Supplementary material of "Superpixel Segmentation using Linear Spectral Clustering"

APPENDIX A
DETAILED PROOF OF THEOREM 1

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

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

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

Proof: This objective function of weighted k-means can be rewritten in the following form.

$$\begin{aligned}
F_{k-m} =& \sum_{k=1}^{K}\sum_{\boldsymbol{p}\in\pi_k} w(\boldsymbol{p})\|\phi(\boldsymbol{p}) - \frac{\sum_{\boldsymbol{q}\in\pi_k} w(\boldsymbol{q})\phi(\boldsymbol{q})}{\sum_{\boldsymbol{q}\in\pi_k} w(\boldsymbol{q})}\|^2 \\
=& \sum_{k=1}^{K}\sum_{\boldsymbol{p}\in\pi_k} w(\boldsymbol{p}) \left(\|\phi(\boldsymbol{p})\|^2 - 2\phi(\boldsymbol{p}) \cdot \frac{\sum_{\boldsymbol{q}\in\pi_k} w(\boldsymbol{q})\phi(\boldsymbol{q})}{\sum_{\boldsymbol{q}\in\pi_k} w(\boldsymbol{q})} + \|\frac{\sum_{\boldsymbol{q}\in\pi_k} w(\boldsymbol{q})\phi(\boldsymbol{q})}{\sum_{\boldsymbol{q}\in\pi_q} w(\boldsymbol{q})}\|^2 \right) \\
=& \sum_{k=1}^{K}\sum_{\boldsymbol{p}\in\pi_k} w(\boldsymbol{p})\|\phi(\boldsymbol{p})\|^2 - \sum_{k=1}^{K}\left(2\sum_{\boldsymbol{p}\in\pi_k} w(\boldsymbol{p})\phi(\boldsymbol{p}) \cdot \frac{\sum_{\boldsymbol{q}\in\pi_k} w(\boldsymbol{q})\phi(\boldsymbol{q})}{\sum_{\boldsymbol{q}\in\pi_k} w(\boldsymbol{q})} - \sum_{\boldsymbol{p}\in\pi_k} w(\boldsymbol{p})\|\frac{\sum_{\boldsymbol{q}\in\pi_k} w(\boldsymbol{q})\phi(\boldsymbol{q})}{\sum_{\boldsymbol{q}\in\pi_k} w(\boldsymbol{q})}\|^2 \right) \\
=& \sum_{k=1}^{K}\sum_{\boldsymbol{p}\in\pi_k} w(\boldsymbol{p})\|\phi(\boldsymbol{p})\|^2 - \sum_{k=1}^{K}\left(2\frac{\|\sum_{\boldsymbol{p}\in\pi_k} w(\boldsymbol{p})\phi(\boldsymbol{p})\|^p}{\sum_{\boldsymbol{p}\in\pi_k} w(\boldsymbol{p})} - \frac{\|\sum_{\boldsymbol{p}\in\pi_k} w(\boldsymbol{p})\phi(\boldsymbol{p})\|^2}{\sum_{\boldsymbol{p}\in\pi_k} w(\boldsymbol{p})} \right) \\
=& \sum_{k=1}^{K}\sum_{\boldsymbol{p}\in\pi_k} w(\boldsymbol{p})\|\phi(\boldsymbol{p})\|^2 - \sum_{k=1}^{K} \frac{\|\sum_{\boldsymbol{p}\in\pi_k} w(\boldsymbol{p})\phi(\boldsymbol{p})\|^2}{\sum_{\boldsymbol{p}\in\pi_k} w(\boldsymbol{p})} \\
=& \sum_{k=1}^{K}\sum_{\boldsymbol{p}\in\pi_k} w(\boldsymbol{p})\|\phi(\boldsymbol{p})\|^2 - \sum_{k=1}^{K} \frac{\sum_{\boldsymbol{p}\in\pi_k}\sum_{\boldsymbol{q}\in\pi_k} w(\boldsymbol{p})\phi(\boldsymbol{p}) \cdot w(\boldsymbol{q})\phi(\boldsymbol{q})}{\sum_{\boldsymbol{p}\in\pi_k} w(\boldsymbol{p})}
\end{aligned} \tag{3}$$

In equation (3), $\boldsymbol{C} = \sum_{k=1}^{K}\sum_{\boldsymbol{p}\in\pi_k} w(\boldsymbol{p})\|\phi(\boldsymbol{p})\|^2$ is a constant independent of the clustering result. Given equations (1) and (2), we have equation (4), from which it can be easily observed that minimizing F_{k-m} is strictly equivalent to maximizing F_{Ncuts}.

$$\begin{aligned}
F_{k-m} &= \boldsymbol{C} - \sum_{k=1}^{K} \frac{\sum_{\boldsymbol{p}\in\pi_k}\sum_{\boldsymbol{q}\in\pi_k} \boldsymbol{W(p,q)}}{\sum_{\boldsymbol{p}\in\pi_k}\sum_{\boldsymbol{q}\in\boldsymbol{V}} \boldsymbol{W(p,q)}} \\
&= \boldsymbol{C} - K \times F_{Ncuts}
\end{aligned} \tag{4}$$

In conclusion, we have proved that by choosing the weighted inner products in the feature space to be equal to the similarity measurement in the input space and by choosing the weight of each point equal to the total weight of edges connecting the corresponding node to the all nodes in the graph, the objective function of weighted k-means and normalized cut are mathematically equivalent.

■

APPENDIX B
KERNEL FUNCTION AND POSITIVITY CONDITION

In this section, we will briefly explain why it is necessary to approximate the similarity function $\widehat{\boldsymbol{W}}(\boldsymbol{p},\boldsymbol{q})$ drawn from Euclidean distance using fourier transform by showing that $\widehat{\boldsymbol{W}}(\boldsymbol{p},\boldsymbol{q})$ is not a kernel function. Equation (8) in our paper demonstrates that the similarity function $\boldsymbol{W}(\boldsymbol{p},\boldsymbol{q})$ must be a kernel function which satisfies the positivity condition. Detailed explanation can be found in [1]. *Theorem* 2 gives the sufficient and necessary condition for a symmetric function to be a kernel function [1].

Theorem 2: Let $\mathcal{X}$ be a finite input space with $\boldsymbol{K}(\boldsymbol{x},\boldsymbol{z})$ a symmetric function on $\mathcal{X}$. Then $\boldsymbol{K}(\boldsymbol{x},\boldsymbol{z})$ is a kernel function if and only if the matrix

$$\mathcal{K} = (K(\boldsymbol{x}_i,\boldsymbol{z}_j))_{i,j=1}^{n} \tag{5}$$

is positive semi-definite.

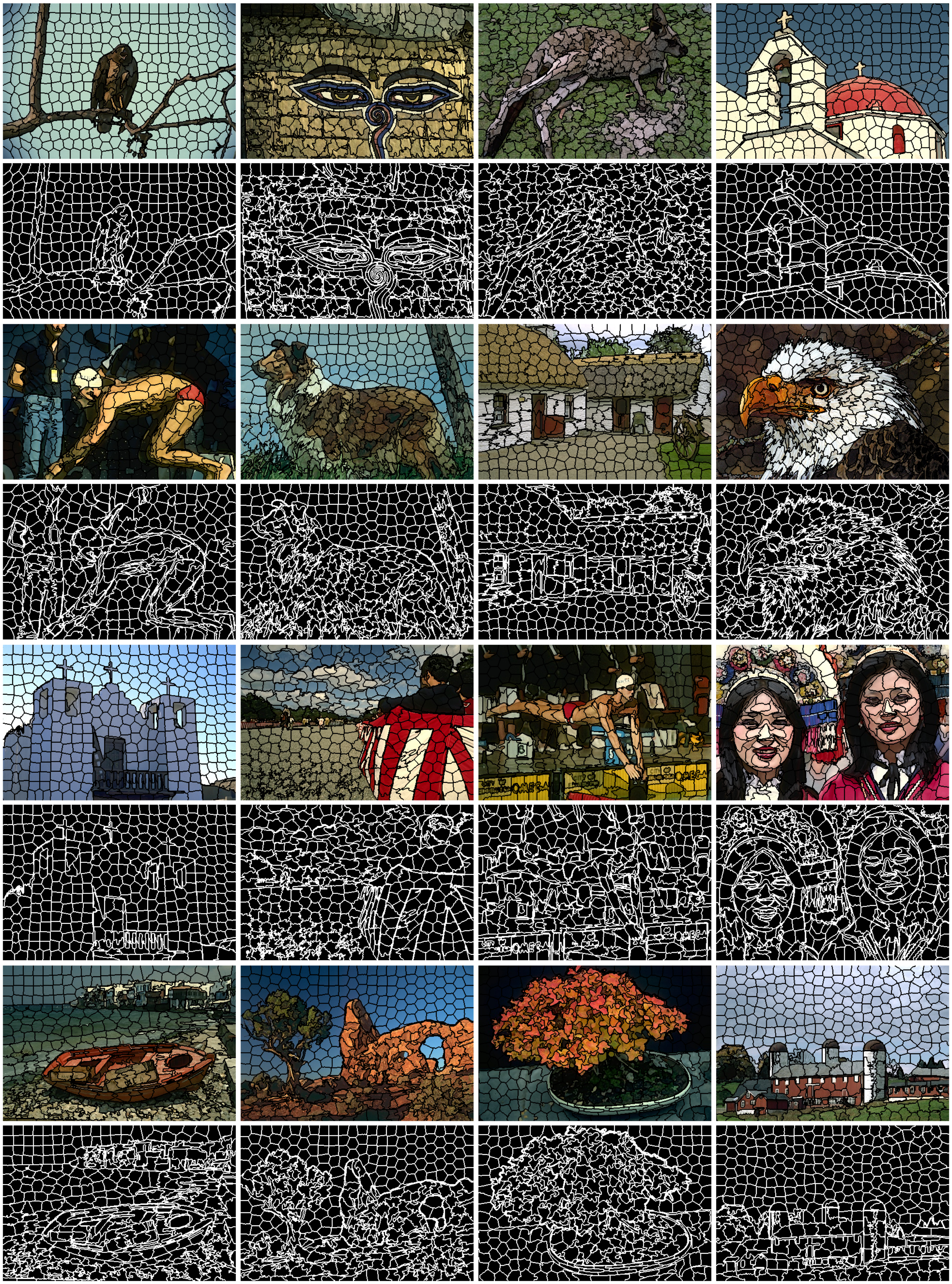
Since $\widehat{\boldsymbol{W}}(\boldsymbol{p},\boldsymbol{q})$ can be rewriten as a nonnegative linear combination of a number of instances of a simple functions $\boldsymbol{g}(\boldsymbol{p},\boldsymbol{q}) = \boldsymbol{1} - (\boldsymbol{p}-\boldsymbol{q})^{\boldsymbol{2}}$, to verify that $\widehat{\boldsymbol{W}}(\boldsymbol{p},\boldsymbol{q})$ is not a kernel function, we need only to prove that function $\boldsymbol{g}(\boldsymbol{p},\boldsymbol{q})$ violates the condition proposed in *Theorem* 2. We prove that $\boldsymbol{g}(\boldsymbol{p},\boldsymbol{q})$ is not a kernel function using a specific example. Given vector $\boldsymbol{X} = (0, 0.5, 1)$, the corresponding kernel matrix $\mathcal{K} = (\boldsymbol{g}(\boldsymbol{X}_i,\boldsymbol{X}_j))_{i,j=1}^{n}$ defined by function $\boldsymbol{g}(\boldsymbol{p},\boldsymbol{q})$ and vector $\boldsymbol{X}$ is shown in equation (6)

$$\mathcal{K} = \begin{pmatrix} 1 & \frac{3}{4} & 0 \\ \frac{3}{4} & 1 & \frac{3}{4} \\ 0 & \frac{3}{4} & 1 \end{pmatrix} = \frac{1}{4} \begin{pmatrix} 1 & -\sqrt{2} & 1 \\ -\sqrt{2} & 0 & \sqrt{2} \\ 1 & \sqrt{2} & 1 \end{pmatrix} \begin{pmatrix} 1-\frac{3\sqrt{2}}{4} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1+\frac{3\sqrt{2}}{4} \end{pmatrix} \begin{pmatrix} 1 & -\sqrt{2} & 1 \\ -\sqrt{2} & 0 & \sqrt{2} \\ 1 & \sqrt{2} & 1 \end{pmatrix} \tag{6}$$

The negative eigenvalue $1-\frac{3\sqrt{2}}{4}$ indicates that $\mathcal{K}$ is not positive semi-definite and therefore $\boldsymbol{g}(\boldsymbol{p},\boldsymbol{q})$ is not a kernel function. As such, $\widehat{\boldsymbol{W}}(\boldsymbol{p},\boldsymbol{q})$ can not be used directly as the similarity function in LSC. To solve this problem, we use fourier transform to approximate $\boldsymbol{g}(\boldsymbol{p},\boldsymbol{q})$ in our method. Fig.1 shows more superpixel segmentation examples using the proposed LSC algorithm.

REFERENCES

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



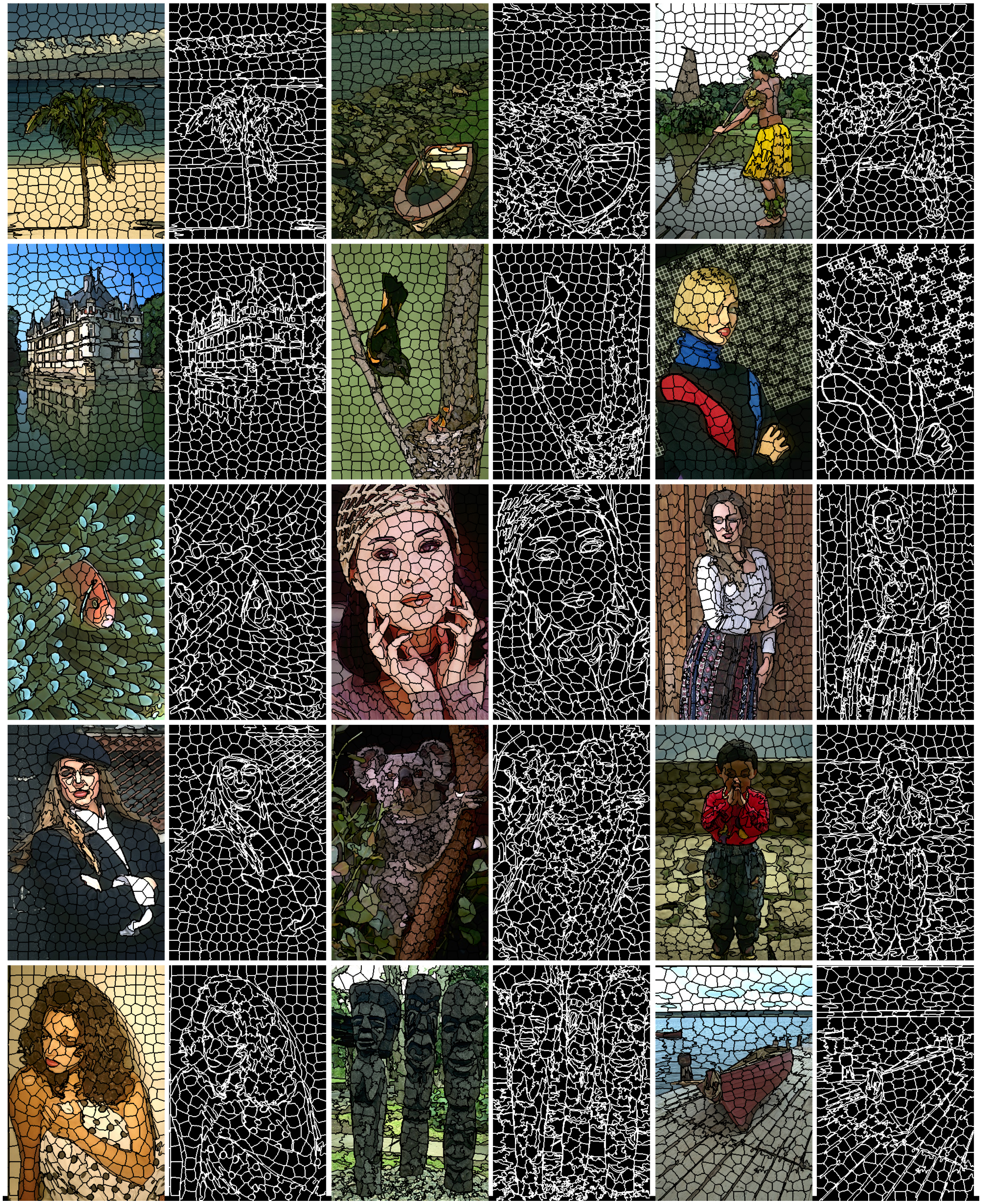


Fig. 1. Images segmented into 400 superpixels using the proposed LSC algorithm.