International Journal on Future Revolution in Computer Science & Communication Engineering ISSN: 2454-4248
Volume: 3 Issue: 11 272 – 278

An Efficient Perceptual of Content Based Image Retrieval System Using SVM and Evolutionary Algorithms

[1*] Dr. M. Ekambaram Naidu,[2]Ravi Kumar Chandu
[1]Professor, dept.of CSE, SRK Institute of Technology, Vijayawada, Andhra Pradesh,
[2] Research Scholar, Rayalaseema University, Kurnool, Andhra Pradesh, India

Abstract: The CBIR tends to index and retrieve images based on their visual content. CBIR avoids several issues related to traditional ways that of retrieving images by keywords. Thus, a growing interest within the area of CBIR has been established in recent years. The performance of a CBIR system mainly depends on the particular image representation and similarity matching operate utilized. The CBIR tends to index and retrieve images supported their visual content. CBIR avoids several issues related to traditional ways that of retrieving images by keywords. Thus, a growing interest within the area of CBIR has been established in recent years. The performance of a CBIR system principally depends on the actual image illustration and similarity matching operate utilized. therefore a replacement CBIR system is projected which can give accurate results as compared to previously developed systems. This introduces the new composite framework for image classification in a content-based image retrieval system. The projected composite framework uses an evolutionary algorithm to select training samples for support vector machine (SVM). to style such a system, the most popular techniques of content-based image retrieval are reviewed initial. Our review reveals some limitations of the existing techniques, preventing them to accurately address some issues.

Keywords*: IR, CBIR (Content Based Image Retrieval), CNN, SVM, Evolutionary Algorithm.*

*****

I. INTRODUCTION

In the domain of content-based image retrieval (CBIR), the retrieval accuracy is essentially based on the discrimination quality of the visual features extracted from images or small patches. Image contents (objects or scenes) may include different deformations and variations, e.g. illumination, scaling, noise, viewpoint, etc, which makes retrieving similar images one of the challenging vision tasks. The typical CBIR approaches consist of three essential steps applied on images: detection of interest points, formulation of image vector, and Similarity/dissimilarity matching A CBIR (content based image retrieval) system performs data mining tasks (search, ranking, classification, etc) on images using the visual content of images only. The CBIR systems are great when dealing with images: without contextual information or meta-data associated to the image. This tends to happen a lot nowadays, with the number of images produced and uploaded on the Internet. Indeed, billions of photos were posted on Internet by now, and the number of public photos uploaded to the Internet everyday is estimated at 2 millions. This thesis introduces new composite framework for image classification in a content based image retrieval system. The proposed composite framework uses an evolutionary algorithm to select training samples for support vector machine (SVM). To design such a system, the most popular techniques of content based image retrieval are reviewed first. Our review reveals some limitations of the existing techniques, preventing them to accurately address some problems.

Then Evolutionary algorithms are identified as an interesting solution to solve the fine grained classification problem. The performances of Evolutionary algorithms for the problem considered are not good enough to use them alone to solve it though. Thus using those together with another algorithmic considered, forming what is called a composite system. Using an evolutionary algorithm to provide training samples to a support vector machine (SVM) seemed to be new in CBIR. This is the first implementation of such method that leads to promising results.

II. LITERATURE REVIEW

Content-based image retrieval at the end of the early years - A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R.Jain - 2000

Presents a review of 200 references in content-based image retrieval. The paper starts with discussing the working conditions of content-based retrieval: patterns of use, types of pictures, the role of semantics, and the sensory gap. Subsequent sections discuss computational steps for image retrieval systems. Step one of the review is image processing for retrieval sorted by color, texture, and local geometry. Features for retrieval are discussed next, sorted by: accumulative and global features, salient points, object and shape features, signs, and structural combinations thereof. Similarity of pictures and objects in pictures is reviewed for each of the feature types, in close connection to the types and means of feedback the user of the systems is capable of giving by interaction. We briefly discuss aspects of system engineering: databases, system architecture, and evaluation. In the concluding section, we present our view on: the driving force of the field, the heritage from computer vision, the influence on computer vision, the role of similarity and of interaction, the need for databases, the problem of evaluation, and the role of the semantic gap.

Content Based Image Retrieval based on Color, Texture and Shape features using Image and its complement - P. S. Hiremath and JagadeeshPujari - 2007

Color, texture and shape information have been the primitive image descriptors in content based image retrieval systems. This paper presents a novel framework for combining all the three i.e. color, texture and shape information, and achieve higher retrieval efficiency using image and its complement. The image and its complement are partitioned into non-overlapping tiles of equal size. The features drawn from conditional co-occurrence histograms between the image tiles and corresponding complement tiles, in RGB color space, serve as local descriptors of color and texture. This local information is captured for two resolutions and two grid layouts that provide different details of the same image. An integrated matching scheme, based on most similar highest priority (MSHP) principle and the adjacency matrix of a bipartite graph formed using the tiles of query and target image, is provided for matching the images. Shape information is captured in terms of edge images computed using Gradient Vector Flow fields. Invariant moments are then used to record the shape features. The combination of the color and texture features between image and its complement in conjunction with the shape features provide a robust feature set for image retrieval. The experimental results demonstrate the efficacy of the method.

Evaluating bag-of-visual-words representations in scene classification - J. Yang, Y.-G. Jiang, A. G. Hauptmann, and C.-W. Ngo - 2007

Based on key points extracted as salient image patches, an image can be described as a "bag of visual words" and this representation has been used in scene classification. The choice of dimension, selection, and weighting of visual words in this representation is crucial to the classification performance but has not been thoroughly studied in previous work. Given the analogy between this representation and the bag-of-words representation of text documents, we apply techniques used in text categorization, including term weighting, stop word removal, feature selection, to generate image representations that differ in the dimension, selection, and weighting of visual words. The impact of these representation choices to scene classification is studied through extensive experiments on the TRECVID and PASCAL collection. This study provides an empirical basis for designing visual-word representations that are likely to produce superior classification performance.

Delving Deep into Rectifiers: Surpassing Human-Level Performance on Image Net Classification - He K, Zhang X, Ren S, et al. - 2015

Rectified activation units (rectifiers) are essential for state-of-the-art neural networks. In this work, we study rectifier neural networks for image classification from two aspects. First, we propose a Parametric Rectified Linear Unit (PReLU) that generalizes the traditional rectified unit. PReLU improves model fitting with nearly zero extra computational cost and little over fitting risk. Second, we derive a robust initialization method that particularly considers the rectifier nonlinearities.

Deeply learned face representations are sparse, selective, and robust - Sun Y, Wang X, Tang X. - 2014

This paper designs a high-performance deep convolution network (DeepID2+) for face recognition. It is learned with the identification-verification supervisory signal. By increasing the dimension of hidden representations and adding supervision to early convolution layers, DeepID2+ achieves new state-of-the-art on LFW and YouTube Faces benchmarks. Through empirical studies, we have discovered three properties of its deep neural activations critical for the high performance: sparsity, selectiveness and robustness. (1) It is observed that neural activations are moderately sparse. Moderate sparsity maximizes the discriminative power of the deep net as well as the distance between images.

Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions - Andoni A, Indyk P. - 2008

In this article, we give an overview of efficient algorithms for the approximate and exact nearest neighbor problem. The goal is to preprocess a dataset of objects (e.g., images) so that later, given a new query object, one can quickly return the dataset object that is most similar to the query. The problem is of significant interest in a wide variety of areas.

Hamming distance metric learning - Norouzi M, Blei D M, Salakhutdinov R. - 2012

Motivated by large-scale multimedia applications we propose to learn mappings from high-dimensional data to binary codes that preserve semantic similarity. Binary codes are well suited to large-scale applications as they are storage efficient and permit exact sub-linear kNN search. The framework is applicable to broad families of mappings, and uses a flexible form of triplet ranking loss. We overcome discontinuous optimization of the discrete mappings by minimizing a piecewise-smooth upper bound on empirical loss, inspired by latent structural SVMs. We develop a new loss-augmented inference algorithm that is quadratic in the code length. We show strong retrieval performance on CIFAR-10 and MNIST, with promising classification results using no more than kNN on the binary codes.

III. RELETD WORK

An image Retrieval (IR) system is a system allowing to perform different data mining tasks on images from a database (classification, retrieval, copy detection, etc). IR systems may use any kind of data available (keywords, GPS

ISSN: 2454-4248
272 – 278

location, website from which the image is extracted, etc) to perform those tasks.

CBIR system is an IR system that uses only the visual content of images as input data. This means that no keywords or any meta-data are used during the process, only information extracted from the pixels of the images is used.

Common tasks that are performed in CBIR.

- Face Detection
- Copy Detection
- Object Detection and Classification
- Image Retrieval
- Image Classification

3.1 General Structure

CBIR systems rely only on information contained in the pixels of the images. The system is expected to find similar images at a semantic level; pixels do not provide such information and do not allow accessing to such a level of abstraction easily. The first task when creating a CBIR system is designing a representation of images, as shown at the top of Figure 1.Therepresentation that holds meaningful information, is extracted for each image in the database. By reducing the size of the description, the cost and complexity of similarity computations can be reduced. Thus the cost of the computation of the decision taken by the Machine Learning (ML) algorithm. This algorithm is used as the search engine in the CBIR system of Figure 1.

Once the representation of each image is extracted from the database, the search engine is learned. The search engine is able to assign images into relevant classes.

Then, the trained Machine Learning (ML) algorithm is applied to the representation of every image of the dataset, assigning them a label. The dataset is then completely classified, and the trained ML algorithm can be applied to classify any new image that is added to the dataset. This concept is shown at bottom of Figure 1.

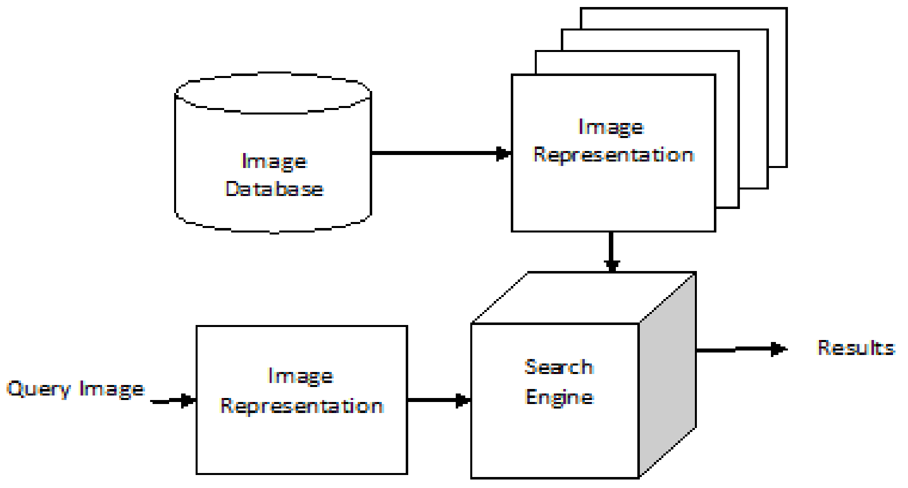

Figure 1

3.2 Comparing Images with Similarity Measures

To be able to compare images, the representation must come with a similarity measure. The similarity measure depends on the image representation. The subsequent sections describe comparison methods.

A. **Histogram Comparison:** Histograms are usually real valued vectors in the image representation context. As such, they can be compared using usual distance functions. More specific measures can bemire suitable to compare vectors. The following sections describe these measures.

B. **Quadratic-Form distance**: The first distance to compare two histograms P and Q is the Quadratic-Form (QF) distance. This measure compares histograms by comparing bins at the same position between each other. Such comparisons assume that the domain of the histograms, the feature vector represented by each bin, is aligned. This condition is most of the time not fulfilled, due to the per image quantization of feature vectors.

C. **Chi-Squared distance**: The Chi-Squared ($\chi 2$) is a famous histogram distance. It is based on the$\chi 2$ statistic test. It considers relative differences instead of absolute differences, thus giving more importance to differences between small bins over differences between large bins.

D. **Bags of Features Comparison**: The BoFs are sometimes used as image representation directly, in particular in the detection of objects covering only a part of an image. In order to compare BoFs, each vector of each BoF must be compared to one another. Those comparisons must then be combined to form a final score. There are several strategies to compute the score from comparison. The voting strategies, for example, are described here.

E. **Voting Strategies**: The voting strategies are fairly simple, and any feature vector within the neighborhood of another has the same contribution to the similarity measure. This is not suitable in CBIR.

F. **Kernels Function**: These are dedicated to any image representation composed of one vector, histogram or not, and are widely used because of their properties. Let *X*be the space in which the raw visual features are defined, also called the input space. There are two ways to define kernel function. The first one is: if there exists an injection φ mapping any $\mathbf{x} \in X$ to vector $\varphi(\mathbf{x})$in an Hilbert space, called the feature space; then, the function k defined by Equation1.1 is a kernel function. K is defined by a dot product in the feature space.

$$k(\mathbf{x}, \mathbf{y}) = (\varphi(\mathbf{x}), \varphi(\mathbf{y})) \qquad (1.1)$$

As kernel functions are dot products in a feature space, they can be used instead of the dot product in the input space to build a decision using those linear classification algorithms. The linear separation is then built in the feature space instead of in the input space. The greater the dimension of the feature space, the higher the probability that the data can be linearly separable in it. This is the so called kernel trick. Then finding the linear separation in the feature space results in finding the non-linear separation in the input space. The computing a kernel function is most of the time not very costly. The kernels have been used a lot in classification tasks and in CBIR and in most of the work using SVMs as the classification algorithms.

I.3 LIMITATIONS OF EXSTING TECHNIQUS

The actual active learning strategies for SVM suffer from one drawback in CBIR. The existing active learning techniques seem to be lacking something to address the fine grained classification problem: exploration. Exploration is the fact of searching areas of the search space that have not been looked for yet. The fine grained classification (exploration) is one of the problems in CBIR.

IV. PROPOSED WORK

This section provides the understanding about the proposed system which is required to develop for the demonstration of the performance study of IR system EAs are population-based meta-heuristic optimization algorithms. An optimization problem consists in finding the parameter vectors for which the optimal of one or several functions are attained. The function to optimize is often called a fitness function or an objective in Eas. Eas cover a wide range of algorithms, including but not limited to Gas , Particle Swarm Optimization (PSO), Differential Evolution(DE) or even firefly algorithm.

4.1 General Structure of EA

The population is initialized first. Once the population is initialized, each individual in the population is evaluated using the fitness function before the evolutionary loop starts. The evolutionary loopis composed of a pre-variation selection process, a variation process, an evaluation process and a post-variation selection process.

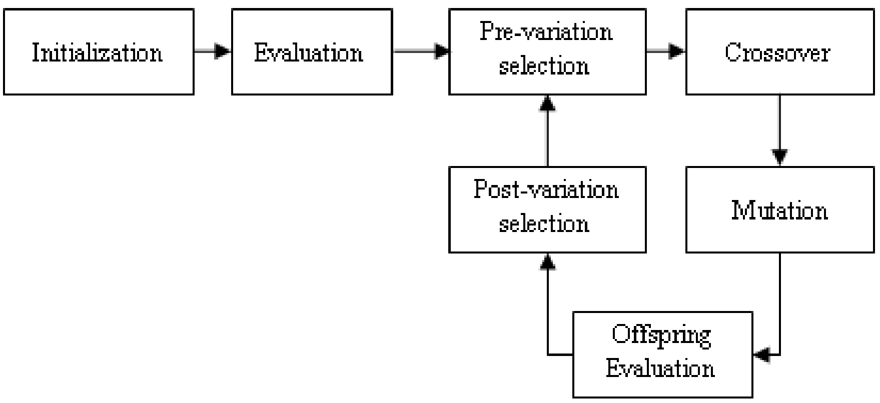

Figure 2

ISSN: 2454-4248
272 – 278

4.2 Multi-Objectives Optimization Problems:

Multi-Objectives Optimization Problems consist in optimizing two or more functions, or objectives, at the same time. The formal notation is shown in definition 1.

Definition 1. Multi-Objective Optimization Problem

Let $F = [f1, f2, \ldots, fn]$,

$minF(x), x \in S$ is a multi-objective optimization problem.

When dealing with such problems, the comparison of individual's performances is harder. Comparison relation Pareto dominance has been built to compare individuals in the objective space in the context of MultiObjectivesOptimization Problems.

Pareto dominance: It basically says that an individual *a* is better than an individual *b* if and only if it is at least as good as *b* in all objectives and strictly better than *b* in at least one objective. If *a* is better than *b* in some objectives and worse in others, then *a* and *b* are considered to be not comparable with respect to the Pareto dominance.

This dominance relation allows comparing individuals between each others, and even allows ranking them. This ranking is important as EAs need a way to know which individuals are better than other during the evolution process and in particular for the selection steps. This ranking is called the non dominated sort.

None dominated sort: As the Pareto dominance is a partial order relation, several individuals may not be comparable. Therefore, a population of individuals is ranked into subsets of incomparable individuals called fronts which are assigned a rank. This ranking process is called the non-dominated sort. It is often used by EAs dedicated to Multi-Objectives Optimization Problems to compute the probability for each individual to be selected during the two selection steps.

When solving a Multi-Objectives Optimization Problem, it is expected that a diverse setoff tradeoffs between objectives is found; therefore, metrics have been designed to measure the diversity of trade-offs presented by a front, so as to rank individuals within a same front by their contribution to the diversity of solutions. Several measures exist; the commonly used measures are crowding distance and the hyper-volume.

Given those tools, Non-dominated Sorting Genetic Algorithm II (NSGA-II) which is well known and quite effective GA, is created to tackle MultiObjectives Optimization Problems. The other algorithms are Parameter less Inching Assisted Non-dominated Sorting Genetic Algorithm-II (PNANSGA-II), Biobjective Differential Evolution (MOBiDE), and Multiobjective Optimization for locating multiple optimal Solutions of Multi-modal Optimization Problems (MOMMOP).

NSGA-II: NSGA-II is a GA dedicated to Multi-Objectives Optimization Problems.The environmental selection is the only step that can't be modified in NSGAII's workflow; every thingelse can be tuned to fit any problem.

4.3 composite system design:

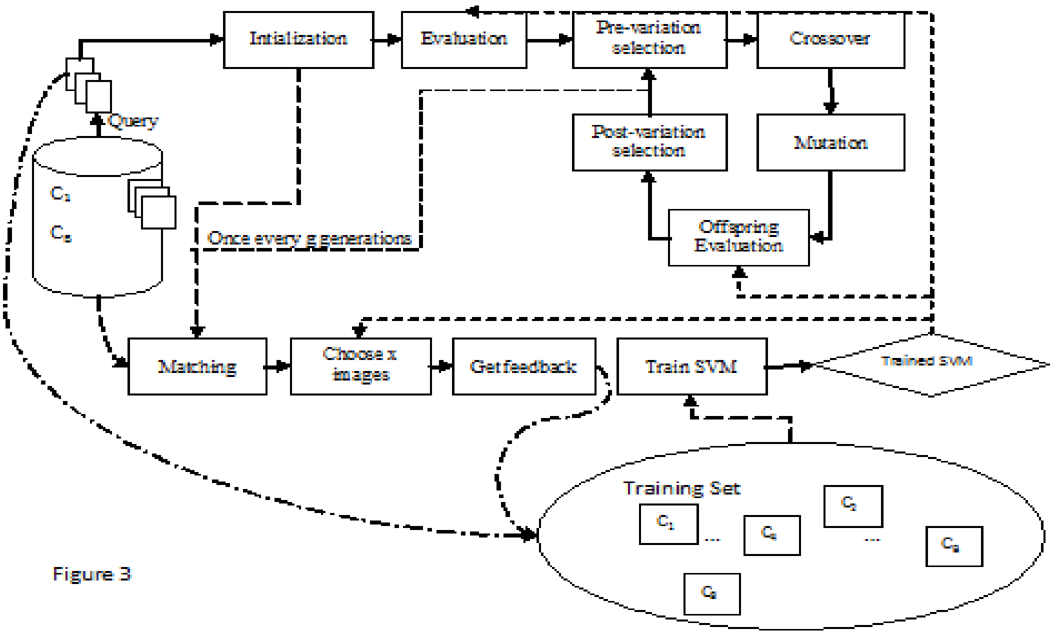

Figure 3

4.3.1Initializing the System: To initialize the learning process, the system needs one query image from each category. So *B* images are drawn from the set of available training samples, one from each category, to form the query images. Those query images are added as the first samples to the training set of the SVM. Then, the initialization process of the EA takes place.

4.3.2 The Relevance Feedback Process: Before the GA loop starts, the relevance feedback process is run, adding some images to the training set and growing it into a relevant first training set. The population of the EA is extracted and processed through the relevance feedback process. This process is composed of 3 steps:

- matching individuals to images from the available training sample set
- selecting which images are selected to get relevance feedback from and to beaded to the training set
- collecting the relevance feedback information and adding the images to the training set

4.4 The Evolutionary Algorithm Process: Once the population and the SVM have been initialized, the EA process can start. The first step is to evaluate the individuals of the population generated during the initialization. This evaluation step is the same as the evaluation of offspring done later in the EA process.

4.4.1 Individual's Evaluation: The EA is expected to converge toward a diverse set of individuals representing interesting elements to add to the SVM training set to improve its classification score. The area of the search space to which the individuals will converge is determined by the functions used to evaluate the individuals. Therefore, those need to be building carefully. The goal would be to identify a diverse set of images from each category in the dataset. Several elements are available to craft the evaluation functions:

- The query image from each category
- The SVM learned so far
- Other individuals

4.4.2 The Evolution Loop: Once the post-initialization evaluation process is done, the evolution loop starts.

Several things are to be defined when implementing such a system:

- image representation
- individuals representation in the EA
- EA evaluation functions
- which EA to use and its components
- algorithm used for the matching
- selection scheme in the relevance feedback process
- Type of SVM used.

Fisher Vectors computed on top of SIFT descriptors are used as the image representation. The SVM is used to learn the multi-class classifier from the training set, in the evaluation of the individuals of the EA and in the selection of images to be added to the training set in this system. The EA is expected to identify training samples for the SVM. Two main objectives are identified: finding images that will quickly improve theism decision, and explore the search space to find new areas with images from each category to better address the ne grained classification. The problem defined is a bi-objective problem. Any multi-objective EA could be used to solve it. NSGA-II is used in this implementation. The two-point crossover is used to generate the offspring. The mutation operator is the simple reset mutation. Relevance Feedback Process is composed of three steps: the matching process, a selection process and finally gathering the feedbacks before adding the images to the training set. In order to lower the impact of the matching on the speed of the system, a fast approximate nearest neighbor search based on Locality Sensitive Hashing (LSH) is used instead.

V. CONCLUSION

This PAPER focuses on one of those new challenges: the fine grained classification. It consists in being able to separate accurately images from class's present inciter class's visual similarities and intra-class visual dissimilarities. This problem turnout to be multi-modal by nature. Indeed, visual descriptions, such as the SIFT descriptors which form the base for the image description used, are built to keep visually similar images close and visually dissimilar images far from each other. Therefore, intra-class visual dissimilarities lead to images from the same class to be scattered through the search space. And inter-class similarities lead images from different classes to be mixed in one place of the search space. Thus they form small groups of images from the same class separated by groups of images from other classes. Those groups of images belonging to the same class but being scattered through the search space are modes for this class. To solve this kind of problem, admix of global and local search is needed. Or, the existing active learning techniques are lacking of global search to deal with such problems. Evolutionary Algorithms are presenting interesting exploration capabilities and were thus considered to tackle the fine grained classification problem. But image descriptions are most of the time of high dimension, more than 1000 dimensional vectors. And it turned out that the performances of EAs dedicated to multi-modal problems are not that good when dealing with high dimensional problems.

REFERENCES

[1] Wan J, Wang D, Hoi S C H, et al. 2014. Deep Learning for Content-Based Image Retrieval: A Comprehensive Study. In Proceedings of the ACM International Conference on Multimedia (MM '14). ACM, New York, NY, USA, 157-166. DOI=http://doi.acm.org/10.1145/2647868.2654948

[2] Ai L, Yu J, He Y, et al. 2013. High-dimensional indexing technologies for large scale content-based image retrieval: a review. Journal of Zhejiang University SCIENCE C. 14(7): 505-520.

[3] K. Ashok Kumar &Y.V.Bhaskar Reddy, "Content Based Image Retrieval Using SVM Algorithm", International Journal of Electrical and Electronics Engineering (IJEEE), ISSN (PRINT): 2231 – 5284, Vol-1, Iss-3, 2012.

[4] AmanbirSandhu , AartiKochhar, "Content Based Image Retrieval using Texture, Color and Shape for Image Analysis" International Journal of Computers & Technology Volume 3, No. 1, AUG, 2012.

[5] Keping Wang Xiaojie Wang, YixinZhong, "A Weighted Feature Support Vector Machines Method forSemantic Image Classification",International Conference on Measuring Technology and Mechatronics Automation (IEEE), 2010.

[6] seng, V.S.; Ming-Hsiang Wang; Ja-Hwung Su;, "A New Method for Image Classification by Using Multilevel Association Rules," Data Engineering Workshops, 2005. 21st International Conference on ,vol., no., pp. 1180, 05-08 April 2005.

[7] P. S. Hiremath and JagadeeshPujari. "Content Based Image Retrieval based on Color, Texture and Shape features using Image and its complement", International Journal of Computer Science and Security, Volume (1) : Issue (4).

[8] A. Foulonneau, P. Charbonnier and F. Heitz. Multireference shape priors for active contours. Int. J. Comput. Vision, vol. 81, no. 1, pages 68-81, 2009. (Cited on pages xi and 10.)

[9] Miguel Arevalillo-Herráez, Francesc J. Ferri and Salvador Moreno-Picot. A hybrid multi-objective optimization algorithm for content based image retrieval. Applied Soft Computing, vol. 13, no. 11, pages 4358-4369, 2013. (Cited on pages xix, 73 and 74.)

[10] Anne Auger, Johannes Bader, Dimo Brockhoff and Eckart Zitzler. Theory of the Hypervolume Indicator: Optimal μ-distributions and the Choice of the Reference Point. In Proceedings of the Tenth ACM SIGEVO Workshop on Foundations of Genetic Algorithms, FOGA '09, pages87-102, New York, NY, USA, 2009. ACM. (Cited on pages xviii and 52.)

[11] Jose Balcázar, Yang Dai and Osamu Watanabe. A Random Sampling Technique for Training Support Vector Machines. In Naoki Abe, Roni Khardon and Thomas Zeugmann, editors, Algorithmic Learning Theory, volume 2225 of Lecture Notes in Computer Science, pages 119-134. Springer Berlin Heidelberg, 2001. (Cited on pages 7 and 29.)

[12] S. Bandaru and K. Deb. A parameterless-niching-assisted bi-objective approach to multimodal optimization. In Evolutionary Computation (CEC), 2013 IEEE Congress on, pages 9-102, June 2013. (Cited on pages xviii, 58, 66 and 67.)

[13] A. Basak, S. Das and K.C. Tan. Multimodal Optimization Using a Biobjective Differential Evolution Algorithm Enhanced With Mean Distance-Based Selection. Evolutionary Computation, IEEE Transactions on, vol. 17, no. 5, pages 666-685, October 2013. (Cited on pages xviii, 58and 63.)

[14] Herbert Bay, Tinne Tuytelaars and Luc Gool. SURF: Speeded Up Robust Features. In Ale Leonardis, Horst Bischof and Axel Pinz, editors, Computer Vision - ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006. Proceedings, Part I, pages 404-417.Springer Berlin Heidelberg, Berlin, Heidelberg, 2006. (Cited on page 12.)

[15] S. Belongie, J. Malik and J. Puzicha. Shape Context: A new descriptor for shape matching and object recognition. In NIPS, 2000. (Cited on pages xi and 10.)

A Brief Author Biography

Dr. Mandalapu Ekambaram Naidu received Ph.D. degree in Computer Science in the faculty of Engineering in 2009 from University of Mysore , karnataka.. He is currently working as Professor & principal in the Department of Computer Science & Engineering, SRK Institute of Technology, Enikepadu, Vijayawada-521108. His current research is focused on Software Engineering, Database Management System, Software Testing Mehodologis.

Ravi Kumar Chandu, Working As An Asst. Professor In The Department Of Computer Science, Cmr Engineering College, Hyderabad . He Obtained His M.Tech. (CSE) In Sam Higginbottom University Of Agriculture Technology & Sciences, Allahabad- Naini, Uttar Pradesh. Hiss Research Interests Lies In Image Processing , Software Engineering .