Searching and Browsing Social Images through *i*AVATAR

Towards Yahoo! Challenge 2010: Novel Image Understanding

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

ABSTRACT

We present a novel graphical social images retrieval system, called *i*AVATAR (**i**nter**A**ctive **V**isu**A**l-representative **TA**gs **R**elationship), that takes a concrete step to address social image search and understanding challenge. *i*AVATAR is the first systematic effort to explore relationships between tags by visual-representativeness, frequency, and co-occurrence, coherently. Through the visualization of tag properties and relationships, *i*AVATAR offers users a visual explanation of the retrieved social images.

1. OVERVIEW

Availability of social tags has significantly facilitated web image search and organization as this rich collection of tags provides more information than we can possibly extract from the content-based algorithms. However, it has been widely recognized that realizing a tag-based image retrieval system is technically challenging as tags associated with an image may describe the image from significantly different perspectives, and are often noisy and imprecise. In this paper, we present the *i*AVATAR system[1] that utilizes tag visual-representativeness [1] to visually explain tagged image search results and guides tag-based image searching and browsing.

Intuitively, a tag is *visual-representative* if it effectively describes the *visual content* of its annotated images. Such tag (e.g., `sky`, `tiger`) easily suggests the scene or object an image describes even before the image is presented to a user. To quantify the visual-representativeness of a tag, we use two distance metrics, namely *cohesion* and *separation*, by measuring (a) how well the set of tagged images presents similar visual content among them, and (b) how distinct the common visual content is with respect to the entire image collection. These two metrics are generic and can be plugged into different distance functions and different image feature representations for computing tag visual-representativeness. Importantly, these measures are parameter free with linear or constant computational complexity. Details on quantifying tag visual-representativeness and evaluation are reported in [1].

Clearly, visual-representativeness of a tag gives users better understanding on the visual coherence of images annotated by the tag and hence the expected image search results. In the following, we introduce the visual interface of *i*AVATAR with an example tag search `cute` and illustrate how *i*AVATAR facilitates superior image retrieval experience with visual explanation of the retrieved social images. Figure 1 depicts the screenshot of *i*AVATAR. Given a search tag t as input (Panel 1 in Figure 1, and we use tag `cute`), it retrieves a ranked list of images, denoted by I_t, that is annotated with t in the image database (Panel 2). A key feature of this system is that for t (resp. for each image $d \in I_t$) it identifies a set of tags 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. Specifically, violet colored nodes are for more visually representative tags and green colored nodes are for less visually representative tags. The query tag (i.e., `cute`) is highlighted in orange. 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 (Panel 3). In our example, the tag `cute` has high co-occurrences with two sets of visually representative tags {`animal` , `cat` , `pet` , `pets` } and {`baby` }, which visually explains the large appearances of animals in the search results in Panel 2. The TRG also shows the relationships between the visually representative tags and less visually representative tags (e.g., `funny` and `pets` in Figure 1). Panel 4 displays a preview of some of the highly visual-representative tags in Panel 3; four images randomly picked for each tag for preview.

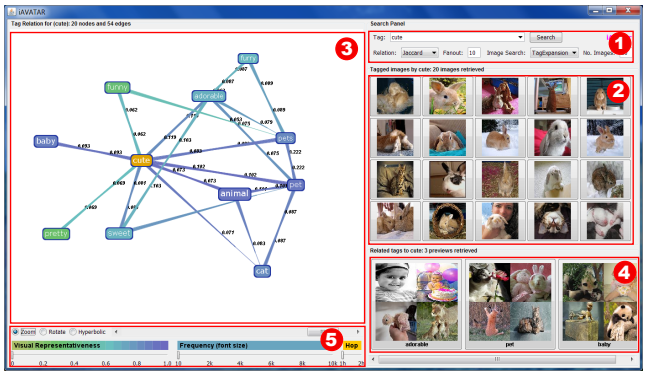

Figure 1: Visual interface of *i*AVATAR with tag search `cute` .

Additionally, *i*AVATAR 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* (Panel 5). 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 *i*AVATAR pave way to superior image retrieval experience as they provide an in-depth understanding of the relationships between tags and the retrieved images.

[1] The system is implemented in Java using open-source libraries TouchGraph, Lucene and JGraphT. Our demonstration will be loaded with 269,648 Flickr images provided by NUS-WIDE dataset: `http://lms.comp.nus.edu.sg/research/NUS-WIDE.htm` .

2. REFERENCES

[1] A. Sun, S. S. Bhowmick. Quantifying Tag Representativeness of Visual Content of Social Images. *In ACM MM*, 2010.