IJCSE International Journal of Computer Sciences and Engineering Open Access
Survey Paper Volume-4, Issue-9 E-ISSN: 2347-2693

Survey on Recent Researches on High Level Image Retrieval

G.Vidya[1*], S.Omprakash[2]

[1]Department of Computer Science [PG], Kongunadu Arts and Science College, Coimbatore, TamilNadu, India
[2]Department of Information Technology, Kovai Kalaimagal College of Arts and Science, Coimbatore, TamilNadu, India

Available online at: www.ijcseonline.org

Received: 27/Aug/2016 Revised: 09/Sept/2016 Accepted: 25/Sept/2016 Published: 30/Sep/2016

Abstract—To obtain retrieval accuracy of content based images retrieval systems, the prime notice is on reduction of 'semantic gaps' between the visual features and human linguistics than designing low-level feature extraction algorithm. This paper elucidates a comprehensive study on recent technical updates in high-level semantic-based image retrieval. Major recent publications are enclosed during this survey covering different aspects of the research during this space, as well as low-level image feature extraction, similarity mensuration, and deriving high-level linguistics options. We have a tendency to establish 5 major classes of the progressive techniques in narrowing down the' linguistics gap': (1) victimisation object metaphysics to outline high-level concepts; (2) victimisation machine learning ways to associate low-level options with question concepts; (3) victimisation relevance feedback to find out user's intention; (4) generating linguistics template to support high-level image retrieval; (5) fusing the evidences from markup language text and also the visual content of pictures for computer network image retrieval. Other connected problems reminiscent of image workand retrieval performance evaluation are mentioned.

Keywords---*CBIR,Feedback,Machine Learning,semantic,Linguistic template*

I. INTRODUCTION

The interest in CBIR has grownup thanks to the restrictions inherent in metadata-based systems, still because the giant vary of doable uses for economical image retrieval. Matter info concerning pictures are often simply searched victimisation existing technology, however this needs humans to manually describe every image within the info. this could be impractical for terribly giant databases or for pictures that square measure generated mechanically, e.g. those from police work cameras. It's conjointly doable to miss pictures that use completely different synonyms in their descriptions. Systems supported categorizing pictures in linguistics categories like "cat" as a taxon of "animal" will avoid the miscategorization drawback, however would require a lot of effort by a user to search out pictures which may be "cats", However square measure solely classified as Associate in Nursing "animal". Several standards are developed to reason pictures, but we still face scaling and miscategorization problems.

Initial CBIR systems were developed looking for databases supported image color, texture, and form properties. When these systems were developed, the requirement for easy interfaces became apparent.

Corresponding Author: *G.VIDYA, vidyagopi.10@gmail.com*

Therefore, efforts within the CBIR field began to embrace humanitarian style that attempted to fulfill the requirements of the user performing arts the search. This generally means that inclusion of question ways which will enable descriptive linguistics, queries which will involve user feedback, systems which will embrace machine learning, and systems which will perceive user satisfaction levels

II. LITERATURE REVIEW

The model-based or template-based ones are the most suitable for accurate local junction detection considering all other approaches. In Deriche and Blaszka present computational approaches for a model-based detection of junctions. A junction model is a 2D intensity function depending on several parameters. Starting from a poorly localized initialization (e.g. from the Harris detector) parameters are then optimized in view of a precise localization. Parida*etal.*suggest a region-based model for simultaneously detecting, classifying and reconstructing junctions. A junction is defined as an image region containing piecewise constant wedges meeting at the central point. This work relies on a template deformation framework and uses minimum description length principle and dynamic programming to obtain the optimal parameters describing the model. Automatic junction detection has been a very active research field over the last four decades.
One of the earliest methods was introduced by , considering corners as points which are not self-similar in an image. The

© 2016, IJCSE All Rights Reserved

similarity of two points is measured by using the sum of squared differences (SSD) between their associated image patches. Harris and Stephens then proposed to approximate the SSD by the autocorrelation in a patch. The resulted cornerness measure is isotropic and has an analytic expansion, called the *Harris matrix*, and is widely used in practice. A large number of detectors rely on this idea, the detection of corners boiling down to the analysis of the Eigen values of this matrix [1].

The body of literature dealing with color descriptors is vast. Some of these descriptors have been normalized in the MPEG7 standard. The color version of the GIST descriptor is widely used for scene recognition. Forthe sake of conciseness, two color signatures will be used as reference namely color GIST and color SIFT . The color SIFT descriptor is the counterpart of SIFT with color integrated. best variant is the Opponent SIFT. Similar to SIFT, it describes regions of interest extracted by a detector. It has mainly be considered for image and video classification tasks. For this reason, it is used in conjunction with the BOW aggregation framework. However, this paper shows that much better results canbe achieved using these descriptors by combining them with the state-of-the-art local descriptor matching method. The main contribution of our paper is a simple color description method, which is used to produce either global or local descriptors. In the first case (global), computing the image signature is much faster than producing a BOWvector the main bottleneck is the computation of a color histogram, which is fast. Vitally improved search results are obtained on the evaluation datasets compared to other global descriptors [2].

We are interested in the problem of semi supervised learning (SSL) for image classification. The task is to design a method that can make use of unlabeled images, while learning classifiers from labeled ones. Recent research in SSL has obtained some success in solving this problem . Most of these methods build themselves upon the local-consistency assumption that data samples with high similarity should share the same label. This assumption allows the geometrical structure of unlabeled data to regularize the classifying functions. While improvements have been reported, these methods share three common drawbacks. First of all, these methods only exploit the local consistency assumption in image feature space, and ignore other prior information. Another reasonable assumption - borne out by our results - is that samples with very low similarity are in high probability come from different classes. We call this the exotic-inconsistency assumption, and design a method to exploit it also for SSL. Furthermore, most previous methods design specialized learning algorithms to leverage the structure of unlabeled data, so users often need to change their learning methods in order to utilize the cheap unlabeled data. This limits the applicability of SSL, as users usually are reluctant to give up their favorite classifiers. Last but not the least, previous methods assume that the unlabeled data are coming from more or less the same distribution as the labeled data. This imposes restrictions as well, as many applications have no prior access to the data to be classified. To overcome these limitations, we depart from the traditional paradigm and propose another route to SSL in this paper. Below, we present our motivations and outline the method. People learn and generalize object classes well from their characteristics, such as color, texture, and size. We also do so by comparing an object with other objects in the world. This is part of Eleanor Rosch's prototype theory , that states that an object's class is determined by its similarity to prototypes which represent object categories. The theory is suitable for transfer learning, where labeled data of other categories are available. An important question is whether the theory can also be used for SSL, with its huge amount of unlabeled data. Our paper investigates this problem[3].

The area of semi-supervised learning has experienced a significant evolution in terms of the adopted models, which comprise complex generative models, self-learning models ,multiview learning models, transductive support vector machines (SVMs) (TSVMs) , and graph-based methods . A survey of semi-supervised learning algorithms is available. Most of these algorithms use some type of regularization which encourages the fact that "similar" features are associated to the same class. The effect of such regularization is to push the boundaries between classes toward regions with low data density, where the usual strategy adopted first associates the vertices of a graph to the complete set of samples and then builds the regularizer depending on variables defined on the vertices. This trend has been successfully adopted in several recent remote sensing image classification studies. For instance, in , TSVMs are used to gradually search a reliable separating hyper-plane(in the kernel space) with a transductive process that incorporates both labeled and unlabeled samples in the training phase. In a semi-supervised method is presented that exploits the wealth of unlabeled samples in the image and naturally gives relative importance to the labeled ones through a graph-based methodology. In kernels combining spectral–spatial information are constructed by applying spatial smoothing over the original hyper-spectral data and then using composite kernels in graph based classifiers[4].

On the one hand, with the increasing spatial resolution, more details (structures and objects) on the Earth's surface emerge in satellite images, invalidating the features proposed for the low-resolution satellite image classification. The intensity and texture cues of entire scenes have proven sufficient for the low resolution satellite image classification. However, for high-resolution satellite images, the structures and objects often dominate their categories, and the whole image representation may sometimes lead to mistakes. The examples in different categories may have

© 2016, IJCSE All Rights Reserved

similar global responses while those in the same category may be different. This observation entails the image representation of focusing on the salient structures for high-resolution satellite images. On the other hand, tremendously acquired satellite images and increasingly demanding category requirements make most classification methods incompetent. State of the art as it is, modern machine learning algorithms (e.g., support vector machine (SVM), boosting, etc.) are incapable of scaling up to thousands of categories as they require a sophisticated learned model for each category. Recently, nonparametric nearest-neighbor schemes have attracted great attention in the image classification community. Despite its popularity, it is still unclear how to define the "true" neighbors of the optical satellite images. Motivated by the two challenges, in this letter, we present a method for the high-resolution satellite image classification, which puts more resources on the salient structures and relaxes the learning phase. The main idea is that our image representation, hereinafter referred to as the biased image representation, is guided by a bottom-up biologically inspired saliency measure. Then, we recognize the image categories based on the coding coefficients of a novel two-layer sparse decomposition model[5].

Extracting nonlinear features by KPLS is a very complex problem in the common situation in remote sensing where relatively few labeled data points are available. Including the information conveyed by unlabeled data via semi-supervised learning can potentially improve the feature extraction task. The semi-supervised framework has recently attracted a considerable amount of theoretical as well as remote sensing applied research. In this paper, we present a new semi-supervised KPLS method for nonlinear feature extraction. Our approach considers modifying the kernel similarity function via a semi-supervised kernel defined on the basis of clustering the analyzed image. Specifically, we propose to combine a standard supervised kernel with a Gaussian mixture model (GMM) clustering algorithm. While the supervised kernel exploits the information conveyed by the labeled samples, the cluster kernel accounts for the structure of the data manifold. The proposed semi-supervised KPLS (SS-KPLS) method is successfully tested in very high resolution and hyper-spectral image classification scenarios[6].

ANN algorithms are typically compared based on the tradeoff between search quality and efficiency. However, this tradeoff does not take into account the memory requirements of the indexing structure. In the case of E2LSH, the memory usage may even be higher than that of the original vectors. Moreover, both E2LSH and FLANN need to perform a final re-ranking step based on exact L2 distances, which requires the indexed vectors to be stored in main memory if access speed is important. This constraint seriously limits the number of vectors that can be handled by these algorithms. Only recently, researchers came up with methods limiting the memory usage. This is a key criterion for problems involving large amounts of data, i.e., in large-scale scene recognition, where millions to billions of images have to be indexed. In, Torralba et al. represent an image by a single global GIST descriptor which is mapped to a short binary code. When no supervision is used, this mapping is learned such that the neighborhood in the embedded space defined by the Hamming distance reflects the neighborhood in the Euclidean space of the original features. The search of the Euclidean nearest neighbors is then approximated by the search of the nearest neighbors in terms of Hamming distances between codes. In, spectral hashing (SH) is shown to outperform the binary codes generated by the restricted Boltzmann machine, boosting and LSH. Similarly, the Hamming embedding method of Jegou et al. uses a binary signature to refine quantized SIFT or GIST descriptors in a bag-of-features image search framework[7].

A number of approaches exist for feature extraction of hyper-spectral images, ranging from unsupervised methods to supervised ones. One of the best known unsupervised methods is principle component analysis (PCA), which is widely used for hyper-spectral images. Recently, some local methods, which preserve the properties of local neighborhoods were proposed to reduce the dimensionality of hyper-spectral images, such as locally linear embedding, Laplacianeigenmap, and local tangent space alignment. Their linear approximations, such as neighborhood preserving embedding (NPE), locality preserving projection (LPP), and Linear Local Tangent Space Alignment (LLTSA) were recently applied to feature extraction in hyper-spectral images. By considering neighborhood information around the data points, these local methods can preserve local neighborhood information and detect the manifold embedded in the high-dimensional feature space. Supervised methods rely on the existence of labeled samples to infer class separability. Two widely used supervised feature extraction methods for hyperspectral images are the linear discriminant analysis (LDA) and nonparametric weighted feature extraction (NWFE). Many extensions to these two methods have been proposed in recent years, such as modified Fisher's LDA, regularized LDA , modified nonparametric weight feature extraction using spatial and spectral information, and kernel NWFE[8].

As a fundamental ingredient of image structures, texture conveys important cues in numerous processes of human visual perception. While, due to the high complexity of the structures in natural images, the modeling of texture is a challenging problem in image analysis and understanding. Over the past decades, tremendous investigations have been made in texture analysis, among which an active topic is developing texture models that can efficiently depict both the statistical and the geometrical aspects of textures and are

© 2016, IJCSE All Rights Reserved

robust to the variations of imaging condition as well. In order to represent the structural aspects of textures, some mathematical tools, such as Gabor or wavelet-like analysis, are used to probe the atomic texture elements such as elongated blobs and terminators in images, and the marginal/joint distributions of the resulted responses are subsequently utilized to describe the statistical arrangement of texture. The strong ability of such mathematical tools to handle multi-scale and oriented structures has made them one of the most popular tool for texture analysis. However, how to efficiently represent the highly geometrical aspects of textures, e.g. sharp transitions and elongated contours, is an open problem. To solve this problem, alternative wavelet-like approaches, e.g. Grouplet and scattering transform, have been elaborated to enable more efficient representations of structured textures. In contrast with explicit models, patch-based method provides another possibility for describing the structured aspects of textures, but it is not trivial to capture the multi-scale nature of textures by this kind of approaches [9].

Within a satellite scene, information is delivered in many forms such as structure, texture, and color. Each individual cue reveals only one aspect of the scene, but in isolation it will not suffice. Given that the information from each cue is ambiguous and incomplete, how to integrate these diverse and complementary features becomes a crucial question. Multiple-kernel learning (MKL).Lanckrietet al. is a prominent type of kernel fusion that makes use of kernel functions defining a measure of similarity between pairs of instances. MKL associates a kernel with each image feature and approximates the optimal feature's kernel as the relative weight for a specified task. However, although the MKL solution is sparse for every class in isolation, it is not sparse in a multi-class situation. Because even the simple baseline methods 'average' and 'product' are highly competitive with MKL, it may be concluded that the performance of MKL might have been overestimated in the past Gehler and Nowozin . One widely used approach is score level fusion, in which scores from different feature channels are fused, opening up new prospects for multi-stage classification. A prominent representative of this approach is a two-stage linear support vector machine (SVM) classification scheme using sparse coding based multiple-feature combinations. In the second stage, the approach simply used score concatenation to combine the intermediate probabilities obtained from the corresponding feature channels in the first stage. Although the approach turned out to work surprisingly well, some improved probability fusion styles ought also to be considered. Another difficulty is choosing a robust score-normalization technique because of the disparate characteristics of the underlying score distributions for different data sources. Since score distributions vary as a function of the classification algorithm, one must normalize the score data before combining them in score-level fusion. *Z*-scores are adaptive normalization techniques that are easy to compute; however, these *Z*-scores are not robust (sensitive to outliers) and are easily impacted by recognition algorithm failure (if one classification algorithm involved in the fusion process cannot produce a correct matching result, it will strongly impact the final result of fusion). 'tanh' estimators are fixed-score normalizations that are considered robust to noise, but are far more difficult to compute than adaptive *Z*-scores. Ideally, a score-normalization method would not only be robust to failure, but would also not be dominated by complex parameter estimation procedures with variation in score distribution. Normalization using *W*-scores is preferable. A statistical extreme value theory(EVT) normalization technique has been used to construct normalized 'multi-attribute spaces' from raw classifier outputs. This method calibrates each raw score to the probability that an image exhibits a given attribute, an approach that has significant applications in multi-attribute searches and target-attribute similarity searches. EVT normalization draws the probabilities from the cumulative distribution function (CDF) of aWeibull distribution (hence the term '*W*-score'). The *W*-score renormalizes the data based on the formal probability of each point being an outlier in the extreme value 'non-match' model, thus enhancing the chance of achieving a successful classification. The resulting probabilities are the normalized *W*-scores, which can be fused together to produce an overall probability or used to perform other tasks. A significant benefit of the approach is that the calibration is done after-the-fact, requiring neither modification to the attribute classifier nor the reference attribute annotations [10].

Comparison Analysis

Despite the fact that the far-reaching independence assumptions are often inaccurate, the naive Bayes classifier has several properties that make it surprisingly useful in practice. In particular, the decoupling of the class conditional feature distributions means that each distribution can be independently estimated as a one dimensional distribution. This in turn helps to alleviate problems stemming from the curse of dimensionality, such as the need for data sets that scale exponentially with the number of features.

Database Setup

The database in the experiments was made up of one thousand video clips downloaded from internet. The total length of video data is approximately one hundred and sixty hrs, including various contents such as news, sports, movie, cartoon, teleplay and natural scene etc. The experiments were performed on Xeon E5410 2.33GHz, 2G memory. The algorithms in were implemented for comparison.

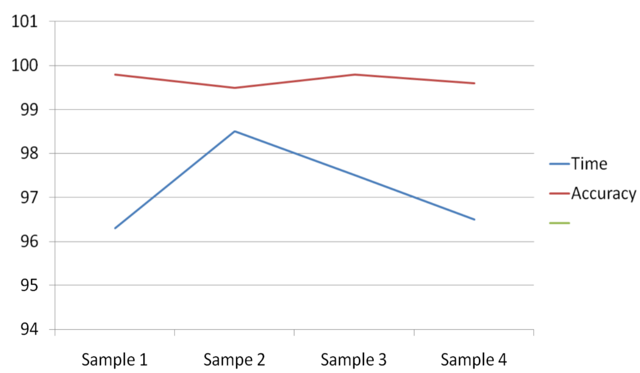
Comparison

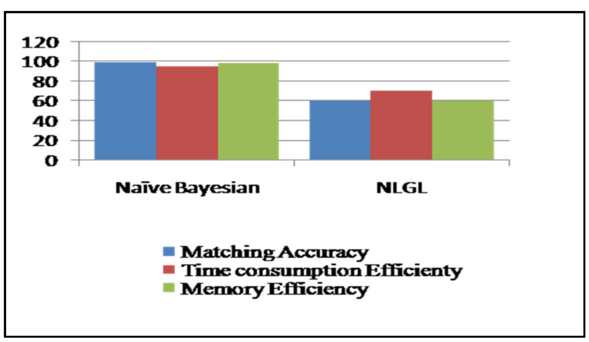
Test Results: The experiments were set to the performance comparison of similarity search. The results of

© 2016, IJCSE All Rights Reserved

recall-precision curves are shown. The method (1)[**Naïve Bayesian**] is the method and the method(2)[**Noise Reduction local and Global algorithm**] stands for the approach. The precision rate goes down when the recall rate rises. Method(1) holds steady precision at 1 while the others drop significantly when the recall rate rises. The method (1) is slightly better than method (2) in precision rate when the recall rate is smaller than 0.60.

The method (1) can hold precision rate at 1 when the recall rate reaches 0.75. The others have the best precision rate at 0.7 when the recall rate is 0.75. When recall rate is approaching to 1, the precision rate of the proposed method decreased a little at 0.9. The others' precision rate are below 0.6 when their recall rate are higher than 0.8. This is because the proposed search approach is according to video components based on the statistics of STD. The similarity search is directed to the corresponding CIT and the dissimilar video can be eliminated effectively by index clustering. The representative video shot images of comparison CBIR





By giving query video, the proposed method can find more similar image with similar images compared with the algorithm (1). Therefore, the better performance can be obtained by proposed method with satisfying recall and precision rate as it is shown. The results of average search time versus the number of images are shown.

The comparison is based on the average generation time of proposed CBIR system. The real query of Semantic CBIR system is extremely fast because it just shows the content. With the growing number of images, the search time increases.

Conclusion

In order to spice up the retrieval accuracy of content-based image retrieval systems, analysis focus has been shifted from coming up with refined low-level feature extraction algorithms to reducing the 'semantic gap' between the visual choices and conjointly the richness of human linguistics. This paper makes a shot to supply a comprehensive survey of the recent technical achievements in high-level semantic-based image retrieval. Major recent publications area unit penned throughout this survey covering all different aspects of the analysis throughout this house, in addition as low-level image feature extraction, similarity measure, and derivation high-level linguistics choices. We've got an inclination to determine five major categories of the progressive techniques in narrowing down the' linguistics gap': (1) mistreatment object philosophy to stipulate high-level concepts; (2) mistreatment machine learning ways that to associate low-level choices with question concepts; (3) mistreatment connection feedback to seek out users' intention; (4) generating linguistics guide to support high-level image retrieval; (5) fusing the evidences from nomenclature text and conjointly the visual content of images for network image retrieval. To boot, other connected issues harking back to image work and retrieval performance analysis are mentioned.

REFERENCES

[1].Gui-Song Xia, Julie Delon, Yann Gousseau "Accurate Junction Detection and Characterization in Natural Images " International Journal of Computer Vision , Volume 106, Issue 1, pp 31–56 ,January 2014.

[2] HervéJégou ,Christian Wengert, MatthijsDouze, "Bag-of-colors for improved image search" Proceeding MM '11 Proceedings of the 19th ACM international conference on Multimedia Pages 1437-1440 , ACM New York, NY, USA ©2011, ISBN: 978-1-4503-0616-4.

[3] .Dengxin Dai, Luc Van Gool , "Ensemble Projection for Semi-supervised Image Classification " ICCV '13 Proceedings of the 2013 IEEE International Conference on Computer Vision Pages 2072-2079 , IEEE Computer Society Washington, DC, USA ©2013

[4] Mamta R. Nagpure, Suchita S. Mesakar, Sonali R. Raut and Vanita P.Lonkar, "Image Retrieval System with Interactive Genetic Algorithm Using Distance", International Journal of Computer Sciences and Engineering, Volume-02, Issue-12, Page No (109-113), Dec -2014,

[5] Dengxin Dai ; Wen Yang "Satellite Image Classification via Two-Layer Sparse Coding With Biased Image Representation" IEEE Geoscience and Remote Sensing Letters ,Volume: 8, Issue: 1,Page(s): 173 – 176, Jan. 2011 .

[6] Emma Izquierdo-Verdiguier ; Luis Gómez-Chova ; Lorenzo Bruzzone ; Gustavo Camps-Valls , " Semi-supervised Nonlinear Feature Extraction For Image Classification"

Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International Date of Conference: 22-27, July 2012.
[7] Jégou H1, Douze M, Schmid C."Product Quantization for Nearest Neighbor Search", IEEE Trans Pattern Anal Mach Intell. 33(1):117-2, Jan 2011.
[8] L. Haldurai and V. Vinodhini, "Parallel Indexing on Color and Texture Feature Extraction using R-Tree for Content Based Image Retrieval", International Journal of Computer Sciences and Engineering, Volume-03, Issue-11, Page No (11-15), Nov -2015,
[9] Aasish Sipani, Phani Krishna and Sarath Chandra, "Content Based Image Retrieval Using Extended Local Tetra Patterns", International Journal of Computer Sciences and Engineering, Volume-02, Issue-11, Page No (11-17), Nov -2014
[10] Shraddha S.Katariya and Ulhas B.Shinde, "Review of Content Based Image Retrieval Using Low Level Features", International Journal of Computer Sciences and Engineering, Volume-04, Issue-03, Page No (91-97), Mar -2016, E-ISSN: 2347-2693

AUTHORS PROFILE

G.Vidya working as an Assistant Professor in Department of Computer Science (PG), Kongunadu Arts and Science College, Coimbatore. She holds her Master of Science in Computer Technology from Anna University. Her research interests are Data Mining ,Computer Networks & Software Engineering.

S.Omprakash is an Assistant Professor in Department of Computer Science, Kovai Kalaimagal College of Arts and Science, Coimbatore. He has completed his Post Graduation and M.Phil degree from Bharathiar University. His research interests are Data Mining & Warehousing.

© 2016, IJCSE All Rights Reserved