Exploring Image Relationships Using Functional Maps

FAN WANG1, RAIF RUSTAMOV2, ADRIAN BUTSCHER3, JUSTIN SOLOMON4, LEONIDAS J. GIUBAS5

Geometric Computing Group, Stanford University
1 fanw@stanford.edu, 2raifrustamov@gmail.com, 3adrian.butscher@gmail.com, 4justin.solomon@stanford.edu, 5guibas@cs.stanford.edu

ABSTRACT
Establishing correspondences between images is challenging, especially when point-to-point matching is hard to obtain due to large variability in object appearances. In this work, instead of computing point-based correspondences between images, we find functional maps between them which can act as information transporters between the images. Functional maps are based on correspondences between local properties or attributes defined over images, and can be found efficiently by solving a linear system. We use functional maps to transfer segmentation from a set of training images to a test image, and obtain results that match or improve other state-of-the-art methods.

1 Problem Setup

1.1 Graph Construction

• Node: over-segmented superpixels by Normalized Cuts
• Edge: the length of the shared boundary of the two superpixels normalized by the average perimeter. Only exists for adjacent superpixels.

Image descriptors can be regarded as functions on the graph \( f \in \mathcal{F}(G, \mathbb{R}) \).

1.2 Reduced Functional Space

Basis selection:
• Local bases: representing meaningful parts of a graph and capturing local distortions.
• Global bases: any function can be well reconstructed by only a few basis functions.

Laplacian eigenfunctions as basis \( \Phi \):
1. Compact.
2. Multi-scale.
Each function can be represented as coefficient vector \( \vec{f} \sim \Phi^T f \).

2 Problem Formulation

2.1 Constraints on Functional Maps

Probe functions: \( f_1, f_2, \ldots \in \mathcal{F}(G, \mathbb{R}), g_1, g_2, \ldots \in \mathcal{F}(G, \mathbb{R}) \);
Coefficient vectors in \( \Phi \) and \( \Phi^T \): \( F = [a_1, a_2, \ldots] \) and \( F^T = [b_1, b_2, \ldots] \); Functional map: \( X_{fi} \sim \vec{f}_i \).

We can recover \( \vec{f} \) by solving the optimization problem
\[ X_{fi} = \arg \min \|X_{fi} - F_i\|_F \]
where \( \| \cdot \|_F \) is the Frobenius norm.

Many natural relationships between images can be incorporated:
• visual descriptors of the graph nodes, such as color.
• SIFT, shape descriptor, bag-of-visual-words, etc.
• point-to-point landmark correspondences between the two images.
• region-to-region correspondence.

We take color features and bag-of-visual-words features as probe functions.

2.2 Regularization

The map \( T \) commutes with Laplacian operator:
\[ L_i(f \circ T i) = (L_i f) \circ T i \forall f \in \mathcal{F}(G, \mathbb{R}) \]

In functional space, applying \( T^i \) followed by \( L_i \) yields
\[ L_i(T^i f) = \Phi_i(\Phi^T_i T^i f)(\Phi^T_i f) = \Phi_i(L_i T^i f) \]

The Laplacian \( L_i \) followed by \( T^i \) yields
\[ T^i(L_i f) = \Phi_i(\Phi^T_i L_i f)(\Phi^T_i f) = \Phi_i(L_i T^i f) \]

The commutation means:
\[ \Phi_i(L_i T^i f) = \Phi_i(X_i T^i f) \]

The regularization can be represented as \( X_i T^i = L_i X_i \).

2.3 Optimization

The optimization problem is formulated as:
\[ \min \|X_i - F_i\|_F + \lambda \|X_i - L_i X_i\|_F \]

When using Laplacian eigenfunctions as basis \( \Phi \), the regularizer becomes simpler:
\[ \min \|X_i - F_i\|_F + \lambda \|X_i - X_i\|_F \]

The commutativity of regularizer can be further rewritten as an element-wise sum:
\[ \|X_i - X_i\|_F^2 = \sum_{i, u} \|\delta_i - X_i^u\|^2 \]

\( X_i^u \) will have significant values on the diagonal and near-diagonal entries, and almost zero everywhere else.

3 Transferring the Segmentation

Training: the ground truth segmentation is an indicator function \( f_{gt} \).
Testing: the transferred segmentation is \( f_{\theta} \).

Merging the Results

Voting mechanism to combine the mapped indicator functions from \( N \) training images:
\[ \hat{f}_{\theta} = \frac{1}{N} \sum_{i=1}^{N} \omega_i f_{\theta_i} \]
The weight \( \omega_i \) is determined based on its distance \( d_i \) to the test image using the GIST descriptor:
\[ \omega_i = e^{-d_i^2/2\sigma^2} \]

4 Experimental Results

Data set: PASCAL VOC challenge 2012
• real-world consumer images from Flickr;
• 20 object classes;
• pixel-level segmented images.

4.1 Parameter Selection

Performance changing with number of basis functions and regularizer weight \( \lambda \).

4.2 Results on PASCAL VOC 2012

<table>
<thead>
<tr>
<th>class</th>
<th>Best</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>background</td>
<td>85.1</td>
<td>82.8</td>
</tr>
<tr>
<td>airplane</td>
<td>65.4</td>
<td>62.8</td>
</tr>
<tr>
<td>bicycle</td>
<td>11.0</td>
<td>20.6</td>
</tr>
<tr>
<td>bird</td>
<td>51.3</td>
<td>48.4</td>
</tr>
<tr>
<td>boat</td>
<td>44.5</td>
<td>43.2</td>
</tr>
<tr>
<td>bottle</td>
<td>58.9</td>
<td>33.3</td>
</tr>
<tr>
<td>bus</td>
<td>60.8</td>
<td>61.3</td>
</tr>
<tr>
<td>car</td>
<td>61.5</td>
<td>48.1</td>
</tr>
<tr>
<td>cat</td>
<td>56.4</td>
<td>60.3</td>
</tr>
<tr>
<td>chair</td>
<td>22.6</td>
<td>14.6</td>
</tr>
<tr>
<td>cow</td>
<td>23.6</td>
<td>60.9</td>
</tr>
<tr>
<td>dining table</td>
<td>22.6</td>
<td>28.1</td>
</tr>
<tr>
<td>dog</td>
<td>47.4</td>
<td>55.4</td>
</tr>
<tr>
<td>horse</td>
<td>37.6</td>
<td>60.5</td>
</tr>
<tr>
<td>motorbike</td>
<td>79.7</td>
<td>51.7</td>
</tr>
<tr>
<td>person</td>
<td>31.9</td>
<td>35.4</td>
</tr>
<tr>
<td>potted plant</td>
<td>35.7</td>
<td>29.7</td>
</tr>
<tr>
<td>sheep</td>
<td>55.3</td>
<td>62.7</td>
</tr>
<tr>
<td>sofa</td>
<td>40.8</td>
<td>33.9</td>
</tr>
<tr>
<td>tv/monitor</td>
<td>34.2</td>
<td>63.1</td>
</tr>
</tbody>
</table>

Analysis
• better on natural objects, such as cat, cow, dog, horse, sheep, etc. Not sensitive to non-rigid deformations.
• worse on bottle, car, person if objects in test images rarely appear in training images.
• low accuracy on bicycle, bottle and chair because of coarse superpixels.