

Scene to Text Conversion and a Cymatics Based Configurable Text Perception

Saeed Mian Qaisa

College of Engineering, Effat University, Jeddah, KSA
sqaisar@effatuniversity.edu.sa

ABSTRACT

This paper propose an original approach of achieving a Cymatics based visual perception of image-extracted text. In this context, an effective approach for automated text detection and recognition for the natural scene images is proposed. The incoming image is firstly enhanced by employing CLAHE and DWT. Afterwards, the text regions of the enhanced image are detected by employing the MSER feature detector. The non-text MSERs are removed by employing the geometrical and contour based filters. The remaining MSERs are grouped into words or phrases by finding out similarities between them. The text recognition is performed by employing an OCR function. The extracted text is sequentially analysed on character by character basis. Each character is converted into a methodical acoustic excitation. Finally, these excitations are converted into the systematic visual perceptions by using the phenomenon of Cymatics. The system functionality is tested with an experimental setup. For the case of studied natural scenes, the suggested approach achieves 80% precision in text localization and 53% precision in end-to-end text recognition. The devised system principle is novel and can be employed in various applications like visual art, encryption, education, integration of impaired people, etc.

Index Terms: Cymatics, text detection and recognition, Optical Character Recognition (OCR).

1. INTRODUCTION

This work focus on realizing a systematic visual perception of the image-extracted texts. Text is the most common way of communication. It is frequently embedded into documents and can be often found into scenes as a means of illustration and explanation. The recent evolution in the performance of mobile devices, in terms of the computational capability and the imaging resolution, and in the robustness of computer vision and pattern recognition techniques makes it viable to effectively address the mystifying problem of text identification in images and videos [17-21]. In this context, the text detection and recognition problems have received an increased attention [22-25].

These automatic image-based text extraction systems are founded on the processes of machine based image enhancement, text detection and recognition. The principle is to firstly enhance the incoming image. It improves the text candidates detection process accuracy. The next step is to verify the text candidates and filter out the non-text ones. The qualified text candidates are segmented. Afterwards, the extracted features of these segments are matched with reference templates. It is done in order to recognize and classify the text candidates. It allows realizing an effective, robust and automated image-based text extraction system [21, 37-40].

The text detection and recognition from images possess many hurdles like the lower quality, degraded data, variations of text layout and fonts, existence of uneven illumination, low resolution, multilingual content, etc. These hurdles result into the low detection rates mostly less than 80 percent [26]. Therefore, the end-to-end text recognition rates of the state of the art methods is mostly

less than 60 percent [19-21]. However, the image enhancement techniques, MSER features extraction and OCR algorithm has solved these shortcomings up to a certain extent in the case of machine printed document images [27]

The process of text recognition from images and videos has diverse applications [22-25, 28-36].

Cymatics is the study of visualizing the effect of acoustic waves [1]. The objective of Cymatics is to make the sound waves and vibrations observable in order to take advantage of the human sense of sight, as it is the most discriminating human sense [1-7].

This work aims to achieve a Cymatics based systematic visual perception of the image-extracted text characters. It can be realized by smartly combining the automatic image-based text detection and recognition modules with the phenomena of Cymatics.

The proposed system tactfully employs the phenomena of Cymatics in order to achieve a systematic visualization of image-extracted texts. The Section 2, describes the devised system principle. Details of different system modules are also discussed in this Section. The system implementation and experimentation results are described in Section 3. Certain aspects and prospects of this work are discussed in Section 4. Section 5, finally concludes the article.

2. MATERIALS AND METHODS

The principle of suggested system is depicted with the block diagram, shown on Figure 1. It shows that the system captures an image via a camera or a scanner. It could be a natural scene image or image of a book page. The captured image is processed and the text, available on this image, is extracted. Later on, the extracted text characters are sequentially converted into systematic excitations via the front-end interface. Furthermore, each excitation is converted into sound and passed to the amplification stage. The amplified sound is transferred to the Cymatics module for a systematic visual perception of the recognized text character. The different system blocks are described in the following sub sections.

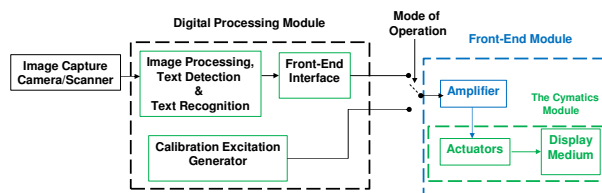


Figure 1: The proposed system block diagram.

2.1. The Image Processing, Text Detection and Recognition Tools

The Text based images can be categorized into two main types: The structured and the unstructured. The structured images contain the graphic text. It is the machine printed text where the position and font of the text is already known. The graphic text is found in captions, subtitles and annotations in video and born-digital images on the web and in Emails [42]. The unstructured images contain text from scenes. This type of text appears in natural scenes and can also

be a hand written script. In this case, the position and font of the text are undetermined and random [40]. Scene text is found in road signs, packages, clothing, touristic notices, etc. [41].

The text detection and extraction from an image or video is a challenging task. Its accuracy depends on the precession of image acquisition and processing chain, employed text enhancement and classification approaches and on the condition of system environment. The system has to encounter various issues like the uneven luminosity, blurring, complexity of scene, distortion, various fonts, multilingual text, etc. [40, 43, 44].

Methods of text detection and recognition can be categorized into two main types, the integrated and the stepwise methods [40]. The text detection and recognition processes are treated separately in stepwise techniques. On other hand, the integrated techniques, deal the text detection and recognition processes together. Both approaches possess advantages and disadvantages. An appropriate choice should be made as a function of the intended application [40].

The employed text detection and recognition chain belongs to the category of stepwise text detection and recognition [40]. The input image is firstly enhanced by converting it into a grayscale image. Later on, an appropriate combination of Contrast-Limited Adaptive Histogram Equalization (CLAHE) with discrete wavelet transform (DWT) is employed to enhance the grayscale image. Firstly, the denoised grayscale image is decomposed into low and high frequency components by using the DWT. Then, the low-frequency components are enhanced by using CLAHE. However, the high-frequency components are kept as it is. It is because of the fact that the high-frequency components correspond to the detail information. They contain most of the original image noises. Therefore, the application of CLAHE can result into the loss of information that is contained by the high-frequency components [51]. Finally, the image is reconstructed by taking inverse DWT of the new components. The over-enhancement is avoided by weighted average of the reconstructed and the original images. This method performs well in preserving the image details while suppressing the noise [51].

The enhanced image is passed through the module of text candidates regions detection. In literature, various approaches have been employed for identifying the text candidates regions like Stroke Width Transform (SWT), Stroke Feature Transform (SFT), External Regions (ERs), Maximally Stable External Regions (MSERs) [40, 49], etc. The MSER features detector is well known for its text detection robustness [44]. In this context, it is employed for detecting the text candidates from the incoming image. The MSER detector is based on the idea of detecting regions with same characteristics within a wide range of thresholds. The intended objects regions that are extracted by the MSER detector, in an incoming image, are called MSERs. All pixels within MSERs have either higher or lower intensity than pixels on their outer boundaries. This property of MSERs explains the term External Regions. It works fine for text because of the consistent colour and high contrast of text with respect to background. It results in a stable intensity profile which allows the MSER detector to easily achieve the Stable External Regions [44].

In natural scenes, certain objects like buildings, paintings and symbols appear like text characters. In this case, some non-text regions can also be detected as text candidates by the MSER detector. In this context the detected text candidates are verified by employing the basic geometric properties and the stroke width variations [45, 46]. The geometric properties of text regions are different from the non-text regions. Therefore, the non-text regions can be filtered out by employing some geometry based rules. Following the work of Neumann and Gonzalez [45, 47], the geometric properties of aspect ratio, eccentricity, Euler number, solidity and extent are employed to filter out the non-text regions [45, 47]. Further non-text

regions removal is achieved by employing the measure of stroke width variations. It is a measure of the width of curves and lines that make up a character. In fact, the text regions tend to have a little stroke width variations as compared the non-text regions [46]. In order to employ the stroke width variations metric a suitable stroke width threshold should be employed as a function of the intended font style. Later on, this threshold is applied on all detected text candidates in order to classify them as text or non-text [46].

The segmentation is a next step after non-text regions filtering. It is assumed that all the detection results, at this stage, are composed of individual text characters. The individual text characters are grouped into words or phrases before proceeding to the recognition stage. It improves the performance of recognition stage and enables to identify the whole words or phrases and to transform the incoming image in meaningful text [49]. The connected component (CC)-based approach is employed in order to properly group the individual text characters into words and sentences. This approach exploits the fact that characters in the same word exhibit similar properties, like nearly constant pixel value, stroke width, mutual distance, etc. These properties are employed for merging individual text regions into words and phrases. In this way, the neighboring text characters are identified and are then placed within the same bounding box [40, 49].

After segmentation the text recognition process is performed. The Optical Character Recognition (OCR) is well known for a robust text recognition [40, 48]. Therefore, it is employed in the devised system. The OCR module, trained on the synthetic data, treats each segmented bounding box separately [50]. It assigns a Unicode label to each bounding box. These labelled elements represent different nodes in a direct acyclic graph. The final and sorted sequence of labels is determined as an optimal path [50]. The OCR employs sets of parameters to classify each text character in the segment. It compares the text candidates parameters with templates. Afterwards, the classification decisions are made on the basis of best match findings [40, 48, 50].

2.2. The Front-End Interface and Calibration Generator

The suggested system functions in two different modes. These are the operation mode and the calibration mode. The choice of mode is made by the user and it is achieved via a software-based switch.

During operation mode, the front-end interface provides a liaison between the Image processing, text detection and recognition unit and the system front-end module. It converts the image-extracted text into a sequence of characters. Afterwards, each character is treated separately in a sequential fashion. Recognition of the incoming text character is made. Then the identified character is converted into a methodical sound via a predefined lookup table. For each identified text character a specific sound is generated. This lookup table is designed, during the calibration process, as a function of the employed display medium and actuator.

In calibration mode, the calibration excitation generator is used to generate a sequence of monotone excitations. Amplitudes and frequencies of these excitations are systematically varied within a predefined range. The choice of this range is a function of targeted application and employed system front-end module characteristics. These excitations are converted into sounds and are applied to the front-end module. In this way excitations which result in distinguishable and repeatable Chladni patterns are registered in an excitation lookup table.

2.3. The Front-End Module

This module composes of an amplifier and a Cymatics module. The Cymatics module consists of an actuator and a display medium.

The amplifier adapts the excitation sound magnitude, generated by the front-end interface. It is done in order to adapt it to the employed actuator acceptable level.

The actuator transforms the augmented excitation acoustic waves into vibrations. These vibrations are employed to excite the display media. It transforms the excitation, generated by actuators, in a visual perception. A unique and repeatable visual pattern is generated corresponding to each specific excitation. In the studied case, the metallic Chladni plates and sand particles are employed as the display media.

According to Chladni's law, for fixed centre flat surfaces, a relationship between the frequency of mode of vibration and complexity of the resulted pattern can be formally expressed by Equation 1 [6]. Where, f_b is the frequency of mode of vibration. C and b are coefficients which depend on the type of surface material. m and n are respectively representing the number of linear and radial nodes. Equation 1, shows that there is a proportional relationship between the frequency and number of both radial and linear nodes. Therefore, when the frequency is increased the number of both nodes increase which results in more complex generated patterns.

$$f_b = C(m + 2n)^b \quad (1)$$

It concludes that for a systematic excitation the generated Cymatics pattern is a function of the employed medium shape, its properties and the excitation characteristics.

3. RESULTS AND DISCUSSION

3.1. Results

The A system prototype is realized in order to demonstrate the proposed system functionality. The commercially available appropriate components are employed to achieve an expeditious prototype development. The system experimentation setup is shown on Figure 2. A personal computer (PC) is employed as a base of the digital processing module (cf. Figure 4). The text detection and recognition chain, the front-end interface and the calibration excitation generator are realized with the MATLAB based specifically developed applications [53].



Figure 2: The proposed system experimentation setup.

The excitation, generated by the front-end interface is converted into sound by using the PC built in speaker. This sound signal is further augmented with the amplifier module. It is done in order to make the excitation compatible with the actuator module. A 100-watt audio amplifier is employed in this context. It possess a Signal to Noise Ratio (SNR) of 103dB and frequency range between 20Hz to 20kHz. The interface between PC and amplifier is realized with an audio cable.

The actuator is realized with a SF-9324 wave driver [54]. The amplified excitation is applied to SF-9324. The interface between amplifier and wave driver is realized with an audio cable with the audio jack on one end and the crocodile connectors on the other end. The SF-9324 frequency range is between 0.1Hz to 5kHz. It can accept a maximum input of 6V. The maximum input current is lim-

ited to 1 ampere. The wave driver transforms the input sound waves into vibrations. Amplitudes and frequencies of these vibrations are directly proportional to the amplitude and frequency of input acoustic waves. The wave driver actuates the Chladni Plate that is employed with sand particles as a display medium [55]. A 24x24cm square metallic Chladni plate is employed in the studied case.

The system prototype functionality is tested with experimentation. A systematic visualization of image-extracted English alphabets is realized. The system functionality is described with the help of its Algorithmic State Machine (ASM) chart, shown on Figure 3.

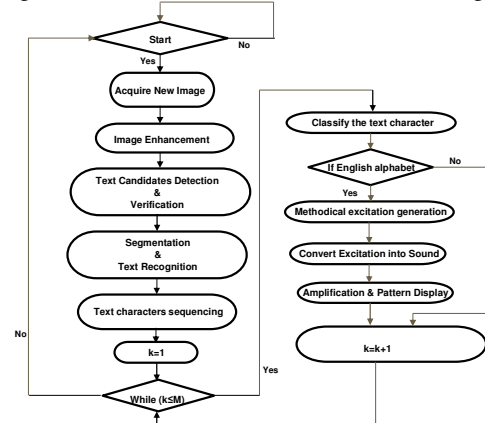


Figure 3: The proposed system ASM (Algorithmic State Machine) chart. Here, k is indexing the extracted text characters and M is the total number of text characters which are extracted from the incoming image.

Figure 3, shows that after initialization, the system acquires a new image. The acquired image is enhanced by employing an effective combination of CLAHE and DWT. The enhanced image is passed through the text candidates detection and verification stage. The text candidates are identified by employing the MSER features detector. Some examples of extracted MSERs from natural scenes are shown on Figure 4. Later on, the extracted MSERs are verified. The text regions verification process filtered out most of the non-text MSERs. It is based on the geometric properties and stroke width variations oriented filters. The output of this stage is shown on Figure 5. The remaining MSERs are considered as true text candidates and are passed to the segmentation and text recognition module. The segmentation is performed by employing the text candidates properties of pixel values, stroke widths and mutual distance. The segmented bounding boxes are recognized by employing the OCR. Results are shown of Figure 6. The next step is to create a sequence of extracted text characters so that they can be processed in a sequential fashion. The total number of extracted text characters, M from the incoming image, is calculated. The following operations are executed, in an iterative fashion, for each extracted character.

Each character is classified as an English or a Non-English alphabet. In the case, if an English alphabet is identified then a methodical excitation is generated by employing the pre-designed lookup table. The employed excitation table, designed during the calibration process, is summarized in Table 1. Later on, this excitation is converted into sound. This sound signal is augmented with the audio amplifier. The augmented excitation is applied to the wave driver. The wave driver actuates the employed square Chladni plate via vibrations. The plate vibrates and in this way a repeatable and specific pattern is created on its surface by using the sand particles. Afterwards, the iteration counter, k , is incremented and the next character is brought to the classification stage. On other hand, if the current character is classified as a Non-English alphabet, then k is incremented and the next character is brought to the classification stage. The process continues in an iterative fashion, unless k be-

comes greater than M. Once $k > M$, becomes true then the next image is acquired by the system and the process continues.

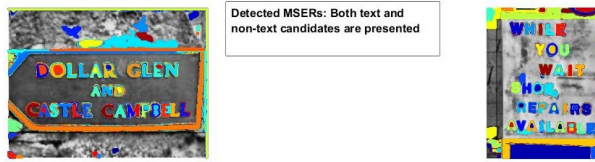


Figure 4: The detected text and non-text MSERs.

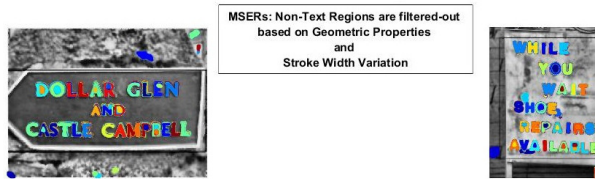


Figure 5: Filtered MSERs: Most of the non-text regions are filtered-out based on the geometrical properties and stroke width variations.

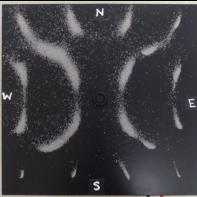
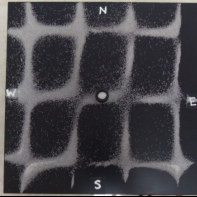
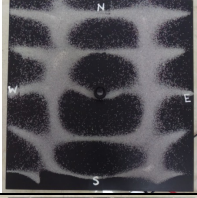
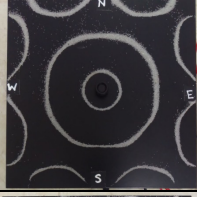
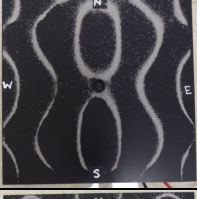
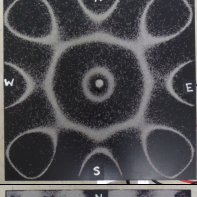
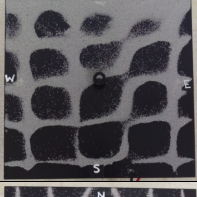



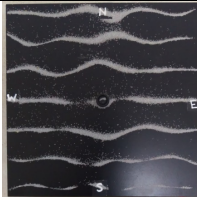
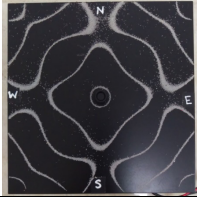
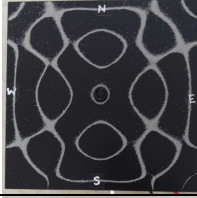
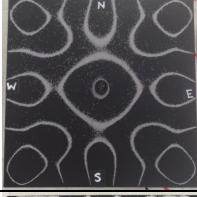
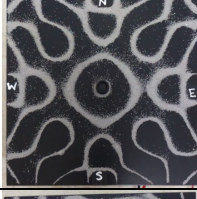


Figure 6: Examples of segmented and recognized texts from natural scenes.

Table 1: Summary of English alphabets, their respective excitation frequencies, amplitude and the generated Chladni patterns. The amplified excitations amplitude is kept constant to 5Volts.

English Alphabet	Chladni Pattern	Freq. (Hz)
A/a		78
B/b		90
C/c		107

D/d		175
E/e		225
F/f		380
G/g		440
H/h		455
I/i		490
J/j		520
K/k		625

L/l		718
M/m		827
N/n		935
O/o		990
P/p		1080
Q/q		1215
R/r		1390
S/s		1488

T/t		1560
U/u		1620
V/v		1945
W/w		2123
X/x		2590
Y/y		3039
Z/z		4273

3.2. Discussion

The proposed image enhancement, text detection and recognition chain performance is evaluated by employing the ICDAR 2013 competition data set [52]. The considered data set images count is increased up to 300 by adding further natural scenes. Each image is processed by employing the method depicted by Figure 3. The suggested approach achieves 80% precision in text localization. Certain sample outputs are shown on Figures 4 and 5. The method achieves

53% precision in text recognition. In this case, a word is considered correctly recognized if all its characters match the ground truth [52]. It confirms that the devised method achieves a better performance in terms of text regions identification as compared to the counter state-of-the-art solutions [40, 50]. In terms of end-to-end text recognition its performance is comparable with the counter state-of-the-art solutions [40, 50, 52]. The main limitations in the performance of suggested method are the scenes with very low resolution and highly noised text characters. Moreover, it also extracts wrong characters in the case of scenes with text like characteristics objects.

Table 1, shows that the devised system achieves a systematic visual perception of each English alphabet. The system is configurable as it generates a specific pattern for each alphabet on the same display medium. It is realized by smartly combining an effective image-extracted text to methodical sound conversion process along with the physical phenomenon of Cymatics. The obtained results have demonstrated the Chladni's law [6]. They present a directly proportional relationship between the excitation frequency and the number of both radial and linear nodes, generated on the square metallic Chladni plate. From Table 1, it is clear that the complexity of generated patterns is increased with the increase in acoustic excitation frequency and vice versa.

A certain number of excitation generators, actuators and display mediums can be employed in parallel and in a variety of topologies as a function of the intended application. Moreover, sound and tangibility can also be embedded in these mediums. In addition, the embedded implementation and miniaturisation of these systems can also be realized by employing the microcontrollers, Field Programmable Gate Arrays (FPGAs), low power wireless interfaces and Microelectromechanical systems (MEMS). This approach possesses a strong potential of opening new prospects in the development of smart fixed, portables and wearables that can be employed in various applications like visual arts, encryption, education, integration of impaired people, architecture, archaeology, etc. The embedded implementation, miniaturisation and investigation of proposed system principle usage for potential applications are rich axis to explore.

4. CONCLUSIONS

An inventive approach of achieving a Cymatics based visual perception of image-extracted text has been proposed. In this context, the image processing, text detection and recognition processes are smartly combined with a methodical acoustic excitation generation. A sequencing of the extracted text characters, from the incoming image, is made.

A system prototype is successfully realised and tested. The proposed text detection and recognition chain has achieved 80% precision in text regions identification. It confirms that the devised method attains a better performance in terms of text regions identification, from natural scenes, as compared to the counter state-of-the-art solutions [40, 50]. The method achieves 53% precision in end-to-end text recognition from the studied natural scenes. In terms of end-to-end text recognition the suggested system performance is comparable with the counter state-of-the-art solutions [40, 50, 52]. The main limitations in the suggested method performance are the scenes of very low resolutions and high noise. Moreover, it also extracts wrong characters in the case of scenes with text like characteristics objects. The proposed system performance can be improved by employing the computationally complex iterative image denoising and resolution enhancement algorithms [56]. It offers a trade-off between the system computational delay, power consumption and precision. An intelligent choice should be made as a function of the intended application.

The employment of signal driven Image digitization, enhancement, text detection and recognition will improve the system efficiency in terms of resources utilization and power consumption [14-16]. The study and integration of these features in the system is another axis to explore.

5. ACKNOWLEDGEMENT

Author is thankful to Eng. S. Toonsi and Eng. M. Allaf for their help during system prototyping and experimentation.

REFERENCES

- [1]. Oh, You Jin, and Sojin Kim. "Experimental Study of Cymatics." *International Journal of Engineering and Technology* 4.4 (2012): 434.
- [2]. Lewis, S. (2010). *Seeing Sound: Hans Jenny and the Cymatic Atlas* (Doctoral dissertation, University of Pittsburgh).
- [3]. Sykes, Lewis. "The augmented tonoscope." Manchester Institute for Research and Innovation in Art and Design (2011).
- [4]. Jenny, H. (2001). *Cymatics: a study of wave phenomena and vibration*. MACROmedia.
- [5]. Rhodenborgh, A. (2016). *Discovering an uncanny world: Cymatics software and the journey to the creation of knowledge within the field of contemporary Cymatics* (Master's thesis).
- [6]. Rossing, Thomas D. "Chladni's law for vibrating plates." *American Journal of Physics* 50.3 (1982): 271-274.
- [7]. Tuan, P. H., Wen, C. P., Chiang, P. Y., Yu, Y. T., Liang, H. C., Huang, K. F., & Chen, Y. F. (2015). Exploring the resonant vibration of thin plates: reconstruction of Chladni patterns and determination of resonant wave numbers. *The Journal of the Acoustical Society of America*, 137(4), 2113-2123.
- [8]. Krasil'nikov, V. A., & Krylov, V. V. (1984). *Introduction to physical acoustics*.
- [9]. Pettersson, Peter. "Cymatics: the science of the future." (2009).
- [10]. Echols, R. *Physics of Music: Making Waves in a Science Classroom*.
- [11]. Misseroni, D., Colquitt, D. J., Movchan, A. B., Movchan, N. V., & Jones, I. S. (2016). Cymatics for the cloaking of flexural vibrations in a structured plate. *Scientific reports*, 6, 23929.
- [12]. Wang, X. Q., & So, R. M. C. (2015). Timoshenko beam theory: A perspective based on the wave-mechanics approach. *Wave Motion*, 57, 64-87.
- [13]. Cooper, Thomas W. "FROM CONSCIOUSNESS TO TECHNOLOGY: CYMATICS, WAVE PERIODICITY, AND COMMUNICATION." *INTEGRATIVE EXPLORATIONS* 2.1 (1994): 52.
- [14]. Qaisar, S. M., Fesquet, L., & Renaudin, M. (2014). Adaptive rate filtering a computationally efficient signal processing approach. *Signal Processing*, 94, 620-630.
- [15]. Qaisar, S. M., Akbar, M., Beyrouthy, T., Al-Habib, W., & Asmatulah, M. (2016, June). An error measurement for resampled level crossing signal. In *Event-based Control, Communication, and Signal Processing (EBCCSP), 2016 Second International Conference on* (pp. 1-4). IEEE.
- [16]. Qaisar, S. M. (2015, June). An event driven technique for filtering computational complexity reduction. In *Event-based Control, Communication, and Signal Processing (EBCCSP), 2015 International Conference on* (pp. 1-4). IEEE.
- [17]. Liu, X. (2008, October). A camera phone based currency reader for the visually impaired. In *Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility* (pp. 305-306). ACM.
- [18]. Chen, X., Yang, J., Zhang, J., & Waibel, A. (2004). Automatic detection and recognition of signs from natural scenes. *IEEE Transactions on image processing*, 13(1), 87-99.
- [19]. Neumann, L., & Matas, J. (2013, December). Scene text localization and recognition with oriented stroke detection. In *Computer Vision (ICCV), 2013 IEEE International Conference on* (pp. 97-104). IEEE.
- [20]. Feild, J. (2014). Improving text recognition in images of natural scenes.
- [21]. Weinman, J. J., Butler, Z., Knoll, D., & Feild, J. (2014). Toward integrated scene text reading. *IEEE transactions on pattern analysis and machine intelligence*, 36(2), 375-387.
- [22]. Shahab, A., Shafait, F., & Dengel, A. (2011, September). ICDAR 2011 robust reading competition challenge 2: Reading text in scene images. In *Document Analysis and Recognition (ICDAR), 2011 International Conference on* (pp. 1491-1496). IEEE.

- [23] Lucas, S. M. (2005, August). ICDAR 2005 text locating competition results. In Document Analysis and Recognition, 2005. Proceedings. Eighth International Conference on (pp. 80-84). IEEE.
- [24] Lucas, S. M., Panaretos, A., Sosa, L., Tang, A., Wong, S., Young, R., ... & Miyao, H. (2005). ICDAR 2003 robust reading competitions: entries, results, and future directions. *International Journal of Document Analysis and Recognition (IJ DAR)*, 7(2-3), 105-122.
- [25] Ye, Q., Gao, W., Wang, W., & Zeng, W. (2003, December). A robust text detection algorithm in images and video frames. In Information, Communications and Signal Processing, 2003 and Fourth Pacific Rim Conference on Multimedia. Proceedings of the 2003 Joint Conference of the Fourth International Conference on (Vol. 2, pp. 802-806). IEEE.
- [26] Yin, X. C., Yin, X., Huang, K., & Hao, H. W. (2014). Robust text detection in natural scene images. *IEEE transactions on pattern analysis and machine intelligence*, 36(5), 970-983.
- [27] Weinman, J. J., Learned-Miller, E., & Hanson, A. R. (2009). Scene text recognition using similarity and a lexicon with sparse belief propagation. *IEEE Transactions on pattern analysis and machine intelligence*, 31(10), 1733-1746.
- [28] Chatzopoulos, D., Bermejo, C., Huang, Z., Butabayeva, A., Zheng, R., Golkarifard, M., & Hui, P. (2017, June). Hyperion: A Wearable Augmented Reality System for Text Extraction and Manipulation in the Air. In Proceedings of the 8th ACM on Multimedia Systems Conference (pp. 284-295). ACM.
- [29] Satyanarayana, P., Sujitha, K., Kiron, V. S. A., Reddy, P. A., & Ganesh, M. (2018). Assistance Vision for Blind People Using k-NN Algorithm and Raspberry Pi. In Proceedings of 2nd International Conference on Micro-Electronics, Electromagnetics and Telecommunications (pp. 113-122). Springer, Singapore.
- [30] Yi, C., Tian, Y., & Arditi, A. (2014). Portable camera-based assistive text and product label reading from hand-held objects for blind persons. *IEEE/ASME Transactions On Mechatronics*, 19(3), 808-817.
- [31] P. Sermanet, S. Chintala, and Y. LeCun, "Convolutional neural networks applied to house numbers digit classification," in Proc. IEEE Int. Conf. Pattern Recognit., 2012, vol. 4, pp. 3288-3291.
- [32] Z. He, J. Liu, H. Ma, and P. Li, "A new automatic extraction method of container identity codes," *IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 1, pp. 72-78, Mar. 2005.
- [33] N. Ezaki, K. Kiyota, B. T. Minh, M. Bulacu, and L. Schomaker, "Improved text-detection methods for a camera-based text reading system for blind persons," in Proc. IEEE Int. Conf. Document Anal. Recognit., 2005, pp. 257-261.
- [34] C. Yi and Y. Tian, "Localizing text in scene images by boundary clustering, stroke segmentation and string fragment classification," *IEEE Trans. Image Process.*, vol. 21, no. 9, pp. 4256-4268, Sep. 2012.
- [35] Q. Ye, Q. Huang, W. Gao and D. Zhao, "Fast and robust text detection in images and video frames," *Image Vis. Comput.*, vol. 23, pp. 565-576, 2005.
- [36] Y. Zhong, H. J. Zhang, and A. K. Jain, "Automatic caption localization in compressed video," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 4, pp. 385-392, Apr. 2000.
- [37] R. Lienhart and A. Wernicke, "Localizing and segmenting text in images and videos," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 4, pp. 256-268, Apr. 2002.
- [38] K. Jung, K. I. Kim, and A. K. Jain, "Text information extraction in images and video: A survey," *Pattern Recognit.*, vol. 37, pp. 977-997, 2004.
- [39] T. Wang, D. J. Wu, A. Coates, and A. Y. Ng, "End-to-end text recognition with convolution neural networks," in Proc. IEEE Int. Conf. Pattern Recognit., 2012, pp. 3304-3308.
- [40] Ye, Q., & Doermann, D. (2015). Text detection and recognition in imagery: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 37(7), 1480-1500.
- [41] P. Shivakumara, A. Dutta, U. Pal, and C. L. Tan, "A new method for handwritten scene text detection in video," in Proc. IEEE Int. Conf. Front. Handwritten Recognit., 2010, pp. 387-392.
- [42] D. Karatzas, S. Robles Mestre, J. Mas, and F. Nourbakhsh, "ICDAR 2011 robust reading competition: Challenge 1: Reading text in born-digital images," in Proc. IEEE Int. Conf. Doc. Anal. Recognit., 2011, pp. 1491-1496.
- [43] J. Liang, D. Doermann, and H. Li, "Camera-based analysis of text and documents: A survey," *Int. J. Doc. Anal. Recognit.*, vol. 7, pp. 84-104, 2005.
- [44] Chen, Huizhong, et al. "Robust Text Detection in Natural Images with Edge-Enhanced Maximally Stable Extremal Regions." *Image Processing (ICIP)*, 2011 18th IEEE International Conference on. IEEE, 2011.
- [45] Neumann, Lukas, and Jiri Matas. "Real-time scene text localization and recognition." *Computer Vision and Pattern Recognition (CVPR)*, 2012 IEEE Conference on. IEEE, 2012.
- [46] Li, Yao, and Huchuan Lu. "Scene text detection via stroke width." *Pattern Recognition (ICPR)*, 2012 21st International Conference on. IEEE, 2012.
- [47] Gonzalez, Alvaro, et al. "Text location in complex images." *Pattern Recognition (ICPR)*, 2012 21st International Conference on. IEEE, 2012.
- [48] Baran, R., Partila, P., & Wilk, R. (2018, January). Automated Text Detection and Character Recognition in Natural Scenes Based on Local Image Features and Contour Processing Techniques. In International Conference on Intelligent Human Systems Integration (pp. 42-48). Springer, Cham.
- [49] Wang, R., Sang, N., & Gao, C. (2015). Text detection approach based on confidence map and context information. *Neurocomputing*, 157, 153-165.
- [50] Neumann, L., & Matas, J. (2015, August). Efficient scene text localization and recognition with local character refinement. In Document Analysis and Recognition (ICDAR), 2015 13th International Conference on (pp. 746-750). IEEE.
- [51] Lidong, H., Wei, Z., Jun, W., & Zebin, S. (2015). Combination of contrast limited adaptive histogram equalisation and discrete wavelet transform for image enhancement. *IET Image Processing*, 9(10), 908-915.
- [52] Lidong, D. Karatzas, F. Shafait, S. Uchida, M. Iwamura, S. R. Mestre, I. Mas, D. F. Mota, I. A. Almazan, L. P. de las Heras et al., "ICDAR 2013 robust reading competition," in ICDAR 2013. IEEE, 2013, pp. 1484-1493.
- [53] Lei, B., Xu, G., Feng, M., van der Heijden, F., Zou, Y., de Ridder, D., & Tax, D. M. (2017). Classification, parameter estimation and state estimation: an engineering approach using MATLAB. John Wiley & Sons.
- [54]. "SF-9324 Wave Driver" Pasco scientific, 2015.
- [55]. "Chladni Plates Kit" Pasco scientific, 2015.
- [56] Cruz, C., Mehta, R., Katkovnik, V., & Egiazarian, K. O. (2018). Single image super-resolution based on Wiener filter in similarity domain. *IEEE Transactions on Image Processing*, 27(3), 1376-1389.
- [57]. Cymatics (2011, 8 December). Retrieved January 5m 2012, from Wikipedia website: <http://en.wikipedia.org/wiki/Cymatics>.
- [58]. Silva, F. (2010). *Secrets in the Fields: The Science and Mysticism of Crop Circles*. Charlottesville, VA: Hampton Roads.
- [59]. Silva, F. (2010). *Secrets in the Fields: The Science and Mysticism of Crop Circles*. Charlottesville, VA: Hampton Roads.
- [60]. Cox, T. J. (n.d.). *Acoustic Diffusers: the Good, the Bad, and the Ugly*. Proceedings of the Institute of Acoustics. Retrieved from: http://www.rpginc.com/docs/5CTechnology/5CWhite/20Papers/5CAcoustic/20Diffusers_The%20Good,%20The%20Bad%20and%20The%20Ugly.pdf
- [61]. Cox, T. J. & D'Antonio, P. (2009). *Acoustic Absorbers and Diffusers: Theory, Design and Application*. Florence KY: Taylor & Francis.
- [62]. Kuttruff, H. (2009). *Room Acoustics*. Abingdon: Spon.