iLike: Integrating Visual and Textual Features for Vertical Search

Yuxin Chen¹, Nenghai Yu², Bo Luo¹, Xue-wen Chen¹

¹ Department of Electrical Engineering and Computer Science
The University of Kansas, Lawrence, KS, USA

² Department of Electrical Engineering and Information Sciences
University of Science and Technology of China, Hefei, China
Motivation

- The problem
  - Huge amount of multimedia information available
  - Browsing and searching is even harder than text
- Text-based image search
Motivation

• Text-based image search
  • Adopted by most image search engines
    – Efficient – text-based index
    – Text similarity, PageRank
  • Some queries work very well
    – Clearly labeled images
    – Distinct keywords
  • Some queries don’t
    – Insufficient tags
    – Gap between tag terms and query terms
    – Descriptive queries: “paintings of people wearing capes”
Motivation

• Content-based Image Retrieval (CBIR)
  • Visual features: color, texture, shape...
  • Semantic gap
    – Low level visual features vs. image content
    – sun -> nice sunshine -> a beautiful day
• Excessive computation: high dimensional indexing?
Motivation

• Put textual and visual features together?
• In the literature: hybrid approaches
  • Text-based search: candidates
  • CBIR-based re-ranking or clustering
• Our idea
  • Connect textual features (keywords) with visual features
  • Represent keywords in the visual feature space
    – Learn users’ visual perception for keywords
Preliminaries

• Data set
  • Vertical search: online shopping for apparels and accessories
  • Text contents are better organized
  • We can associate keywords and images with higher confidence
  • In this domain, text description and images are both important

• Data collection
  • Focused crawling: 20K items from six online retailers
    – Mid-sized hi-quality image with text description
  • Feature extraction
    – 263 low-level visual features: color, texture and shape
    – Normalization
Representing keywords

- **Keywords**
  - Image -> Human perception -> text description
  - Perception is subjective, the same impression could be described through different words
  - Calculating text similarity (or distance) is difficult - distance measurements (such as cosine distance in TF/IDF space) do NOT perfectly represent the distances in human perception.
Representing keywords

• *Items share the same keyword(s) may also share some consistency in selected visual features.*

• *If the consistency is observed over a significant number of items described by the same keyword, such a set of features and their values may represent the human “visual” perception of the keyword.*
Representing keywords

- Example: checked
Representing keywords

• Example: floral
Representing keywords

• For each term, we have
  • Positive set: items described by the term
  • Negative set: items not described by the term

• “Good” features
  • are coherent with the human perception of the keyword
  • have consistent values in the positive set
  • show different distributions in the positive and negative sets

• How do we identify “good” features for each keyword?
  • Compare the distributions in the positive and negative sets...
Representing keywords

- Distribution of visual features (term="floral")
Kolmogorov-Smirnov test

- Two sample K-S test
  - Identify if two data sets are from same distribution
  - Makes no assumptions on the distribution
  - Null hypothesis: two samples are drawn from same distribution
  - P-value: measure the confidence of the comparison results on the null hypothesis.
    - Higher p-value -> accept the null hypothesis -> insignificant difference in the positive and negative sets -> “bad” feature
    - Lower p-value -> reject the null hypothesis -> statistically significant difference in the positive and negative sets -> “good” feature
Weighting visual features

- The inverted p-value of Kolmogorov-Smirnov test could be used as weight for the feature
- “floral”: 

![Graph showing weighting of visual features]
Weighting visual features

- More examples: “shades”
Weighting visual features

- More examples: "cute"
Query expansion and search

- User employs text-based search to obtain an initial set
- For each item in the initial set:
  - Load the corresponding weight vector for each keyword
  - Obtain an expanded weigh vector from the textual description.

\[ \tilde{q}'(Item_i, Query) = \tilde{q}_i \times (\alpha \cdot \tilde{w}_Q + \beta \cdot \tilde{w}_E) \]
Query expansion and search

• Query: “floral”
• Initial set:
Query expansion and search

- CBIR-query vectors
Query expansion and search

- iLike-query vectors
Results

“Floral”
Results

- **iLike**: our approach
- **Baseline**: Pure CBIR
- **Query**: “floral”

We are able to infer the implicit user intension behind the query term, identify a subset of visual features that are significant to such intension, and yield better results.
Visual thesaurus

- Statistical similarities of the visual representations of the text terms

<table>
<thead>
<tr>
<th>Words</th>
<th>First Few Words in Visual Thesaurus</th>
</tr>
</thead>
<tbody>
<tr>
<td>feminine</td>
<td>bandeau, hipster, breezy, pregnancy, hem, lifestyle, braid, comfy, femininity.</td>
</tr>
<tr>
<td>flirty</td>
<td>flirt, bikini, vibrant, effortlessly, pointelle, dressy, edgy, splashy, swimsuit</td>
</tr>
<tr>
<td>gingham</td>
<td>subtle, sparkly, floral, gauze, glamour, sassy, surplice, beautifully, pajama</td>
</tr>
<tr>
<td>trendy</td>
<td>adorn, striking, playful, supersoft, shiny, nancy, ladylike, cuddly, closure</td>
</tr>
<tr>
<td>pinstripe</td>
<td>smock, sporty, khaki, pleat, oxford, geometric, gauzy, ruffle, chic, thong</td>
</tr>
<tr>
<td>embroider</td>
<td>suede, crochet, versatility, ultra, corduroy, spectrum, softness, faux, crease</td>
</tr>
<tr>
<td>twill</td>
<td>complement, plaid, contour, logo, decorative, buckle, classically, tagless</td>
</tr>
</tbody>
</table>
Conclusion and future work

• iLike: find the “visual perception” of keywords
• Better recall compared with text-based search
• Better precision: understand the needs of the users

• Better “understanding” of keywords: NLP?
• More features?
• Segmentation: feature+region?
Thank you!

Questions?