A Co-Saliency Model of Image Pairs

IEEE Transaction on Image Processing
vol. 20, No. 12, 2011
Hongliang Li, and King Ngi Ngan

Presented by Bong-Seok Choi

School of Electrical Engineering and Computer Science
Kyungpook National Univ.
Abstract

- **Goal of proposed method**
  - Detecting co-saliency from image pair
    - Extracting similar objects from image pair

- **Proposed method**
  - Using co-saliency model
    - Single-Image Saliency Map (SISM)
      - Describing local attention
        » Using three saliency detection techniques
    - Multi-Image Saliency Map (MISM)
      - Using co-multilayer pyramid
      - Describing each node in graph
        » two types of visual descriptor (color and texture)
        » Evaluation similarity between two nodes
          » Using SimRank algorithm
Introduction

- Visual attention model
  - Saliency based visual attention model
    - Making multi-scale image features into single saliency map
    - Using MRF by integrating computational visual attention mechanism

- Previous method of extracting visual attention
  - Detecting saliency point
    - Based on center-surround mechanism
  - Measuring visual saliency
    - Using Site Entropy Rate
  - Context-aware saliency detection
  - Global contrast based method
Detecting saliency object from image pair

- Applying computer vision and multimedia
  - Common pattern discovery
  - Image matching and Co-recognition
- Procedure of detecting saliency object
  - Measuring degree of similarity
  - Extracting object by grouping together similar pixels
Similar work with proposed method

- Co-segmentation method
  - Aim to segment similar object
    - Matching common part of histogram
  - Minimizing energy with MRF term

Proposed perceptual model

- Entity in pair of images as co-saliency
  - Strong local saliency with region in pair
  - Region pair should exhibit high similarity of features
    - Intensity, color, texture, or shape
Proposed co-saliency model
- Combination of SISM and MISM
  - SISM model
    - Itti`s saliency model
    - Frequency-tuned saliency (FTA)
    - Spectral residual saliency (SRA)
  - MISM model
    - Finding co-salient object from image pair
      » Performing co-multilayer by image pyramid decomposition
      » Computing distance of node-pair
        » Using color and texture descriptor
        » Computing similarity score
        » Using normalized single-pair SimRank algorithm
Proposed Method

- Single-Image Saliency Map
  - Achieving robust saliency detection
    - Weighted saliency detection method
      - Combining several saliency map
        » Itti`s saliency model
        » Frequency-tuned saliency (FTA)
        » Spectral residual saliency (SRA)
      - Corresponding single saliency map

\[
S_i = \sum_{j=1}^{J} \omega_j \cdot \mathcal{N}(S_{ij})
\]  

(1)

where \( \mathcal{N}(S_{ij}) \) denotes j th normalized saliency map
\( \omega_j \) denotes weight with \( \sum_{j=1}^{J} \omega_j = 1 \)
- Illustration of single image saliency map
  - Comparison with each method

Fig. 1. Example of the single-image saliency map. (a) Original image amira. (b) saliency map by itti's method. (c) Saliency map by FTA method. (d) Saliency map by SRA method. (e) proposed single-image saliency map.
Multi-Image Saliency Map

- Goal of MISM
  - Extracting multi-image saliency information

- Definition of Multi image saliency map

\[
S_g(I_i(p)) = \max_{q \in I_j} \text{sim}(I_i(p), I_j(q))
\]

where \( p \) and \( q \) denote entities in images \( I_i \) and \( I_j \).

\( \text{sim}(\ ) \) represents a function that measures similarity between two entities.
- Block diagram of proposed multi image saliency detection
  - Pyramid decomposition
  - Feature extraction
  - SimRank optimization
  - Multi-image saliency computation

Fig. 2. Block diagram of the multi-image saliency extraction
Pyramid decomposition of an image pair

- Decomposing image pair into multiple segmentation
  - Grouping pixels into “superpixels”
    » Roughly homogeneous in size and feature
  - Computing region of finer pyramid resolution by region of coarse level
    » Coarse level as parents region
    » Sub-region as children region

Region feature extraction

- using two properties for region descriptor
  - Color descriptor
    » Describing color variation in region
  - Texture descriptor
    » Describing texture property in region
Fig. 3. Block diagram region feature extraction (e.g., the region with yellow color).
• Creating color descriptor of region
  - Using RGB, L*a*b*, YCbCr color space
  - Representing pixel as 9-dimensional color vector
    » Combining three color space
  - Quantizing pixels in image pair into N clusters
    » Using k-means clustering algorithm
  - Computing histogram each region by counting number of codeword
    » Representing color descriptor by N bins of histogram
• Creating texture descriptor of region
  − Extract \( p \times p \) patches from color images
  − Vectorization of each patch
    » Single vector size \( p^2 \)
  − Quantizing pixels in image pair into \( M \) clusters
    » Using \( k \)-means clustering algorithm
  − Combining series of histogram of patchwords
    » Measuring frequency of patchwords
    » Creating texture descriptor
  − Final texture descriptor

\[
f'(k) = [H_{3 \times 3}(k), H_{5 \times 5}(k), H_{7 \times 7}(k), \ldots]
\]  \( (3) \)

where \( H_{i \times i}(k) \) denotes histogram computed for \( k \)th region of size \( i \times i \)
The Co-Multilayer Graph Representation

- Designing co-multilayer graph

\[ G = (V, E) \quad \text{with nodes } v \in V \text{ and edges } e \in E \]

Fig. 3. Our co-multilayer graph model.
• Representing edges
  - Weight function to each edge of graph
    » Given N nodes, get N(N-1)/2 links between nodes.
    » Considering edges between nodes within adjacent layer
  - Representing weight for edge

  \[
  \omega_{ij} = \begin{cases} 
  \exp(-\theta_f d(f_i, f_j)), & \text{if } l_i - l_j = -1 \text{ or } l_i - l_j = 0 \\
  0, & \text{if } \|l_i - l_j\| > 1 \text{ or } l_i > l_j 
  \end{cases}
  \]

  with

  \[
  d(f_i, f_j) = \chi^2(f_i, f_j) = \sum_{z=1}^{Z_f} \frac{(f_i(z) - f_j(z))^2}{f_i(z) + f_j(z)}
  \]

  where \(i\) and \(j\) denote two nodes.
  \(f_i\) and \(f_j\) denote color texture descriptor for regions.
  \(Z_f\) denote dimensional number of descriptor.
  \(\theta_f\) is constant, controls strength of weight.
  \(\chi^2()\) denote chi-square distance
Normalized simrank similarity computation

- Computing similarity score of two region nodes
  - Similarity score between object a and b

\[
s(a, b) = \frac{C}{|\text{In}(a)||\text{In}(b)|} \sum_{i=1}^{\text{In}(a)} \sum_{j=1}^{\text{In}(b)} s\left(\text{In}_i(a), \text{In}_j(b)\right)
\]

(6)

where C is decay factor between 0 and 1

\[|\text{In}(a)| \text{ and } |\text{In}(b)| \text{ denote number of in-neighbors } \text{In}(a) \text{ and } \text{In}(b)\]

for nodes a and b

- Normalization of SimRank score to measure similarity

\[
s^* (a,b) = \frac{s(a, b)}{\max\{s(a,a), s(b,b)\}}
\]

(7)
Multi-image saliency map

» Substituting eq.(7) into eq.(2)

\[ S_g(I_i(p)) = \max_{q \in I_j} s^*(I_i(p), I_j(q)) \] (8)

where p and q denote region nodes in image pair \((I_i, I_j)\)
Co-saliency Map

- Extracting co-saliency map for image pair \((I_i, I_j)\)
  - Combining two saliency maps eq.(1) and eq.(8)

\[
SS(I_i(p)) = \alpha_1 \cdot S_t(I_i(p)) + \alpha_2 \cdot S_g(I_i(p)) \\
= \alpha_1 \cdot S_t(I_i(p)) + \alpha_2 \cdot (\alpha_3 \cdot S_g^c(I_i(p)) + \alpha_4 \cdot S_g^t(I_i(p))) \\
= \beta_1 \cdot S_t(I_i(p)) + \beta_2 \cdot S_g^c(I_i(p)) + \beta_3 \cdot S_g^t(I_i(p)), \\
\text{for all } p \in R\{I_i\}
\] (9)

where \(\beta_j\) is a constant with \(\beta_1 + \beta_2 + \beta_3 = 1\) that is used to control impact of SSIM and MISM on image co-saliency.

\(S_g^c\) and \(S_g^t\) denote MISM obtained by color and texture descriptors.

<table>
<thead>
<tr>
<th>Symbols</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>(I_i)</td>
<td>The ith image</td>
</tr>
<tr>
<td>SS</td>
<td>Co-saliency map</td>
</tr>
<tr>
<td>(S_t)</td>
<td>Single-image saliency map</td>
</tr>
<tr>
<td>(S_g^c)</td>
<td>MISM by color descriptors</td>
</tr>
<tr>
<td>(S_g^t)</td>
<td>MISM by texture descriptors</td>
</tr>
<tr>
<td>(\beta_1)</td>
<td>Weight for the SISM</td>
</tr>
<tr>
<td>(\beta_2, \beta_3)</td>
<td>Weights for the MISM</td>
</tr>
</tbody>
</table>
Experiments

- Detection result of image pairs

Fig. 5. (a) Original image pairs. (b) Ground truth masks.
Configuration of each image sequence

Fig. 6. (a) The test images (i.e., banana, amira, and dog). (b) SISMs. (c) MISM. (d) Co-saliency maps by our method. (e) Co-saliency images w.r.t. (d).
Fig. 7. Experimental results for single objects. (a)-(b) and (e)-(f): Original image pairs. (c)-(d) and (g)-(h): Results by our method.
- Performance evaluation

**Table 2.** Performance Evaluation by Object Criterion

<table>
<thead>
<tr>
<th>Image Pair</th>
<th>Results of [14]</th>
<th>Results of [15]</th>
<th>Our Results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Recall</td>
<td>Precision</td>
<td>$F$</td>
</tr>
<tr>
<td>banana</td>
<td>0.2760</td>
<td>0.6404</td>
<td>0.4908</td>
</tr>
<tr>
<td>amira</td>
<td>0.3989</td>
<td>0.7143</td>
<td>0.6040</td>
</tr>
<tr>
<td>kim</td>
<td>0.3005</td>
<td>0.9055</td>
<td>0.6182</td>
</tr>
<tr>
<td>stone</td>
<td>0.4941</td>
<td>0.6717</td>
<td>0.6203</td>
</tr>
<tr>
<td>dog</td>
<td>0.1919</td>
<td>0.3617</td>
<td>0.3003</td>
</tr>
<tr>
<td>llama</td>
<td>0.5390</td>
<td>0.7396</td>
<td>0.6811</td>
</tr>
<tr>
<td>face</td>
<td>0.0051</td>
<td>0.0132</td>
<td>0.0096</td>
</tr>
<tr>
<td>flower</td>
<td>0.3868</td>
<td>0.4215</td>
<td>0.4129</td>
</tr>
<tr>
<td>sign</td>
<td>1.0000</td>
<td>0.7236</td>
<td>0.7729</td>
</tr>
<tr>
<td>horse</td>
<td>0.2515</td>
<td>0.6783</td>
<td>0.4874</td>
</tr>
<tr>
<td>coke</td>
<td>0.4737</td>
<td>0.9720</td>
<td>0.7821</td>
</tr>
<tr>
<td>man</td>
<td>0.7763</td>
<td>0.6430</td>
<td>0.6695</td>
</tr>
</tbody>
</table>
– Result of multiple objects

Fig. 8. Experimental results for multiple objects. (a)-(b): Original image pairs. (c)-(d): Results by our method.
– Evaluation of other image

Fig. 9. results for 210 images. (a) Precision-recall bars for adaptive thresholds. (b) Precision-recall curves for varying thresholds.
Fig. 10. Comparison of results of co-segmentation with other methods. First row: Original image pairs including stone, amira, llama, and horse. Second row: Results by the method [28]. Third row: Results by the method [27]. Fourth row: Results by our method.
Fig. 11. Illustration of tracking accuracy in sequence “traffic condition”: the Euclidean distance between the estimated objection position and the ground truth is plotted against frame numbers.
Fig. 12. Illustration of tracking accuracy in sequence “traffic condition”: the Euclidean distance between the estimated objection position and the ground truth is plotted against frame numbers.
Discussion and Conclusion

Goal of proposed method

- Detecting co-saliency from image pair
  - Extracting similar objects from image pair
- Using co-saliency model
  - Single-Image Saliency Map (SISM)
    - Describing local attention
      » Using three saliency detection techniques
  - Multi-Image Saliency Map (MISM)
    - Using co-multilayer pyramid
    - Describing each node in graph
      » two types of visual descriptor (color and texture)
      » Evaluation similarity between two nodes
        » Using SimRank algorithm