# AN OVERVIEW OF IMAGE SEGMENTATION TECHNIQUES

IN Fabris-Rotelli[1] and J-F Greeff*[1]

[1]Department of Statistics, University of Pretoria, 0002, Pretoria, South Africa,
inger.fabris-rotelli@up.ac.za

## ABSTRACT

We present an overview of some image segmentation techniques, employed to extract regions of interest. Examples, with comparisons, are presented for Iterative Selection, Balanced Histogram, Otsu's method, Wellner algorithm, Integral Image algorithm, Gaussian mixtures and Iterated Conditional Modes.

## 1. INTRODUCTION

Image segmentation plays an integral part within computer vision, since the identification of regions of interest is usually the first step in extracting useful information from an image. We provide an overview of possible methods of segmenting an image, including examples of results. Applications of segmentation techniques include medical imaging (Pham et al., 2000), image compression (Vaisey & Gersho, 1992), facial recognition (Jain & Park, 2009), scene classification (Debba et al., 2008), handwriting recognition (Shapiro & Stockman, 2002) and lane detection (Sezgin & Sankur, 2004). A good overview of applications can be found in Shapiro & Stockman (2002).

Many methods exist that aim at segmenting images (Shapiro & Stockman, 2002; Mobahi et al., 2011): thresholding to create a black and white (BW) image; clustering iterative methods, such as $k$-means; compression-based segmentation to minimise the data coding length; histogram-based methods to identify segments; edge detection; region growing to group neighbouring pixels with similar luminosities; and split-and-merge methods where the image is repeatedly split into heterogeneous squares and merged into homogeneous squares. These methods are not mutually exclusive, Otsu's method is both a histogram-based and clustering method.

Throughout the paper we will refer to an image matrix of pixel luminosities of size $n \times m$ as $I$, where $I(i, j)$ is the luminosity of pixel $(i, j)$ with $i = 1, 2, ..., n$ and $j = 1, 2, ..., m$, noting that $I\left(i, j\right)$ could be a vector of luminosity values.

*J-F Greeff is the corresponding and presenting author. Thanks to the Department of Statistics and Statomet for support and funding.

## 2. THRESHOLDING

*Thresholding* creates BW images from gray-scale images. In its simplest form, all pixels with intensities below an appropriately chosen threshold level $T$ are considered to be background pixels, represented by white pixels; all pixels with intensities above $T$ are considered the foreground, represented by black pixels. This is termed threshold above, while threshold below considers pixels below $T$ to be the foreground (Shapiro & Stockman, 2002). Other thresholding techniques include band thresholding, $p$-tile thresholding, optimal thresholding and adaptive thresholding (Henden, 2004). Standard thresholding suffers from lack of robustness in the presence of non-stationary and correlated noise, ambient illumination, variation of intensity levels within the objects and the background, inadequate contrast, the object size relative to the image size, and a lack of an objective performance measure (Sezgin & Sankur, 2004). Suggested solutions include noise reduction techniques (Nachtegael et al., 2001), adaptive thresholding, and cropping of the image to make the object the focus of the image.

Ridler & Calvard (1978) present a special case of $k$-means clustering, called *Iterative Selection* (IS), to find an optimal $T$. The algorithm is as follows **1.** Assume some initial binary segmentation, **2.** calculate the mean luminosity of the foreground, $\mu_b^{(0)}$, and background, $\mu_f^{(0)}$, **3.** calculate the threshold as $T^{(0)} = \frac{1}{2}\left(\mu_b^{(0)} + \mu_f^{(0)}\right)$, **4.** segment the image using $T_0$, **5.** repeat steps 2 to 4, where $\mu_b^{(\alpha)}$, $\mu_f^{(\alpha)}$ and $T^{(\alpha)}$ are the mean luminosities and threshold value at iteration $\alpha$, until some predetermined stop criteria is met.

The *Balanced Histogram* (BH) method is presented by Anjos & Shahbazkia (2008). It is a recursive model in which the histogram is balanced to find a weighted mid-point of the grayscale luminosities in the set $\mathscr{I} = \{I(i,j)\,\forall\,(i,j)\}$[1]. Define $\mathscr{I}_l^{(0)} = \left\{I(i,j) \in \mathscr{I} : I(i,j) < m^{(0)}\right\}$ and $\mathscr{I}_u^{(0)} = \left\{I(i,j) \in \mathscr{I} : I(i,j) \geq m^{(0)}\right\}$ for some $m^{(0)}$, and construct the image histogram as $f(t) = \frac{\text{card}\{(i,j):I(i,j)=t\}}{n \times m}$ for all $t \in \mathscr{I}$ The method can be summarised as **1.** choose an initial point $m^{(0)}$ for the weighted mid-point of the histogram domain, **2.** calculate the weights of the histogram on either side of $m^{(0)}$ as $w_l = \sum_{t \in \mathscr{I}_l^{(0)}} f(t)$ and $w_u = \sum_{t \in \mathscr{I}_u^{(0)}} f(t)$, **3.** remove weight from the outer-edge of the heavier side of the histogram, that is if $w_l < w_u$, then let $\mathscr{I}_u^{(1)} = \mathscr{I}_u^{(0)} \setminus \max \mathscr{I}_u^{(0)}$ and $\mathscr{I}_l^{(1)} = \mathscr{I}_l^{(0)}$, or if $w_l \geq w_u$, then let $\mathscr{I}_l^{(1)} = \mathscr{I}_l^{(0)} \setminus \min \mathscr{I}_l^{(0)}$ and $\mathscr{I}_u^{(1)} = \mathscr{I}_u^{(0)}$, **4.** recalculate the mid-point $m^{(1)}$ of the new histogram domain $\mathscr{I}^{(1)} = \mathscr{I}_l^{(1)} \bigcup \mathscr{I}_u^{(1)}$, **5.** repeat steps 3 to 5, until $\mathscr{I}^{(\alpha)}$ is a single value which is then taken to be the threshold for this image.

*Otsu's Method* (OM) is popular for selecting the optimal threshold by minimising the within-cluster luminosity variance within the foreground and background clusters. It can be shown that this is equivalent to maximising the between-cluster variance, given by $\sigma_B^2(T) = p_1(T)\,p_2(T)\,[\mu_1(T) - \mu_2(T)]^2$, where $f(t)$ is the image histogram, $p_1(T) = \sum_{t<T} I(t)$, $p_2(T) = 1 - p_1(T)$, $\mu_1(T) = \sum_{t<T} t f(t)$, and $\mu_2(T) = \sum_{t \geq T} t f(t)$. This reduces the effort needed to obtain the optimal threshold by only requiring estimation of cluster means and cluster relative frequency (Otsu, 1979; Shapiro & Stockman, 2002). Together with its

[1]The definition of $\mathscr{I}$ used within this algorithm overcomes the problem of extreme values within the domain of the histogram, which may cause the algorithm to converge to an incorrect solution.

effectiveness this method is popular in practice and literature and is included in many software packages.

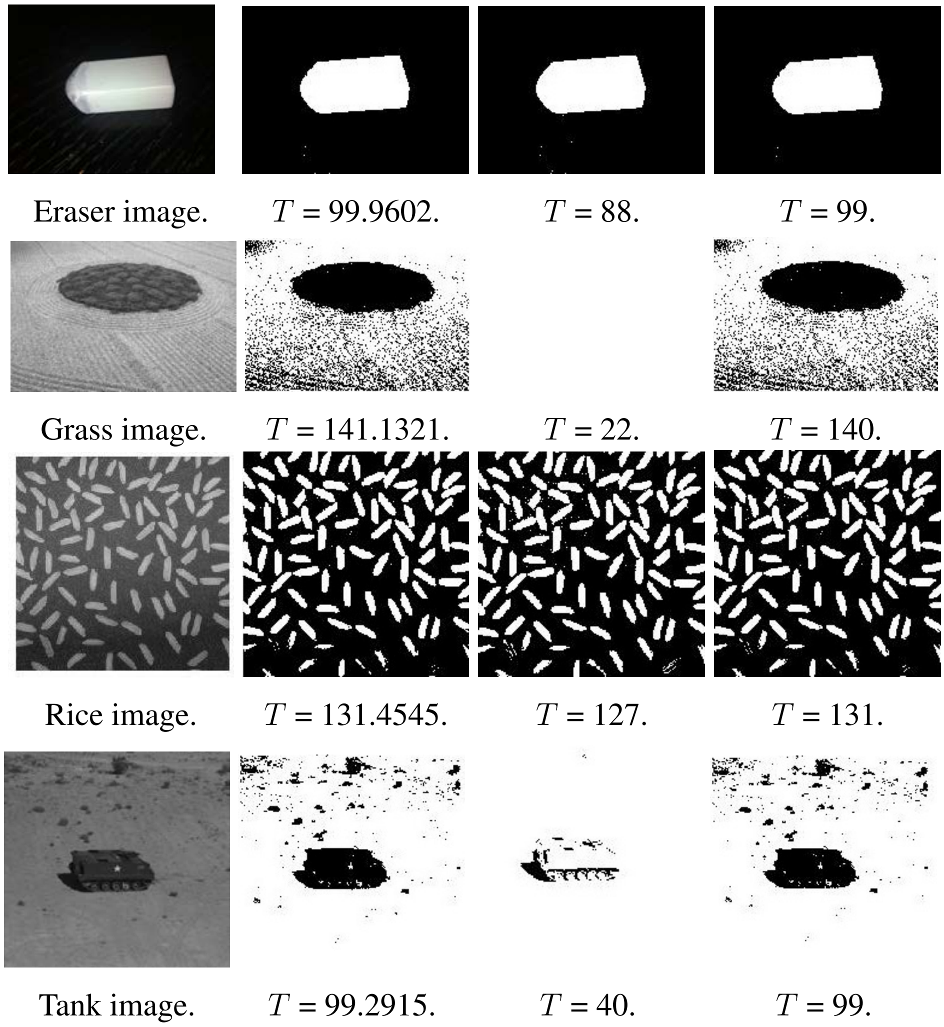



**Fig. 1**: Image results for each original image (left) based on Iterative Selection (center-left), Balanced Histogram (center-right) and Otsu's method (right). The threshold in each case is $T$.

In Figure 1 we see that BH underperformed compared to OM and IS and that $T$ for each image is significantly different. This can be attributed to the simplicity of BH, which does not take into account the skewness of the image histograms. It is possible to obtain a less skewed distribution by cropping the image but this requires user interaction. In the rice image this problem does not occur due to the approximately equal number of light and dark pixels. We see that the thresholds and results obtained using IS and OM are similar. For the eraser image the observed segmentation is very good due to the stark background/foreground contrast, cluster luminosity homogeneity, and lack of ambient illumination. The textured grass image presents a bigger challenge due to greater cluster luminosity variability. The application of a noise reduction filter (Lim, 1990) to the segmentation provides an improvement by removing the texture effect. The rice image illustrates the issue with ambient light. The grain edges near the bottom edge are not as clearly defined as those in the centre due to a contrast reduction by the shadow. Adaptive thresholding (Section 3) provides a solution. The tank image shows the effect of the relative size of foreground and background, as well as the effect of lack of contrast. When applied to a cropped tank image, a better segmentation is obtained; because large sections of the misclassified shrubbery, with similar luminosities to the tank, are removed, and because the size of the tank relative to the background is increased.

## 3. ADAPTIVE THRESHOLDING

In *Adaptive thresholding* the threshold changes according to the position of the pixel being thresholded. The threshold may be different for each individual pixel, or constant over regions in the image; enabling robustness to ambient light changes (Wellner, 1993). We provide an overview of two algorithms, the Wellner algorithm (WA) and the Improved Image algorithm (IIA).

In Wellner (1993)'s algorithm a pixel's threshold is determined as a fraction of the average neighbourhood pixel luminosity. Define $s \in \mathbb{N}$ to be the size of the neighbourhood and $t \in (0,1)$ as the relative threshold parameter. Wellner discusses methods for determining the average pixel luminosity in each neighborhood, and how to deal with image boundaries. However, the Wellner algorithm only considers pixels in the same row meaning information above and below the pixel is ignored. Define the neighbourhood of $(i,j)$ as $N_s(i,j) = \{(a,b) \in I : a \in \left[i-\frac{s}{2}, i+\frac{s}{2}\right], b = j\}$ and the average neighbourhood luminosity as $\mu_s(i,j) = \frac{\sum_{(i,j)\in N_s(ij)} I(ij)}{\text{card}(N_s(ij))}$. A pixel is classified as background (foreground) if $I(i,j) < (1-t)\,\mu_s(i,j)$ $(I(i,j) \geq (1-t)\,\mu_s(i,j))$. Wellner (1993) and our own investigation found that $s = \frac{m}{8}$ and $t = 0.15$ provide satisfactory results for a range of images. Bradley & Roth (2007) propose an improvement on WA, the Improved Image algorithm (IIA). They calculate the average luminosity within a rectangular neighbourhood around the pixel to be thresholded. To lighten the computational load they introduce the integral image $J(i,j) = \sum_{a\leq i}\sum_{b\leq j} I(a,b)$ and compute $J$ only once. They redefine the neighbourhood $N_s(i,j) = \{(a,b) \in I : a \in [i-s/2, i+s/2], b \in \left[j-\frac{s}{2}, j+\frac{s}{2}\right]\}$ with $i_1(i_2) = \min_a(\max_a)N_s(i,j)$, $j_1(j_2) = \min_b(\max_b)N_s(i,j)$, so that the average neighbourhood luminosity is $\mu_s(i,j) = \frac{J(i_2,j_2)-J(i_2,j_1)-J(i_1,j_2)+J(i_1,j_1)}{(i_2-i_1)(j_2-j_1)}$.

In Figure 2 we illustrate the marked improvement of IIA over WA and OM, with two images subject to ambient light. The WA seems more robust to ambient light. Misclassification does still occur near the shadow borders and areas with a long run of similar luminosities. The latter is addressed by IIA. A clear improvement is seen when comparing the results for WA and IIA. The misclassification in the top and bottom sections of the QR codes are not observed in the IIA results, and the misclassification in the centre along the shadow edges disappears. Overall, the QR codes are more accurately represented, even near the shadow edges. In the journal image, the line below the title is more accurately segmented, with no missing sections. The text is generally more clearly defined and the dots running perpendicular to the title are mostly removed. One important aspect to highlight is the fact that both WA and IIA extract information not visible to the human eye. This is clearly seen in the journal image, where initially not visible parts of the title and text are extracted. In regions where shadows create a uniform band of luminosities, a slight deviation in luminosity still triggers a segmentation result, making the technique useful in data recovery or image compression. An investigation into a smaller neighbourhood size, $s$, and larger relative threshold, $t$, could yield valuable insight into extracting such information within these regions. Bradley & Roth (2007) showed that IIA is approximately 2.5 times slower than WA, but this is negligible in practice.

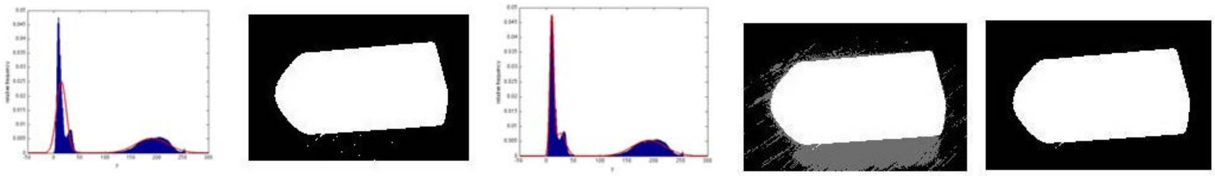
**Fig. 2**: Image results for each image (left) based on Otsu's method (center-left), the Wellner algorithm (center-right) and the Integral Image algorithm (right). (Journal image from Bradley & Roth (2007).)

## 4. GAUSSIAN MIXTURES

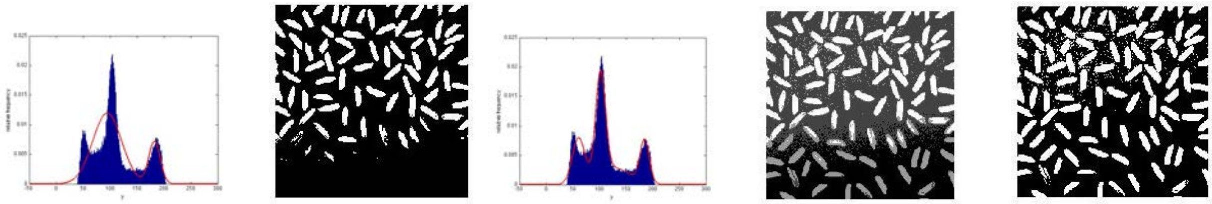
*Gaussian Mixtures* (GM) are a special case of mixture distributions, described by McLachlan & Peel (2000), in which a distribution is obtained by superimposing two or more normal distributions by means of a linear transformation. The probability density function of such a GM is given by $f(\boldsymbol{x}) = \sum_{k=1}^{K} \pi_k f_k(\boldsymbol{x}) = \sum_{k=1}^{K} \pi_k (2\pi)^{-\frac{d}{2}} |\Sigma_k|^{-\frac{1}{2}} \exp\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_k)' \Sigma_k^{-1} (\boldsymbol{x}-\boldsymbol{\mu}_k)\}$ for $\boldsymbol{x} \in \mathbb{R}^d$ where each of the $K$ components of $f$ are normally distributed in $d$-dimensions, with mean vector $\boldsymbol{\mu}_k$ and covariance matrix $\Sigma_k$; and have mixing coefficients $\pi_k \in [0, 1]$, such that $\sum_{k=1}^{K} \pi_k = 1$. By altering $K$, we are able to fit $f$ to any continuous distribution with an arbitrary degree of accuracy (Bishop, 2006). In grayscale image segmentation we take $d = 1$ and $K$ the number of clusters to identify; whereas for more general scenarios, such as colour images, we let $d \geq 1$, specifically $d = 3$ for RGB images. The problem is then to fit $f$ to the observed pixel luminosities by estimating the $3K$ parameters of $f$, through maximum likelihood estimation and expectation-maximisation estimation (Bishop, 2006). The responsibilities of pixel $(i, j)$ are given by $\gamma(z_k) = E[z_k | I(i, j)] = \frac{\pi_k f_k(I(i,j))}{\sum_{i=1}^{K} \pi_i f_i(I(i,j))}$ for $k = 1, 2, ..., K$, that is, the conditional expected values of the elements $z_1, z_2, ..., z_K$, indicating to which cluster the pixel belongs, given the pixel's luminosity. Pixel $(i, j)$ is clustered to the cluster with the highest responsibility (Bishop, 2006).

Figure 3 presents image histograms with the fitted GM overlaid, and the image segmentation. For each image on the left $K = 2$ was used, while for those on the right $K = 3$ (eraser image) and $K = 4$ (rice image) was used. In the cases of $K > 2$ we have also grouped two or more clusters to form a BW image with only two clusters. The results of the eraser image with $K = 2$ are similar to those obtained using OM. For $K = 3$, however, we observe that the GM extracts the grain of the wooden desk from the background. There are three distinct peaks in the distribution for the image histogram, implying at least three clusters. Merging the two clusters removes almost all of the misclassified pixels around the bottom of the image.

We see that GM with $K=2$ underperforms OM in the rice image, due to the ambient light. The histogram displays three distinct peaks, and a section of the domain with a nearly constant frequency, suggesting at least four clusters in the image, supported by the well-fitting GM, with $K=4$. Intuitively, this is expected since there are four natural clusters - the grains, background, and inside and outside the shadow. In this case the GM outperforms even the IIA which is specifically designed to deal with ambient light.



$\boldsymbol{\pi} = (0.59; 0.41)'$ and $\boldsymbol{\mu} = (15; 190)'$. $\boldsymbol{\pi} = (0.41; 0.18; 0.41)'$ and $\boldsymbol{\mu} = (10; 27; 191)'$.



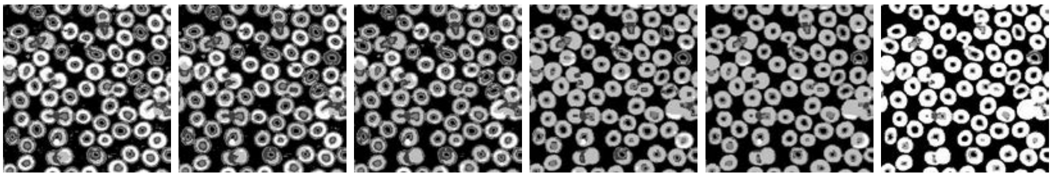
$\boldsymbol{\pi} = (0.81; 0.19)'$ and $\boldsymbol{\mu} = (95; 183)'$. $\boldsymbol{\pi} = (0.23; 0.4; 0.22; 0.16)'$ and $\boldsymbol{\mu} = (61; 102; 137; 185)'$.

**Fig. 3**: Histogram of the pixel luminosities with the fitted GM (left) and the image results (right). In the case of more than 2 clusters, the appropriate clusters are merged to create two clusters (far right).

## 5. ITERATED CONDITIONAL MODES

*Iterated Conditional Modes* (ICM) is an algorithm introduced by Besag (1986) to reduce the noise in dirty pictures. It takes into account both features of each pixel and spatial information based on a Markov Random Field of each pixel to be clustered (Debba et al., 2008). ICM, within the context of noise removal, is based on the assumption that neighbouring pixels tend to have similar luminosities, or other features, and that each pixel is corrupted independently with a given probability. Consider an image $I$ in which there are $K$ clusters of pixels which we would like to detect and extract. Then, for each iteration of the algorithm, indexed by $\alpha$, define $\omega_{ij}^{(\alpha)}$ the cluster of pixel $\boldsymbol{x}=(i,j)$; $C_k^{(\alpha)}=\left\{\boldsymbol{x} : \omega_{ij}^{(\alpha)}=k\right\}$ the set of pixels belonging to cluster $k$; $N_k^{(\alpha)} = \text{card}\left(C_k^{(\alpha)}\right)$ the number of pixels in cluster $k$; $N_{ij}^{(\alpha)}(k) = \text{card}\left(C_k^{(\alpha)} \cap N(\boldsymbol{x})\right)$ the number of pixels in a neighbourhood $N(\boldsymbol{x})$, of $\boldsymbol{x}$ in cluster $k$; $\boldsymbol{\mu}_k^{(\alpha)} = \frac{1}{N_k^{(\alpha)}} \sum_{\boldsymbol{x}\in C_k^{(\alpha)}} I(\boldsymbol{x})$ the $d$-dimensional mean vector of cluster $k$; $\nu^{(\alpha)} = \frac{1}{n\times m} \sum_{k=1}^{K} \sum_{\boldsymbol{x}\in C_k^{(\alpha)}} \left(I(\boldsymbol{x})-\boldsymbol{\mu}_k^{(\alpha)}\right)' \left(I(\boldsymbol{x})-\boldsymbol{\mu}_k^{(\alpha)}\right)$ the total within-cluster variance. ICM minimises the total within-cluster variance, by assigning and reassigning each pixel in the image to a class, while taking spatial information into account. We proceed as follows, **(1)** initialise the parameters using only the luminosity of each pixel. We used a multivariate $K$-means clustering procedure (see Hartigan & Wong (1979)); **(2)** calculate $C_k^{(\alpha)}$, $N_k^{(\alpha)}$, $\boldsymbol{\mu}_k^{(\alpha)}$ and $\nu^{(\alpha)}$ for each $k = 1, 2, ..., K$; **(3)** calculate $\Lambda(\boldsymbol{x}, k) = \left(I(\boldsymbol{x})-\boldsymbol{\mu}_k^{(\alpha)}\right)' \left(I(\boldsymbol{x})-\boldsymbol{\mu}_k^{(\alpha)}\right) - \beta\nu^{(\alpha)} N_{ij}(k)$ and find $k^* = \arg\min_k\{\Lambda(\boldsymbol{x}, k)\}$ for all $\boldsymbol{x}$; **(4)** set $\omega_{ij}^{(\alpha+1)} = k^*$ for each pixel in the image to reclassify the pixel to cluster $k^*$; **(5)** repeat 2 through 4 until $C_k^{(\alpha+1)} = C_k^{(\alpha)}$ for all $k = 1, 2, ..., K$ or until some predetermined stop criterion is met. If the algorithm converges the sets $C_k$ are the clusters of pixels, grouped according to their likely cluster membership, taking into account the spatial location of the pixel.

The function $\Lambda$ is similar to the function minimised within the $K$-means framework, but with a second term, called the spatial penalisation term (Debba et al., 2008), allowing for the inclusion of spatial information. In effect, we are reducing the within-cluster sum of square deviations by a multiple of the number of pixels in the neighbourhood which are in the cluster $k$. Since if a pixel is surrounded by many pixels which belong to cluster $k^*$, it is more likely that that pixel also belongs to cluster $k^*$, due to our first assumption, we increase the likelihood that $k^*$ will minimise $\Lambda$, by reducing the size of $\Lambda(\boldsymbol{x}, k^*)$ by a constant related to our spatial information. Debba et al. (2008) suggest that a good choice for the parameter $\beta$ is 1.5. A larger value leads to a smoother image based more heavily on spatial data, while a lower value leads to a clustering similar to $K$-means. Figure 4 illustrates the effect of the $\beta$ parameter. Overall, it was found that ICM is an effective segmentation method due to the inclusion of spatial information as well as well as more robust than adaptive thresholding.



$\beta = 0.5$ $\beta = 1$ $\beta = 1.5$ $\beta = 2$ $\beta = 2.5$ $\beta = 5$

**Fig. 4**: Comparison of the effect of the $\beta$ parameter in ICM with $K = 5$.

## 6. CONCLUSION

We have presented an overview of image segmentation techniques for grayscale images. Iterative Selection, the Balanced Histogram, Otsu's method, the Wellner algorithm, the Integral Image algorithm, Gaussian mixtures and Iterated Conditional Modes were discussed. We conclude for general segmentation Otsu's method is the preferred technique. It is, however, not robust in the presence of ambient light, in which case we recommend the use of the Integral Image algorithm. A more robust alternative for including spatial information is provided by the Iterated Conditional Modes algorithm. The Gaussian mixtures algorithm provides an effective method for fitting a distribution to the image histogram. It provides a good segmentation, especially when multiple clusters are identified and grouped. Due to their short processing times the BH and WA are useful as an initial segmentation for more complex methods.

## REFERENCES


Anjos, A. & Shahbazkia, H. (2008). Bi-Level Image Thresholding - A Fast Method. *Biosignals*, 2, 70–76.

Besag, J. (1986). On the statistical analysis of dirty pictures. *J R Stat Soc*, 48(3), 259–302.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer, first edition.

Bradley, D. & Roth, G. (2007). Adaptive Thresholding using the Integral Image. *J Graphics, GPU, Game Tools*, 12(2), 13–21.

Debba, P., Stein, A., van der Meer, F., Carranza, E., & Lucieer, A. (2008). Field Sampling from a Segmented Image. In *ICCSA 2008*, volume 5072 of *Lect Notes Comput Sc* (pp. 756–768). Springer Berlin / Heidelberg.

Hartigan, J. & Wong, M. (1979). A K-means clustering algorithm. *Appl Stat-J Roy St C*, 28, 100–108.

Henden, P. (2004). Exercise in Computer Vision. unpublished manuscript, NTNU Faculty of IT.

Jain, A. & Park, U. (2009). Facial marks: Soft biometric for face recognition. In *P 16th IEEE Int Conf Image Process* (pp. 37–40).

Lim, J. S. (1990). *Two-Dimensional Signal and Image Processing*. Englewood Cliffs, N.J.: Prentice Hall.

McLachlan, G. J. & Peel, D. (2000). *Finite Mixture Models*. New York: Wiley.

Mobahi, H., Rao, S., Yang, A., Sastry, S., & Ma, Y. (2011). Segmentation of Natural Images by Texture and Boundary Compression. *Int J Comput Vision*, 95(1), 86–98.

Nachtegael, M., Van der Weken, D., Van De Ville, D., Kerre, E., Philips, W., & Lemahieu, I. (2001). An overview of classical and fuzzy-classical filters for noise reduction. In *10th IEEE Int Conf Fuzzy Sys*, volume 1 (pp. 3–6).

Otsu, N. (1979). A Threshold Selection Method from Gray-Level Histograms. *IEEE T Syst Man Cyb*, SMC-9(1), 62–66.

Pham, D., Xu, C., & Prince, J. (2000). Current Methods in Medical Image Segmentation. *Annu Rev Biomed Eng*, 2, 315–337.

Ridler, T. W. & Calvard, S. (1978). Picture Thresholding Using an Iterative Selection Method. *IEEE T Syst Man Cyb*, 8(8), 630–632.

Sezgin, M. & Sankur, B. (2004). Survey over image thresholding techniqies and quantitative performance evaluation. *J Electron Imaging*, 13(1), 146–165.

Shapiro, L. & Stockman, G. (2002). *Computer Vision*. Upper Saddle River, NJ: Prentice Hall.

Vaisey, J. & Gersho, A. (1992). Image compression with variable block size segmentation. *IEEE T Image Process*, 40(8), 2040 –2060.

Wellner, P. (1993). *Adaptive Thresholding for the DigitalDesk*. Technical Report EPC-1993-110, Rank Xerox.