Caranx: Scalable Social Image Index Using Phylogenetic Tree of Hashtags

Yusheng Xie[1] Zhuoyuan Chen[2] Ankit Agrawal[1] Wei-keng Liao[1] Alok Choudhary[1]
[1]: Northwestern University, Evanston, IL USA
[2]: Adobe Research, San Jose, CA USA
[1]: {yxi389,ankitag,wkliao,choudhar}@eecs.northwestern.edu [2]: zhuchen@adobe.com

ABSTRACT

Most existing image indexing techniques rely on Scale Invariant Feature Transformation (SIFT) for extracting local point features. Applied to individual image, SIFT extracts hundreds of numerical vectors. The vectors are quantized and stored in tree-like data structures for fast search. SIFT-based indexing can exhibit weakness under certain non-rigid transformations, which are common among real world applications. For example, SIFT often cannot recognize a face as the same with different expressions (e.g. giggling vs. crying). Non-Rigid Dense Correspondence (NRDC) addresses such drawbacks of SIFT. However, directly using NRDC incurs an impractical amount of computation in large-scale image indexing. We present a novel idea here that uses social hashtags to organize the images into a phylogenetic tree (PT). We provide an efficient algorithm to build/search the PT, and show that using PT structure can effectively avoid unnecessary NRDC computation. The resulting image index provides more accurate and diversified search results[1].

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous

Keywords

Image index, Tag, Instagram, Phylogenetic tree, SIFT, NRDC

1. INTRODUCTION

Recent growth of online social networks have literally changed the way people communicate. Indeed, websites like Facebook and Twitter have been recently extensively studied. However, it is often overlooked that social networks like Facebook have profoundly and almost tacitly changed the way people produce, perceive, and share images. Everyday 300 million photos are shared on Facebook alone[2]. Most of the socially *relevant* images are digitally tagged, commented, liked, retweeted, etc. In short, they are *socially interactive*. It is going to more difficult for an average user to come across any photo *without* any social context. Images are no longer arrays of pixels; instead, social context is now inseparable from the images. The presence of rich social information impels the CARANX[3] project.

Figure 1: The two highlighted Starbucks logos are partially recognized by NRDC as visual words.

2. RELATED WORK

Most existing image indexing techniques rely on SIFT [3], a technique that extracts local point feature vectors from individual images. For example, Google street view uses similar techniques for seamlessly stitching the landscapes from view to view. The leading *scalable* index approach based on SIFT is full feature representation (FFR). FFR fully stores feature vectors as compact numerical vectors. Tree structures can enable fast search over the vectors. When a query sends an image and its SIFT features, the search is done through vector similarity [1]. While FFR+SIFT has the potential to scale to even billions of images (100s TBytes $\sim$ 1 PBytes) [1], it suffers from following drawbacks. SIFT is not superb with non-rigid transformation or differences in lighting conditions, both of which are common among real world applications. For example, Alice uploads a set of photos of her New Zealand trip to Facebook. Photos taken in the morning/evening or photos of Alice making different facial expressions (e.g. calm face vs. giggling) would pose serious challenges to SIFT.

[1]This work is supported in part by the following grants: NSF awards CCF-0833131, CNS-0830927, IIS-0905205, CCF-0938000, CCF-1029166, and OCI-1144061; DOE awards DE-FG02-08ER25848, DE-SC0001283, DE-SC0005309, DESC0005340, and DESC0007456; AFOSR award FA9550-12-1-0458.

[2]http://visual.ly/impact-photos-facebook-engagement
[3]https://github.com/yvesx/caranx

SC'13 Companion, November 17-22, 2013, Denver, Colorado, USA.
ACM 978-1-4503-1030-7/11/11.

2.1 Non-Rigid Dense Correspondence

To circumvent the limitations of SIFT, alternative patch-matching algorithms are developed. Non-Rigid Dense Correspondence (NRDC) [2] is originally designed for pairs of images that contain similar (foreground) scenes captured by cameras of different lenses and point-of-shoot under non-rigid transformations. NRDC can address the two aforementioned drawbacks of SIFT. Figure 2 shows example match results by NRDC algorithm.

3. IMAGE INDEXING WITH NRDC

However, NRDC cannot be directly applied for indexing (social) photos at large scale because NRDC, unlike SIFT, is a pairwise type of operation. Suppose $T = \{t_1, t_2, \ldots, t_{|T|}\}$ is a database of 1 billion photos. NRDC will need to work on the quadratic set $\{\{t_i, t_j\} | t_i \neq t_j$, and $t_i, t_j \in T\}$. In other words, SIFT would need to run 10^9 times on T while NRDC would require to run $\binom{1B}{2} \approx 5 \times 10^{17}$ times. As a pairwise algorithm, NRDC cannot scale to index large databases of images. In this work, we exploit the tag information among a large number of social photos. We first organize the social tags into a PT structure. We propose a new probabilistic algorithm (Algorithm 1) that, taking advantage of the PT, drastically reduces the amount of NRDC pairwise computation for indexing images. Another probabilistic algorithm (Algorithm 2) is proposed to enable fast, accurate, and diversified query result retrieval.

Data Characteristics A large amount of images that we care about are photos of friends, family, events, etc. They are often tagged (multiple times) and shared through social media websites. CARANX utilizes the tag information in photos, which is key to scaling up image retrieval application. Table 1 summarizes the initial dataset we are working with.

Phylogenetic Tree of #tags A phylogenetic tree (PT) is a branching diagram showing the inferred evolutionary relationships among biological species based upon the pairwise similarity. CARANX proposes to use a PT to organize #tags, which, like biological species, evolve over time. Certain #tags like "#HappyNewYear2013" quickly gains and loses popularity. #tag is a well-studied phenomenon[4], we believe PT is a reasonable model for #tags. To produce a pairwise distance matrix, D, for the #tags, we measure the similarity between #tag t and #tag v by the number of sample photos that have both t and v; then take the inverse. Known algorithms like Unweighted Pair-Group Method with Arithmetic Means (UPGMA) can solve for the PT structure from D. Similar ideas are not mentioned in previous works like [1] or Caltech 101 dataset[4] most likely because a few years ago #tags were not invented for most social photos.

Dense Visual Words The data structure in CARANX is based on the Bag-of-Visual-Words (BVW) model. BVW is not as sophisticated as other hierarchical decomposition methods (HDM) [1], but we think it is the right choice here for two reasons. First, most HDMs (e.g., Kd-Tree) apply directly to a large amount of high-dimensional SIFT feature vectors because they can provide excellent locality properties for vector range queries. This advantage is not relevant for us because CARANX does not use any range query on the numerical vectors. Second, geometric consistency check, for which HDMs are good at, plays a big part in SIFT-based methods. CARANX does not perform post-query geometric consistency checks because NRDC-discovered visual words are often large, dense (10% ∼ 50% of image size), and already have confidence label. Figure 2 shows an example of a NRDC dense visual word.

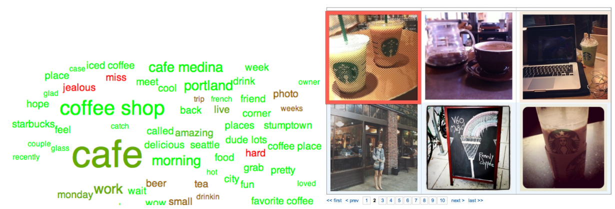
[4]http://www.vision.caltech.edu/Image_Datasets/Caltech101/

Algorithm 1: Overview of CARANX index

Build #tag distance matrix D from common images
Build phylogenetic tree PT from D by UPGMA algorithm
Extract pairwise NRDC visual words for images in each node (and its vicinity) of PT
Register matched visual words with #tags and images

Algorithm 2: Overview of CARANX search

Identify likely branch $PT[b]$ of PT to search
Perform NRDC match between query image and stored visual words in $PT[b]$
Return images associated with the matched words



(a) (b)

Figure 2: (a) #tags closely located to "#Starbucks" in the phylogenetic tree. (b) Sample retrieval results for a Starbuck query image.

Table 1: Caranx experiment data as of 2013 June.

Source	Instagram public data
Images	54.8M
#tags (unique)	∼200M (∼5M)
Unique users	15.4M

4. PRELIMINARY RESULTS

Figure 2(a) shows the #tags that are most closely related to "#Starbucks"; bigger fonts mean shorter distance in terms of phylogenetic tree position. Figure 2(b) shows examples of retrieval results when using CARANX (image in red is the query). In future work, we will focus on performance assessment and correctness verification.

5. REFERENCES

[1] M. Aly. Searching Large-Scale Image Collections. Master's thesis, Caltech, Pasadena, CA, USA, June 2011.

[2] Y. HaCohen, E. Shechtman, D. B. Goldman, and D. Lischinski. Non-rigid dense correspondence with applications for image enhancement. *SIGGRAPH*, 2011.

[3] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60:91–110, 2004.

[4] L. Yang, T. Sun, M. Zhang, and Q. Mei. We know what @you #tag: does the dual role affect hashtag adoption? WWW '12, pages 261–270. ACM, 2012.