Illumination Normalization and Blackboard Segmentation for Presenter Tracking Systems: A Literature Review

Charles Fitzhenry
Honours, Department of Computer Science
University of Cape Town
Rondebosch, 7701, South Africa
FTZCHA002@myuct.ac.za

ABSTRACT
This literature review forms part of a larger project for developing a lecture tracking system for capturing lectures. The project is divided into three sections, namely, pre-processing, lecture tracking and virtual cinematography. This literature review, is concerned with the pre-processing stage that deals with gain correction and segmentation of the blackboard from the video stream. Several research papers are surveyed and algorithms that would potentially be feasible to implement in the final project, are highlighted.

CCS Concepts
• **Computing methodologies → Computer vision problems**

• **Computing methodologies → Video segmentation**

• **Computing methodologies → Tracking**

Keywords
Presenter Tracking; 4K Video; Light Normalization; Blackboard Tracking; Blackboard Segmentation

1. INTRODUCTION
The concept of recording lectures in classrooms has been implemented as an accepted practice at universities and institutions worldwide, and has gained a lot of traction in the last few years [1, 2]. It has been especially implemented widely across Massive Open Online Courses (MOOCs), making distance learning very prevalent today [3]. There are many different terms used to refer to the concept of lecture capturing, but the general definition interprets it as the use of any technology to record what a lecturer does and says during a class, and then making this digitally available to students [4]. Software and hardware is used to capture the lecture as well as any other presentation medium such as presentation slides used during the lecture. Since the lecturer does not have to wear specific equipment to be detected by the camera there is a need to utilise methods from the field of computer vision, most notably tracking by the video source only [5].

There have been several studies that consider the advantages and disadvantages of recording lectures and making them available to students. The majority of studies have however identified more advantages than disadvantages. For example, Al Nashash and Gunn [6] observed that students are able to learn at their own pace and review lectures as many times as needed, allowing students to engage more with the material. Vassar [4] indicates that students can also overcome the challenge of taking notes while trying to concentrate on the content being explained and assists students when studying for assessments. Other than the benefits experienced by students, institutions can also benefit by making courses available at lower prices, since students do not need to attend a live lecture.

At the University of Cape Town (UCT), the Centre for Innovation in Learning and Teaching (CILT) is responsible for learning and teaching challenges at UCT and, among other services, have implemented a lecture capturing system that is used across the campus. CILT is currently using Pan-Tilt-Zoom (PTZ) cameras in the lecture venues along with an overview camera that is driven by a Raspberry Pi to control the PTZ cameras movements. The LectureSight [5] system is used to track the lecturer in real time. Because the tracking is done in real-time, the virtual cinematographer (VC) heuristics often produce undesirable results due to the non-deterministic behaviours produced by different lecturers. To overcome these shortcomings, CILT has endeavoured on a new project that uses Ultra High Definition (UHD) 4K cameras that have a resolution of 3840x2160 pixels (P) or 8 Mega Pixels (MP). These cameras capture the entire lecture area; however, their video streams are very large in size and CILT currently does not have Virtual Panning (VP) software that produces results that closely resemble that of a real cinematographer. Their goal for this project is to develop a software system that would take the 4K video stream in, track the lecturer and output only the segment containing the lecturer and the blackboard in a 720P resolution.

This project forms part of the Honours in Computer Science Module at UCT. The project has been divided into three sections namely, pre-processing of the raw video, lecture tracking and a VC module. This paper is concerned with the first section, pre-processing. The pre-processing section entails illumination normalization for varying lighting conditions in the video that is caused by different situations, such as the use of a data projector. It also will extract the active blackboard being used by the lecturer to export as a separate video stream with a much lower frame rate.

This paper examines current algorithms and techniques available to achieve the goal of this section of the project, that is illumination normalization and image segmentation.

2. ILLUMINATION CORRECTION
Illumination correction (also known as gain correction) is an umbrella term used to refer to a set of techniques that aim to normalize or compensate for variations of lighting conditions that occur in the scene [7, 8]. In the context of object recognition and segmentation, non-uniform illumination poses a complex challenge to achieving positive results in facial recognition and segmentation processes [8, 9]. Ruiz-del-Solar *et al*. [10] and Zou *et al*. [11] suggest separating the illumination algorithms into three categories based on their objective. These are namely illumination normalization, illumination-invariant feature extraction and illumination modelling. Illumination normalization attempts to normalize the lighting in the video to enable feature extraction and the output video shows these lighting enhancements. Illumination-invariant looks at techniques of extracting features regardless of

light condition and the output video is unchanged. Illumination modelling creates a set of features from different images that have varying illumination [8]. Because the video output from this project will be viewed by students, it is sensible to use the first category of algorithms, illumination normalization. This will benefit the second stage (Tracking) in the larger project and will also provide a video that is more appealing to users. The following sections will explore past approaches and techniques to illumination normalization that are considered both efficient and accurate.

Santamaria and Palacios [12] suggested that the algorithms for illumination normalization be divided into two categories, namely global normalization methods and local normalization methods. Global normalization methods are used when the image must retain the modifications in the final output. Local normalizations are only applied to a parts of the image which are poorly lit and thus have the disadvantage that the output is not necessarily realistic; instead the normalization is performed to enhance a section of the image for recognition of features. An example use case is facial recognition. Because this project requires the image correction for both image recognition as well as for the final output, that needs to be visually appealing to the users, only the local normalization techniques are discussed in the following sections.

2.1 Histogram Equalization

Histogram Equalization (HE) is a technique used to adjust the intensities of images to enhance contrast. The HE technique has been referenced by Hummel [13] as early as 1974 and although the original HE algorithm was only able to correct the lighting of grayscale images, the algorithm has been extended and improved extensively since then. In 1992 Trahanias and Venetsanopoulos [14] successfully extended the HE algorithm to colour images, in a technique called 3-D Histogram Equalization. This proposed method could not be measured quantitatively because the evaluation criteria for the performance of such a method depends on the visual appearance of the final image produced. However, the method successfully enhanced the images without producing any colour artefacts as opposed to other similar techniques. Although this method has been shown to produce good results, one significant drawback is the computational complexity, that is $O(n^3)$ where n is the discreet number of grey levels (or bins) in the histogram [14]. Naik and Murthy [15] identified a problem with this approach that shifts the three primary colour components Red, Green, and Blue (RGB) unequally, resulting in a change of the pixel's hue. They proposed an improvement that extended to colour images and preserved the hue during correction, and eliminated the gamut problem, where pixels go beyond the displayable bounds after the process. The latest contributions and research to this technique are from Wong *et al.* [16] and Wang *et al.* [17]. Wong *et al.* indicates the shortcomings in all the current approaches, that is, they often produce undesirable results caused by over-enhancement and when limited to only intensity-enhancements, the colour and saturation aspects of the image do not get enhanced. To solve this, a pipelined approach that incorporates colour stretching, HE, magnitude compression and saturation maximization, is proposed. The experiments show that this algorithm solves all the previous challenges associated with HE without affecting performance [16]. Wang *et al.* take a slightly different approach to achieve the same results. In previous HE approaches, histograms were constructed and a mapping was determined, upon which the pixels of the source image were then adjusted. This new approach instead aims to adjust the image pixel values directly and the resultant histogram can be redistributed [17].

2.2 Histogram Specification

Histogram Specification (also known as Histogram Matching or Histogram Fitting) (HS/HM/HF) is a generalization of HE [12] and was initially proposed by Woods and Gonzalez [18]. HS transforms the histogram of a given image to a specified shape, that is usually the histogram of a well illuminated image [12, 19]. The algorithm works by computing the histograms of two images as well as their Cumulative Distribution Function (CDF) [20]. Rolland *et al.* [20] proposed a new approach referred to as the Optimal Cumulative Distribution Function Matching (OCM) algorithm. This approach reduces computational time by utilising a lookup table (LUC) that is built from the two functions needing to be matched. The algorithm improved the time complexity from $O(NK)$ to $O(N \log_2 N)$, where N is the number of pixels in the image and K is the number of grey values (or bins) in the histogram. There is, however, a drawback associated with the CDF algorithm. The discrete case fails because the functions are not always guaranteed to be exactly invertible and the algorithm cannot produce the desired histogram perfectly. To achieve the desired shape, Coltuc *et al.* [21] proposed a solution that involves ordering the image pixels based on the intensity values of their local means [19]. While Wan *et al.* [22] acknowledges that the OCM method is by far the most successful exact pixel ordering algorithm, they identify a significant flaw in this method. The algorithm does not take into consideration that for the general case, most of the image data falls into the high-frequency intensity range and not into the low-frequency range and because the OCM algorithm only uses the local mean to correct intensity, it works well for smooth regions, however, it does not pick up edge regions well. This is solved by implementing a Wavelet-Based (WB) method, that considers both the local mean by using a low-pass filter and the local edge information by using a high-pass filter [19, 21]. The experiments show that the WB method improves the image correction when compared to the OCM method. It also shows an improvement in computational time and when processing a 2048x2048 image, the WB method took 26.88 seconds as opposed to the OCM method that took 36.64 seconds. On the same size image, the pixel ordering failure rate was greater than 70% using the OCM method as opposed to 0.3% when using the WB method. There is a downside to this WB method; it amplifies noise as indicated by Nikolova *et al.* [23], who also proposed a Variational Approach (VA) to solve these problems. This VA was then further generalised to work on colour images by Nikolova [24], who proposed a hue and range preserving colour adjustment. Currently the latest contribution to this research is from Pierre *et al.* [25], who firstly proposes a new method for enhancing greyscale images by performing a difference calculation of the average local contrast values between the original and enhanced images. They then generalise this approach to colour images.

2.3 Gamma Intensity Correction

Shan *et al.* [26] have presented a Gamma Intensity Correction (GIC) algorithm that builds on the idea of the HE and HS techniques and attempts to normalize an image to a given intensity level derived from a sample image. This is done by calculating the gamma value, for each pixel over the image in the xy space, using the following formula:

$$I'_{xy} = G(I_{xy}, \gamma^*) \tag{1}$$

Where γ^* is calculated by the following algorithm that aims to minimise difference between original and the image provided I_0:

$$\gamma^* = \min_{\gamma}\left(\sum_{x,y}[G(I_{xy}, \gamma) - I_0(x, y)]^2\right) \tag{2}$$

Where I_{xy} is the images' grey intensity value at the pixel position (x, y) and:

$$G(I_{xy}, \gamma) = c \cdot I_{xy}^{\frac{1}{\gamma}} \quad (3)$$

Is the transform function for gamma, γ is the coefficient of gamma and c is a parameter for stretching the grey values over the histogram.

Using faces from the Yale and Harvard database found that the proposed GIC method interestingly shows that the HE method recognises faces better than GIC by 1%. However, another approach known as Quotient Illumination Relighting (QIR) was proposed and achieved a 60.1% recognition rate under the same conditions. The downside of this approach is that the lighting modes of the image is known.

3. BLACKBOARD SEGMENTATION

Image Segmentation (IS) is the process of dividing an image into multiple semantically related segments and is the initial stage of image processing and analysis [27]. Blackboard Segmentation is a specialization of IS and is used to segment the blackboard(s) from a video stream. Because the project is going to consider creating a third video stream of just the blackboard(s) that are used, it would require the principles and techniques of IS. The following section explores related work in this field.

Onishi *et al.* [28] proposed a method for segmenting regions of text from a blackboard by using the recorded lecture videos. They identify that it is important to capture a shot of the blackboard that is legible to students and they also note how conventional approaches show either the lecturer or the whole blackboard only, making it difficult to see what is written on the board. Writing on the blackboard is detected by using edge detection (Sobel) functions. The algorithm incrementally pans the section as the lecturer adds new text to the board. The experimental results show that the technique achieves accurate segmentation results 91.5% of the time. Liu *et al.* [29] suggest an improvement that does not assume a static board, but instead can accommodate moveable boards under varying lighting conditions. They also consider that the board colour can change due to the accumulation of chalk dust. Their method can also be applied to whiteboards. They model the colour distribution of the board and this solves the challenges faced by techniques that assume a uniform background colour of the board. Wallick *et al.* [30, 31] has independently proposed similar techniques to [29] and note that the problem of segmenting blackboard images does not require the creation of new fundamental computer vision techniques, but instead using already existing techniques in a way that would achieve desirable results. There is, however, currently no research found that directly explores the possibility of extracting the entire actively used blackboard and all methods discussed all extract only a subsection of the board where the lecturer is currently writing.

4. DISCUSSION

The literature reviewed for each of the sections required by the final project have shown that there are a wide range of available algorithms to solve both challenges, namely illumination normalization and blackboard segmentation. However, many of the research methods used very different methods for testing their performance and how well they solve the problem. This renders it challenging to directly compare the various techniques. However, each new contribution to existing techniques have concluded that they have improved on the previous techniques and this gives a reason to consider using the latest research. In illumination correction, Wong *et al.* [16] and Wang *et al.* [17] both provide very creditable work in the field of histogram normalization and may be explored further for use in this project.

For blackboard segmentation, there is not as much research directly relating to the topic. However, Wallick *et al.* [30] suggested that this problem of blackboard segmentation is a specific case of the larger image segmentation problem, and that existing computer vision approaches can be used for segmenting the blackboard.

5. CONCLUSIONS

This literature review aimed to gain a deeper understanding of existing techniques that could be used to solve the tasks in the larger project. Because the algorithms and techniques identified in this paper did not all have comparable metrics, a suggestion for future work in this project would be to perform tests by using these algorithms for illumination correction on 4K videos and evaluating the feasibility in terms of time complexity and how well the lighting is corrected. A set of tests could be constructed using various approaches from those covered in this paper, and the best technique can then be identified. For the blackboard segmentation, the current techniques do not extract the entire board, but instead only a subsection on which the lecturer is currently writing. As future work, these algorithms can be extended and enhanced to achieve the goals for the blackboard segmentation phase of the project. At the end of these tests, a final evaluation will need to be done, to select the best techniques for both sections that would yield the best results for the overall project.

6. REFERENCES

[1] González-Agulla, E., Alba-Castro, J. L., Canto, H. and Goyanes, V. Galitracker: Real-time lecturer-tracking for lecture capturing. In Anonymous *Multimedia (ISM), 2013 IEEE International Symposium.* (Anaheim, CA, December 9-11,2013). IEEE, 2013, 462-467.

[2] Demetriadis, S. and Pombortsis, A. E-lectures for flexible learning: A study on their learning efficiency. Educational Technology & Society, 10, 2 (April 2007), 147-157.

[3] Kay, J., Reimann, P., Diebold, E. and Kummerfeld, B. MOOCs: So Many Learners, So Much Potential... IEEE Intelligent Systems, 28, 3 (May-June 2013), 70-77. DOI=http://doi.ieeecomputersociety.org/10.1109/MIS.2013.66.

[4] Vassar, P., Havice, P. A., Havice, W. L. and Brookover, R. The Impact of Lecture Capture Presentations in a Distributed Learning Environment in Parks, Recreation, and Tourism Management. Schole, 30, 1 (2015).

[5] Wulff, B. and Fecke, A. Lecturesight-an open source system for automatic camera control in lecture recordings. In Anonymous *Multimedia (ISM), 2012 IEEE International Symposium on.* (Irvine, CA, December 10-11, 2012). IEEE, 2012, 461-466.

[6] Al-Nashash, H. and Gunn, C. Lecture Capture in Engineering Classes: Bridging Gaps and Enhancing Learning. Educational Technology & Society, 16, 1 (January 2013), 69-78.

[7] Brown, M. and Lowe, D. G. Automatic panoramic image stitching using invariant features. International journal of computer vision, 74, 1 (August 2007), 59-73.

[8] Ochoa-Villegas, M. A., Nolazco-Flores, J. A., Barron-Cano, O. and Kakadiaris, I. A. Addressing the illumination challenge in two-dimensional face recognition: a survey. IET Computer Vision, 9, 6 (December 2015), 978-992.

[9] Mustafa, W. A., Yazid, H. and Bin Yaacob, S. Illumination normalization of non-uniform images based on double mean filtering. In Anonymous *Control System, Computing and Engineering (ICCSCE), 2014 IEEE International Conference on.* (Batu Ferringhi, November 28-30, 2014). IEEE, 2014, 366-371.

[10] Ruiz-del-Solar, J. and Quinteros, J. Illumination compensation and normalization in eigenspace-based face recognition: A comparative study of different pre-processing approaches. Pattern Recog. Lett., 29, 14 (July 2008), 1966-1979. DOI=http://dx.doi.org/10.1016/j.patrec.2008.06.015.

[11] Zou, X., Kittler, J. and Messer, K. Illumination invariant face recognition: A survey. In Anonymous *Biometrics: Theory, Applications, and Systems, 2007. BTAS 2007. First IEEE International Conference on.* (Crystal City, VA, September 27-29, 2007). IEEE, 2007, 1-8.

[12] Santamaría, M. V. and Palacios, R. P. Comparison of illumination normalization methods for face recognition. (2005).

[13] Hummel, R. A. Histogram modification techniques. Computer Graphics and Image Processing, 4, 3 (September 1975 1975), 209-224. DOI=http://dx.doi.org.ezproxy.uct.ac.za/10.1016/0146-664X(75)90009-X.

[14] Trahanias, P. and Venetsanopoulos, A. Color image enhancement through 3-D histogram equalization. In Anonymous *Pattern Recognition, 1992. Vol. III. Conference C: Image, Speech and Signal Analysis, Proceedings., 11th IAPR International Conference on.* (The Hague, August 30-September 3, 1992). IEEE, 1992, 545-548.

[15] Naik, S. K. and Murthy, C. Hue-preserving color image enhancement without gamut problem. Image Processing, IEEE Transactions on, 12, 12 (December 2003), 1591-1598. DOI=10.1109/TIP.2003.819231.

[16] Wong, C. Y., Jiang, G., Rahman, M. A., Liu, S., Lin, S. C., Kwok, N., Shi, H., Yu, Y. and Wu, T. Histogram Equalization and Optimal Profile Compression based Approach for Colour Image Enhancement. Journal of Visual Communication and Image Representation, (April 2016). DOI=http://www.sciencedirect.com/science/article/pii/S1047320316300529.

[17] Wang, W., Chen, C. and Ng, M. K. An image pixel based variational model for histogram equalization. Journal of Visual Communication and Image Representation, 34 (January 2016), 118-134. DOI=http://dx.doi.org/10.1016/j.jvcir.2015.10.019.

[18] Woods, R. and Gonzalez, R. Real-time digital image enhancement. In Anonymous *IEEE Proceedings.* (May, 1981). 1981, 643-654.

[19] Liu, H., Yang, M., Gao, Y. and Cao, L. Fast Local Histogram Specification. Circuits and Systems for Video Technology, IEEE Transactions on, 24, 11 (2014), 1833-1843.

[20] Rolland, J. P., Vo, V., Bloss, B. and Abbey, C. K. Fast algorithms for histogram matching: Application to texture synthesis. Journal of Electronic Imaging, 9, 1 (2000), 39-45.

[21] Coltuc, D., Bolon, P. and Chassery, J. Exact histogram specification. Image Processing, IEEE Transactions on, 15, 5 (2006), 1143-1152.

[22] Wan, Y. and Shi, D. Joint exact histogram specification and image enhancement through the wavelet transform. IEEE Trans. Image Process., 16, 9 (2007), 2245-2250.

[23] Nikolova, M., Wen, Y. and Chan, R. Exact histogram specification for digital images using a variational approach. Journal of Mathematical Imaging and Vision, 46, 3 (2013), 309-325.

[24] Nikolova, M. *A fast algorithm for exact histogram specification. Simple extension to colour images.* Springer, 2013.

[25] Pierre, F., Aujol, J., Bugeau, A., Steidl, G. and Ta, V. Variational Contrast Enhancement of RGB Images. (2015).

[26] Shan, S., Gao, W., Cao, B. and Zhao, D. Illumination normalization for robust face recognition against varying lighting conditions. In Anonymous *Analysis and Modelling of Faces and Gestures, 2003. AMFG 2003. IEEE International Workshop on.* (October 17, 2003). IEEE, 2003, 157-164.

[27] Cheng, H. D., Jiang, X. H., Sun, Y. and Wang, J. Color image segmentation: advances and prospects. Pattern Recognition, 34, 12 (12 2001), 2259-2281. DOI=http://dx.doi.org/10.1016/S0031-3203(00)00149-7.

[28] Onishi, M., Izumi, M. and Fukunaga, K. Blackboard segmentation using video image of lecture and its applications. In Anonymous *Pattern Recognition, 2000. Proceedings. 15th International Conference on.* (Barcelona, September 30, 2000). IEEE, 2000, 615-618.

[29] Tiecheng, L. and Kender, J. R. Spatial-temporal semantic grouping of instructional video content. In Anonymous *Image and Video Retrieval.* Springer, 2003, 362-372.

[30] Wallick, M. N., Gleicher, M. L. and Heck, R. M. Obtaining a Mid-Level Representation of Handwriting without Semantic Understanding. (2003).

[31] Wallick, M. N., Heck, R. M. and Gleicher, M. L. Marker and chalkboard regions. In *Proceedings of Mirage.* 2005, 223-228.