

International Journal of Wavelets, Multiresolution and Information Processing
 © World Scientific Publishing Company

A parameterless scale-space approach to find meaningful modes in histograms - Application to image and spectrum segmentation

Jérôme Gilles

*Department of Mathematics and Statistics, San Diego State University
 5500 Campanile Drive, San Diego, CA 92182-7720, USA
 jgilles@mail.sdsu.edu*

Kathryn Heal

*Department of Mathematics, University of California, Los Angeles,
 520 Portola Plaza, Los Angeles, CA 90095, USA
 kathryn.heal@ucla.edu*

Received (Day Month Year)

Revised (Day Month Year)

Accepted (Day Month Year)

Published (Day Month Year)

In this paper, we present an algorithm to automatically detect meaningful modes in a histogram. The proposed method is based on the behavior of local minima in a scale-space representation. We show that the detection of such meaningful modes is equivalent in a two classes clustering problem on the length of minima scale-space curves. The algorithm is easy to implement, fast and does not require any parameter. We present several results on histogram and spectrum segmentation, grayscale image segmentation and color image reduction.

Keywords: Histogram; meaningful modes; scale-space; segmentation.

AMS Subject Classification: 60G35, 65D18, 68U10, 94A12

1. Introduction

Despite an extensive literature on this topic, image segmentation remains a difficult problem in the sense there does not exist a general method which works in all cases. One reason is in that the expected segmentation generally depends on the final application goal. Generally researchers focus on the development of specific algorithms according to the type of images they are processing and their final purpose (image understanding, object detection, ...). Different types of approaches were developed in the past which can be broadly classified into histogram based, edge based, region based and clustering (and mixes between them).

Although the more conceptually straightforward types of histogram methods are usually less efficient, they are still widely used because of their simplicity and the few computational resources needed to perform them. Despite their drawbacks,

these types of methods satisfy important criteria for computer vision applications. The idea behind such methods is that the final classes in the segmented image correspond to “meaningful” modes in an histogram built from the image characteristics. For instance, in the case of grayscale images, each class is supposed to correspond to a mode in the histogram of the gray values. Finding such modes is basically equivalent to finding a set of thresholds separating the mode supports in the histogram.

The body of literature regarding histogram-based segmentation is steadily increasing. Within this literature two main philosophies emerge, based on different approaches to the problem: techniques using histograms to drive a more advanced segmentation algorithm or techniques based on the segmentation of the histogram itself. For instance, Chan et al.⁵ compare the empirical histograms of two regions (binary segmentation) by using the Wasserstein distance in a levelset formulation. In Yuan et al.,²⁴ local spectral histograms (i.e obtained by using several filters) are built. The authors show that the segmentation process is equivalent to solving a linear regression problem. Based on their formalism, they also propose a method to estimate the number of classes. In Puzicha et al.,²⁰ a mixture model for histogram data is proposed and then used to perform the final clustering (the number of clusters is chosen according to rules based on statistical learning theory). In Sural et al.,²¹ it is shown that the HSV color space is a better color representation space than the usual RGB space. The authors use k -Means to obtain the segmentation. They also show that this space can be used to build feature histograms to perform an image retrieval task. Another approach, called JND (Just Noticeable Difference) histogram, is used in Bhoyar et al.³ to build a single histogram which describes the range of colors. This type of histogram is based on the human perception capabilities. The authors propose a simple algorithm to segment the JND histogram and get the final classes. This method still has some parameters and therefore we do not see it as being optimal. In Kurugollu et al.,¹⁴ the authors build 2D histograms from the pairs RB-RG-GB (from the RGB cube), then segment them by assigning each histogram pixel to their most attractive closest peak. The attraction force is based on distances to each histogram peak and their weights. This last step consists of the fusion of each segmentation to form a global one. In Yildizoglu et al.,²³ a variational model embedding histograms is proposed to segment an image where some reference histograms are supposed to be known. While all these methods can have several parameters, can be difficult to implement and computationally expensive, an interesting approach was investigated in Delon et al.⁶ The authors propose a fully automatic algorithm to detect the modes in an histogram H . It is based on a fine to coarse segmentation of H . The algorithm is initialized with all local minima of H . The authors defined a statistical criteria, based on the Grenander estimator, to decide if two consecutive supports correspond to a common global trend or if they are parts of true separated modes; this criteria is based on the ϵ -meaningful events theory (see Delon et al.⁶ or Desolneux et al.^{8,9}). The (parameterless) algorithm can be resumed as following: start from the finest segmentation given by all local

minima, choose one minimum and check if adjacent supports are part of a same trend or not. If yes then merge these supports by removing this local minima from the list. Repeat until no merging is possible. This work is extended to color images in Delon et al.⁷ the previous algorithm is applied to each of the different components H,S,V. The segmentation results of this application are used to initialize a k -Means algorithm to get the final segmentation. While this approach provides a fully automatic algorithm, it becomes computationally expensive (both in terms of time and memory) for histograms defined on a large set of bins.

One could think of a spectrum as a histogram that counts the occurrence of each frequency within a signal/function. Through this lens, we see that histogram mode detection methods can also be used to identify “harmonic modes” in Fourier spectra. For instance, the ability to find the set of supports of such modes is fundamental to build the new empirical wavelets proposed in Gilles et al.^{10,11} Although harmonic modes and histogram modes are intuitively comparable, it is essential to remember that the expected behavior of a spectrum can vary dramatically from that of a histogram. For instance, spectra are generally less regular than classic histograms. With spectra it can be more difficult to have an a priori idea of how many modes should be identified during segmentation. These characteristic differences will many times affect the performance of a segmentation method; thus the intended applications (spectral vs. histogram) must be taken into consideration before selecting/evaluating any methods.

In this paper, our aim is to propose a parameterless algorithm to automatically find meaningful modes in an histogram or spectrum. Our approach is based on a scale-space representation of the considered histogram which permits us to define the notion of “meaningful modes” in a simpler way. We will show that finding N (where N itself is unknown) modes is equivalent to perform a binary clustering. This method is simple and runs very fast. The remainder of the paper is organized as follows: in section 2, we recall the definition and properties of a scale-space representation. In section 3, we expose how the scale-space representation can be used to automatically find modes in an histogram. In section 4, we present several experiments in histogram segmentation, grayscale image segmentation and color reduction as well as the detection of harmonic modes in a signal spectrum. Finally, we draw conclusions in section 5.

2. Scale-space representation of a function

2.1. Continuous scale-space

Let a function $f(x)$ be defined over an interval $[0, x_{max}]$ and let the kernel $g(x; t) = \frac{1}{\sqrt{2\pi t}} e^{-x^2/(2t)}$ where t is called the scale parameter. Then the operator defined by ($*$ denotes the standard convolution product)

$$L(x, t) = T_t[f](x) = g(x; t) * f(x) \quad (2.1)$$

is a continuous scale-space representation (see Lindeberg,¹⁵ Witkin²²) of f if it fulfills the following axioms:

- Linearity: $T_t[f_1 + f_2](x) = T_t[f_1](x) + T_t[f_2](x)$,
- Shift invariance: $T_t[S_{\Delta_x}f](x) = S_{\Delta_x}T_t[f](x)$ where $S_{\Delta_x}f(x) = f(x - \Delta_x)$,
- Semi-group structure: $T_{t_2}[T_{t_1}[f]](x) = T_{t_1+t_2}[f](x)$,
- Kernel scale invariance, positivity and normalization,
- Non-creation of local extrema,

some complimentary axioms are needed when the dimension is larger than one, see Lindeberg.¹⁵ The scale-space representation can be interpreted in the following manner: when t increases, $L(x, t)$ becomes smoother in the sense that this operation removes all “patterns” of characteristic length smaller than \sqrt{t} . It is proven that the Gaussian kernel is the only kernel (in the continuous case) which fulfills these axioms (see Lindeberg¹⁵).

2.2. Discrete scale-space

From this point on, we will consider a sampled version of the function f . Hence, we need a discrete scale-space representation of f . Although the Gaussian kernel is the only admissible kernel in the continuous case, several kernel choices are viable in the discrete case. In this paper, we adopt the sampled Gaussian kernel:

$$L(m, t) = \sum_{n=-\infty}^{+\infty} f(m - n)g(n; t), \quad (2.2)$$

where

$$g(n; t) = \frac{1}{\sqrt{2\pi t}} e^{-n^2/2t}. \quad (2.3)$$

It is important to note that the semi-group property is not valid for a discretized Gaussian kernel except if the ratio t_2/t_1 is odd (see Proposition 12 in Lindeberg¹⁶). In practice we use a truncated filter in order to have a finite impulse response filter:

$$L(m, t) = \sum_{n=-M}^{+M} f(m - n)g(n; t), \quad (2.4)$$

with M large enough that the approximation error (measured by $2 \int_M^\infty g(u, t) du$, see Lindeberg¹⁶) of the Gaussian is negligible. A common choice is to set $M = C\sqrt{t} + 1$ with $3 \leq C \leq 6$ (meaning that the filter’s size is increasing with respect to t). In our experiments we fix $C = 6$ in order to ensure an approximation error smaller than 10^{-9} .

2.3. Discretization of the scale parameter

For numerical implementation purposes, we sample the scale parameter t in the following manner: $\sqrt{t} = s\sqrt{t_0}$ where $s = 1, \dots, S_{max}$ are integers. We choose to start

with $\sqrt{t_0} = 0.5$ because it corresponds to half the distance between two samples of f . Moreover, since we work with finite length signals, there is no interest to go further than $\sqrt{t_{max}} = x_{max}$. Finally, we want to perform a finite maximum number of steps, denoted S_{max} , to go from the initial scale to the final one. Thus we can write: $\sqrt{t_{max}} = S_{max}\sqrt{t_0}$ which implies $S_{max} = 2x_{max}$.

3. Scale-space histogram segmentation

3.1. Meaningful scale-space histogram modes

Our objective is to find meaningful modes in a given histogram; hence we first need to define the notion of “meaningful mode”. We will begin by defining what is a mode, and explaining its representation in the scale-space plane. Let us consider an histogram like the one depicted in figure 1 (a). It is clear that to find boundaries delimiting consecutive modes is equivalent to finding intermediate valleys (local minima) in the histogram. In order to be able to define the notion of “meaningful”

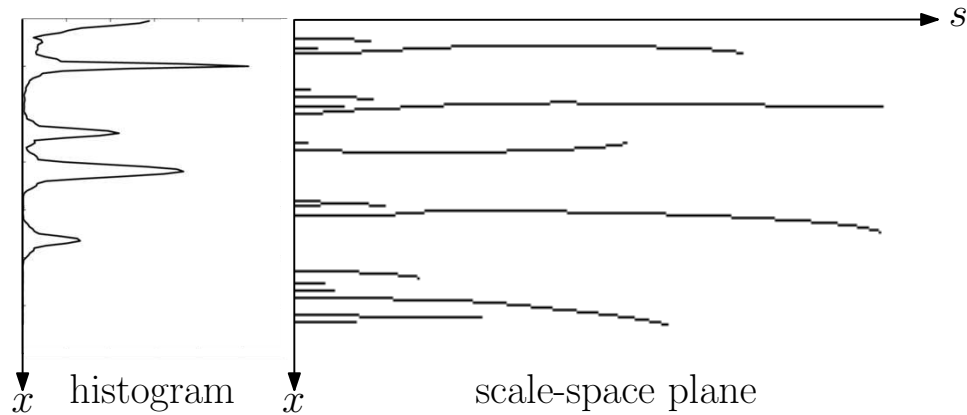


Fig. 1. Example of an histogram (on left, rotated by 90°) and its scale-space plane representation (on right). Each initial local minima (for $s = 1$) give rise of scale-space curves of different lengths.

modes, we will use a pivotal property of scale-space representations. The following notion is already used by the computer vision community for edge detection applications. The number of minima with respect to x of $L(x, t)$ is a decreasing function of the scale parameter t , or the scale-step parameter s as defined in the previous section, and no new minima can appear as t increases. Figure 1 gives an example of an histogram and its scale-space representation. Observe that each of the initial (for $s = 1$) minima generates a curve in the scale-space plane.

Let us fix some notations. The number of initial minima will be denoted N_0 , and each of these local minima defines a “scale-space curve” C_i ($i \in [1, N_0]$) of length

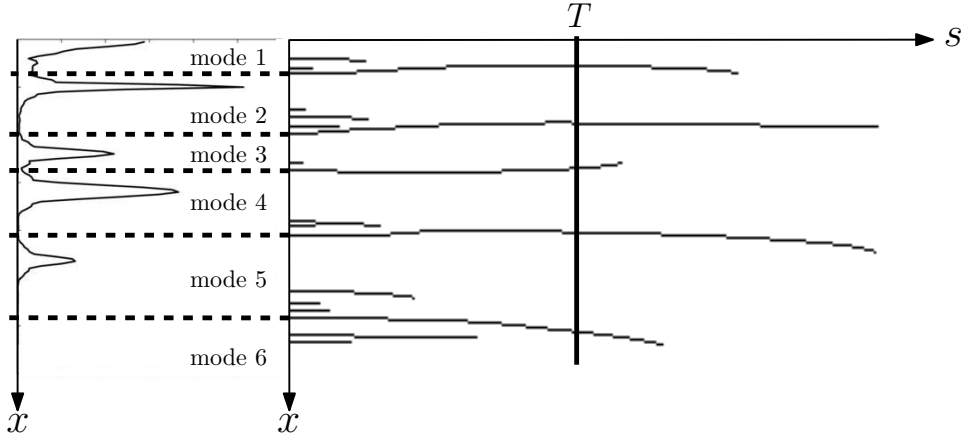


Fig. 2. Histogram modes detection principle: all scale-space curves larger than the threshold T provide boundaries (dashed lines) of supports of different modes.

L_i . The (integer) length L_i is understood as the life span of the minimum i (and not as the arc length of C_i) and is defined by

$$L_i = \max\{s/\text{the } i\text{-th minimum exists}\}$$

We can now define the notion of “meaningful modes” of an histogram.

Definition 1 *A mode in an histogram is called meaningful if its support is delimited by two local minima corresponding to two long (i.e above a certain length T) scale-space curves C_i .*

As a consequence, finding meaningful modes is equivalent to finding a threshold T such that scale-space curves of length larger than T are the curves corresponding to minima delimiting modes’ supports. This principle is illustrated in figure 2. This means that the problem of finding such modes is equivalent to a two class clustering problem on the set $\{L_i\}_{i \in [1, N_0]}$. The following three sections explore independent ways to automatically determine such optimal threshold T . Section 3.2 presents a probabilist approach providing an analytical expression for T , an existing algorithmic method is used in section 3.3 and section 3.4 suggests to use a k -Means ($k = 2$) clustering algorithm to find T .

3.2. Probabilist approach

Probabilistic models are often used to help solve detection problems. In order to use a probabilistic approach within our scale-space framework, we must adapt our definition of “meaningful mode”. The following definition is inspired from the Gestalt theory ideas developed in Desolneux et al.^{8,9}

Definition 2 Given a positive small number ϵ , the i -th local minimum will to be said ϵ -meaningful if the length L_i of its associated scale-space curve C_i is larger than a threshold T . Hence, by considering L_i as a random variable and for a given probability distribution, we have

$$\mathbb{P}(L_i > T) \leq \epsilon. \quad (3.1)$$

Based on this definition, the following proposition gives an explicit expression of T when the considered distribution law is a half-normal distribution.

Proposition 1 Let L_i be independent random variables (where $1 \leq L_i \leq L_{max}$) and ϵ be a positive small number. Define H_L to be the histogram representing the occurrences of the lengths of a scale-space curve, and suppose \mathbb{P} follows a half-normal distribution. Then:

$$T \geq \sqrt{2\sigma^2} \operatorname{erf}^{-1} \left(\operatorname{erf} \left(\frac{L_{max}}{\sqrt{2\sigma^2}} \right) - \epsilon \right)$$

(where $\sigma = \sqrt{\frac{\pi}{2}} \operatorname{Mean}(H_L)$ according to the definition of the half-normal distribution).

Before we give the proof of this proposition, let us comment on the choice of this distribution law. In many problems, a Gaussian law is chosen as the default distribution law. In our context, we know that the random variables L_i are always positives and in many experiments we can observe that they follow a monotone decreasing law hence our choice of the half-normal law (see Azzalini²). However, we emphasize that other exponential laws could also be considered. In the Gestalt theory, the parameter ϵ corresponds to “how much” an event is meaningful. The idea is to fix the value of ϵ to be very small. In this paper we choose the same $\epsilon = 1/n$ for all cases to avoid introducing any new parameters. Of course it is possible to use ϵ as a parameter. This point of view will not be explored in this paper as we seek a parameterless algorithm.

Proof. Considering the half-normal distribution, we have

$$\begin{aligned} \mathbb{P}(L_i > T) &= \int_T^{L_{max}} \sqrt{\frac{2}{\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx \\ &= \left[\int_0^{L_{max}} \sqrt{\frac{2}{\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx - \int_0^T \sqrt{\frac{2}{\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx \right] \\ &= \operatorname{erf} \left(\frac{L_{max}}{\sqrt{2\sigma^2}} \right) - \operatorname{erf} \left(\frac{T}{\sqrt{2\sigma^2}} \right), \end{aligned}$$

then applying (3.1) gives

$$\begin{aligned} \operatorname{erf} \left(\frac{L_{max}}{\sqrt{2\sigma^2}} \right) - \operatorname{erf} \left(\frac{T}{\sqrt{2\sigma^2}} \right) &\leq \epsilon \\ \Leftrightarrow T &\geq \sqrt{2\sigma^2} \operatorname{erf}^{-1} \left(\operatorname{erf} \left(\frac{L_{max}}{\sqrt{2\sigma^2}} \right) - \epsilon \right). \end{aligned}$$

This concludes the proof. \square

3.3. Otsu's method

In Otsu,¹⁹ the author proposed an algorithm to separate an histogram H_L (defined as in the previous section) into two classes H_1 and H_2 . The goal of Otsu's method is to find the threshold T such that the intra-variances of each class H_1, H_2 are minimal while the inter class variance is maximal. This corresponds to finding T (an exhaustive search is done in practice) which maximizes the between class variance $\sigma_B^2 = W_1 W_2 (\mu_1 - \mu_2)^2$, where $W_r = \frac{1}{n} \sum_{k \in H_r} H(k)$ and $\mu_r = \frac{1}{n} \sum_{k \in H_r} k H(k)$, see Otsu¹⁹ for details.

3.4. k -Means

The k -Means algorithm (see Hartigan et al¹²) is a very popular clustering algorithm; it aims to partition a set of points into k clusters. In our context, we apply the k -Means algorithm to the histogram H_L to get the two clusters H_1, H_2 (meaningful/non-meaningful minima). This is equivalent to solving the following problem:

$$(H_1, H_2) = \arg \min_{H_1, H_2} \sum_{r=1}^2 \sum_{H(k) \in H_r} \|H(k) - \mu_r\|^2, \quad (3.2)$$

where μ_r is the mean evaluated over all points in the cluster H_r . In this paper, we experiment with both the ℓ^1 and ℓ^2 norms^a and two types of initialization: random or uniformly distributed. In practice, it is usual to run the k -Means several times and keep the solution that provides the smallest minimum of (3.2) (in our experiments we chose to perform ten iterations).

4. Experiments

In this section, we present several types of experiments to illustrate the relevance of the proposed approach. We first start to present the detected modes in 1D histograms. The used histograms are of two main categories: grayscale image histograms and Fourier spectra (which can be viewed as histograms showing the amount of each frequency). For each experiment we plot the detected boundaries (dashed lines) on the considered histogram.

Then we address the grayscale image segmentation problem which split the image into a certain number of class. Each class corresponds to the pixel grayscale values which lie in a specific modes detected on the image histogram by our algorithm.

Finally, we present a simplified version of the algorithm presented in⁷ to perform some color reduction in a given color image.

^athe ℓ^p norm of a sequence $x = \{x_k\}_{k=1 \dots K}$ is defined by $\|x\|_{\ell^p} = \left(\sum_{k=1}^K |x_k|^p \right)^{1/p}$

	$x16$	$x21$	Sig_1	Sig_2	Sig_3	EEG	Textures
$\ell^2 - k$ -Means (Random)	6	2	1	2	3	11	4
$\ell^2 - k$ -Means (Uniform)	6	2	1	2	3	11	4
$\ell^1 - k$ -Means (Random)	6	2	1	2	3	11	3
$\ell^1 - k$ -Means (Uniform)	6	2	1	2	3	11	3
Otsu	6	3	2	3	4	12	5
Half-Normal law	6	3	2	2	3	30	3

Table 1. Number of detected boundaries per histogram for each detection method.

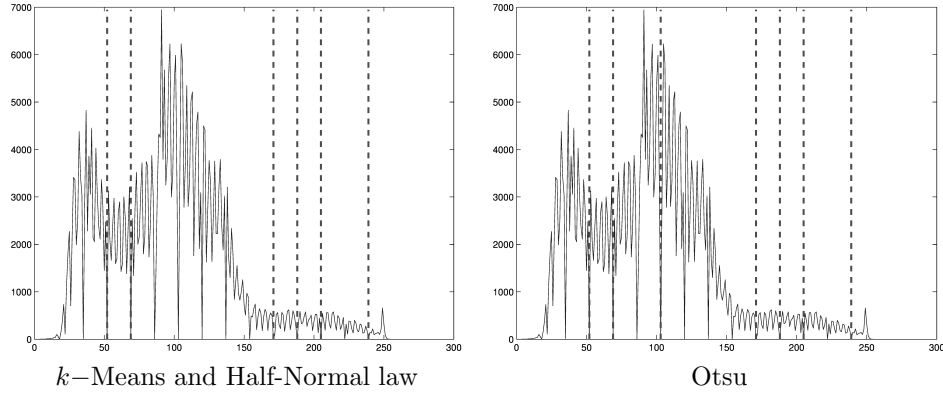
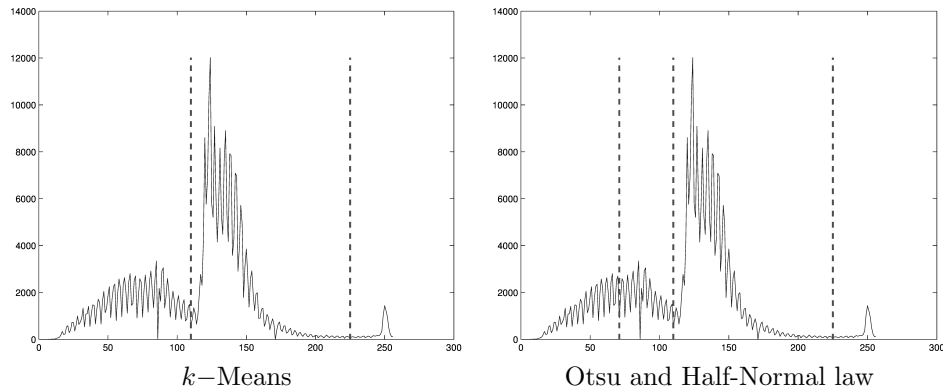
4.1. 1D histogram segmentations

In this section, we present results obtained on 1D histograms by the method described in this paper. In figures 3 and 4, the used histograms correspond to histograms of grayscale values of two input images (denoted, as in the Kodak database, $x16$ and $x21$, respectively). Each histogram contains 256 bins. Figures 5, 6, 7, 9 and 8 depict the segmentation method performed on various Fourier spectra rather than on classical histograms. These spectra were introduced in Gilles¹⁰ and Gilles et al¹¹ with the labels $sig1$, $sig2$, $sig3$, EEG and $Textures$, respectively. Table 1 contains the number of boundaries found by the proposed method for each experiment.

We can observe that for $x16$ and $x21$ all methods give acceptable sets of boundaries. For $sig1$, $sig2$, $sig3$ and $Textures$ spectra, Otsu's method and $\ell^2 - k$ -Means seem to provide the most consistent results throughout the different cases. Except for $x16$ and $Textures$, the half-normal distribution give similar results as Otsu's method. It can also be observed that the type of initialization (uniform or random) of the k -Means algorithm has no influence on the obtained boundaries. Moreover, except for the $Textures$ case, the $\ell^1 - k$ -Means and $\ell^2 - k$ -Means provide exactly the same results. Let us now shift our focus to the special case of EEG spectrum. This spectrum is associated with an electroencephalogram (EEG) signal and is much more complicated than usual spectra. As such it is difficult to have an a priori idea of a relevant number of modes as well as where they may occur. However, we can see that Otsu's method and k -Means give very similar outputs. Notice that, the half-normal distribution generate a significantly higher number of modes.

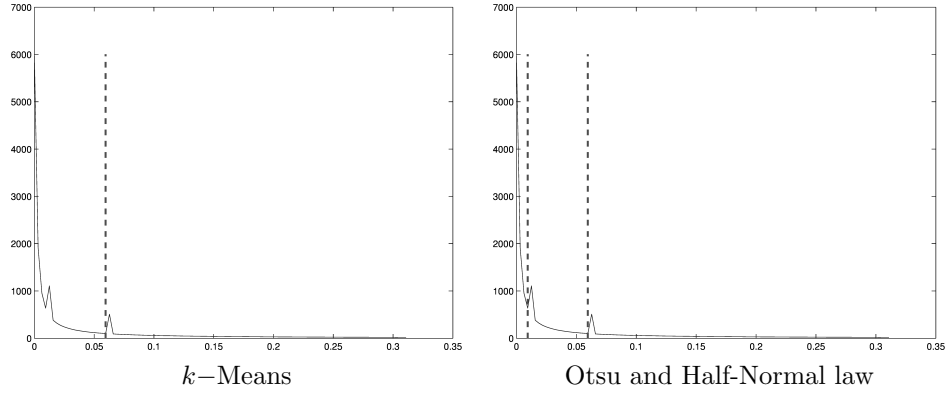
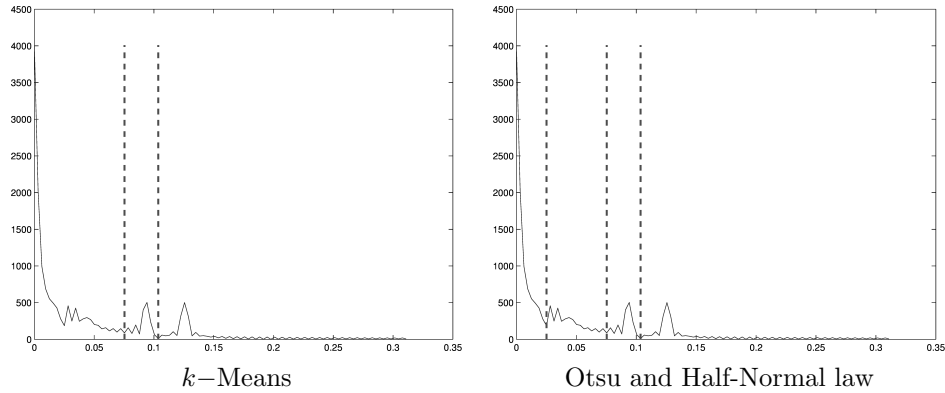
4.2. Grayscale image segmentation

In this section, we illustrate the use of the proposed method for grayscale image segmentation purposes. Here we consider images where classes provided by a segmentation algorithm correspond to distinct modes in the image histogram. Restricting our focus to such images forces us to avoid the segmentation of textured images; as of yet, pixel intensities alone are not sufficient to characterize textures (we refer the reader interested to texture segmentation to Acharyya,¹ Liu¹⁷ and

Fig. 3. Boundaries (dashed lines) for $x16$ image histogram.Fig. 4. Boundaries (dashed lines) for $x21$ image histogram.

Muneeswaran¹⁸). For images following this assumption, this segmentation problem can be easily solved by our method; we segment the grayscale values histogram of the image which will automatically provide a certain number of classes. Based on the previous section on 1D histograms, we choose to use Otsu's method in all following experiments (as well as in the next section). In these experiments, we consider only visual evaluations. A quantitative assessment of this segmentation approach is out of the scope of this paper.

In figures 10 and 11, we present the images corresponding to the previous $x16$ and $x21$ histograms (the original image is on left and the segmented one on right). In both cases, this simple segmentation algorithm gives pretty good results as, for instance, it can separate important features in the images (clouds, sky, house's roof, light house, ...). It is also to notice that main segmentation algorithms in the liter-

Fig. 5. Boundaries (dashed lines) for *Sig1* Fourier spectrum.Fig. 6. Boundaries (dashed lines) for *Sig2* Fourier spectrum.

ature need to, a priori, know the number of classes while our method automatically finds the number of classes.

4.3. Image color reduction

In Delon et al,⁷ the authors use their histogram segmentation algorithm to reduce the number of colors in an image. Their method uses the following steps: first the image is converted into the HSV color system. Secondly, a first segmentation is obtained by segmenting the histogram of the V component. Thirdly, for each previous obtained class they segment the corresponding S's histograms. This step gives them a more refined set of color classes. Finally, they perform the same step, but on the H's histograms of each new class. The final set of color classes is provided as an initialization to a k -Means algorithm that performs the final color extraction. In

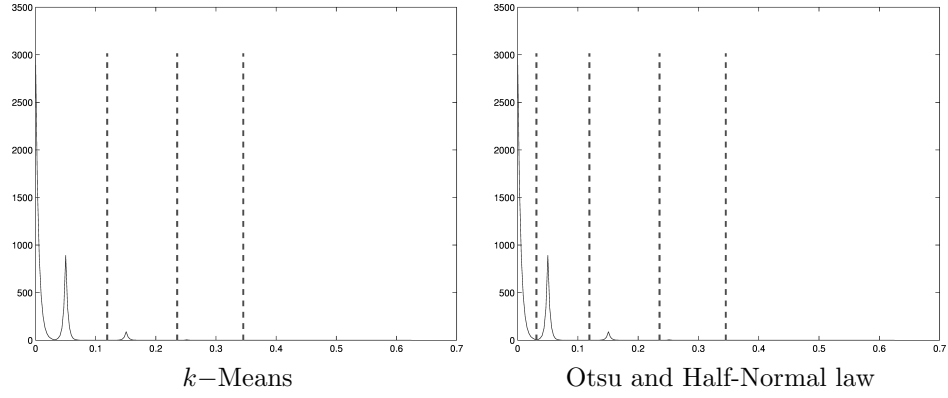


Fig. 7. Boundaries (dashed lines) for *Sig3* Fourier spectrum.

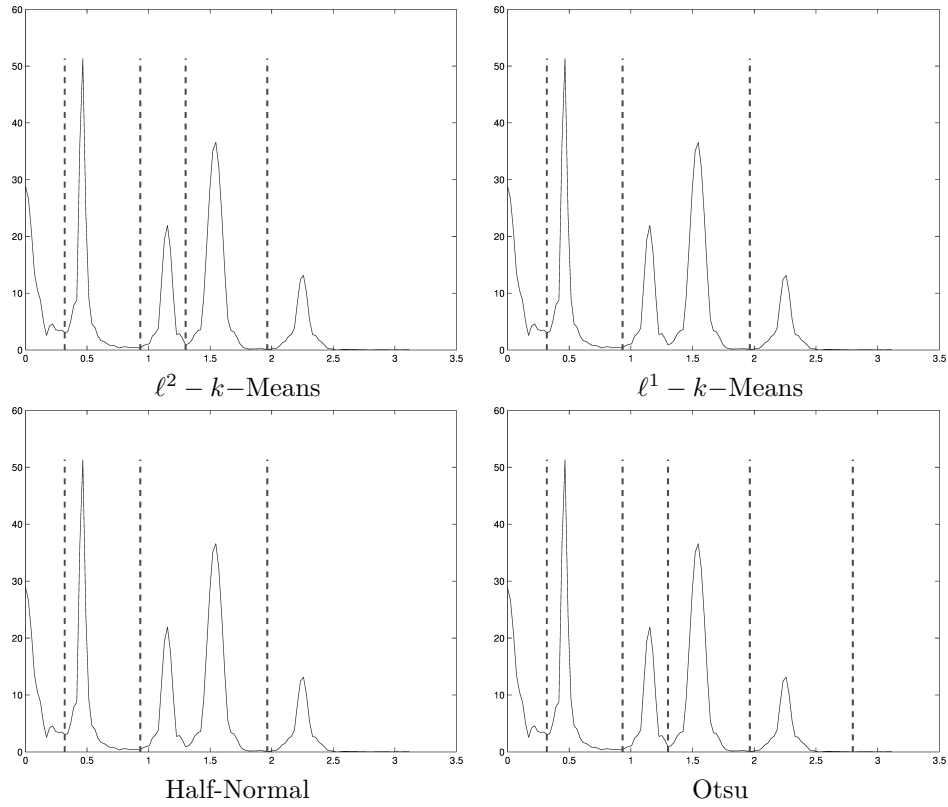


Fig. 8. Boundaries (dashed lines) for *Textures* Fourier spectrum.

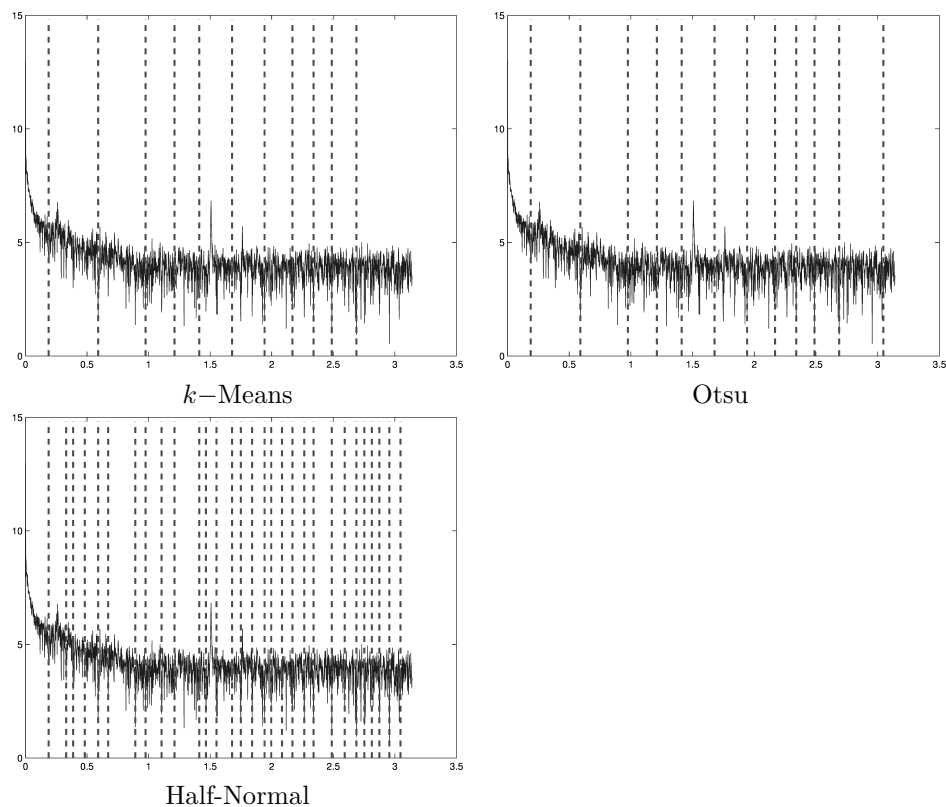


Fig. 9. Boundaries (dashed lines) for *EEG* Fourier spectrum.

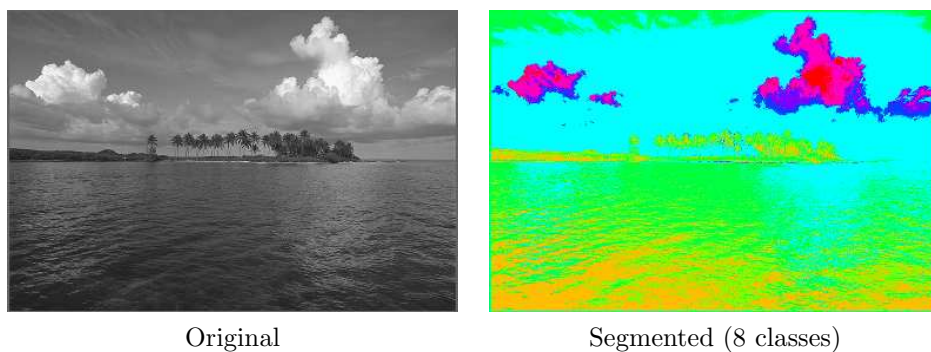


Fig. 10. Grayscale image segmentation: $x16$ case. The corresponding histogram and its set of modes are given in figure. 3 left.

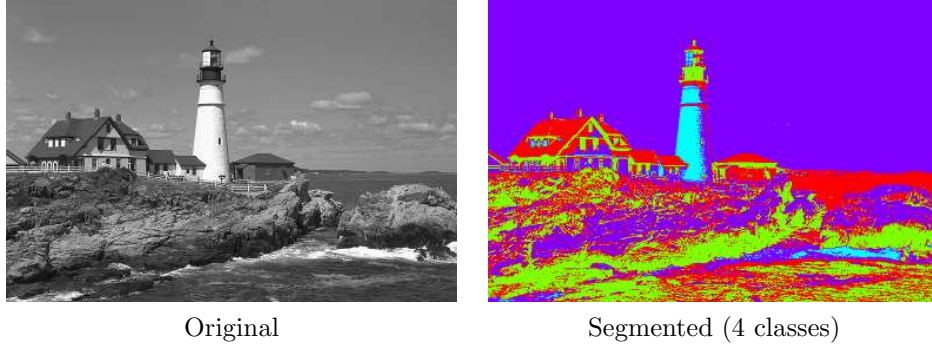


Fig. 11. Grayscale image segmentation: $x21$ case. The corresponding histogram and its set of modes are given in figure. 4 right.

practice, the HSV conversion and the segmentation of the V's histogram is sufficient to keep the most important colors in the image. In the presented experiments, we consider only the first step of the previous method: we apply our histogram segmentation approach to the V component to get a new \tilde{V} and then recompose a color image from the $HS\tilde{V}$ representation.

In figure 12, 13, 14 and 15, we show some example of such color reduction. We can see that this simple algorithm performs very well by reducing the number of colors used in the image but still retains the image's significant features.



Fig. 12. Color reduction: $c15$ case. The V component is reduced to four classes.

5. Conclusion

In this paper, we proposed a very simple and fast method to find meaningful modes in an histogram or a spectrum. The algorithm is based on the consistency of local minima in a scale-space representation. We show with several experiments that this



Fig. 13. Color reduction: c16 case. The V component is reduced to four classes.



Fig. 14. Color reduction: c21 case. The V component is reduced to five classes.



Fig. 15. Color reduction: c22 case. The V component is reduced to four classes.

method efficiently finds such modes. We also provide straightforward image segmentation and color reduction results. It would be of great interest to perform a

complete evaluation of the proposed segmentation method on well-known datasets with quantitative metrics rather than using visual inspections.

A future avenue of inquiry could be characterizing the behavior of the scale-space curves compared to the intrinsic characteristics of the input histogram. We are particularly interested in the application of this new method to EEG spectra, as success with EEG could have profound implications within the neuroscience community. One noteworthy result of our research is that the Scale-Space method yields a segmentation on EEG spectra that for the most part agrees with the traditionally-employed set of neural spectral bands (delta, theta, alpha, beta, gamma). Next our research will move towards performing our algorithm on a large dataset of EEG signals.

Finally, it will be interesting to extend the proposed approach to find higher dimensional modes in larger dimension histograms or spectra.

Acknowledgements

This work was done while the first author was with the Department of Mathematics at UCLA and is partially funded by the following grants NSF DMS-0914856, ONR N00014-08-1-119, ONR N00014-09-1-360, the UC Lab Fees Research and the Keck Foundation. All images are part of the Kodak dataset.¹³ The authors thank the reviewers for their suggestions which permit to improve this manuscript.

References

1. M. Acharyya and M. K. Kundu, Extraction of Noise Tolerant, Gray-Scale Transform and Rotation Invariant Features for Texture Segmentation Using Wavelet Frames, *International Journal of Wavelets, Multiresolution and Information Processing*, **6**(3) (2008), 391–417.
2. A. Azzalini, Further results on a class of distributions which includes the normal ones, *Statistica*, **46** (1986), 199–208.
3. K. Bhoyar and O. Kakde, Color image segmentation based on JND color histogram, *International Journal of Image Processing*, **3**(6) (2010), 265–384.
4. J. M. Buhmann and J. Puzicha, Unsupervised learning for robust texture segmentation, *Performance Characterization in Computer Vision, Computational Imaging and Vision Series*, **17** (2000), 195–209.
5. T. Chan and S. Esedoglu and K. Ni, Histogram Based Segmentation Using Wasserstein Distances, *Scale Space and Variational Methods in Computer Vision, Lecture Notes in Computer Science Series*, **4485** (2007).
6. J. Delon and A. Desolneux and J-L. Lisani and A-B. Petro, A Nonparametric Approach for Histogram Segmentation, *IEEE Transactions on Image Processing*, **16**(1) (2007), 253–261.
7. J. Delon and A. Desolneux and J-L. Lisani and A-B. Petro, Automatic color palette, *Inverse Problems and Imaging*, **1**(2) (2007), 265–287.
8. A. Desolneux and L. Moisan and J-M. Morel, From Gestalt Theory to Image Analysis: A Probabilistic Approach, *Springer-Verlag, Interdisciplinary Applied Mathematics*, **34** (2008).
9. A. Desolneux and L. Moisan and J-M. Morel, Maximal Meaningful Events and Applications to Image Analysis, *Annals of Statistics*, **31**(6) (2003), 1822–1851.

10. J. Gilles, Empirical Wavelet Transform, *IEEE Transactions on Signal Processing*, **61**(16) (2013), 3999–4010.
11. J. Gilles and G. Tran and S. Osher, 2D Empirical transforms. Wavelets, Ridgelets and Curvelets Revisited, *SIAM Journal on Imaging Sciences*, **7**(1) (2014), 157–186.
12. J.A. Hartigan and M.A. Wong, A K-Means Clustering Algorithm, *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **28**(1) (1979), 100–108.
13. Kodak, Kodak Lossless True Color Image Suite, <http://r0k.us/graphics/kodak/>
14. F. Kurugollu and B. Sankur and A.E. Harmanci, Color image segmentation using histogram multithresholding and fusion, *Image and Vision Computing*, **19**(13) (2001), 915–928.
15. T. Lindeberg Scale-space theory in computer vision, Springer (1994).
16. T. Lindeberg Scale-space for discrete signals, *IEEE Transactions of Pattern Analysis and Machine Intelligence*, **12**(3) (1990), 234–254.
17. J. Liu and Y. Y. Tang, Adaptive Image Segmentation With Distributed Behavior-Based Agents, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, **21**(6) (1999), 544–551.
18. K. Muneeswaran and L. Ganesan and S. Arumugam and P. Harinarayan, A Novel Approach Combining Gabor Wavelet and Moments for Texture Segmentation, *International Journal of Wavelets, Multiresolution and Information Processing*, **3**(4) (2005), 559–572.
19. N. Otsu, A threshold selection method from gray-level histograms, *IEEE Transaction on Systems, Man and Cybernetics*, **9**(1) (1979), 62–66.
20. J. Puzicha and T. Hofmann and J. M. Buhmann, Histogram Clustering for Unsupervised Image Segmentation, *Proceedings of CVPR'99* (1999), 602–608.
21. S. Sural and G. Qian and S. Pramanik, Segmentation and histogram generation using the HSV color space for image retrieval, *Proceedings of International Conference on Image Processing (ICIP)*, **2** (2002), 589–592.
22. A. P. Witkin, Scale-space filtering: A new approach to multi-scale description, *IEEE International Conference on Acoustic, Speech, and Signal Processing (ICASSP)*, **9** (1984), 150–153.
23. R. Yildizoglu and J-F. Aujol and N. Papadakis, A Convex Formulation for Global Histogram Based Binary Segmentation, *9th International Conference on Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR)*, Lund (Sweden), (2013).
24. J. Yuan and D. Wang and R. Li, Image segmentation using local spectral histograms and linear regression, *Pattern Recognition Letters*, **33**, 615–622, 2012.