jet INTL JOURNAL OF ELECTRONICS AND TELECOMMUNICATIONS, 2010, VOL. 56, NO. 1, PP. 63-72
Manuscript received November 27, 2009; revised March, 2010. 10.2478/v10177-010-0008-1

Pseudo Bayesian and Linear Regression Global Thresholding

Khalid Aboura

***Abstract*—Classification is an important task in image analysis. Simply recognizing an object in an image can be a daunting step for a computer algorithm. The methodologies are often simple but rely heavily on the thresholding of the image. The operation of turning a color or gray-scale image into a black and white image is a determining step in the effectiveness of a solution. Thresholding methods perform differently in various problems where they are often used locally. Global thresholding is a difficult task in most problems. We highlight a pseudo Bayesian and a linear regression global thresholding methods that performed well in an engineering problem. The same approaches can be used in biomedical applications where the environment is better controlled.**

***Keywords*—Classification, Image Thresholding, Probability, Linear Regression.**

I. Introduction

CLASSIFICATION is an important task in image analysis. Simply recognizing an object in an image can be a daunting step for a computer algorithm. Images come with variation and noise. Classification is a field that evolved significantly in recent years due to the interest in pattern recognition and machine learning. While statisticians developed the field of *Statistical Classification*, engineers, physicists and computer scientists developed *Data Algorithmic* approaches that do not require modeling. Both approaches lead to successes and failures. Classification is applied in many fields. In medical imaging, statistical methods can be used to classify brain tissue in magnetic resonance (MR) images [21]. Manual recognition by a medical expert of the three brain tissue types, white matter, gray matter, and cerebrospinal fluid, is a time consuming task. The images are three dimensional, and the volume of data involved is large. An example of a data algorithmic approach in classification is the use of acoustic emission for the investigation of local damage in materials and the application of neural networks to the study of the acoustic signals [26]. In the mid 1980's, neural networks, along with decision trees provided two new powerful algorithms for fitting data. Breiman (2001) [15] argues for the goodness of the data algorithmic approach, using examples of large problems he encountered as a consultant. Classifiers from both schools have been compared on many problems and no single classifier outperforms the others. The classifier performance depends greatly on the characteristics of the data. To take advantage of the strengths of both methods, Osl et al. (2008) [19] propose an algorithm that combines a logistic regression model with a non-parametric classification method, the k-nearest neighbors. In this article we present probability models for classification problems in image analysis. *Neural Networks* have been consistently applied to one problem, when in fact a simple probability model provided a performing solution. We use one of the examples to highlight the need for good thresholding and the difficulty in obtaining performing global thresholds. We then highlight a pseudo Bayesian and a linear regression global thresholding methods that performed well in an engineering problem. The same approaches can be used in biomedical applications where the environment is better controlled.

This work was supported by the School of Computing and Communications, University of Technology Sydney.

K. Aboura is with the School of Civil and Environmental Engineering, University of Technology Sydney, 15 Broadway, Ultimo, NSW 2007, Australia. (e-mail: khalid.aboura@uts.edu.au).

II. Probabilistic Solution for a Character Recognition Problem

A pattern recognition problem that has been worked on extensively is the recognition of characters in license plates. A licence plate recognition system is designed for the automatic identification of a vehicle through its license plate. The system consists of a series of steps starting with the detection of a vehicle, the capture of images and the recognition of characters in the license plate. The last step involves image analysis in three parts; (i) the localization of the license plate in the image, (ii) the segmentation of characters from the localized license plate region and (iii) the recognition of those characters. These steps need to be performed automatically and require good algorithms. Many solutions have been applied and most of them use a learning approach. A historical data set is collected and used in the training and validation of a selected algorithm. For the character recognition, the data set consists of binary images of license plate characters extracted from images of vehicles. The characters are visually inspected and classified in the 36 possible classes {A, B, C, ..., X, Y, Z, 0, 1, ..., 8, 9}. Each set of characters is split into a training set and a validation set. The problem is to develop an algorithm that can recognize an extracted character as one of the 36 possible characters. It is a classification problem and the solution is known as an optical character recognition algorithm. Many approaches are used in the resolution of this problem. The most common ones are the correlation-based template matching [22] and neural networks [30]. Other methods are feature based, use pattern mapping or are based on the Hausdorff distance. Binary classifiers are also used as well as the Hidden Markov model.

A. Neural Network

Neural networks (NN) haven been applied successfully in many prediction and classification problems. In license plate

recognition, they are used to localize the license plate in the image and to recognize the extracted characters of the plate. A NN is trained to recognize by feeding it a set of inputs to which the outputs are known. Training data is taken from historical records. The NN processes the inputs one by one and compares the resulting outputs against the desired outputs. Errors are calculated and weights which control the strength of network connections are adjusted at each iteration. The training is stopped once the NN reaches a satisfactory level of recognition. The set of final weights is used for processing new data.

B. Template Matching

The other approach most used in character recognition in license plate recognition is template matching. It is a technique in image analysis for scanning an image template until part of it matches an image at hand. There are many variants in the application of template matching to character recognition. In its simplest form, the image, in its binary form is compared with same size parts of the template image using a suitable metric. The metric can be the euclidian distance or a correlation measure between the pixels of the image and the template. The cross-correlation, a statistical measure, can be a metric for template matching. Template matching is a method based on the minimization of a distance between two images.

C. A Simple Probabilistic Solution

Let Z be a random variable that represents statistical features of the character in a binary image processed for recognition. In a search for a performing statistic Z, Aboura (2008) [11] arrived at the conclusion that the values of the pixels in the binary image hold all the information needed to recognize the character. Z is defined as the multidimensional vector of the values of all the pixels in the image. Each of these values is either 0 or 1, the input image being binary. For each pixel, [11] applies the Bernoulli probability model $\theta_i^{Z_i}(1-\theta_i)^{1-Z_i}$, Z_i being the value of Z at pixel i. Making the assumption of conditional independence of the pixel values given an image, the likelihood function is

$$\mathcal{L}(\theta_i(C)) = \text{Prob}(Z|\theta_i(C)) = \prod_{i=1}^{|Z|} \theta_i^{Z_i}(1-\theta_i)^{1-Z_i} \qquad (1)$$

where $|Z|$ is the cardinal, or vector size of Z. $|Z|$ equals the total number of pixels in the image. To estimate the proportion $\theta_i(C)$ for pixel i in images of character C, a number of approaches are available, among them the average, and maximum likelihood estimator:

$$\hat{\theta}_i(C) = \sum_{j=1}^{N_C} x_{i,j}/N_C \qquad (2)$$

where N_C in equation (2) is the size of the training set for character C, and $x_{i,j} = \{0 \text{ or } 1\}$ is the value of pixel i for image j of the training set. This is done for each character C. From a computational point of view, the assessment of the likelihood parameters is very simple. For each character C, all the images of the training set are added. They are matrices of the same size. The resulting matrix sum is then divided by the size of the set, N_C, automatically providing a matrix of estimates $[\hat{\theta}_i(C)]_{i=1}^{|Z|}$. This is a simple operation, inexpensive computationally, that replaces the training of a neural network. Figure 1 shows the matrix $[\hat{\theta}_i(C)]_{i=1}^{|Z|}$ for character $C = K$.

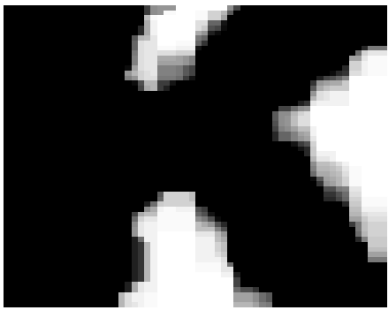


Fig. 1. Likelihood image of character K

Each value of the pixel i of the image in Figure 1 is the estimate $0 \leq \hat{\theta}_i(K) \leq 1$. It is the estimate of the parameter of the Bernoulli model of equation (1) for pixel i, for character K. Values of 1 signify that the corresponding pixels are always present in the foreground of C (in black in Figure 1). Values 0 mean that the corresponding pixels are always background in C (in white in Figure 1). The different shades of gray in between correspond to pixels i that are present in C with a probability $0 < \hat{\theta}_i(C) < 1$. One observes that the K in the image is not a perfect one. The extracted images of characters are most often taken at angles and subject to many sources of noise and deformation, making the problem a stochastic one. The image is the addition of many images, and the variation is captured in the summation. The colors of the image in Figure 1, presented for illustration, do not show all the nuances in the values. But the matrix $[\hat{\theta}_i(C)]_{i=1}^{|Z|}$ is computed with double precision accuracy and provides estimates of the probabilities of the foreground existence.

Once the likelihood function is constructed, it is used to recognize characters in a simple operation. Let z be the realization of the statistic Z for a binary image that has received similar cleaning, cropping and resizing as have the images of the training set. Then

$$\text{Prob}(C|z) = \frac{\text{Prob}(z|C)\text{Prob}(C)}{\sum_{S=A}^{S=9} \text{Prob}(z|S)\text{Prob}(S)} \qquad (3)$$

where $\text{Prob}(z|C)$ is assessed from the model of equation (1). The posterior probability distribution of equation (3) ranks the characters $A, B, \ldots, X, Y, Z, 0, 1, \ldots, 8, 9$ for their likelihood of being the character in the image being treated. The character with the highest posterior probability is selected. This method provides a full probabilistic approach. To further improve the speed and increase the accuracy, we note that maximizing the product of two bounded positive values is equivalent to maximizing their sum. Ignoring any prior probability influence, the score function $f(C)$ of equation (4) is used to determine the

most likely character candidate:

$$\max_{C=A,B,\ldots,9} \mathrm{f}(C) = \sum_{i=1}^{|Z|} \{\hat{\theta}_i(C)^{z_i} + (1-\hat{\theta}_i(C))^{1-z_i}\} \quad (4)$$

This simple and fast method yielded excellent results. As such, it fails to distinguish fully between some characters like 2 and Z, 5 and S, 1 and I, B and 8, and O, 0, D and Q. However, using the same logic and applying it exclusively to parts of the image, a 97% reliability was reached.

III. Statistical Analysis of a Localization Signal

License plate numbers show frequencies that lead themselves to a spectral analysis. Parisi et al. (1998) [28] propose an application of the *Discrete Fourier Transform* (DFT) to find the license plate. DFT is a transform for Fourier analysis of finite-domain discrete-time signals that can be computed efficiently in practice using a *Fast Fourier Transform* (FFT) algorithm. The authors in [28] scan the image horizontally and vertically and use the average periodogram estimate of the signal as a statistic to detect the license plate region. Acosta (2004) [4] proceeds similarly and computes the periodogram estimate using the FFT, for each row to locate vertically, then each column to locate horizontally. Rather than using the whole range of frequencies, the author in [4] computes the average of the periodogram over some frequencies where the license plate signal is expected to dominate. We introduce a formal statistical analysis of the localization signal.

A. Image Signal Data

The DFT is used in signal processing to analyze the frequencies contained in a sampled signal. In our approach [10], the original image is scanned with a small window that slightly overestimates the license plate size. The study was done on images taken for vehicles within a certain distance. This gives a fairly good estimate of the size of the license plate in the image. The scanned region was taken to be 30 x 120 for 637 x 480 images. The scan window starts at the upper left corner of the image and is moved 5 pixels at a time to the right until it reaches the end of the image. It is then lowered 5 pixels and moved again to the right, and so on, until the whole image is scanned. In each iteration, the original image in the window is transformed into gray scale, G=.2989*R+.5870*G+.1140*B, using the RGB components of the window image. Hysteresis thresholding [8] is applied to obtain a binary image of the scanned window. Hysteresis thresholding was found to be the best in highlighting characters in a license plate image region, if applied locally. That is if the size of the thresholded region is not too big, which happened to be near perfect in our case. We observed this fact empirically over thousands of license plate regions of size 30 x 120. Once thresholding is applied, the columns of the resulting binary matrix are summed up and normalized by the height of 30, to provide the discrete signal **S** for that scanned region, over the 120 discrete points representing the pixels on the x-axis. Figure 2 shows the signal of a license plate region.

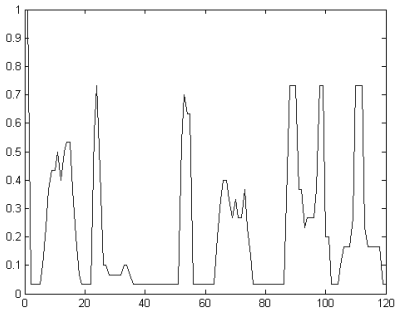


Fig. 2. Signal **S** of the scanned region

The DFT of the signal $\mathbf{S} = \{S(j), j = 1, \ldots, 120\}$ is defined in equation (5):

$$Y(k) = \sum_{j=1}^{120} S(j)\ e^{-2\pi i(j-1)(k-1)/N} \quad k = 1, \ldots, 120 \quad (5)$$

or $Y(f) = Y(k)$ where $f = k/120$. The power spectrum is a measurement of the power at various frequencies, defined in this case as $P(f) = |Y(f)|^2/120$. Due to its symmetry, the curve $P(f)$ (Fig. 3) is split in two to provide the power function.

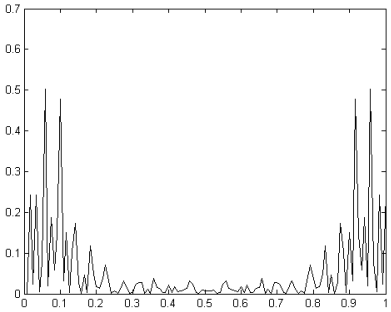


Fig. 3. Power spectrum of the image signal **S**

B. The Statistical Model

The behavior of the power spectrum of the scanned region shows a significant increase in magnitude at some frequencies, for scanned parts of the image that contain the license plate or parts of it. This is due to the periodicity in the signal generated by the characters of the license plate, as apparent in Figure 2. Summing the power at these frequencies allows for the ranking of the scanned regions for the likelihood of hosting the license plate. The regions with maximum summed power are considered candidate regions. For example, [4] chooses to sum the power of the FFT signal over a range of frequencies. While this approach does work, it requires the specification of the domain of frequencies over which to sum the power spectrum. This domain may change from one recognition system deployment to another. We introduce a statistical procedure that does not require the specification of the power spectrum range. We define functions of the power spectrum and analyze them statistically using probability models built with historical

data. Using likelihood ratio tests, we determine whether each scanned region is a candidate. Bayesian posterior ranking of scores allows us to determine areas in the image in which the license plate is located.

1) Statistics of the Signal: Several images were used as historical data. Each image was visually inspected for the existence of a car in it. For each scanned region, we consider the power spectrum function $p(f) = P(f)/120, f = .025, .0333, \ldots, .5083$ as being the range of the power function divided by 120. The frequencies correspond to $k = 3, \ldots, 61$, and the division by 120 was arbitrary and does not affect the results. We compute the following five statistics:

- Strength of the Signal: The strength of the signal is the summation of the power spectrum over the range of frequencies, i.e. $SS = \sum_f p(f)$.
- Normalized Maximum Amplitude: We consider the maximum amplitude of the power spectrum divided by the strength, $MS = \max_f p(f)/SS$.
- Frequency of Maximum Amplitude: Let $FM = \hat{f}$ such that $p(\hat{f}) = MS$.
- Frequency Center: We compute a statistic and call it the frequency center $FC = \sum_f (f * p(f))/SS$. It is a power weighted average of all the frequencies.
- Frequency Spread: The frequency spread is a measure of variation of the frequency of the signal, weighted by the power spectrum, $FS = \sum_f (f - FC)^2 * p(f)/SS$.

2) The Likelihood Model: Let the event $\theta = 1$ be the event that the region has the license plate, and let D be data that provides information about θ. By the laws of probability, $\text{Prob}(\theta|D) \propto \text{Prob}(D|\theta)\text{Prob}(\theta)$. The first step is to build the probability model $\mathcal{L}(\theta) = \text{Prob}(D|\theta)$, the likelihood function of θ. This is done using the historical set of images. These images are taken through the same steps that a scanned region is: (i) gray scale conversion, (ii) thresholding, (iii) projection of foreground to obtain a signal, and (iv) FFT and power spectrum computations. The five statistics are computed to form the data $D = (SS, MS, FM, FC, FS)$. It was found that, except for FM, the Normal distribution is a reasonable model for each of the statistics. The histogram of the frequency of the maximum amplitude FM shows bi-modality. It is modeled with a mixture of Normal distributions (Fig. 4).

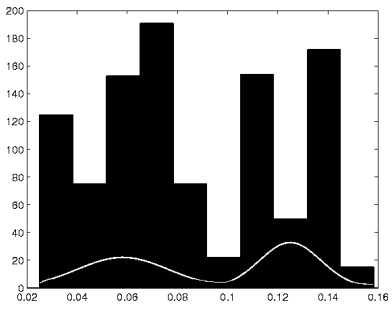


Fig. 4. Frequency of maximum amplitude as a Mixture of Normals

Using conditional independence assumptions, we build the likelihood model $\mathcal{L}(\theta) = \text{Prob}(D|\theta)$, starting with $\theta = 1$,

$$\begin{aligned}\text{Prob}(D|\theta = 1) &= \text{Prob}(MS, FM, FC, FS, SS|\theta = 1) = \\ &= \text{Prob}(FM|MS, SS, \theta = 1)\text{Prob}(MS|SS, \theta = 1) \times \\ &\quad \text{Prob}(FS|FC, SS, \theta = 1)\text{Prob}(FC|SS, \theta = 1) \times \\ &\quad \text{Prob}(SS|\theta = 1).\end{aligned}$$

Similarly, a large set of historical data on scanned regions that are not license plate regions is gathered. Over 19,000 images of non plate data are collected and cut into 30 x 120 sections. The 5 statistics are computed and analyzed to obtain $\text{Prob}(D|\theta = 0)$, the alternative likelihood.

C. Scores and Likelihood Ratios

The assessment of the likelihood function is the essential part of the statistical methodology. Once $\mathcal{L}(\theta)$ is constructed, it can be used in the selection of candidate regions for the license plate. The candidate regions are found by maximizing the likelihood; $\max_d \text{Prob}(d|\theta = 1)$, where d is the data of the scanned region, realization of D. We ranked all the scanned regions to obtain the first 100 regions with a descending score. Figure 5 shows successive candidate regions for one image, pasted together left to right and top to bottom for illustration. The method performed relatively well for most images. In this

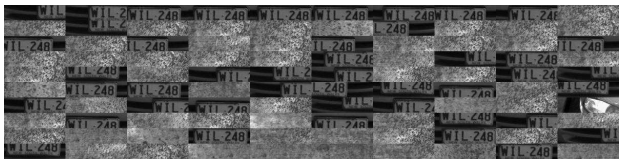


Fig. 5. Candidate license plate regions

example, the first nine regions with the highest scores contain part of the license plate and are therefore successful in locating the region of the license plate in the image. Supplemental localization processing will extract the exact license plate region using this first localization. But this approach did not work for all images. It failed in a number of cases, for example when there is a strong signal similar to that of a license plate existing in the image. It failed in a non negligible percentage of the images processed. The approach does not use information about the alternative. It does not employ the alternative likelihood model $\text{Prob}(D|\theta = 0)$. Conducting a full Bayesian analysis, we arrived at a methodology that results in a reliable localization of the license plate.

1) Bayesian Analysis: To rank all the scanned regions of an image according to their likelihood of being the license plate, a score is assigned to the area. The score is the posterior probability $\text{Prob}(\theta = 1|d)$, computed for each scanned area as follows:

$$\text{Prob}(\theta = 1|d) = \frac{1}{\delta}\,\text{Prob}(d|\theta = 1)\text{Prob}(\theta = 1) \quad (6)$$

$\delta = \text{Prob}(d|\theta = 1)\text{Prob}(\theta = 1) + \text{Prob}(d|\theta = 0)\text{Prob}(\theta = 0)$. It is not trivial, but possible, to provide prior knowledge for the scanned regions. Texture and colors in the region can be used to construct a prior probability. In our case, we choose to ignore any differentiating prior knowledge, making the prior

probabilities $\text{Prob}(\theta = 1) = \text{Prob}(\theta = 0) = 1/2$. This leads to

$$\text{Prob}(\theta = 1|d) = \frac{\text{Prob}(d|\theta = 1)}{\text{Prob}(d|\theta = 1) + \text{Prob}(d|\theta = 0)} \quad (7)$$

This is the posterior probability that the scanned region is the license plate. Similarly, the posterior probability that the region is not the license plate is

$$\text{Prob}(\theta = 0|d) = \frac{\text{Prob}(d|\theta = 0)}{\text{Prob}(d|\theta = 1) + \text{Prob}(d|\theta = 0)} \quad (8)$$

Using this approach, the reliability of the localization process was improved (Fig. 6).

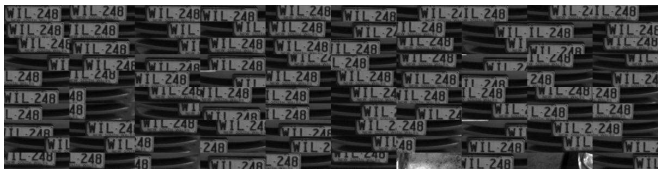


Fig. 6. Candidate regions using the Bayesian approach

D. Locating the License Plate

In Figure 7, the upper-left corner location of these candidate regions are dotted in a scattered plot over an area the size of the original image. The larger concentration area of dots in this plot corresponds to the license plate location in the image. This pattern repeats itself in all images. To locate the license plate,

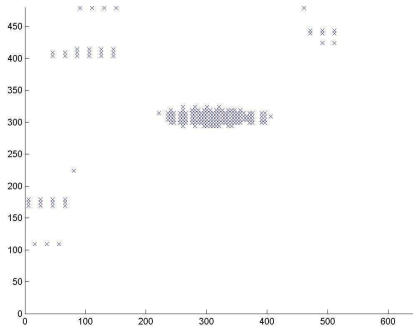


Fig. 7. Scattered plot of the upper-left corner of the candidate regions

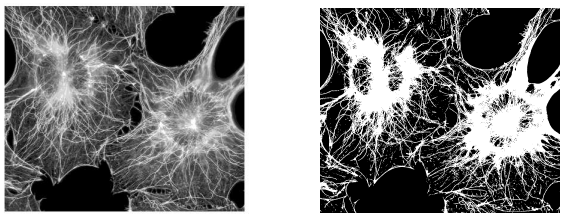
additional steps are taken in [14], where the scatter points are joined to form clusters and a statistical analysis is performed on their binary images. A likelihood function is built and used to determine precisely the location of the license plate region. The overall approach yielded a 94% reliability in the case of our data, based on a set of 1000 images. This approach shows the applicability of a probability model to a signal in the localization of a region in an image.

E. Importance of Thresholding

Image thresholding is a classification problem where the pixels of an image are divided into two classes: foreground and background. Thresholding creates binary images from grey-scale ones by turning some pixels to zero (black) and the other pixels to one (white). A digital image is a matrix of values consisting of the colors of the pixels. For a $n \times m$ image, let $I_{i,j}$ represents the color of pixel (i, j), $i = 1, \ldots, n$, $j = 1, \ldots, m$. In the red green blue (RGB) image color space, $I_{i,j}$ can take one of $256^3 = 16,777,216$ possible colors. This range creates an extremely large number of combinations of pixel values making up the image. To work in this image space makes it nearly impossible to have reasonable processing times. A reduction in dimension is unavoidable and it comes first in the form of a grayscale, or gray level, transformation. The color image is reduced to a grayscale image through the application of one of a number of transformations that can take the pixel value from $[0 \;\; 255]^3$ to $[0 \;\; 255]$. Even in this one dimensioned pixel space, the application of some search algorithms can be tedious and would require long processing times. In some image analysis problems such as the search for an object in the image, or the definition of a contour, the difficulty may be greatly reduced by switching to a black and white image, without too much loss in efficiency. In a black and white image, each pixel can have only one of two values, 0 for black and 1 for white, as compared to a whole interval $[0 \;\; 255]$, or 256 values for a gray pixel. When this difference is taken to a $n \times m$ image, where n and m are in the hundreds, it makes a great difference. The operation of turning a gray-scale image into a black and white image is done through thresholding. All three steps in license plate recognition rely on thresholding, or binarization, of the original image or parts of it. Many approaches exist for plate localization and most achieve a good result if the thresholding is adequate. Thresholding in license plate recognition has often been done empirically and seldom mentioned in the reporting of results. In [12], this topic is discussed and a new thresholding method is introduced. In the methodology we introduce here for the localization of the license plate, we improve the accuracy and the speed of the results using the thresholding method of [12]. In this article, we review the new global thresholding method. But first we introduce the topic in biomedical applications where it is often conducted manually or semi-automatically.

IV. Thresholding in Biomedical Applications

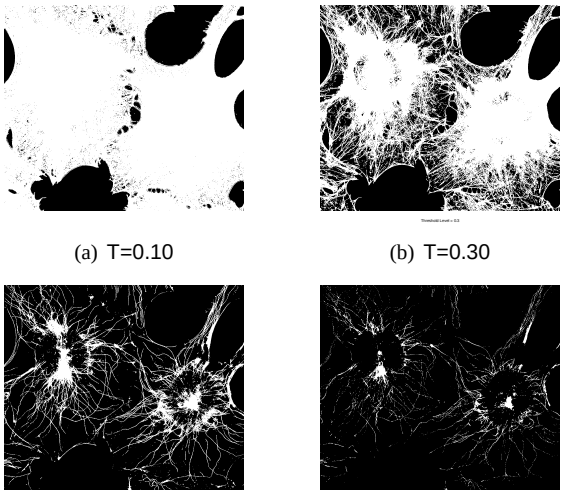
To illustrate the thresholding problem, we take the image of cells used in biomedicine (Figure 8a) and taken from the image bank of the National Institutes of Health, US Department of Health and Human Services [23]. The original image is in color, the actin purple, the microtubules yellow, and the nuclei green. And these are labeled in these cells by immunofluorescence. We first take the image through a grayscale transformation using the formula $J = .2989 * R + .5870 * G + .1140 * B$ on the RGB components of the image. Then we use a well known thresholding method, the Otsu method [24] and apply it to the grayscale image to obtain the black and white image in Figure 8b. The Otsu method is a clustering-based method. The algorithm assumes two classes of pixels, say foreground and background. It then computes the optimal threshold that minimizes the weighted sum of within-class variances of the foreground and background pixels. This method gives satisfactory results when the numbers of pixels in each class



(a) Original Image (b) Otsu Thresholding

Fig. 8. Cell image (NIH [23])

are close to each other. The Otsu method still remains one of the most referenced thresholding methods. In the case of Figure 8a, it yielded an optimal threshold of 0.4275. To better understand threshloding, we apply the simplest thresholding method. The color image is first converted to a gray scale image using the formula $J = .2989 * R + .5870 * G + .1140 * B$ on the RGB components of the image. In a simple approach, thresholding at level $T \in [0, 1]$ means that all pixel values J are classified as background (black=0) if $J/255 < T$ and foreground (white=1) otherwise[1]. The image is divided into background and foreground using the threshold level T. we vary the threshold T from its optimal value and set it to 0.10, 0.30, 0.70 and 0.90. The resulting black and white images are show in Figure 9. Varying the threshold T from 0 to 1 takes



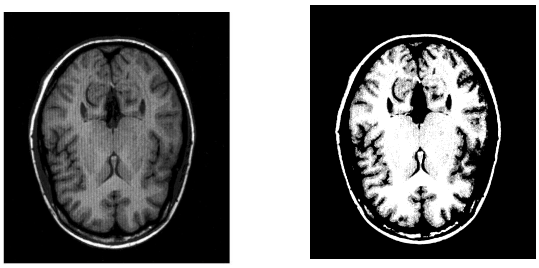
(a) T=0.10 (b) T=0.30

(c) T=0.70 (d) T=0.90

Fig. 9. Different thresholds

the resulting image from an all white image to a completely dark one. There is an 'optimal' threshold level that separates best the desired features in the image. In this example, the Otsu level of 0.4275 seems to provide a good black and white image, although 'good' here depends on the context, that is the goal of the image analysis problem.

Another biomedical example requiring thresholding is the automatic segmentation of the caudate nucleus (CN) region from human brain magnetic resonance images (MRI). The issue of automation is important in the analysis of MRIs. Manual segmentation requires significant time on the part of expert medical staff. In addition, manual inspection is prone to human errors. In the case of this particular biomedical problem, existing solutions such as SnAP [29], a software package for CN segmentation available in the public domain, require manual input for a number of tasks including thresholding. This highlight the importance of thresholding. It is a significant step in any image analysis. All ensuing results depend on the effectiveness of the separation of pixels in the image before applying any search or localization algorithm. Figure 10 shows a 2D section of a MRI brain image, which comes in color originally, and its corresponding Otsu thresholded binary image.



(a) Original Image (b) Otsu Thresholding

Fig. 10. MRI brain image (NIH [23])

A. Local thresholding

A common problem with global thresholding is the changes in illumination across an image. Parts of the image appear to be brighter and some parts darker regardless of the objects being photographed. This illumination can be natural or man made and has to do with the angle of the shot, the time of the day and other factors, some random. This variation of illumination renders the application of a global threshold difficult. A commonly used solution is to apply thresholds locally. Instead of having a single global threshold, the threshold is allowed to vary across the image. The hysteresis thresholding approach [8] is a local thresholding method that can be used efficiently in a number of problems, for example in angiography where vessel imaging is conducted after the injection of a radiopaque substance [2]. Figure 11 shows the segmentation results obtained for a retina image. The first column shows the original image, the second the ground truth (hand segmentation) and the third the segmentation result using Hysteresis thresholding. A number of thresholding methods exist (see surveys [27], [18] and [5]).

V. Pseudo Bayesian Global Thresholding

In license plate recognition, hysteresis thresholding performs poorly when applied to the whole image. It is used locally around the license plate area where it is 'optimal', in that it brings out the characters in the foreground. In this section, we review a global thresholding method [12] suitable for license plate recognition. We call it a Pseudo Bayesian

[1]The color black can be used to represent background or foreground, depending on the convention adopted.

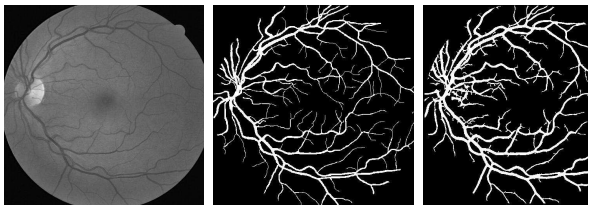


Fig. 11. The hysteresis thresholding approach (Condurache and Aach [2])

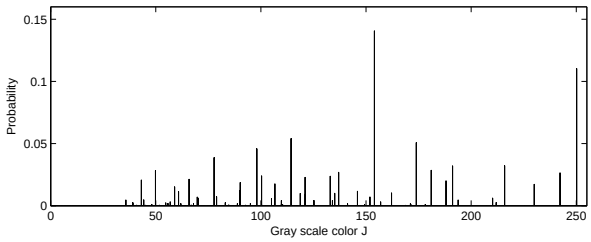


Fig. 12. Probability distribution of the major gray colors

method as it mimics the use of a proper Bayesian method. The solution was found accidentally while searching for a probability model. It uses information from a color spectrum histogram representation of the image. The notion of the reduction of the number of colors in an image and their representation in a histogram has been in use for some time in a number of image analysis problems; detection of image-to-image similarity (Chen and Wong 1999 [32]), in optimal color composition matching of images (Hu and Mojsilovic 2000 [9]), tracking people in video surveillance (Lu and Tan 2001 [31]), tracking people (Piccardi and Cheng 2005 [20]). A normalized geometric distance between two points in the RGB space can be used to create clusters of colors. The centers of the clusters define the major colors and these major colors are ranked by the sizes of their respective clusters. The image colors are reduced to a limited number of major colors (for example, 15 to 100) without losing much accuracy on representing an object in the image with those colors. We use the same approach in defining a color histogram for each image. This information in the image is treated as data, and we mimic the Bayesian combination of prior information and image data to provide an estimate of the 'optimal' thresholding level of the image.

A. Prior information and image data

A set of images was studied. The license plate regions in the images were identified visually and thresholded using the hysteresis method. Then each license plate image was run through the straightforward thresholding where a level $T \in [0,1]$ separates the gray scale version of the image into two sets $\{0,1\}$. T was varied from 0 to 1, with .01 increments. The minimization of squared errors was used to select that T^{opt} value that yielded a thresholded image that most resembles the result of the hysteresis approach. This T^{opt} value is considered to be the 'optimal' threshold level. The major color histogram is calculated from the image of a vehicle. Let c represent the vector of colors on the x-axis of the histogram. These are the major colors found in the image. c is a vector of RGB colors. Let J be its gray scale transform, that is $J = .2989{*}c(.,1){+}.5870{*}c(.,2){+}.1140{*}c(.,3)$, where $c(.,1), c(.,2)$ and $c(.,3)$ are the RGB colors, respectively, of the corresponding color c. Figure 12 shows the resulting histogram in the form of a probability distribution for the intensities (gray colors) in an image. J is then sorted in a descending order into $u = \{u_1, \ldots, u_n\}$, where $u_i > u_j$ for $i > j$, as shown in figure 13 where the colors J in the histogram are the u_i's, starting with the maximum major gray color 250.18 and finishing with the smallest (darkest) gray color, .298 in the example. There are $n = 282$ values in the example. Note that n can be set arbitrarily, depending on the problem. A color image can possess a very large number of colors. One can limit the number of major colors. In our implementation, we accept as a many as 400 major colors. In this fashion, most likely, all major colors in an image will be captured by the color histogram. Given the u_i's, the image

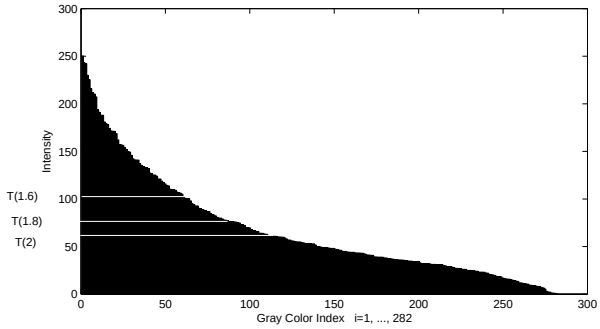


Fig. 13. Histogram gray colors ranked in decreasing order

data is generated the following manner. Let $Rs = 1.6, \ldots, 2$, with increments of .004. Let $T(Rs)$ be the threshold value $u_m/255$, such that m is the first i that satisfies equation (9), as shown in figure 13.

$$\sum_{i=1}^{m} u_i > \frac{\sum_{i=1}^{n} u_i}{Rs}, \tag{9}$$

Then these 101 image sampled values, $T(Rs),\ Rs = 1.6, 1.604, \ldots, 2$, are the image data.

B. Pseudo-Bayesian thresholded image

The prior information is a discrete probability distribution derived from historical data, $p = \{p_i\}$, such that $p_i = \text{Prob}(T^{opt} = T_i),\ T_i = 0, 0.01, , \ldots, 1$. The image data $T(Rs),\ Rs = 1.6, 1.604, \ldots, 2$ is rounded off to match the discrete sample space $T_i = 0, 0.01, , \ldots, 1$, and its frequency is summarized in a probability distribution $q = \{q_i\}$, such that $q_i = \text{Prob}(T(Rs) = T_i),\ T_i = 0, 0.01, , \ldots, 1$. In an ad-hoc manner that mimics the combination of likelihood and prior in Bayes' theorem, we multiply these two distribution in equation (10) to obtain a distribution Q:

$$Q = \frac{p.q}{\sum_i p_i.q_i} \tag{10}$$

Using this distribution, we generate an image we call a *Pseudo-Bayesian Thresholded* image by simply thresholding the original image for each threshold level $T_i = 0, 0.01, , \ldots, 1$, and multiplying the image with the coresponding $Q_i = p_i.q_i / \sum_j p_j.q_j$, then summing all the resulting images. This is the equivalent of taking the expected value over the optimal threshold with the posterior distribution. Figure 14 shows an example of the resulting thresholded image.

Fig. 14. Pseudo-Bayesian thresholded image

The image in figure 14 has pixel values between 0 and 1. To make it a fully thresholded image, we threshold again, this time in the middle, by turning all pixels of value less than 0.5 into 0 and the others into 1. This approach yielded good results. In about 89.8% of the cases, this thresholding gave a high level of separation of the characters in the license plate.

VI. Linear Regression Global Thresholding

The information contained in the image data used in the pseudo Bayesian approach provides the major part of the solution. However the method is ad-hoc and does not make full use of the color information contained in J. It uses the intensities found in the image, but not the frequencies at which these intensities exist in the image. To incorporate the full color information, a formal statistical approach is adopted by [12] in the development of a thresholding method based on a linear regression model. A set of 9 explanatory variables is used. The first source of explanation was the same one used to generate the image data in the above thresholding method. Instead of varying Rs from 1.6 to 2 and sampling within, the intensity and therefore the threshold $T(Rs)$ is taken at 6 values $Rs = 4, 2.66, 2, 1.6, 1.33, 1.14$. These variables provide the information about the intensities in the image. To use the full color information, the histogram values corresponding to the intensities given by the color histogram are accumulated from the smallest intensity to the highest. Then the three intensities that provide the 25%, 50% and 75% of the accumulated sum are used as the remaining 3 variables in the linear regression. These 9 variables plus the intercept proved to be the most effective regression in the case of the data. A training set of 200 images was used to estimate the model parameters. Figure 15 shows the performance of the model over the 800 images test data. The method is compared to the Otsu method, which performs poorly globally, to the optimal threshold assessed manually and to a a fixed threshold $T = 0.5$. Where the model appears to perform poorly, the images were inspected visually. Due to the varying color image composition, illumination and brightness, more often than not the poor performance was not as bad as the error in the model. All the 1000 images were

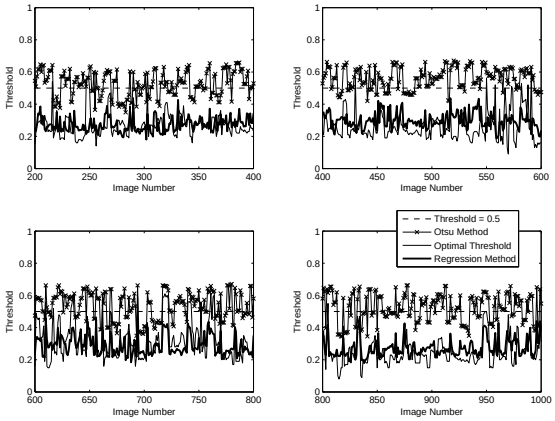


Fig. 15. Thresholds for the validation data

inspected visually for each of the three thresholding methods. The regression approach yields a 93.4% reliability figure in the case of our data. Figure 16 shows the example of a thresholded image using the regression approach.

Fig. 16. Linear Regression thresholded image

VII. Probability and Bayes Theorem

We call the first global thresholding approach a "Pseudo Bayesian" solution. This is because we mimic the use of Bayes theorem in an ad-hoc manner that works. We construct a prior distribution from the data and combine it with data from images in the same fashion Bayes theorem would use that data in a likelihood function. We do that without having constructed the probability model of the likelihood function. The likelihood model is implied in the image, and the image

data q of section V-B is used as its values. This solution was arrived at by accident, and it shows in fact that Bayes theorem works wonders even when mishandled.

Bayes' Theorem is a powerful inference tool, derived from the laws of probability. In 1933, the Russian mathematician Andrey Nikolaevich Kolmogorov introduced an axiomatic approach, *Grundbegriffe der Wahrscheinlichkeitsrechnung* [1], turning the probability concept into a mathematical theory. Andrei Kolmogorov work set out the axiomatic basis for modern probability theory. Kolmogorov organized a theory Emile Borel [1871-1956] had created many years earlier by combining countable additivity with classical probability. In Kolmogorovs axioms and in his way of relating his axioms to the world of experience, they were traces of the work of many others such the work of Borel, the work of Maurice Frechet [1878-1973], and that of Francesco Cantelli [1875-1966], Alexander Chuprov [1874-1926], Paul Levy [1886-1971], Wladyslaw Steinhaus [1887-1972], Stanislaw Ulam [1909-1984] and von Mises [1883-1953] [7]. A Russian translation of [1] by G. M. Bavli, appeared in Moscow in 1936, with a second edition, slightly expanded by Kolmogorov with the assistance of A. N. Shiryaev, in 1974, and a third edition in 1998. An English translation by N. Morrison appeared under the title *Foundations of the Theory of Probability* (Chelsea, New York) in 1950, with a second edition in 1956 [7].

Probability theory is a modern mathematical theory, considered a subfield of measure theory. Probability models are used to solve problems that have a relevant amount of uncertainty in them. Bayes' Theorem is a probability rule for inverse probability computation. In what seems to be a simple law, it contributed to the resolution of countless problems in estimation, inference, and prediction. The formulation of the theorem is due to Thomas Bayes [1702-1761] and was rediscovered by Laplace [1749-1827] in 1774. Laplace also derived the Law of Total Probability. This law is another important probability derived rule, also used to compute the denominator, or normalizing constant in Bayes theorem. Bayes' theorem allows the updating of probability in light of new information. If $P(E)$ is the initial probability for an event E of interest, such as "Optimal threshold value is 0.567" and D is the newly acquired data, then $P(E)$ is changed into $P(E|D)$ through Bayes' Theorem to reconcile past and current information (See Singpurwalla (2006) [25]):

$$P(E|D) = \frac{P(E,D)}{P(D)} = \frac{P(D|E)P(E)}{P(D)}. \tag{11}$$

Bayes' theorem is known as the "law of inverse probability", having the ability to assess $P(D|E)$, one can turn $P(E)$ into $P(E|D)$, since $P(D)$ can be calculated using the law of total probability [25]:

$$P(D) = \sum_i P(D|E_i)P(E_i), \tag{12}$$

where $\{E_i\}$ represents an exhaustive and exclusive partition of the set of all possible outcomes.

A. Bayes Theorem and Subjective Probability

Bayes theorem is a powerful tool, but it requires the specification of a prior distribution. In many problems, one starts with $P(E)$, called a prior probability, or as is often the case, a prior distribution, and computes $P(E|D)$, where D is the set of data. D often comes in pieces over time, $(D_1, D_2, \ldots)$ such as radar measurements in *Target Tracking*. Each time new information D_j arrives, $P(E|D_j, D_{j-1}, \ldots, D_1)$ is obtained from $P(E|D_{j-1}, \ldots, D_1)$ using Bayes' theorem. This mechanism provides for a powerful recursive probabilistic updating that proved successful in many problems. For example, in target tracking, it is the basis of many effective algorithms for tracking targets in all sorts of conditions [6]. However, one must start with $P(E)$. In a field like target tracking or image analysis, data are abundant and the effect of $P(E)$ is small and does not affect the solution much after a while. For example, in the character recognition problem, we chose a uniform prior in equation (3), that is $P(C) = 1/36$ for any C. In the license plate region localization problem we chose $P(\theta = 1) = P(\theta = 0) = 1/2$ (Eq. 6). It doesn't matter much as data takes over quickly. But when there is little data, the subjective input weights in significantly on the final answer. Some dictates the use of *Subjective Probability* in defining the prior. The approach is an axiomatic theory for defining the meaning of probability that lead to the development of the *Bayesian* school of statistics in the 20th century. While mostly successful in the application of their methods, *Subjective Bayesians* encounter resistance to their definition of probability (See Aboura (2009) [13]).

Bayes' theorem is a probability law, and as such, it is not subject to controversy. In the image analysis solution we presented, we used Bayes' theorem and therefore presented Bayesian solutions. This does not make us Subjective Bayesians. Subjective Bayesians define probability as 'personal' probability. This is where the controversy occurs.

B. Subjective Probability

The meaning of probability took many forms over the course of centuries. From James Bernoulli's [1654-1705] notion of probability, to Laplace's [1749-1827] definition of probability, to Venn [1834-1923] and Von Mises [1883-1953] frequentist interpretation, to de Finetti [1906-1985] and Savage [1917-1971] subjective probability, the meaning of probability has been constantly questioned and continues to be. In the example of a flip of a coin, the probability of 0.5 could be arrived at through three possible reasonings. The first one uses the symmetry of the coin and can be used in many other situations. Not all situations lead themselves to this assessment. In the second reasoning, the probability is taken to be the limit of the frequency of the outcome if the coin is flipped infinitely in similar conditions. This provides the basis for the frequentist view of probability. While this definition of probability prevailed for a long time and is liked by many in different scientific fields for its rigorous definition, it lacks a major quality; it does not apply to all situations. The idea of a 'personal' probability, or 'subjective' probability saw the day with the work of Frank Ramsey [1903-1930] and Bruno de Finetti [3] early in the 20th century in Europe, and that of Leonard Jimmie Savage [16] in the United States in mid century. It was developed over the course of decades, with its origin crossing

centuries. The full theoretical axiomatic approach of Savage is described in Degroot (1970) [17]. Aboura (2009) [13] argues against the use of subjective probability. First the author shows a case in bridge maintenance optimization where the likelihood model for expert opinion can be developed leading to a 'good' subjective prior distribution, then states the reason for the inapplicability of the approach. The author reviews the foundations of the definition of subjective probability and highlights flaws of the theory. He concludes that there is no convincing definition of probability and while the work of de Finetti and Savage is admirable, probability still remains to be defined properly.

VIII. Conclusion

Converting an image into a binary image is a fundamental step in image analysis. It has several applications in the biomedical field. There is a large variety of methods, but the problem of finding a global threshold remains in the case of many problems. We highlight an approach based on the application of a linear regression model where the explanatory variables are color intensity informative variables extracted from the original images. The method succeeds in the good segmentation of the image in an engineering problem where outdoor conditions increase the difficulty of separating the foreground from the background. The purpose of this article is to bring the success of the method to the attention of researchers in the bio-medical imaging field. Simple approaches can be used to develop effective global thresholding methods.

Acknowledgment

The author would like to thank Ryszard Klempous and Zenon Chaczko for their efforts in making possible the 4th International Conference on Broadband Communication, Information Technology & Biomedical Applications, held in Wroclaw, Poland.

References

[1] A. N. Kolmogorov, *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Berlin: Julius Springer, 1933.

[2] A. P. Condurache and T. Aach, "Vessel segmentation in angiograms using Hysteresis thresholding", in *Proc. of the Ninth IAPR Conference on Machine Vision Applications*, Tsukuba Science City, Japan, 2005, pp. 269-272.

[3] B. de Finetti Bruno, "La Prevision: ses lois logiques, ses sources subjectives", *Annals de l'Institut Henri Poincare*, vol. 7, pp. 1-68, 1937.

[4] B. D. Acosta, "Experiments in image segmentation for automatic US license plate recognition", Master's thesis, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, USA, 2004.

[5] C.-I Chang, Y. Du, J. Wang, S.-M. Guo and P. D. Thouin, "Survey and comparative analysis of entropy and relative entropy thresholding techniques", *IEE Proceedings Vision, Image and Signal Processing*, vol. 153, no. 6, pp. 837-850, 2006.

[6] D. Musicki,, R. Evans and S. Stankovic, "Integrated probabilistic data association", *IEEE Transactions on Automation and Control*, vol. 39, no. 6, pp. 1237-1241, 1994.

[7] G. Shafer and V. Vladimir, "The sources of Kolmogorovs Grundbegriffe", *Statistical Science*, vol. 21, no. 1, pp. 70-98, 2006.

[8] J. Canny, "A computational approach to edge detection" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679-698, 1986.

[9] J. Hu and A. Mojsilovic, "Optimal color composition matching of images", in *Proc. of the 15th International Conference on Pattern Recognition*, Barcelona, Spain, 2000, vol.4, pp. 47-50.

[10] K. Aboura, "Stochastic analysis of discrete fourier transform data for license plate localization", Technical Report, School of Computing and Communications, University of Technology Sydney, 2007, unpublished.

[11] K. Aboura, "Probabilistic reasoning still the shortest path in many problems", *International Journal of Computing, Anticipatory Systems*, vols 20-21-22, 2008.

[12] K. Aboura, "Automatic Thresholding of License Plate", *International Journal of Automation and Control*, vol. 2, nos. 2-3, 2008.

[13] K. Aboura, "The Inapplicability of the Concept of Subjective Probability", presented at the 9th International Conference on Computing Anticipatory Systems, Liege, Belgium, Aug. 3-8, 2009, to be published in the *International Journal of Automation and Control*, 2010.

[14] K. Aboura and R. Alhmouz, "License plate localization using a statistical analysis of discrete fourier transform signal", Technical Report, School of Computing and Communications, University of Technology Sydney, 2007, unpublished.

[15] L. Breiman, "Statistical Modeling: The Two Cultures", *Statistical Science*, vol. 16, no. 3, pp. 199-231, 2001.

[16] L. J. Savage, *The Foundations of Statistics*, 2nd edition. New York: Dover, 1954.

[17] M. H. Degroot, *Optimal Statistical Decisions*. McGraw-Hill, 1970

[18] M. Sezgin and B. Sankur, "Survey over image thresholding techniques and quantitative performance evaluation", *Journal of Electronic Imaging*, vol. 13, no. 1, pp. 146-165, Jan. 2004.

[19] M. Osl, C. Baumgartner, B. Tilg, and S. Dreiseitl, "On the combination of logistic regression and local probability estimates", presented at the Third International Conference on Broadband Communications, Informatics & Biomedical Applications, Pretoria, South Africa, Nov. 23-26, 2008.

[20] M. Piccardi and E. D. Cheng, "Multi-frame moving object track matching based on an incremental major color spectrum histogram matching algorithm", in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 3, pp. 19-24, 2005.

[21] M. B. Cuadra, L. Cammoun, T. Butz, O. Cuisenaire and J. Thiran, "Comparison and validation of tissue modelization and statistical classification methods in T1-weighted MR brain images", *IEEE Transactions on Medical Imaging*, vol. 24, no. 12, pp. 1548-1565, Dec. 2005.

[22] M.-A. Ko and Y.-M. Kim, "License plate surveillance system using weighted template matching", presented at the 32nd Applied Imagery Pattern Recognition Workshop, Washington, DC, USA, Oct. 15-17, 2003.

[23] National Institutes of Health, US Department of Health and Human Services, NIH Image Bank, http://media.nih.gov/imagebank/about.asp.

[24] N. Otsu, "A threshold selection method from gray-level histogram", *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62-66, Jan. 1979.

[25] N. D. Singpurwalla, *Reliability and risk. A Bayesian perspective*. John Wiley and Sons, 2006.

[26] N. Godin, S. Huguet and R. Gaertner, "Integration of the Kohonen's self-organising map and k-means algorithm for the segmentation of the AE data collected during tensile tests on cross-ply composites", *NDT & E INTERNATIONAL*, vol. 38, no. 4, pp. 299-309, 2005.

[27] P. K. Sahoo, S. Soltani, A. K. C. Wong and Y. Chen, "A survey of thresholding techniques", *Computer Vision, Graphics and Image Processing*, vol. 41, pp. 233-260, 1988.

[28] R. Parisi, E. D. Di Claudio, G. Lucarelli and G. Orlandi, "Car plate recognition by neural networks and image processing", in *Proc. of the 1998 IEEE International Symposium on Circuits and Systems*, Monterey, CA, USA, 1998, vol. 3, pp. 195-198.

[29] SnAP, University of Pennsylvania, University of North Carolina, Chapel Hill, ETH Zurich, 2006. Available: http://www.itksnap.org/

[30] S.-L. Chang, L.-S. Chen, Y.-C. Chung and S.-W. Chen, "Automatic license plate recognition", *IEEE Transactions on Intelligent Transportation Systems*, vol. 5, no. 1, pp. 42-53, March 2004.

[31] W. Lu and Y.-P. Tan, "A color histogram based people tracking system", in *Proc. IEEE International Symposium on Circuits and Systems*, vol. 2, pp. 137-140, 2001.

[32] Y. Chen and E. K. Wong, "Augmented image histogram for image and video similarity search", in *Proc. of the 7th Conference on Storage and Retrieval for Image and Video Databases*, San Jose, California, USA, 1998, vol. 3656, pp. 523-532.