Scene Completion of Selfie Photos based on Image Search

Woong Jick Youn
2016.06.16

Advisor: Sung-Eui Yoon
Images on Social Network Service

- Selfies are focused on their faces
- How can we expand the background?
- Can we make a larger selfie with scene completion?
Scene Completion on Large-scale Images

- Scene Completion Using Millions of Photographs
  [Hays and Efros, SIGGRAPH 07]
- Data-driven approach with millions of images
Scene Completion on Large-scale Images

- Finding visually similar images with nearest neighbor search
- But, expensive for a large image collection
- To solve the scalability issue, we can consider ANN search

Images from Scene Completion Using Millions of Photographs
Approximate Nearest Neighbor (ANN) Search

- Characterizes the neighborhood rather than identifying the exact neighborhood themselves
- Finds the NN with high probability
- Can be much faster, but missing some nearest matches
Our Approach

- Enlarges a selfie with scene completion through image search
- Expands upward and both sides with image patches
- Reduces memory footprint by ANN search
Overview of Our Approach

① Extracts GIST and LBP-HSV descriptors
② Product quantization
③ Patch-based image search
④ Poisson image editing for image completion
⑤ Iterates ③ and ④ for all the patches
Overview of Our Approach

1. Extracts GIST and LBP-HSV descriptors
2. Product quantization
3. Patch-based image search
4. Poisson image editing for image completion
5. Iterates ③ and ④ for all the patches
1. Descriptors

- **GIST**
  - Extracts in 8 x 8 blocks with 4 scales (4,608 dim)

- **LBP-HSV**
  - HSV color space for texture information
  - Extracts in 8 x 8 blocks (3,776 dim on each HSV channel)
1. Descriptors

- **GIST**
  - Extracts in 8 x 8 blocks with 4 scales (4,608 dim)
  - 32GB from 1M images

- **LBP-HSV**
  - Extracts in 8 x 8 blocks (3,776 dim on each HSV channel)
  - 78GB (26GB on each HSV channel) from 1M images

- Over 100GB for two features

- Scalability issue

- Solve this issue with product quantization
Overview of Our Approach

① Extracts GIST and LBP-HSV descriptors

② **Product quantization**

③ Patch-based image search

④ Poisson image editing for image completion

⑤ Iterates ③ and ④ for all the patches
Product Quantization (PQ)

- One of the ANN search techniques
- Vector quantization with block-wise Cartesian product
- Decomposes the original vector space into disjoint subspaces
- Effectively generates clusters in each subspace
Product Quantization (PQ)

- Product quantization on both descriptors
  - Generates 4 codebooks
  - 64 subspaces (Utilizing 8 x 8 spatial bins of descriptors)
  - Each subspace is quantized to 256 codewords

Images from Product quantization for nearest neighbor search
Overview of Our Approach

① Extracts GIST and LBP-HSV descriptors
② Product quantization
③ Patch-based image search
④ Poisson image editing for image completion
⑤ Iterates ③ and ④ for all the patches
③ Patch-based Image Search

- Generates a target region by a user input
- Expand the boundary regions to marginal regions
  - Reduces drastic color changes
- A small patch searches a similar image by PQ
Patch-based Image Search

- Distance computation for search by PQ
  - The distance of GIST descriptor
    \[ d_{\text{gist}}(x, y) = \sqrt{\sum_{i=1}^{m} \| x - q_i(y) \|^2} \]
  - The distance of LBP-HSV descriptors
    \[ d_h(x, y) = \sqrt{\sum_{i=1}^{m} \| x - q_i(y) \|^2} \]
    \[ d_s(x, y) = \sqrt{\sum_{i=1}^{m} \| x - q_i(y) \|^2} \]
    \[ d_v(x, y) = \sqrt{\sum_{i=1}^{m} \| x - q_i(y) \|^2} \]
  - The distance the query and image collection
    \[ dist(x, y) = d_{\text{gist}}(x, y) + d_h(x, y) + d_s(x, y) + d_v(x, y) \]
Overview of Our Approach

① Extracts GIST and LBP-HSV descriptors
② Product quantization
③ Patch-based image search
④ Poisson image editing for image completion
⑤ Iterates ③ and ④ for all the patches
④ Poisson Image Editing

- Seamless Image Cloning by a Closed Form Solution of a Modified Poisson Problem [SIGGRAPH Asia 12]
  - Can control the level of color adaptation
  - Suppresses drastic color changes

Images from Seamless Image Cloning by a Closed Form Solution of a Modified Poisson Problem
4 Poisson Image Editing

- Blends similar image on the boundary regions only
- Preserves the color in blended regions
- Can keep the visual information on the rest of the region

Patch image + Similar image from dataset → Result of image completion
Overview of Our Approach

① Extracts GIST and LBP-HSV descriptors
② Product quantization
③ Patch-based image search
④ Poisson image editing for image completion
⑤ Iterates ③ and ④ for all the patches
5 Iteration

- Steps for patch-based image completion
Dataset

- MIRFlickr-1M
- One million Flickr images
- Excludes grayscale images for the dimension consistency
- After excluding, 936,800 images are used

Images of MIRFlickr-1M dataset
Result

- Memory efficient
  - Descriptors are compressed to 250MB from 110GB (450x)
  - Runnable on a typical computer
- Contains a creative content
- Avoids the duplication problem caused by a copy and paste manner

![Target image](image1.png)  ![Final result](image2.png)
Compared our result with content-aware fill in Photoshop CC
Result

- Contains a creative content

Query image  Content-aware fill  Ours
Result

- Avoids the duplication problem

Query image | Content-aware fill | Ours
Conclusion

- Scene completion based on product quantization
- Memory efficiency of image search
- Extrapolates the background where visual information does not exist
- Generates a creative image
Future work

- A larger image collection with tags
- Graph matching algorithm
- Fine matching for improving side patches
Publication

- 이미지 검색을 활용한 자가촬영 사진의 배경 합성

윤웅직, 윤성의

IPIU 2016
Thank you

Any questions or comments?
Nearest Neighbor Search

- Extracts image descriptors
- Compares the distance between descriptors
- Finds nearest neighbor features from all the database images

\[ NN(x) = \arg \min_y \| x - y \| \]

Nearest neighbour matching
- expensive to do for all frames
Nearest Neighbor Search

- NN search is expensive for a large image collection
- Matching two images (1k 128-d SIFT descriptors)
  NN search in 0.4 seconds
- To solve the scalability issue, we can consider ANN search

<table>
<thead>
<tr>
<th># of images</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>~ 7 mins</td>
</tr>
<tr>
<td>10,000</td>
<td>~ 1h 10 mins</td>
</tr>
<tr>
<td>10 million</td>
<td>~ 115 days</td>
</tr>
</tbody>
</table>
Approximate Nearest Neighbor (ANN) Search

- Characterizes the neighborhood rather than identifying the exact neighborhood themselves
- Finds the NN with high probability
- Can be much faster, but missing some nearest matches
- E.g. LSH, ITQ, PQ ...
Scene Completion

- BiggerPicture: Data-Driven Image Extrapolation Using Graph Matching [Wang et al., SIGGRAPH Asia 14]
- Hierarchically segmented to build graphs in multi levels
- Graph matching algorithm
Scene Completion

Input image

Scene Descriptor

Image Collection

20 completions

Context matching + blending

200 matches

Slides from Scene Completion Using Millions of Photographs
Result

- Compare with extrapolation method

Query image:  
- BiggerPicture: 5mins for segmentation, 5secs for graph matching, 3secs for stitching and alignment
- Ours: 13secs for extrapolation with low memory footprint

Wang et al.: 

Ours: