POINTS OF INTEREST EXTRACTION USING PAIRWISE CLUSTERING AND SPATIAL FEATURES

Dimitrios Besiris[a], Nikolaos A. Laskaris[b], E. Zigouris[a], George Economou[a] and Spiros Fotopoulos[a]

[a]Electronics Laboratory, Department of Physics, University of Patras
Patras 26500, Greece.
phone: + 30 2610 996058, fax: + 30 2610 997456, email: spiros@physics.upatras.gr

[b]Artificial Intelligence & Information Analysis Laboratory, Department of Informatics
Aristotle University of Thessaloniki, Thessaloniki 54124, Greece.

ABSTRACT

In this work, the idea of local features extraction from image data based on points of interest, is revised. The method is based on a nonparametric pairwise clustering algorithm and the application of Hubert's test statistic. The clustering algorithm iteratively partitions the input image data until it finally converges to 2 classes. On the other hand the use of Hubert's test guarantees that the 2 classes in the feature space are associated with a well organized structure in the image plane. Both algorithms utilize the dissimilarity matrix of the input data. The validity of the approach is demonstrated by applying the method to an image retrieval system.

1. INTRODUCTION

In content based image retrieval applications (CBIR), the query object usually covers only a fractional part of the image. As a result, global image descriptors are inefficient and a more local solution must be considered, where each region of the image is analyzed independently. A colour image [10] can be considered that it is composed of two kinds of regions: smooth and structured areas. Smooth areas are regions with no interest from information point of view, since these present little chromatic variability. On the other hand, structured areas are regions containing the maximum amount of colour information, necessary for a successful recognition, and are referred as points of interest. Points of interest are arranged in three categories, according to the shape information these provide: junctions of Y, X, T type, L corners and edges with horizontal, vertical or without orientation.

Localized invariant features have been applied to image retrieval applications since these offer robustness to partial occlusion, varying illumination conditions and background changes. However, local features are not always sufficient for image discrimination without colour consideration, as colour is one of the most intuitive features for visual recognition. Points of interest proved [1], [2], [3] to be robust to geometric and illumination transformations, and their colour and shape information contributed to the extraction of invariant and distinguished image features.

In this work we propose a new method for colour feature extraction from local image regions [4] that contain points of interest. The identification of points of interest is seen as the detection of multimode colour regions, areas with informative colour structure, localizing the query object in the image. The method includes four steps:

- Step 1: Structured areas detection

The image is organized in uniform rectangular regions and the proximity matrix of each one is computed. The procedure is presented in Section 2.

- Step 2: Multimode regions detection

The colour structure of regions is clustered using a non parametric bipartition algorithm, that provides compact colour modes, as shown in Section 3.

- Step 3: Points of interest identification

Regions with more than one mode can be considered as points of interest. Hubert's test detects the structure of colour modes in the plane, providing information about the shape of the regions. The procedure is presented in Section 4.

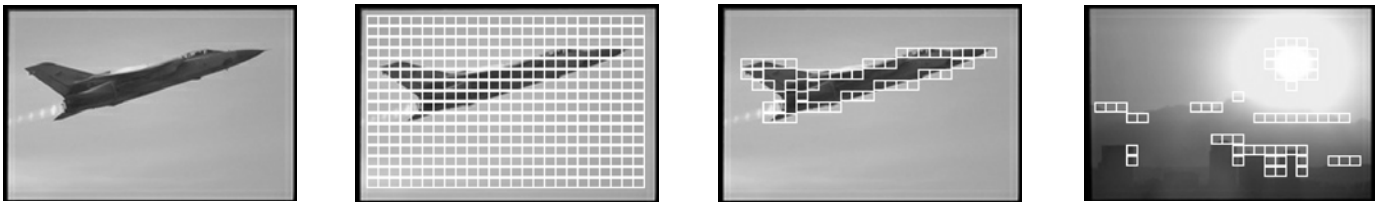
- Step 4: Colour features extraction

From the selected points of interest a set of colour features are extracted, as shown in Section 5.

Finally in Section 6 the quality of the feature extraction process is evaluated by applying the technique to a CBIR system.

2. REGION AND PROXIMITY MATRIX COMPUTATION

Regions are formed using uniform non-overlapping square windows that cover the whole image plane in a grid structure, as shown in Figure 1(b). The exact window shape is not critical, but the choice of a square size facilitates simple and fast processing of data. For an image with m×n pixel dimensions, the number of extracted regions is the quotient of $\frac{n \cdot m}{b^2}$, where b is the size of the square region.



(a) (b) (c) (d)

Figure 1: (a) original image, (b) grid with b=7, (c) regions with $\overline{M}$ > threshold, (d) regions with $\overline{M}$ > threshold, for an image with small colour distribution.

For each region the pairwise proximity distance matrix $M=d_{ij}$, i,j= 1,…,b^2, is computed, as the dissimilarity measure of every pixel pair of the region's data set, $f_i=(R_i, G_i, B_i)$, i=1,…, b^2. The most common dissimilarity measure is the Euclidean distance:

$$d_{ij} = D(f_i, f_j) = \sqrt{(R_i - R_j)^2 + (G_i - G_j)^2 + (B_i - B_j)^2}$$

, or often given in squared form that simplifies computations:

$$d_{ij} = D(f_i, f_j) = (f_i - f_j) \cdot (f_i - f_j)^T$$

Clearly for smooth areas the mean value of the proximity matrix given by: $\overline{M} = \frac{1}{b^4} \sum_i \sum_j d_{ij}$, i,j=1,…$b^2$, is very small (zero in case of colour-constant areas), while for points of interest this value is larger, as a result of their colour structure. Selecting the mean value of all regions $\overline{M}_k$, where k=1,…, $\frac{n \cdot m}{b^2}$, as a threshold we can disregard those ones with $\overline{M}_k < \frac{b^2}{n \cdot m} \sum_k \overline{M_k}$ and focus only to the regions containing points of interest, as shown in Figure 1(c). Selecting a fixed threshold for all images is not appropriate, since colour distributions of different images can differ considerably. Such a case is illustrated in Figure 1(d). Although this method demands the computation of all regions proximity matrix, experimental results shown that more than half of regions are discarded, alleviating the computations needed during the detection process.

3. PAIRWISE CLUSTERING ALGORITHM

Given the proximity matrix $M = (d_{ij})$,i,j=1,…,b^2 informative of the region's colour structure, the algorithm extracts the two most prominent clusters of the data set in the colour feature space. The algorithm iteratively employs a two-step transformation [7] on the proximity matrix. The first step of the transformation represents each pixel by its relation to all others, and the second step re-estimates the pairwise distances using a statistically motivated proximity measure on these representations. Using this transformation, the algorithm iteratively partitions the data points, until it finally converges to two distinguished clusters.

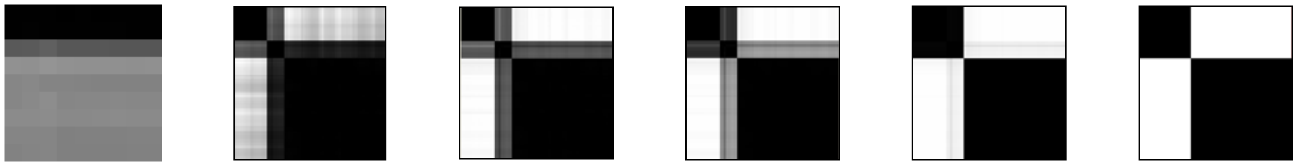
- Normalization step. For each pixel's colour vector i, let $d_i=(d_{i1},d_{i2},\ldots,d_{ib^2})^t$ is the vector of proximities of pixel i to all the other colour vectors. For each i d_i is normalized. The resulting vector is thus $p_i=(p_{i1},p_{i2},\ldots,p_{ib^2})=\left(\frac{d_{i1}}{\sum_k d_{ik}}, \frac{d2}{\sum_k d_{ik}}, \ldots, \frac{d_{in}}{\sum_k d_{ik}},\right)$
- Re-estimation step. The proximity matrix is re-estimated for every pair of data i and j, so that: $M^{new} = (d_{ij}^{new})_{i,j=1,5\ b^2} \Rightarrow M^{new} = T^k(M)$

, where $T^k(M)$ denote the k-fold composition of T, starting with M. The iterative transformation of the proximity matrix is given by the set of equations:

$$d_{ij}(t+1) = \sum_k D[p_i(t+1), p_j(t+1)]\text{, k=1,…}b^2$$

, where D denotes the Euclidean distance of vectors p_i and p_j. In each re-estimation step of the algorithm, the estimated values of the proximity matrix lie in the range [0, 1], and are distributed in two groups. As the number of iterations increases these groups start to separate, forming two compact set of values. After a sufficient large number of iterations, depending on the image, the estimated proximity matrix M^{new}, consists of only two values. The above described procedure is illustrated in Figure 2. Clearly, for regions with no texture, that have one or two modes, only one iteration is needed, as the nearest colour vectors are addressed to the same cluster. Thus each region can now be transformed to a binary image, corresponding each binary vector $d_i^{new} = (d_{i1}^{new}, d_{i2}^{new}, 5\ , d_{ib^2}^{new})$ of the proximity matrix in one of the two clusters.

For colour image clustering, the number of iterations needed to attain a stable condition is critical, as it increases considerably computation time, thus reducing the performance of the mode detection process. For a small number of iterations the binary mode of the proximity matrix can be approached to a certain limit. Using the mean value of the estimated proximity matrix as threshold, the two distributed values conclude two binary ones. Based on these considerations only a few iterations are required, since additional ones can only marginally influence the clustering result.



(a) (b) (c) (d) (e) (f)

Figure 2: (a) the image, (b) the original proximity matrix, (c) the re-estimated proximity matrix after k=1 iteration, (d) after k=2 iterations, (e)after k=4 iterations, (f) after k=5 iterations.

In order to estimate the minimum number of iterations required achieving a stable clustering state, the following study was conducted using the COREL [5] image data base. For all regions the pairwise clustering algorithm is applied to all regions, constructing for each iteration a binary image $[X_{ij}]^z$, i,j=1,…,b and z=1,…,k. The classification errors occurred, as a mismatch of $[X_{ij}]^z$, z=1,…,k-1 and $[X_{ij}]^k$, where k is the maximum number of iterations needed, for the correct bipartition of the image data set. The study includes different values of b (region size) and the results are presented in table 1.

Width	**Number of errors**	0	1	2	3	4
7	**Number of Iterations**	10	8	7	3	5
8		10	8	7	7	6
9		11	9	8	7	7
10		11	9	8	7	7
11		12	9	9	8	8
12		12	10	9	8	8
13		13	11	10	9	9

Table 1: The number of iterations according to region's size and classification errors.

After the bipartition process a cluster validation measure must be computed, since the 2 classes in the feature space must be well separated. The validity coefficient [6] used is:

$$V = \frac{\overline{M_1} - \overline{M_2}}{D[\overline{X_1}, \overline{X_2}]}$$

, where $\overline{M}_Z = \frac{1}{b_Z^2} \sum_i \sum_j d_{ij}$, z=1,2 and i,j=1,… b_z^2

, b_z is the size of the z cluster and $\overline{X_z}$ is the mean value of the cluster z colour vectors. The validity index is determined as the proportion of the inner distance versus the outer distance of the colour vectors in the two clusters. Small values of V indicate a good separation, while large values indicate the possible presence of more than one colour mode. Clearly for a bimodal region, only 1 bipartition step is needed.

3. HUBERT'S Γ STATISTIC

Hubert's test is a statistical criterion for similarity evaluation between two proximity matrices [8]. Given two proximity matrices M_z, z=1,2, containing the distance measurements of two datasets, as components of the same space, the statistic is defined by their serial correlation:

$$\Gamma = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (d_{ij})_1 (d_{ij})_2 \ .$$

Denote that Γ's definition uses only the upper diagonal entries of M_z, since proximity matrices are symmetrical, reducing the required computations. In its normalized form Γ_n, Hubert's statistic is the sample correlation between the two matrices entries:

$$\Gamma_n = \frac{2}{n(n-1)M_{1s}M_{2s}} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left((d_{ij})_1 - M_{1m}\right) \cdot \left((d_{ij})_2 - M_{2m}\right)$$

, where M_{1m}, M_{2m} and M_{1s}, M_{2s} denote respectively the mean value and the standard deviation of M_1 and M_2. The range of Γ depends on the ranges of values in the proximity matrices and on the number of entries, while Γ_n is always within [-1, 1]. Unusually large absolute values of this statistic suggest that the two matrices are in accordance with each other.

The selection of spatial information, as entry of the second data set, provides the structure information of pixels on the image plane. The feature proximity matrix M_2 is replaced with the spatial distance matrix. Hubert's statistic can compare the two different structures and output a characteristic value for each data structure type. To indicate spatial information the use of 2-dimensional image indexing scheme is suggested. For example for b=3, region's index is: $\begin{bmatrix} 11 & 12 & 13 \\ 21 & 22 & 23 \\ 31 & 32 & 33 \end{bmatrix}$. The according spatial information is symmetrical across the first diagonal and provides distance matrix values also symmetrical to vertical and horizontal orientation. In this way Γ value is independent of image rotation, between horizontal and vertical directions.

By definition, Hubert's statistic depends both on image structure as well as on chromatic information, providing different values for characteristic types of regions. Γ value is proved to be almost independent of region's size, since each increase or decrease of size by 1, changes the value of Γ by a proportion 0.28%. Smooth areas contain monochromatic information (unimodal regions), there is no structure present and as a result the Γ statistical value approaches zero. Points of interest are associated with multimodal regions. More analytical bimodal regions include different types of edges and corners. Areas with vertical or horizontal orientation produce the same Γ values, irrespective of colour information. In case where the cardinality of the pixels in the two classes is different, the Γ value decreases. As a result corners present smaller values than edges, as shown in Table 2.

Number of modes	**Type**	**Range of Γ values**	
		With no texture	**With texture**
1	Smooth areas	0	~|0.08|
2	Horizontal-vertical edges	(0.26, 0.445)	~(0.1, 0.46)
	Diagonal edges	(0.24, 0.39)	~(0.16, 0.41)
	Corners L	(0.18, 0.29)	~(0.1, 0.32)

Table 2: Hubert's Γ statistic typically measurements for unimodal and bimodal regions with different structure.

Multimodal regions contain junctions with variant shapes. Different colour content results to different Γ values, since the colour information in the feature space contributes to a variant spatial structure in the plane. As a result for multimodal regions no firm conclusions can be extracted.

For bimodal regions with no texture the Hubert's test results to stable values, ranging from 0.15 to 0.45. For these regions the Γ value can distinguish the shape of the area, as the colour information has no effect.

5. FEATURE EXTRACTION

From multimodal regions, a number of invariant local colour features can be extracted. In this study to simplify the analysis we will describe feature extraction from bimodal neighborhoods only.

Clearly feature extraction is focused on points of interest as regions having rich information content. Points of interest can be identified with bimodal regions forming well separated classes (V small) and compact spatial arrangement (Γ big). The number of points of interest to be detected on an image depends on its colour distribution, and on the size of the region and the query object.

For each point of interest, the extracted colour feature consists of the resulting pair of colour vectors f_i, i=1,2 associated with the bimodal region clusters. The two vector are computed using the mean colour value in each cluster.

The colour feature vector used in this work is the projection of the two colour vectors f_1 and f_2 to a 6D feature space:

$$f = [(R_1, G_1, B_1), (R_2, G_2, B_2)]$$

To order the two vectors, the one with the largest magnitude value, is used first. The evaluation of points of interest recognition and features extraction is examined at section 6, using a CBIR system.

6. EVALUATION

Quality evaluation of the extracted points of interest is done using the Multivariate Wald-Wolfowitz test [9], [11]. The ensemble of images utilized to produce the experimental results presented in this work is part of the COREL image collection [5]. The specific data set is a heterogeneous subset from the Corel gallery, including D=800 still colour images of 24bpp each, given in portable pixel map format of sizes [192×128] or [128×192] pixels. The entire Corel database contains a wide variety of images, from animals and plants to landscapes and natural images. The utilized data set was formed by pre-assigning the images into 13 distinct classes. 45 among these images included in the data set of the query images. In the experimental results the performance score was averaged across all queries.

The evaluation of an image retrieval system measures the efficiency and accuracy in information retrieval based on the computation of two quantities: precision (Pr) and recall (Re).

The performance of our point of interest detection technique was examined by varying the number of extracted colour features. Evaluation results are presented in Figure 3, indicating retrieval performance for varying region sizes.

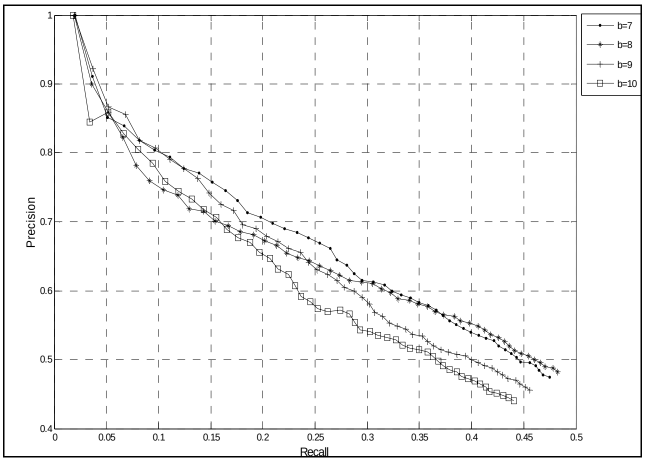


Figure 3: Comparison of retrieval performance via the Precision versus Recall diagram for varying region size.

Precision results can be derived from the measurements on the curves line according to the 10^{th} value, as the precision computed for the first 10 most similar images. Using a region's size b=7 the precision computed by the WW-test is 0.758 with 35 points of interest selected.

7. CONCLUSION

We have proposed a local feature extraction method from bimodal areas of an image that are considered as points of interest. The selection of points of interest is based on two non parametric algorithms: pairwise clustering algorithm, which clusters the data set of the region in the feature space and Hubert's test statistic, which detects coherent spatial organizations of chromatically similar pixels on the image plane. Both algorithms use as input the proximity data matrix of pixel colour values. The use of proximity matrix (only for small region's size) and the bipartition technique makes the method suitable for compact hardware implementation. The experimental results show that the presented feature extraction method, although it is based only on bimodal regions, can significantly contribute towards to an efficient visual retrieval system.

8. ACKNOWELEGMENTS

This work was financed by the European Social Fund (ESF), Operational Program for Educational and Vocational Training II (EPEAEK II), and particularly the Program "New graduate programs of University of Patras".

REFERENCES

[1] H. Shao, T. Svoboda, T. Tuytelaars and L. V. Gool, "HPAT indexing for fast object/ scene recognition based on local appearance", in *Int'l Conf. on Image Video Retrieval*, 2003, pp. 71-80.

[2] J. Wang, H. Zha and R. Cipolla, "Combining Interest Points and Edges for Content-based Image Retrieval", in *Proc. IEEE Int. Conf. on Image Processing*, 2005, pp. 1256-1259.

[3] V. Gouet and N. Boujemaa, "Object-based queries using color points of interest", in *Proc. IEEE Workshop on Content-based Access of Image and Video Libraries*, 2001.

[4] J. G. Matas, D. Koubaroulis and J. Kittler, "Colour Image Retrieval and Object Recognition Using the Multimodal Neighbourhood Signature", in *Proc. European Conf. on Computer Vision, LNCS*", 2000 vol. 1842, p. 48-64.

[5] Corel Stock Photo Library, Corel Corp., Ontario, Canada.

[6] Turi, R.H. and Ray S., "An application of clustering in colour image egmentation", (ICARCV 2000), Singapore, 5-8 December 2000, IEEE.

[7] S. Dubnov, R. E. Yaniv, Y. Gdalyahu, E. Schneidman, N. Tishby and G. Yona, "A new Nonparametric Pairwise Clustering Algorithm Based on Iterative Estimation of Distance Profiles", *Machine Learning*, 47, 35-61, 2002.

[8] J. T. Chien, C. H. Huang and S. J. Chen, "Compact decision trees with cluster validity for speech recognition", in *Proc. Int. Conf. on Acoustics, Speech and Signal Processing 2002*, May 13-17. 2002, vol. 1, pp. 873-1-873-6.

[9] J.H. Friedman and L.C. Rafsky , "Multivariate Generalizations of the Wald-Wolfowitz and Smirnov Two-Sample Tests", Annals of Statistics, vol. 7, no. 4, pp. 697-717, July 1979.

[10] K. Mikolajczyk , T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, L. Van Gool, "A comparison of Affine Region Detectors", in *Int. Journal of Computer Vision, January 2006.*

[11] Ch. Theoharatos, N. Lakaris,G. Economou and S. Fotopoulos, "A Generic Schemefor Color Image Retrieval based on the Multivariate Wald-Wolfowitz test", in *IEEE Transactions on Knowledge and Data Engineering* 17 (6), pp. 808-819, 2005.