STEREOTYPE-BASED SEMANTIC EXPANSION FOR IMAGE RETRIEVAL

Jungin Lee[1], *OSung Kwon*[2], *Youngwoon Lee*[2], *Sund-Eui Yoon*[1,2]

Korea Advanced Institute Science and Technology
[1]Division of Web Science and Technology, [2]Dept. of Computer Science
{ezdevil,miruce78,lywoon89,sungeui}@gmail.com

ABSTRACT

We present a novel, stereotype-based semantic expansion approach to identify various image sets that stereotypically represent different aspects of a given keyword. Specifically, given an adjective keyword query, our method expands it to a set of noun sub-keywords, which are stereotypical examples that can be described by the given adjective (e.g., "cute" to "{infant, kitten, ...}"). We also perform a similar process for noun keywords with adjectives (e.g., "infant" to "{cute, sweet, ...}"). To perform such expansion, we use Google Books n-grams, a new corpus of 500 million books. We harvest stereotypical relationships among nouns and adjectives by utilizing useful lexical patterns such as similes on n-grams. To demonstrate benefits of our method, we have applied our method to text-based image retrieval. Our method shows a diverse set of images given tested keywords. According to a small scale user study with 12 participants, our method shows a higher recall ratio of what a user wants to find, compared to returning images only from original keywords.

1. INTRODUCTION

For text-based image retrieval, users typically provide a keyword. Most of prior techniques then identify images that have the exactly same tag to the given keyword [1]. In this case we may not have diverse output results given a keyword. For example, given a keyword "tall", Google Image Search [2] returns four images showing tall and short people together and one tall statue in the top five results (see images under "tall" keyword in (a) of Fig. 4). This is mainly because those images are frequently appeared in its image database and their associate texts match exactly to the keyword "tall". To ameliorate this problem, Google Image Search provides related words (e.g., "tall man" and "tall people") to the given keyword "tall". Nonetheless, Google Image Search provides only a limited set of related keywords that are identified as frequently co-occurring terms from recent user query inputs.

To address this problem, query expansion techniques have been proposed to improve recall while achieving precision given a query [3, 4]. At a high level, these techniques expand an initial keyword to its related keywords, while avoiding topic drifting. Query expansion studied in text retrieval can be classified broadly as lexical and statistical approaches. Lexical query expansions take advantage of global relationships between words such as holonym (i.e. part-whole relationship), which can be commonly derived from various knowledge bases such as WordNet [5]. Such global relationship between words have been recently adopted for efficient object recognition [6] as prior knowledge for semantic similarity in the visual domain based on Bag-of-visual-Words (BoWs) [8], a visual concept corresponding to words in text retrieval.

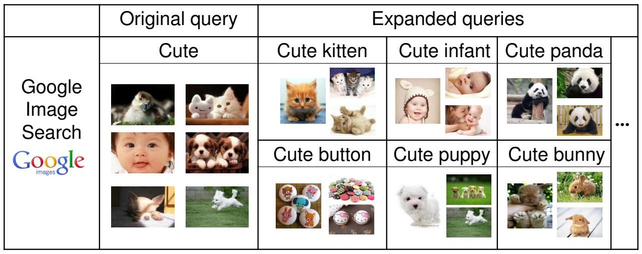


Fig. 1: This figure shows results of Google Image Search [2] with a keyword and its expanded sub-keywords based on our method. Our approach is able to effectively identify different, stereotypical concepts of the keyword.

Statistical-driven query expansion approaches, on the other hand, identify word relationships by looking into term co-occurrences. These techniques identify co-occurring terms globally from all the data in a corpus or locally tailored to given queries. Statistical approaches have been also actively adopted for text-based image retrieval. Also, a similar concept, identifying co-occurring terms, is used in the visual domain [9].

Departing from query expansion techniques employed in text-based image retrieval, we propose a novel, *stereotype-based semantic expansion* approach. Our approach starts from identifying cultural, stereotypical word relationships that are not typically represented in WordNet. Specifically, we use Google Books n-grams [10] that are extracted from five million digitized books containing about 4% of books ever printed, as our knowledge base for identifying stereotypical semantic relations between words. This new type of data sets has not been widely investigated for image retrieval, compared to WordNet.

n-grams are defined by n consecutive words that are frequently identified on the corpus. "boy" and "3.14159" are examples of 1-grams. "chicken and egg" is an example of 3-grams. Moreover, each n-grams are supplemented by its published year, match count, page count, and volume count. Various stereotypical semantic relationship (e.g., an association between "cute" and "baby") from Google Books n-grams can be extracted by looking into useful lexical patterns such as similes (e.g., "as cute as a baby").

Based on the identified semantic relationships, we expand a given keyword into a set of sub-keywords that have stereo-

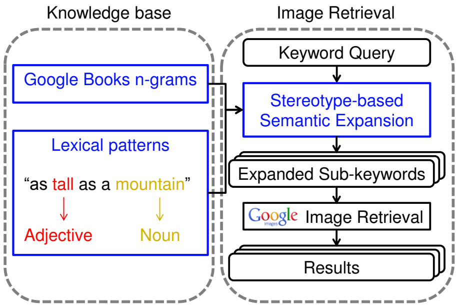


Fig. 2: This figure shows the overall structure of our stereotype-based semantic expansion. Components shown in the blue boxes are our novel contributions.

typical relationships. Especially, we expand an adjective keyword to a set of nouns whose stereotypical characteristic can be described by the given adjective keyword (e.g., expand "fast" into "{arrow, bullet, train, ...}"). In the similar manner, we also expand a noun keyword to a set of adjectives that stereotypically capture properties of the noun (e.g., "diamond" to "{transparent, expensive, shiny, ...}").

We have implemented our stereotype-based semantic expansion and applied it to image retrieval, to demonstrate its benefits. In image retrieval, compared to searching images given original keywords, we can identify more diverse sets of images that represent different characteristics of the original keywords, by expanding them into sub-keywords (Fig. 1). This in turn results in a higher recall ratio of what a user wants to find, according to a conducted user study. Our semantic expansion technique can be easily adopted in existing text-based image retrieval systems. Also, our method can be used together with listing frequently co-occurring terms derived from statistical approaches. By integrating our method with suggesting frequently co-occurring terms, we can explore unusual concepts and their images that are stereotypically related to the given keyword.

2. STEREOTYPE-BASED SEMANTIC EXPANSION

The overall structure of our stereotype-based semantic expansion is shown in Fig. 2. In this work we focus on keywords whose form is either nouns or adjectives. We first describe our expansion technique for adjective keywords, followed by noun ones.

For an adjective keyword, we aim to identify related nouns (or objects) that can be stereotypically described by the given adjective. For example, given an adjective "fast", we attempt to identify "arrow", "leopard", "bullet", etc. To implement our stereotype-based semantic expansion for adjectives, we need to identify word pairs that have such stereotypical relationship. Inspired by the work of Veale and Hao [11] we identify stereotypical relationships by looking into similes such as "as X as Y" (e.g., "as fast as an arrow") or "about as X as Y" among 4- and 5-grams. By searching the lexical patterns of those similes from Google Books n-grams we can have stereotypical relationship between over thousands of nouns and adjectives.

Based on the computed relationships between adjectives and nouns, we perform our stereotype-based semantic expansion. One keyword can be expanded to many (e.g., 100) sub-keywords in practice. Since some of them are not strongly related to the given keyword, we cull sub-keywords whose frequency is low in n-grams. Interestingly, we have found that term-frequency distributions of most sub-keywords approximately follow the well-known power law (Fig. 3), except for keywords that are expanded to a few sub-keywords (e.g., "spicy" to "{salsa, tamale}"). Given the power law, we accept sub-keywords in an order of decreasing frequency until the accumulated frequency of accepted sub-keywords is bigger than a certain percentage threshold, *an accepted percentage*, of the total frequency of all the sub-keywords. We then cull the rest of the sub-keywords. By doing so, we can report nouns that contain what a user wants in a likelihood close to the accepted percentage without having so many sub-keywords.

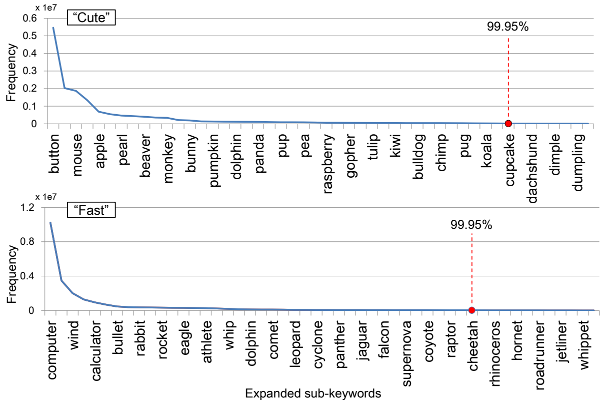


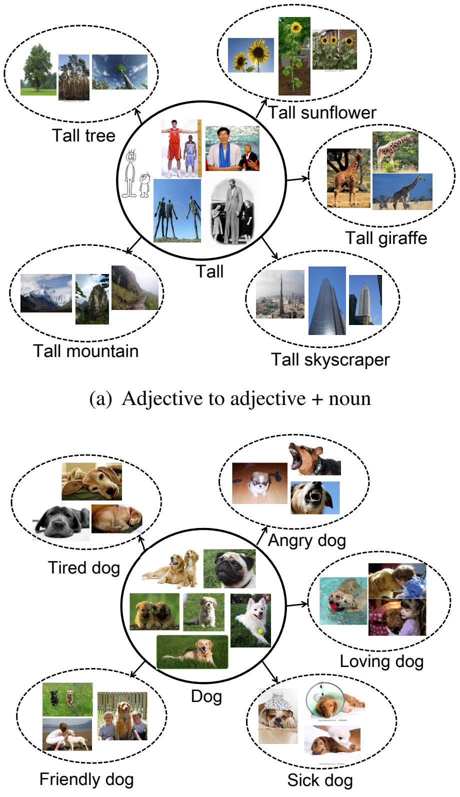
Fig. 3: Frequency distributions of keywords "cute" and "fast".

Each expanded sub-keyword for an adjective keyword represents a particular object represented by a noun. The noun sub-keyword itself is useful for identifying images related to the given adjective keyword, but searching images only with a noun may cause a topic drift. In order to efficiently avoid such drift, we place the original adjective in front of each noun sub-keywords. Given the example of the adjective keyword "fast", we replace it to "{fast computer, fast horse, fast wind, ...}". We then retrieve images with the expanded phrases. We found that using these expanded phrases is more useful than using only expanded nouns, since they can identify images that emphasize stereotypical features of the given adjective keyword.

In the same manner of applying our stereotype-based semantic expansion to adjective keywords, we can expand noun keywords into a list of adjective keywords. An example of such expansion is to replace "diamond" to "{expensive, transparent, shiny, ...}". Expanding nouns to a list of adjective keywords is, however, not directly useful for image retrieval. We therefore place each expanded adjective in front of the given noun. Given the example of the adjective keyword "diamond", we replace it with "{expensive diamond, transparent diamond, shiny diamond, ...}", and then retrieve images with the expanded phrases.

3. RESULTS AND DISCUSSIONS

We have used Google Image Search (GIS) [2] as our image retrieval engine. To show benefits of our method we have tested



(a) Adjective to adjective + noun

(b) Noun to adjective + noun

Fig. 4: Representative images downloaded among top-10 images by Google Image Search with original keywords "tall" and "dog" and their expanded phrases by our method.

ten different keywords (seven adjectives[1] and three nouns[2]). These keywords are chosen, since their meaning is intuitive to any users, but it is challenging to specify a particular image or object, especially for adjective keywords. We have also tested only ten keywords, mainly because it is non-trivial to download many images from GIS [3] and we need to manually classify those downloaded images for comparisons as elaborated on later. We expand these keywords based on our stereotype-based semantic expansion. For the expansion, we set the accepted percentage to be 99.95%, to cover most of important concepts and cull sub-keywords locating in the long tail with small frequency. In this setting, the tested keywords are expanded to 30.7 sub-keywords on average. Top-20 sub-keywords expanded from the tested keywords are shown in Table 1 in the supplementary report.

For each keyword, we expand it to sub-keywords and download top-10 images from GIS for each expanded sub-keyword. We have also downloaded images with the original keyword as the same number of images to those of downloaded images with the expanded sub-keywords; the total number of downloaded images is thus approximately 6 K $(= 10 \cdot 30.7 \cdot 10 \cdot 2)$. For images downloaded with each sub-keyword, we classify them to be in the class of the sub-keyword. For all the downloaded images with the original keyword, we manually classify them. For this manual classification, we first check whether an image belongs to the class of each sub-keyword. If the image does not belong to any of classes of sub-keyword, we assign a new class. For example, given a keyword "tall", we get an image that belongs to "tall ship", which is not covered by expanded sub-keywords. Nonetheless, most of them are clearly assigned to classes of expanded sub-keywords. On average, 84.1% of images downloaded with original keywords are manually classified into one of their expanded sub-keywords. The rest of images are assigned to new classes. For each original keyword, we have to create 3.57 new classes, which are 12% of the number of expanded sub-keywords on average. Detailed information about each keyword query is shown in Fig. 5.

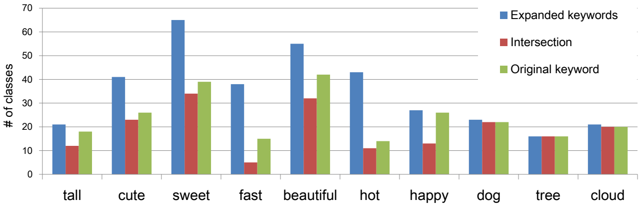


Fig. 5: This figure shows the number of different classes where downloaded images with original keywords and their expanded sub-keywords belong to. Intersection indicates classes reported both by searching an original keyword and by its expanded sub-keywords.

In order to study how much different classes of images are returned by original keywords and their expanded sub-keywords, we compute *diversity*, which measures how much portions of these different classes out of all the identified classes are covered by images retrieved by an original keyword and its expanded sub-keywords. The main reason why we measure the diversity is because the diversity value is likely to correlate with the probability that an user gets what he/she attempts to find by providing keywords. One may concern that even though some expanded sub-keywords are off the topic of the original keyword and thus actually noisy labels, they can increase the diversity measure. Nonetheless, we have identified that expanded sub-keywords of the tested keywords are weakly or strongly related to their original keywords; see Table 1 in the supplementary report. For example, "button" is the top expanded sub-keyword of the tested "cute" keyword. We initially thought that this expanded keyword is off the topic, but we realized that it is expanded thanks to the common English idiom of "as cute as a button". This example demonstrates a unique characteristic of our stereotype-based semantic expansion.

We have found that on average our stereotype-based semantic expansion reports 92.04% diversity, which is 16.93% higher than that (75.11%) of results with the original keyword. When adjective keywords are given, our method achieves the diversity of 84.08%, which is 30.83% higher than that (53.25%) achieved by using the original keywords without running our expansion method. For noun keywords, we have found that original keywords themselves are enough to identify diverse sets of images. Specifically, diversity values w/ and w/o our expansion method are 100% and 96.9% respectively. Overall our method shows significant improve-

[1] 'tall', 'cute', 'sweet', 'fast', 'beautiful', 'hot', and 'happy'

[2] 'dog', 'tree', and 'cloud'

[3] GIS blocks any access from a particular IP address when a computer with it attempts to download many images.

ment for effectively identifying diverse sets of images representing different aspects of adjective features. Some of image retrieval results with original keywords and their expanded sub-keywords are shown in Fig. 4.

To further verify the usefulness of our method, we have conducted a user study for measuring a level of user satisfaction of our method with the seven adjective queries. For this study we show top-10 images, called Group A, retrieved from each original keyword, and another set of 10 images, called Group B, retrieved from top-10 expanded keywords from each original keyword. We have chosen 12 applicants who has the computer science background and an experience on using image retrieval. We give each keyword to a participant and ask him or her to imagine an image or an object from the keyword. We then show Group A and B together and ask the participant whether each group has the image (or the object) that the participant imagined; all the asked images are shown in the supplementary report. We then measure the recall ratio of each group. Group B computed by our method reported 72.29% recall ratio of what a participant had in mind, while Group A reported only 59.52% recall ratio. The higher recall ratio of our method is mainly thanks to a diverse set of images related to the given keyword and clearly indicates the usefulness of our method.

3.1. Discussions

GIS [2] uses a keyword suggestion technique given a query keyword. There have been no public articles explaining on how to GIS identifies related words, but it has been conjectured that GIS maintains co-occurring terms among recently used search keywords and their output results (i.e. associated texts around the identified images), and reports more frequently related keywords given the query keyword. In this aspect, GIS is considered to utilize a drastically simplified form of n-grams. In addition there are previous text-based image retrieval techniques [9] that utilize co-occurring terms like GIS.

Our approach differs in which our method filters co-occurring terms based on lexical patterns and extracts stereotypical relationships among words. As a result, our approach can explore unusual concepts that cannot be captured by simply suggesting frequently co-occurring terms [4]. In addition, our method can be integrated with various methods that train classifiers (e.g., SVMs) based on visual features such as BoWs extracted from images collected by text queries for providing a more coherent set of images [12] or utilize image and text information jointly [13] to achieve semantically better results.

4. CONCLUSION

We have presented a novel, stereotype-based semantic expansion technique for identifying diverse sets of images that stereotypically represent different concepts of given keywords. Our key contribution is to harvest stereotypically related concepts from Google Books n-grams by looking into useful lexical patterns such as similes.

There are many interesting future research directions. In this paper we have showed potential benefits of our method in a small-scale experiment. We would like to optimize our method and test it in a large-scale configuration to further verify its benefits. We would like to extend our method more deeply into the visual domain by applying the same concept of our method to the visual domain considering the visual content of images [7]. This can be done by mapping stereotypical concepts to visual concepts. Finally, it will be interesting to see how concepts vary over a period of time and different cultures.

5. ACKNOWLEDGEMENTS.

We would like to thank Tony Veale and SGLab members for constructive discussions. This work was supported in part by MCST/KOCCA/CT/R&D 2011, DAPA/ADD (UD110006MD), MEST/NRF/WCU (R31-2010-000-30007-0), MSRA, and MEST/NRF (2012-0009228).

6. REFERENCES

[1] Ritendra Datta *et al.*, "Image retrieval: Ideas, influences, and trends of the new age," *ACM Computing Survey*, vol. 40, no. 2, pp. 1–60, 2008.

[2] Google, "Google image search," http://images.google.com.

[3] Gerard Salton and Chris Buckley, "Improving retrieval performance by relevance feedback," in *Readings in information retrieval*, pp. 355–364. 1997.

[4] Apostol Natsev *et al.*, "Semantic concept-based query expansion and re-ranking for multimedia retrieval," in *ACM Multimedia*, 2007.

[5] Fellbaum Christiane, "Wordnet," in *Theory and Applications of Ontology: Computer Applications*, pp. 231–243. 2010.

[6] Jia Deng, A.C. Berg, and Li Fei-Fei, "Hierarchical semantic indexing for large scale image retrieval," in *CVPR*, june 2011, pp. 785 –792.

[7] J.-P. Heo, Y. Lee, J. He, S.-F. Chang, and S.-E. Yoon, "Spherical Hashing," in *CVPR*, 2012.

[8] D. Nistér and H. Stewénius, "Scalable recognition with a vocabulary tree," in *CVPR*, 2006.

[9] Xiaogang Wang, Ke Liu, and Xiaoou Tang, "Query-specific visual semantic spaces for web image re-ranking," in *CVPR*, june 2011, pp. 857 –864.

[10] J.-B. Michel et al., "Quantitative analysis of culture using millions of digitized books," *Science*, 2010.

[11] Tony Veale and Yanfen Hao, "Making lexical ontologies functional and context-sensitive," in *ACL*, 2007.

[12] D. Grangier and S. Bengio, "A discriminative kernel-based approach to rank images from text queries," *IEEE TPAMI*, vol. 30, no. 8, pp. 1371 –1384, 2008.

[13] A.C. Berg *et al.*, "Understanding and predicting importance in images," in *CVPR*, 2012.

[4] Top 10 related words in GIS, "fast" = "{food, fashion, and furious, and furious 6, track, five, follower, slow, animal, and furious 5}". Top10 sub-keywords with our expansion, "fast" = "{computer, horse, wind, laser, calculator, arrow, bullet, microwave, rabbit, missile}".