Superpixel Segmentation using Linear Spectral Clustering

Zhengqin Li[1], Jiansheng Chen[1]
[1]Department of Electronic Engineering, Tsinghua University.

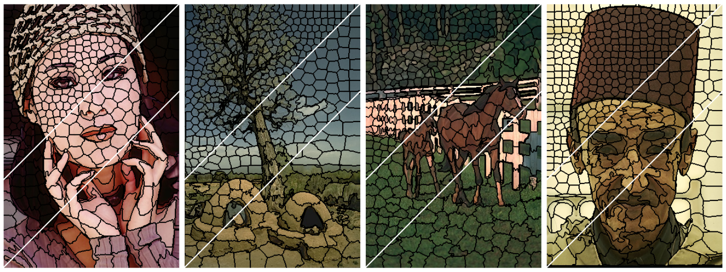

Figure 1: Images [4] segmented into 1000/500/200 superpixels using the proposed LSC algorithm.

Superpixel segmentation is an increasingly popular image preprocessing technique used in many computer vision applications. Many different superpixel segmentation algorithms have been proposed[5][2][3] and the following properties of superpixel segmentation are generally desirable. First, superpixels should adhere well to the natural image boundaries. Second, superpixel segmentation should be of low computational complexity. Last, global image information should be considered appropriately. Utilizing the perceptually important global clues to group pixels into semantically meaningful regions usually lead to substantial increases in complexity [5]. As a result, most practical superpixel segmentation algorithms are mainly based on the local image information only[3]. These methods may fail to correctly segment image regions with high intensity variability [2].

To address these issues, we present in this paper a superpixel segmentation algorithm called Linear Spectral Clustering (LSC). In LSC, we map each image pixel to a point in a ten dimensional feature space in which weighted K-means is applied for segmentation. Non-local information is implicitly preserved due to the equivalence between the weighted K-means clustering in this ten dimensional feature space and normalized cuts in the original pixel space. Simple weighted K-means clustering in the feature space can be used to optimize the segmentation cost function defined by normalized cuts. **Corollary 1** gives the sufficient conditions for the objective functions of weighted K-means and normalized cuts to share the same optimum point. A simplified version of proof is given in our paper.

Corollary 1 *Optimization of the objective functions of weighted K-means and normalized cuts are mathematically equivalent if both (1) and (2) hold.*

$$\forall\ p,q \in V,\ w(p)\phi(p)\cdot w(q)\phi(q) = W(p,q) \tag{1}$$

$$\forall\ p \in V,\ w(p) = \sum_{q\in V} W(p,q) \tag{2}$$

Among the two sufficient conditions, (2) can be easily fulfilled, while fulfilling (1) requires a careful selection of the similarity function W. Equation (1) can be rewritten as (3), in which W can be regarded as a kernel function. Therefore, W must satisfy the positivity condition [1]. Also, it must be separable to allow an explicit expression of the mapping function ϕ.

$$\phi(p)\cdot\phi(q) = \frac{W(p,q)}{w(p)w(q)} \tag{3}$$

To find a suitable form for $W(p,q)$, we investigate the widely used Euclidean distance based pixel similarity measurement. Given two pixels $p = (l_p,\alpha_p,\beta_p,x_p,y_p)$ and $q = (l_q,\alpha_q,\beta_q,x_q,y_q)$ in the combination of CIELAB and image plane space with normalized coordinates, a similarity measurement between them can be defined as (4), in which $\widehat{W}_c$ and $\widehat{W}_s$ are used to measure color similarity and space proximity respectively.

$$\begin{aligned}
\widehat{W}(p,q) &= C_c^2\cdot\widehat{W}_c(p,q) + C_s^2\cdot\widehat{W}_s(p,q)\\
\widehat{W}_c(p,q) &= 2.55^2\left[2-(\alpha_p-\alpha_q)^2-(\beta_p-\beta_q)^2\right] + \left[1-(l_p-l_q)^2\right]\\
\widehat{W}_s(p,q) &= \left[2-(x_p-x_q)^2-(y_p-y_q)^2\right]
\end{aligned} \tag{4}$$

Although $\widehat{W}(p,q)$ has very clear physical meaning in measuring pixel similarity, it cannot be directly used in our method because it does not satisfy the positivity condition [1] required by (3). To solve this problem, we try to find a proper approximation of $\widehat{W}(p,q)$ using Fourier transform.

$$\begin{aligned}
\widehat{W}(p,q) &= C_s^2\left[g(x_p-x_q)+g(y_p-y_q)\right] + C_c^2\left[g(l_p\right.\\
&\quad \left. -l_q) + 2.55^2\left(g(\alpha_p-\alpha_q)+g(\beta_p-\beta_q)\right)\right]\\
g(t) &= 1-t^2 \approx \frac{32}{\pi}\cos\frac{\pi}{2}t,\quad t\in[-1,1]
\end{aligned} \tag{5}$$

$\widehat{W}(p,q)$ is a nonnegative linear combination of a number of $g(t)$ which can be expanded as a uniformly convergent Fourier series. The coefficients of this series converge to 0 very quickly at a speed of $(2k+1)^3$. Therefore, $g(t)$ can be well approximated by the first term in the series as is expressed in (5). Omitting the constant multiplier $32/\pi$, $\widehat{W}(p,q)$ can be approximated by $W(p,q)$ defined in (6). According to the properties of cosine function, $W(p,q)$ is positive definite and can be directly written in the inner product form shown in (1), in which ϕ and w are defined in (7).

$$\begin{aligned}
W(p,q) &= C_s^2\left[\cos\frac{\pi}{2}(x_p-x_q)+\cos\frac{\pi}{2}(y_p-y_q)\right] + C_c^2\left[\cos\frac{\pi}{2}(l_p-l_q)+\right.\\
&\quad \left. 2.55^2\left(\cos\frac{\pi}{2}(\alpha_p-\alpha_q)+\cos\frac{\pi}{2}(\beta_p-\beta_q)\right)\right]
\end{aligned} \tag{6}$$

$$\begin{aligned}
\phi(p) &= \frac{1}{w(p)}(C_c\cos\frac{\pi}{2}l_p, C_c\sin\frac{\pi}{2}l_p, 2.55C_c\cos\frac{\pi}{2}\alpha_p,\\
&\quad 2.55C_c\sin\frac{\pi}{2}\alpha_p, 2.55C_c\cos\frac{\pi}{2}\beta_p, 2.55C_c\sin\frac{\pi}{2}\beta_p,\\
&\quad C_s\cos\frac{\pi}{2}x_p, C_s\sin\frac{\pi}{2}x_p, C_s\cos\frac{\pi}{2}y_p, C_s\sin\frac{\pi}{2}y_p)\\
w(p) &= \sum_{q\in V} W(p,q) = w(p)\phi(p)\cdot\sum_{q\in V} w(q)\phi(q)
\end{aligned} \tag{7}$$

Until now, we have explicitly define a ten dimensional feature space in (7) so that weighted K-means clustering in the feature space is equivalent to normalized cuts in the input space. We perform superpixel segmentation by applying weighted K-means in the feature space. Pixels assigned to the same cluster form a superpixel. Experiments show that LSC performs equally well or better than state of the art superpixel segmentation methods in terms of several commonly used evaluation metrics in image segmentation.

[1] N. Cristianini and J.S. Taylor. *An introduction to support vector machines and other kernel-based learning methods*. Cambridge University Press New York, NY, USA, 2000.

[2] P.F. Felzenszwalb and D.P. Huttenlocher. Efficient graph-based image segmentation. *International Journal of Computer Vision*, 59(2):167–181, 2004.

[3] A. Levinshtein, A. Stere, K.N. Kutulakos, D.J. Fleet, S.J. Dickinson, and K. Siddiqi. Turbopixel: fast supepixels using geometric flow. *IEEE Trans. on PAMI*, 31(12):2209–2297, 2009.

[4] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. *Proc. of ICCV*, 2:416–423, 2001.

[5] X. Ren and J. Malik. Learning a classification model for segmentation. *Proc. of ICCV*, 1:10–17, 2003.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.