Each cluttered scene contained 6 different objects. Then, 10 objects were randomly selected which participated in exactly one of the cluttered scenes. These objects were recorded in isolation against a white background yielding the query set. The query set $N_2 = 10$ was matched against the database of $N_1 = 30$ images. Although the data set is small, some important results can be observed, see Figure 1.b, showing the accumulated ranking percentile differentiated for various similarity functions.

As can be expected, $\mathcal{D}_{xor}$ provides poor retrieval accuracy in the presence of clutter. This is because $\mathcal{D}_{xor}$ is sensitive to false negatives. False negatives are introduced by hue-hue edges coming from objects surrounding the object described by the query image.

From the results we can see that $\mathcal{D}_{rms}$ and $\mathcal{D}_{and}$ give high image retrieval accuracy in the presence of object clutter. $\mathcal{D}_{rms}$ and $\mathcal{D}_{and}$ are insensitive to false negatives.

5 Discussion and Conclusion

Similarity functions were presented to express to what extent hue-hue edge pattern-cards are similar. The aim was to get to similarity functions allowing for fast and accurate pattern-card matching.

Without any object clutter, experimental results show very high retrieval accuracy for $\mathcal{D}_{xor}$ (which is a metric). $\mathcal{D}_{rms}$ and $\mathcal{D}_{and}$ give similar but slightly worse retrieval performance. Further, the experimental results show that hue-based pattern-card matching is robust to a change in viewpoint up to 75^o. The effect of occlusion occurs only gradually beyond 50% blanking for all similarity functions. In the presence of object clutter in the scene, preliminary results reveal that high image retrieval accuracy is given by $\mathcal{D}_{rms}$ and $\mathcal{D}_{and}$. It is concluded that for accurate image retrieval $\mathcal{D}_{xor}$ is appropriate when there is no object clutter. $\mathcal{D}_{rms}$ and $\mathcal{D}_{and}$ are appropriate in the presence of object clutter. No constraints have been imposed on the images and the camera imaging process other than that images should be taken from colored objects illuminated by average day-light color. It is our point of view that these conditions are acceptable for a large variety of applications yielding a promising hue-based pattern-card matching method for viewpoint independent image retrieval.

References

[1] Baddeley, A. J., *An Error Metric for Binary Images*, In Proceedings of Robust Computer Vision, W. Forstner and S. Ruwiedel (eds.), Karlsruhe, pp. 59 - 78, 1992.

[2] Funt, B. V. and Finlayson, G. D., *Color Constant Color Indexing*, IEEE PAMI, 17(5), pp. 522-529, 1995.

[3] Gevers T. and Smeulders A.W.M., *Enigma: An Image Retrieval System*, In proceedings of the 11th IAPR International Conference on Pattern Recognition, J. Kittler (ed.), The Hague, The Netherlands, IEEE Computer Society Press, 1992.

[4] Gevers, T. and Smeulders, A. W. M., *Evaluating Color and Shape Invariant Indexing for Consumer Photography*, In Proceedings of Visual'96: The First International Conference on Visual Information Systems, Melbourne, Australia, 1996.

[5] Gevers, T., *Color Image Invariant Segmentation and Retrieval*, PhD Thesis, ISBN 90-74795-51-X, University of Amsterdam, The Netherlands, 1996.

[6] Kender, J. R., *Saturation, Hue, and Normalized Colors: Calculation, Digitization Effects, and Use*, Technical Report, Department of Computer Science, Carnegie-Mellon University, 1976.

[7] Lambert J. H., Photometrica Sive de Mensura de Gratibus Luminis, Colorum et Umbrae, Eberhard Klett, Augsberg, Germany, 1760.

[8] Niblack, W., Barber, R., Equitz, W., Flickner, M., Glasman, E., Petkovic, D. and Yanker, P., *The QBIC Project: Querying Images by Content Using Color, Texture, and Shape*, In Proceedings of Storage and Retrieval for Image and Video Databases, SPIE, 1993.

[9] Pentland, A., Picard, R. W. and Sclaroff, S., *Photobook: Tools for Content-based Manipulation of Image Databases*, In Proceedings of Storage and Retrieval for Image and Video Databases II, 2, 185, SPIE, Bellingham, Wash. pp. 34-47, 1994

[10] Swain, M. J. and Ballard, D. H., *Color Indexing*, International Journal of Computer Vision, Vol. 7, No. 1, pp. 11-32, 1991.