Interactive Image Search with Attribute-based Guidance and Personalization

Adriana Kovashka Kristen Grauman
The University of Texas at Austin
{adriana, grauman}@cs.utexas.edu

Abstract

Interactive image search, where a user initiates a query with keywords and refines it via feedback, can be enhanced using attributes. To minimize the user's effort, we propose to let the system guide a user's attribute-based comparative feedback with "pivot" exemplars that most reduce the system's uncertainty. Since humans vary in how they perceive the link between a named property and image content and might disagree on an attribute's presence, we further show how to efficiently learn user-specific attribute models. We demonstrate that attribute adaptation and system-driven feedback allow the user to quickly find his desired target.

1. Introduction

Visual attributes have proven useful as a middle ground for communication between users and retrieval systems during image search. Attributes are high-level semantic properties of objects, and have been used for multi-attribute keyword search in [6]. In [5], we show how *comparative feedback* based on attributes can refine search results more efficiently than traditional relevance feedback [9]. After typing a query for "black high heels", the user can refine the search results by making feedback statements on relative attributes, e.g., "I want shoes like these, but *more pointy*."

However, there are two key questions that remain to be addressed in order to make this form of feedback as useful as possible. First, what are the *most informative* images on which the user should give feedback? Further, how can we ensure the system's and user's models of attributes *align*? We study these questions in detail in our recent work at ICCV 2013 [4, 3], and here briefly overview our findings.

2. Guiding Feedback with Attribute Pivots

The user may often not know what feedback would be most useful for the system. Further, the images estimated to be most relevant and shown at the top of the results page in a given iteration may not be most informative as anchors for feedback. Thus, we propose a new form of interaction where the system engages the user in a *relative* 20-questions-like game, and the answers the user provides are visual comparisons. Unlike other active selection methods, ours only evaluates a small number of candidate questions by exploiting the ordinal structure of relative attributes.

Figure 1. Our image search approach actively requests feedback on selected images in terms of visual attribute comparisons.

A user initiates a search (e.g., with keywords or a sample image), and our approach then refines the results. It interacts with the user through questions of the form: "Is the image you are looking for *more*, *less*, (or *equally*) A than image I?", where A is a semantic attribute and I is an exemplar from the database being searched (see Figure 1). Our goal is to generate the series of such questions that will most efficiently narrow down the relevant images in the database, so that the user finds his target.

For each attribute $1, \ldots, M$, we learn a relative attribute [8] ranking function $a_m(I_i)$ that maps the descriptor for image I_i to a real-valued attribute strength, using training data collected on Amazon Mechanical Turk. We then construct a binary search tree for each attribute m. The tree recursively partitions all database images into two balanced sets, where the key at a given node is the median relative attribute value occurring within the set of images passed to that node. The "pivot" image I_{p_m} for attribute m starts out as the root of m's tree, and is updated in accordance with user feedback for this attribute.

The output of our system is a sorting of all database images according to their predicted relevance. To compute relevance scores, we evaluate how well each image satisfies the total set of feedback statements given by the user. We use information-theoretic entropy to select one of the pivots $\{I_{p_1}, \ldots, I_{p_M}\}$ as the anchor for feedback at the next iteration. Specifically, we choose the pivot $I^*_{p_m}$ which minimizes the expected entropy of the system after adding feedback on that pivot (weighed by the likelihood of each possible user response). See [4] for more details.

We evaluate our approach on three datasets: Shoes [1, 5], Scenes [7], and Faces [6], using up to 14K images and 11 attributes. After five iterations, our method achieves a target percentile rank (which represents search accuracy) of 96.8%. In contrast, the status quo approach of seeking feed-

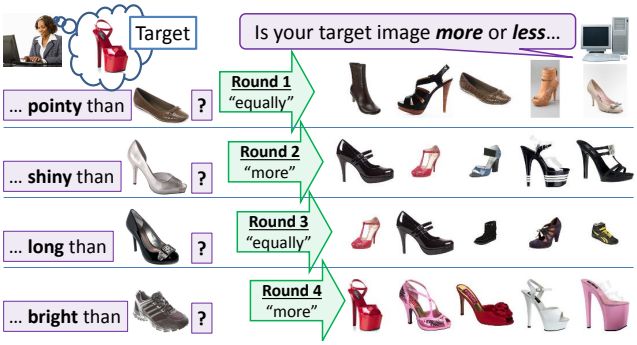

Figure 2. Using the user's sequential feedback on the left, we retrieve the images on the right at the top of the results list.

back on the top-ranked images achieves 89.8%, and an exhaustive active approach which evaluates all possible pairs of images and attributes achieves 88.7%. The latter requires up to 10 min to select its next question, while our method takes < 1 sec. Thus, in addition to regularizing the selection and hence boosting search accuracy, the attribute trees enable our method to run in real time. In Figure 2, we show an example of a search performed by a user on MTurk, where the user found her target among 14K images in just four iterations. Our method saves up to 11 iterations and 70 seconds of user time per query.

3. Personalizing Search with Adaptation

We showed how to let the system request the most useful feedback, but any information that the user provides which is misinterpreted by the system will negatively impact search accuracy. This might happen if the user's mental models of the attribute terms involved in a search are different than the system's models. Attributes can often be subjective due to the imprecision of the attribute vocabulary and to differing user perceptions of visual content (see Figure 3 for examples). Yet existing approaches [2, 6, 1, 8] learn monolithic attribute functions under the assumption that one model per attribute is sufficient for all users.

We next propose a transfer learning approach to learn personalized models of attributes, which ensure that the system understands a user's queries and feedback as intended. We first learn a *generic* model of an attribute by distributing the labeling effort across many annotators (the "crowd") and training on the resulting majority-voted data from multiple annotators. Then, for a given user, we adapt the parameters of this generic model with minimal supervision from this user, such that the new model accounts for any user-specific labeled data. We use a large-margin formulation and a regularizer preferring user-specific parameters that do not deviate greatly from the generic parameters. We employ adaptive formulations of SVM and ranking SVM for binary and relative attributes, respectively.

We collect user-specific labels on a set of diverse images

Figure 3. Visual attribute interpretations vary slightly from viewer to viewer. For example, 5 viewers *confidently* declare a shoe as formal (left) or more ornamented (right), while 5 others confidently declare the opposite! We propose to adapt attribute models to take these differences in perception into account.

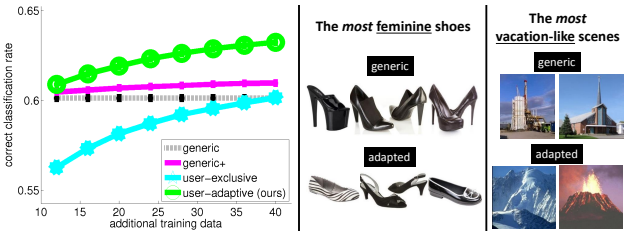

Figure 4. (a) Accuracy of predicting perceived attributes. (b, c) Images with high strength of "feminine" and "vacation-like".

or ones with labels that are frequently disagreed upon. For relative attributes, we also develop two techniques to automatically infer user-specific labels from the user's search history. We test on 10 shoe and 12 scene attributes. In Figure 4 (a), we show that between 12 and 40 user-specific labels allow our adaptation approach (in green) to achieve much higher accuracy on a held-out test set from each user, compared to using either the generic data or a single user's data exclusively. In Figure 4 (b) and (c), we show that the adapted "feminine" attribute captures a user's perception that *flatter* dark shoes are more feminine than high-heeled dark shoes. For another user, mountain scenes are more "vacation-like" than those with architectural features.

Personalization of attribute models affects not just prediction accuracy but also the success of search. We experimentally verify that adapting the models for attributes that are used during search improves search accuracy. See [3] for details.

References

[1] T. L. Berg, A. C. Berg, and J. Shih. Automatic Attribute Discovery and Characterization from Noisy Web Data. In *ECCV*, 2010.
[2] A. Farhadi, I. Endres, D. Hoiem, and D. Forsyth. Describing Objects by Their Attributes. In *CVPR*, 2009.
[3] A. Kovashka and K. Grauman. Attribute Adaptation for Personalized Image Search. In *ICCV*, 2013.
[4] A. Kovashka and K. Grauman. Attribute Pivots for Guiding Relevance Feedback in Image Search. In *ICCV*, 2013.
[5] A. Kovashka, D. Parikh, and K. Grauman. WhittleSearch: Image Search with Relative Attribute Feedback. In *CVPR*, 2012.
[6] N. Kumar, P. Belhumeur, and S. Nayar. Facetracer: A Search Engine for Large Collections of Images with Faces. In *ECCV*, 2008.
[7] A. Oliva and A. Torralba. Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope. *IJCV*, 42:145–175, 2001.
[8] D. Parikh and K. Grauman. Relative Attributes. In *ICCV*, 2011.
[9] X. Zhou and T. Huang. Relevance Feedback in Image Retrieval: A Comprehensive Review. *Multimedia Systems*, 2003.