Object and Scene Recognition Using Color Descriptors and Adaptive Color KLT

Volkan H. Bagci[1], Mariofanna Milanova[1], Roumen Kountchev[2], Roumiana Kountcheva[3], and Vladimir Todorov[3]

[1] Computer Science Department, UALR, 2801 S. University Ave., Little Rock, Arkansas 72204, USA
[2] Department of Radio Communications, Technical University of Sofia, Bul. Kl. Ohridsky 8, Sofia 1000, Bulgaria
[3] T&K Engineering, Mladost 3, Sofia 1712, Pob.12, Bulgaria
{Vhbagci,mgmilanova}@ualr.edu, rkountch@tu-sofia.bg, {todorov_vl,kountcheva_r}@yahoo.com

Abstract. With the emergence and explosion of huge image databases there is an increasing necessity for effective methods to assess visual information on the level of objects and scene types. A wide variety of Content – Based Image Retrieval (CBIR) systems already exists. As a key issue in CBIR, similarity measure quantifies the resemblance in contents between a pair of images. Depending on the type of features, the formulation of the similarity measure varies greatly. The primary goal of our study is to reduce the computation time and user interaction. The secondary goal is to reduce the semantic gap between high level concepts and low level features. A third goal is to evaluate system performance with regard to speed and accuracy. In the proposed study transform color after statistical transform, such as the Adaptive Color Karhunen Loeve Transform (ACKLT) is used as a color descriptor. The results are showing the advantage of the new algorithm for ACKLT in comparison with the YCrCb color model. Based on the experimental results, we concluded that correct selection of descriptors invariant to light intensity and light color changes affects object and scene category recognition.

Keywords: content-based image retrieval, Adaptive Color Karhunen Loeve Transform.

1 Introduction

All current Content–Based Image Retrieval (CBIR) systems retrieving stored images from a collection by comparing features automatically extracted from the images themselves. The system then identifies those stored images whose feature values match those of the query most closely and displays the result on the screen. CBIR system is using machine learning techniques and image descriptions to distinguish objects and scene categories. For real world scenes, there can be large variations in viewing and lighting conditions and this complicates the image descriptions. Koen and al. in [1] discussed the invariance properties and the distinctiveness of color descriptors.

G. Salvendy, M.J. Smith (Eds.): Human Interface, Part II, HCII 2011, LNCS 6772, pp. 355–363, 2011.

There are two basic types of color spaces – deterministic and statistical. The deterministic transformations such as YCrCb, YUV, YIQ, CMYK [2, 3, 4] are calculated using fixed coefficients and require less computations but their disadvantage is that they are not adapted to each individual image that is being processed. In the other type – the statistical transforms such as the Adaptive Color Karhunen-Loeve Transform (KLT) [2] the generated color space is adapted to the statistical properties of each image or group of images that is being transformed but the disadvantage is that it requires more computations than the deterministic color systems. That gives better quality of the restored image, less correlation of the components, etc [5, 6]. In [9] the new method for color segmentation of human faces using KLT approximation with a matrix of fixed coefficients. In this case the KLT is of relatively low computational complexity, but it is not fully adapted to the local statistics of the face colors.

In the proposed system texture features and color features for computing the similarity between query and database images are used. The invariance properties of color descriptors are explored. In statistical transforms, (in our case, the Adaptive KLT) the components are adapted to each image that is being transformed. Therefore, the transform color descriptor is adapted to the statistical information of the image. The results show the advantages of the new algorithm for Adaptive Color KLT in comparison with the YCrCb color model. The most important feature of the Adaptive Color KLT is that it ensures strong components decorrelation. Based on the experimental results the conclusion is that correct selection of descriptors invariant to light intensity changes and light color changes affects object and scene category recognition .

On Fig. 1 is shown the Module Block Diagram of the proposed system. The new approach includes first, the extraction of color descriptors based on Adaptive Color Karhunen Loeve Transformation (KLT); and second, the matching and classification is implemented. The system calculates the distance similarly between the image request and images stored in the database. As an output the system calculates the accuracy of object recognition and displays sorted images in ascending order of distance similarity.

This paper is organized as follows: In section 2 the algorithm for Adaptive Color KLT is presented. In section 3 the recognition module is given. The experimental setup is presented in section 4. Finally, in section 5, conclusions are drawn.

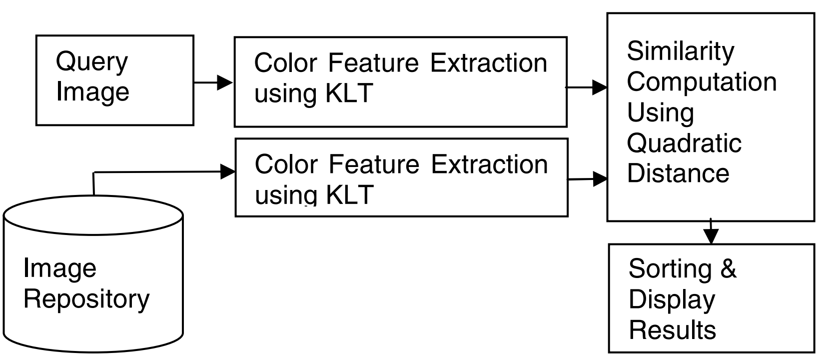

Fig. 1. Color Feature Extraction Block Diagram

2 Color Transformation Based on KLT

The proposed algorithm is a complete analytical solution to the problem of the color transform based on the KLT. It is based of the method presented in [1]. The algorithm is simplified so that to reduce the necessary computations of the color transform.

Transforming an RGB image into the new color format is made by the following steps following the proposed algorithm presented in Fig 2, blocks (1)-(10), which is the forward algorithm for the Adaptive Color KLT:

Step 1: Determination of the primary color vectors $\vec{C}_s$ for each pixel from the original RGB image, where s is the current pixel and S the total number of the pixels in the image, therefore $S = M \times N$, where M and N are the image height and width.

Step 2: Calculation of mean values of the colors R, G and B - Fig. 2, block (4). The mean values are necessary for the computation of the covariance matrix in the next step.

Step 3: Calculation of the image covariance matrix:

$$[K_C] = \left[\frac{1}{S}\sum_{s=1}^{S}\vec{C}_s\vec{C}_s^t\right] - \vec{m}_c\vec{m}_c^t = \begin{bmatrix} k_{11} & k_{12} & k_{13} \\ k_{21} & k_{22} & k_{23} \\ k_{31} & k_{32} & k_{33} \end{bmatrix} \tag{1}$$

Where the coefficients $k_{i,j}$ are calculated using Fig. 2, block (4). The covariance matrix is a diagonal matrix so therefore the eigenvalues are always real numbers.

Step 4: Calculation the coefficients (a, b, c) of the characteristic equation of the covariance matrix,

$$\det| k_{ij} - \lambda\delta_{ij} | = \lambda^3 + a\lambda^2 + b\lambda + c = 0 \tag{2}$$

using equations Fig. 2, blocks (5).

Step 5: Calculation of the eigenvalues of the characteristic equation defined in the previous step. Given that the covariance matrix $[K_C]$ is a diagonal matrix the eigenvalues can be defined by the “Cardano” relations or the so called trigonometric equations [10] Fig.2, block (6) and (7). Where we have the condition

$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0$.

Step 6: Calculation of the eigenvectors of the covariance matrix $[K_C]$: Fig. 2, block (8). Here A_m, B_m, D_m, P_m are the coefficients used for the computations and $m = 1,2,3$. From the eigenvectors we form the transformation matrix $[\Phi]$:

$$[\Phi] = \begin{bmatrix} \vec{\Phi}_1^t \\ \vec{\Phi}_2^t \\ \vec{\Phi}_3^t \end{bmatrix} = \begin{bmatrix} \Phi_{11} & \Phi_{12} & \Phi_{13} \\ \Phi_{21} & \Phi_{22} & \Phi_{23} \\ \Phi_{31} & \Phi_{32} & \Phi_{33} \end{bmatrix} \tag{3}$$

1 [R], [G], [B] of size M×N = S

2 $$\vec{C}_s = [R_s, G_s, B_s]^t \text{ for } s = 1,2,..,S$$

3 $$\bar{R} = \frac{1}{S}\sum_{s=1}^{S} R_s;\ \bar{G} = \frac{1}{S}\sum_{s=1}^{S} G_s;\ \bar{B} = \frac{1}{S}\sum_{s=1}^{S} B_s$$

4 $$k_1 = \frac{1}{S}\sum_{s=1}^{S}(R_s - \bar{R})^2;\ k_2 = \frac{1}{S}\sum_{s=1}^{S}(G_s - \bar{G})^2;\ k_3 = \frac{1}{S}\sum_{s=1}^{S}(B_s - \bar{B})^2;$$
$$k_4 = \frac{1}{S}\sum_{s=1}^{S}(R_s - \bar{R})(G_s - \bar{G});\ k_5 = \frac{1}{S}\sum_{s=1}^{S}(R_s - \bar{R})(B_s - \bar{B});\ k_6 = \frac{1}{S}\sum_{s=1}^{S}(G_s - \bar{G})(B_s - \bar{B})$$

5 $$a = -k_1 - k_2 - k_3;\ b = k_1 k_2 + k_1 k_3 + k_2 k_3 - k_4^2 - k_5^2 - k_6^2;$$
$$c = k_1 k_6^2 + k_2 k_5^2 + k_3 k_4^2 - k_1 k_2 k_3 - 2 k_4 k_5 k_6.$$

6 $$p = -(a^2/3) + b < 0,\ q = 2(a/3)^3 - (ab)/3 + c,\ \varphi = \arccos\left[-q/2\sqrt{(|p|/3)^3}\right]$$

7 $$\lambda_1 = 2\sqrt{\frac{|p|}{3}}\cos\left(\frac{\varphi}{3}\right) - \frac{a}{3};\ \lambda_2 = -2\sqrt{\frac{|p|}{3}}\cos\left(\frac{\varphi+\pi}{3}\right) - \frac{a}{3};\ \lambda_3 = -2\sqrt{\frac{|p|}{3}}\cos\left(\frac{\varphi-\pi}{3}\right) - \frac{a}{3}$$

8 $$A_m = (k_3 - \lambda_m)[k_5(k_2 - \lambda_m) - k_4 k_6],\qquad B_m = (k_3 - \lambda_m)[k_6(k_1 - \lambda_m) - k_4 k_5],$$
$$D_m = k_6[2k_4 k_5 - k_6(k_1 - \lambda_m)] - k_5^2(k_2 - \lambda_m),\qquad P_m = \sqrt{A_m^2 + B_m^2 + D_m^2};$$
$$\Phi_{m1} = A_m/P_m;\ \Phi_{m2} = B_m/P_m;\ \Phi_{m3} = D_m/P_m \qquad m=1,2,3$$

9 $$\begin{bmatrix} L_{1s} \\ L_{2s} \\ L_{3s} \end{bmatrix} = \begin{bmatrix} \Phi_{11} & \Phi_{12} & \Phi_{13} \\ \Phi_{21} & \Phi_{22} & \Phi_{23} \\ \Phi_{31} & \Phi_{32} & \Phi_{33} \end{bmatrix} \begin{bmatrix} R_s \\ G_s \\ B_s \end{bmatrix}$$
for $s = 1,2,..,S$

10 $$\vec{L}_s = [L_{1s}, L_{2s}, L_{3s}]^t \text{ for } s = 1,2,.., S$$

Fig. 2. Block diagram of the algorithm for Adaptive Color KLT

Step 7: Performing the color transform using the already generated transformation matrix $[\Phi]$ to obtain the transformed color vectors $\vec{L}_s = [L_{1s}, L_{2s}, L_{3s}]^t$ using the equation Fig, 2, block (9). Where again s is the current pixel that is being transformed and S is the total number of the pixels in the image.

Step 8: Perform an adaptive quantization of the obtained matrix $[L_1]$ to comply with the limits of 8 bits per pixel or 256 unique values in the matrix using the equations from Fig. 2, block (9). Here $h_{L_1}(t)$ is the histogram calculated for the first component of the Adaptive Color KLT, t_k^c is the center of gravity of the part of the histogram $h_{L_1}(t)$ between levels t_k and t_{k+1} of the component L_1 (k=1,2,..,K for K - number of quantization levels). After adaptive quantization we have three matrices $[\hat{L}_1],[\hat{L}_2],[\hat{L}_3]$ that comply with the limit of 8 bits per pixel or 24 bpp for each pixel in the image.

3 Object and Scene Recognition

The proposed system comprises two basic steps of image processing, described below.

3.1 Feature Extraction

The color features vector of query image and database images are computed. Then these features of a query image and the features of database images will be used either to find the smaller distances or large similarity.

In this work is offered to apply the direct ACKLT on the color vectors $\vec{C}_s = [R_s, G_s, B_s]^t$ for s=1,2,..,T using the equation Fig, 2, block (9). In result they are transformed into vectors $\vec{L}_s = [L_{1s}, L_{2s}, L_{3s}]^t$, which are after that normalized regarding their module.

3.2 Image Classification

The relevant images are sorted out based on the dissimilarity and displayed in ascending manner. That is if distance measures are used they are displayed starting from shorter to longer distances and if similarity measure is used they are displayed starting with higher similarity to lower similarity.

The main aim of classification is to identify the characteristics that indicate the group to which each case belongs. It can also be used to understand the existing data and to predict how new instances will behave. For example it can be used to predict cases like whether individuals can be classified as likely to respond to a direct mail solicitation, vulnerable to switching over to a competing long distance phone service, or a good candidate for a surgical procedure.

Classification models are created by examining already classified data (cases) and inductively finding a predictive pattern. The existing cases may be from an historical database, such as images that belong to particular class. They may come from an experiment in which a sample of the entire database is tested in the real world and the results used to create a classifier. For example, a sample of a mailing list would be sent an offer, and the results of the mailing used to develop a classification model to be applied to the entire database. Sometimes an expert classifies a sample of the database, and this classification is then used to create the model which will be applied to the entire database.

In this study a well known algorithm called Nearest-Neighbor algorithm is implemented . In this algorithm when an image is given it is compared to each and every image in the data base using a distance measure. The “unknown” or input image is said to be classified as belong to the class to which its distance will be shortest. It can also be used using similarity measure. If similarity measure is calculated we have to consider the largest value for similarity.

To compare the distance between two manifolds we are using the distance measure variant of the Hausdorff metric. The proposed distance measure can handle changes in duration and is invariant to temporal shifts.

Given two manifolds from two images $A = [a_1.a_2.......a_n]$ and $B = [b_1.b_2....b_m]$, we define

$$d(A,B) = \frac{1}{n}.\sum_{j=1}^{n} \min \left\| \frac{a_j}{\|a\|} - \frac{b_i}{\|b_i\|} \right\| \quad (4)$$

To ensure similarity, was used the following distance measure, presented in [8]:

$$D(A,B) = d(A,B) + d(B,A) \quad (5)$$

For the final activity classification was adapted the Nearest Neighbor classifier (NN). It is assumed that *T* represents a test image and R_i represents the reference image of the class I. Then, the test image is classified into the class *I* which will minimize the similarity distance between the test image and the reference image,

$$c = \arg\min D_i(T, R_i)\,, \quad (6)$$

where D is the similarity measure described in (5).

4 Experiments and Results

We used the Amsterdam Library of Object Images (ALOI) dataset [7] Wide – baseline Stereo Full Color collection for experiments. This collection includes 2250 images and 750 objects.We use 30 different objects under various illumination conditions. Different lighting conditions are presented in the AIOI such as : objects lighted by different number of white lights, object rotation images and images with different levels of JPEG compression.

Object Type (Code)	True					Type						Total
	0	1	2	3	4	5	6	7	8	9	X*	
0 (257)	3											3
1 (284)		2									**1**	3
2 (337)			3									3
3 (406)				2							1	3
4 (503)					2						1	3
5 (602)						2					1	3
6 (706)							3					3
7 (836)								3				3
8 (909)									3			3
9 (979)										3		3
X*												0
Total	3	2	3	2	2	2	3	3	3	3	4	30

Fig. 3. Confusion Matrix

On Fig 3 is shown the confusion matrix.

R = E / S = 4 / 30 = 0.13

Where E is a number of wrong classifications and S is a total number of objects.

A = 1 – R = 1 – 0.13 = 0.87 (87%)

Confusion matrix uses the test data including 3 closest images based on each test image but do not neglect the test image itself in results.

According to test results, system performs well based on the database. When the system misses an image, it's observed that missed image is close to the location of the current neighbor among other 2200+ neighbors. In the future we aim to test the accuracy of the method based on different lightning conditions, different temperature and wide rotation angles. Another test would be to use multiple levels of image resolution to understand its effect in accuracy based on the levels of resolution.

On Fig. 4 are shown randomly chosen test images.

Fig. 4. Randomly Chosen Test images

5 Conclusions

There are two types of color transforms – deterministic (RGB, YcrCb, YUV, YIQ, CMYK) and statistical (ACT – Adaptive Color KLT Transform). The deterministic transforms are defined by fixed equations with fixed coefficients which are not changing for each image. The statistical transforms (ACT) the components are adaptive for each image that is being transformed. Therefore the transform matrix generated by the algorithm is adapted to the statistical information of the image that is being transformed. The new approach permits reliable object detection and identification in various positions, lighting conditions and viewpoints.

Acknowledgments. This paper was supported by the System Research and Application (SRA) Contract No. 0619069. This work was also supported in part by the Joint Research Project Bulgaria-Romania (2010-2012): "Electronic Health Records for the

Next Generation Medical Decision Support inRomanian and Bulgarian National Healthcare Systems".

References

1. Koen, E., van de Sande, A., Gevers, T., Snoek, C.: Evaluating Color Descriptors for Object and Scene Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 32(9), 1582–1596 (2010)
2. Kountchev, R., Kountcheva, R.: New Method for Adaptive Karhunen-Loeve Color Transform. In: Proc. of 9th Intern. Conf. on Telecommunications in Modern Satellite, Cable and Broadcasting services (TELSIKS 1909), Nish, Serbia, October 7- 9, pp. 209–216 (2009)
3. Ivanov, P., Kountchev, R.: Comparative analysis of Adaptive Color KLT and YCrCb for representation of color images. In: Proc. of ICEST (2010)
4. Pratt, W.: Digital Image Processing. Wiley Interscience, New York (2007)
5. Fleury, M., Downton, A., Clark, A.: Karhunen–Loeve Transform – Image Processing. University of Essex, Wivenhoe Park (1997)
6. Dony, R.: The Transform and Data Compression Handbook. In: Rao, K., Yip, P., Raton, B. (eds.) Karhunen-Loève Transform. CRC Press, Boca Raton (2001)
7. Geusebroek, M., Burghouts, G., Smeulders, A.: The Amsterdam library of object images. Int. Journal of Computer Vision 61(1), 103–112 (2005), `http://staff.science.uva.nl/~aloi/`
8. Masoud, O., Papanikolopoulos, N.: A method for human action recognition. Image and Vision Computing 21(8), 729–743 (2001)
9. Ionita, M., Corcoran, P.: Benefits of using decorrelated color information for face segmentation/tracking. Advances in optical technologies. Hindawi Publishing Corporation, ID 583687 (2008)