An Efficient and Effective Region-based Image Retrieval Framework

Feng Jing
Tsinghua University
Outline

- Content-based image retrieval
- Region-based image retrieval
- Overview of our framework
- Technologies
  - Image representation;
  - Indexing techniques;
  - Learning techniques
- Experiments
Content-based Image Retrieval

- **Motivation**
  - Text-based image retrieval
    - Manual annotation is a tedious task.
    - “An image is worth a thousand words”
    - Some visual contents are difficult to describe

- **Goal**
  - Retrieve images based on their content
  - What is content:
    - Automatically extracted visual features.
    - Color, texture, shape and so on.
  - Query by an example image or sketch
Content-based Image Retrieval

- **Challenges**
  - High dimensionality of features
    - Dimension of color histogram > 100;
    - Storage and indexing.
  - Gap between features and semantics.
    - Similar features might represent different semantic concepts;
    - Images of similar semantics might have very different features.
Content-based Image Retrieval
To shorten or bridge the gap

- **Relevance feedback**
  - Originally developed for text retrieval
  - Supervised active learning technique.

- **Better representation**
  - Region-based representation.
Region-based Image Retrieval (RBIR)

- **Motivation**
  - Global representation vs. region representation;
  - Represent and retrieve images at the granularity of region (Object in ideal case);
  - Perception of human visual system.

- **Typical RBIR systems**
  - Netra (UCSB)
  - Blobworld (Berkeley)
  - Walrus (Bell-lab)
  - SIMPLIcity (Stanford)
Key Issues

- **Image similarity measure**
  - Region-to-region (Blobworld, Netra)
    - Sensitive to image segmentation;
  - Image-to-image (SIMPLIcity, Walrus)
    - All the regions;
    - Importance of regions is area percentage.
- **Our solution**
  - Earth Mover’s Distance (EMD);
  - The importance of regions is learned from user’s feedback.
Key Issues (Cont.)

- **Efficiency issues**
  - Storage
  - Indexing
    - Tree structure: R*-tree (Walrus);
    - Clustering techniques: SIMPLIcity & Netra.
- **Our solution**
  - Region codebook (storage);
  - Inverted files (Indexing).
Key Issues (Cont.)

- Relevance feedback
  - Global feature-based RF is fruitful.
  - Little attention for region-based RF;
  - Our solution
    - Query Point Movement;
    - Support Vector Machines;
    - Region importance learning.
Overview Of The Framework

1. Pretreatment
   - Region-Based Retrieval

2. Region-Based Retrieval
   - All Positive Examples
   - Learning RIs
   - Updating CRIs
   - End

3. Relevance Feedback
   - User Feedback
   - Neg examples?
     - Yes
       - Neg & Pos Examples
     - No
       - Pos Examples
   - Most Positive Images
   - Learning SVM Classifier
   - Learning Optimal Query

4. Region Weighing
   - Ranking Using EMD
   - Selected Images
   - Indexing

5. Clustering
   - Codebook
   - Modified Inverted Files

6. Image Database
   - Image Segmentation and Region Feature Extraction
   - Representation: Compact & Sparse
Pretreatment

- **Image segmentation**
  - HSEG (region growing).

- **Region representation**
  - Low-level visual feature:
    - Color moments
      - CIE Luv color space
      - Three scales of moments (9)
    - Any other feature is available.
  - Importance (weight)
    - Constraint: sum be 1.
Region codebook design:
- K-Means;
- Regions of all the images in the database
- The size of the codebook (K);
- Codeword properties:
  - The center;
  - Indexing structure (modified inverted file).
Pretreatment (Cont.)

Examples of the Codewords

…………

…………

…………

…………

………

………

………

………

………

………

………

………

………

………
Image Representation

- **Image encoding:**
  - Low level feature → codeword index

- **Compact Representation:**
  - A region set;

- **Sparse (Vector) Representation:**
  - A N dimensional vector;
  - N is the number of the codewords;
  - Number of regions vs. number of codewords.
Image Similarity Measure

- **Compact representation**
- **Earth Mover’s Distance (EMD)**
  - Originally introduced as a flexible similarity measure between multidimensional distributions;
  - Based on the minimal cost that must be paid to transform one distribution into another;
  - Matches perceptual similarity well;
  - Can operate on variable-length representations.
Indexing Using Modified Inverted Files

- **Tree structures:**
  - Curse of dimensionality;
  - The performance of R*-trees degrades by a factor of 12 as the number of dimensions increases from 5 to 10.

- **Inverted Files (IF):**
  - Widely used in text retrieval community;
  - The IF of a codeword are the IDs of the images contain it.
  - Indexing using IF
    - Find out the images within at least one of the IF of the query’s codewords;
    - Only sort these images with other images neglected
Modified Inverted Files:

- Expand codewords according to importance.
- Expand $\lfloor w \cdot k \rfloor$ codewords to a region with importance $w$.
- The upper bound $k$ is 10 currently.
Relevance Feedback

- **Query Point Movement (QPM):**
  - Only positive examples.
  - Vector representation;
  - Optimal query:
    \[
    \overline{I}_{opt} = \frac{\sum_{k=1}^{n} \overline{I}_k}{n}
    \]
  - Indexing using modified IFs.
Relevance Feedback

- **Support Vector Machine (SVM)**
  - Both negative and positive examples;
  - Classification problem;
  - A classifier that separates positive examples from negative ones.
  - Compact representation;
  - Why SVM
    - strong theoretical foundations
    - excellent empirical successes
New Kernel

- Motivation:
  - Common kernels (Gaussian) depend on $L_p$ norm or inner product in the input space.

- Generalized Gaussian kernel:
  \[ k_{GGaussian}(x, y) = \exp(-d(x, y)/2\sigma^2) \]
  
- EMD-based kernel:
  \[ k_{GEMD}(x, y) = \exp(-EMD(x, y)/2\sigma^2) \]
Region Importance Learning

- Basic assumption: Important regions should appear more times in the positive images.
Some definitions:

- Region $R_i$ is similar to region $R_j$, if $d(R_i, R_j) < \varepsilon$,
- $s(R, I) = \begin{cases} 1, & \text{if one region of image } I \text{ is similar to } R \\ 0, & \text{otherwise} \end{cases}$
- **Region Frequency (RF):** $RF_i = \sum_{j=1}^{n^+} s(R_i, I_j^+)$
- **Inverse Image Frequency (IIF):**

\[
IIF_i = \log(N \sqrt{\sum_{j=1}^{N} s(R_i, I_j)})
\]
Region Importance Learning (Cont.)

- **Region Importance (RI):**

\[
RI_i = \frac{RF_i \times IIF_i}{\sum_{j=1}^{n} (RF_j \times IIF_j)}
\]

the *RI* of all the positive images can be learned.
Cumulate Region Importance

**Why to cumulate?**

- The region importance ($RI$) is similar to common users.
- The feedback information from users is often incomplete.
- The accumulation can make the $RI$ more robust and meaningful.
How to cumulate?

- The cumulated importance of region after \( l (>0) \) updates is:

\[
CRI_i(l) = \frac{CRI_i(l-1) \cdot (l-1) + RI_i}{l}
\]

- \( RI_i \) is the latest \( RI \).
- \( CRI_i(0) \) is initialized to be area percentage.
- Once updated, area percentage is ignored.
Experimental setup

- Image DB size: 10,000 (from Corel)
- Query number: 200 (10 categories with each 20 queries)
- Test categories:
Experimental setup

- For each of the query images, 5 iterations of user-and-system interaction were carried out.
- At each round of feedback, the system examined the top 30 images.
- When the top $N$ images are considered and there are $R$ relevant images, the precision within top $N$ images is defined to be $P(N) = R / N$.
- Accuracy = the average precision within top 50 images, i.e. average $P(50)$. 
We use 400 (the first turning point) as the size of the codebook.
## Indexing Evaluation

### Search Time

<table>
<thead>
<tr>
<th>Method</th>
<th>Search Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEQ</td>
<td>351 ms</td>
</tr>
<tr>
<td>IF</td>
<td>70 ms</td>
</tr>
<tr>
<td>MIF</td>
<td>91 ms</td>
</tr>
</tbody>
</table>

### Average Precision vs. Scope

- **MIF**
- **IF**
- **SEQ**
Relevance Feedback Evaluation

- No RF in most of the existing RBIR systems
- Same color moment feature
- Performance of SVM is better
- The speed of QPM is fast

![Graph showing accuracy vs. number of iterations for RQPM, GQPM, RSVM, and GSVM.](image)
Region Weighting Evaluation

- To learn the CRIs, 10,000 random query and feedback sessions were carried out.
- An example:

Cumulated region importance (red and italic) and area percentage of the regions of an image.
Region Weighting Evaluation

Scope vs. Average Precision

- **AP**
- **CRI**

Scope values range from 10 to 100.
Region Weighting Evaluation

Accuracy vs. Number of Iterations for QPM(CRI), QPM(AP), SVM(CRI), and SVM(AP).
Conclusions

- **Region codebook**
  - Save storage;
  - Facilitate indexing and relevance feedback.

- **Modified inverted files**
  - Save retrieval time.

- **Relevance feedback**
  - Query point movement;
  - Support Vector Machines.

- **Region importance learning**
  - Reasonable, reflects semantic importance;
  - Cumulated for future use.
Related Publications


Q&A

http://scenery.nease.net

Thanks!