Adaptive Nonparametric Image Parsing

IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, T-CSVT, VOL. 25, NO. 10, OCTOBER 2015

Tam V. Nguyen,
Can yi Lu,
Jose Sepulveda,
Shuicheng Yan,

Department For Technology, Innovation And Enterprise, Singapore Polytechnic, Singapore
Outline

• Introduction

• Overview
  • Locality-Aware Retrieval Set
  • Adaptive Nonparametric Superpixel Classification
  • Contextual Smoothing

• Experiments

• Conclusion
Introduction

Liu et al. [4] proposed a nonparametric image parsing method based on estimating scale-invariant feature transform (SIFT) flow, a dense deformation field between images. Given a test and a training image, the annotated category labels of the training pixels are transferred to the test ones via pixel correspondences.

However, inference via pixelwise SIFT flow is currently very complex and computationally expensive.

Introduction

Tighe and Lazebnik [5] further transferred labels at the level of superpixels. In this scheme, given a test image, the system searches for the top similar training images based on global features. The superpixels of the most similar images are obtained as a retrieval set. Then, the label of each superpixel in the test image is assigned based on the corresponding $k$ most similar superpixels in the retrieval set.

Introduction

There are several shortcomings in existing nonparametric methods.

1. quite difficult to get globally similar images to form the retrieval set.
2. $k$ is fixed empirically in advance in such a nonparametric image parsing scheme.

Main issues

1. how to get a good retrieval set
2. how to choose a good $k$ for initial label transfer
Introduction

1. We propose the *locality-aware retrieval set*. The *locality-aware* retrieval set is extracted from the training data based on **superpixel** matching similarities, which are augmented with feature extraction for better differentiation of local superpixels.

2. We propose an *adaptive* method

To set the sample-specific $k$ as the number of the **fewest nearest neighbors** that similar training superpixels can use to get their best category label predictions.
Outline

- Introduction
- Overview
  - Locality-Aware Retrieval Set
  - Adaptive Nonparametric Superpixel Classification
  - Contextual Smoothing
- Experiments
- Conclusion
Overview

Test Image

Super-pixels

Over-segmentation

Locality-aware Retrieval Set

Adaptive Nonparametric Super-pixel Classification

Contextual Smoothing

Final Labeling
Locality-Aware Retrieval Set

For nonparametric image parsing, one important step of parsing a test image is to find a retrieval set of training images.

A good retrieval set should contain images of a similar scene type as that of the test image, along with similar objects and spatial layouts.

Unlike [5] where global features are used to obtain the retrieval set, we utilize the superpixel matching.
Locality-Aware Retrieval Set

Retrieval set is selected based on local similarity measured over superpixels.

We utilize linear discriminant analysis (LDA) [24] for feature reduction to a lower feature dimension.

Locality-Aware Retrieval Set

\[ \hat{x} = Wx \]

\( x \in \mathbb{R}^{n \times 1} \) original feature vector of the superpixel dimension

\( \hat{x} \) corresponding feature vector after the feature reduction

\( W \) transformation matrix
Locality-Aware Retrieval Set

LDA looks for the directions that are most effective for discrimination by minimizing the ratio between the intra-category ($S_w$) and inter-category ($S_b$) scatters.

$$W^* = \arg \min_W \frac{|W^T S_w W|}{|W^T S_b W|}$$

$$S_w = \sum_{i=1}^{N} (x_i - \bar{x}^c_i)(x_i - \bar{x}^c_i)^T$$

$$S_b = \sum_{c=1}^{N_c} n_c (\bar{x}^c - \bar{x})(\bar{x}^c - \bar{x})^T$$

- $N$ # of superpixels in all training images
- $N_c$ # of categories
- $n_c$ # of superpixels for the $c$th category
- $x_i$ feature vector of one training superpixel
- $i=1...N$
- $c_i$ category label of the $i$th superpixel in the training images
- $\bar{x}$ mean of feature vector
- $\bar{x}^c$ mean of the $c$th category
Locality-Aware Retrieval Set

LDA looks for the directions that are most effective for discrimination by minimizing the ratio between the intra-category ($S_w$) and inter-category ($S_b$) scatters.

$$W^* = \arg \min_W \frac{|W^T S_w W|}{|W^T S_b W|}$$

$$S_w = \sum_{i=1}^{N} (x_i - \bar{x}^c_i)(x_i - \bar{x}^c_i)^T$$

$$S_b = \sum_{c=1}^{N_c} n_c (\bar{x}^c - \bar{x})(\bar{x}^c - \bar{x})^T$$

Projection matrix $W^*$ is composed of the eigenvectors of $S_w^{-1}S_b$ at most $N_c-1$ eigenvectors with nonzero real corresponding eigenvalues. The dimensionality of $W$ is $N_c - 1 \times n_x$.

$$\hat{x} = Wx$$
Algorithm 1 Locality-Aware Retrieval Set Algorithm

1: parameters: \( n_q, n_t, N_I, Q, T \).
2: The unique index set \( S = \emptyset \).
3: \( v = 0 \in \mathbb{R}^{N_I} \).
4: for \( i = 1 : n_q \) do
5: \[ [\eta_i, \Delta_i] \leftarrow \text{KNN}(Q_i, T, k_m); \]
6: \( \eta_i \leftarrow \eta_i \setminus S; \)
7: if \( \eta_i \neq \emptyset \) then
8: \( \eta_i \leftarrow \text{REFINEINDEXSET}(\eta_i); \)
9: \( I_i \leftarrow \text{FINDIMAGEINDEX}(\eta_i); \)
10: \( v(I_i) \leftarrow v(I_i) + 1./\Delta_i(\eta_i); \)
11: \( S \leftarrow S \cup \eta_i; \)
12: end if
13: end for
14: \( v = \text{NORMALIZEANDSORT}(v). \)
15: \( k_r = \arg \min_u \frac{\sum_{j=1}^{n_t} v_{ij}}{\sum_{j=1}^{N_I} v_j} \geq \tau. \)
16: return top \( k_r \) training images.

- \( n_q \) # of superpixels in the test image
- \( n_j^t \) # of superpixels for the \( j \) th training image
- \( N_I \) # of training images

one superpixel in a training image is matched with only one superpixel of the test image.  
The unique index set (stores the indices of the already matched superpixels) 
\( v \) similarity vector between the test image and all training images 
\( Q \in \mathbb{R}^{(N_c-1)\times n_q} \) feature matrix for all the superpixels in the test image 
\( T \in \mathbb{R}^{(N_c-1)\times (\sum_j n_j^t)} \) feature matrix for all the superpixels in the training set
Algorithm 1 Locality-Aware Retrieval Set Algorithm

1: parameters: \( n_q, n_t, N_I, Q, T \).
2: The unique index set \( S = \emptyset \).
3: \( v = 0 \in \mathbb{R}^{N_I} \).
4: for \( i = 1 : n_q \) do
5: \([\eta_i, \Delta_i] \leftarrow \text{KNN}(Q_i, T, k_m);\)
6: \( \eta_i \leftarrow \eta_i \setminus S; \)
7: if \( \eta_i \neq \emptyset \) then
8: \( \eta_i \leftarrow \text{REFINEINDEXSET}(\eta_i); \)
9: \( I_i \leftarrow \text{FINDIMAGEINDEX}(\eta_i); \)
10: \( v(I_i) \leftarrow v(I_i) + 1./\Delta_i(\eta_i); \)
11: \( S \leftarrow S \cup \eta_i; \)
12: end if
13: end for
14: \( v = \text{NORMALIZEANDSORT}(v). \)
15: \( k_r = \arg \min_u \frac{\sum_{j=1}^{N_I} v_j}{N_I} \geq \tau. \)
16: return top \( k_r \) training images.

\( \eta_i \) indices of the returned nearest superpixels of the \( i \) th test superpixel \( \Delta_i \) corresponding distances of the returned nearest superpixels to the \( i \) th test superpixel

17: function \( \text{REFINEINDEXSET}(\eta, \Delta) \)
18: \( d = \infty \in \mathbb{R}^{N_I}. \)
19: \( \Gamma = \emptyset. \)
20: for \( i = 1 : |\eta| \) do
21: if \( d(m(\eta_i)) > \Delta_i \) then
22: \( d(m(\eta_i)) = \Delta_i; \)
23: else
24: \( \Gamma = \Gamma \cup i; \)
25: end if
26: end for
27: return \( \Gamma. \)
28: end function
Algorithm 1 Locality-Aware Retrieval Set Algorithm

1: parameters: $n_q$, $n_t$, $N_I$, $Q$, $T$. 
2: The unique index set $S = \emptyset$. 
3: $v = 0 \in \mathbb{R}^{N_I}$. 
4: for $i = 1:n_q$ do 
5:    $[\eta_i, \Delta_i] \leftarrow \text{KNN}(Q_i, T, k_m)$; 
6:    $\eta_i \leftarrow \eta_i \setminus S$; 
7:    if $\eta_i \neq \emptyset$ then 
8:        $\eta_i \leftarrow \text{REFINEINDEXSET}(\eta_i)$; 
9:        $I_i \leftarrow \text{FINDIMAGEINDEX}(\eta_i)$; 
10:        $v(I_i) \leftarrow v(I_i) + 1./\Delta_i(\eta_i)$; 
11:        $S \leftarrow S \cup \eta_i$; 
12:    end if 
13: end for 
14: $v = \text{NORMALIZEANDSORT}(v)$. 
15: $k_r = \arg \min_u \frac{\sum_{j=1}^u v_j}{N_I} \geq \tau$. 
16: return top $k_r$ training images.

\(\eta_i\) indices of the returned nearest superpixels of the $i$ th test superpixel $\Delta_i$ corresponding distances of the returned nearest superpixels to the $i$ th test superpixel.

29: function FINDIMAGEINDEX($\eta$) 
30:    $\Gamma = \infty \in \mathbb{R}^{\left|\eta\right|}$. 
31: for $i = 1:\left|\eta\right|$ do 
32:    $\Gamma_i = m(\eta_i)$; 
33: end for 
34: return $\Gamma$. 
35: end function
Algorithm 1 Locality-Aware Retrieval Set Algorithm

1: parameters: $n_q, n_t, N_I, Q, T$.
2: The unique index set $S = \emptyset$.
3: $v = 0 \in \mathbb{R}^{N_I}$.
4: for $i = 1:n_q$ do
5:     $[\eta_i, \Delta_i] \leftarrow \text{KNN}(Q_i, T, k_m)$;
6:     $\eta_i \leftarrow \eta_i \setminus S$;
7:     if $\eta_i \neq \emptyset$ then
8:         $\eta_i \leftarrow \text{REFINEINDEXSET}(\eta_i)$;
9:         $I_i \leftarrow \text{FINDIMAGEINDEX}(\eta_i)$;
10:        $v(I_i) \leftarrow v(I_i) + 1. / \Delta_i(\eta_i)$;
11:        $S \leftarrow S \cup \eta_i$;
12:    end if
13: end for
14: $v = \text{NORMALIZEANDSORT}(v)$.
15: $k_f = \arg\min_u \frac{\sum_{j=1}^{n_t} v_j}{\sum_{j=1}^{N_I} v_j} \geq \tau$.
16: return top $k_f$ training images.

36: function \text{NORMALIZEANDSORT}(v)
37: \hspace{1em} $\Gamma = \infty \in \mathbb{R}^{|v|}$.
38: for $i = 1:|v|$ do
39: \hspace{2em} $\Gamma_i = v_i / \min(n_i^t, n_q)$;
40: end for
41: $\Gamma = \text{sort}(\Gamma)$.
42: return $\Gamma$.
43: end function
Algorithm 1 Locality-Aware Retrieval Set Algorithm

1: **parameters:** $n_q$, $n_t$, $N_I$, $Q$, $T$.  
2: The unique index set $S = \emptyset$.  
3: $v = 0 \in \mathbb{R}^{N_I}$.  
4: **for** $i = 1 : n_q$ **do**  
5: $[\eta_i, \Delta_i] \leftarrow \text{KNN}(Q_i, T, k_m)$;  
6: $\eta_i \leftarrow \eta_i \setminus S$;  
7: **if** $\eta_i \neq \emptyset$ **then**  
8: $\eta_i \leftarrow \text{REFINEINDEXSET}(\eta_i)$;  
9: $I_i \leftarrow \text{FINDIMAGEINDEX}(\eta_i)$;  
10: $v(I_i) \leftarrow v(I_i) + 1. / \Delta_i(\eta_i)$;  
11: $S \leftarrow S \cup \eta_i$;  
12: **end if**  
13: **end for**  
14: $v = \text{NORMALIZEANDSORT}(v)$.  
15: $k_R = \arg\min_u \frac{\sum_{j=1}^{u} v_j}{\sum_{j=1}^{N_I} v_j} \geq \tau$.  
16: **return** top $k_R$ training images.
Adaptive Nonparametric Superpixel Classification

Adaptive nonparametric superpixel classification aims to overcome the limitation of the traditional $k$-NN algorithm, which usually assigns the same number of nearest neighbors for each test sample.

Our improved $k$-NN algorithm focuses on looking for the suitable $k$ for each test sample.
Adaptive Nonparametric Superpixel Classification

Each training image \( t \) retrieved by the superpixel matching process is considered as one test image, while the left \( N_I - 1 \) images in the training set are referred to the corresponding training set.

We perform superpixel matching to obtain the retrieval set for \( t \) and assign the label \( l^k_i \) of the \( i \) th superpixel by the majority vote of the \( k \) nearest superpixels in the retrieval set

\[
    l^*_i = \arg \max_{l_i} L(k, l_i)
\]

\[
    L(k, l_i) = \frac{P(i|l_i, k)}{P(i|l^*_i, k)} = \frac{n(l_i, \text{NN}(i, k)) / n(l_i, D)}{n(l^*_i, \text{NN}(i, k)) / n(l^*_i, D)}.
\]

\( \# \) of superpixels with class label \( l_i \) in the \( k \) nearest superpixels of the \( i \) th superpixel in the retrieval set.
Adaptive Nonparametric Superpixel Classification

Then, we compute the **per-pixel accuracy** of each retrieved training image $t$ for different $k$s from 1 to 50.

$A_{tk}$ : per-pixel performance of the training image $t$ with the parameter value $k$. 

![Histogram of Image Frequency](image-url)
Adaptive Nonparametric Superpixel Classification

For each test image,

\[ k^* = \arg \max_k \sum_{t=1}^{k_r} A_{tk} \]

per-pixel performance of the training image \( t \) with the parameter value \( k \)

Based on selected \( k^* \), the initial label of a superpixel in the test image is obtained in the same way as in [4].
Adaptive Nonparametric Superpixel Classification

similar images should share the same $k$
Contextual Smoothing

In general, the initial labels for the superpixels may still be noisy, and these labels need be further refined with global context information.

Therefore, the initial labels are smoothened with an Markov random field (MRF) energy function defined over the field of pixels:

\[
E(l) = \sum_{i \in V} E_d(i, l_i) + \lambda \sum_{e_{ij} \in E} E_s(l_i, l_j)
\]

- **data term**
- **smoothing term**

all pixels in the image  \( \sum_{i \in V} E_d(i, l_i) \)

edges connecting adjacent pixels  \( \sum_{e_{ij} \in E} E_s(l_i, l_j) \)
Contextual Smoothing

Data term

\[ E_d(i, l_i) = -\log L(k^*, l_{sp(i)}) \]

Smoothing term

\[ E_s(l_i; l_j) = -\xi_{ij} \times \log \left( \frac{P(l_i | l_j) + P(l_j | l_i)}{2} \right) \times \delta[l_i \neq l_j] \]

\[ \xi_{ij} = \frac{\nabla_{ij}}{\sum_{e_{pq} \in E} \nabla_{pq}} \quad \text{(normalized gradient value)} \]

\[ \nabla_{ij} = \|I(i) - I(j)\|^2 \]
Outline

• Introduction
• Overview
  • Locality-Aware Retrieval Set
  • Adaptive Nonparametric Superpixel Classification
  • Contextual Smoothing
• Experiments
• Conclusion
Experiments

Data sets:

1. SIFT Flow [4]
   2488 training images and 200 test images (256 × 256 pixels)
   33 semantic labels

2. 19-Category Label Me [28].
   250 training images and 100 test images
   19 categories

\(\lambda\) is set as 16 in the contextual smoothing process.

\(k_m\) and \(\tau\) are set as 1000 and 0.3

Performance Comparison Of Our Algorithm With Other Algorithms On The SIFTflow Data Set [4].

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Per-Pixel (%)</th>
<th>Per-Category (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Parametric Baselines</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tighe et al. [14]</td>
<td>78.6</td>
<td><strong>39.2</strong></td>
</tr>
<tr>
<td>Parabct et al. [16]</td>
<td>78.5</td>
<td>29.6</td>
</tr>
<tr>
<td><strong>Nonparametric Baselines</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liu et al. [4]</td>
<td>74.8</td>
<td>–</td>
</tr>
<tr>
<td>Tighe et al. [5]</td>
<td>76.3</td>
<td>28.8</td>
</tr>
<tr>
<td>Tighe et al. [5] (adding geometric information)</td>
<td>76.9</td>
<td>29.4</td>
</tr>
<tr>
<td>Myeong et al. [11]</td>
<td>76.2</td>
<td>29.6</td>
</tr>
<tr>
<td>Eigen et al. [17]</td>
<td>77.1</td>
<td>32.5</td>
</tr>
<tr>
<td><strong>Our Proposed Adaptive Nonparametric Algorithm</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Super-pixel Classification</td>
<td>77.2</td>
<td>34.9</td>
</tr>
<tr>
<td>Contextual Smoothing</td>
<td><strong>78.9</strong></td>
<td>34.0</td>
</tr>
</tbody>
</table>

Experiments

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Per-Pixel (%)</th>
<th>Per-Category (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k = 1$</td>
<td>70.2</td>
<td>31.9</td>
</tr>
<tr>
<td>$k = 5$</td>
<td>76.6</td>
<td>34.8</td>
</tr>
<tr>
<td>$k = 10$</td>
<td>77.5</td>
<td>34.6</td>
</tr>
<tr>
<td>$k = 20$</td>
<td>77.8</td>
<td>33.5</td>
</tr>
<tr>
<td>$k = 30$</td>
<td>77.9</td>
<td>33.3</td>
</tr>
<tr>
<td>$k = 40$</td>
<td>77.9</td>
<td>30.6</td>
</tr>
<tr>
<td>$k = 50$</td>
<td>77.8</td>
<td>29.5</td>
</tr>
<tr>
<td>$k = 60$</td>
<td>77.5</td>
<td>28.6</td>
</tr>
<tr>
<td>$k = 70$</td>
<td>77.8</td>
<td>28.5</td>
</tr>
<tr>
<td>$k = 80$</td>
<td>77.5</td>
<td>28.2</td>
</tr>
<tr>
<td>$k = 90$</td>
<td>77.1</td>
<td>27.2</td>
</tr>
<tr>
<td>$k = 100$</td>
<td>76.9</td>
<td>26.8</td>
</tr>
<tr>
<td><strong>Adaptive $k$ in Our Algorithm</strong></td>
<td><strong>78.9</strong></td>
<td><strong>34.0</strong></td>
</tr>
</tbody>
</table>

Performance Comparison Of Different Ks And Our Algorithm On The SIFTflow Data Set [4].
## Experiments

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>76.3 (28.8)</td>
</tr>
<tr>
<td>SuperParsing [5]</td>
<td>76.4 (31.2)</td>
</tr>
<tr>
<td>SuperParsing + LDA + Global Matching + (fixed $k = 20$)</td>
<td>77.8 (33.5)</td>
</tr>
<tr>
<td>SuperParsing + LDA + Super-pixel Matching + ($k = 20$)</td>
<td>78.9 (34.0)</td>
</tr>
</tbody>
</table>

Performance Comparison Of Different Settings On The SIFTflow Data Set [4].
Experiments
Experiments

Normalized Discounted Cumulative Gain (NDCG) [29]

\[
\text{NDCG}@k_r = \frac{1}{Z} \sum_{i=1}^{k_r} \frac{2^{\text{rel}(i)} - 1}{\log(i + 1)}
\]

whether the scene of the returned image is relevant

<table>
<thead>
<tr>
<th>Retrieval Set Algorithm</th>
<th>NDCG</th>
</tr>
</thead>
<tbody>
<tr>
<td>GIST-based matching [4]</td>
<td>0.83</td>
</tr>
<tr>
<td>Global matching [5]</td>
<td>0.85</td>
</tr>
<tr>
<td>Super-pixel matching</td>
<td>0.88</td>
</tr>
</tbody>
</table>

Experiments

Similar images should share the same $k$

<table>
<thead>
<tr>
<th>Scene Class</th>
<th>Mean No. of Categories</th>
<th>Selected $k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coast</td>
<td>3.8</td>
<td>12</td>
</tr>
<tr>
<td>Forest</td>
<td>2.5</td>
<td>36</td>
</tr>
<tr>
<td>Highway</td>
<td>6.5</td>
<td>6</td>
</tr>
<tr>
<td>Inside City</td>
<td>7.2</td>
<td>12</td>
</tr>
<tr>
<td>Mountain</td>
<td>2.6</td>
<td>22</td>
</tr>
<tr>
<td>Open Country</td>
<td>3.9</td>
<td>14</td>
</tr>
<tr>
<td>Street</td>
<td>7.5</td>
<td>6</td>
</tr>
<tr>
<td>Tall Building</td>
<td>3.3</td>
<td>43</td>
</tr>
</tbody>
</table>

Mean Number Of Categories And The Correspondingly Selected $k$ Of Each Scene Class On The SIFTflow Data Set
Experiments

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Per-Pixel (%)</th>
<th>Per-Category (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Parametric Baselines</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jain et al. [28]</td>
<td>59.0</td>
<td>–</td>
</tr>
<tr>
<td>Chen et al. [30]</td>
<td>75.6</td>
<td>45.0</td>
</tr>
<tr>
<td><strong>Nonparametric Baselines</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mycong et al. [6]</td>
<td>80.1</td>
<td>53.3</td>
</tr>
<tr>
<td><strong>Adaptive Nonparametric Algorithm</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Super-pixel Classification</td>
<td>80.3</td>
<td>53.3</td>
</tr>
<tr>
<td>Contextual Smoothing</td>
<td><strong>82.7</strong></td>
<td><strong>55.1</strong></td>
</tr>
</tbody>
</table>

Performance Comparison Of Our Algorithm With Other Algorithms On The 19-category LabelMe Data Set [28].
Experiments
Outline

• Introduction

• Overview
  • Locality-Aware Retrieval Set
  • Adaptive Nonparametric Superpixel Classification
  • Contextual Smoothing

• Experiments

• Conclusion
Conclusion

This paper has presented a novel approach to image parsing that can take advantage of adaptive nonparametric superpixel classification.

To the best of our knowledge, we are the first ones to exploit the locality-aware retrieval set and adaptive nonparametric superpixel classification in image parsing.