

Survey On Novel Techniques for Effective Image Search Based On Users Intention

S. Aarif Ahamed¹, B. A. Vishnupriya¹, V. Venkateshwaradevi¹

First Year M.E (CSE)¹, Assistant Professor¹

Department of Computer Science & Engineering

M.I.E.T. Engineering College¹, M.A.R. College of Engineering and Technology¹, Oxford Engineering College¹
Tiruchirapalli, TamilNadu, India.

Abstract

Image search is a specialized data search used to find images. To search for images, a user may provide query terms such as keyword, image file/link, or click on some image, and the system will return images "similar" to the query. The similarity used for search criteria could be meta tags, color distribution in images, region/shape attributes, etc. Many commercial Internet scale image search engines use only keywords as queries. Users type query keywords in the hope of finding a certain type of images. The search engine returns thousands of images ranked by the keywords extracted from the surrounding text. It is well known that text-based image search suffers from the ambiguity of query keywords. The keywords provided by users tend to be short. They cannot describe the content of images accurately. The search results are noisy and consist of images with quite different semantic meanings. For example, if a user wants to search for an "apple" image, he/she may request a query search using the keyword "apple" to the corresponding image search engine. The meanings of the word "apple" include apple fruit, apple computer, and apple ipod. The search results will contain different categories, such as "green apple," "red apple," "apple logo," and "iphone" because of the ambiguity of the word "apple". This leads to ambiguous & noisy search results which are not satisfactory to fulfill the user query request. In order to solve the ambiguity, additional information has to be used to capture users' search intention. A comprehensive survey and study of various techniques for effective image retrieval based on users' search intention has been presented in this paper.

Keywords - Image Search, Intention, Visual, Web Image Search, Clustering, User Interface

1. Introduction

Image Search is a process of retrieving and displaying relevant images based on user's queries from a database. Image searching techniques can be broadly classified in to two types Text-Based Image Retrieval (TBIR) and Content-Based Image Retrieval (CBIR).

Text-Based Image Retrieval (TBIR) uses text descriptions to retrieve relevant images based on Time, location, events, and objects. Users type query keywords in the hope of finding a certain type of images. The search engine returns thousands of images ranked by the keywords extracted from the surrounding text. It is well known that text-based image search suffers from the ambiguity of query keywords. The keywords provided by users tend to be short. They cannot describe the content of images accurately. The search results are noisy and consist of images with quite different semantic meanings. For example, if a user wants to search for an "apple" image, he/she may request a query search using the keyword "apple" to the corresponding image search engine. The meanings of the word "apple" include apple fruit, apple computer, and apple ipod. The search results will contain different categories, such as "green apple," "red apple," "apple logo," and "iphone" because of the ambiguity of the word "apple". This leads to ambiguous & noisy search results which are not satisfactory to fulfill the user query request.

Content-based image retrieval (CBIR), also known as query by image content (QBIC) and content-based visual information retrieval (CBVIR) [7][8] is the application of computer vision techniques to the image retrieval problem, that is, the problem of searching for digital images in large databases. The term has been used to describe the process of retrieving desired images from a large collection on the basis of syntactical image features. The techniques, tools and algorithms that are used originate from fields such as statistics, pattern recognition, signal processing, and computer vision. Extracting images based on image content involves following levels:

Level 1: Retrieval by primitive features such as color, texture, shape and spatial location.

Level 2: Retrieval of objects of given type. Example: find the picture of the flower.

Level 3: Retrieval of abstract attributes that involves high level reasoning. Example: 'find picture of a baby smiling'.

The paper is organized as follows. In section 2, the study of various techniques for effective image retrieval based on users' search intention will be introduced. Finally, section 3 is a conclusion.

2. Literature Survey

2.1 Improving Web-based Image Search via Content Based Clustering

The main intention of content based image retrieval is that when a user submits a query image, it retrieves the images that are mostly relevant to the content. Nadav Ben et al [6] achieved the content based image retrieval by introducing a new approach name ReSPEC (Re-ranking Sets of Pictures by Exploiting Consistency).

ReSPEC is composed of two main methods. First, based on the user query image (keyword) the image search engine (Google, Yahoo,...etc) retrieves the images then, clusters the results based on extracted image features, and returns the cluster that is inferred

to be the most relevant to the search query. Secondly, ranks the results that are most relevant to the user query images. Figure 1, represent various stages of images retrieval using ReSPEC approach.

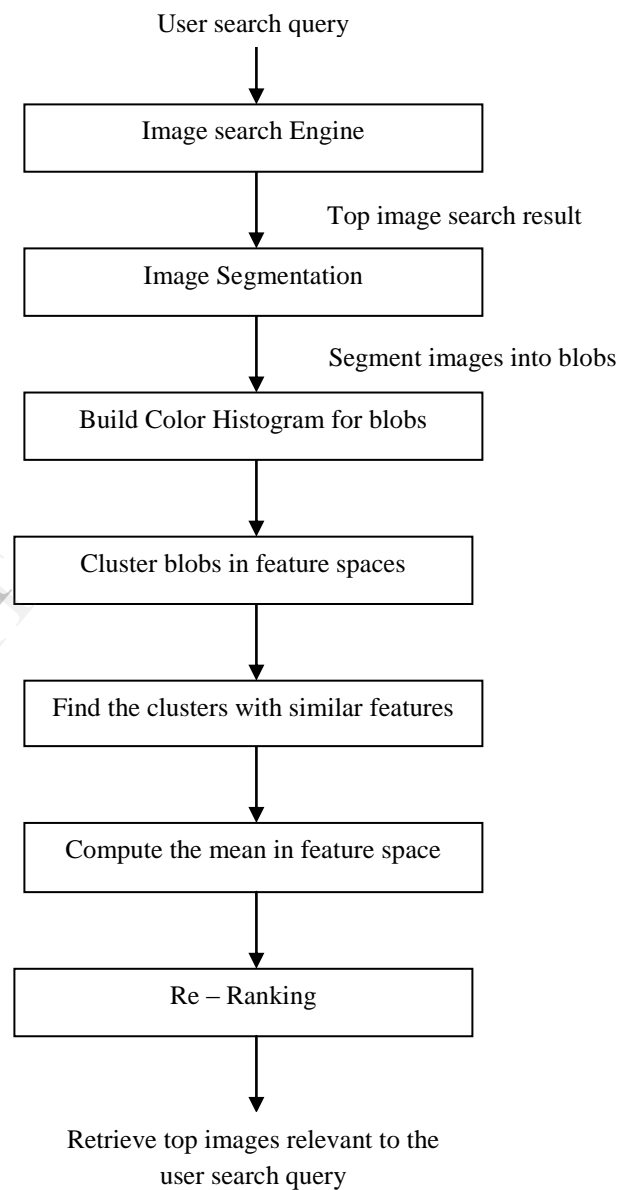


Figure 1: Data flow diagram of ReSPEC approach

2.1.1 Image Segmentation

In image segmentation each images collected from the image search engine has been broken into regions of

resemblance, with the intuition that each of these regions is a separate object in the image by using a graph based approach. Each pixel is a node in the graph with undirected edges connecting its adjacent pixels in the image. Each edge has a weight encoding the similarity of the two connected pixels. The partitioning is done such that two segments are merged only if the dissimilarity between the segments is not as great as the dissimilarity inside either of the segments.

2.1.2 Feature Selection

In order to obtain a measure of how similar image blobs are to one another, good features are needed to represent the blobs. Color histograms in HSV color space used to represent the image features. To form a feature vector for each blob, histograms are built for the H, S and V channels, with 15 bins each, and then concatenated together to form a 45 dimensional feature vector. Although histograms have clear advantage over taking the mean color of all the pixels in the blob, there is an inherent problem. For example, consider the following three histograms for hue: $X = (1; 0; 0; 0; 0; 0; 0; 0; 0)$; $Y = (0; 1; 0; 0; 0; 0; 0; 0; 0)$; $Z = (0; 0; 0; 1; 0; 0; 0; 0; 0)$. It is clear that X and Y are more similar in hue. Nevertheless, the distance between X and Y, and X and Z are equal.

2.1.3 Mean Shift Clustering in Feature Space

The next step in the system is to cluster the blobs, according to their extracted features with the hope that the object of interest will form the largest cluster. Since some of the blobs will represent garbage, it is difficult to predict the number of clusters that are present. Hence, a standard clustering approach such as k-means is not appropriate.

The mean shift clustering algorithm, which is an iterative gradient ascent method for finding local density maxima, was used instead. The main idea behind mean shift is to treat the points in the d-dimensional feature space as an empirical probability density function where dense regions in the feature space correspond to the local maxima or modes of the underlying distribution. For each data point in the feature space, one performs a gradient ascent procedure on the local estimated density until convergence. The stationary points of this procedure represent the modes of the distribution. Furthermore, the data points associated (at least approximately) with the same stationary point are considered members of the same cluster [6].

2.1.4 Re - ranking the Images

After obtaining the “significant” cluster in feature space, the mean is computed. The rest of the images are then resorted based on the distance of their blobs to this mean. Since each image could potentially contain more than one blob, the closest blob in each image is used. Chi-squared distance comparisons are used in the re-sorting because it is known that for histograms, a chi-squared distance measure yields better results than L2 distance.



Figure 2 (a): Search result before ranking



Figure 2 (b): Search result after ranking

The result is a re-ranking of the images from the original search engine. Figure 2 (a), (b) shows the collection of images before and after re-ranking.

2.2 IntentSearch: Interactive On-line Image Search Re-ranking

The previous approach has some demerits as follows:

1. Very small set of personal photographs, and saw some promising results. Due to the fact that our feature extraction is limited to color only, however, we found this to not be discriminative enough to sort large collections of highly variable photos.
2. Some irrelevant images appear after re-ranking.

In order to overcome these drawbacks Jingyu Cui et al [4] introduce a new approach that uses adaptive visual similarity to re-rank the text based search results. A query image is first categorized into one of several predefined intention categories, and a specific similarity measure is used inside each category to combine image features for re-ranking based on the query image and automatically implied user intention. In addition to searching a more flexible interface is provided to let users browse and play with all the images in the current search session, which makes web image search more efficient and interesting. This

can be addressed using two interfaces which allow user to either re-rank in Live Image Search, or browse in Rank Collage. Figure 3, shows the Data flow diagram of On-line Image Search Re-ranking

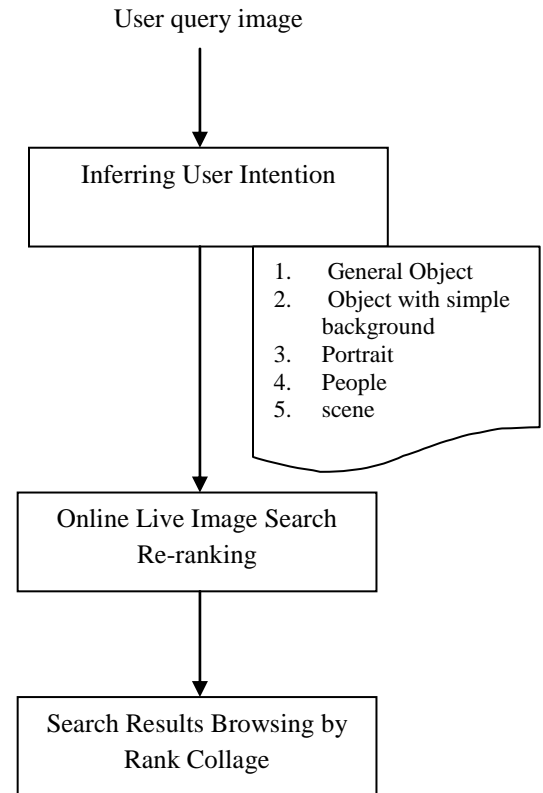


Figure 3: Data flow diagram of On-line Image Search Re-ranking

2.2.1 Inferring User Intention

Human can easily categorize images into high level semantic classes, such as scene, people, or object. Based on these general images are classified into typical intention categories as: 1. General Object. Images containing close-ups of general objects; 2. Object with Simple Background; 3. Scene. Scenery images; 4. Portrait. Images containing portrait of a single person; 5. People. Images with general people inside, and are not "Portrait".

The attributes for intention categorization [3] includes:

1. Face existence - Whether the image contains faces. ("Face", "Portrait")
2. Face number - Number of faces occurred in the image. ("Face", "Portrait")
3. Face size - The percentage of the image frame taken up by the face region. ("Portrait")
4. Face position - Coordinate of the face center relative to the center of the image. ("Portrait")
5. Directionality - Kurtosis of Edge Orientation Histogram. The bigger the Kurtosis is, the stronger the image shows directionality. ("Scene", "General object", and "Object with simple background")
6. Color Spatial Homogeneousness - Variance of values in different blocks of Color Spatialet describing whether color in the image is distributed spatially homogeneously. ("Scene")
7. Edge Energy -Total energy of edge map obtained from Canny Operator on the image. ("General object", "Object with simple background")

With these attributes, train a C4.5 decision tree on an image set with manually labeled intentions. The training process decides decision boundaries of the intention categories in the feature space defined by those attributes, and intention of a new input image is easily decided by applying the rules of the decision tree to it.

2.2.2 On-line Live Image Search Re-ranking

After typing a query keyword, the original result of Live Image Search based on text is presented to user. The user can then drag an image to the "Key Image" pad, and initiate a content-based query. Our background algorithm infers the best Intention for the query image, and then gives the re-ranked results based on an adaptive feature set. The user can either drag another image from current result to the "Key Image" pad for another round of query, or drag it to the "Additional Images" pad to let the system update the results[4]. Figure 4, shows Live Image Search Re-ranking.

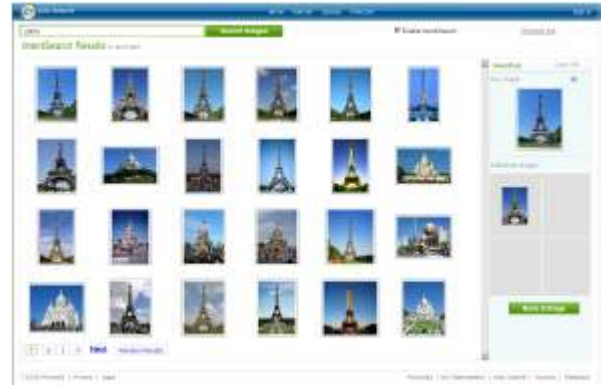


Figure 4: Live Image Search Re-ranking

2.2.3 Search Results Browsing by Rank Collage

User can switch to a Rank Collage view anytime to browse the whole set of images in current search session. All images are presented in a collage, with the images near the center being bigger and more relevant to user query, and images further from the center being less relevant. When a new image is dragged to the center, a new round of search is started with the new image as query. Endless zooming, various operations on a single image, and multiple search results side-by-side comparison are also supported. Figure 5, shows Search Results Browsing by Rank Collage [4].



Figure 5: Search Results Browsing by Rank Collage

2.3 IntentSearch: Capturing User Intention for One-Click Internet Image Search

The above mentioned technique have some drawbacks as follows: It is difficult to interpret users' search intention only by query keywords and this leads to ambiguous and noisy search results which are not satisfactory. It is important to use visual information in order to solve the ambiguity in text-based image retrieval. In order to overcome this, Xiaou Tang et al [1] proposed an Internet image search approach which requires the user to click on one query image with minimum effort and images from a pool retrieved by text-based search are re-ranked based on both visual and textual content. The users will tolerate one-click interaction, which has been used by many popular text-based search engines. For example, Google requires a user to select a suggested textual query expansion by one-click to get additional results. Figure 6, shows that various steps involved in capturing the users' search intention from this one-click query image.

2.3.1 Steps involved in capturing the users' search intention from this one-click query image:

2.3.1.1 Adaptive similarity

Designing a set of visual features to describe different aspects of image, how to integrate various visual features to compute the similarities between the query image and other images is an important problem. An Adaptive Similarity is introduced to deal with a user always has specific intention when submitting a query image. For example, if the user submits a picture with a big face in the middle, most probably he/she wants images with similar faces and using face-related features is more appropriate. The query image is first categorized into one of the predefined adaptive weight categories, such as "portrait" and "scenery." Inside each category, a specific pretrained weight schema is used to combine visual features adapting to this kind of images to better re-rank the text-based search result. This correspondence between the query image and its proper similarity measurement reflects the

user intention. This initial re-ranking result is not good enough and will be improved by the following steps [1].

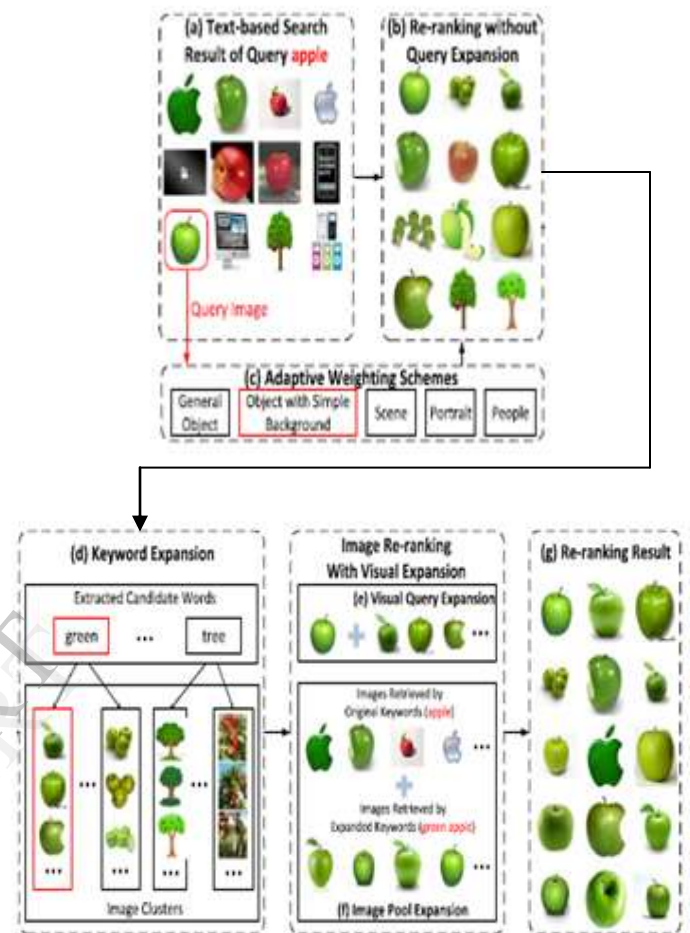


Figure 6: Capturing the users' search intention from this one-click query image

2.3.1.2 Keyword expansion

Query keywords input by users tend to be short and some important keywords may be missed because of users' lack of knowledge on the textual description of target images. Here, query keywords are expanded to capture users' search intention, inferred from the visual content of query images, which are not considered in traditional keyword expansion approaches [1]. A word w is suggested as an expansion of the query if a cluster of images are visually similar to the query image and all contain the same word

w¹. The expanded keywords better capture users' search intention since the consistency of both visual content and textual description is ensured.

2.3.1.3 Image pool expansion

The image pool retrieved by text-based search accommodates images with a large variety of semantic meanings and the number of images related to the query image is small. In this case, re-ranking images in the pool is not very effective. Thus, more accurate query by keywords is needed to narrow the intention and retrieve more relevant images. A naive way is to ask the user to click on one of the suggested keywords given by traditional approaches only using text information and to expand query results like in Google "related searches." This increases users' burden. Moreover, the suggested keywords based on text information only are not accurate to describe users' intention. Keyword expansions using both visual and textual information will better capture users' intention. They are automatically added into the text query and enlarge the image pool to include more relevant images. Feedback from users is not required.

2.3.1.4 Visual query expansion

One query image is not diverse enough to capture the user's intention. In Step 2, a cluster of images all containing the same expanded keywords and visually similar to the query image are found. They are selected as expanded positive examples to learn visual and textual similarity metrics, which are more robust and more specific to the query, for image re-ranking. Compared with the weight schema in Step 1, these similarity metrics reflect users' intention at a finer level since every query image has different metrics. Different from relevance feedback, this visual expansion does not require users' feedback.

3. Conclusion

Image search is a specialized data search used to find images. To search for images, a user may provide query

terms such as keyword, image file/link, or click on some image, and the system will return images "similar" to the query. This paper presents an overview of various techniques for effective image retrieval based on users' search intention. The extensive study of these three techniques reveals that "IntentSearch: Capturing User Intention for One-Click Internet Image Search" proves more profitable comparatively than the other two techniques. This approach involves user to click in the first step without increasing users' burden. This makes it possible for Internet scale image search by both textual and visual content with a very simple user interface. Further optimization could be done to this technique improve the quality of the retrieved images.

4. References

1. Xiaoou Tang, Fellow, IEEE, Ke Liu, Jingyu Cui, Student Member, IEEE, Fang Wen, Member, IEEE, and Xiaogang Wang, Member, "IntentSearch: Capturing User Intention for One-Click Internet Image Search" IEEE IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 34, NO. 7, JULY 2012
2. F. Jing, C. Wang, Y. Yao, K. Deng, L. Zhang, and W. Ma, "Igroup: Web Image Search Results Clustering," Proc. 14th Ann. ACM Int'l Conf. Multimedia, 2006.
3. J. Cui, F. Wen, and X. Tang, "Real Time Google and Live Image Search Re-Ranking," Proc. 16th ACM Int'l Conf. Multimedia, 2008.
4. J. Cui, F. Wen, and X. Tang, "IntentSearch: Interactive On-Line Image Search Re-Ranking," Proc. 16th ACM Int'l Conf. Multimedia, 2008.
5. "Bing Image Search," <http://www.bing.com/images>, 2012.
6. N. Ben-Haim, B. Babenko, and S. Belongie, "Improving Web-Based Image Search via Content Based Clustering," Proc. Int'l Workshop Semantic Learning Applications in Multimedia, 2006.
7. http://en.wikipedia.org/wiki/Image_retrieval
8. http://en.wikipedia.org/wiki/Content-based_image_retrieval