5.1 Introduction

Binarization techniques, which use global, local, or adaptive thresholding, are the simplest methods for text localization. These methods are widely used for image segmentation, as these images usually include black characters on a white background, thereby enabling successful segmentation based on thresholding. The goal of the binarization module is to separate the pixels of a localized text region into categories of text and background. The output of the module is a binary image of the localized region suitable for input into a VOCR system. For document images, simple thresholding is typically sufficient to convert a gray scale image into a binary image suitable for OCR. This technique assumes that the text and background colors are uniform (typically black on white) so that the image histograms are bimodal. In video, however, text often appears against complex, nonuniform backgrounds. The text color may also vary due to uneven illumination of scene text, anti-aliasing, or due to bleeding caused by video compression. These problems are further compounded by the low resolution of video images, in which character may be two pixel or less in width. Due to these factors, it was found that algorithms which rely on histogram bimodality [1][2] are generally unsuccessful for video images. Noise introduced during the scanning process may cause the grayscale image to have more than two gray levels. However, the histogram of the grayscale image
is still system, with one peak corresponding to text pixels and the other corresponding to background pixels. Thresholding (either locally or globally) can be performed at a well-chosen gray level in the valley of the two peaks to give accurate binarization results. Many techniques have been studied for finding the ideal threshold. Figure 5.1 illustrates this binarization process for an image. The histogram for the fragment of a document figure in (a) is shown in (b). The strong peak at about corresponds to background pixels, while the weaker peaks around corresponds to text pixels. Main difficulties in binarization are related to the estimation of the appropriate threshold to separate text characters from the background. There is no difficulties in binarization if the text is correctly extracted.

Projected a text segmentation technique to extract text from any type of camera grabbed frame image or video. Colour based segmentation methodology was used to link consecutive pixels in the same direction by exploiting the general text properties. Light Edge Enhancement (LEE) was used to find a set of consecutive candidate points and enhance the edge between them. Next, heavy edge enhancement (HEE) was applied to remove or reduce motion blur from camera image sequences. This helped to treat camera images and video frames in the same manner [27].
Figure 5.1 Illustrates this Histogram and Median Filter

Segmentation: - An image or video frames may be defined as two dimensional light intensity function \( f(x, y) \) where \( x \) and \( y \) be the spatial plane co-ordinate and the amplitude of \( f \) at any pair of co-ordinates \( (x, y) \) is called as intensity or gray level of the image at that point. When \( x, y \) and the amplitude values of \( f \) are finite, desecreate quantities then we call the image as digital images and these images can be of various types such as multi-resolution, multi-spectral, halftone images etc. The multi-resolution, multi-spectral images perceives very dense information(with respect to \( f \) and \( x, y \) co-ordinates) and recognizing the objects from such type of image is an important aspect in analysis and processing of such images. Therefore the effective processing of such (multi-resolution, multi-spectral) images can be done by dividing them into subsequent
parts by image segmentation and processing each segment independently for recognizing an objects in the segment. There is currently substantial and growing interest in the field of video frames understanding where several research groups are developing and designing systems to automatically process and extract texture information from video frames. A thresholding technique must be able to segment a digitised image into different objects with similar properties. There are general technique for image segmentation and limit the scope of the problem to document images.

5.2 Review of prior binarization work

This section summarizes some of the past approaches to binarization of text in images and video frames. I have classified these algorithms into four main approaches to binarization: global thresholding, local thresholding, color clustering, and neural networks. The algorithms belonging to each approach are now discussed.

• J. Ohya et al [3] use a combined detection/binarization stage to extract characters from scene images. Text is assumed to be either black or white. Regions of the image with bimodal histograms are assumed to be text regions. Local thresholding is performed on these regions using the threshold selection algorithm described in [4]. Shape and size heuristics are applied to filter out non-text strokes.

• Lee and Kankanhalli [7] also use a combined detection/binarization stage. After quantizing the gray levels in the image, detection is performed by
searching for strokes with uniform gray level. Each potential character is thresholded using the gray level of its boundary. Post-processing removes components with suspicious aspect ratios, contrast, and fill ratios.

- Winger et al [6] use a modified form of Niblack’s Multiple and Variable Thresh-holding scheme [23], which employs variable thresholds based on mean local pixel intensity. After calculating the variance, the modified scheme uses a different multiplier and exponent.

Logical level binarization algorithm proposed by Kamel and Zhao [8] has proven to be fairly successful for this step. The logical level algorithm was developed to extract character strokes from complex backgrounds in document images such as cash amounts from noisy check images. This method is successful for extracting character strokes from gray scale images of video frames, provided that the text gray levels are darker than the background. If the text is lighter than the background, the localized region must be negated before applying the method.

Methods described in [9][10] perform poorly for text appearing against complex backgrounds. A variant of the method published in [11] is selected. The method performs logical level binarization on both the positive and the negative of the video frame, and the decision of determining the correct polarity is delayed until later. Connected components are then found in both binarized images.

Yih-Ming et.al [12] proposed a scheme to extract the caption text from various sports videos. Iteratively temporal averaging approach was used in caption
extraction process. To improve the image quality and to reduce noise spatial-image analysis was performed. Threshold value was determined using binarization process based on the global mean and the standard deviation of the gray level of the averaged video image [21]. Binarization may lead to holes and disconnectivity on video captions with blurred background. This was cured by morphological processing.

Antani et al. [17] used a number of algorithms for extracting text. They developed a binarization algorithm similar to Messelodi and Modena’s approach [18]. To refine the binarization result, a filtering stage is employed based on color, size, spatial location, topography, and shape. It is not easy to determine whether text is lighter or darker than the background. Therefore, many approaches assume that the text is brighter than the background. However, this is not always true. Therefore, these extraction algorithms operate on both the original image and its inverse image to detect text regions.

Chen et al. [19] proposed a text enhancement method which uses a multi-hypotheses approach. Text regions are located using their former work [20], where text-like regions are detected using horizontal and vertical edges. Text candidates are localized using their base line locations that are filtered by a Support Vector Machine. Localized text regions, which are gray scale images, are transferred to an EM-based segmentation stage.

### 5.3 An algorithm for Text Binarization in Video Frames

An algorithm is now presented for binarization of text in video.
There are several motivations for analyzing more than one frame during the binarization process. Lossy video compression methods introduce noise, but the noise varies from frame to frame. Simple temporal averaging can reduce such noise. Temporal integration is also helpful for background removal. Text often remains stationary while the background behind it changes or moves. Or the text may move, causing the background behind the text to change. In either case, temporal averaging can be used to smooth out the background. Global thresholding is a common approach. Algorithms in this category determine some grayscale threshold, and apply it to all pixels in a localized text region[25]. The methods differ in the strategy for choosing the threshold, and in the preprocessing and post-processing steps. A threshold is computed based on the pixels underneath the window. The pixel at the center of the window is then binarized based on this threshold. The binarized images are examined and a choice is made based on statistical information computed from the components of each image, such as the similarity in character height, character aspect ratio, vertical position, and horizontal spacing as shown in figure 5.2. The image with the more text-like features is then chosen.
Global Thresholding:

If the pixel values of the components and those of the background are fairly consistent in their respective values over the entire image, then a single threshold value can be found for the image. This use of a single threshold for all image pixels is called global thresholding. Processing methods are that automatically determine the best global threshold value for different images. The most way to automatically select a global threshold is by use of a histogram of the pixel intensities in the image. The intensity histogram plots the number of pixels with values at each intensity level. For an image with well-
differentiated foreground and background intensities, the histogram will have
two distinct peaks. The valley between these peaks can be found as the
minimum between two maxima, and the intensity value there is chosen as the
global threshold that best separates the two peaks [6]. Median filtering is one
kind of smoothing technique and effective at removing noise in smooth patches
or smooth regions of a text, but adversely affects edges. Often though, at the
same time as reducing the noise in an image, it is important to preserve the
edges [21]. A wide variety of ICA algorithm is used for a global thresholding
method. To find the IC that contains the foreground text, we examine the
connected components (CC) in the binarization of each IC. For each binarized
image, we extract the following features from the CCs: average aspect ratio,
variance of CC size, and the deviation from linearity of their centroids. A simple
linear classifier is designed to separate the text and non-text classes in the above
feature space. After binarization, we identify the connected components and
remove non-text portions based on size and aspect ratio. An effective method
to binarize text from colored scene text images with reflective, shadowed and
specular background. By using a blind source separation technique followed by
global thresholding, we are able to clearly separate the text portion of the image
from the background [26].
References


