An Integrated Approach of Transform Techniques & Subspace Model for Web Image Retrieval

Shrinidhi.S.Acharya[1]   Syeda Fiza[1]   Aruna kumari[1]   Monika.U[1]

[1]Dept of ECE, SJB Institute of Technology, Bangalore-560060, India.

shrinidhiacharya.s@gmail.com, fiza.syeda02@gmail.com.

Abstract— **In this paper, the problem of efficiently extracting low level features for image representation in content based image retrieval is addressed. Wavelet transform being very popular and effective in representing objects with isolated point singularities, but failed to represent line singularities. We introduce a novel framework to extract appropriate features like color, shape and texture representing edges and other singularities along lines which effectively fills the gap between high level and low level semantics. Ridgelet transform is applied on a segmented image and statistical features such as mean and standard deviation are extracted from each of the Ridgelet sub-bands there by generating feature vectors. PCA is applied to reduce the dimensionality of the data and Euclidian distance is used as similarity measure between database and query images. The algorithm also supports for the retrieval of similar images to the given query image from Google web server. Four benchmarking datasets are considered to conduct the experiment and result shows that the success rate is improved much when compared with traditional methods.**

Keywords— Low level features, semantic gap, line singularities, Ridgelet transform, feature vectors, Content Based Image Retrieval (CBIR), similarity measures, Principal Component Analysis (PCA).

I. INTRODUCTION

With the evolution of digital technology, there has been a significant increase in the use of digital images and pictures for storage and communication in electronic format. In this context, content-based image retrieval systems (CBIR) have become very popular for browsing, searching and retrieving images from a large database of digital images with minimum human intervention. The research community are competing for more efficient and effective methods as CBIR systems may be heavily employed in serving time critical applications in scientific and medical domains.

The visual features of images, such as color [1], texture [2][4], and shape features [3&5] have been analysed to represent and index image contents, color considered to be most dominant low level feature which occupies a small portion of total number of colors in the whole color space. Several color descriptors have been approved in the MPEG-7 Final Committee Draft [6], Jianmin et al. proposed a simple, low-cost and fast algorithm to extract dominant colour features directly in DCT domain without involving full decompression to access the pixel data [7]. Color histograms do not incorporate spatial adjacency of pixels in the image and may lead to inaccuracies in the retrieval. Shape information

based on histogram of significant edges contained in image can be extracted using canny edge operator, Further Anil K Jain et al. Proposed CBIR technique which used shape descriptors and color model [8].

Texture analysis, as recently been observed that different objects are best characterized by different texture methods which deals with the spatial distribution of gray values. Statistical features such as mean and standard deviation is extracted from the wavelet decomposed sub-bands, Apart from these statistical features, co-occurrence features also extracted in order to increase the correct classification rate [9]. Ruofei Zhang et al. [10] proposed effective and efficient content based image retrieval by presenting a novel indexing and retrieval methodology that integrates color, texture, and shape information for the indexing and retrieval. WangXing-yuan et al. [11] presents an effective color image retrieval method based on texture, which uses the colour co-occurrence matrix to extract the texture feature and measure the similarity of two color images. Wavelet features contain only a fixed number of directional elements, independent of scale [13], following this reason new sparse geometrical image representation such as ridgelets are identified as multiscale orientation-selective transforms [12]. Ridgelet transform allows representing edges and other singularities along lines in a more efficient way, in terms of compactness of the representation, than traditional transformations, such as the wavelet transform, for a given accuracy of reconstruction [14].

As World-Wide Web is growing in an exploding rate, search Engines become indispensable tools for any users who look for information on the Internet, and web image search is no exception. Web image retrieval has been explored and developed by academic researchers as well as commercial companies, including academic prototypes, additional search dimension of existing web search engines and web interfaces to commercial image providers [27]. Web search especially web image retrieval is one of the most needed and challenging task being faced nowadays as there is an enormous dependence on internet web in every walk of life based on technology.

The paper is organized as follows: In section II segmentation process, the theory of continuous ridgelet transform (CRT) and digital implementation aspect of ridgelet transform, Principal component analysis, indexing and retrieval of images from web are explained. Experimental results are discussed in Section III. Conclusions are drawn at end.

II. PROPOSED METHODOLOGY

In this paper, we have been motivated to assimilate 1-D wavelets along lines in radon domain because of the fact that singularities are habitually connected together along edges or contours in image. In this regard, we initially apply segmentation algorithm on four popular and widely used benchmarking image datasets (Caltech_101, Caltech_256, Corel_1k, Corel_10k) to obtain segmented image with straight edges. Later, Ridgelet transform is applied to construct feature vector by extracting orthogonal properties such as line singularities or edges in the segmented image. Finally, for classification purpose we use four different similarity distance measure techniques in reduced feature space using PCA. Following subsections furnishes the detailed explanation of segmentation as pre-processing, feature extraction and classification stages.

A. Segmentation

Image segmentation is a fundamental process in many image, video, and computer vision applications. Image segmentation is the process of partitioning a digital image into multiple segments. The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain visual characteristics. The result of image segmentation is a set of segments that collectively cover the entire image, or a set of contours extracted from the image. Each of the pixels in a region are similar with respect to some characteristic or computed property, such as color, intensity, or texture. It is considered to be a critical step towards content analysis and image understanding. Following steps are considered in segmentation process: 1. Smoothing: Blurring of the image to remove noise. 2. Finding gradients: The edges should be marked where the gradients of the image has large magnitudes. 3. Non-maximum suppression: Only local maxima should be marked as edges. 4. Double thresholding: Potential edges are determined by thresholding. 5. Edge tracking by hysteresis: Final edges are determined by suppressing all edges that are not connected to a very certain (strong) edge as shown in Fig. 1.

Fig. 1. Original and segmented image

B. Feature Extraction

As we all know, an image contents basically configure about color, texture, shape, region etc...Extraction of these contents from an image is possible through various methodologies and the process is termed as feature extraction.

1) Wavelets

Wavelets are used to detect the zero dimensional point singularities in an image. Wavelet transform is used for the time-frequency analysis, which is essential for singularity detection. The wavelet transform can be expressed by the following equation:

$$F(a, b) = \int_{-\infty}^{\infty} f(x) \bigcup_{(a,b)}^{*}(x)dx \qquad (1)$$

Where the '*' is the complex conjugate symbol and function ψ is some function. This function can be chosen arbitrarily provided that obeys certain rules. When processing a 2-D image, the wavelet analysis is performed separately for the horizontal and the vertical directions. Thus, the vertical and the horizontal edges are detected separately. The 2D discrete wavelet transform (DWT) decomposes the images into sub-images, 3 details and 1 approximation. The approximation looks similar to the input image but only 1/4 of original size. The 2-D DWT is an extension of the 1-D DWT in both the horizontal and the vertical direction. We label the resulting sub-images from an octave as LL (the approximation or we say the smoothing image of the original image which contains the most information of the original image), LH (preserves the horizontal edge details), HL (preserves the vertical edge details), and HH (preserves the diagonal details which are influenced by noise greatly), according to the filters used to generate the sub-image. HL means that we used a high pass filter along the rows, and a low pass filter along the columns whereas in LH, low pass filter is applied along the rows, and a high pass filter along the columns. This process can repeat continuously by putting the first octave's LL sub-image through another set of low pass and high pass filters. These iterative procedures construct the multi-resolution analysis.

2) Radon Transforms

The Radon transform is used to detect features within an image. Given a function $A(x, y)$, the Radon transform is defined as:

$$R(\theta, \rho) = \int_{-\infty}^{\infty} A(\rho\cos\theta - s\sin\theta, \rho\sin\theta + s\cos\theta)\, ds \quad (2)$$

Equation 2 describes the integral along a line s through the image, where p is the distance of the line from the origin and theta is the angle from the horizontal. Radon transform in two dimensions, is the integral transform consisting of the integral of a function over straight lines. Radon further included formulas for the transform in three-dimensions, in which the integral is taken over planes.

The Radon transform computes projections of an image matrix along specified directions. A projection of a two-dimensional function f (x, y) is a set of line integrals. The Radon function computes the line integrals from multiple sources along parallel paths, or beams, in a certain direction. The beams are spaced 1 pixel unit apart. The radon function takes multiple, parallel-beam projections of the image from different angles by rotating the source around the center of the image.

Radon transform basically has properties like:

1) Symmetry. 2) Linearity. 3) Rotation. 4) Shifting. 5) Scaling. 6) Convolution.

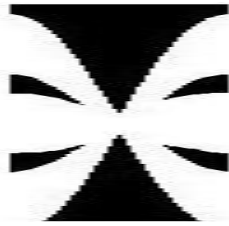Fig. 2 shows the output when radon transform is applied on a segmented image.

Fig. 2. Result obtained applying radon on segmented image

3) *Ridgelet transform*

In this paper, we propose a discrete Ridgelet transform that achieves both invertibility and non-redundancy. In fact, our construction leads to a large family of orthogonal and directional bases for digital images, including adaptive schemes. As a result, the inverse transform is numerically stable and uses the same algorithm as the forward transform. Given an integrable bivariate function f(x), its continuous Ridgelet transform (CRT) in R^2 is defined by,

$$CRT_f(a, b, \theta) = \int_{R^2} \Psi_{(a,b,\theta)}(x) f(x) dx \qquad (3)$$

Where the Ridgelet $\Psi_{(a, b, \theta)}(x)$ in 2-D are defined from a Wavelet-type function in 1-D $\Psi(x)$ as

$$\Psi_{(a,b,\theta)}(x) = a^{-1/2} \Psi((x_1 \cos\theta + x_2 \sin\theta - b)/a) \qquad (4)$$

Fig. 3 shows an example of Ridgelet function, which is oriented at an angle θ and is constant along the lines $x_1 \cos\theta + x_2 \sin\theta = $ const.
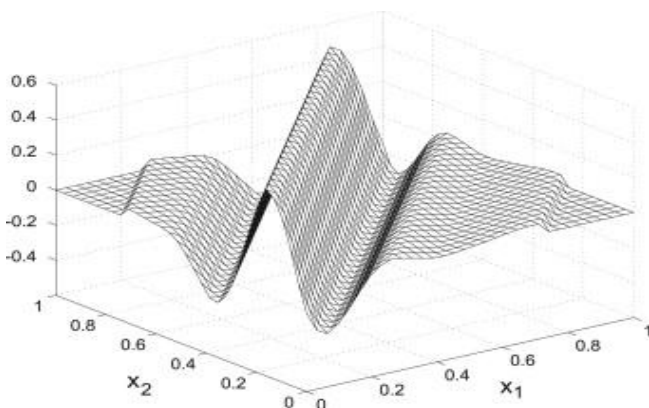

Fig. 3. An example Ridgelet function $\psi_{a, b, \theta} = (x_1, x_2)$.

C. *Principal Component Analysis (PCA):*

The problem of excessive dimensionality can be reduced by projecting high dimensional data onto a lower dimensional space that best represents the data in a lest-squares sense.

Let a set of n d-dimensional samples $X_1, X_2 .... X_n$ represented by a single vector X_0 such that squared error criterion function is as small as possible given by:

$$J_0(X_0) = \sum_{k=1}^{n} \| X_0 - X_k \|^2 \qquad (5)$$

One dimensional representation of the dataset onto a line running through the sample mean (m) can be written as X = m+ae, where a is a unit vector in the direction of the line. To find the best direction of the line we compute scatter matrix 'S' which is given by

$$S = \sum_{k=1}^{n} (X_k - m)(X_k - m)^t \qquad (6)$$

We now project the data onto a line through the sample mean in the direction of the Eigen vector of the scatter matrix having the largest Eigen value. For d^l – dimensional data projection we have

$$X = m + \sum_{i=1}^{d^l} a_i e_i \qquad (7)$$

coefficients a_i are called as principal coefficients, geometrically if we plot the data points $X_1, X_2 .... X_n$ as forming a d-dimensional, then the Eigen vectors of the scatter matrix are the principal axes of that hyper ellipsoid.

In our proposed methodology we have adopted two phases, training phase and testing phase. Training phase consists of the images present in the datasets and Test phase comprises of the query image which is the interested image for retrieval.

Training phase Algorithm:
Step 1: Read the RGB image.
Step 2: Apply segmentation technique.
Step 3: Convert the image to grayscale.
Step 4: Resize the image to 50*50 pixels.
Step 5: Apply the 2D Ridgelet Transform
Step6: Generate feature vector for training images.

Testing phase Algorithm:
Step1: Read the query image.
Step2: Apply segmentation technique.
Step3: Convert the image to grayscale.
Step4: Resize the image to 50*50 pixels.
Step5: Apply the 2D Ridgelet Transform.
Step6: Generate feature vector for query image.
Step7: Calculate the Euclidean distance for query image and trained dataset.
Step8: Increment the count for every image detected & calculate Recognition Accuracy.

D. *Indexing and Web image retrieval:*

Once the transform technique is successfully implemented and tested for large image datasets. We manually annotated the dataset images with its category and created a log file. For effective retrieval we have to process the image and store their corresponding feature vectors in the log file. Query Image is

processed and Feature vector is obtained. If the obtained Feature Vector matches the log file then corresponding query image is converted into text and passed to a search engine.

III. EXPERIMENTAL RESULTS AND DISCUSSIONS

As said earlier, wavelets are good at catching zero dimensional point singularities but fail to extract line singularities in higher dimensions. So to overcome this disadvantage of wavelets, Ridgelet transform is introduced. Here line singularities in frequency domain is converted to point singularities in radon domain. Now by applying wavelet transform, the point singularities are efficiently extracted.

By implementing the below procedure in Matlab R2007b installed in a system of Processor Intel(R) core(TM) i5-2450M CPU @2.50GHz with 4.00GB memory, the Results specified below for different datasets are obtained.

This section presents the results based on four popular and widely used benchmarking image datasets: Caltech-101, Caltech-256, Corel-1k and Corel-10k consisting of various object categories.

We initialize our system by reading the RGB image and applying 2-D Ridgelet transform. Ridgelet features are collected from segmented image by applying 1-D wavelet in radon domain and vectorizing each image coefficients from 2-D to 1-D. The resultant feature vector is extensively high (10,000 elements per image in a vector), so in order to reduce the dimension we use PCA technique. Further, for classification four different distance measure techniques are applied in reduced feature space to measure the similarity between train feature dataset and query feature vector.

As per the standard procedural settings, we divided the entire dataset into 15 and 30 images/category as training phase and remaining images for testing. We obtained an average of per class recognition under each category for all runs in each stage. Performance analysis for well-known techniques & benchmarking datasets are explained in the following sub-sections.

By taking Caltech-101 as example, below lies the various subbands on which the accuracy values are calculated and estimated. The accuracy values are found to be high in the LH and HL subbands which represents the vertical edge details and horizontal edge details of an image respectively. So the results for other datasets are calculated for the two subbands of LH & HL which lies in the range of 7001 to 17000 wavelet coefficients.

A. Caltech 101

Caltech – 101 dataset comprising of 9,159 images of 101 different categories of natural scenes (animals, butterfly, chandelier, Garfield, cars, flowers, human face, etc.). Each category consisting images in the range of 31 to 800 with most of the images centered, occluded, affected by corner artifacts, and with large intensity variations has made this dataset most challenging [18].
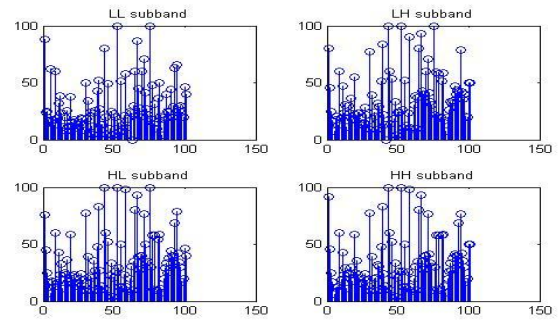

Fig. 4. Histograms of wavelet for different subbands.

An experimental results exhibited in TABLE I. proves that the proposed model of Ridgelet transform with PCA is superior compared with the most popular techniques found in the literature considering similar dataset and experimental procedures.

With reference to the TABLE I. We notice that proposed Ridgelet with PCA has obtained classification rates of 38.20% and 47.95% for 15 & 30 trained images per category respectively in comparison with conventional classifiers mentioned in [16, 17, and 18].

In our experiment, we have randomly tested 5800 for 15 train images and 4600 for 30 train images selected from all categories, ranging between 15 to 200 images in each.

TABLE I. Performance analysis for Caltech_101 dataset in %.

Method	15 train images	30 train images
Serre et al.[16]	35	42
Holub et al.[17]	37	43
Berg et al.[18]	45	-
Proposed	38.20	47.95

Results of the proposed methodology are found to be superior compared to the results cited in [16 & 17] and competitive with Berg et al [18].

B. Caltech_256

Caltech_256 is a challenging set of 256 object categories containing 29768 images in total [21]. Caltech–256 was generated by downloading Google images and manually screening out with different object classes. It presents high variations in intensity, clutter, object size, location, pose, and also increased the number of category with at least 80 images per category.

In order to evaluate the performance of proposed method we followed the standard experimental procedure [20] of labeling first 15 & 30 images per category for generating train feature vector and remaining images as test. But in our algorithm we have randomly considered the test images in each category not exceeding 50 in number for 30 train images and not exceeding 65 images per category for 15 train images. TABLE II. Shows the performance analysis of the Caltech-256 dataset compared with other methodologies.

TABLE II. Performance analysis of Caltech_256 dataset in %.

Method	15 train images	30 train images
Van et al.[19]	-	27.17
Griffin et al.[20]	28.3	34.1
Sancho et al.[21]	33.1	40.1
Proposed	29.60	37.96

Results of our proposed method is found with leading recognition accuracy when compared to most popular techniques proposed in [19 & 20] and is competitive compared with Sancho et.al [21].

Fig. 5 illustrates few samples of Caltech-101 and Caltech-256 datasets with high classification rate for 30 train images per category.
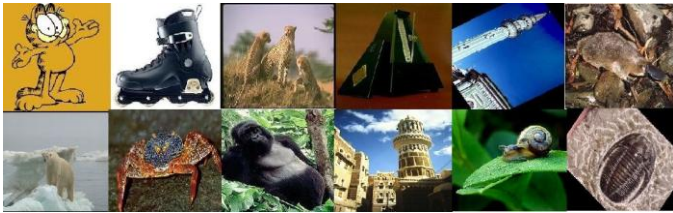

Fig. 5. Sample images of Caltech dataset with high classification rate (>80%).

C. Corel

Dataset consists of 1000 images been chosen from the Corel photo gallery and grouped into 10 categories with 100 similar images each and a dataset with 9788 images grouped under 108 different categories. There are many algorithms and experiments being carried out on this dataset in order to learn more about its specifications. As proposed in [22] distributions of the segments from the different semantic classes can be seen to be highly concentrated on different SOM surfaces.

TABLE III. Performance analysis of Corel dataset in %.

Method	15 train images	30 train images
Gerald[24]		
Histogram intersection	73	75
Spatial chromatics	74	75
Henning[25]	83(for 20 train images)	
Proposed		
Corel_1k	63	67
Corel_10k	18	22

Note: In the methods mentioned above, they have considered the Corel datasets with categories varying from 30 to 62 in number.

Fig. 6 represents some of the sample images of Corel dataset with high recognition accuracy for 30 train images.
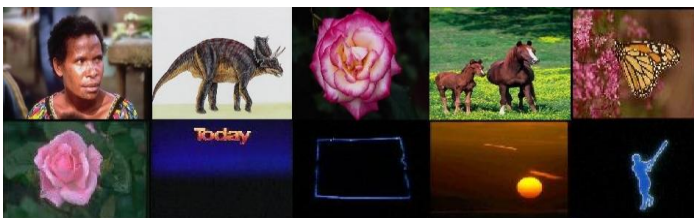

Fig. 6. Sample images of Corel dataset with high classification rate (80%).

In [25] a dataset with 61 categories each with 100 images is considered and different algorithms are implemented and the retrieval results are found to be up to 83%. Hence we can declare that our proposed method, which has considered 108 categories with each having up to 100 images have got the leading results compared to other popular methods which is rated to be 67 and 22% for 1k and 10k datasets respectively.

After the performance analysis of the ridgelet algorithm on all the datasets, the query image is tested to be retrieved from the Google web server. Similar page in the browser (as shown in Fig. 7) is displayed for given query images. 7(a) represents the input query images and 7(b) represents the resultant search images got from Google web dataset.
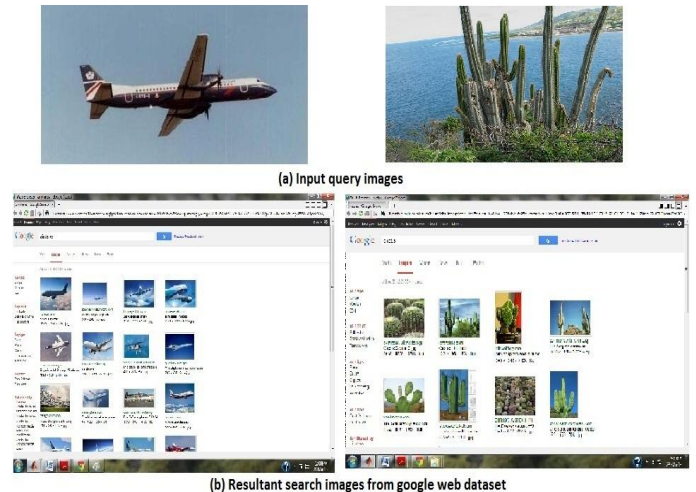

Fig. 7. (a) Input query images. (b)Resultant search images from Google web dataset.

IV. CONCLUSION

In this paper Ridgelet transform which deal effectively with line singularities in 2-D is introduced. It allows representing edges and other singularities along lines in a more efficient way compared to wavelet transform and also the issue of texture classification based on Ridgelet transform has been analyzed and proved to be superior compared to other conventional methods. We also demonstrated our method on Web image retrieval application by successfully retrieving similar images from the Google dataset. But for complex images, where edges are mainly along curves and there are texture regions (which generate point discontinuities), the Ridgelet transform is not optimal. Another scheme is to use the Ridgelet transform as the building block in a more localized construction such as the curvelet transform.

REFERENCES

[1] Stehling, R. O., Nascimento, M. A., and A. X . Falcao . On Shapes of Colors` for Content-based Image Retrieval. In ACM International Workshop on Multimedia Information Retrieval (ACM MIR'00), 2000, 171-174.

[2] M. Flickner et al., Query By Image and Video Content: The QBIC System. IEEE Computer, 28, 9 (1995), 23-32.

[3] Jing, ,M. Li,H. J. Zhang and B. Zhang, An Effective Region-based Image Retrieval Framework, In Proceedings of the Tenth ACM international conference on Multimedia, 2002, 456-465.

[4] Sami Brandt, Jorma Laaksonen and Erkki Oja, Statistical Shape Features in Content-Based Image Retrieval, 2000 IEEE, PP. 1062-1065.

[5] M. Safar, C. Shahabi and X. Sun, Image Retrieval by Shape: A Comparative Study, In Proceedings of IEEE International Conference on Multimedia and Expo (ICME'00), 2000, 141-144.

[6] ISO/IEC 15938-3/FDIS Information Technology—Multimedia Content Description Interface—Part 3 Visual Jul. 2001, ISO/IEC/JTC1/SC29/WG11 Doc. N4358.

[7] Jianmin Jiang, Ying Weng, PengJie Li, Dominant colour extraction in DCT domain, Image and Vision Computing 24 (2006) 1269–1277.

[8] Anil K Jain, Aditya Vailaya, Image Retrieval using Color and Shape , Department of Computer Science Michigan State University, May 1995.

[9] Arivazhagan, S., 2004. Some studies on texture image analysis using wavelet transform and its usage in few industrial applications. Ph.D. Thesis.

[10] Ruofei Zhang, Zhongfei (Mark) Zhang,A Clustering Based Approach to Efficient Image Retrieval, Proceedings of the 14th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'02), 2002.

[11] WangXing-yuan , ChenZhi-feng, YunJiao-jiao, An effective method for color image retrieval based on texture, Computer Standards & Interfaces Volume 34 (2012) Pages 31–35.

[12] E.J. Cand`es and D.L. Donoho. Ridgelets: the key to high dimensional intermittency? Philosophical Transactions of the Royal Society of London A, 357:2495–2509, 1999.

[13] E.J. Cand`es. Ridgelets: theory and applications. PhD thesis, Stanford University, 1998.

[14] Patrizio Campisi, Alessandro Neri, Gaetano Scarano, 2002. Model based rotation invariant texture classification. In: IEEE International Conference on Image Processing, pp. 117–120

[15] Kristen Grauman and Trevor Darrell, Pyramid Match Kernels: Discriminative Classification with Sets of Image Features (version 2),Massachusetts Institute of Technology,Computer Science and Artificial Intelligence Laboratory,Cambridge, MA, USA

[16] T. Serre, L. Wolf, T. Poggio. Object recognition with features inspired by visual cortex.(2005). In CVPR, San Diego.

[17] A. Holub, M. Welling, P. Perona. Exploiting unlabelled data for hybrid object classification. (2005). In NIPS Workshop on Inter-Class Transfer, Whistler, B.C.

[18] Hao Zhang, Berg A.C, Maire. M, Malik, J, SVM-KNN: Discriminative Nearest Neighbor Classification for Visual Category Recognition. (2006). IEEE-CVPR, vol-2, pp 2126-2136.

[19] J. C. Van Gemert, J.-M. Geusebroek, C. J. Veenman, A. W. M. Smeulders. Kernel codebooks for scene categorization. In ECCV, 2008.

[20] G. Griffin, A. Holub, P. Perona. Caltech 256 object category dataset. (2007). Technical Report UCB/CSD-04-1366, California Institute of Technology.

[21] Sancho Mc Cann and David G. Lowe, Local Naive Bayes nearest Neighbor for Image Classification. (2012). IEEE CVPR. pp 3650 – 3656.

[22] Jorma Laaksonen, Ville Viitaniemi and Markus Koskela EMERGENCE OF SEMANTIC CONCEPTS IN VISUAL DATABASES, Neural Networks Research Centre, Helsinki University of Technology, P.O.Box 5400 , FI-02015 TKK, FINLAND.

[23] Serge Belongie, Chad Carson, Hayit Greenspan, and Jitendra Malik, Recognition of Images in Large Databases Using a Learning Framework Computer Science Division, University of California at Berkeley, CA 94720.

[24] Gerald Schaefer1 and Michal Stich1,2 UCID – An Uncompressed Color Image Database, School of Computing and Mathematics, The Nottingham Trent University, Nottingham, United Kingdom.

[25] HenningM¨uller, Stephane Marchand-Maillet, Thierry Pun, The truth about Corel - evaluation in image retrieval,Computer Vision Group, University of Geneva.

[26] Nidhi Singh,Kanchan Singh, Ashok K.Sinha, A Novel Approach for Content Based Image Retrieval, Procedia Technology4 (2012) 245 – 250,C3IT-2012.

[27] Wei-Hao Lin, Rong Jin, Alexander Hauptmann, Web Image Retrieval Re-Ranking with Relevance Model, Language Technologies Institute, School of Computer Science, Carnegie Mellon University,Pittsburgh, PA, 15213 U.S.A.