Scientific Journal Impact Factor (SJIF): 3.518
International Journal of Modern Trends in Engineering and Research
ISSN (ONLINE) : 2349 - 9745
ISSN (PRINT) : 2393 - 8161

Intent Search: Capturing User Intention for One-Click Internet Image Search

Pooja Sharma[1] and Mrunali Vaidya[2]
[1,2]Computer Science, Ballarpur Institute of Technology

Abstract: Web-scale image search engines (e.g., Google image search, Bing image search) mostly rely on surrounding text features. It is difficult for them to interpret users' search intention only by query keywords and this leads to ambiguous and noisy search results which are far from satisfactory. It is important to use visual information in order to solve the ambiguity in text-based image retrieval. In this paper, we propose a novel Internet image search approach. It only 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. Our key contribution is to capture the users' search intention from this one-click query image in four steps. 1) The query image is categorized into one of the predefined adaptive weight categories which reflect users' search intention at a coarse level. Inside each category, a specific weight schema is used to combine visual features adaptive to this kind of image to better re-rank the text-based search result. 2) Based on the visual content of the query image selected by the user and through image clustering, query keywords are expanded to capture user intention. 3) Expanded keywords are used to enlarge the image pool to contain more relevant images. 4) Expanded keywords are also used to expand the query image to multiple positive visual examples from which new query specific visual and textual similarity metrics are learned to further improve content-based image re-ranking. All these steps are automatic, without extra effort from the user. This is critically important for any commercial web-based image search engine, where the user interface has to be extremely simple. Besides this key contribution, a set of visual features which are both effective and efficient in Internet image search are designed. Experimental evaluation shows that our approach significantly improves the precision of top-ranked images and also the user experience.
Keywords: Image search, intention, image re-ranking, adaptive similarity, keyword expansion.

I. INTRODUCTION

Today, the flow of the large amount of multimedia information is significant; which increases the need to manipulate and access them as quickly as possible. 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. For example, the average query length of the top 1,000 queries of image search is 1.368 words, and 97 percent of them contain only one or two words. They cannot describe the content of images accurately. The search results are noisy and consist of images with quite different semantic meanings.

Fig. 1 shows the top rankedimages from Bing image search using "apple" as query. They belong to different categories, such as "green apple," "red apple," "apple logo," and "iphone" because of the ambiguity of the word "apple." The ambiguity issue occurs for several reasons. First, the query keywords' meanings may be richer than users' expectations. For example, the meanings of the word "apple" include apple fruit, apple computer, and apple ipod. Second, the user may not have enough knowledge on the textual description of target images. For example, if users do not know "gloomy bear" as the name of a cartoon character (shown in Fig. 2a) and they have to input "bear" as query to search images of "gloomy bear." Lastly and most importantly, in many cases it is hard for users to describe the visual content of target images using keywords accurately.

Inordertosolvetheambiguity,additionalinformationhas tobeusedtocaptureusers' searchintention. Onewayistextbased keyword expansion, making the textual description of the query more detailed. Existing linguistically-related methods find either synonyms or other linguistic-related wordsfromthesaurus,orfindwordsfrequentlyco-occurring with the query keywords. For example, Google image search provides the "Related Searches" feature to suggest likely keyword expansions. However, even with the same query keywords, the intention of users can be highly diverse and cannotbeaccuratelycapturedbytheseexpansions.Asshown in Fig. 2b, "gloomy bear" is not among the keyword expansions suggested by Google "related searches."

Fig. 1. Top-ranked images returned from Bing image search using "apple" as query.

(a) Images of "gloomy bear"

(b) Google Related Searches of query "bear".

Fig. 2. (a) Images of "gloomy bear." (b) Google related searches of query "bear."

Another way is content-based image retrieval with relevance feedback. Users label multiple positive andnegative image examples. A query-specific visual similarity metric is learned from the selected examples and used to rank images. The requirement of more users' effort makes it unsuitable for web-scale commercial systems like Bing image search and Google image search in which users' feedback has to be minimized.

II. LITERATURE SURVEY

Image search is a type of data search that determines images. For searching required images, a user may give several disparate query terms. For example keyword or click on few images. Further the system will determine images almost similar to the query. The search engines, such as Google, Yahoo, Bing and so on, generally depends upon the textual query.

User will enter the keyword for the particular desired image. It is assumed that keywords in the search engine should be able to match at least one among the group of common tags. On this basis, the system retrieves and displays the group of resultant images. But for these types of searching methods, it is not easy to capture the user's intention swiftly. The string based searching methodology is basically used by these systems. Thus, the result generated from these systems may contain the noisy results. Thus, it is mandatory to use visual information to get the easy access. This method is well known as text based keyword expansion.

III. EXISTING METHODOLOGY

In the existing system, the textual based keyword expansion is used making the description of the query more detailed than necessary. Also, the recent search engines find either linguistic-related words or synonyms from various online dictionaries like thesaurus that co occurs with the keywords.

For instance, "Related Searches" feature is provided in the Google image search to suggest likely keyword expansions. However, the intention of the users, even with the similar keywords, may not be accurately captured by expansions. On the other hand, the image search works well for content that is usually described on the web. And therefore, the result will be more relevant and better than the textual based keywords searching.

We design and adopt a set of features that are both effective in describing the visual content of images from different aspects, and efficient in their computational and storage complexity. Some of them are existing features proposed in recent years. Some new features are first proposed by us or extensions of existing features. It takes an average of 0.01 ms to compute the similarity between two features on a machine of 3.0 GHz CPU. The total space to store all features for an image is 12 KB. More advanced visual features developed in recent years or in the future can also be incorporated into our framework.

3.1. Existing Features

- **SIFT**: For describing regions around Harris interest points, we adopt 128-dimensions SIFT.
- *Gist*: We use Gist that works well for scenery images and characterizes the holistic appearance of an image.
- **Daubechies Wavelet**:To characterize the texture properties in the image, we use the second order moments of wavelet coefficients in various frequency bands.
- **Histogram of Gradient(HoG):** Histogram of Gradientis effective for images with long and strong edges and reflects distributions of edges over different parts of an image.

3.2.New Features

- **Attention Guided Color Signature:** Color signature describes the color composition of an image. After clustering colors of pixels in the LAB color space, cluster centers and their relative proportions are taken as the signature. We propose a new Attention Guided Color Signature (ASig) as a color signature that accounts for varying importance of different parts of an image. We use an attention detector to compute a saliency map for the image, and then perform k-Means clustering weighted by this map. The distance between two ASigs can be calculated efficiently using the Earth Mover Distance algorithm.
- **Color Spatialet (CSpa):** We design a novel feature, Color Spatialet, to characterize the spatial distribution of colors.
- **Multilayer Rotation Invariant (EOH):** Edge Orientation Histogram describes the histogram of edge orientations. We incorporate rotation invariance when comparing two EOHs, rotating one of them to best match the other. This results in a Multilayer Rotation Invariant EOH (MRI-EOH). Also, when calculating MRI-EOH, a threshold parameter is required to filter out the weak edges. We use multiple thresholds to get multiple EOHs to characterize image edge distributions at different scales.
- **Facial Feature:** Face existence and their appearances give clear semantic interpretations of the image. We apply face detection algorithm to each image, and obtain the number of faces, face sizes, and positions as features to describe the image from a “facial” perspective.

IV. PROPOSED SYSTEM

In this Proposed Methodology, we propose a novel Internet image search approach. In this the user is only required to give one click on a query image. The resultant pool of images that is retrieved by text based is re-ranked on the basis of visual and textual query, similar to the query image (as shown in figure 3). We believe that one click interaction, which has been used by many popular text-based search engines, would be tolerated by many users. For example, the user is required to get the additional results by one click interaction of the suggested keywords. But the key problem that is to be solved is how to capture user intention from the one-click query image.

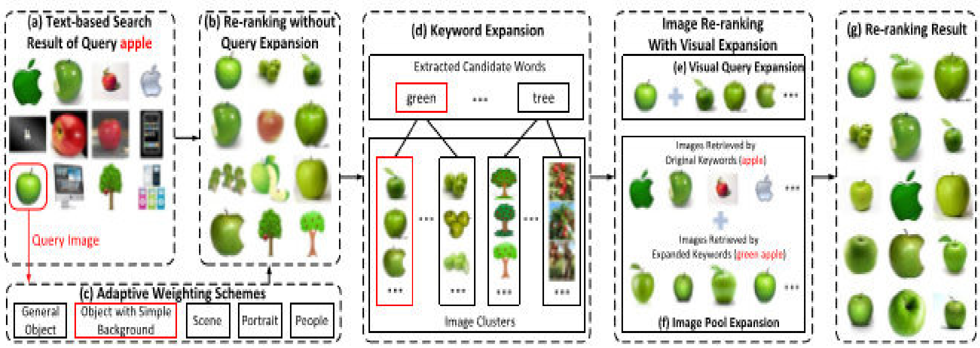

Fig. 3. An example to illustrate our algorithm. The details of steps (c)-(f) are given in Sections 3.3, 3.4, 3.5, and 3.6.

3.3.New Approach:

- Multi-Layer Rotation Invariant EOH
- Color Spatialet
- Attention Guided Color Signature
- Facial Feature

3.4.Modules:

3.4.1. Image Search.

Today, various Internet scale image search methods are text-based but are limited by the fact that query keywords cannot describe image content precisely. The visual features for evaluating image similarity are used by content-based image retrieval. But the major challenge of the content-based image retrieval is to study the visual similarities that will reflect the identical images. It can be learnt by a large training set of the relevance of pairs of images.

3.4.2. Query Categorization.

The query categories we considered, in this module, are: Object with Simple Background, General Object, Scenery Images, People and Portrait. To train a C4.5 decision tree for query categorization, we use 500 manually labeled images, 100 for each category. the number of faces in the image, existence of faces, the coordinate of the face center relative to the center of the image, the percentage of the image frame taken up by the face region, Color Spatial Homogeneousness (variance of values in different blocks of Color Spatialet, Section 3.2), Directionality (Kurtosis of Edge Orientation Histogram, Section 3.2), Edge Spatial Distribution (the variance of edge energy in a 33 regular block of the image, characterizing whether edge energy is mainly distributed at the image center) and total energy of edge map obtained from Canny operator.

3.4.3. Visual Query Expansion.

The goal of visual query expansion is to obtain multiple positive example images to learn a visual similarity metric. It is more robust and more specific to the query image. For example, the query keyword entered by user is “Paris” and the query image is an image of “Eiffel tower”. The re-ranking result of the query search is on visual similarities without visual expansion; with many irrelevant images among the top-ranked images. This happens because the visual similarity learnt from one simple query image is not strong enough.Such irrelevant images can be filtered out, by adding more positive examples to learn a more robust similarity metric, such irrelevant images can be filtered out. In a traditional way, adding additional positive examples was typically done through relevance feedback, which required more users’ labeling burden. We aim at developing an image re-ranking method which only requires one-click on the query image and thus positive examples have to be obtained automatically.

3.4.4. Images Retrieved by Expanded Keywords.

In this module, considering efficiency, image search engines, such as Bing image search, only re-rank the top *N* images of the text-based image search result. If the query keywords do not capture the user's search intention accurately, there are only a small number of relevant images with the same semantic meanings as the query image in the image pool. Visual query expansion and combining it with the query specific visual similarity metric can further improve the performance of image re-ranking.

IV. CONCLUSION

In this paper, we propose a novel Internet image search approach which only requires one-click user feedback. Intention specific weight schema is proposed to combine visual features and to compute visual similarity adaptive to query images. Without additional human feedback, textual and visual expansions are integrated to capture user intention. Expanded keywords are used to extend positive example images and also enlarge the image pool to include more relevant images. This framework makes it possible for industrial scale image search by both text and visual content. The proposed new image re-ranking framework consists of multiple steps, which can be improved separately or replaced by other techniques equivalently effective. In future work, this framework can be further improved by making use of the query log data, which provides valuable co-occurrence information of keywords, for keyword expansion. One shortcoming of the current system is that sometimes duplicate images show up as similar images to the query. This can be improved by including duplicate detection in the future work. Finally, to further improve the quality of re-ranked images, we intend to combine this work with photo quality assessment work to re-rank images not only by content similarity but also by the visual quality of the images.