A Picture is Worth a Thousand Tags: Automatic Web Based Image Tag Expansion

Andrew Gilbert (a.gilbert@surrey.ac.uk, www.AndrewJohnGilbert.co.uk) Richard Bowden (R.Bowden@surrey.ac.uk) Faculty of Engineering and Physical Sciences, University Of Surrey, Guildford, Surrey, GU2 7XH

Problems

- Ever increasing amounts of images captured and stored by users,
- How to organise them for retrieval?
- Often images only tagged with a few high level keywords
- Tagging images is time consuming and limited to the users knowledge of the image

Aim

- Automatically enrich images with new unseen complex contextual text annotations
- Use associated web pages to collect unsupervised dataset,
- No large labelled groundtruth

Traditional Tags

- Proposed Tags

<table>
<thead>
<tr>
<th>Road</th>
<th>Sky</th>
<th>Building</th>
<th>Big Ben</th>
<th>Bus</th>
</tr>
</thead>
<tbody>
<tr>
<td>London</td>
<td>London</td>
<td>Charles Barry</td>
<td>London</td>
<td></td>
</tr>
</tbody>
</table>

Approach

1. A small subset of existing image tags, are used as internet searches
2. The text and image information from the resultant web pages, form a textual and visual bag of words model and is symbolised for data mining.
3. Association rule mining is then used to identify rules which relate words to the visual contents.
4. Unseen images that contain these visual features can then be annotated with the related textual features

Image Feature Descriptor

- A compact generic representation of the media
- Symbolised to allow mining to take place
- Invariant to the feature type
- For a given image from a webpage, a feature descriptor is formed of both the visual features (visual Bag of Words) and text (traditional textual Bag of Words)
- Features include SIFT, GIST and Colour Histogram
- The feature descriptor weighted histogram is symbolised based on its frequency in a “Transaction”

Removal of Frequent words

- Within a webpage there will be a large amount common text that has little to do with the semantic meaning.
- The most frequent and un discriminative words are identified by the mining and removed

Material

- Given an unseen image, if a subset compound of image features matches a rule, the image can be appended with the text tag from that rule

Results—ImageTag

- To identify co-occurring words within the Transactions form from the web page images
- Uses A Priori association rule mining
- Identifies all possible rules Features-> text tag
- The most frequent rules will be common across multiple text tags, so find discriminative unique rules instead
- Use the “confidence” of the rule: The number of cases in which the rule is correct relative to the number of cases in which it is applicable.

Results ImageCLEF 2010

- On average, each image is annotated with 12.8 additional complex linguistic text tags.
- Standard image annotation /concept dataset, “sky, water, portrait, night, sun, red, love, old, cute, and red”
- Use 8K training images to learn rules and Annotate unseen images compared to groundtruth

Conclusion

- The approach is able to create test tags - that provide more information than manual user tagging.
- This is because the images are being tagged with the keywords harvested from detailed internet web pages.