

***k*-Means image segmentation using Mumford–Shah model**

Nilima Shah,^{a,b} Dhanesh Patel,^a and Pasi Fränti^{b,*}

^aThe Maharaja Sayajirao University of Baroda, Department of Applied Mathematics,
Faculty of Technology and Engineering, Vadodara, Gujarat, India

^bUniversity of Eastern Finland (UEF), School of Computing, Joensuu, Finland

Abstract. *k*-Means (KM) is well known for its ease of implementation as a clustering technique. It has been applied for color quantization in RGB, YUV, hyperspectral image, Lab, and other spaces, but this leads to fragmented segments as the pixels are clustered only in the color space without considering connectivity. The problem has been attacked by adding connectivity constraints, or using joint color and spatial features (r, g, b, x, y), which prevent fragmented and nonconvex segments. However, it does not take into account the complexity of the shape itself. The Mumford–Shah model has been earlier used to overcome this problem but with slow and complex mathematical optimization algorithms. We integrate Mumford–Shah model directly into KM context and construct a fast and simple implementation of the algorithm. The proposed approach uses standard KM algorithm with distance function derived from Mumford–Shah model so that it optimizes both the content and the shape of the segments jointly. We demonstrate by experiments that the proposed algorithm provides better results than comparative methods when compared using various error evaluation criteria. The algorithm is applied on 100 images in the Weizmann dataset and two remote sensing images. © The Authors. Published by SPIE under a Creative Commons Attribution 4.0 International License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JEI.30.6.063029](https://doi.org/10.1117/1.JEI.30.6.063029)]

Keywords: image segmentation; clustering; *k*-means; Mumford–Shah segmentation; optimization algorithms.

Paper 210347 received Jun. 17, 2021; accepted for publication Dec. 8, 2021; published online Dec. 24, 2021.

1 Introduction

Image segmentation is an essential preprocessing task in many computer vision problems. It handles the problem of segregating regions of interest of an image that possess homogeneity and are spatially connected. These segregated regions may represent objects and thus help to recognize parts of satellite images, or patterns in biometric images, or detect diseased areas in MRI/x-ray images. Segmentation is also extensively used in automation and other artificial intelligence applications. Hence, profound research is still carried out and numerous image segmentation techniques are available in the literature, but a universal method applied to any type of image related to any image processing problem is not known.

1.1 Image Segmentation

Let the function $g(x, y)$ represent light intensity at any point (x, y) of domain $\Omega \in \mathbb{R}^2$ where an image is captured. Image is then defined as $g = \{(x, y) | (x, y) \in \mathbb{R}^2\}$. The RGB color representation of an image is mathematically represented by $g(x, y) \in \mathbb{R}^3$. The aim is to find regions $\Omega_1, \Omega_2, \dots, \Omega_n$ for $n \in \mathbb{N}$, of a domain $\Omega \in \mathbb{R}^2$, corresponding to different objects or their parts or shadows in the image. In summary, the segmentation problem has the following definition: Suppose Ω be a domain of h , find $\Omega_1, \Omega_2, \dots, \Omega_n$ such that

*Address all correspondence to Pasi Fränti, franti@cs.uef.fi

$$\Omega = \bigcup_{i=1}^n \Omega_i,$$

and g varies smoothly over the regions Ω_i , $i = 1, 2, \dots, n$, but discontinuously over the boundaries between Ω_i , $i = 1, 2, \dots, n$.

In terms of approximation theory, we can represent it as: find an approximation f of g in such a way that each f_i represents the region Ω_i via a piece-wise smooth function.

Thresholding, clustering, and classification methods,^{1–3} region-based methods,⁴ and edge-based active contour methods^{5–7} are different approaches to image segmentation.⁸ The segmentation problem has also been considered as an energy minimization problem. But to date, obtaining a generalized solution that works without *a priori* facts about the object elements, its characteristics, such as shape, color, texture, appearance of shadows, and overlapping of objects, is an open problem.

1.2 Classical *k*-Means Clustering for Image Segmentation

Clustering is defined as an unsupervised learning task, where one needs to identify a finite set of unknown categories called clusters.⁹ Clustering can also be applied to image segmentation via grouping the pixels by minimizing intrasegment similarity and by maximizing intersegment similarity.

One of the categories of clustering is hard clustering, where a pixel belongs to one and only one cluster. It is applicable to data sets that have sharp boundaries between clusters.¹⁰ This is suitable also for image segmentation where the goal is to group the pixels into nonoverlapping regions.

The most popular algorithm in hard clustering is classical *k*-means (KM).⁹ It is simple to implement, and its computational cost is low, which makes it suitable for large data sets such as images. Panwar et al.¹¹ observed that performance of KM segmentation improves with increase in the number of clusters when mean square error and peak signal to noise ratio are used to measure the quality of the reconstruction.

Applying KM based on the pixel values, it corresponds to color quantization. However, this is rarely enough for image segmentation because there is no control of the shapes and convexity of the segments, which can therefore become very fragmented.

KM algorithm has also other drawbacks –such as the number of clusters k must be known *a priori*.¹² The algorithm is also sensitive to the initialization, which can lead to an inferior clustering result. However, there are plenty of solutions in literature to attack these problems. First, if the number of clusters needs to be solved, we need to revise the original minimization function so that the number of clusters is no longer a parameter but a variable that is also solved.¹³ The quality of clustering can be significantly improved using better initialization and by repeating the algorithm 100 times.¹⁴ If not enough, a simple modification is to use random swap algorithm instead,¹⁵ which practically never gets stuck into an inferior local minimum.

To use KM for image segmentation, the procedure is as follows: First, randomly select k pixels from the image as the initial cluster centroids. Second, segment the image by labeling every pixel by finding the minimum-squared Euclidean distance to every centroid. Third, calculate new centroids by averaging the pixels belonging to the same clusters. The steps two and three are repeated until the centroids stabilize. This procedure aims at optimizing sum-of-squared error (SSE) between the pixels and their nearest centroid

$$\text{SSE} = \sum_i \|f - g\|^2.$$

However, using KM as such leads to color quantization where every pixel is mapped to the nearest color regardless of its location in the image and the mapping of its neighbors. In image segmentation, it is desired that the pixels close to each other in the image would be assigned into the same segment; and vice versa, pixels far away would be assigned to different segments.

The algorithm is shown as follows. It will find local minimum for the SSE function.

Algorithm 1 Classical KM

X : original image, a set of N data vectors (image of $N \leftarrow n_1 \times n_2$ RGB vectors)
 k : number of clusters
 C_I : initialized k cluster centroids
 C : final cluster centroids of k -clustering
 P : $\{p_i | i = 1, \dots, N\}$, cluster label of image X
 $KM(X, C_I) \rightarrow (C, P)$

REPEAT

$C_{prev} \leftarrow C_I$;

//Generate optimal partitions

FOR all $i \in [1, N]$ DO

FOR all $j \in [1, k]$ DO

$SSE_j \leftarrow SSE(x_i, c_j)$;

$p_i \leftarrow \arg \min_{1 \leq j \leq k} (SSE_j)$

//Calculate optimal centroids

FOR all $j \in [1, k]$ DO

$c_j \leftarrow \text{Average of } x_i, \text{ whose } p_i = j$;

UNTIL $C = C_{prev}$

Many solutions also implement modifications to the KM algorithm. Theiler and Gisler¹⁶ proposed a variant called *contig-KM*. It calculates within-cluster variance (V) to measure compactness of the cluster, and discontinuity (D) at a single pixel as the fraction of its neighbors that are in a different cluster. *Contig-KM* minimizes a linear combination of compactness and discontinuity as $\lambda D + (1 - \lambda)V$ where the parameter λ defines the relative importance of these two properties. They state that spatial contiguity and spectral compactness properties are nearly orthogonal, i.e., meaning that we can make considerable improvements in the one with only minimal loss in the other. Unlike conventional KM, *Contig-KM* updates the centroids immediately after a pixel changes its segment. Its time complexity is the same as that of KM: $O(NkdI)$, where d is the number of attributes ($d = 3$ with RGB images, $d = 103$ to 204 with the remote sensing images), and I is the number of iterations.

Fuzzy *c*-means (FCM) has also been applied to hyperspectral image (HSI) segmentation using normalized spatial distance in addition to the normalized color distance.¹⁷ The difference of FCM to KM is that the pixels can belong to multiple segments with a membership value in range $[0, 1]$. It is expected to provide smoother clustering result at the cost of higher time complexity of $O(Nkd^2I)$. However, the results reported in Ref. 17 show that the result of standard FCM (0.69) is worse than that of KM (0.72) and an additional processing step called Albedo recovery was applied before clustering to reach better accuracy (0.81).

Lindsten et al.¹⁸ proposed relaxation of KM method formulated as a convex optimization problem which does not require the number of clusters to be specified beforehand. Instead, a regularization parameter is used to control the trade-off between model fit and the number of clusters. They show that due to relaxation and convexification, the original NP-hard KM problem becomes efficiently solvable but this means time complexity of $O(n^3)$ using semi-definite programming, which is quite high compared with KM.

A two-step KM-based method was developed by Mignotte.¹⁹ The steps include detexturing the input image and then using spatially constrained KM segmentation. The spatial constraint is a set of color values around the pixel. The time complexity is the same as that of KM added by the

detexturing process and the analysis of the spatial constraint. The method was reported to perform competitively with the state-of-the-art methods at the time.

A lot of research has been done so far; however, there are many issues that still need to be addressed. Connectivity constraint can still generate complex shape. Spatial constraint tends to keep the segments small but it can prevent to find segments of variable sizes. Many solutions implement modifications to the standard KM algorithm, which itself should not be necessary if a good algorithm and proper minimization function is used. In this paper, we attack the problem by employing minimization function, which strictly controls the complexity of the shape via the Mumford–Shah functional.

1.3 Mumford–Shah Model

Mumford–Shah model is widely used and extensively studied energy minimization model for image segmentation. It was introduced by Mumford and Shah in 1989²⁰ with the aim at segmenting the image into nearly homogeneous regions, separated by consistent boundaries. It consists of finding the minimum of a functional, which is formed by combining the length of the boundaries, the gradient of a smoothed version of the image and the difference of the true and smoothed images.²¹ The Mumford–Shah functional is defined as follows:

$$E(f, \tau) = \mu^2 \iint_{\Omega} \|f - g\|^2 dx dy + \iint_{\Omega - \tau} \|\nabla f\|^2 dx dy + \lambda L(\tau), \quad (1)$$

where function f approximates image g which is differentiable over the regions Ω_i

$$\Omega = \bigcup_{i=1}^n \Omega_i \cup \tau.$$

Here, τ is the set of arcs forming the boundaries, μ and λ are the positive weights that control the quality of the approximation and the coarseness of the segmentation, respectively, and $L(\tau)$ denotes the total length of all the arcs belonging to τ . A large value of λ will lead to fewer boundaries. The goal is to minimize E , i.e., find f and τ so that E is minimized; the smaller the value of E , the better is the segmentation.

The first term in Eq. (1) ensures that f approximates the image g , i.e., gives us homogeneous regions. The second term ensures that only smooth changes appear inside the segments. The third term prevents boundaries from growing very large.

If the segments are homogeneous, the second term is no longer needed in Eq. (1). We then obtain a simplified, piecewise constant Mumford–Shah model

$$E(f, \tau) = \mu^2 \iint_{\Omega} \|f - g\|^2 dx dy + \lambda L(\tau). \quad (2)$$

There has been extensive use of these models in the areas of image denoising, image segmentation, and image restoration.²² Algorithmic techniques to minimize the Mumford–Shah functional includes the use of Douglas–Rachford algorithm (MS-DR) in Ref. 23, and several implicit method used by researchers in Refs. 24–29. However, the existing approaches are rather slow, the algorithms can be complex, and they may result in poor segmentation result if the parameters are not tuned properly.

In this paper, we show how simple KM algorithm can be used to minimize Mumford–Shah functional. It has all the same benefits as other Mumford–Shah-based methods but with faster and simpler optimization algorithm. Besides the simplicity and speed, the other benefits are that it avoids complex shapes, does not produce numerous tiny segments, and no algorithmic patches are needed as the standard KM algorithm itself can be applied by simple modifications of the partition step adopted with the shape constraint.

The method can be applied to any system where low level color-based segmentation is needed as the first step of the process. These include scene classification, detection of human activity in visual surveillance, assigning a name to a human face in image, classifying

handwritten characters, and identifying land cover in satellite images. In this paper, we demonstrate how the method applies to remote sensing images.

The rest of the paper is organized as follows. Section 2 introduces the proposed KM segmentation using Mumford–Shah model. Section 3 explains the evaluation criteria. In Sec. 4, we compare segmented images achieved through the proposed KM using Mumford–Shah model with the ground truth images using criteria discussed in Sec. 3. In Sec. 5, we use the compactness measure to evaluate multiphase segmentation of three methods: proposed KM using Mumford–Shah model, KM, and regularized KM (reg-KM). Section 6 shows the application of the proposed approach to HSIs. Conclusions are drawn in Sec. 7.

2 *k*-Means Using Mumford–Shah Model

The goal is to divide the image into m segments by optimizing certain criteria. Image segmentation consists of two contradicting criteria: (1) similarity criterion such as variance, which measures goodness of the segment and (2) spatial connectivity criterion that penalizes pixels belonging to different segment than their neighbors. KM includes only the similarity criterion, the intracluster variance.

A reg-KM using Mumford–Shah was proposed by Ref. 30 combining the intensity and the spatial information together. The model calculates reg-KM energy using two terms. First term introduces the notion of size of the cluster in the regularization term using the scale. Scale is obtained as $P(A)/|A|$ where $P(A)$ denotes the perimeter and $|A|$ denotes the area of the cluster. Second term measures the intracluster dissimilarity of the clusters in the same way as in the KM. Their results show better classification of objects when using the combined approach compared with using only the intensity values. However, the regularization term is optimized when, for a given perimeter, the area is maximized. This happens for round shapes whereas we should not direct the optimization toward any given shape. The time complexity of reg-KM is $O(Nk^2d)$.

We next present a new variant called Mumford–Shah KM (MS-KM). For spatial criterion, we introduce a new term adopted from the simplified Mumford–Shah functional in Eq. (2). We use the total boundary length of the segments. It drives the segmentation toward simplicity but not directly toward any certain shape. We next outline how this can be done within the KM.

When generating the optimal partition in KM, the algorithm decides into which segment a pixel will be assigned. Using Mumford–Shah function, the optimal assignment depends not only on the pixel value as in standard KM, but also on its effect on the boundary length. Fortunately, this can be calculated from image gradients easily. The result is a label for every pixel in the image to indicate which segment it is assigned to.

The segmentation process goes as follows: First, we select m random centroids and initialize the gradient at each pixel of image as 0. We obtain the label image P [see Fig. 1(a)] using Mumford–Shah distance (MSD) defined in our algorithm (see Sec. 2.2). At the first iteration the boundary length of each pixel is set to 0. Next, we calculate new centroids of each segment by averaging the pixels belonging to the same segment. The assignment and centroid steps are then repeated until maximum iterations have been reached, or until the Mumford–Shah functional stop improving ($\Delta E \leq 0$).

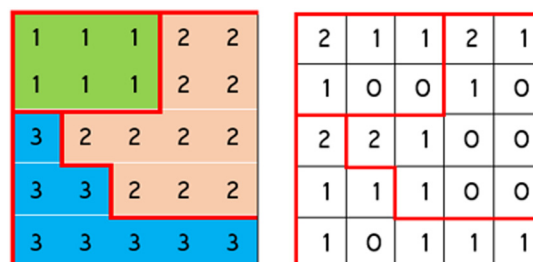


Fig. 1 (a) Labeled image and (b) gradient of the labeled image.

The Mumford–Shah functional E is calculated at the end of each iteration as follows:

$$E_{\text{iter}} \leftarrow \sum_{1 \leq i \leq N} \{\text{SSE}_i\} + \lambda * \text{BL},$$

where sse_i is the total squared error between every pixel to its cluster centroid, and BL is the total boundary length of all the segments.

This functional is an extension of the standard SSE criterion. Similar λ -based approach is widely used in other multiobjective optimization problems. It jointly optimizes both the content and the border of the segments, and therefore, it is expected to provide more robust segmentation than simpler heuristics such as boundary smoothing of the connected components. For more details about the theoretical properties of Mumford–Shah functional, we refer to Refs. 21 and 23, and for the properties of KM algorithm, we refer to Ref. 14.

2.1 Calculating Gradient and Boundary Length of the Segment

Gradient of the image is directional change in intensity or color in the image. Mathematically, it can be defined as the derivative between the pixel value and its surrounding. In image processing, it can be obtained by calculating the change in the horizontal and in the vertical direction separately. Here, instead of intensity or color, we calculate gradient from the labeled image [see Fig. 1(a)]. We use zero padding on the boundaries. Also, the gradient is calculated in backward direction. Hence, the right and bottom borders of the image are not marked as boundary. Gradient of the pixel at point (x, y) is calculated as

$$\text{gp}(x, y)_{\text{new}} = \text{gp}(x, y)_{\text{nb left}} + \text{gp}(x, y)_{\text{nb top}} = |p_{(x-1, y)} - p_{(x, y)}| + |p_{(x, y-1)} - p_{(x, y)}|, \quad (3)$$

$$\text{where } \text{gp}(x, y)_{\text{nb}} = \begin{cases} 1, & |p_{\text{nb}} - p_{(x, y)}| > 0 \\ 0, & |p_{\text{nb}} - p_{(x, y)}| = 0 \end{cases}.$$

We can conclude the boundary length directly from the gradients. The total boundary length is computed by summing up the gradients of all the pixels in the label image

$$\text{BL} = \sum_{1 \leq x, y \leq N} \{\text{gp}_{(x, y)}\}. \quad (4)$$

In Fig. 1, if we trace the boundary marked with red, we will find borders of length 21 pixels, in total. The same result can be obtained by summing up the gradient values using Eq. (4).

2.2 Distance Calculation in MS-KM

From the second iteration onward, when deciding the cluster label to be assigned to a pixel, we need to find out its effect on the boundary length (here gradient) when assigned to a particular




Assigned to	$p_{(x, y)}$ in green	$ p_{\text{nb}} - p_{(x, y)} $		gp_{nb}		total $\text{gp}_{(x, y)} =$ $\text{gp}_{\text{nb left}} +$ $\text{gp}_{\text{nb top}}$
		Left	Top	Left	Top	
Cluster 4		2	3	1	1	2
Cluster 2		0	5	0	1	1
Cluster 7		5	0	1	0	1

Fig. 2 Changed label of the pixel and the corresponding changes in gradients.

Algorithm 2 KM using Mumford–Shah

X : input image, a set of N data vectors (image of $N \leftarrow n_1 \times n_2$ RGBvectors)

k : number of clusters

λ : weighting factor controlling the coarseness of the segmentation

C : initialized k cluster centroids

C : final cluster centroids

iter: energy functional of Mumford–Shah equation

blen: boundary length of the segments

MS-KM (X, C, λ) $\rightarrow (C, P)$

FOR all $i \in [1, N]$ DO

$p_i \leftarrow 0$;

 iter $\leftarrow 1$;

REPEAT

 // generate optimal partitions

 GP \leftarrow ImageGradient (P);

 FOR all $i \in [1, N]$ DO

 FOR all $j \in [1, k]$ DO

$MSD_j \leftarrow$ MumfordShahDist (GP _{i} , x_i , c_j);

$p_i \leftarrow \arg \min_{1 \leq j \leq k} MSD_j$

$MSD_i \leftarrow \min_{1 \leq j \leq k} \{MSD_j\}$

 blen \leftarrow CalculateBoundaryLength (GP);

$E_{\text{iter}} \leftarrow \sum_{1 \leq i \leq N} \{SSE_i\} + \lambda * \text{blen}$ //Compute functional E

 // Calculate optimal centroids

 FOR all $j \in [1, k]$ DO

$c_j \leftarrow$ Average of x_i , whose $p_i = j$;

 iter \leftarrow iter + 1;

UNTIL ($\Delta E_{\text{iter}} < 0$) OR (iter = maxiter)

MumfordShahDist (g, x, c) \rightarrow MSD

//Calculate gradient for the pixel x

IF $|p_{\text{nb}} - p_x| > 0$

 gpx $\leftarrow 1$

ELSE

 gpx $\leftarrow 0$

MSD $\leftarrow (x - c)^2 + \lambda * \text{gpx}$

cluster. The boundary length of the pixel will be obtained as gradient using Eq. (4). It is further illustrated in Fig. 2.

The decision about which segment the pixel is assigned to, is calculated by MSD, which is calculated as follows:

$$\text{MumfordShahDist}(x, y, j) \leftarrow (g(x, y) - c_j)^2 + \lambda * gp_j(x, y),$$

where λ is the weighting factor controlling the coarseness of the segmentation. In other words, the distance is the squared Euclidean distance between the pixel and the centroid of the clusters plus its new gradient value within that segment (weighted by λ). Thus, the distance incorporates two different criteria—similarity and spatial convexity. The algorithm for MS-KM is given in Algorithm 2. Its main structure is the same as KM containing both the optimal partitioning and calculating the new centroids.

Even though the $\lambda \cdot gp$ part is fixed during a single iteration, any change of the partition stage will influence changes in the gp part of the equation. This will eventually create different segmentation than KM. We further illustrate this with an example image of size 10×10 in Fig. 3. Figure 4 shows the effect of the MSD on the small red object of the sample image.

Original	KM		MS-KM		
Initial	Iteration 1 (final)	Initial	Iteration 1	Iteration 2 (final)	
Blen	22	22	20	12	
SSE	4.00E-04	4.00E-04	5.7864	10.9107	
E	16.5004	16.5004	20.7864	19.9107	
ΔE	—	—	—	0.8757	

Fig. 3 Sample image and results with KM and proposed MS-KM method.

Iteration 1		Iteration 2	
SSE with centroid 0	SSE with centroid 0	Blen with label 0	MSD with centroid 0
SSE with centroid 1	SSE with centroid 1	Blen with label 1	MSD with centroid 1
Labeled		Labeled after iteration 2	

Fig. 4 Merging of left pixel of small red object into cluster 1 from above sample image where $\lambda = 0.83$.

3 Experimental Evaluations

To evaluate the performance of our algorithm in case of two-phase segmentation, we use region-based error metrics defined in Ref. 31 and structural similarity index.³² For multiphase segmentation, we have used compactness measure.³³

3.1 Bidirectional Consistency Error

Considering S_1 as segmentation obtained by an algorithm, and S_2 as one of the ground truth, we calculate

$$E(S_1, S_2, p_i) = \frac{|R(S_1, p_i)/R(S_2, p_i)|}{|R(S_1, p_i)|}, \quad (5)$$

a value in the range $[0, \infty]$, where zero represents no error. Here, $R(S, p_i)$ is the set of pixels corresponding to the region in segmentation S that contains pixel p_i . Here, local error measure is not symmetric, and it gives refinement measure in one direction only. It is to be noted that $E(S_1, S_2, p_i) = 0$ if S_1 is a refinement of S_2 at pixel p_i , but not the opposite. Using Eq. (5), the following four evaluation measures can be calculated:

$$\text{GCE}(S_1, S_2) = \frac{1}{n} \min \left\{ \sum_i E(S_1, S_2, p_i), \sum_i E(S_2, S_1, p_i) \right\}, \quad (6)$$

$$\text{LCE}(S_1, S_2) = \frac{1}{n} \sum_i \min \{ E(S_1, S_2, p_i), E(S_2, S_1, p_i) \}, \quad (7)$$

$$\text{BCE}(S_1, S_2) = \frac{1}{n} \sum_i \max \{ E(S_1, S_2, p_i), E(S_2, S_1, p_i) \}, \quad (8)$$

$$\text{BCE}^*(S_1, S_2) = \frac{1}{n} \sum_i \min_k \{ \max \{ E(S_1, S_2, p_i), E(S_2, S_1, p_i) \} \}. \quad (9)$$

Equations (6)–(8) represent the global consistency error (GCE), local consistency error (LCE), and bidirectional consistency error (BCE), respectively. Equation (9) is the measure (BCE*) obtained by computing the minimum error at every pixel, which is consistent with each ground truth segmentation S_k of that image. We will use BCE*.

3.2 Structural Similarity Index

Structural similarity index (SSIM)³² measures the visual quality between two images. It is calculated as a weighted combination of three factors luminance, contrast, and structural similarity. It can be expressed as

$$\text{SSIM}(x, y) = [l(x, y)]^\alpha \cdot [c(x, y)]^\beta \cdot [s(x, y)]^\gamma,$$

where l is the luminance, c is the contrast, s is the structural similarity, and α , β , and γ are the positive constants. The three factors are calculated separately as follows:

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1},$$

$$c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2},$$

$$s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3},$$

where μ_x and μ_y are the local means, σ_x and σ_y are the standard deviations, and σ_{xy} is the cross-covariance between the two images x and y , respectively. If $\alpha = \beta = \gamma = 1$, then SSIM simplifies to

$$\text{SSIM} = \frac{(2\mu_x\mu_y + C_1) \cdot (2\sigma_x\sigma_y + C_2)}{(\mu_x^2 + \mu_y^2 + C_1) \cdot (\sigma_x^2 + \sigma_y^2 + C_2)}.$$

We use SSIM to calculate the similarity between the image generated by the segmentation result, and the image generated from the ground truth segmentation.

3.3 Compactness Measure

Assessment of segmentation quality has been analyzed using combination of shape and compactness.³³ In Ref. 34, the authors used the idea of compactness in the area of measuring super pixel compactness. Several approaches have been proposed for measuring shape compactness of discrete regions. In Ref. 35, a review of several methods for measuring shape circularity and compactness was done. They also survey some of the methods when scaling transformation is applied to discrete regions.

Compactness is defined as the ratio of the area of an object to the area of a circle with the same perimeter

$$\text{Compactness} = \frac{\text{Perimeter}^2}{4 \cdot \pi \cdot \text{area}}.$$

The measure takes a minimum value of 1 for a circle. Objects that have complicated, irregular boundaries have larger compactness.

3.4 Adjusted Rand Index

The Rand index (RI) computes a similarity measure between two segmentation results. All pairs of pixels are considered and pairs that are assigned in the same or different segments in the obtained segmentation and its ground truth are counted.

The RI score is then “adjusted for chance” to get adjusted RI using the following formula:

$$\text{ARI} = (\text{RI} - \text{Expected_RI}) / (\max(\text{RI}) - \text{Expected_RI}).$$

The adjusted RI is thus ensured to have a value close to 0.0 for random labeling independently of the number of segments and pixels and exactly 1.0 when the segmentations are identical.

4 Results for Two Phase Segmentations

We used nine random images from the Weizmann dataset³⁶ depicting one object in the foreground and their three ground truth segmentations. Figure 5 shows the original image and one of the three ground truths. The resulting segmentations obtained by KM, reg-KM, MS-DR, and the proposed MS-KM are shown in Fig. 6. The experiments were conducted on a desktop with Intel Core i5 processor with 2.50 GHz speed, 8 GB RAM, 64-bit Windows 10 operating system. The results of KM, MS-DR, and MS-KM depend on the choice of the initial centroids. We therefore executed KM 20 times, calculated the BCE* error for every result, and then selected the one with lowest BCE* value to obtain the best KM result. This was the used as the input to MS-DR and the proposed MS-KM. The number of clusters was selected to match the ground truth. The reg-KM method uses the change in the energy directly, that each pixel is moved to the phase which locally minimizes the energy compared with the previous phase. The parameter used for regularization gives an option of creating a new cluster. If the value of the parameter is large, it increases the number of clusters. Here in Fig. 6, we show the results for only two segments generated by reg-KM, by selecting the value of the parameter between 0.1 and 0.5. Visually, the results show that the proposed method gives improved results with inclusion of

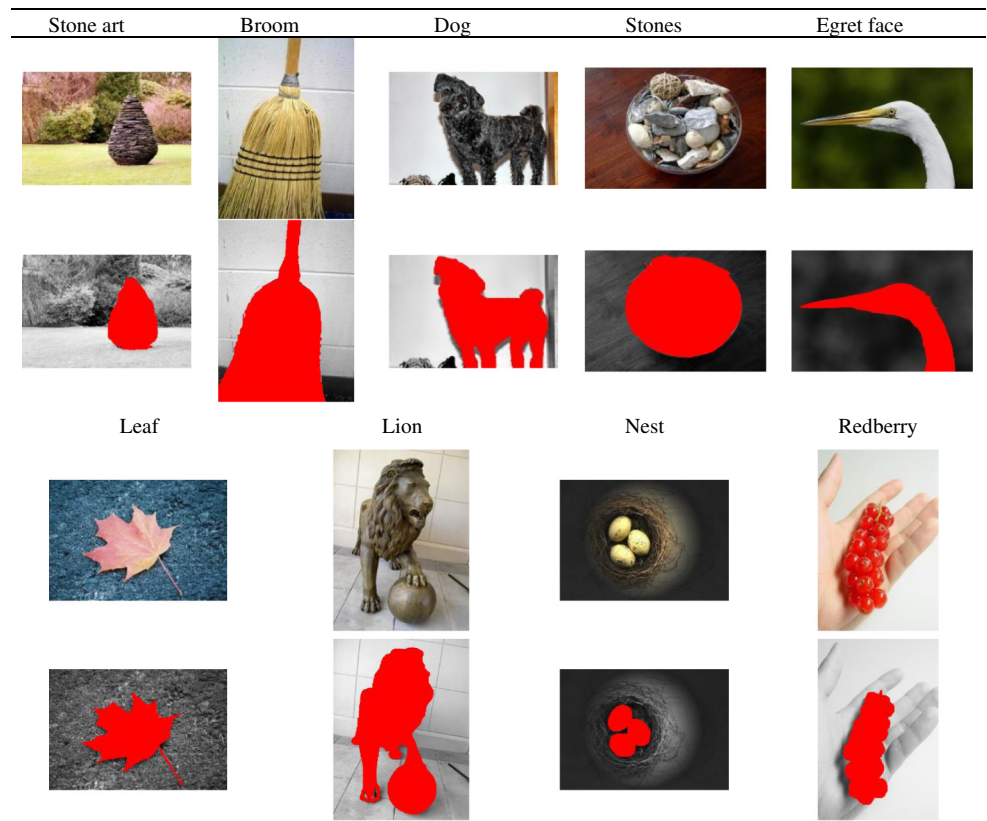


Fig. 5 Original and one of the three ground truths.

scattered minute segments into bigger segments. Also, the proposed method gives smoother boundaries, which can be seen in partial detailed images of Fig. 7.

Results of the BCE* error measure are given in Table 1. We can see that the proposed method reduces the error of KM and reg-KM in eight out of the nine cases. The higher values of SSIM in Table 2 indicate better quality segmentation using the proposed approach. Figure 8 shows the effect of different values of λ on the segmentation. The larger is the λ value, the smoother are the segments. Eventually, too many details will be lost.

We tested statistical significance of the results for the nine images in Table 2 by running KM and MS-KM 10 times with different randomly chosen centroids as the initial solution. We used one-tailed Wilcoxon signed-rank test for paired samples as normality of our distribution is not satisfied. The observed value test-statistic $W = 1$ and the critical value at $N = 9$ ($p < 0.05$) is of $W = 8$, which shows that the difference between KM (0.69) and MS-KM (0.79) is statistically significant. However, the difference between MS-DR (0.78) and MS-KM (0.79) is not significant.

The processing times are compared in Table 3. The proposed method is almost as fast as KM (25% slower), and significantly than MS-DR (10 times faster) and reg-KM (600 times faster). All KM variants share the same time complexity of $O(NkdI)$ except reg-KM, which uses slow $O(Nk^2d)$ initialization. MS-DR also uses KM as the initialization but the following proximal splitting step by Douglas–Rachford has a higher number of iterations and more operations in each step, which together make the algorithm slower than the faster KM variants.

Figure 9 shows the effect of λ on the results. As the images have nothing in common, we cannot predict much about the value of λ except that the value 0.5 gives us the average result in all the cases. Figure 10 shows the compactness measure when KM and MS-KM methods are applied to 100 images of the Weizmann dataset. Resolution of the images in Weizmann dataset vary from 300×170 (lowest) to 300×400 pixels (highest). For the proposed MS-KM method, we have taken $\lambda = 0.5$ which gives average results. The results can be even better if appropriate λ is taken for each image. These results are also statistically significant according to ANOVA test ($p < 0.00001$).

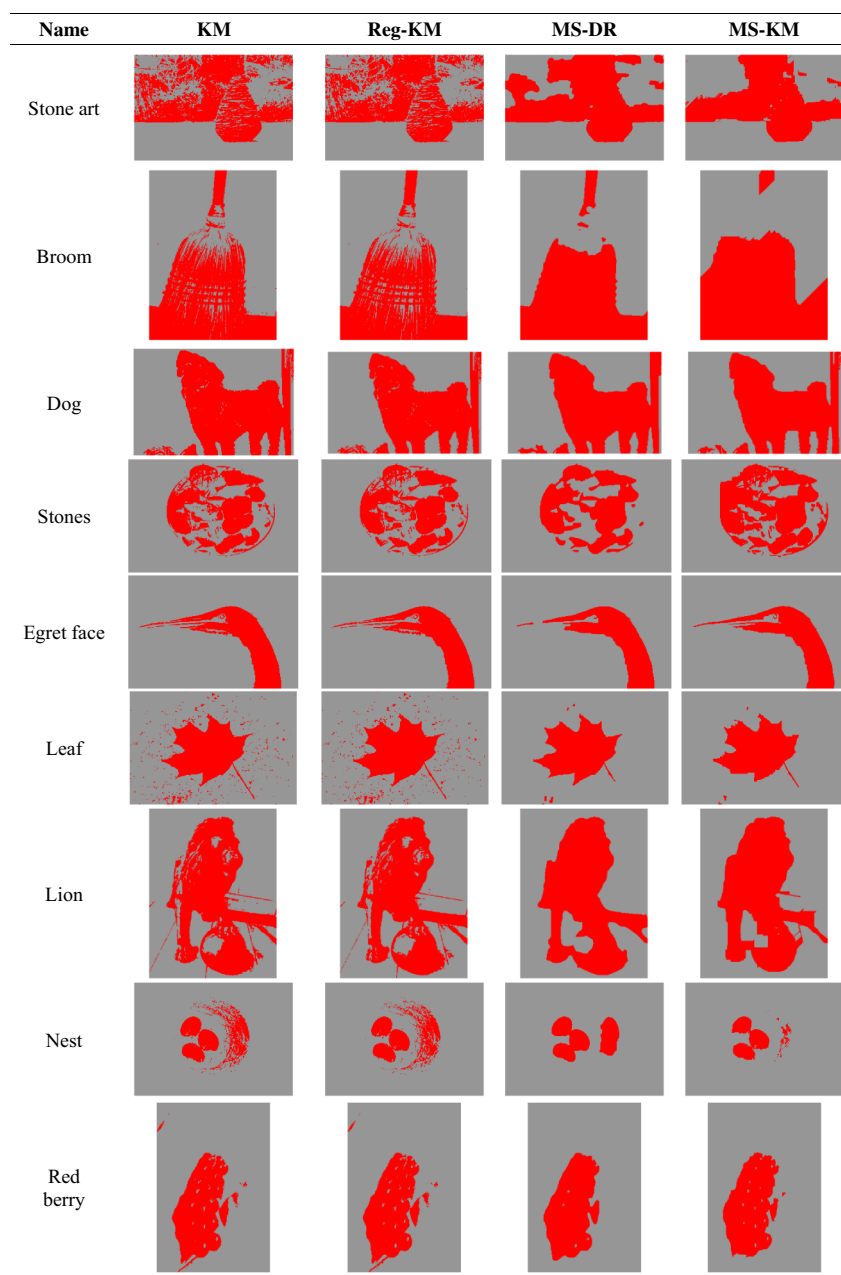


Fig. 6 Results of KM, Reg-KM, MS-DR, and the proposed MS-KM method.

5 Results for Multiphase Segmentation

We have used three selected images: horse, cows, and flowers. Figure 11 shows the resulting segmentations obtained by KM, the proposed MS-KM, and reg-KM as per.³⁰ Visually, the results show that the proposed method gives less fragmented segments with smooth and regular boundaries.

We can observe from Figs. 12–14 that the compactness value of all clusters by the case of proposed MS-KM method is significantly smaller than that of the other techniques. We also observe that the results of KM and reg-KM techniques are almost similar, except Fig. 13 where reg-KM works better.

We also note that all KM variants tend to produce fragmented results despite the spatial constraint. The reason is not the spatial constraint being too weak, but actually the opposite, being too strong. Figure 15 demonstrates the issue with MS-KM using larger 4-pixel neighborhood gradient instead of the proposed 2-pixel neighborhood. It is theoretically more solid but

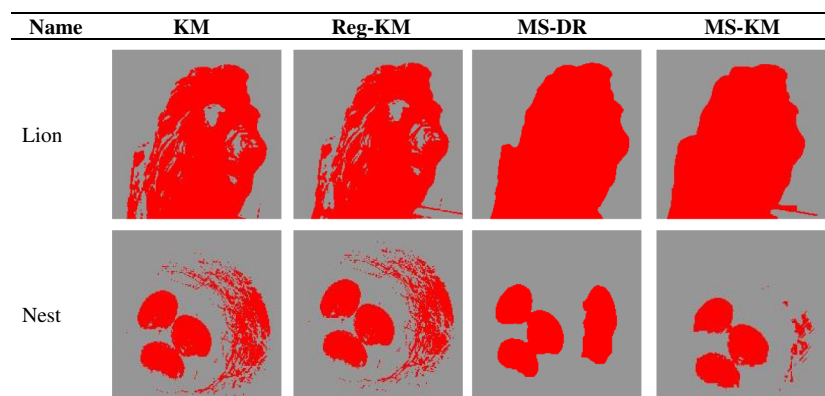


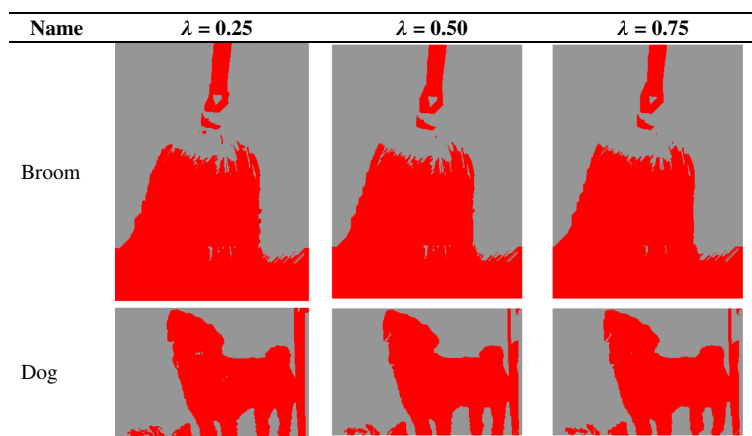
Fig. 7 Partial images showing inclusion of minute segments into bigger segments and boundary smoothing.

Table 1 Comparison using BCE* (the lower the better).

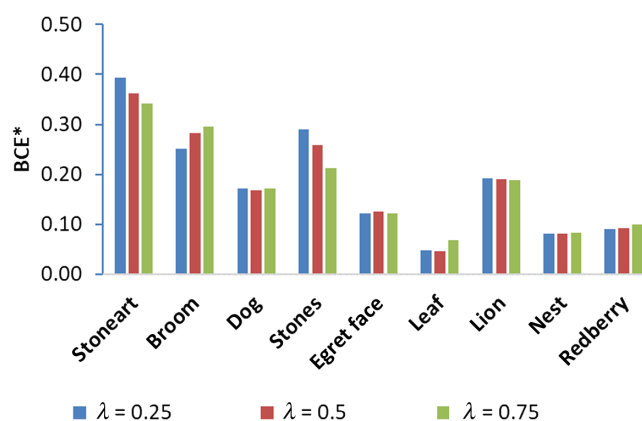
Name	KM	Reg-KM	MS-DR	Proposed MS-KM
Stone art	0.49	0.50	0.43	0.36
Broom	0.23	0.25	0.21	0.28
Dog	0.17	0.21	0.18	0.17
Stones	0.32	0.38	0.33	0.26
Egret face	0.13	0.14	0.07	0.13
Leaf	0.07	0.09	0.02	0.05
Lion	0.22	0.30	0.19	0.19
Nest	0.12	0.12	0.08	0.08
Redberry	0.10	0.14	0.08	0.09
Average	0.21	0.24	0.18	0.18

Table 2 Comparison using structural similarity index (the higher the better).

Name	KM	Reg-KM	MS-DR	Proposed MS-KM
Stone art	0.37	0.37	0.58	0.64
Broom	0.67	0.67	0.84	0.78
Dog	0.77	0.77	0.80	0.80
Stones	0.64	0.64	0.64	0.68
Egret face	0.95	0.95	0.95	0.95
Leaf	0.49	0.49	0.66	0.64
Lion	0.70	0.70	0.81	0.85
Nest	0.80	0.80	0.86	0.87
Redberry	0.86	0.86	0.91	0.89
Average	0.69	0.69	0.78	0.79

**Fig. 8** Output for proposed MS-KM for various values of λ .**Table 3** Runtime in seconds.

Name	KM	Reg-KM	MS-DR	MS-KM
Stone art	0.16	144.27	2.19	0.31
Broom	0.14	440.77	1.79	0.29
Dog	0.05	64.42	1.42	0.12
Stones	0.07	40.39	4.06	0.08
Egret face	0.04	20.25	2.87	0.18
Leaf	0.09	26.15	2.58	0.39
Lion	0.10	199.50	2.27	0.29
Nest	0.09	15.78	1.36	0.08
Redberry	0.09	112.40	1.81	0.10
Average	0.16	118.21	2.26	0.20

**Fig. 9** Effect of λ for proposed MS-KM method on BCE* (the lower the better).

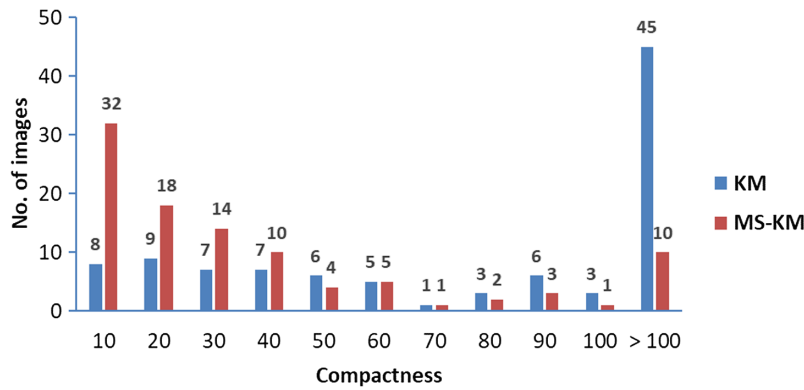


Fig. 10 Frequency distribution of the compactness scores provided by KM and the proposed MS-KM method for $\lambda = 0.5$ (the lower the compactness the better).

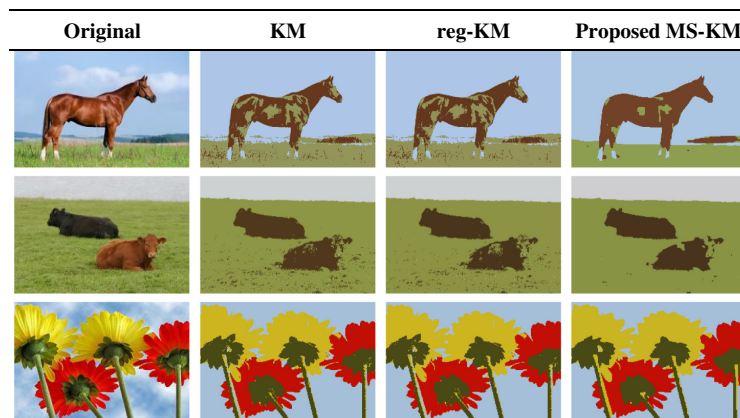


Fig. 11 Output for various methods.

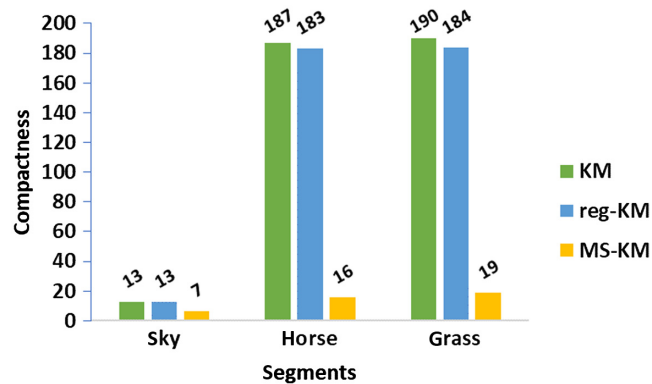


Fig. 12 Compactness for horse image.

will lead to fragmentation in practice. The reason is the KM initialization, which contains many isolated subsegments, and KM with too strong spatial constraint is unable to erode these away as well as the simpler 2-pixel gradient.

6 Application to HSIs

In this section, we apply the proposed MS-KM method to HSIs. Unlike RGB color image, which stores the intensity value of three bands only (red, green, and blue), HSIs can have intensity

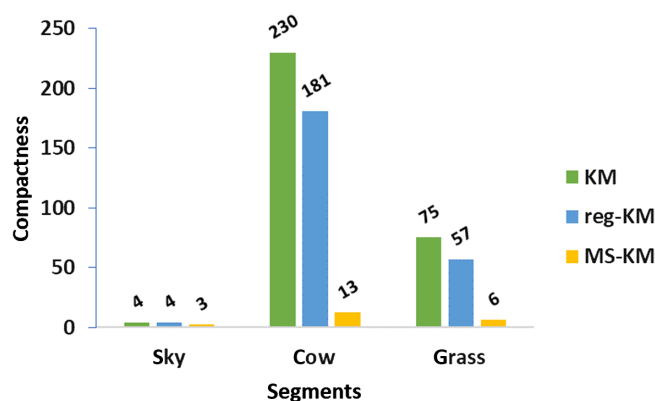


Fig. 13 Compactness for cow image.

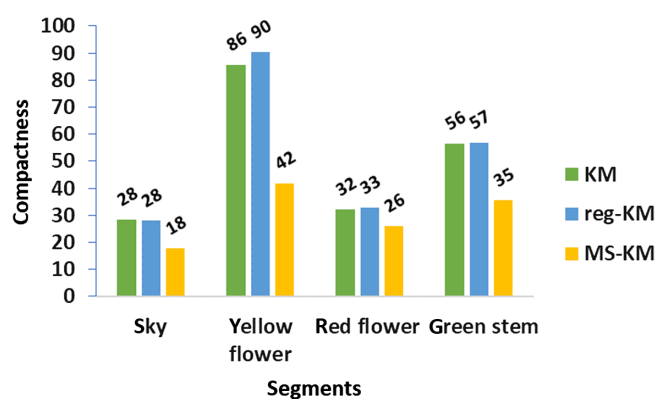


Fig. 14 Compactness for flowers image.

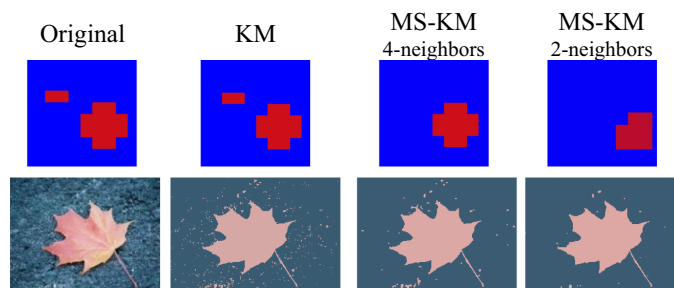


Fig. 15 Fragmentations of the KM variants due to local optimality of isolated subsegments.

information of more than 100 spectral bands, which are in discrete intervals of 5 to 10 nm across the visible to the infrared region. The spectral resolution is very high due to narrow band gap, which leads to high dimensionality. However, the results in Ref. 15 have shown that the performance of KM depends mostly on the overlap of the clusters in the feature space and is less dependent of the dimensionality. This is also the case with the HSI images, which makes KM as a suitable choice here.

We have used two images. First image is AVIRIS-Salinas HSI captured by 224-band AVIRIS sensor in 1998 over agricultural fields in the Salinas Valley, California.³⁷ The number of spectral bands used is 204 (discarding 20 water absorption bands) and the total number of pixels is 512×217 . Its ground truth with 16 classes is shown in Fig. 16. There are also unclassified data in the ground truth, which are shown by white color.

The second image is Pavia University HSI acquired by the ROSIS sensor over Pavia, northern Italy.³⁷ The number of spectral bands is 103, and the number of pixels is 610×610 , but some of the samples contain no information and have been discarded before the analysis. Hence, the actual resolution is 340×610 pixels. The ground truth has nine classes. In this image, there are also unclassified data in the ground truth, which are shown by white color in Fig. 16.

As a preprocessing step, we normalize bands using the following formula:

$$\text{norm_val} = (\text{val} - \text{band_min}) / (\text{band_max} - \text{band_min}).$$

The number of clusters is set to 16 for the AVIRIS-Salinas image and to 9 for the Pavia University image according to the number of classes in their ground truth data. The value of λ is set to 0.5. The results of KM method and the proposed MS-KM are shown in Fig. 16. The implementation of the reg-KM did not support HSIs. As expected, the segments generated by the proposed method are much less fragmented than that of generated by KM. More detailed analysis will follow.

Table 4 shows the BCE* error, adjusted Rand index (ARI), and SSIM for the KM and the proposed method. In case of Pavia University image, the BCE* values are 0.27 by KM and 0.18 by MS-KM. Smaller BCE* values indicate better accuracy. The ARI values are very close to 1.0 by both methods. While the proposed method improves based on BCE*, the segmentation is worse when measured by SSIM (0.94 versus 0.87). While MS-KM provides more convex and visually pleasant segments, the color difference of those fragmented pixels is so high that they cause measurable degradation in SSIM.

In case of AVIRIS-Salinas image, both BCE* values are close to 0.50 (0.47 and 0.46). While RI values are still high (0.92 and 0.93), the results indicate that, regardless of the algorithm, segmenting the image by spectral features alone is not sufficient even when using >100 bands. This is evidenced by the fact that even the less fragmented segmentation by the proposed method did not provide significant improvement according to any of the used measure.

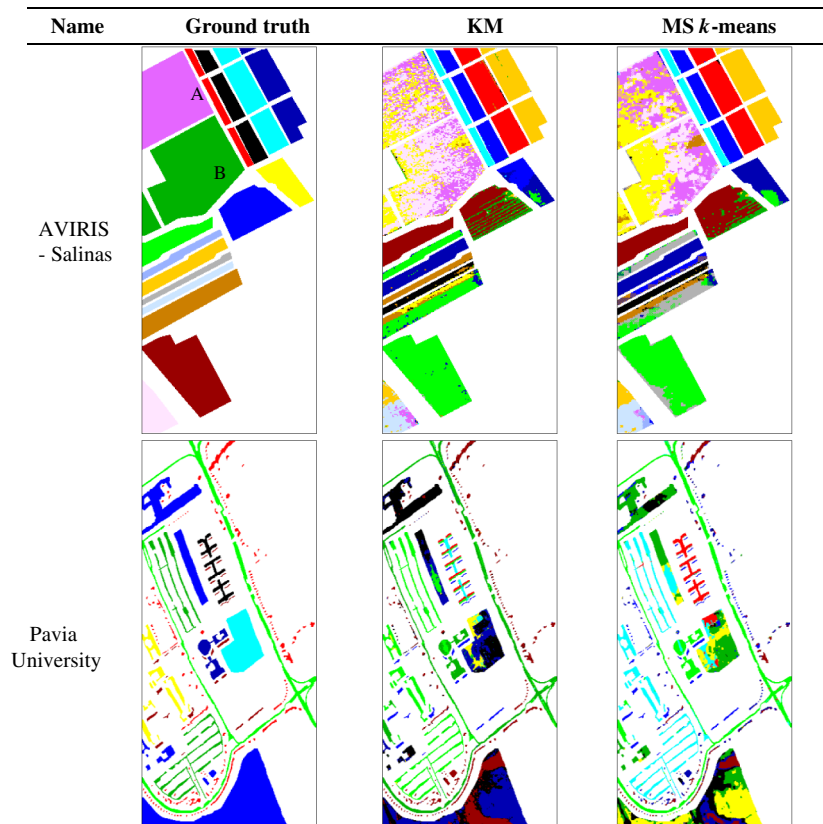


Fig. 16 Results of the proposed MS-KM.

Table 4 Various error metrics for KM and MS-KM method.

Name	KM			MS-KM		
	BCE*	RI	SSIM	BCE*	RI	SSIM
AVIRIS-Salinas	0.47	0.92	0.63	0.46	0.93	0.62
Pavia University	0.23	0.97	0.94	0.23	0.98	0.87

The authors of Ref. 38 have concluded that their nearest-neighbor density-based (NN-DB) methods perform better than density-based spatial clustering of applications with noise (DBSCAN), fast sparse affinity propagation (FSAP), and fuzzy C-means (FCM) regardless the chosen number of clusters (obtained or initialized). For AVIRIS-Salinas image, the highest ARI for all the methods mentioned in Ref. 38 is ~ 0.7 . They reported that classes A and B cannot be clearly identified by any of the methods in their tests, and by any methods in papers published earlier. We have observed the same phenomenon with our tests.

The quality assessment of the segmentation results of HSIs requires precise and unbiased ground truth.³⁹ When analyzed, the spectral signatures of the ground truth (GT) classes reveal that several classes exhibit high heterogeneity. Especially for Pavia University HSI, it is observed by the authors that the GT classes are not perfectly homogeneous and do not correctly represent the diversity and variability of the land cover.

7 Conclusions

The approach is introduced by embedding the well-known Mumford–Shah model into KM. The key contribution is to combine KM with Mumford–Shah model. Its main benefits are: (1) better segmentation performance than other models used with KM and (2) significantly faster optimization than that of the existing optimization methods for Mumford–Shah model. The results of the proposed approach are better than KM and reg-KM techniques in terms of shorter boundary lengths, and 25% better SSIM-value than Reg-KM (0.24 versus 0.18). The proposed method is almost as fast as KM (0.16 versus 0.20), 10 times faster than MS-DR (2.26 versus 0.20), and 600 times faster than reg-KM (118 versus 0.20).

The proposed method has two limitations. First, algorithm has the weakness of KM of being sensitive to the initialization which can lead to suboptimal clustering results and also fragmented segmentation. These could be overcome by better initialization and repeats,¹⁴ or by adding random swap step around the KM as in Ref. 40. The second limitation is that it was tested only with color features (RGB and HSI), which are not good enough in all applications. For example, the HSI examples showed that some segments cannot be modeled by the color spectrum. This could be overcome by applying detexturing method as a preprocessing¹⁹ or extending the method to be used with texture or some application-specific features. These are points for future research.

Acknowledgments

The authors declare no conflict of interest.

References

1. Y. Yang and S. Huang, “Image segmentation by fuzzy c-means clustering algorithm with a novel penalty term,” *Comput. Inf.* **26**, 17–31 (2007).
2. Y. A. Tolias and S. M. Panas, “Image segmentation by a fuzzy clustering algorithm using adaptive spatially constrained membership functions,” *IEEE Trans. Syst. Man Cybern. - Part A: Syst. Hum.* **28**, 359–369 (1998).
3. F. Gibou and R. Fedkiw, “A fast hybrid k-means level set algorithm for segmentation,” in *Proc. 4th Annu. Hawaii Int. Conf. Stat. Math.* (2002).

4. J. Martin, *Implementing the Region Growing Method Part 1: The Piecewise Constant Case*, Naval Air Warfare Center Weapons Division, China Lake, California (2002).
5. T. Chan and L. Vese, “An active contour model without edges,” *Lect. Notes Comput. Sci.* **1682**, 141–151 (1999).
6. T. Chan and L. Vese, “Active contours without edges,” *IEEE Trans. Image Process.* **10**, 266–277 (2001).
7. T. Chan and L. Vese, “A level set algorithm for minimizing the Mumford–Shah functional in image processing,” in *Proc. Var. and Level Set Methods in Comput. Vision Workshop* (2001).
8. N. Shah, D. Patel, and A. Jivani, “Review on image segmentation, clustering and boundary encoding,” *Int. J. Innov. Res. Science, Eng. Technol.* **2**, 6309–6314 (2013).
9. V. K. Dehariya, S. K. Shrivastava, and R. C. Jain, “Clustering of image data set using K-means and fuzzy K-means algorithms,” in *Int. Conf. Comput. Intell. Commun. Networks* (2010).
10. F. Z. Kettaf, D. Bi, and J. P. Asselin de Beauville, “A comparison study of image segmentation by clustering techniques,” in *Proc. Third Int. Conf. Signal Process. (ICSP’96)* (1996).
11. P. Panwar and G. Gopal, “Image segmentation using K-means clustering and thresholding,” *Image* **3**, 1787–1793 (2016).
12. P. Lukac et al., “Simple comparison of image segmentation algorithms based on evaluation criterion,” in *Proc. 21st Int. Conf. Radioelektron.* (2011).
13. Q. Zhao and P. Fränti, “Editorial: WB-index: a sum-of-squares based index for cluster validity,” *Data Knowl. Eng.* **92**, 77–89 (2014).
14. P. Fränti and S. Sieranoja, “How much can k-means be improved by using better initialization and repeats?” *Pattern Recognit.* **93**, 95–112 (2019).
15. P. Fränti and S. Sieranoja, “K-means properties on six clustering benchmark datasets,” *Appl. Intell.* **48**, 4743–4759 (2018).
16. J. Theiler and G. Gisler, “Contiguity-enhanced k-means clustering algorithm for unsupervised multispectral image segmentation,” in *Opt. & Photonics* (1997).
17. P. Azimpour et al., “Hyperspectral image clustering with Albedo recovery Fuzzy C-Means,” *Int. J. Remote Sens.* **41**(16), 6117–6134 (2020).
18. F. Lindsten, H. Ohlsson, and L. Ljung, “Just relax and come clustering!: a convexification of k-means clustering,” Technical report from Automatic Control at Linköpings Universitet, LiTH-ISY-R-2992 (2011).
19. M. Mignotte, “A de-texturing and spatially constrained K-means approach for image segmentation,” *Pattern Recognit. Lett.* **32**, 359–367 (2011).
20. D. Mumford and J. Shah, “Optimal approximations by piecewise smooth functions and associated variational problems,” *Commun. Pure Appl. Math.* **42**, 577–685 (1989).
21. D. Mumford and J. Shah, “Boundary detection by minimizing functionals,” in *IEEE CVPR* Vol. 17, pp. 22–26 (1985).
22. L. Bar et al., *Mumford and Shah Model and Its Applications to Image Segmentation and Image Restoration*, 2 ed., pp. 1097–1154, Springer Science+Business Media LLC (2011).
23. N. Shah, D. Patel, and P. Fränti, “Fast Mumford–Shah two-phase image segmentation using proximal splitting scheme,” *J. Appl. Math.* **2021**, 1–13 (2021).
24. L. Ambrosio and V. Tortorelli, “Approximation of functional depending on jumps by elliptic functional via t-convergence,” *Commun. Pure Appl. Math.* **43**, 999–1036 (1990).
25. E. Brown, T. Chan, and X. Bresson, “Convex formulation and exact global solutions for multi-phase piecewise constant Mumford–Shah image segmentation,” *UCLA CAM Report* (2009).
26. X. Cai, R. Chan, and T. Zeng, “Image segmentation by convex approximation of the Mumford–Shah model,” *UCLA CAM Report* (2012).
27. W. Fleming and R. Rishel, “An integral formula for total gradient variation,” *Arch. Math.* **11**, 218–222 (1960).
28. G. Koepfler, C. Lopez, and J. Morel, “A multiscale algorithm for image segmentation by variational method,” *SIAM J. Numer. Anal.* **31**, 282–299 (1994).
29. T. Pock et al., “A convex relaxation approach for computing minimal partitions,” in *IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 810–817 (2009).

30. S. Kang, B. Sandberg, and A. Yip, “A regularized *k*-means and multiphase scale segmentation,” *Inverse Probl. Imaging* **5**, 407–429 (2011).
31. D. Martin et al., “A database of human segmented natural images and its application to evaluating Seg Algo and measuring ecological statistics,” in *Proc. Eighth IEEE Int. Conf. Comput. Vision, ICCV* (2001).
32. A. C. Brooks, X. Zhao, and T. N. Pappas, “Structural similarity quality metrics in a coding context: exploring the space of realistic distortions,” *IEEE Trans. Image Process.* **17**, 1261–1273 (2008).
33. H. Tonbul and T. Kavzoglu, “The effect of shape and compactness parameters on segmentation quality: a case study,” in *Int. Symp. GIS Appl. Geogr. & Geosci.* (2017).
34. A. Schick, M. Fischer, and R. Stiefelhagen, “Measuring and evaluating the compactness of superpixels,” in *Proc. 21st Int. Conf. Pattern Recognit. (ICPR2012)* (2012).
35. R. S. Montero and E. Bribiesca, “State of the art of compactness and circularity measures,” *Int. Math. Forum* **4**, 1305–1335 (2009).
36. S. Alpert et al., “Image segmentation by probabilistic bottom-up aggregation and cue integration,” in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.* (2007).
37. M. Graña, M. A. Veganzons, and B. Ayerdi, “Computational Intelligence Group University of the Basque Country (UPV / EHU),” http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes#Salinas_scene.
38. C. Cariou, S. Le Moan, and K. Chehdi, “Improving K-nearest neighbor approaches for density-based pixel clustering in hyperspectral remote sensing images,” *Remote Sens.* **12**, 3745 (2020).
39. K. Chehdi and C. Cariou, “Learning or assessment of classification algorithms relying on biased ground truth data: what interest?” *J. Appl. Remote Sens.* **13**, 034522 (2019).
40. P. Fränti, “Efficiency of random swap clustering,” *J. Big Data* **5**(13), 1–29 (2018).

Nilima Shah is an assistant professor at The Maharaja Sayajirao University of Baroda, Vadodara, Gujarat, India. She received her bachelor’s degree in physics and master’s degree in computer applications from The Maharaja Sayajirao University of Baroda in 1993 and 1997, respectively. Currently, she is pursuing a PhD at the University of Eastern Finland.

Dhanesh Patel is a professor of applied mathematics at The Maharaja Sayajirao University of Baroda, Vadodara, Gujarat, India. He received his master’s degree in mathematics and a PhD in applied mathematics from The Maharaja Sayajirao University of Baroda. He also received his master’s degree in industrial mathematics from the University of Kaiserslautern, Kaiserslautern, Germany, in 2001. He was a DAAD fellow from 1999 to 2001. His research interests include mathematics in industry, computational fluid dynamics, numerical methods for PDE, image processing, and mathematical biology.

Pasi Fränti received his master’s and PhD degrees from University of Turku, Finland, in 1991 and 1994, respectively. He has been tenure professor in the University of Eastern Finland since 2000 where he is currently the leader of the machine learning group. He has published 100 papers in refereed journals and 177 papers in refereed conferences. He has gained over 7000 Google scholar citations. His research interests include clustering algorithms, location-based services, machine learning, data mining, and optimization and analysis of health care services.