Parsing World's Skylines using Shape-constrained MRFs

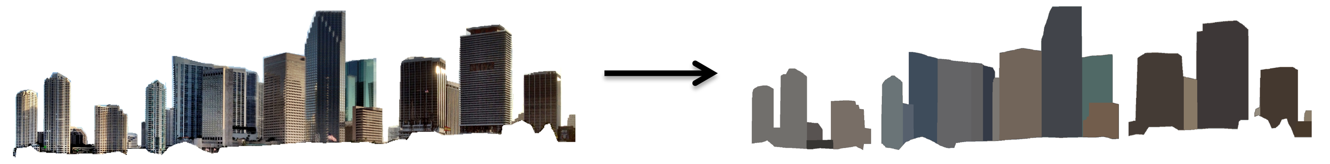
Rashmi Tonge[1], Subhransu Maji[2], C. V. Jawahar[1]

[1]IIIT, Hyderabad, India [2]TTI, Chicago

CVPR 2014

Problem and contributions

An approach for **segmenting** buildings in skyline images



We use priors on **topology** and **shape** of the buildings to develop a MRF solver that is **10x faster** and **more accurate** than a graph-cut based approach.

Energy minimization formulation (MRF)

For a set of pixels **P** and set of possible labels **L**, energy of labeling $F: P \rightarrow L$, is defined as :

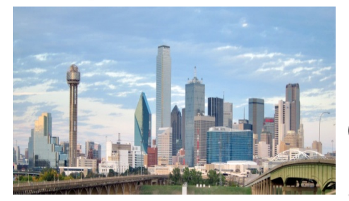
$$E(F) = \sum_{p \in P} D_p(F_p) + \sum_{p,q \in N} V_{pq}(F_p, F_q)$$

Where $V_{pq}(a,b) = \lambda exp\left(-\gamma (I_p - I_q)^2\right).1(a \neq b)$ and I_p denotes image intensity at pixel *p*. The optimal labeling can be obtained by, $F^* = argmin_f E(f)$

Skyline-12 dataset

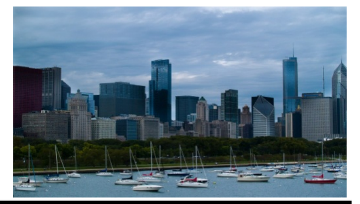
Consisting of **120** high resolution skyline images, with **10** images from each of the **12** cities - Chicago, Dallas, Frankfurt, Hong Kong, Miami, New York, Philadelphia, Seattle, Shanghai, Singapore, Tokyo and Toronto.

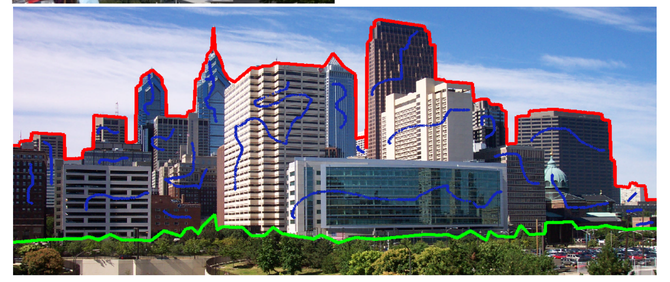
Code and dataset available at :
http://ttic.uchicago.edu/~smaji/projects/skylineParsing



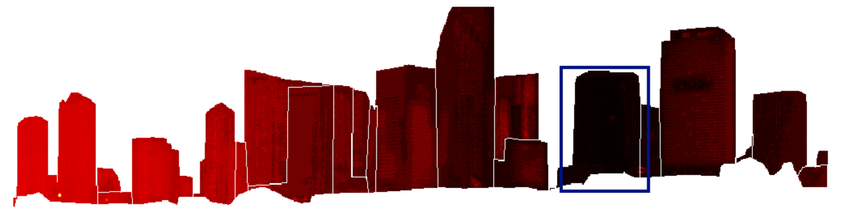

Images within each city are divided into *training, testing and validation* in 3:3:4 ratio.


Annotations of *upper boundary*, *lower* boundary and *seeds* are provided for each image, as seen below:

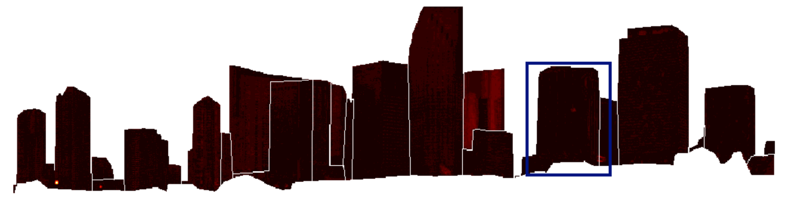



Region representation

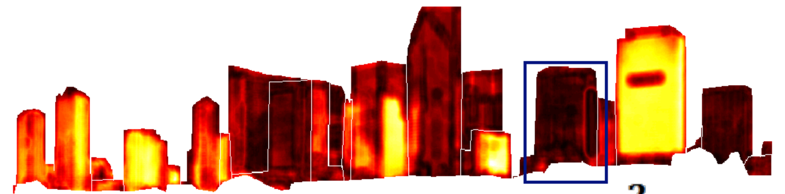


$$D_p(b) = \alpha\left(\beta \min_k C_p(b,k) + (1-\beta)\min_k T_p(b,k)\right) + (1-\alpha)S_p(b)$$

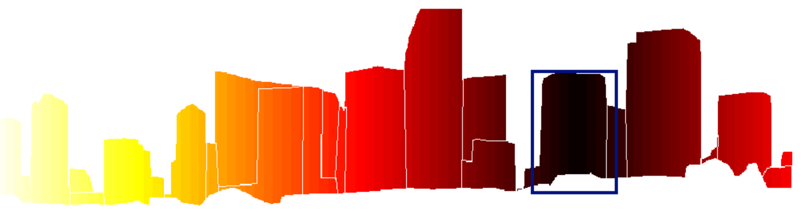
Unary term $D_p(b)$ for pixel ***p***, building ***b***.



Color $C_p(b,k)$ modeled with Gaussian Mixture Models representing color contribution at pixel ***p*** in the k^{th} cluster in building ***b***.



Texture $T_p(b,k)$ is modeled as χ^2 distance of the local histogram at pixel ***p*** from the mean histogram at k^{th} cluster center for building ***b***.

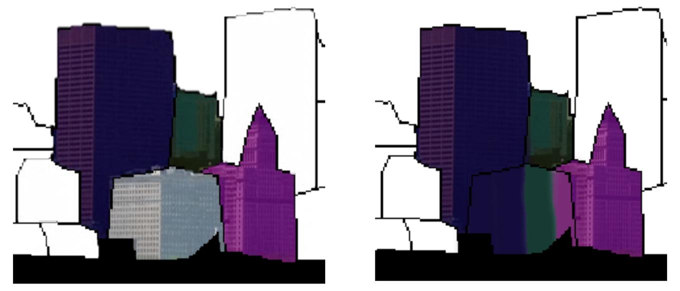


Spatial $S_p(b)$ modeled as horizontal distance of pixels ***p*** from center of the building ***b***.

Approach

Topological structure

Given label **α**, background beneath it can be obtained by copying labels *top* to *bottom*, allowing us to simultaneously expand and contract regions with label **α**, i.e. only the upper boundary needs to be estimated per building.



Foreground α and background

Input to the algorithm

image, upper/lower boundary, seeds

Method (1) Order the buildings based on the lowest seed pixel.
(2) Iteratively refine upper boundary.

Algorithm 1 Greedy skyline segmentation

Require: data D, pairwise V, boundary (l, u)

```
Initialize, initial labeling F from unary labels
for iter := 1 to K do
    Initialize, frontier f ← l
    for α := 1 to N do
        Ω_α ← upperBoundary(α, F, D, V, f, u)
        f ← max(f, Ω_α)
        F ← updateLabels(F, Ω_α)
    end for
end for
```

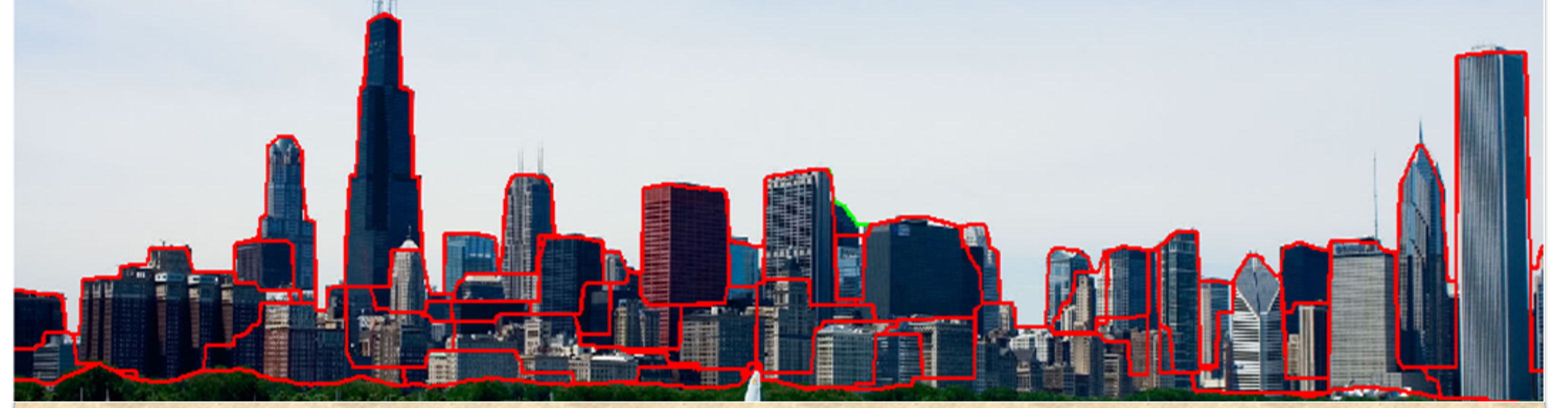
We propose three fast methods to obtain the upper boundary – **rectangle MRF, tiered MRF and refined MRF.** Each one solves a *binary segmentation problem*, as in **α-expansion**. Standard solver using graph-cuts requires $O(m^3n^3)$ time for a **mxn** image.

Rectangle MRF $O(mn^2)$

Upper boundary is a rectangle parameterized by *(left, right, top)*. Brute-force search for optimal rectangle can be done in $O(1)$ time per value using integral images, hence optimal can be found in $O(mn^2)$ time. Additionally can enforce width, height and aspect constraints.

Tiered MRF $O(m^2n)$

Upper boundary is 'x-monotonic', i.e. intersects each column only once. Optimal path can be computed using dynamic programming in $O(m^2n)$ time [1,2]. Fast, but cannot enforce shape priors efficiently.



Refined MRF $O(mn^2 + m^2d)$

Only *refine* the upper boundary *within* the width of the building. Constrains the overall segmentation and *improves* accuracy. Given a building of width **d** the refinement can be computed in $O(m^2d)$ time, for a total of $O(mn^2 + m^2d)$ per building.

References

1. P. F. Felzenszwalb and O.Veksler. *Tiered scene labeling with dynamic programming*, *CVPR*, 2010
2. Y. Zheng, S. Gu, and C. Tomasi. *Fast tiered labeling with topological priors*, *ECCV*, 2012
3. R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Su¨sstrunk. *SLIC Superpixels compared to state-of-the-art superpixel methods*, *IEEE PAMI*, 2012
4. P. F. Felzenszwalb and D. P. Huttenlocher. *Efficient graph-based image segmentation*, *IJCV*, 2004
5. P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik. *Contour detection and hierarchical image segmentation*. *IEEE PAMI*, 2011

Experimental evaluation

Evaluation metric

Let G_I and P_I denote ground truth and predicted labeling.
Let G_I^i and P_I^i denote the set of pixels labeled as building ***i***.
The average overlap for the image with ***N*** buildings is defined as :

$$AO(G_I, P_I) = \frac{1}{N}\sum_{i=1}^{N} \frac{G_I^i \cap P_I^i}{G_I^i \cup P_I^i}$$

We report mean average overlap (MAO) scores over the *test* set.

Interactive setting

Seeds are provided as input along with upper/lower boundary

Automatic setting

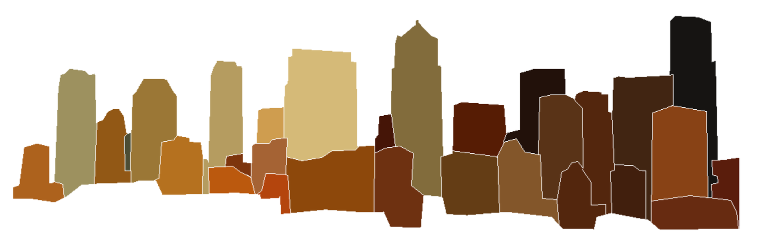
Only the upper/lower boundaries are provided, i.e., no seeds. We start with several automatic segmentation algorithms and *improve* over them using the shape and topological priors we proposed.

For evaluation we first perform a matching between the ground truth and segmented labels $m: N \rightarrow M$. Accuracy is measured under this matching:

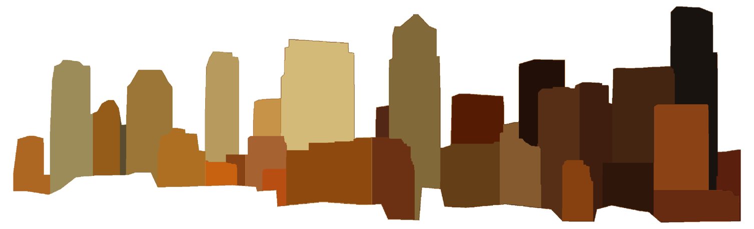
$$AO(G_I, P_I) = \max_{m \in M} \frac{1}{N}\sum_{i=1}^{N} \frac{G_I^i \cap P_I^{m(i)}}{G_I^i \cup P_I^{m(i)}}$$

Interactive Setting			Automatic Setting		
Method	MAO	Speed/img	SLIC[3]	Graph based[4]	gPb[5]
Standard MRF	62.3 %	69.5 s	24.56 %	20.17 %	26.35 %
Tiered MRF	59.4 %	7.5 s	27.22 %	25.86 %	31.51 %
Rectangle MRF	62.0 %	**5.5 s**	27.33 %	27.87 %	32.79 %
Refined MRF	**63.4 %**	9.2 s	27.30 %	27.42 %	33.13 %

Original Skyline

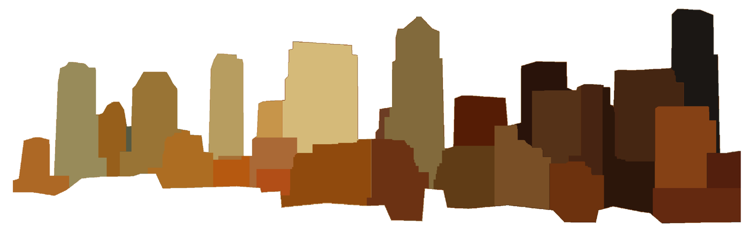


Ground Truth



Refined MRF

Rectangle MRF



Tiered MRF

Automatic Refined MRF