Autonomous OCR Dictating System for Blind People

Christos Liambas
Dep. of Mathematical,Physical & Computational Sciences
Engineering School, Aristotle University of Thessaloniki
Thessaloniki, Greece
Christos.Liambas@gmail.com

Miltiadis Saratzidis
Department of Electrical & Computer Engineering
Engineering School, Aristotle University of Thessaloniki
Thessaloniki, Greece
M.Saratzidis@gmail.com

***Abstract*—In this study, the main idea is the development of an autonomous mobile system for dictating text documents via image processing algorithm for blind people. The system is constituted by the Raspberry Pi 2B - the mobile processing unit- and a pair of specially designed glasses with an HD camera and Bluetooth headset. The blind user should hold the book open (two pages) with his hands stretched straight at the level of his eyes; then a calibration procedure takes place in order to capture the best image. Therefore, 1-D signal transformation of the above image is produced in order to filter every text line. Finally, every word of each text line is identified via an OCR (Optical Character Recognition) method and the user hears it via TTS (Text To Speech) procedure.**

Keywords— Algorithm; Optimization; OCR; Text to Speech; Visually Impaired

I. INTRODUCTION

The number of Visually Impaired (VI) people worldwide is approximately 285 million [1], in other words more than 3.86% of the entire population. So far devices for improving their understanding of the environment have been invented as well as methods like the Braille, which was introduced as an option for studying engraved text. But this method has two major issues. Firstly, very few books are modified into Braille and secondly only the minority of the blind population can read Braille, in actual fact and according to surveys carried out recently, fewer than 10% of the US legally blind people can read Braille [2]. So, an algorithm has been developed and used on a custom hardware implementation in order for the blind people to read printed books in the same way as normal readers do.

II. PROBLEM STATEMENT

The main problem is the development of an algorithm which automatically "reads" every kind of printed book or written material (e.g. magazines) and turns it into speech, under certain circumstances.

Also, the above described problem encounters the following difficulties:

1) Hardware limitations (CPU, RAM, camera resolution)
2) Noise in the data (light conditions, reflections, blurring, shadows, finger existence etc.)
3) Data variations according to different font styles, font sizes and languages.

III. SYSTEM DESCRIPTION

The system is constituted by the *Raspberry Pi* 2B*,* as mobile processing unit, and a pair of glasses equipped with an HD camera, Bluetooth headset and LED light (Fig. 1). The aim of the hardware is to help VI people to have the experience in a natural way, without many wires or big cameras.

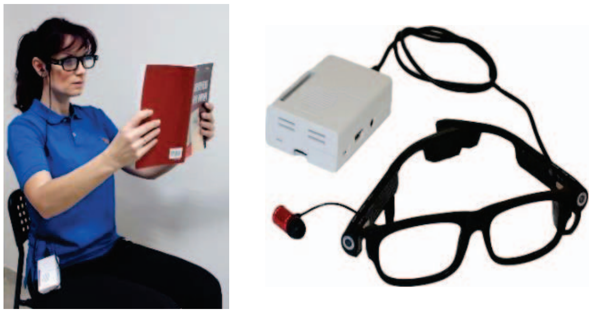


Figure 1. A typical use of the proposed system.

IV. PRIOR WORK

Since the beginning of the 20th century, various devices have been created in order to help VI people to read books. Some of the initial works, according to the literature, are the *Optophone* [3] and *Optacon* [4], which use a sensory substitution to translate the black and white text into time-varying chords of tones. Nowadays, smartphone applications and more advanced devices -which are also wearable- have been developed in order to help VI people to read text material by using OCR (Optical Character Recognition) and TTS (Text To Speech) technologies.

A representative device is *OrCam* [5] which has many hardware similarities with the proposed system, where both of them are constituted by a pair of glasses with micro camera and a processing unit. *OrCam* is designed to recognize not only text in specific printed material but also real life objects.

Another remarkable device is the *FingerReader* [6], a specially designed wearable ring with a micro camera, which is used in a similar way as the one used by VI people reading braille.

Also, mobile applications have been implemented [7] [8] [9] that take advantage of the modern cell phones as mobile processing units.

978-1-5090-2432-2/16/$31.00 ©2016 IEEE

V. ALGORITHM

The proposed method is inspired by the way that a normal reader is using the books. For that reason a human centered approach was invented, in order to achieve the highest degree of the same natural experience between a normal reader and a VI person.

Thus, a micro camera is placed in a pair of glasses near the eyes with a hidden processing unit and a book in the VI person's hands. The most important benefit is that the VI person would be more comfortable holding the book in a natural position like a normal reader. Furthermore, the best overall results are achieved by holding the book with both hands, because this minimizes the curviness of the book pages. By utilizing all the above ideas an algorithm was developed and it is constituted of three main phases:

1) Calibration
2) Line Separation
3) Data preparation for OCR and TTS

A. First phase

In the first phase of the algorithm the calibration procedure takes place. The main goal is to locate the best point of view for processing the book pages, according to the camera's relative position.

The blind user should hold the book open (two pages) with his hands stretched straight at the level of his eyes and then he starts to move the book towards him slowly but steadily (Fig. 1). During this process, video is being captured and analyzed continuously (320x240 resolution with 10fps), in order to guide the VI person via voice commands on how to improve the position of the book according to the requirements of the Best Position (BP). Best Position consists of the two book pages with the least possible rotation that should occupy the maximum height of the processing image but at the same time no part of the pages can be left out.

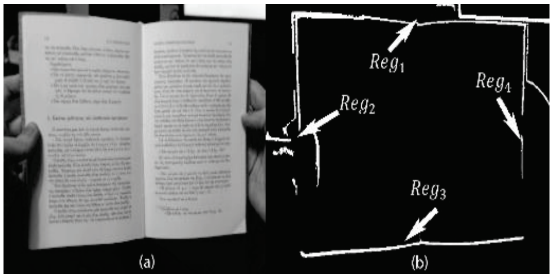


Figure 2. (a) One frame of the Book (b) Corresponding trace.

So, the guiding procedure leads to the BP by analyzing the moving traces of the book edges in a binary image, called Trace Image (TI) with $X \in [1, TI_{Width}]$ and $Y \in [1, TI_{Height}]$. TI is computed by applying Otsu's threshold method [10] in the subtracted video frames, as shown in Fig. 2b.

In order to separate the regions $Reg_s, s \in [1,4]$, a machine learning technique is applied on the TI image data and more specifically the Support Vector Machine (SVM). By using only two features, this learning technique, is taking the advantage of a considerable classifier according to the literature with respect to the data [11] [12]. Also, it should be noted that the processing time of the hypothesis training is not concerned in the overall performance due to the fact that the training procedure takes place only once, during the pre-processing phase.

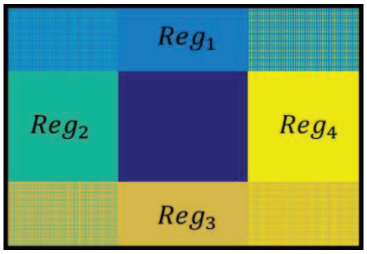


Figure 3. Input image for the SVM training method.

The main training feature is the spatial coordinates of white pixels and for that reason a separation procedure is required in order to classify them to the corresponding regions Reg_s , $s = 1, \dots, 4$. Taking an objective approach to classify the corner point's means that it should be defined a hypothesis with suitable boundaries between the overlapping areas (Fig. 3). This is done by applying a Gaussian kernel on the data as shown in Figure 4.

Finally, for optimization reasons and without affecting computed results, the noise in the TI image is not removed intentionally.

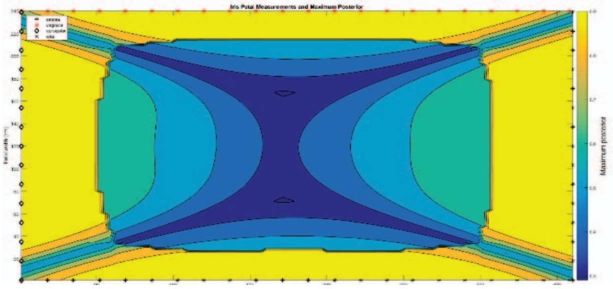


Figure 4. SVM after training.

Thus every white pixel in the TI image is classified in one of the four regions by using the above described SVM model. Then for every region $Reg_s, s \in [1,4]$ the algorithm calculates the center of mass (X_s, Y_s). Also at this point the rotation φ of the book is calculated. This is being done by calculating the angle μ (Fig. 5) from the (X_1, Y_1) and (X_2, Y_2) :

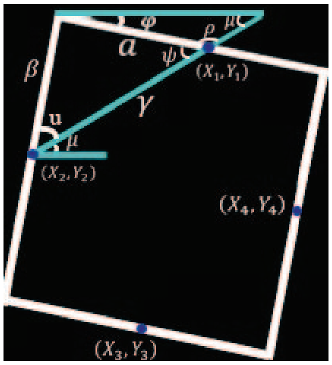


Figure 5. A book position rotated by φ.

$$\beta = \frac{\sqrt{(X_1 - X_3)^2 + (Y_1 - Y_3)^2}}{2}$$

$$\gamma = \sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2}$$

$$\cos(u) = \frac{\beta}{\gamma}$$

$$\psi = 90^\circ - u \qquad (1)$$

$$\psi + \rho = 180^\circ \overset{(1)}{\Rightarrow} \rho = 90^\circ + u$$

$$\mu = \tan^{-1}\frac{(Y_2 - Y_1)}{(X_2 - X_1)}$$

$$\varphi = 90^\circ - \mu - \cos^{-1}\left(\frac{\beta}{\gamma}\right)$$

So, the guiding voice commands for improving the book position are produced according to the rotation φ , the center of mass (X_s, Y_s) and the BP requirements. As every page of the book is turned, the calibration procedure is repeated until the new BP is found, then an image with higher resolution (1280x720, grayscale) is captured and the book can be closed. The algorithm crops that image, called Best Image (BI) with $X' \in [1, BI_{Width}]$ and $Y' \in [1, BI_{Height}]$, which contains only the pages of the book without background. The width of BI is smaller than 1280 but almost the same height, according to the BP requirements.

B. Second phase

In the second phase, the algorithm locates the text lines for each one of the two pages BI_{left} $(x \in [1, BI_{Width}/2]$ $, y \in [1, BI_{Height}])$ and BI_{right} $(x' \in [BI_{Width}/2, BI_{Width}/4]$ $, y' \in [1, BI_{Height}])$ contained in the BI respectively. This is done with the help of the Horizontal Projection (HP) and Vertical Projection (VP) signals and without loss of generality the above analysis is based on the BI_{left} image. So, HP signal is defined as the sum of the grayscale intensities in every column, whereas the VP signal is defined as the sum of the grayscale intensities in every row (Fig. 6):

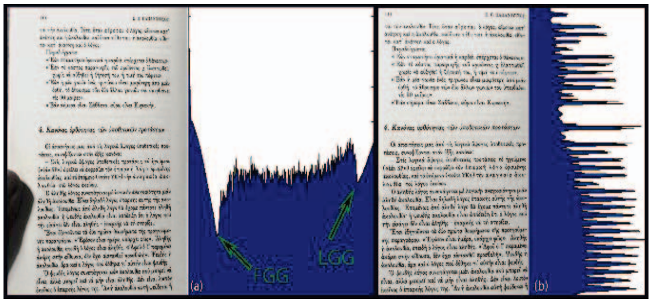


Figure 6. (a) A page with the corresponding HP signal. (b) The cropped page without the finger and the corresponding VP Signal.

$$HP(a) = \sum_{y=1}^{BI_{Height}} (255 - BI_{left}\ (x,y))\ , a \in [1, BI_{Width}/2]$$

$$VP(\xi) = \sum_{x=FGG}^{LGG} (255 - BI_{left}\ (x,y))\ , \xi \in [1, BI_{Height}]$$

Is should be crystal clear that a procedure takes place when the HP signal is produced, in order to remove finger existence and grayscale gradient in the most left and right parts of the BI_{left} image respectively. So, this kind of data is excluded completely by detecting First Great Gap (FGG) and the Last Great Gap points (LGG) in the HP signal:

$$FGG = min\{HP(a)\},\ a \in [1, BI_{Width}/4]$$

$$LGG = min\{HP(a)\},\ a \in [BI_{Width}/4, BI_{Width}/2]$$

The above boundaries are based on the basic assumption that the finger could not be extended to more than half of the BI_{left} image. Also, the maximum density of the grayscale gradient –which is caused by the book formulation in the center- could not be extended in the center of the page.

Following the above procedure, the VP signal is produced between the boundaries of the FGG and LGG (without the left and the right problematic areas).

It should be noted that the peaks of the produced VP signal, which derive from the length of the text lines, significantly deviate from the noise.

The gradient G between the adjacent signal values is computed by the following equations:

$$\frac{dVP}{d\xi} = G(\xi, \xi + 1) = \frac{VP(\xi + 1) - VP(\xi)}{\xi + 1 - \xi} =$$

$$= VP(\xi + 1) - VP(\xi)$$

The VP signal contains P_k peaks, where k is the number of k_{th} peak which is limited between two valleys (V_k^{min}, V_k^{max}). Thus a peak P_k notices the existence of a text line -which is a local maximum- and the corresponding boundaries are represented by (V_k^{min}, V_k^{max}) in Fig. 7a.

$$P_k \begin{cases} G(\xi - 1, \xi) > 0 \ \ and\ G(\xi + \omega, \xi + \omega + 1) < 0 & (1) \\ and\ G(\xi, \xi + 1) \approx 0, \ldots, G(\xi + \omega - 1, \xi + \omega) \approx 0 & \\ \mathrm{G}(\xi - 1, \xi) > 0 \text{ and } G(\xi, \xi + 1) < \ 0 & (2) \end{cases}$$

$$V_k^{min} \begin{cases} G(\xi, \xi + 1) \approx 0 \ \ and\ G(\xi + 1, \xi + 2) > 0 & (1) \\ G(\xi, \xi + 1) < 0 \ \ and\ G(\xi + 1, \xi + 2) > 0 & (2) \end{cases}$$

Also the V_k^{max} is calculated respectively. So, the VP signal is formed into triangles T_k by using the corresponding triplets $(P_k,\ V_k^{min}, V_k^{max})$, where a triangulation method takes place for surveying purposes (Fig. 7b). First of all, the VP signal is used for filtering the good and the bad data. Good data corresponds to the higher level values which include text lines in most cases that are useful for the OCR procedure (last phase of the algorithm). On the other hand, bad data corresponds to lower level values which include noise in most cases.

Throughout the proposed formulation, the following difficulties were identified, but were effectively overcome without human intervention (completely automatically without subjective parameterizations):

1) High variations without following a specific “pattern”
2) Every text line may correspond to two or more peaks in the VP signal, because of the horizontal compactness of the letters.
3) Unknown data features and number of text lines
4) Prior results cannot be used as a guide for future ones because every single text line is examined as a unique case.
5) Noise in the data.

So, for every pair of triangles T_k, T_{k+1} (two adjacent peak areas), an analysis takes place about the variability of the vertices’ size. The main idea is to compare adjacent triangles in order to combine and consider them as one peak and not seperated peaks. But under specific circumstances some triangles Tk should be left untouched. The criteria for determining the discrimination is the following:

The pair T_k, T_{k+1} is considered one peak area $max\{P_k, P_{k+1}\}$ with valleys $V_k^{min}$and V_{k+1}^{max}, if and only if $\left|P_k - V_k^{min}\right| \gg |P_k - V_k^{max}|$ and $\left|P_{k+1} - V_{k+1}^{min}\right| \ll |P_{k+1} - V_{k+1}^{max}|$. Of course, V_k^{max}may be too close or even equal with V_{k+1}^{min} (Fig. 7c).

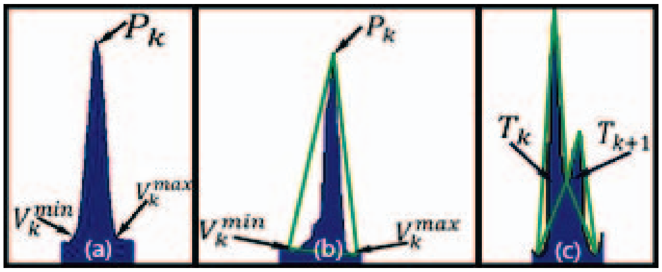


Figure 7. (a) Representation of the peak P_k and the corresponding bounds $V_k^{min}$and V_k^{max}(b) The corresponding triangle T_k (c) Two adjacent triangles T_k,T_{k+1}that are considered as one peak area.

As a result, by triangulating the VP signal with the above method showed that each pair of valleys (V_k^{min}, V_k^{max}) corresponds to the boundaries of every text line, which distinguishes the text line from the background.

However, there is a special issue concerning the lower level value peaks which may include short text lines (e.g. end of paragraphs) or noise. Clearly, for such a situation, a boundary value that triggers this kind of signals cannot be constructed; thus this issue is discussed in the next section.

C. Third Phase

In the third phase, separation and binarization of words takes place for every text line, so that they can be used by the OCR and TTS procedures. Initially, in every column x the difference (D_x) between the minimum and maximum intensity (In_{min} , In_{max}) is calculated.

The aim is to separate “Words” (W) and “Spaces” (S) in two different groups according to the D_x of every text line that is detected from the previous section. Group W contains only columns with letters in contrast to group S that contains only columns without letters (only background). It is worth noting that the variation of intensity in a column which intersects with a letter is much greater than the variation of intensity in an area without a letter. So, columns which contain letters maximize the D_x parameter in contrast to columns without letters. Thus, in order to classify the columns, the most representative algorithms K-Means and K-Medoids [13] were examined and analyzed, based on their initial approach, but only the K-Medoids method has worked, because it is more robust to noise and outliers [14][15].

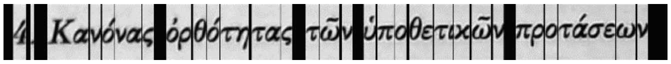

4. Κανόνας ὀρθότητας τῶν ὑποθετικῶν προτάσεων

Figure 8. A typical text line with the identified spaces.

Moreover, by grouping the adjacent columns which minimize D_x parameter, the corresponding widths are classified again into two groups with the K-Medoids method, in order to distinguish the spaces between the letters and the spaces between the words. Consequently, all the useful characters of every text line are detected and grouped into words, because the distance (width of the adjacent columns) between the characters which are contained in a word is much smaller than the distance between the words (Fig. 8). A local threshold is applied in every word by using the Otsus method [10], where an example of the final text line is shown in Figure 9.

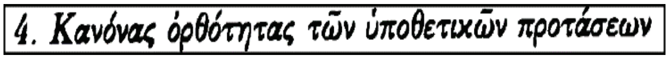

4. Κανόνας ὀρθότητας τῶν ὑποθετικῶν προτάσεων

Figure 9. The final binary image.

In the final stage of the algorithm, the OCR procedure is applied, which is an implementation of the method described in [16]. Therefore, TTS procedure [17] enables the system’s voice function to dictate the recognized words to the VI person via Bluetooth headset.

D. Optimization

In order to decrease the processing time, some optimization techniques were applied, especially in the case where the algorithm is executed on a mobile processing unit (e.g. raspberry pi).

An enormous reduction of the captured video data occurs during calibration phase, with resolution 320x240@10fps (instead of the default 1080p@30fps). So, the required time of the calibration procedure is 0.2915sec for every trigger, instead of 6.89 sec for the default configuration.

During the dictation process, the algorithm continuously analyses future text data which will be used for dictating. This is vital because the time for dictating a word is approximately 1.65sec, which is much longer compared to the time of processing every word which is approximately 0.09sec.

VI. COMPUTATIONAL RESULTS

To evaluate the performance of the proposed algorithm which has been implemented in MATLAB, a test bed of images is proposed with different types of books and environmental conditions. The computational experiments have been conducted on a raspberry pi 2B with 900MHz quad-core processor and 1GB RAM DDR2.

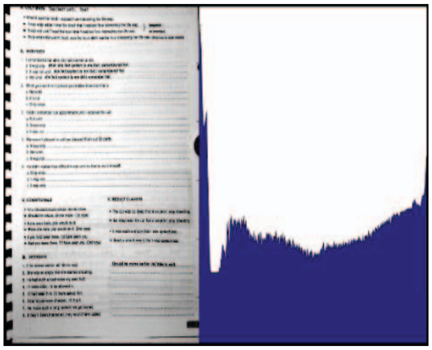


Figure 10. (a) Workbook (b) Corresponding HP Signal.

In the first instance, a special photocopied book with spiral connected pages is used. The figures 10b and 11b shows the corresponding HP and VP signals, which indicate several issues that derived from the initial captured book image. Firstly, the right margin is not cropped correctly by the HP signal (Fig. 10b), in contrast to the left side of the book, because of the following two reasons:

1) Non-uniform lighting conditions and
2) Inconsiderable distance between the finger and the text.

Also, it is important to note, that the OCR procedure is unable to identify the answer area (dots) as a non-text area and the user is not informed for this special case.

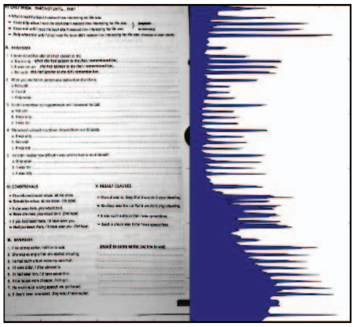
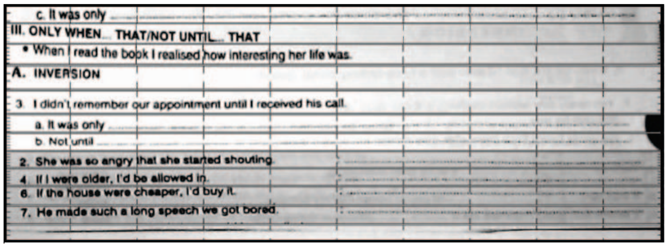


Figure 11. (a) Cropped Image (b) The corresponding VP signal.

Additionally, one more issue is the illumination variation of the image. The light source is located near the center of the book page; therefore the intensity of the light spreads radially. For that reason the VP Signal (Fig. 11b) has an overall curve form. Despite this illumination problem, the bounds of the lines are computed correctly (Fig . 12) by processing the produced peaks of the VP signal.



c. It was only
III. ONLY WHEN... THAT/NOT UNTIL... THAT
• When I read the book I realised how interesting her life was.
A. INVERSION
3. I didn't remember our appointment until I received his call.
a. It was only
b. Not until
2. She was so angry that she started shouting.
4. If I were older, I'd be allowed in.
6. If the house were cheaper, I'd buy it.
7. He made such a long speech we got bored.

Figure 12. Separated text lines.

In the following example, VP and HP signals were produced in a completely dark room. The aim is to evaluate the algorithm's performance by using only the LED light of the glasses. This is possible because it is known that the VI persons are not using light sources in their privacy.

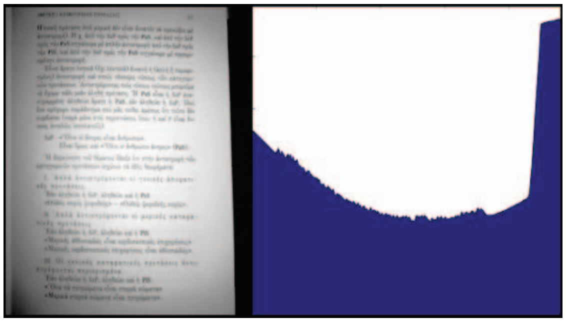


Figure 13. (a) Book page with poor light conditions (b) Corresponding HP signal.

The above described conditions have a great impact on the HP and VP signals. More specifically the HP signal has a greater contrast between the book and the background, so the cropping procedure is very precise (Fig 13b). Also, VP signal appears to have a curved form, because it is affected by the position of the LED light and what it falls upon (Fig 14b). But, the proposed algorithm is taking advantage of the area near the peaks of the VP signal, which are still produced correctly.

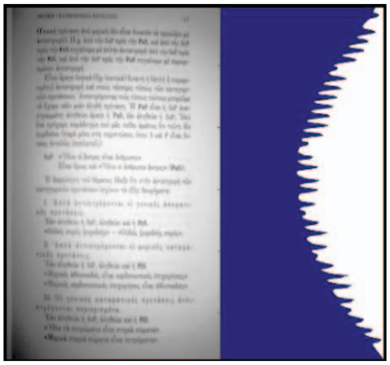


Figure 14. (a) Cropped image (b) Corresponding VP Signal.

A book page in ancient Greek (Aristotle, Nicomachean Ethics) was chosen in order to evaluate the performance of the algorithm in a multitoned language (Fig. 15). In some

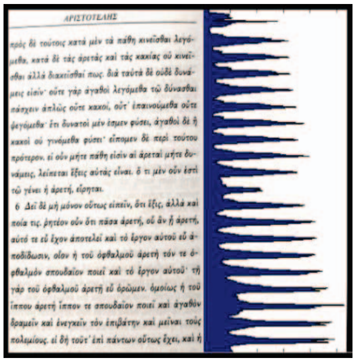
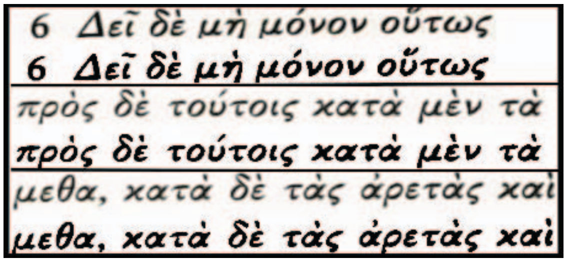


Figure 15. A book written in ancient Greek language, with the corresponding VP signal.

languages, pitch accent is important, because a word's meaning could be different, depending on which syllable is stressed.

The text lines are found despite the irregular VP signal and transformed into binary images as shown in Figure 16. Binarization is successfully calculated on the letters as well as in the tones.



6 Δεῖ δὲ μὴ μόνον οὔτως
6 Δεῖ δὲ μὴ μόνον οὔτως
πρὸς δὲ τούτοις κατὰ μὲν τὰ
πρὸς δὲ τούτοις κατὰ μὲν τὰ
μεθα, κατὰ δὲ τὰς ἀρετὰς καὶ
μεθα, κατὰ δὲ τὰς ἀρετὰς καὶ

Figure 16. Three text lines with the corresponding binary images in ancient Greek language.

Another instance with special issues is the scientific books, which contain figures, arrays and math equations (Fig. 17). Thus a page from a mathematical book was used in order to evaluate the performance under these special circumstances.

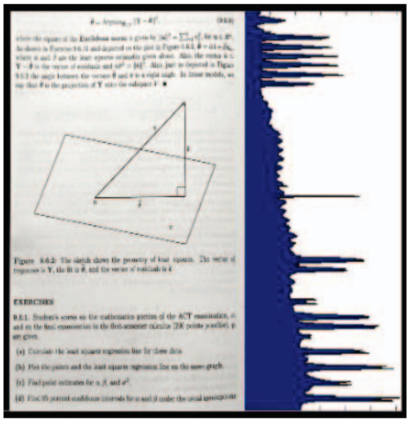
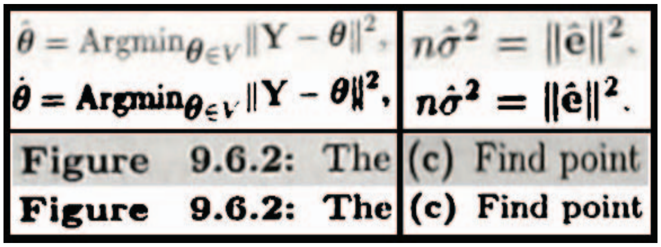


Figure 17. A mathematical book with the corresponding VP Signal.

In this example the VP signal (Fig. 17) was affected by the vertical lines, which produce abnormalities to the VP signal. As expected the line separation procedure accepts the above peak incorrectly. However, the final result is not affected, because the third phase rejects this case. Finally, the words were transformed (Fig. 18) and identified by the OCR procedure. On the other hand the equations were transformed into binary form but not identified by the OCR.



$\hat{\theta} = \mathrm{Argmin}_{\theta \in V} \|Y - \theta\|^2$,	$n\hat{\sigma}^2 = \|\hat{e}\|^2$.
$\hat{\theta} = \mathrm{Argmin}_{\theta \in V} \|Y - \theta\|^2$,	$n\hat{\sigma}^2 = \|\hat{e}\|^2$.
Figure 9.6.2: The	(c) Find point
Figure 9.6.2: The	(c) Find point

Figure 18. Four text lines with the corresponding binary images from a scientific book.

In the last instance, a book page is used which contains mixed text and picture, as shown in Figure 19. There are some noticeable issues such as the illumination variation and the grayscale background of the text lines that appeared after the first ten lines. Obviously, these special issues have a great impact on the VP signal.

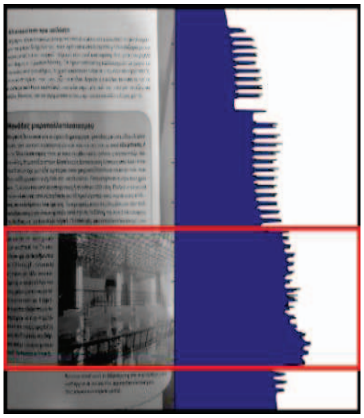


Figure 19. A book page with a picture next to the text lines and the corresponding VP signal.

Firstly, the background intensity is changing after the 10_{th} line to more dark values and consequently it causes the rise of the values in the VP signal to a higher level. Secondly, the source of light produces a constantly upward trend in the VP signal. Also, the picture (red box) creates a variable intensity that corrupts the peaks and makes it harder to detect them. However, the algorithm was able to detect it despite the corruption.

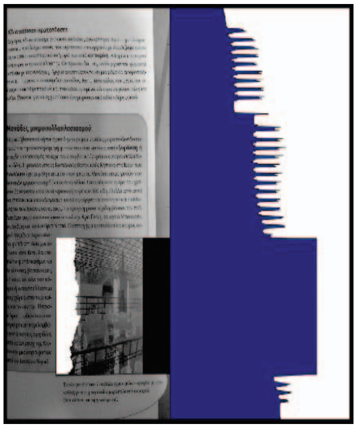


Figure 20. The intensity of the picture was changed intensively in order to achieve a non-discriminate signal.

Finally an artificial example is produced in order to illustrate the conditions that the proposed algorithm doesn't work in. If a picture with variable intensity eliminates the peaks of the VP signal and creates a flat signal (fig. 20) the algorithm fails to locate the text lines.

This device was used on both visually impaired and completely blind people, with similar results. First of all, the users note smooth experience in the calibration procedure as well as in the dictation of the words. The calibration procedure takes approximately 20sec at maximum.

Also, it should be mentioned that the users successfully follow the voice commands during the calibration procedure, so there is a successfully "communication" between of them. A remarkable point is the capability for storing captured book page images in order to dictate them at a later time. Finally, the most important thing is that a user of this device feels exactly the same as a normal reader who reads a book with glasses.

By summarizing the general results for the separation of the lines and the OCR identification rates were approximately 94.2% and 88.7% respectively.

VII. COMPARISON

At this point a comparison takes place between the proposed device and the prior works. As mentioned in section III, *OrCam* has many hardware similarities with the proposed device, but also some noticeable differences. The main disadvantage is that the users with total loss of vision would have some limitations with this device, because it is developed to identify what the user's finger is pointing to. Also, this device is not designed for identifying words from a book for which the blind person can't use both hands to stretch the book and avoid any possible curviness of the pages that causes poor results.

Another wearable device is the *FingerReader* that uses a specially designed wearable ring with a micro camera. The most important advantage is the fact that this device has the ability to read in indistinguishable places like the spine of the book. On the other hand, the drawback is that the user must have continuous contact with the book in contrast with the proposed system where the user's hands are freed after the calibration procedure. The user is listening to the already captured data that is stored and analyzed in the system's RAM, so it allows for doing multiple tasks.

Mobile applications are widely used, but these approaches have some drawbacks. First of all, mobile hardware variation, depending on the phone, the accuracy and the processing time varies. Furthermore the user should use his hand as the basis for the camera thus the calibration procedure becomes unstable with focus problems and it is very hard to handle it (find the proper rotation, height and position). Also a crucial disadvantage of the hand calibration method is that the user has his hands occupied and this makes the use of a book very difficult, unpleasant and with poor results. Finally it is important to note that the user doesn't have to buy an additional device and for this reason the cost is significantly less.

VIII. CONCLUSION AND FURTHER RESEARCH

Some concluding remarks and suggestions are obtained for the proposed mobile system. The main idea is a device which is constituted by the Raspberry Pi 2B, a pair of glasses with an HD camera and Bluetooth headset that automatically dictates text books. The camera is capturing data continuously and the algorithm is analyzing texts without any human intervention. The successfully recognized words are dictated via a Bluetooth headset to the visually impaired person. With the contribution of the optimization techniques that have been developed, the computational cost has decreased so that it can be implemented in a mobile system.

However, there are some remarkable modifications which can extend the usefulness of the suggested system:

1) Implementation of an improved OCR method that can identify mathematical equations, hand writing and ancient Greek language as well as other languages (multilingual support).
2) Maximizing the utilization of the hardware via the parallel core processing techniques.
3) Ability for analyzing various text sources such as monitors and newspapers.
4) A major improvement for processing text data from road names and signs under different circumstances.

REFERENCES

[1] World Health Organization: visually impaired worldwide, August 2014.

[2] NBC Article, "fewer blind Americans learning to use braille", 2009.

[3] d'Albe, E. F. "Optophone". Proceedings of the Royal Society of London. Series A 90, 619, pp. 373–375, 1914.

[4] Linvill, J. G., and Bliss, J. C., "A direct translation reading aid for the blind". Proc. of the IEEE 54, 1, pp. 40–51, 1966.

[5] OrCam Technologies Ltd, "OrCam - See for Yourself", June 2014.

[6] R. Shilkrot, J. Huber, C. Liu, P. Maes, and N. S. Chandima, "FingerReader: A Wearable Device to Support Text Reading on the Go," CHI '14 Ext. Abstr. Hum. Factors Comput. Syst., no. VI, pp. 2359–2364, 2014.

[7] Roberto Neto, Nuno Fonseca , "Camera Reading For Blind People" , ELSEVIER , 2014.

[8] KNFB reader application, 2016.

[9] Voice application from AppStore, 2016.

[10] Otsus, N., "A Threshold Selection Method from Gray-Level Histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 9, No. 1, pp.62-66, 1979.

[11] Diego Alejandro Salazar1, Jorge Iván Vélez2, Juan Carlos Salazar, "Comparison between SVM and Logistic Regression: Which One is Better to Discriminate?", 2012.

[12] Ming-Chang Lee, and Chang To, "Comparison of Support Vector Machine and Back Propagation Neural Network in
Evaluating the Enterprise Financial Distress", (IJAIA), 2010.

[13] Singh, Shalini S., and N. C. Chauhan. "K-means v/s K-medoids: A Comparative Study", National Conference on Recent Trends in Engineering & Technology, 2011.

[14] Wyld, David C. and Wozniak, Michal and Chaki, Nabendu and Meghanathan, Natarajan and Nagamalai, Dhinaharan, "Advances in Computing and Information Technology", *Springer Berlin Heidelberg*, Vol. 198, pp 472-481, 2011.

[15] Jin, Xin and Han, Jiawei, "*Encyclopedia of Machine Learning*", Springer US, pp 564-565, 2010.

[16]OCR. Proceedings of the International Workshop on Multilingual OCR, 2009.

[17] Microsoft Win32 Speech API (SAPI) 5.1.