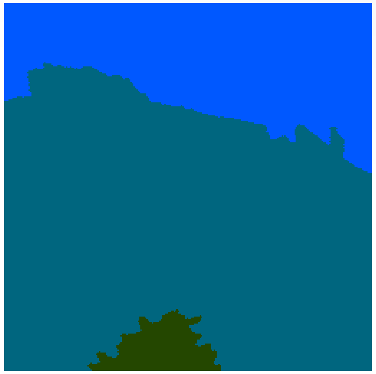


Nonparametric Scene Parsing with Adaptive Feature Relevance and Semantic Context

Gautam Singh
Jana Kosecka

Emine Gül DANACI
N12129719
Master Student

Related Works

- J. Tighe and S. Lazebnik. Superparsing: Scalable nonparametric image parsing with superpixels. In ECCV (5), pages 352–365, 2010.

- D. Eigen and R. Fergus. Nonparametric image parsing using adaptive neighbor sets. In CVPR, pages 2799–2806, 2012.

- D. Ramanan and S. Baker. Local distance functions: A taxonomy, new algorithms, and an evaluation. PAMI, 33(4):794–806, 2011.

Approach

- Local patches (gradient orientation, color, location features)
- k-NN
- Semantic Labelling

Semantic segmentation with small superpixels

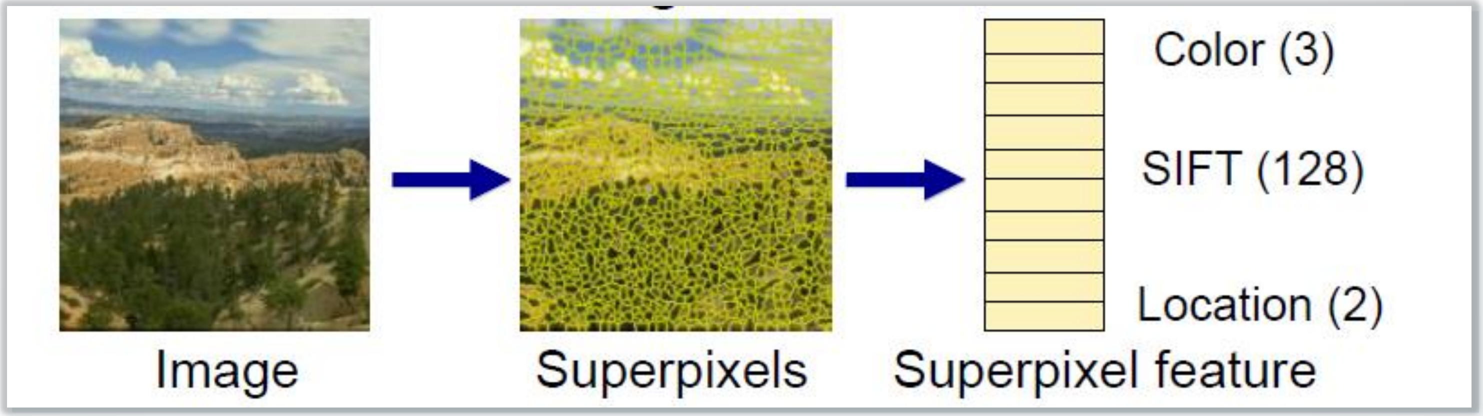
Posterior probability of labelling L

$$P(\mathbf{L}|\mathtt{A}) = \frac{P(\mathtt{A}|\mathbf{L})\, P(\mathbf{L})}{P(\mathtt{A})}.$$

Estimated value of labelling L as MAP

$$\operatorname*{argmax}_{\mathbf{L}} P(\mathbf{L}|\mathbf{A}) = \operatorname*{argmax}_{\mathbf{L}} P(\mathbf{A}|\mathbf{L})\, P(\mathbf{L}).$$

Superpixels and features



Retrieval set

- Global image features
  - GIST
  - Spatial Pyramid of quantized SIFT
  - RGB-Color Histograms

- Euclidean distance from the query image

- Ranking

- Re-ranking

Appearance likelihood

Naive Bayes assumption

$$P(\mathbf{A}|\mathbf{L}) \approx \prod_{i=1}^{S} P(\mathbf{a}_i|l_i).$$

Label likelihood score

$$L(\mathbf{a}_i, l_j) = \frac{n(l_j, N_{ik})/n(l_j, G)}{n(\bar{l_j}, N_{ik})/n(\bar{l_j}, G)}$$

- $\bar{l_j} = L \setminus l_j$ is the set of all labels excluding l_j;
- N_{ik} is a neighbourhood around $\mathbf{a}_i$ with exactly k points in it;
- $n(l_j, N_{ik})$ is the number of superpixels of class l_j inside N_{ik};
- $n(l_j, G)$ is the number of superpixels of class l_j in the set G

Weighted k-NN

Color, SIFT and location

$$d_f^{ij} = [d_c^{ij}, d_s^{ij}, d_l^{ij}]^\top$$

Weighted distance

$$d_w^{ij} = w^\top d_f^{ij} \qquad w = [w_1, w_2, w_3] \in \Re^3$$

Weight computation

i: feature channel
N: neighbourhood

x: query point
c: influence of weights

$$r_i(\mathbf{z}) = \sum_{l_j=1}^{nL} \frac{(P(l_j|\mathbf{z}) - \bar{P}(l_j|x_i = z_i))^2}{\bar{P}(l_j|x_i = z_i)}$$

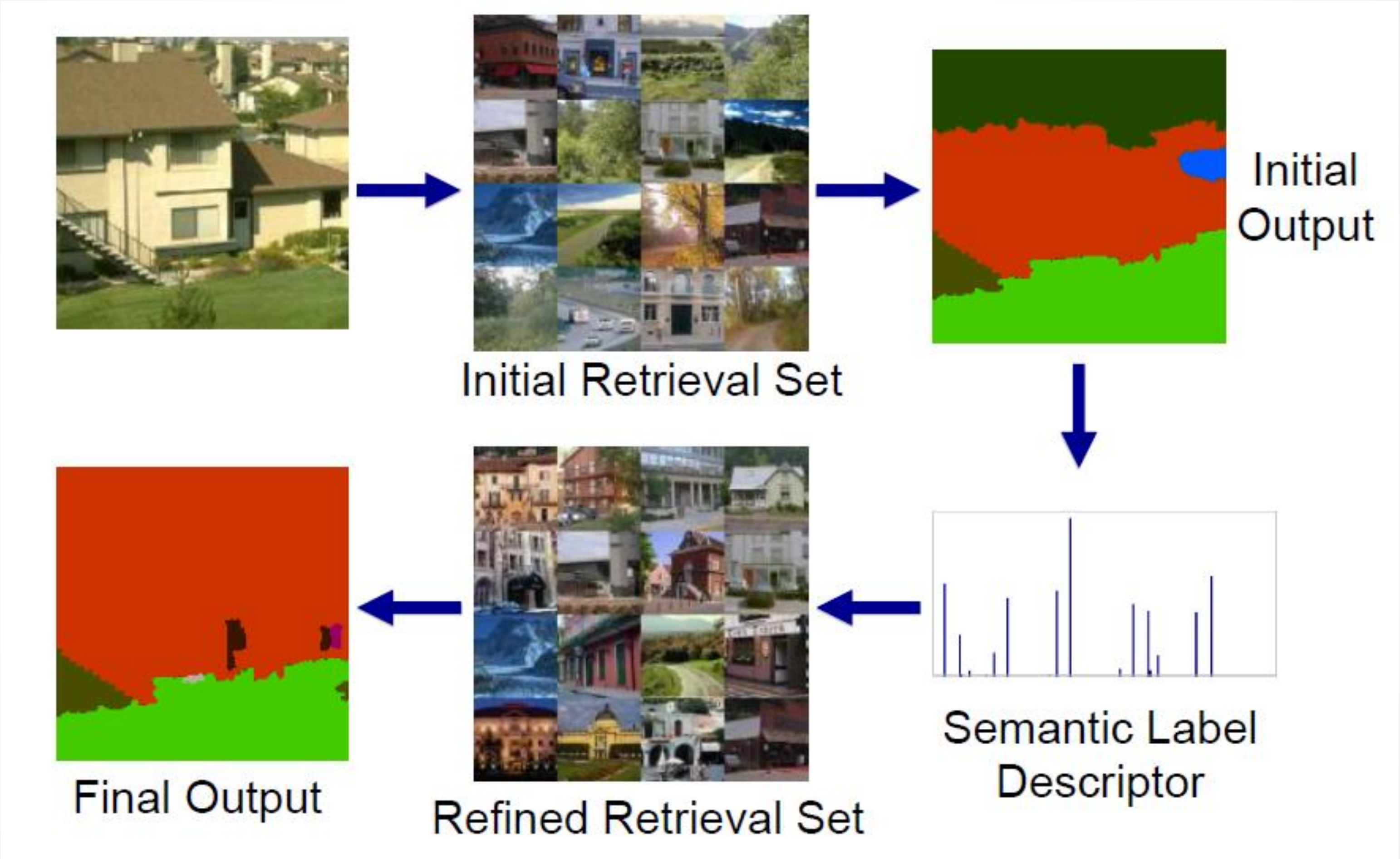
$$\bar{r}_i(\mathbf{x_0}) = \frac{1}{|N(\mathbf{x_0})|} \sum_{\mathbf{z} \in N(\mathbf{x_0})} r_i(\mathbf{z})$$

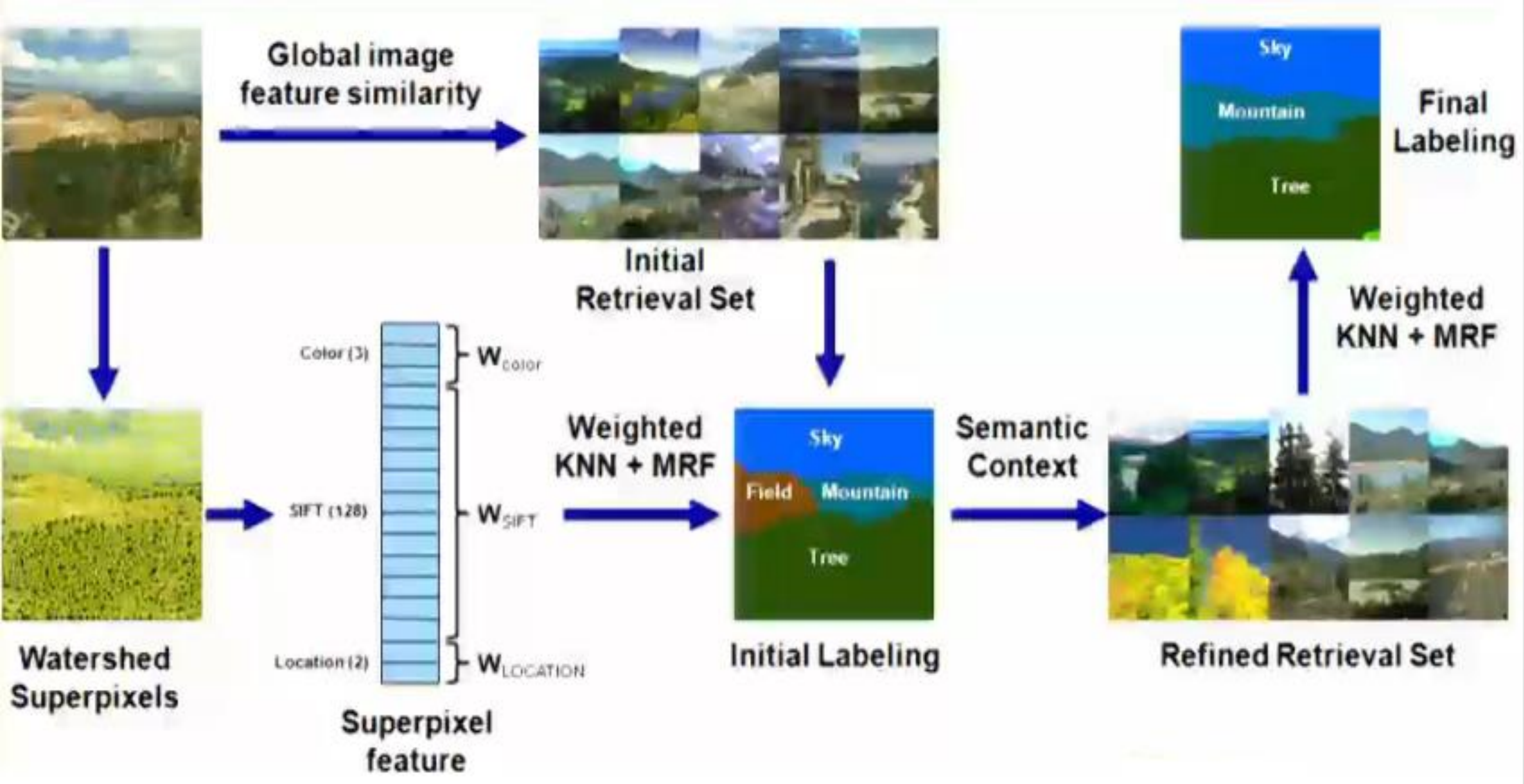
$$w_i(\mathbf{x_0}) = \frac{exp\,(cR_i(\mathbf{x_0}))}{\sum_{p=1}^{m} exp\,(cR_p(\mathbf{x_0})}$$

Query based algorithm

1) *Initialize* **w** *to uniform weights*
2) *At query point x_0, find set N_0 of K_0 nearest neighbors using current* **w**
3) *For each feature channel, compute relevance estimate using points in N_0*
4) *Update* **w** *using relevance estimates from 3*
5) *Repeat steps 2-4 (five times in our experiments)*

Semantic retrieval set




Global image feature similarity
Initial Retrieval Set
Sky
Mountain
Tree
Final Labeling
Weighted KNN + MRF
Color (3)
SIFT (128)
Location (2)
W_{color}
W_{SIFT}
$W_{LOCATION}$
Weighted KNN + MRF
Sky
Field
Mountain
Tree
Semantic Context
Watershed Superpixels
Superpixel feature
Initial Labeling
Refined Retrieval Set

Experiments

- Datasets
  - SiftFlow – 2688 images with 33 semantic categories
    - 2488 for training 200 for testing
  - SUN09 – fully labeled, 107 semantic categories
    - 4352 for training 4310 for testing
  - Google Street View – labeled data, 320 images
    - 160 for training 160 for testing
  - Stanford Background – semantic and geometric labeled 715 images
    - 572 for training 143 for testing

Results

System	Per-Pixel	Per-Class
Liu et al. [15]	76.7	-
Tighe et al. [26]	76.9	29.4
Eigen et al. [4]	77.1	32.5
UKNN-MRF	75.6	27.9
WKNN-MRF	77.2	29.3
WLKNN-MRF	78.5	32.0
WAKNN-MRF	79.2	33.8
WKNN-MRF (with HOG)	76.7	27.4

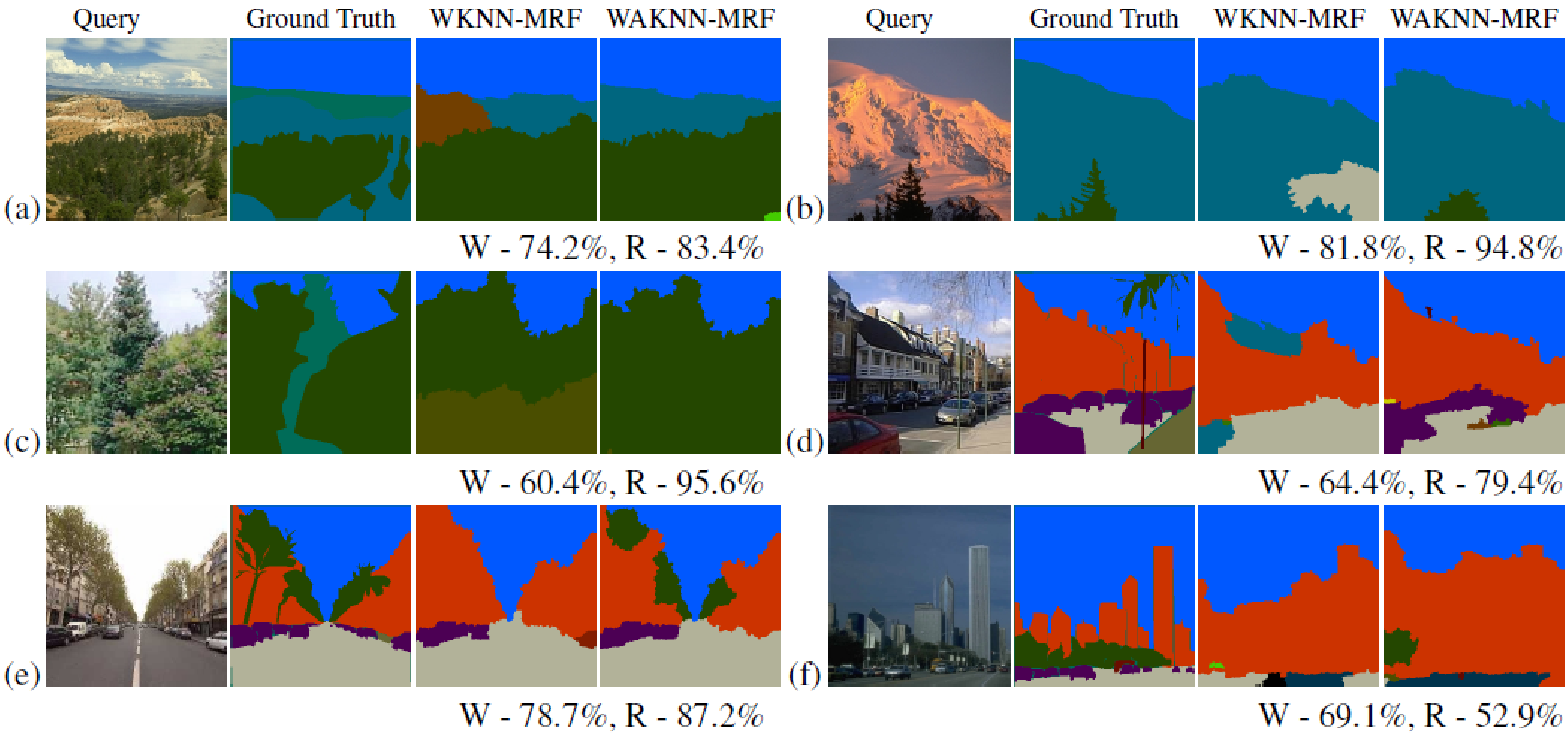
Table 1. Semantic labelling performance on the SiftFlow dataset

Results

Dataset and System	Per-Pixel	Per-Class
SUN09		
Choi et al. [1]	33.0	10.6
[25] CascALE Expert	49.3	16.7
[25] CascALE Sharing	52.8	15.2
WKNN-MRF	49.5	8.7
WAKNN-MRF	53.1	12.1
Google Streetview		
Zhang et al. [32]	88.4	80.4
Zhang et al. [31]	93.2	73.1
Singh et al. [24]	94.4	81
WAKNN-MRF	93.7	76.4
Stanford background		
[6] Pixel CRF	74.3	66.6
[6] Region Energy	76.4	65.5
[20] Leaf Level	72.8	58.0
[20] Hierarchy	76.9	66.2
WKNN-MRF	73.6	61.2
WAKNN-MRF	74.1	62.2

Table 2. Performance on the SUN09, Google-Streetview and Stanford background datasets.

Results


Query
Ground Truth
WKNN-MRF
WAKNN-MRF
Query
Ground Truth
WKNN-MRF
WAKNN-MRF
(a)
W - 74.2%, R - 83.4%
(b)
W - 81.8%, R - 94.8%
(c)
W - 60.4%, R - 95.6%
(d)
W - 64.4%, R - 79.4%
(e)
W - 78.7%, R - 87.2%
(f)
W - 69.1%, R - 52.9%

THANK YOU!