**iAVATAR: An Interactive Tool for Finding and Visualizing Visual-Representative Tags in Image Search**

Aixin Sun  
Sourav S Bhowmick  
Yao Liu  
School of Computer Engineering, Nanyang Technological University, Singapore  
axsun,assourav@ntu.edu.sg

**ABSTRACT**

Tags associated with social images are valuable information source for superior image search and retrieval experiences. Due to the nature of tagging, many tags associated with images are not visually descriptive. Consequently, presence of these noisy tags may reduce the effectiveness of tags’ role in image retrieval. To address this problem, we demonstrate iAVATAR (interActive VisuAl-representative TAgs Relationship) system that uses the notion of Normalized Image Tag Clarity (NITC) to find visual-representative tags. A visual-representative tag effectively describes the visual content of the images. Further, we visually demonstrate relationships between popular tags and visual-representative tags as well as co-occurrence likelihood of a pair of tags associated with a search tag or image using tag relationship graph (TRG). We demonstrate various innovative features of iAVATAR with a real-world dataset and show that it enriches users’ understanding of various important tag features during image search.

1. INTRODUCTION

With the advances in digital photography and social media sharing web services, a huge number of multimedia content is now available online. Most of these services enable users to annotate images with free tags (e.g., aircraft, lake, sky). A key consequence of the availability of such tags as meta-data is that it has significantly facilitated web image search and organization as this rich collection of tags provides more information than we can possibly extract from content-based algorithms. However, it has been widely recognized that realizing a tag-based image retrieval system is technically challenging due to noisy and imprecise nature of tags [2]. Two similar images may be associated with significantly different sets of tags from different users. Further, tags associated with an image may describe the image from significantly different perspectives. For example, consider a photo uploaded by Sally which she took using her Canon camera at Sentosa when she traveled to Singapore in 2009. This image may be annotated by tags such as Canon, 2009, Singapore, beach, sentosa, and many others. Notice that some of the tags (e.g., 2009 and Canon) do not effectively describe the visual content of the image. Consequently, presence of these noisy tags may reduce the effectiveness of tags’ role in image retrieval. Needless to say that “de-noising” tags has been recently identified as one of the key research challenges in [2].

In this demonstration, we present a novel graphical noisy tag-aware social images retrieval system, called iAVATAR (interActive VisuAl-representative TAgs Relationship), that takes a concrete step to address the above challenge. Given a search tag $t$ as input, iAVATAR retrieves a ranked list of images, denoted by $T$, that is annotated with $t$ in the image database. A key feature of this system is that for $t$ (resp. for each image $d \in T$) it identifies a set of visual-representative tags [8] related to $t$ (resp. $d$) and how these tags are associated with other related tags using a color-coded tag relationship graph (TRG). Each tag is a labeled colored node in the TRG where the font size of the label and the color intensity of the node are proportional to the tag frequency and visual-representativeness, respectively. A pair of nodes is connected by a labeled edge if the corresponding tag pair co-occur together among images in the dataset beyond certain threshold.

Intuitively, a tag is visual-representative if it effectively describes the visual content of the images. A visual-representative tag (e.g., sky, sunset) easily suggests the scene an image may describe even before the image is presented to a user. On the other hand, tags like 2009 and Asia often fail to suggest anything meaningful with respect to the visual content of the annotated image. Clearly, identification of visual-representative tags from all tags assigned to images enables end-users to eliminate noisy tags. Further, when a user selects a visual-representative tag $t_v$ in the TRG, iAVATAR retrieves a fraction of images associated with $t_v$ to validate that it indeed describes the visual content of the images.

Additionally, iAVATAR supports two interactive graphical features: the filtering mechanism and the difference viewer. The filtering mechanism enables users to filter or expand the TRG to view different sets of tags associated with $t$ (resp. $d$) and their relationships based on different threshold values for visual-representativeness, tag frequency, and hop distance. The difference viewer provides a graphical view of the effects of different types of tag co-occurrence measures (e.g., cosine, Jaccard coefficient, KL divergence) on the tag relationships in the TRG. Such interactive features of iAVATAR pave way to superior image retrieval experience as they not only enrich users with the knowledge of noisy or non-noisy tags but also provides an in-depth understanding of the relationships between tags associated with the retrieved images.

2. RELATED SYSTEMS AND NOVELTY

To the best of our knowledge, this is the first work that uses visual-representativeness as a new dimension to better understand tag properties. The proposed finding of visually-representative tags...
exploits the techniques in the area of query performance prediction in Web search [3]. A query is unambiguous if all its matched documents are topically cohesive; analogously, a tag is visually-representative if its associated images are visually cohesive.

Tag cloud is the most widely adopted tag visualization technique. In a tag cloud, the tags are often ordered in alphabetical order and the font sizes are proportional to their frequencies. Other than frequency, tags have also been visualized by their spatial/temporal aspects. Map-based tag cloud visualizes tags on top of a map interface with geo-referenced social images [1]. The temporal evolution of tags within the Flickr is visualized in [4] where the font size of the tag is proportional to its interestingness derived from the frequency evolution of the tag along the timeline. To visualize the relationships between tags, [6] displays tag clouds on top of a topographical image so that related tags are closer to each other. However, none of the existing approaches explore relationships between tags by visual-representativeness, frequency, and co-occurrence, coherently.

3. SYSTEM OVERVIEW

The iAVATAR system is implemented in Java using open-source libraries TouchGraph, Lucene and JGraphT. Figure 1 shows the system architecture of iAVATAR and mainly consists of the following modules.

The iAVATAR GUI Module: Figure 2 depicts the screenshot of the visual interface of iAVATAR. It consists of five main panels. A user may formulate a search query in several ways. He may enter a search tag in Panel 1 and specify the number of images he may wish to view in the Images field in Panel 1, such as "sun", "sky", or "clouds" among others. If any of the preview tags is clicked, then the clicked tag becomes the new search tag. Note that we can also view the preview of a specific visual-representative tag in the TRG by clicking on the corresponding node.

Lastly, the Filtering Panel (Panel 5) enables the user to filter the nodes in the TRG in real time by modifying the thresholds of visual-representativeness and tag frequency. It also allows a user to view related tags that are one or two hops away from the search tag.

The Image Retriever Module: Given a search tag, this module retrieves a ranked list of images from the image repository that are annotated (or related) with the given tag. A user may choose a ranking method by selecting from the drop down menu associated with the Image Search field in Panel 1, such as TFIDF-based and tag expansion-based ranking methods. The default approach is random ranking where as long as an image is annotated by the searched tag, it has equal probability of being displayed in Panel 2.

The Tag Features Extractor Module: This module is the core engine of iAVATAR and consists of the following submodules.

The Tag Frequency Retriever Module: This module is used to compute the frequencies of tags in the image repository and store them in the tag features database (Tag DB). Tag frequency is the number of images a tag \( t \) is associated with in the given dataset.

The Visual-Representative Tags Finder Module: This module finds visual-representative tags from the image dataset and store them in Tag DB. Intuitively, a tag is visually-representative if all the images annotated with the tag are visually similar to each other. We use the notion of normalized tag clarity score (NTIC) to measure the visual-representativeness of a tag. We briefly describe NTIC here. The reader may refer to [8] for details.

We consider a tag to be a keyword query and the set of images annotated with the tag are the retrieved documents based on a boolean
using the Jelinek-Mercer smoothing with Equation 4. The estimated tag language model is further smoothed.

In Equation 3, we use “image” and “document” interchangeably and use d to denote an image. We now define the notion of image tag clarity. Let D be the set of images and T ⊆ D be the set of images annotated by a tag t. Let w be an arbitrary visual word in the vocabulary. The image tag clarity score of t, denoted by ITC(t), is defined as the Kullback–Leibler (KL) divergence between the tag language model (P(w|T)) and the collection language model (P(w|D)). It is expressed by the following equation.

\[
ITC(t) = KL(T||D) = \sum_w P(w|T) \log_2 \frac{P(w|T)}{P(w|D)}
\]

The collection language model P(w|D) is estimated from the relative visual word frequency in D. The tag language model P(w|T) is estimated using Equation 2, where P(d|T) reflects the relative closeness of the image d to T’s centroid defined in Equation 3.

\[
P(w|T) = \sum_{d \in T} P_{ml}(w|d)P(d|T)
\]

\[
P(d|T) = \frac{\varphi(d, T)}{\sum_{d \in T} \varphi(d, T)}
\]

\[
\varphi(d, T) = \prod_{w \in d} P_s(w|T)P_{ml}(w|d)
\]

In Equation 3, \(\varphi(d, T)\) is a centrality function which defines the similarity between an image d to T, adopted from [5]. Let \(P_{ml}(w|d)\) be the relative word frequency of w in image d. Let \(P_s(w|T)\) be the tag language model estimated from the expected word frequency in the tagged images with equal importance \(1/T\), i.e., \(P_s(w|T) = \sum_{d \in T} P_{ml}(w|d)\). Then \(\varphi(d, T)\) is defined to be the weighted geometric mean of word generation probabilities in T shown in Equation 4. The estimated tag language model is further smoothed using the Jelinek-Mercer smoothing with \(\lambda = 0.99\).

\[
P_{\text{smoothed}}(w|T) = \lambda P(w|T) + (1 - \lambda) P(w|D)
\]

In tagging, the tag distribution follows a power-law distribution with a small set of tags much more frequently used than other tags [8]. To overcome the impact of tag frequency, we applied zero-mean normalization to the image tag clarity scores. The expected image tag clarity score with respect to t is computed by randomly assigned dummy tags with the same frequency to images in the dataset. Let \(f(t)\) be the frequency of a tag t in the image dataset. Let \(\mu(f(t))\) and \(\sigma(f(t))\) be the expected tag clarity and standard deviation obtained by assigning multiple dummy tags having the same frequency \(f(t)\). Then, the normalized image tag clarity score, denoted by NITC(t), is given by Equation 6. A tag t is considered visual-representative if NITC(t) ≥ 3 (i.e., the tag clarity is 3 standard deviations away from the expected tag clarity of randomly assigned tags of the same frequency).

\[
\text{NITC}(t) = \frac{\text{ITC}(t) - \mu(f(t))}{\sigma(f(t))}
\]

The proposed tag language model can be estimated in \(O(N)\) time for a tag associated with N images and requires at most three scans of the images (for computing Equations 2, 3, and 4). Note that the expected tag clarity scores need to be computed only once for a given dataset. As NITC(t) values are in the range of \((-\infty, +\infty)\), the values are further normalized into [0,1] using a sigmoid function in iAVATAR. The evaluation of tag visual representativeness will be reported in a separate study.

**The TRG Constructor Module:** Given the visual-representative tags and tag frequencies, the objective of this module is to construct the tag relationship graph (TRG) of the search tag or image. It consists of two submodules as follows.

**The Node Constructor Module:** The nodes of a TRG of a given search tag or image is constructed by this module. Each node is labeled with a tag t and the font size of the label is proportional to the frequency of t. That is, the larger the font the more frequent is the tag, similar to most tag clouds. A node is color-coded based on its visual-representativeness (except for the search tag whose node is highlighted in orange color). The more visually representative (higher NITC value) a tag is the darker is the color (the spectrum of color codes used in iAVATAR is shown in Panel 5) of its node. For example, Figures 2 and 3 depict two examples of TRGs for the search tags sunset and flickr, respectively. Notice that
the TRG of flickr has many lighter colored nodes as most the associated tags are not visually representative. In contrast, the TRG of sunset has many violet colored nodes as the associated tags have high NITC values.

The Edge Constructor Module. This module constructs edges between the nodes in TRG. As the label of an edge represents the tag co-occurrence score of the connected tag pair, it first computes the tag co-occurrence of a pair of tags \( (t_a, t_b) \) related to the search tag (or image). Let \( f(t_a \land t_b) \) be the number of images tagged by both \( t_a \) and \( t_b \). Let \( N \) be the number of images in the given dataset. Then, the tag co-occurrence values are computed using one of the measures listed in Table 1, where \( f(t_a) \) denotes \( t_a \)'s frequency. A user can choose one of these measures using the drop down list associated with the Relation field in Panel 1. Observe that it is possible for \( t_a \) to co-occur with a large number of \( t_b \)s. Hence, iAvatar lets users control the number of pairs of \( (t_a, t_b) \) to view by specifying the fanout in Panel 1. For instance, in Figure 2 the fanout is specified as 10. Given a fanout \( k \), at most top-\( k \) \( (t_a, t_b) \) pairs are selected based on the tag co-occurrence scores and labeled edges are added between these pairs in the TRG. The color of an edge is determined by the average NITC value of the tag pair.

Figure 4 depicts the TRGs of the sunset search tag for three tag co-occurrence measures. Observe that the TRG's structure can significantly change with the choice of tag co-occurrence metric. It is easy to see that this feature of iAvatar enables users to visualize the effect of a specific tag co-occurrence measure on the TRG.

The Tag Filter Module. Finally, the objective of this module is to provide users flexibility to filter or expand the TRG based on different features of the tags. Currently, it supports tag frequency, visual-representativeness, and HOP distance-based filters (Panel 5). A user can modify the threshold of visual-representativeness (resp. tag frequency) by dragging the Visual Representativeness (resp. Tag Frequency) slider in Panel 5. Only tags whose visual-representativeness (resp. tag frequency) are greater than the threshold are displayed in the TRG. The HOP filter controls expansion of the TRG by determining whether tags that are related to the search tag or image indirectly by two hops shall be displayed. If they are displayed, then the nodes on the second hop are highlighted with orange-colored border. A node is shown as a second-hop node if it is related to at least two first hop nodes (to avoid the situation of too many nodes in Panel 3).

5. REFERENCES