# Robust Path-Based Spectral Clustering with Application to Image Segmentation

Hong Chang & Dit-Yan Yeung
Hong Kong University of Science and Technology
Clear Water Bay, Kowloon, Hong Kong
{hongch,dyyeung}@cs.ust.hk

## Abstract

*Spectral clustering and path-based clustering are two recently developed clustering approaches that have delivered impressive results in a number of challenging clustering tasks. However, they are not robust enough against noise and outliers in the data. In this paper, based on M-estimation from robust statistics, we develop a robust path-based spectral clustering method by defining a robust path-based similarity measure for spectral clustering. Our method is significantly more robust than spectral clustering and path-based clustering. We have performed experiments based on both synthetic and real-world data, comparing our method with some other methods. In particular, color images from the Berkeley Segmentation Dataset and Benchmark are used in the image segmentation experiments. Experimental results show that our method consistently outperforms other methods due to its higher robustness.*

## 1. Introduction

Clustering has been among the most active research topics in machine learning and pattern recognition. While many traditional clustering algorithms have been developed over the past few decades [2, 10], some new clustering algorithms emerged over the last few years give very promising results on some challenging tasks. Among them are *spectral clustering* [12, 14, 15, 16] and *path-based clustering* [4, 5, 6], which have demonstrated excellent performance on some clustering tasks involving highly nonlinear and elongated clusters in addition to compact clusters.

Despite the promising performance demonstrated by these algorithms, there exist situations when they do not perform very well. The poor clustering result is due mainly to the choice of the affinity matrix, which is usually defined in a way similar to the Gaussian kernel based on inter-point Euclidean distance in the input space. While the combined use of path-based clustering and spectral clustering, referred to as *path-based spectral clustering* here, seems to be very

effective, we will show later in the paper that this combined method, like the separate use of spectral clustering or path-based clustering, is not robust enough against noise and outliers which commonly exist in real-world data.

In this paper, based on robust statistical techniques [9], we propose a novel scheme to make path-based (spectral) clustering more robust. Our work is built upon the recent work of Fischer et al. [4, 5, 6]. We devise an M-estimator and use it to define a robust path-based similarity measure which takes into account the existence of noise and outliers in the data and hence brings about robustness in the method.

The rest of this paper is organized as follows. Some related work is briefly reviewed in Section 2. In Section 3, we propose a robust path-based similarity measure based on robust statistics, with which a robust path-based spectral clustering algorithm can be devised. Experimental results on synthetic data as well as color image segmentation are presented in Sections 4 and 5, respectively, comparing our method with non-robust methods. Finally, some concluding remarks are given in the last section.

## 2. Related Work

Some related clustering methods have been proposed in the literature. Besides spectral clustering and path-based clustering, there also exist some other methods that are capable of finding elongated structures. Hierarchical agglomerative clustering based on the single linkage algorithm [10] is efficient and has been widely used, but it is well known that this method is highly sensitive to outliers. Mean shift clustering [1] and spectral clustering have shown good performance in some clustering tasks. However, both of them are sensitive to the preselection of the kernel bandwidth parameter.

Fischer and Buhmann have a more recent paper which uses bagging (or *b*ootstrap *agg*regat*ing*) with path-based clustering to address the robustness issue [3]. Like other bagging applications, data resampling is used to reduce data variance so that the results are less sensitive to noise. However, since many bootstrap samples are involved, efficiency

has to be sacrificed in exchange for effectiveness of the method. Moreover, some bootstrap samples of the input data may still contain outliers which can affect the clustering results.

## 3. Robust Path-Based Clustering

### 3.1. Path-Based Similarity Measure

The path-based dissimilarity measure was originally proposed in [6]. Following their formulation, we define a closely related path-based similarity measure which will be extended to a robust version in the next two subsections.

We denote the data set of $n$ points by $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$. The data points can be represented as a fully connected graph with $n$ vertices corresponding to the $n$ points. Each edge $(i, j)$ in the graph is assigned a weight $s'_{ij}$ reflecting the original similarity between $\mathbf{x}_i$ and $\mathbf{x}_j$ analogous to the Gaussian kernel:

$$s'_{ij} = \begin{cases} \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right) & \text{for } i \neq j \\ 0 & \text{for } i = j, \end{cases} \quad (1)$$

except that $s'_{ii}$ is equal to 0 rather than 1 for $i = 1, 2, \ldots, n$. Here, the scaling parameter $\sigma$ controls how fast $s'_{ij}$ falls off with the Euclidean distance between $\mathbf{x}_i$ and $\mathbf{x}_j$. Recall that this is actually one common way of defining the affine matrix in spectral clustering [12], where $\sigma$ is a somewhat sensitive parameter. While this parameter is usually pre-specified, Ng et al. [12] proposed a method for choosing $\sigma$ automatically.

Let $\mathcal{P}_{ij}$ denote the set of all paths from vertex $i$ to vertex $j$ through the graph. For each path $p \in \mathcal{P}_{ij}$, the *effective similarity* $s^p_{ij}$ between vertices $i$ and $j$ (or the corresponding data points $\mathbf{x}_i$ and $\mathbf{x}_j$) is the minimum edge weight along the path. We define the total similarity $s_{ij}$ between vertices $i$ and $j$ as the maximum among all path-based effective similarities $s^p_{ij}$'s for paths $p$'s in $\mathcal{P}_{ij}$:

$$s_{ij} = \max_{p \in \mathcal{P}_{ij}} \left\{ \min_{1 \leq h < |p|} s'_{p[h]p[h+1]} \right\},$$

where $p[h]$ denotes the $h$th vertex along the path $p$ from vertex $i$ to vertex $j$ and $|p|$ denotes the number of vertices that $p$ goes through.

From the definition, the similarity between points $\mathbf{x}_i$ and $\mathbf{x}_j$ should be large when the two points belong to the same cluster, and small when they belong to different clusters. However, in the latter case, if there exist some outliers between the two clusters, the similarity between points residing in different clusters may become much larger than it should be. In other words, the path-based similarity measure is very sensitive to noise and outliers.

### 3.2. Robust Estimator Based on M-Estimation

In this subsection, we devise a robust estimator based on the concept of M-estimation in robust statistics [9].

Let us regard the neighbors of a data point $\mathbf{x}_i$ as realizations from an estimator of $\mathbf{x}_i$. Then the average squared distance of the neighbors from $\mathbf{x}_i$ can be seen as the squared residual error of the estimator on $\mathbf{x}_i$. Thus the total squared error $E$ is given by

$$\begin{aligned} E &= \sum_{i=1}^{n} \left( \frac{1}{|\mathcal{N}_i|} \sum_{\mathbf{x}_j \in \mathcal{N}_i} \|\mathbf{x}_i - \mathbf{x}_j\|^2 \right) \\ &= \sum_{i=1}^{n} \sum_{\mathbf{x}_j \in \mathcal{N}_i} \left( \frac{\|\mathbf{x}_i - \mathbf{x}_j\|}{\sqrt{|\mathcal{N}_i|}} \right)^2, \end{aligned}$$

where $\mathcal{N}_i$ denotes the neighborhood of $\mathbf{x}_i$ and $|\mathcal{N}_i|$ denotes the number of neighbors it contains.

In order to make use of robust estimation techniques, we express the optimization problem as minimizing the total weighted squared error

$$E_w = \sum_{i=1}^{n} \sum_{\mathbf{x}_j \in \mathcal{N}_i} a_{ij} \left( \frac{\|\mathbf{x}_i - \mathbf{x}_j\|}{\sqrt{|\mathcal{N}_i|}} \right)^2.$$

Following the ideas of Huber, the least squares estimator is replaced by a robust estimator that minimizes

$$E_\rho = \sum_{i=1}^{n} \sum_{\mathbf{x}_j \in \mathcal{N}_i} \rho \left( \frac{\|\mathbf{x}_i - \mathbf{x}_j\|}{\sqrt{|\mathcal{N}_i|}} \right) = \sum_{i=1}^{n} \sum_{\mathbf{x}_j \in \mathcal{N}_i} \rho(e_{ij}),$$

where $\rho(\cdot)$ is some convex function and $e_{ij}$ is an error residue. Using the Welsch function

$$\rho(e_{ij}) = \frac{c^2}{2} \left[ 1 - \exp(-(e_{ij}/c)^2) \right],$$

for some parameter $c > 0$, the weight function can be defined as $a(e_{ij}) = \frac{\rho'(e_{ij})}{e_{ij}} = \exp(-(e_{ij}/c)^2)$. Therefore, in our problem, the weight $a_{ij}$ for the $j$th neighbor of $\mathbf{x}_i$ is

$$a_{ij} = \exp\left( -\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{|\mathcal{N}_i| c^2} \right).$$

### 3.3. Robust Path-Based Similarity Measure

By summing up the weights $a_{ij}$'s of all neighbors $\mathbf{x}_j \in \mathcal{N}_i$ and setting $c = \sqrt{\frac{2}{|\mathcal{N}_i|}} \sigma$, we obtain a weight $w'_i$ for each point $\mathbf{x}_i$ which can be expressed solely based on the original similarity values:

$$w'_i = \sum_{\mathbf{x}_j \in \mathcal{N}_i} s'_{ij}.$$

The weight of a data point is large if many data points are in its vicinity, and small if very few other points are close to it. Neighborhood can be defined in a number of different ways. Although some methods may be better than others, we keep it simple in this paper by using the same neighborhood size for all points in a data set, so that the neighborhood is just large enough to include at least two neighbors in each neighborhood. Note that the justification for using local neighborhoods is based on the argument that the weights should depend on the local geometry rather than the global structure of the whole data set.

To make the weights insensitive to $\sigma$, we normalize each weight as $w_i = w_i'/\max_{\mathbf{x}_i \in \mathcal{X}} w_i'$, so that all weights fall within the range $(0, 1]$. While a large value for $w_i$ indicates that $\mathbf{x}_i$ is likely to be inside a compact cluster, a small value indicates that $\mathbf{x}_i$ is an outlier.

Therefore, the robust path-based similarity measure can be expressed as:

$$s_{ij} = \max_{p \in \mathcal{P}_{ij}} \left\{ \min_{1 \le h < |p|} w_{p[h]} w_{p[h+1]} s_{p[h]p[h+1]}' \right\}. \quad (2)$$

This measure can reflect the genuine similarity between $\mathbf{x}_i$ and $\mathbf{x}_j$ even when outliers exist. If there exists a path from $\mathbf{x}_i$ to $\mathbf{x}_j$ going through only points with high values of $w_{p[h]}$'s and $s_{p[h]p[h+1]}'$'s, then the total similarity should be high and hence $\mathbf{x}_i$ and $\mathbf{x}_j$ are likely to belong to the same cluster. On the other hand, if all paths between $\mathbf{x}_i$ and $\mathbf{x}_j$ contain at least one low value of either $w_{p[h]}$ or $s_{p[h]p[h+1]}'$, then the total similarity should be relatively low, implying that $\mathbf{x}_i$ and $\mathbf{x}_j$ either belong to different clusters or are themselves outliers.

We define the robust similarity matrix $\mathbf{S} = [s_{ij}]$ and use it in place of the affinity matrix $\mathbf{A} = [s_{ij}']$ commonly used in spectral clustering [12]. Besides its robustness property, $\mathbf{S}$ is no longer sensitive to the parameter $\sigma$ as pointed out above and to be illustrated further through sensitivity analysis in Section 4.2. Our robust path-based spectral clustering algorithm is simply the ordinary spectral clustering algorithm using the robust similarity measure described above to define the affinity matrix. This seemingly minor modification is in fact very significant as dramatically improved clustering results can be obtained.

## 4. Experiments on Synthetic Data

To assess the efficacy of our algorithm for clustering tasks, we first perform some experiments on synthetic data.

### 4.1. Data Sets and Results

Although both standard spectral clustering and path-based clustering can find the two clusters in the 2-circle data set as shown in Figure 1(a), they are no longer robust enough to give satisfactory results if some noise points are added. Figure 1(b)–(d) compare these two methods with our robust path-based spectral clustering algorithm when 30 noise points are added. Note that some noise points located between the two circles end up connecting the two circular clusters. Figure 1(b) shows that spectral clustering cannot give good result for this data set.[1] Due to the existence of noise points, the dissimilarity measure used by the original path-based clustering algorithm gives much lower dissimilarity values than they should be to point pairs residing in different circles. As a consequence, its result as shown in Figure 1(c) is not satisfactory either. Our method gives very satisfactory result, as shown in Figure 1(d), that agrees well with human judgement. This shows that the robust similarity measure is very effective in reducing the influence of the outliers.

We also conduct some experiments on a 3-cluster data set. As shown in Figure 2(a), the data set consists of a circular cluster with an opening near the bottom and two Gaussian distributed clusters inside. Each cluster contains 100 data points. This data set is similar to the one shown in Figure 1(b) of [12], except that ours is more difficult. Unlike the previous example, we do not add artificial noise points to this data set. However, the Gaussian clusters can be seen as having Gaussian noise points which tend to connect the clusters together. Figure 2(b) and (c) show that both spectral clustering and path-based clustering cannot find the three clusters correctly. Again, our robust path-based spectral clustering algorithm gives much more satisfactory result, as shown in Figure 2(d). However, some inter-cluster points are mis-clustered. Our method is successful in assigning lower weights to these points (and hence essentially detecting them as outliers). As a result, they have relatively lower similarity values to other points within a cluster than those points closer to the Gaussian centers or on the incomplete circle. Spectral clustering will cluster such points incorrectly, as can also be seen in the 2-circle data set in Figure 1. If the robust path-based similarity measure is used with $k$-means clustering rather than spectral clustering, this problem actually does not happen and gives better result. However, in general, spectral clustering gives better results than $k$-means and hence is generally a better choice.

We further conduct experiments on a synthetic data set that can be seen as a simplified version of the image segmentation problem to be studied in the next section. Figure 3(a) shows a square image consisting of four regions, contaminated with some random noise points shown as small color squares of four different hues. Each small square corresponds to a point in the RGB color space, as shown in Figure 3(b). The clustering results based on different methods are shown in Figure 3(c)–(e). This again

---

[1]The noise level of this data set is higher than that of a similar data set in Figure 1(g) of [12].
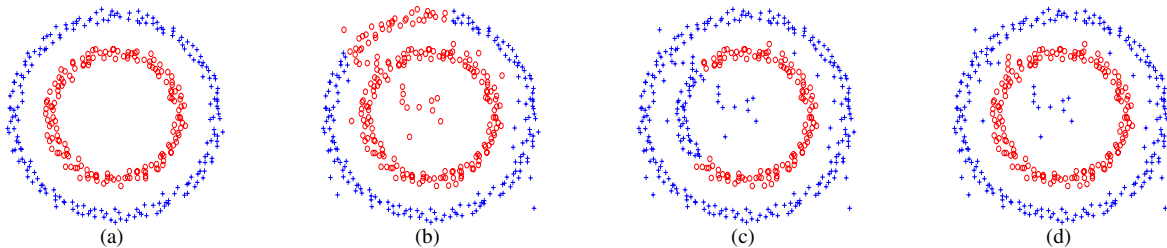
**Figure 1. Clustering results for a noisy 2-circle data set: (a) original data set before noise points are added; (b) spectral clustering result; (c) path-based spectral clustering result; and (d) robust path-based spectral clustering result.**
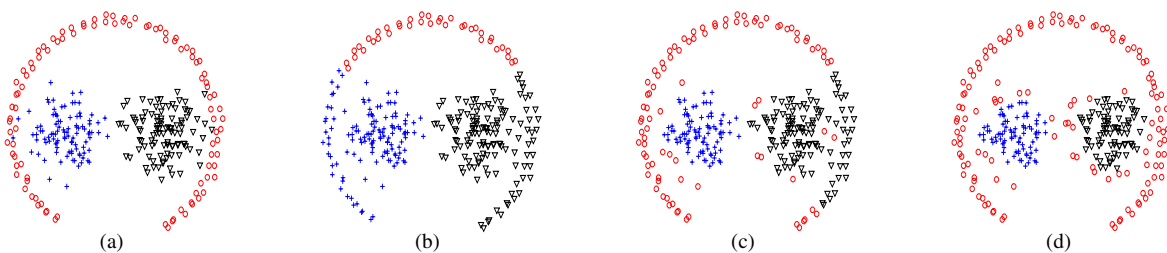


**Figure 2. Clustering results for a 3-cluster data set: (a) original data set; (b) spectral clustering result; (c) path-based spectral clustering result; and (d) robust path-based spectral clustering result.**

shows that our method is superior to other methods.

## 4.2. Sensitivity Analysis of Parameters

Similar to standard spectral clustering, our method also requires choosing the scaling parameter $\sigma$. We argued above that our method is not sensitive to the setting of $\sigma$ while the standard spectral clustering method is. In this subsection, we elaborate this point by performing sensitivity analysis to see how the clustering result varies with $\sigma$.

For simplicity, we only report results for the 2-circle data set, which are representative of the general case. Figure 4(a)–(c) show three 2-circle data sets of varying noise levels, with no noise in (a) (same as that in Figure 1(a)), 20 noise points in (b) and 30 noise points in (c) (same as that in Figure 1(b)–(d)). We compare robust path-based spectral clustering with standard spectral clustering and path-based spectral clustering on a range of different $\sigma$ values. The clustering quality is measured using the Rand index [13] with respect to the ground truth clustering partition.

We can see that the standard spectral clustering algorithm is fairly sensitive to the choice of $\sigma$. With $\sigma^2 \in [0.018, 0.096]$, this method can correctly find the two clusters in the noise-free data set, as the Rand index curve in Figure 4(d) shows. When noise points are added, the reli-

able range of $\sigma$ decreases with the increase of noise points (Figure 4(e) and (f)). For the noisy data set with 30 noise points, while there exists a very small range of $\sigma$ values in which standard spectral clustering can find the correct clusters, this range is so narrow that it is very hard to locate and hence the method generally fails. As for path-based spectral clustering, which can give good results (with a wide range of $\sigma$ values) for clean data, it is very difficult to get satisfactory results for noisy data sets. On the other hand, our method can find the correct clusters for a very wide range of $\sigma$ values on all three data sets (Figure 4(d)–(f)). It fails only when $\sigma$ gets too small. This shows that our method is easier to use and is more stable than standard spectral clustering.

Another parameter setting of our method that may affect the clustering results is the way of defining the neighborhoods $\mathcal{N}_i$'s. In our experiments, we use the same neighborhood size for all points in a data set so that the neighborhood is just large enough to include at least two neighbors in each neighborhood. This is not the only way to define the neighborhoods though. For example, we may use the mean edge length of the neighborhood graph derived from a data set to determine the radius of each neighborhood. The neighborhood graph can be constructed by connecting $\mathbf{x}_i$ and $\mathbf{x}_j$ if $\mathbf{x}_i$ is one of the $K$ nearest neighbors of $\mathbf{x}_j$ or if $\mathbf{x}_j$ is one of the
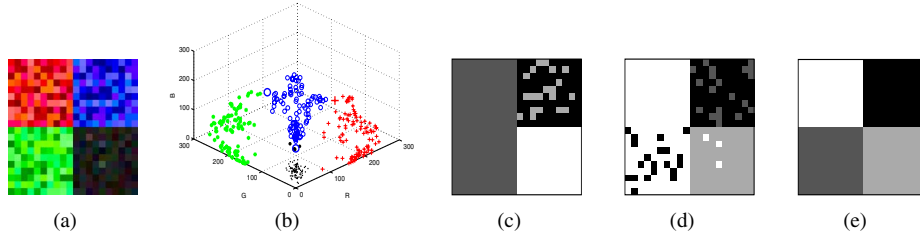
**Figure 3. Segmentation results for a synthetic image: (a) image with some noise points; (b) feature vectors in RGB color space; (c) spectral clustering result; (d) path-based spectral clustering result; and (e) robust path-based spectral clustering result.**
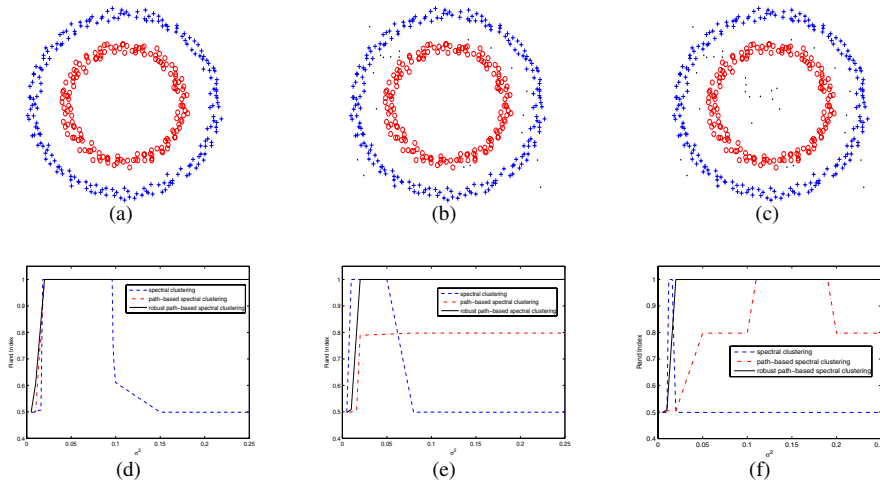


**Figure 4. Effect of $\sigma$ on clustering results: (a) 2-circle data set without noise points; (b) 2-circle data set with 20 noise points; (c) 2-circle data set with 30 noise points; (d)–(f) spectral clustering results (blue dash line), path-based spectral clustering results (red dash-dot line) and robust path-based spectral clustering results (black solid line) on (a)–(c).**

$K$ nearest neighbors of $\mathbf{x}_i$ based on the Euclidean distance. Under this alternative setting, which we have also tried, our method gives similar results to those shown above.

## 5. Experiments on Color Image Segmentation

While the results presented in the previous section are very promising, we would like to test our robust path-based spectral clustering algorithm on real-world clustering problems too. In particular, we study the color image segmentation problem in this section.

### 5.1. Color Image Segmentation

Image segmentation tries to parse natural images into homogeneous tiles corresponding to different surfaces and ob-

jects. Homogeneity is usually defined as pairwise similarity between pixels or regions. Recent work on pairwise clustering, including spectral clustering [14, 15, 16] and path-based clustering [4, 3, 6], has demonstrated impressive results for image segmentation.

In this section, we compare our robust path-based spectral clustering algorithm with several non-robust methods for image segmentation based on color and spatial features from the images.

### 5.2. Experimental Settings

The images used in our experiments are from the Berkeley Segmentation Dataset and Benchmark [11]. As in [4, 3, 6], we formulate image segmentation as a data clustering problem based on sparse proximity data. For each image
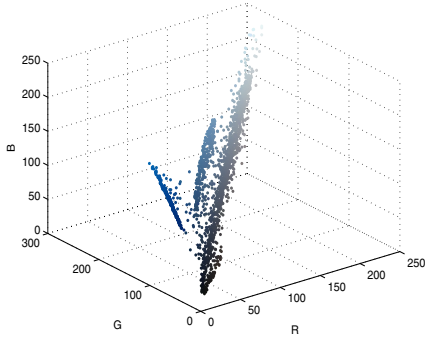
**Figure 5. Image patches from the bottommost image of Figure 6(a) shown as feature vectors in the RGB color space.**

of 154401 ($321 \times 481$) pixels, we consider 1855 ($35 \times 53$) overlapping patches with each patch of size $13 \times 13$ pixels.

Similarity between patches is computed based on color and spatial features. We use a relatively simple feature representation scheme in our experiments. For each image, the average color of each patch is represented as a point or feature vector $\mathbf{x}_i^c, i = 1, \ldots, 1855$ in the RGB color space. Figure 5 shows all the feature vectors for the bottommost image of Figure 6(a). We can see that the sky, lake and mountain form three manifolds with elongated structures. To segment the three regions correctly, a clustering algorithm should not only find the elongated structures but should also be sufficiently robust even at the presence of outliers.

Besides color features, we also include some spatial features as in [1]. The spatial features are simply the two coordinates of each patch in the image lattice, denoted as $\mathbf{x}_i^s, i = 1, \ldots, 1855$. The overall feature vector $\mathbf{x}_i$ for patch $i$ is the concatenation of the vectors for color and spatial features, i.e., $\mathbf{x}_i = ((\mathbf{x}_i^c)^T, \lambda(\mathbf{x}_i^s)^T)^T$, where the superscript $^T$ denotes vector/matrix transpose and $\lambda > 0$ specifies the relative importance of the spatial information for defining the similarity measure.

### 5.3. Image Segmentation Results

Figure 6 shows the segmentation results for four images on natural scenes. From left to right, the five columns show the input color images, segmentation results based on $k$-means clustering, spectral clustering, path-based spectral clustering, and robust path-based spectral clustering, respectively. The segmentation results are shown with different gray levels representing different segments. In these images, color changes in the sky, water, grass land and mountain form elongated manifold structures, such as the exam-

ple shown in Figure 5. It can be seen that the robust path-based spectral clustering algorithm outperforms all other methods. Although path-based clustering can also cluster elongated structures, it is not robust enough against noise and outliers. Specifically, the existence of outlier points between manifolds in the feature space ends up connecting the corresponding color segments together.

For more quantitative comparison of the different methods, we make use of the global consistency measure (GCE) and local consistency measure (LCE) proposed by [11]. These two measures quantify the consistency between different segmentation results and have been validated through extensive experiments. With these measures, deviation or refinement from a reference segmentation result (typically human segmentation result) can be quantified by an error measure, but not simply regarded as an incorrect result in the form of a binary decision. Between GCE and LCE, GCE has more stringent requirements since it requires all local refinements to be in the same direction while LCE allows refinements in different directions in different parts of the image.

We perform image segmentation experiments on a subset of 40 color images from the Berkeley segmentation database.[2] Each image has at least five human segmentation results available in the database. For each image, we compare the segmentation results based on different clustering methods with each human segmentation result. Figure 7 depicts the comparison results, where the distributions of GCE and LCE are shown as histograms. The horizontal axis of each histogram shows the range of GCE or LCE values, while the vertical axis indicates the percentage of comparisons. From the subfigures, we can see that human segmentation results for the same image are fairly consistent. The segmentation results using $k$-means clustering, spectral clustering and path-based clustering algorithms have much higher errors. On the other hand, the robust path-based clustering algorithm can give significantly lower errors.

## 6. Concluding Remarks

In this paper, we have presented a robust path-based spectral clustering algorithm that makes the combined use of spectral clustering and path-based clustering more robust against noise and outliers. We summarize some major advantages of our method here. First, it is robust in the sense that Equation (2) can give a reliable measure of the inter-point similarity even at the presence of noise and outliers. Therefore, based on the robust similarity measure, our method is much more robust than standard spectral clustering and path-based clustering on noisy data sets and data sets involving slighted coupled clusters. Second, it is also

---

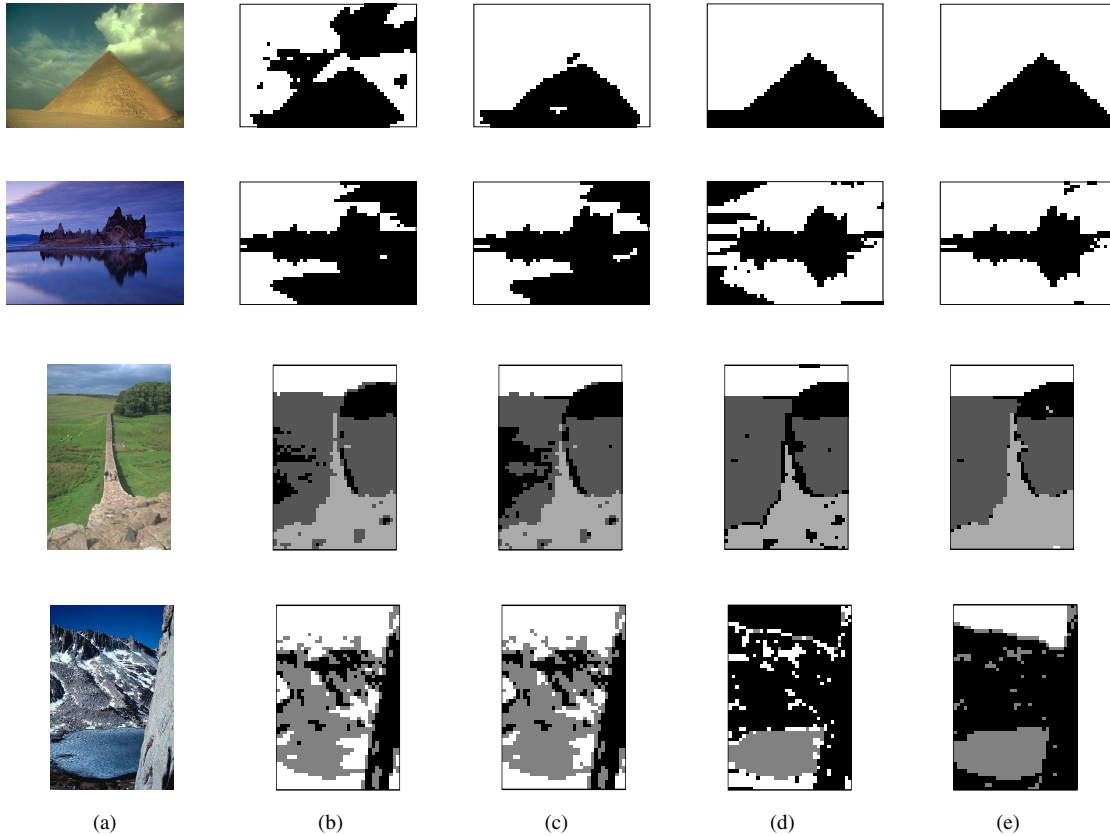[2]More segmentation results are available in `http://www.cs.ust.hk/˜hongch/image-segmentation.htm`.

**Figure 6. Color image segmentation results: (a) input images; (b) $k$-means clustering results; (c) spectral clustering results; (d) path-based spectral clustering results; and (e) robust path-based spectral clustering results.**

robust in the sense that the algorithm is not sensitive to the scaling parameter $\sigma$ while the standard spectral clustering algorithm is. The implication of this desirable property is that our method is easier to use and is more stable, making it a good candidate as a general clustering method. Third, it is easy for our method to be extended to the semi-supervised learning setting, which allows it to handle some even more challenging cases with the aid of supervisory side information. This makes our method even more powerful for solving some highly challenging clustering problems.

Despite its promising performance, there is still room for us to further enhance our method. To trade accuracy for efficiency, we perform our image segmentation experiments on image patches instead of pixels, as was done in [4, 3, 6]. As a consequence, the segmentation results are relatively crude compared with human segmentation results. Although some approximation methods have been proposed [5], we can look for even more efficient approximation methods for specific clustering tasks. Recently, Fowlkes

et al. [7] proposed an approximation technique based on sampling to alleviate the computational burden of spectral grouping. Inspired by their ideas, one possible extension is to explore the possibility of solving the segmentation problem for a small subset of pixels first and then extrapolating this solution to the full set of pixels in the image. Moreover, image segmentation methods typically make use of combined features, including color, brightness and texture, in patch or gradient forms [8]. We will consider incorporating more feature types into our method to further improve its image segmentation accuracy.
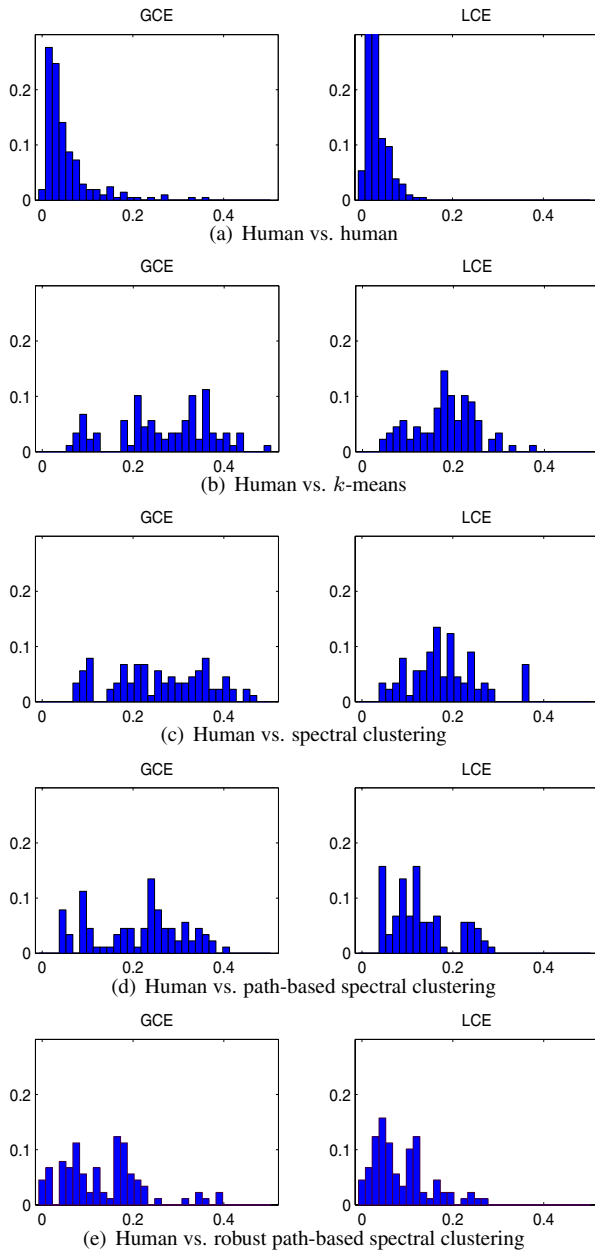
## Acknowledgments

**Figure 7. Histograms of the distribution of errors (GCE and LCE) for different segmentation methods. Human segmentation results compared to results based on: (a) human segmentation; (b) $k$-means clustering; (c) standard spectral clustering; (d) path-based spectral clustering; and (e) robust path-based spectral clustering.**

## References

[1] D. Comaniciu and P. Meer. Mean shift: a robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619, 2002.

[2] R. Duda, P. Hart, and D. Stork. *Pattern Classification*. Wiley, New York, NY, USA, 2nd edition, 2001.

[3] B. Fischer and J. Buhmann. Bagging for path-based clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(11):1411–1415, 2003.

[4] B. Fischer and J. M. Buhmann. Path-based clustering for grouping of smooth curves and texture segmentaion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(4):513–518, 2003.

[5] B. Fischer, V. Roth, and J. M. Buhmann. Clustering with the connectivity kernel. In *Advances in Neural Information Processing Systems 16*. 2004.

[6] B. Fischer, T. Zöller, and J. M. Buhmann. Path based pairwise data clustering with application to texture segmentation. *Energy Minimization Methods in Computer Vision and Pattern Recognition*, 2134:235–250, 2001.

[7] C. Fowlkes, S. Belongie, F. Chung, and J. Malik. Spectral grouping using the Nyström method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(2):214–225, 2004.

[8] C. Fowlkes, D. Martin, and J. Malik. Learning affinity functions for image segmentation: combining patch-based and gradient-based approaches. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 54–61, 2003.

[9] P. Huber. Robust regression: asymptotics, conjectures, and Monte Carlo. *Annals of Statistics*, 1(5):799–821, 1973.

[10] A. Jain and R. Dubes. *Algorithms for Clustering Data*. Prentice Hall, Englewood Cliffs, NJ, USA, 1988.

[11] 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. In *Proceedings of the Eighth IEEE International Conference on Computer Vision*, volume 2, pages 416–425, 2001.

[12] A. Y. Ng, M. I. Jordan, and Y. Weiss. On spectral clustering: analysis and an algorithm. In *Advances in Neural Information Processing Systems 14*. 2002.

[13] W. Rand. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 66:846–850, 1971.

[14] J. Shi and J. Malik. Normalized cuts and image segmentation. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 731–737, San Juan, Puerto Rico, 17–19 June 1997.

[15] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.

[16] Y. Weiss. Segmentation using eigenvectors: a unifying view. In *Proceedings of the Seventh IEEE International Conference on Computer Vision*, volume 2, pages 975–982, Kerkyra, Greece, 20–27 September 1999.