WhittleSearch: Image Search with Relative Attribute Feedback

Adriana Kovashka\(^1\)  
Devi Parikh\(^2\)  
Kristen Grauman\(^1\)  
\(^1\)The University of Texas at Austin  
\(^2\)Toyota Technological Institute Chicago (TTIC)

**Problem**
- Existing relevance feedback methods too restrictive
- User cannot specify what is relevant or irrelevant about marked images

**Our Idea**
- Allow user to "whittle away" irrelevant regions of the image space through precise statements relating their envisioned target and exemplar images.

**Learning to Predict Relative Attributes**
1) Obtain ordered image pairs \(O_m = \{(i,j)\}\) and unordered pairs \(E_m = \{(i,j)\}\)
2) For each attribute \(m\), learn a ranking function
   \[ r_m(x) = \text{scores}_m(x) \text{ such that:} \]
   \[ \forall (i,j) \in O_m: \text{scores}_m(x_i) > \text{scores}_m(x_j) \]
   \[ \forall (i,j) \in E_m: \text{scores}_m(x_i) = \text{scores}_m(x_j) \]
3) Rank images based on how many constraints they satisfy:
   \[ \text{score} = \sum \text{constraint satisfaction} \]

**Relative Attribute Feedback**
1) User selects some images and marks how they differ from the desired image, thus defining constraints: "I want [objects] that are [more/less] [attribute name] than the [object] in this image."
2) Update the scores for each image in the database, using these:
   \[ \text{scores} = \text{scores} + 1 \] if more natural than this
   \[ \text{scores} = \text{scores} + 0 \] if less natural than this
3) Rank images based on how many constraints they satisfy:
   \[ \text{score} = \sum \text{constraint satisfaction} \]

**Hybrid Feedback**
Integrate relative attribute and binary feedback by learning relevance ranking function.

**Constraint Generation**
- Is similar to or dissimilar from
- Is more or less than?
- Binary feedback baseline
- Relative attribute feedback

**Results**
- **Datasets**
  - Shoes: 14,658 images from Attribute Discovery; augmented with 10 attributes; features: GIST+color
  - CSR: 2,688 images from Outdoor Scene Recognition; 6 attributes; features: GIST
  - PubFig: 772 images from Public Figures; 11 attributes; features: GIST+color

- **Evaluation Metrics**
  - \( \text{NDCG@50} \) correlation between method’s ranking and ground truth ranking (high = good)
  - Ground truth ranking: images ranked by their distance to the secret image in learned feature space

- **Impact of iterative feedback**
  - More open than
  - Less ornaments than
  - Less high at heel than

- **Impact of amount of feedback**
  - More bright than with fewer constraints.

See our paper for additional results with human-given constraints, keyword search, impact of relative images, and impact of supervision type (comparing categories vs. instances).

**Conclusions**
- Proposed method allows user to communicate precisely how results compare with mental model
- Refines image search results more effectively, often with less human effort