Linking the Text and Visual Features in Vertical Image Search Engine: A Survey

Nayana N Kumar (PG Scholar)
Department of CSE
Vemana IT
Bengaluru-34, India

Usha K (PG Scholar)
Department of CSE
Vemana IT
Bengaluru-34, India

Jayashree L K (Selection Grade)
Department of CSE
Vemana IT
Bengaluru-34, India

***Abstract*— The image search based on the content is a challenging problem in the current trend on the Internet; this is mostly due to the semantic gap that exists between and high-level terms and low-level visual features. And the excessive computation brought by huge amount of images and high dimensional features. The concept of presenting ILIKE, which is a vertical image search engine that integrates both visual features from image content, and textual features from web pages for better image search. Here the problem is tackled by trying to identify the meaning of each text term in the visual feature space and re-weighting visual features according to their significance to the query content.**

Keywords — Vertical search engine, retrieval, tagging

I. INTRODUCTION

With the development of the internet, large volumes of multimedia data have become popular and they are obtainable in Online. Multimedia data's that involves text, images, video, audio or combination of all these .The images are extracted from the web based on the visual contents. But there are some primary key challenges in extracting images based on their contents. (1), Visual qualities of the images are not related with the content because there occurs a semantic gap. (2)It is a tedious task to deal with the very large scale image database because text-indexing is easier than indexing the large dimensional data (3) for the general user it is very difficult to sketch a good query. So the new approach applies, ILIKE as a vertical search Engine. Mainly this search engine is a product which is implemented for attire (apparels) shopping. This engine combines visual features [1][2][3][4]and text and performance of image retrieval is improved. In this search engine, there is much better chance to combine visual and textual features. So this paper presents a vertical search engine- ILIKE, where textual and visual contents linkup and corresponds with each other. In this approach, we identify the relationships that coexist between image features extracted from product pictures and textual features extracted from product descriptions. Then both types of features are combined to build a bridge across the semantic gap. There are three contributions technically: 1) bridging the semantic gap by combining textual and visual features. 2) Bridging the user intention gap between user information necessities. 3) By evaluating the representations of keywords in the visual feature space, it is able to identify the semantic relationships of the terms. Now in this approach the system will be able to recognize or grasp user's perception or the visual intentions for search terms, and then applying such intents to leverage on relevance ranking and assessment. And automatically generating a thesaurus based on the visual semantics of words. So our system improves the search performance.

II. RELATED WORKS

A. *Retrieving of Images Based on Content and tagging of images*

In Existing image retrieval systems the images are manually annotated with metadata, and they use text-based retrieval to search on tags. But manual annotation increases time complexity for very large scale image databases. And it is difficult to describe images accurately with a set of keywords. The primary challenge in retrieving of images based on content is the semantic gap that exists between the Low-level visual features and the High-level image. To solve these type of issues, Content Based Image Retrieval systems [5][6][7][8] were developed. And next tagging of images is done automatically by adding tags and available metadata for images. Automatic image tagging techniques is considered as a classification problem, In order to build a classifier which identifies the mapping between the low-level image features and the images with the tags [13][14][15]. Here the goal is to train the classifier by assigning some testing samples with the highest likelihood. The approach of Text-image interaction methods makes use of visual information in annotating the images. The approach of automatic tagging seems to be efficient when there are keywords with frequent occurrence and having strong visual similarities. But it is difficult to annotate the images with more specification and visually less similar keywords. To overcome the difficulty of manual Tagging and in improving the quality of the image tags, many automatic tag recommendation systems are developed [9][10][11][12]. In current trend number of growing social network sites allows tagging of photos and sharing. These methods are used to develop fully automatic and folksonomic tag recommendation systems. This system leverages the set of vocabulary from a group of users, which is less susceptible to noise than an individual's subjective annotation, resulting in high-quality image tags.

B. *Searching of images on the Web*

At present the web image search engines like Google and Bing depend on textual metadata. By taking textual queries and matching them with the metadata that are associated with the images, like URL, image file name, and other text surrounding in the webpage containing the image. But the textual information provided for an image may not define the image content, and it is very difficult to describe visual content using text, the performance of retrieving the metadata-based searches are still considered to be poor. There are more efficient text-based methods to associate the semantic information with the images to improve the search performance [18][19][20]. Here a two-stage Hybrid approach is been introduced for a text-based search using several prototypes for Content-based image search for the web [21][22][23]. First it will generate an intermediate result with low precision and high recall, and then applying CBIR to cluster will rerank the results. The main idea of applying CBIR for text based search seems to be a better alternative for clustering or reranking the results. And there are also image retrieval systems that work on offline with domain specific collections of images, like personalized albums, flower image search, arts images search and so on. Hence these perspectives make use of domain specific knowledge in image feature selection, measuring of similarity, and preprocessing of images [24][25]. For example, personal album searches may depend on face recognition while flower image searches may rely on shape, color and texture features to improve search performance.

III. SYSTEM ARCHITECTURE

The system architecture of the ILIKE vertical search engine. The ILIKE system consists of three major Components. As shown in Fig.1:

(1) CRAWLER

(2) PROCESSOR

(3) A SEARCH COMPONENT.

The summarization of the system procedure as is follows:

1 When the user inputs a query text terms to the browser.

2 The input for each text-based term, the image search engines, like Google Image Search, PicSearch,

Bing, AltaVista Image Search, will forward to the Crawler.

3 Then the Crawler sends the query to each search engine and fetches the product pages from different retailer websites.

4 The parser will collect descriptions for items and it will generate the terms by indexing them.

5 Simultaneously the image processor will extract the visual features from the items Using the URLs, the Image Crawler recovers the images from the Internet to construct the initial image set;

6 Then integrating the textual features is done in a reweighting scheme and construction of visual thesaurus for each text term is done .Feature Extractor computes the content of image feature vectors for all images in the initial image set.

7 The search component provides a query interface and to browse the views of search result.

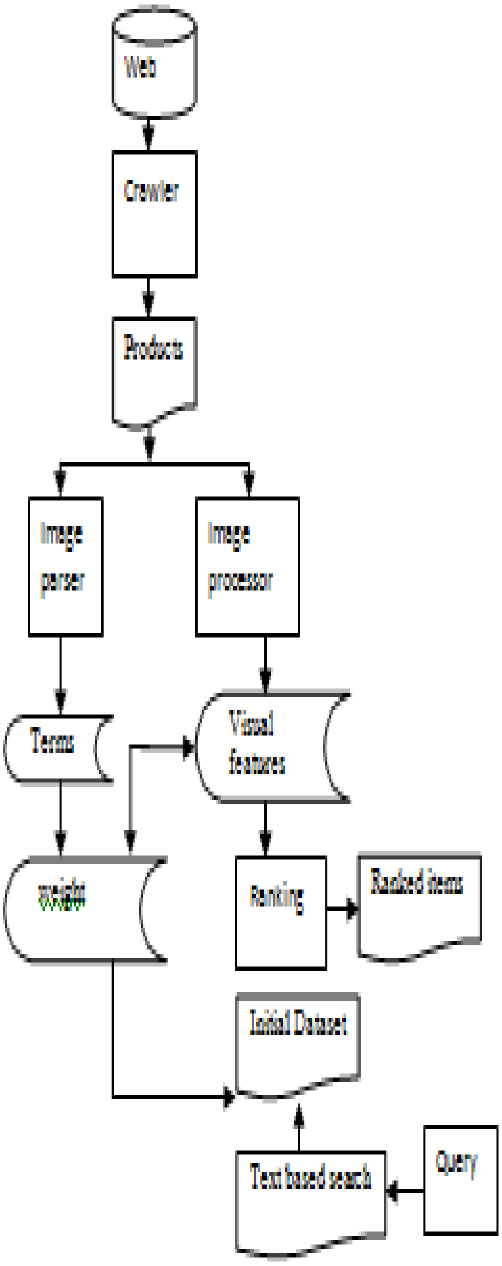

Fig.1. System architecture of the ILIKE vertical search engine

THE METHOD

The roles of textual feature space and visual feature space are complementary in multimedia. Textual information better illustrate the logical meaning, while visual features play a dominant role at the physical level. They are isolated by the semantic gap, which is the at most barrier in content-based image retrieval.

A. Keywords Representation

The textual analysis is a estimate of the commentator's recognition of the image content. There are some complications using only text features to recover the mixtures of image/textual terms. Furthermore, calculating text similarity is hard—distance calculation do NOT perfectly represent the distances in human recognition. Though, they are equally different in contents of textual representation. To make up the lack of pure QBIC approaches, we traverse the connections between visual and textual feature subspaces. The text description constitutes the commentator's recognition of the visual features. Furthermore, if the Stability is observed over a significant number of contents reported by the similar keyword, such a set of features and their Significance may constitute the human "visual" recognition of the keyword. if contents with different explanations demonstrate a different values on these selected visual features, we can further confirm the connection between the contents and visual features. let us look see at the items with the keyword "dotted" in their illustration (shown in Fig 2). There are very unique texture features. They also differ in other features, like color and shape. So the term "dotted" is used to sketch certain texture features. When a user explore with this content, her beginning is to find such texture features, not about the color or shape.

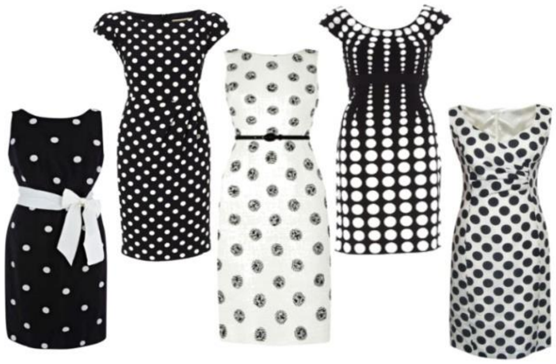

Fig.2. some items with keyword"dotted" in their explanation

In this approach, many contents could be connected with such a "visual meaning." In ILIKE, the first step is to identify such "visual meanings" automatically.

B Weighting Visual Features

If we have other way, it illustrates two Groups: 1) positive: N1 contents that have the keyword in their explanation, and 2) negative: N2 contents that do not have the keyword. In this approach, if the meaning of a keyword is rational with a visual feature, its N1 values in the positive group should indicate a different distribution than the N2 values in the negative group. Furthermore, the feature values in the positive group move to indicate a little variance, while values in the negative group are expanded. When we contrast distributions, we do not make such expectations. In the experimental results, we will show that ILIKE is able to recover such items without the false hits. Kolmogorov-Smirnov (K-S) test captures the difference of two distributions across the dimensions of feature vectors. Human perception is important of the keyword. Contents with and without the keyword has statistically distinct values on the visual features, and such features are not connected with the keyword. Keywords of visual features are reweighed for each, we extend the features of keyword that are significant, while disappears the others. For the convenience of discussion we group the visual features, and they might overlaps with each other. Large value can bearer generated by statistically varying negative and positive samples. In this approach, when users search with term "pattern," we can consider that she is interested in texture features; hence the other color and shape features are given less importance. Here , further can be retrieve contents with similar visual presentation , but does not have the particular term ("pattern") in their explanation. It is difficult to describe the human visual perception for some keywords. Fortunately, this approach is still capable of judge such intention. It is not easy for a user to summarize the characteristics of "dotted" contents. However, when we go through, the visual meaning is obvious. "Dotted" contents share some different distributions in the Color and shape, while they are expanding in Strength and high-frequency in the textual features. In order to create a enough coverage of an image's meaning of semantic, we attempt to expand the part of quality selection.

C Finding of words Visually[Thesaurus]

Thesauri are used over a large area in data retrieval, in the query expansion of linguistic preprocessing. Although higher quality of thesauri is manually generated, it is very labor intensive of developing process. Meanwhile, by using statistical thesauri can be generated. In ILIKE, different types of visual thesaurus can be generated, based on the visual space in the phrase distributions that is the statistical similitude of the visual description of the phrase. In ILIKE, two phrases are compared in terms of "visual semantics" if they are used to define visually similitude contents. Since terms are used to define many contents, the similarity is assessed statistic significance across all the contents defined by both terms. We can also see that some no adjective terms show similarity in moderate with many other terms. We remove the high-frequency terms by post processing. the terms with the similarity set of significant feature element but, by the consistently opposite values . Meantime, the meaning of “white” and “pale” are similar, and the “gray” is different. In the dictionary we enumerate the term-wise similarity, to generate the “visual Word Net”. Some examples are In the Table 1. This visual thesaurus can be used for query for search engines existing in text-based product.

D process of Weight Vector

As we have given, product explanation could be very instinctive because of personal tastes. Different retailers may use different terms to tag similitude objects. Due to synonyms, we can observe false negatives In the sets of negative. A false negative is an content that: 1) is actually applicable to the content, 2) with the positive contents demonstrates the similar visual features 3) describes the synonym of term and categorizes the terms of negative set. The visual thesaurus can help to find both antonyms and synonyms. Synonyms described by merging contents, so we can reduce the false positive Contents caused by the synonyms; so, we can observe higher stability on notable features, and can get higher weights. In ILIKE, for all the terms in the dictionary we generate visual thesaurus. Later, for each Term, we add the contents explanation by its positive set and its top synonyms. According to the updated negative / positive Sets We recalculate the new vector weight. The negative and positive sets from the combined set are shown in the distributions. Combined positive set is narrower and cleaner can seen in the feature distribution. Combining similitude keywords in the thesaurus of visual, the quality of the vector weights can be improved.

E Feature Quality and Correlation

In QBIC, the low-level visual features is widely Used widely for image annotation and selection in feature. if low weight for all the terms in the dictionary produces, it is “useless” hence in weighted queries it will always have a very low value. High weight for all terms can be produced, so it is not a good feature because it does not represent any definite semantic meaning. Hence, we do not find the feature that is significant for the keywords. In Section 4.b, weight vector for each keyword is generated. For each feature in visual, weight values is collected across all Keywords. Here we have the positive and negative contents. For some terms good feature produces low weights, and high Weights for the others. With the higher entropy we can observe the features of semantic meaning. The initial feature is to distinguish the negative and positive sets for some terms. This is accurate with the QBIC literature. Features may be related to each other. In ILIKE, if there is similarity set of keywords are significant, and for the others it is insignificant, they are correlated. For correlations in the selected features in visual. we do not find the feature that is significant for the keywords. It can be seen that features are more independent, That is same type of features in moderate correlations. There is stronger correlations among PC and CF features. So the vector of new weight can be calculated. It introduces computational overhead in ILIKE, but the effect on search accuracy is very limited.

F Search and Query Expansion

In ILIKE, we use classic text-based search to get an initial set. For the keywords in users query, the system loads its comparable weight, correlation Feature: features are independent; some PC and CF features are related. It reformulates a seed query to improve to retrieving the performance in information retrieval operations in the search engine. Finding the synonym of words. And searching the synonyms as well in the search and query of the items or contents. The original query terms are reweighted. Search the query to match additional documents. The goal of expansion in the query this regard increases the recall, and potentially precision increases.

Table 1: Visual Thesaurus

Words	*Words in visual thesaurus*	*words in thesaurus after first iteration*
Saree	print,embroier	border,short,print,paint, embroider
Sports	outdoor,fitnes	adventure,kits,fitness, fashion
footwear	casual,sandals	formal,sports,new, casual,wedge,lace

IV. CONCLUSION AND DISCUSSIONS

The purpose of this survey is to provide an overview of the functionality of ILKIE, a vertical search engine for apparel shopping. The main goal is to combine the visual features and textual to improve search performance. So text terms are represented in the visual feature space, to develop a text-guided weighting scheme for visual features. This weighting scheme assumes user intention from query terms, and magnifies the visual features that are significant towards such intention. Hence ILIKE is capable and effective in bridging the semantic gap. Through the comprehensive user study, ILIKE has demonstrated outstanding performance for a large number of descriptive terms. In some cases, it does not work well for some keywords. Many of such words have abstract meaning and are unlikely to be included in queries (e.g., zip, logo).Finally to conclude, by combining textual and visual features, ILIKE is able to pick “good” features that reflect users’ perception, and therefore is effective for vertical search.

REFERENCES

[1] B. Luo, X. Wang, and X. Tang, “A World Wide Web Based Image Search Engine Using Text and Image Content Features,” Proc. IS&T/SPIE, vol. 5018, pp. 123-130, 2003.

[2] Y. Chen, N. Yu, B. Luo, and X.-w. Chen, “iLike: Integrating Visual and Textual Features for Vertical Search,” Proc. ACM Int’l Conf. Multimedia, 2010.

[3] X. Tang, K. Liu, J. Cui, F. Wen, and X. Wang, “Intentsearch:Capturing User Intention for One-Click Internet Image Search,”IEEE Trans. Pattern Analysis Machine Intelligence, vol. 34, no. 7,pp. 1342-1353, July 2012.

[4] X. Wang, K. Liu, and X. Tang, “Query-Specific Visual SemanticSpaces for Web Image Re-Ranking,” Proc. IEEE Conf. ComputerVision Pattern Recognition (CVPR), June 2011.

[5] W.-Y. Ma and H.-J. Zhang, “Content-Based Image Indexing andRetrieval,” Handbook of Multimedia Computing, CRC Press, 1998.

[6] D. Cai, X. He, Z. Li, W.-Y. Ma, and J.-R. Wen, “HierarchicalClustering of WWW Image Search Results Using Visual, Textualand Link Information,” Proc. 12th ACM Int’l Conf. Multimedia,2004.

[7] A.W.M. Smeulders, S. Member, M. Worring, S. Santini, A. Gupta, and R. Jain, “Content-Based Image Retrieval at the End of the

[8] Early Years,” IEEE Trans. Pattern Analysis Machine Intelligence,vol. 22, no. 12, pp. 1349-1380, Dec. 2000.

[9] M.S. Lew, N. Sebe, C. Djeraba, and R. Jain, “Content-BasedMultimedia Information Retrieval: State of the Art and Challenges,”ACM Trans. Multimedia Computing, Comm., and Applications,vol. 2, no. 1, pp. 1-19, 2006.

[10] S. Wang, F. Jing, J. He, Q. Du, and L. Zhang, “IGroup: PresentingWeb Image Search Results in Semantic Clusters,” Proc. SIGCHIConf. Human Factors in Computing Systems, 2007.

[11] J. Jeon, V. Lavrenko, and R. Manmatha, “Automatic ImageAnnotation and Retrieval Using Cross-Media Relevance Models,”Proc. ACM SIGIR Conf. Research and Development in InformationRetrieval, 2003.

[12] J. Li and J.Z. Wang, “Real-Time Computerized Annotation ofPictures,” IEEE Trans. Pattern Analysis Machine Intelligence, vol. 30,no. 6, pp. 985-1002, June 2008.

[13] R. Datta, D. Joshi, J. Li, James, and Z. Wang, “Image Retrieval:Ideas, Influences, and Trends of the New Age,” ACM ComputingSurveys, vol. 39, article 5, 2006.

[14] H. Tamura, S. Mori, and T. Yamawaki, “Textural FeaturesCorresponding to Visual Perception,” IEEE Trans. Systems Manand Cybernetics, vol. SMC-8, no. 6, pp. 460-473, June 1978.

[15] R.M. Haralick, K. Shanmugam, and I. Dinstein, “Textural Featuresfor Image Classification,” IEEE Trans. Systems Man and Cybernetics,vol. SMC-3, no. 6, pp. 610-621, Nov. 1973.

[16] J. Cui, F. Wen, and X. Tang, “Real Time Google and Live ImageSearch Re-Ranking,” Proc. 16th ACM Int’l Conf. Multimedia, 2008.

[17] L. Wu, L. Yang, N. Yu, and X.-S. Hua, “Learning to Tag,” Proc.18th Int’l Conf. World Wide Web (WWW), Apr. 2009.

[18] G. Carneiro, A.B. Chan, P.J. Moreno, and N. Vasconcelos,“Supervised Learning of Semantic Classes for Image Annotationand Retrieval,” IEEE Trans. Pattern Analysis Machine Intelligence,vol. 29, no. 3, pp. 394-410, Mar. 2007.

[19] S. Raimondo, S. Simone, C. Claudio, and C. Gianluigi, “Prosemantic Features for Content-Based Image Retrieval,” Proc. Seventh Int’l Workshop Adaptive Multimedia Retrieval, 2009.

[20] L.S. Kennedy, S.-F. Chang, and I.V. Kozintsev, “To Search or toLabel?: Predicting the Performance of Search-Based AutomaticImage Classifiers,” Proc. ACM Int’l Workshop Multimedia InformationRetrieval (MIR), 2006.

[21] Z.-H. Zhou and H.-B. Dai, “Exploiting Image Contents in WebSearch,” Proc. 20th Int’l Joint Conf. Artificial Intelligence (IJCAI),2007.

[22] X. Li, L. Chen, L. Zhang, F. Lin, and W.-Y. Ma, “Image Annotation by Large-Scale Content-Based Image Retrieval,” Proc. 14th Ann.ACM Int’l Conf. Multimedia, 2006

[23] F. Jing, C. Wang, Y. Yao, K. Deng, L. Zhang, and W.-Y. Ma,“IGroup: Web Image Search Results Clustering,” Proc. 14th ACMInt’l Conf. Multimedia, 2006.

[24] J. Cui, F. Wen, and X. Tang, “Intentsearch: Interactive On-LineImage Search Re-Ranking,” Proc. 16th ACM Int’l Conf. Multimedia,2008.

[25] L. Zhang, Y. Hu, M. Li, W. Ma, and H. Zhang, “EfficientPropagation for Face Annotation in Family Albums,” Proc. 12thACM Ann. Int’l Conf. Multimedia (Multimedia), 2004.

[26] B. Manjunath, J.-R. Ohm, V. Vasudevan, and A. Yamada, “Colorand Texture Descriptors,” IEEE Trans. Circuits and Systems forVideo Technology, vol. 11, no. 6, pp. 703-715, June 2001.

[27] Z. Wang, Z. Chi, and D. Feng, “Fuzzy Integral for Leaf ImageRetrieval,” Proc. IEEE Int’l Conf. Fuzzy Systems (FUZZ), 2002.

[28] J.-X. Dua, X.-F. Wang, and G.-J. Zhang, “Leaf Shape Based PlantSpecies Recognition,” Applied Math. and Computation, vol. 185,pp. 883-893, 2007.

[29] I. Kompatsiaris, E. Triantafyllou, and M. Strintzis, “A World WideWeb Region-Based Image Search Engine,” Proc. 11th Int’l Conf.Image Analysis and Processing (ICIAP), 2001.

[30] H. Lieberman, E. Rozenweig, and P. Singh, “Aria: An Agent forAnnotating and Retrieving Images,” Computer, vol. 34, no. 7,pp. 57-62, July 2001.

[31] C. Wang, L. Zhang, and H.-J. Zhang, “Learning