Image segmentation using CRF

Santosh K, Srikanth M

Overview

- Given an image, segment it into objects and label them.
- We follow the approach described in [2].

Figure 1: Sample input and segmented output images [2].

Approach

1. Pre-segment the image into patches at different scales.
2. Extract colour, texture, and SIFT features for each patch.
3. Compute global feature vector.
4. Use SVM to predict class of each patch.
5. Build a CRF model, where the patch labels are dependent on each other.
6. Apply threshold on the posteriors from the CRF to obtain final segmentation.

Step 1: Pre-segmentation

- Pre-segment the image using graph cut minimization formulation [1].
- Segmentation is done at different scales (finer to coarser) by adjusting the parameters of the algorithm.

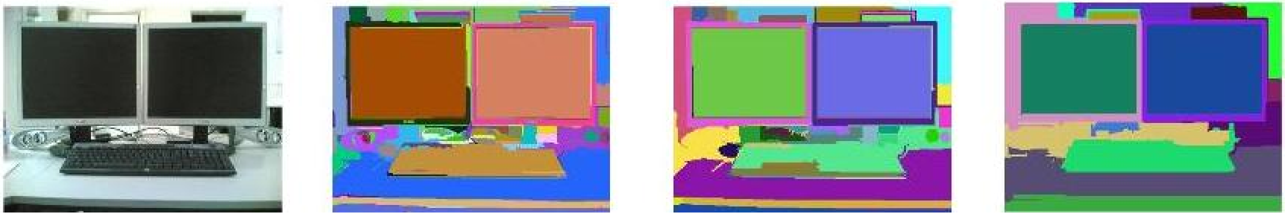

Figure 2: Example pre-segmented image on different scales.

Step 2: Feature extraction

1. Colour features are extracted by applying HSV transformation and computing histogram.
    1. Hue and saturation are binned to 10 x 10 2D histogram.
    2. Values are binned into 10-bin histogram.
    3. Results in 110 dimension feature vector for each patch.
2. Texture features using Gabor filters (we did not understand it yet).
3. Extracted SIFT key points and descriptors for each image and used k-means to cluster the descriptors into 1000 clusters.
    1. We compute histogram over each patch, using the VQ indices of the descriptors.
    2. Results in a 1000 dimension feature vector for each patch.

Next steps: Building CRF

1. Class label for each patch cannot be predicted independent of other patches.
2. Each patch is dependent on other:
   - Spatial relation in the image.
   - Overlap between patches in each scale.
3. The dependency between patches is represented with the help of a tree.
4. Child nodes represent patch in finer level and parent node represents the patch in coarser level.

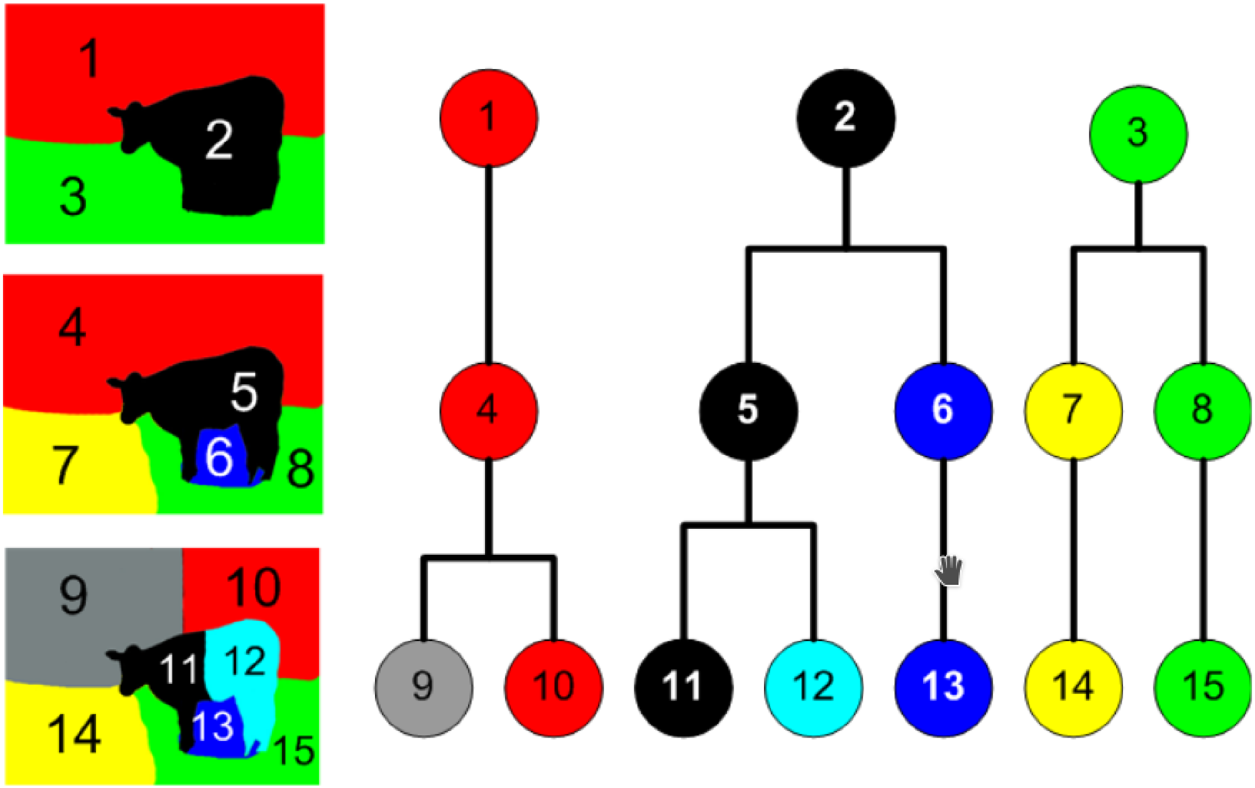

Figure 3: Example of a tree built from patches at different scales [3].

Building CRF

1. If $\mathbf{x}$ denotes the feature vector for a patch p and y_p denotes the corresponding label, then CRF for an image with P patches is represented (factorized as follows:)

$$P(\mathbf{y}, \mathbf{X}) \propto \prod_{p=1}^{P} \phi(y_p, \mathbf{x}) \prod_{q \in \pi(p)} \varphi(y_p, y_q) \tag{1}$$

   where $\pi(p)$ denotes the child nodes of patch p.

2. The potential function ϕ represents local evidence for labelling patch p to one the object classes k, and φ represents the coupling (similarity) between patches at two different levels (parent and child from the tree).
3. The goal is to learn the potential functions ϕ and φ, from the training data (images and corresponding labels with boundaries).

Learning potential functions: ϕ

1. $\phi(y_p, \mathbf{x})$ can be obtained from a classifier (support vector machines).
$$\phi(y_p, \mathbf{x}) \propto \phi(y_p \mid \mathbf{x}, \boldsymbol{\theta}) p(\mathbf{x} \mid \boldsymbol{\theta}) \tag{2}$$
where $\boldsymbol{\theta}$ are the parameters of the classifier.
2. SVM with χ^2 kernel is observed to be good for histogram features.
3. Further convert the output of the classifier into probabilities using softmax function as,
$$\phi(y_p = k \mid \mathbf{x}) = \frac{\exp\{a_l \,\mathrm{svm}(\mathbf{x}, \mathcal{C}_k) + b_l\}}{\sum_{\forall j \in K} \exp\{a_l \,\mathrm{svm}(\mathbf{x}, \mathcal{C}_j) + b_l\}} \tag{3}$$
where l is the scale level, $l = \{1, 2, 3\}$, and b_l is the scaling parameter of the χ^2 kernel.

Learning potentials: φ

1. The pairwise coupling or similarity between patches in two consecutive levels (i. e., for parent p at level l and child q at level $l+1$) is defined as,

$$\varphi(y_p, y_q) = \begin{pmatrix} e^{\gamma l} & e^{-\gamma l} \\ e^{-\gamma l} & e^{\gamma l} \end{pmatrix} \tag{4}$$

2. Here, inference is done using junction tree algorithm (yet to understand).

References

P. F. Felzenszwalb and D. P. Huttenlocher.
Efficient graph-based image segmentation.
Int. J. Comput. Vision, 59(2):167–181, Sept. 2004.

N. Plath, M. Toussaint, and S. Nakajima.
Multi-class image segmentation using conditional random fields and global classification.
In *Proceedings of the 26th Annual International Conference on Machine Learning*, ICML '09, pages 817–824, New York, USA, 2009. ACM.

J. Reynolds and K. Murphy.
Figure-ground segmentation using a hierarchical conditional random field.
In *Proceedings of the Fourth Canadian Conference on Computer and Robot Vision*, CRV '07, pages 175–182, Washington, DC, USA, 2007. IEEE Computer Society.