International Journal of Control and Automation
Vol.8, No.8 (2015), pp.399-408
http://dx.doi.org/10.14257/ijca.2015.8.8.37

Annotation Based Image Retrieval using GMM and Spatial Related Object Approaches

Monica Hidajat

School of Computer Science, Bina Nusantara University
monica.hidajat@yahoo.com

Abstract

Image annotation and retrieval has been a popular research topic for decades. Based on published journals from 2012 until 2015, a lot of research and studies has been focused on Content Based Image Retrieval (CBIR). In most cases, CBIR systems that use an image as the input query always face a problem called semantic gap due to the use of low-level features for similarity matching. The semantic gap will consequently reduce the performance of the CBIR systems. To overcome this problem, a new class of CBIR known as Annotation Based Image Retrieval (ABIR) has been developed. The ABIR systems, based on the annotation process of the images, employ bag of words for query. However it is found that employing pure bag of word only is not adequate to solve semantic problem. In this study, an attempt to use a Gaussian Mixture Model (GMM) based approach and spatial related information of the annotated objects has been performed in order to improve the performance of the ABIR systems. From the retrieval experiments, it is found that ABIR could achieve average performance up to 88% which is two times better than the CBIR systems in reducing the semantic problem.

Keywords: *ABIR, GMM, VSM, Spatial Information, Semantic Problem*

1. Introduction

The massive of digital data collection becomes a new challenge for retrieval technology. Based on IDC iView, digital data is predicted to grow from 130 exabytes in 2005 to 40,000 exabytes, or 40 trillion gigabytes in 2020 (more than 5,200 gigabytes for every man, woman, and child in 2020) [4]. Therefore, the urgency of managing digital data becomes extremely important. Multimedia retrieval system is a method to manage digital data. Digital data can be divided into image, sound, text, and video. In this research, the media Almost 200 Content Based Image Retrieval (CBIR) systems have been studied and explored in the last decades. CBIR is a retrieval method which is based on image low level features like color, shape, or texture. Unfortunately, CBIR fails to meet user expectation. The main factor is the semantic gap [6].

Semantic gap is a gap between high level information semantic (keyword, text descriptor, *etc*.,) and results of low level features extraction (color, shape, *etc*.,) [7]. This problem is crucial because high level information semantic is meaningful and effective for image retrieval. One solution for semantic gap is Annotation Based Image Retrieval (ABIR) [5]. ABIR is a new class retrieval methods that is based on text or user query. The main step in ABIR can be divided into two, automatic image annotation and query processing. Automatic image annotation can be done through image segmentation and labeling processes.

One of popular method used in segmentation image is Gaussian Mixture Model (GMM). GMM is reliable to do segmentation process. The first reason is GMM is one of Parametric Probabilistic Model. The second is GMM assign probability of each point to each cluster (soft clustering) [10].Unfortunately multi label image of ABIR is still not sufficient for reducing semantic problem. Therefore spatial related object information

ISSN: 2005-4297 IJCA

may be added. Spatial related object information is added to describe position of the objects and relation among them. In this paper, an attempt to use GMM and spatial information approaches will be described. The GMM is used for image segmentation process and SVM for labelling the segmented image. Moreover the use of spatial information will facilitate user to search image more detail.

2. Literature Review

Multimedia refers to a digitized collection of media types used together and computer readable representation. Multimedia can be divided into text, image, audio, and video (Blanken, Vries, Blok, & Feng, 2007).The method to manage multimedia is multimedia retrieval system. In terms of image type is called Image retrieval. It is a technology of process for browsing, searching, and retrieve image from large image database. The options of image retrieval are random by browsing, search by example, search by text, and navigation with customized image categories [11].

Content Based Image Retrieval (CBIR) is the retrieval of interested images from image collections by example image or based on visual properties of images. The visual properties used are often low-level features, such as colors, textures, shapes *etc*., [6]. On the other hand, humans who search digital collections express their information needs by using high-level concepts such as keywords. The gap between low level features (image) with high level concept (text) is called semantic gap [2] and it is main problem of CBIR.

Annotation Based Image Retrieval (ABIR) is cross-medium type retrieval since user queries are often in the form of text and the search targets are images [6].ABIR is a bridge between high level (words) and low level features (color, texture, shape, *etc*., in image). Main step in ABIR can be divided into two *i.e*., automatic image annotation and query processing.

Image annotation is process for assign relevant labels described image. The goal of image annotation is find relation between image visual features and semantic, then label the image with words [12]. Image annotation is a difficult task for two main reasons: The first is the well-known pixel-to-predicate or semantic gap problem, which points to the fact that it is hard to extract semantically meaningful entities using just low level image features, *e.g*., color and texture. The second difficulty arises due to the absence of correspondence between the keywords and image regions in the training data [8]. Annotating image manually need big cost, time consuming, *etc*. Therefore, there is a need to automatically assign labels to an image called automatic image annotation (AIA) [3].

Automatic image annotation can be done by doing image segmentation based on low level feature image and assigning label for each segmentation that produced. One of method to do automatic image annotation is using machine learning techniques. Machine learning techniques are commonly used for the image classification and image feature analysis such as segmentation *etc*. In this image annotation method, a general term is usually called ‘blob’. A blob is a part of an image with a vocabulary meaning. The image is separated into blobs based on the region or cluster and the corresponding blobs are labeled with a word [9].

3. Methodology

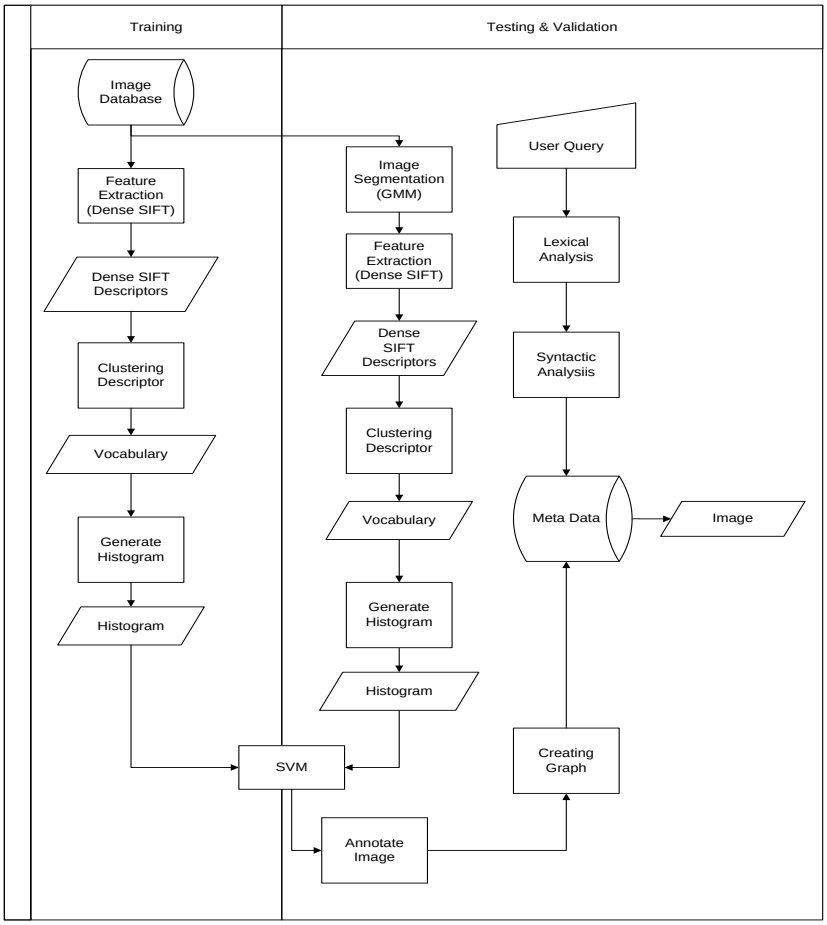


Figure 1. Annotation Based Image Retrieval Methodology

The proposed methodology for this study is using Gaussian Mixture Model (GMM) as clustering feature image and image segmentation and Multiclass Support Vector Machine as classifier for labelling. There are two main phases in this study. The first phase is the training phase using SVM (Support Vector Machine) to create model for each keyword. The following steps are:

1. Feature will be extracted and transform using dense SIFT.
2. Dense SIFT descriptor is clustered using K-means to create vocabulary.
3. Creating Histogram from the vocabulary.
4. Training model based on vocabulary histogram using SVM.

The second phase is the testing and validation phase for automatic image annotation and image retrieval. In testing and validation phase, the two main steps are image annotation and retrieval. The following is steps in image annotation:

1. Color space image transforms from RGB into HSV.
2. Image will be segmented using GMM into k object. GMM is implemented using Expectation – Maximization algorithm. EM is a widely used method for estimating the parameter set of models using incomplete data. The EM algorithm produces a sequence of estimates by applying the following steps [10]:
   a. Initialize Gaussian parameters: means μk, covariance Σk, and mixing coefficients πk.
   b. E – step: Evaluate the Responsibilities.

c. M – step: Re-estimate parameter: means μk, covariance Σk, and mixing coefficients πk.
d. Evaluate likelihood. If likelihood or parameters converge, stop.

3. Each object will be extracted and transformed using dense SIFT.
4. Dense SIFT descriptor transformed into vocabulary histogram and classified using SVM.
5. The labelled image is processed to get spatial relation between object using graph matrix and position of each object. The graph matrix and object position result are saved in metadata.

The following is steps in retrieval phase:

1. User query (text) is processed using lexical analysis. User query is trimmed to get token words.
2. Token words is processed using syntactic analysis. The token words is analyze to change it into graph model.
3. The graph query is matched with all graph in metadata.
4. System is going to retrieve and show the matched image to user.

4. Result and Discussion

Data used for training, testing, and validation in this study is LAMDA data sets. This following is detail of dataset:

Table 1. Population and Sample Dataset

Training (#category)	Testing & Validation	Keywords
84	457	dessert, mountains, sea, sky, trees

The following are characteristics of data:

Table 2. Dataset Characteristics

Category	Value
Color Space	RGB
Min Width	83px
Min Height	46px
Max Width	800px
Max Height	640px
Bit Depth	24bit

Result of retrieval testing can be evaluated by computes precision and recall. This precision can be computed by comparing number of data correctly retrieved and total number of data retrieved. Meanwhile recall can be computed by comparing number of data correctly retrieved and total number of data relevant in collection.

$$\text{precision} = \frac{\text{Number of relevant images retrieved}}{\text{Total number of images retrieved}} \quad (1)$$

$$\text{recall} = \frac{\text{Number of relevant images retrieved}}{\text{Total relevant images in collection}} \quad (2)$$

$$\text{F} - \text{measure} = \frac{2 * precision * recall}{\text{precision} + \text{recall}} \quad (3)$$

In this research, we compare retrieval result of ABIR and CBIR. CBIR system use color histogram as low level features. Color histogram is matched with all color histogram image in database and sorted based on similarity.

4.1. Segmented & Labelling

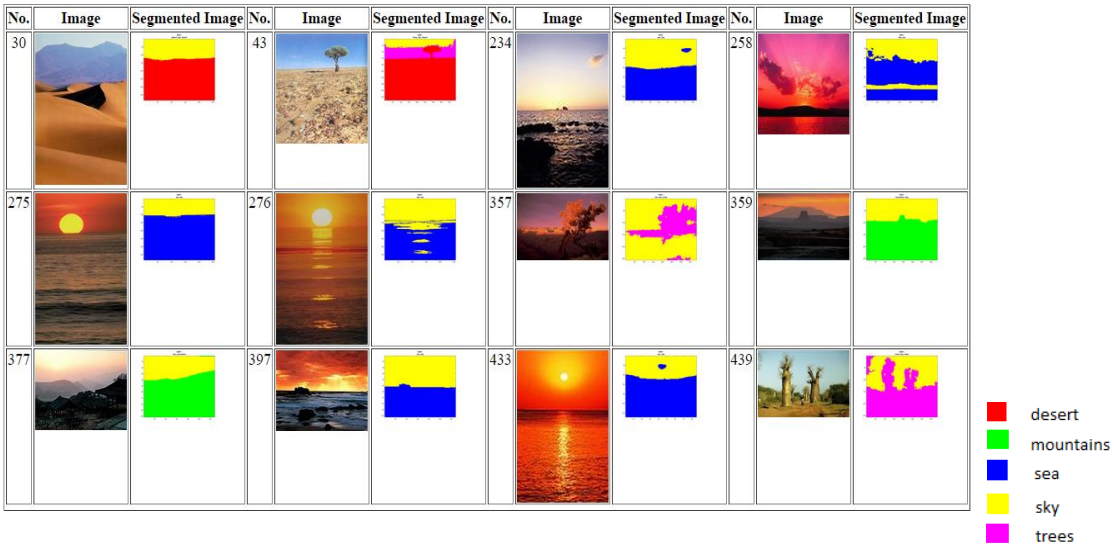


Figure 2. Sample of Segmented & Labelling Image

There are 5 keywords for annotation labels (dessert, mountains, sea, sky, and trees).Image has RGB color space, 24 bit depth, and various dimension pixel. Images were segmented using GMM and were labeled using SVM. Color space image is transformed from RGB into LAB to get color layout descriptor. The pixel value of transformed image is clustered into k objects by GMM using EM algorithm. The k objects is classified using Multiclass SVM based on model on training phase. The result of this two processes can be seen at figure 2. For example image 30 has segmented into 2 class *i.e.*, dessert and sky.

4.2. Graph based Object Spatial Information

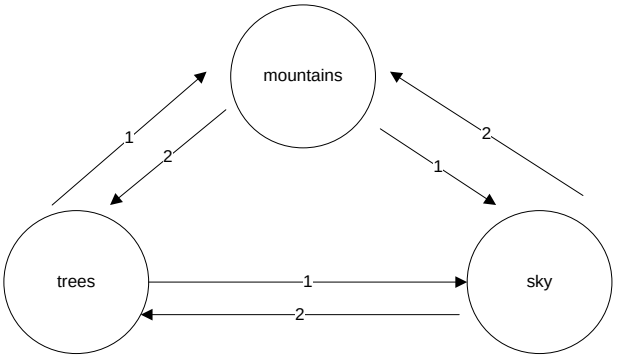
The segmented image can be explored to get spatial information. Spatial information can be divided into position and relation between objects. There are 5 area to discover, top, middle, bottom, left, and right. Machine reads pixel per pixel to save every label in certain area. Therefore every area have information which label exist. The steps algorithm to get relation between object information can be divided into two stages. First stage check vertical relation (above and under) and second stage check horizontal relation (left of and right of). Based on illustration above, every object has own relationship value of other object. For example, if A is left of B, then B is right of A. The value truly depend on viewpoint of object.

Table 3. Relation Value for Image Graphing

Keyword	Value
Under	2^0
Above	2^1
Left Of	2^2
Right Of	2^3

Relation value is calculated by accumulate existed relation value (Table 3).For example object A is under and left of object B, then relation value between A and B becomes $2^0 + 2^2 = 5$.The following is graph sample of image:

Table 4. Relation Value for Image Graphing

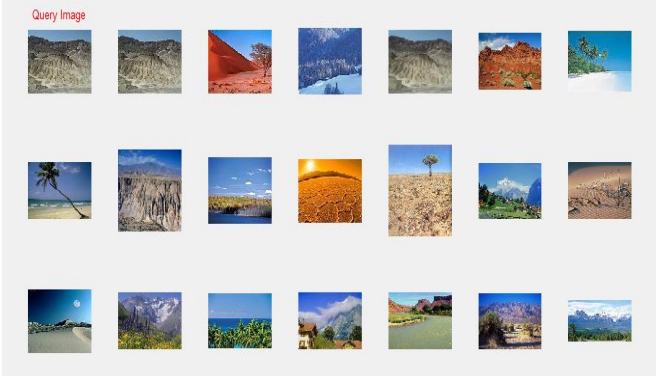
Image	Graph Image
	

The results of graph based object spatial information can be seen at table 4. In table 4, the relation between mountains and sky is mountains under the sky or sky above the mountains. Therefore, the relation between mountains and sky is 2^0 while the relation between sky and mountains is 2^1.

4.3. ABIR vs CBIR

The following is result of Query 1:

Table 5. Result of Query 1

Query : **mountains under the sky**	**Query:** 
 (19 relevant images)	 (10 relevant images)

Based on result above, ABIR successfully show 19 related images (95%), but CBIR only show 10 related images (50%). CBIR using color histogram to matching and sorting image based on similarity. Therefore the result of CBIR show existence of semantic gap. Besides ABIR is more reliable to reduce semantic problem because using human annotation as model to classify.

The following is retrieval results of ABIR and CBIR:

Table 6. Retrieval Results of ABIR vs CBIR

# Query	Query	ABIR	CBIR	ABIR (%)	CBIR (%)
1	mountains under the sky	19	10	95%	50%
2	mountains above the sea	19	4	95%	20%
3	sea on bottom	17	13	85%	65%
4	trees above desert	6	3	55%	15%
5	tree in the middle	20	7	100%	35%
6	tree under the mountains and tree above dessert	19	3	95%	15%
7	tree under the mountains	14	2	70%	10%
8	desert under the mountains	14	2	82%	10%
9	tree under the sky	20	8	100%	40%
10	mountain on the top	20	10	100%	50%

The retrieval result is limited with maximum 20 images. Table 4-6 shows precision percentage of each query. The results show inconsistent result because ABIR depends on annotation process. The annotation of tree label is not less accurate than the others. Query 1, 2, 3, 5, 6, 8, 9, and 10 of ABIR show higher precision significantly than CBIR precision. Besides Query 4 and 7 show higher precision than CBIR but not significant. Some results of CBIR has low precision. It may cause of the query image has the same color histogram with image in database.

CBIR able to retrieve 31% average precision of related images. However, ABIR able to retrieve almost 88% average precision of related images. ABIR is significantly reduce semantic problem in CBIR. The result show CBIR is not sufficient to reduce semantic problem because CBIR using low level features to matching. While ABIR is sufficient because using label (semantic level) to matching.

a. Precision, Recall, and F-measure

Summary of precision, recall, and F-measure:

Table 7. Summary of Precision, Recall, and F- measure

# Query	Recall	Precision	F-measure
1	0.68902439	0.965812	0.80427046
2	0.694915	0.931818	0.7961165
3	0.73764259	0.94634146	0.82906
4	0.461538	0.66666667	0.54545455
5	0.63218391	1	0.77464789
6	0.57142857	0.95652174	0.71544715
7	0.571429	0.956522	0.71544715
8	0.875	0.823529	0.848485
9	0.61728395	0.961538	0.7518797
10	0.71794872	1	0.8358209

Semantic problem is a popular issue in information retrieval topic. In this research, we proposed spatial information to reduce semantic problem. Spatial information can be saved using Graphing. Graphing was usually used for image matching, but in this research we use graphing to save relation between object for each image. In our design, we proposed retrieval system using spatial information so user can get specific image.

In this research, the evaluation can be done by calculate recall, precision, and F-measure. F-measure value is range value to describe the level of semantic level. The higher F-measure shows the less semantic problem because trained image has been annotated by human (semantic level) as a benchmark to evaluate the result of model.

Based on result above, the range of precision is 66.67% - 100%, the range of recall is 46.15% - 66.67%, and the range of F-measure is 54.54% - 84.85%. Therefore the average of F-measure is 76.17%. The main factor of inaccurate is the system shows image based on sorted image id. The proposed method is sufficient to reduce semantic problem in image retrieval topic because the average level of F-measure is approximately 7.

5. Conclusion and Recommendation

Semantic problem is common problem in information retrieval topic. In this study, an attempt to use a Gaussian Mixture Model (GMM) based approach and spatial related information of the annotated objects has been performed in order to improve the performance of the ABIR systems. GMM is used to segment images into objects. All segmented objects are annotated using SVM and extracted spatial information into graph.

Based on recall and precision evaluation, the range of precision is 66.67% - 100%, the range of recall is 46.15% - 66.67%, and the range of F-measure is 54.54% - 84.85%. Therefore, the average F-measure is 76.17%. The level of F-measure in this research is approximately 7. It means that this method is sufficient to be implemented in image retrieval.

Average retrieval of CBIR is 31% and average retrieval of ABIR is 88%. It shows ABIR is more reliable than CBIR in reduce semantic problem. Although this research has a good method to reduce semantic problem, but there are few things can be added or improved. Annotation process need to be improved to increase recall and precision. In this study, the system shows image based on image id. This condition cause unrelated image can be showed as first or second data. The solution of this condition is relevance feedback feature. Relevance feedback is usually used to evaluate annotation and set priority for each image.

Acknowledgements

The author would like to thank Dr. Diaz D. Santika for his fruitful discussion throughout the supervision process. This research use dataset from LAMDA (http://lamda.nju.edu.cn/).

References

[1] Blanken, H., De, P., Blok, H., & Feng, L. (2010). Multimedia Retrieval. Amsterdam: Springer.
[2] Clinchant, S., Ah-Pine, J., & Csurk, G. (2011). Semantic Combination of Textual and Visual Information in Multimedia Retrieval. ACM International Conference on Multimedia Retrieval, Article No. 44.
[3] Feng, Y., & Lapata, M. (2008). Automatic Image Annotation Using Auxiliary Text Information. Association for Computational Linguistics, 272-280.
[4] Gantz, J., & Reinse, D. (2012). THE DIGITAL UNIVERSE IN 2020: Big Data, Bigger Digital Shadow s, and Biggest Growth in the Far East. IDC iView: IDC Analyze the Future.
[5] Inoue, M. (2004). On the need for annotation-based image retrieval. Research Gate.
[6] Joshi, N. k., & Khedekar, R. (2013). Narrowing Semantic Gap in Content-based. PARIPEX - INDIAN JOURNAL OF RESEARCH, 98-99.
[7] Liu, Y., Zhanga, D., Lua, G., & Wei-Ying. (2007). A survey of content-based image retrieval with high-level semantics. Pattern Recognition, 262-282.

[8] Makadia, A., Pavlovic, V., & Kumar, S. (2010). Baselines for Image Annotation. International Journal of Computer Vision, Volume 90, 88-105.
[9] Nagarajan, S., & Saravanan, S. (2012). Content-based Medical Image Annotation and Retrieval using Perceptual Hashing Algorithm. IOSR Journal of Engineering, 814-818.
[10] Robotka, Z., & Zempleni, A. (2009). Image Retrieval Using Gaussian Mixture Models. Sect. Comp 31, 93-105.
[11] Rui, Y., Huang, T., & Chang, S.-F. (1999). Image Retrieval: Current Techniques, Promising Directions, and Open Issues. Journal of Visual Communication and Image Representation, 39-62.
[12] Tsuboshita, Y., Kato, N., & Fukui, M. (2012). Image Annotation Using Adapted Gaussian Mixture Model. International Conference on Pattern Recognition, 1346-1350.