

Anonymous Lecture Theatre Occupancy Measurement using Low Resolution Thermal Camera

Kieran Hitchcock

Department of Computer Science
University of Canterbury
Christchurch, New Zealand
khi48@uclive.ac.nz

Richard Green

Department of Computer Science
University of Canterbury
Christchurch, New Zealand
richard.green@canterbury.ac.nz

Abstract— This paper proposes a method to accurately count the occupancy of a lecture theatre. This method aims to create the concept of a system that can be remotely set up at the front of a lecture theatre using a thermal camera and Raspberry Pi. Lecture theatre occupancy is useful to know in order to track how attendance varies during the year, semester, and even during the day. By tracking occupancy, lectures that do not maximize the capacity of the lecture theatre can be rescheduled to smaller theaters in order to free up the larger theatres. Furthermore, by tracking room occupancy, other smart systems can be implemented such as controlling heating, lighting or systems to help in evacuations. Prior studies have looked at using other detection methods such as colour cameras, passive infrared (PIR), and thermal cameras. This method used a low-resolution thermal camera (160x120 pixels) to detect people. Thermal cameras used body heat to detect people and do not require distinguishing features typically used by color cameras. Therefore, thermal cameras can preserve the anonymity of the lecture attendees. Using Green's theorem, the area containing thermal blobs was identified. Each blob's position on the y-axis, shape, and area was then calculated and used to determine whether the blob was representative of an individual person. As each lecture theatre has different dimensions, parameters within the proposed method needed to be adjusted accordingly. The proposed method achieved a people counting accuracy of 100% and 96% from two different lecture theaters. A common problem with room occupancy detection methods is the inability to work in a range of different locations. This method had a similar problem but can be resolved with some slight modifications. This method will likely require far less data and can be more easily retrained to work in other lecture theatres compared to other methods such as neural networks. For future research, this method could benefit from further research into working in a range of different theatres to try and overcome this problem. Automatic scaling of these parameters could be achieved by measuring the size of the largest blob in the frame. Additionally, future development could be done to implement this method to run off of a Raspberry Pi and collect information.

Keywords— *People Detection, Thermal Camera, Low Resolution, Lecture Theatre, Infrared, Room Occupancy*

I. INTRODUCTION

Lecture theatre occupancy can be used to determine how well the theatre is being utilized. This can allow the theatres to be optimized such by moving poorly attended classes to smaller lecture theatres to free up the larger and more useful theatres. Furthermore, it can be used to determine whether the room is occupied and can be used to adjust lighting or heating or help in evacuations.

Typically, colour cameras have been used to predict the number of people present in crowds. However, colour cameras require high-resolution images for accurate

prediction. This means distinguishing features, and therefore individual people, can be tracked. As this could potentially be used maliciously, it is unethical to set up a colour camera in every lecture theatre.

Thermal cameras detect body heat in the form of infrared radiation. People show up as large blob-like features and blob detection can be used to identify individuals. Thermal cameras work independently of visible light conditions, unlike colour cameras, making them a more robust detection method. Additionally, thermal camera detection method would not have to undergo complicated image processing such as facial detection and could be less intensive.

Very low-resolution thermal cameras, such as a Passive Infrared (PIR) sensor, are commonly used to detect moving people and are typically used over doorways to count people entering and leaving [1], [2]. Multiple PIR sensors would be required in order to track a single lecture theatre's traffic and they would not provide any positional information of the occupants in the room. It would be more efficient to use a single, higher resolution thermal camera that could capture an entire lecture theatre and count individuals.

A range of studies have looked at using high resolution thermal cameras to count people in crowded areas [3], [4]. As thermal cameras are very expensive compared to colour cameras, using the lowest resolution thermal camera possible would make the concept more financially viable. Additionally, the lower resolution also reduces the possibility of any unique/distinguishing features of people being identified to preserve anonymity.

This paper proposes a method to estimate the number of people present in a theatre using a single Lepton 3 thermal camera. Using Green's theorem [5], the area containing thermal blobs was identified. Each blobs position on the y-axis, shape, and area was then calculated and used to determine whether the blob was representative of an individual person. This paper looked at creating a proof of concept that could be later developed to run remotely off a Raspberry Pi in the front of a lecture theatre.

This paper firstly outlines some of the previous research into this field and the limitations of the methods used. Secondly, this paper outlines the proposed method used to measure the occupancy of the lecture theatres. Finally, this paper compares how well this method worked compared to similar attempts and indicates where future research could improve the method to achieve better results.

II. BACKGROUND RESEARCH

For people detection with high-resolution images, identifying features such as faces, or arms can be used to separate individual people. As the Lepton 3 has a relatively low-resolution compared to the size of the lecture theatre, people appeared as simple blobs. Therefore, blob detection methods were researched to try and count individual people.

A. Blob Detection

Thermal cameras show people as brightly coloured spots as they are significantly warmer than the surroundings, as shown in Figure 1.

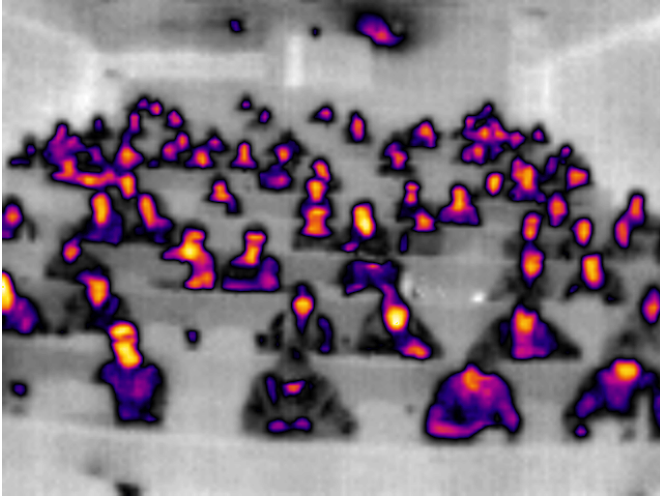


Fig. 1. Thermal image of the University of Canterbury's A2 lecture theatre

Separating the relevant foreground containing people from the background is trivial with blob detection methods. M. Berger and A. Armitage [6] used Laplacian of Gaussian to detect individual blobs while M. Arif [7] used Green's Theorem. Both blob detection methods found the edges of the blobs and created an enclosing contour. While there was little difference in function between the two methods, Green's Theorem was more compatible with the software used in the proposed method.

B. Blob Separation

Due to the low resolution and front facing view of the thermal camera, people sitting behind one another in a lecture theatre tend to merge together into one larger blob. A number of methods were researched that work to separate connected blobs. Furthermore, the blob sizes decreased higher in the lecture theatre as less thermal radiation was detected from people sitting further away.

1) Morphology

M. Berger and A. Armitage [6] used morphology operations to separate out merged blobs into individual blobs representative of people. The original, merged blobs were compared to a pre-set maximum and minimum areas that each person was expected to form. If the blobs area was larger than A_{max} , then the blob was eroded until it was less than the A_{max} . If the blob area was less than A_{max} , but larger than A_{min} , then it was dilated back to A_{max} . If the area was below A_{min} , then it was erased from the original image and believed to be

noise. The flow diagram representing his method is outlined in Figure 2.

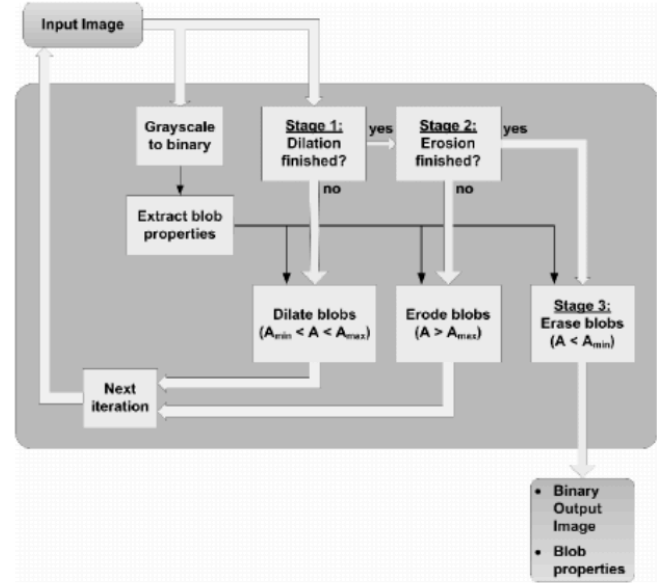


Fig. 2. Morphology area adjustment used to separate merged blobs

This method was easy to implement and had easily adjustable parameters. By adjusting the values of A_{max} and A_{min} , a variety of constant blob sizes could be detected. However, this method only worked when every blob was similar sized and did not vary to a large degree. This method also used a top down camera view to ensure people formed similar blobs as well as minimizing the amount of blob merging that occurred. Additionally, this method only worked when the blobs were only slightly overlapping. When the blobs overlapped, even to a relatively minor degree, then blobs failed to separate into their individual components. This method had the advantage of not having to operating in real time. A single image can be taken, and the system could run as long as required to determine the occupancy.

2) Background Subtraction

Other studies have looked at using background subtraction and double subtraction with thermal cameras to count moving crowds [8]. The difference between two consecutive frames showed where objects have moved. In a crowd, this is a good indication of individual people. By counting the number of blobs in the frame differences, a good estimation of people can be achieved.

The largest limitation of this method is that it relies upon people moving. Once the students are seated, they are mostly stationary, and this method would not give accurate results. People could be tracked as they entered the room and took their seats, but it would require keeping track of people already seated and people entering and exiting. Furthermore, it would be required to operate in real time to capture and process the images to calculate the frame differences. This could be difficult on a system with limited processing power, such as a Raspberry Pi.

3) Neural Networks

Neural networks were commonly used for blob separation as shown in multiple papers. A.-W. Negied [9] used a neural

network to identify people in low resolution images the thermal cameras, while Y. Xue [11] used a neural network to detect merged cell nuclei. While different applications, the same root problem occurs of trying to separate blobs. There are also a number of papers using neural networks on images from both colour and thermal cameras [12].

The major limitation of these projects was the inability to working in changing environments. The studies had all had relatively constant blob sizes and did not use images with drastically different dimensions. This does not mean it is impossible to train it to detect different sized blob, the neural network would have to be trained to work in different locations and change its scaling parameters accordingly. This would require a large amount of training and reinforcement. Y. Xue [11] trained their neural network with over 500 images to detect similar sized blobs consistently. It is likely that a lot more training data would be needed in order to ensure that a neural network reliably worked when used in a range of different lecture theatres.

4) Genetic Algorithm

I. Amin [10] used images from a coloured camera on a moving crowd. The people were extracted from the crowd by using background subtraction and were identified using blob detection. By repetitive thresholding and blob area checking, the blobs were separated and then counted. A genetic algorithm was employed to try and optimise the threshold and critical area sizes. This method is similar to that employed by the morphology blob separation method [6].

This method worked well and had an effective way to find the required values to optimise the blob detection method. This method also only considered blobs of equal sizes and also failed to account for stationary people.

5) Watershed

Y. Xue [11] and J. Davis [13] used a watershed algorithm to separate nuclei in clustered cells. This method required removing the background before heavily dilating the image. Once dilated, a distance transform was taken to find the centre of each blob, as shown in Figure 3. By thresholding the distance transform, the centre of the blobs could be identified.

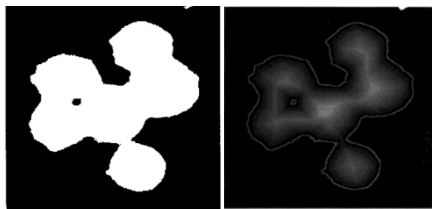


Fig. 3. Watershed algorithm, dilation (left) and distance transform (right)

Initial testing of this method showed that it mainly worked with blobs that were all of similar sizes. This was a major limitation in lecture theatres where people close and far away from the camera merge together into one blob. This method tended to remove the smaller person completely.

6) Ellipse Fitting

X. Bai [14] took a different approach to the separation of cell. Ellipse fitting was used in order to determine the boundaries of each cell in a large, merged cell. In this method, the large, merged blob is separated out into a of ellipses that could form the larger blob, as shown in Figure 4.

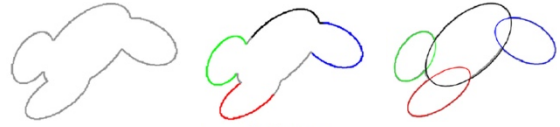


Fig. 4. Diagram example of ellipse fitting used to separate merged cells

While this method could detect different sized blobs, it would not detect irregular sized shapes. This meant it would struggle to identify full bodies and partially occluded people. A majority of the thermal blobs were not of a regular shape in lecture theatres, and therefore would not conform to an elliptical shape.

III. PROPOSED METHOD

The proposed method aimed to try and improve upon the previous limitations of prior research by using similar ideas and combining them together. The proposed method is outlined in the flow diagram in Figure 5 and in more detail in the following sections.

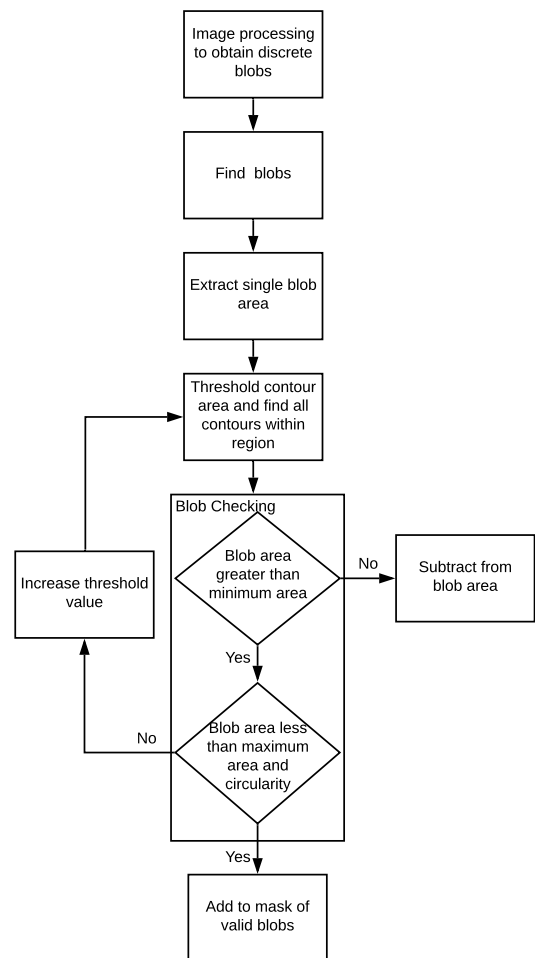


Fig. 5. Flow diagram of this papers proposed method to separate merged blobs

Thermal images were collected from two lecture theatres, A2 and A5. The A2 lecture theatre is significantly larger than A5 holding more people and has people seated further away. Therefore, the size of the blobs formed by people were significantly different and was accounted for in the method.

A. Method Steps

1) Image Processing

The method used to capture the thermal images used third party software [15] came pre-thresholded and with a colour map added, as shown in Figure 1. The pre-thresholding made easier to identify where the main hotspots are, however, it contained the semi-warm regions black. This made it difficult to threshold and separate out the exact region where people were. Therefore, relatively complicated methods were required in order to extract the thermal blobs.

Initially, the image was converted to HSV and the saturation and value matrices were separated as done in image segregation of satellite images [16]. The saturation matrix and value matrix were combined with a bitwise and to extract the regions of interest.

Further research has shown that using the third-party software is not required as the Lepton 3 camera works as a regular USB camera. Images collected as a regular USB camera have different colour scheme and are not directly compatible with this proposed method.

a) Differential Thresholding

A large number of the hot spots extracted contained multiple people and there were undesirable blobs caused by the projector and lights at the top of the image. The image containing the regions of interest was then split into a number of sections along the y-axis. A similar idea was used in a sports arena [17]. The higher up the region was in the image, the larger the threshold was applied to it. This worked well as a first step to separate some of the merged blobs. Additionally, the thermal blobs formed by lights and other noise were successfully removed.

2) Find Blobs

Once the image was initially processed, all of the blobs were found in the image using Green's Theorem [5]. However, this detected false positives, such as noise from people's feet, and many large blobs that consisted of multiple people.

3) Extract Single Blob/Thresholding Blob Area

The initial blob area was extracted from the pre-processed image. The threshold on the blob area was increased until its characteristics conformed to one of the required characteristics. As the threshold increased, the original blob area would split up into multiple smaller blobs, as shown in Figure 6.

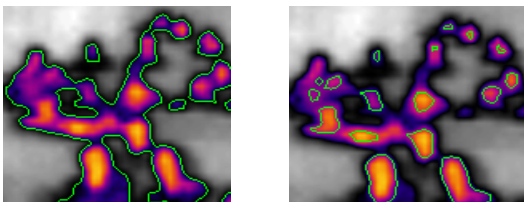


Fig. 6. Example blob a large blob before separation (left) and after (right).

All of the new blobs that are formed are then checked individually. The threshold on the original blob area is increased until the blobs are no longer too large or is too small.

4) Blob Comparison

From each blob that was extracted, the area, centre coordinates and minimum enclosing circle was determined. People further away in a lecture theatre have a smaller heat signature, as shown in Figure 1. Therefore, the maximum and minimum allowable area would need to change depending on the blob's location on the y-axis.

The optimal equations to describe the maximum and minimum allowable areas were initially determined by a linear relationship between the largest blob containing a single person and the smallest blob containing a single person. A number of different equations for the minimum and maximum area of blobs were tested.

For smaller lecture theatres, such as A5 lecture theatre, the optimal equations were:

$$A_{min} = -3.0 * (height - cy) + 1170 \quad (1)$$

$$A_{max} = -0.8 * (height - cy) + 2570 \quad (2)$$

$$Circularity = 0.30 \quad (3)$$

Where height is the number of pixels in the image and cy is y-coordinate of the center of the blob. The image processing had the y-axis flipped so the y value increased by moving lower in the image. The (height-cy) fixed this inversion so the y value increased by moving up the image

For larger and more complicated lecture theatres, the best results were achieved by using a quadratic formula for the A_{max} . The optimal equations for the larger lecture theatres, such as A2, were:

$$A_{min} = -0.9 * (height - cy) + 310 \quad (4)$$

$$A_{max} = 0.06 * (height - cy)^2 - 41 * (height - cy) + 7700 \quad (5)$$

$$Circularity = 0.34 \quad (6)$$

For every blob, the A_{max} and A_{min} values were recalculated using the y-coordinate of the centre of the blob. If the blob's area was determined to be larger than A_{max} , then the threshold value increased, and another thresholding would occur on the original blob area. Additionally, if the blob did not meet the minimum require circularity then the threshold value would also be increased, and another thresholding would be completed.

5) Subtract from Blob Area

If the blob area is determined to be too small compared to the A_{min} determined at the centre of the blob, then it was removed from the original blob area. This excludes it from being counted incorrectly as a person and also prevents it being checked on every iteration, saving processing time.

6) Add to Mask of Valid Blobs

If the blob was determined to be neither too small nor too large, then it was added to the mask of valid blobs. It was also

removed from the blob area to prevent it being check in every other iteration.

B. Advantages of Proposed Solution

This method uses a unique approach to detect a range of different sized blobs in a lecture theatre. Prior studies have only designed blob detection methods with blobs of similar sizes. This method makes use of simple equations to define whether a blob is likely to be an individual person or not. As these equations are quick to change, they can be easily modified to work in any lecture theatre. This has a significant advantage over other methods, such as genetic algorithms or neural networks as they would require significant training and large amounts of data before they could operate in a new environment.

IV. RESULTS

This method was implemented using a 160x120 Lepton 3 thermal camera. The image processing was done using OpenCV Python version 4.1.0 on Mac OS running on 2.4 GHz Intel Core i5 and 8 GB of RAM.

A. Detection Accuracy

A person missed detection when the method failed to separate blobs into individual people, or when the thresholding was too aggressive and the person to be classified as noise. Examples of both cases are shown in Figure 7.

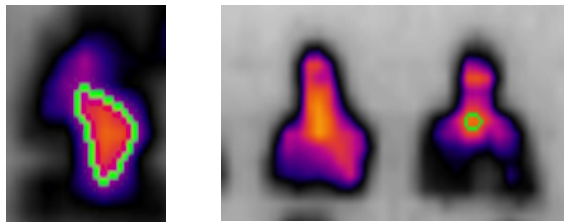


Fig. 7. Missed person detection examples. Failed separation (left) and blob classified as noise (right).

A false positive occurred when a blob was split into multiple smaller blobs, as shown in Figure 8. This caused two people to be counted when there was only one.

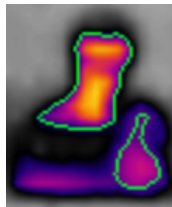


Fig. 8. Example of false positives caused by blob separation.

Optimal equations describing the A_{max} , A_{min} and circularity were designed for lecture theatres A2 and A5 as shown in equations 1-6. The accuracy of each these equations when used for their designed lecture theatre are shown in Table I. The equations were also tested using images from the other lecture theatre to determine how well they would work in a general sense.

TABLE I. LECTURE THEATRE OCCUPANCY ACCURACY

Equations Used	Lecture Theatre	People Correctly Detected	People Missed	False Positives
A2 Optimized Equations	A2	72/75	3	3
	A5	14/16	2	9
A5 Optimized Equations	A2	40/75	35	0
	A5	16	0	0

The results from Table I show a very high accuracy in successful detection rates in most cases.

When using the optimised equations, the method achieved a 100% successful detection rate in the A5 lecture theatre with no people missed or false positives. In the A2 lecture theatre, only 96% were correctly detected. However, the number of false positives and people missed were equal. Therefore, the total number of people detected was correct, but was achieved erroneously.

When using the wrong set of optimised equations, the proposed methods accuracy fell apart quickly. The correct detected rate in A5 stayed high at 88% but had a large number of false positives that would skew the number of people detected. The correct detection rate in A2 dropped significantly to only 54% but had no false positives.

The results from the propose method were also compared to the other detection methods outlined in the Background Research section in Table II.

TABLE II. COMPARISON OF PROPOSED METHOD AGAINST PRIOR STUDIES

Detection Method	Proposed Method (Optimised Avg)	Morphology [6]	Background Subtraction [8]	Neural Network (Thermal) [9]
Accuracy	98%	90%	96%	98%

As shown in Table II, the proposed method did very well compare to prior research and was actually more effective than the main study it was based off [6]. The neural network method was more accurate than the proposed method, however it would likely be much more difficult to create and set up for different lecture theatres.

The future aim of this project would be to have it installed into a lecture theatre and running off a Raspberry Pi. Therefore, the time taken to process each image was calculated and used as a rough estimation to the algorithm's computation requirements. Using optimised equations in the images taken from A5 containing 16 people, processing took 0.12s. Using optimised equations on the images taken from A2 containing 75, processing took 0.6s. This showed that the method was more computationally expensive with more people present. While it is likely to take significantly longer on a Raspberry Pi, due to lower processing power, it is still likely to be able to make many estimations during a lecture. Additionally, by using single images, the system does not have to operate in real time. As long as a single image can be processed within the time of a single lecture, the system

should be able to accurately measure the occupancy of the lecture theatre.

V. CONCLUSION

The results from the testing in multiple lecture theatres showed that the proposed method could be used to accurately measure the occupancy of a lecture theatre. The method proposed in this paper managed to achieve an accuracy rate of 100% and 96% in two different lecture theatres. This method was unique as it was used to detect different sized blobs. Previous studies had only focused on detecting blobs that were similar sized. Even with this new features, the proposed method achieved better blob detection compared previous methods such as morphology operations [6] and background subtraction [8]. These methods had previously achieved accuracy of 90% and 96% respectively. Other methods using neural networks [9] managed to achieve a higher accuracy of 98% which was greater than the proposed method.

This shows that neural networks are a more effective method in blob detection, but have trouble transferring to new environments, such as different sized lecture theatres. This problem could be overcome, but it is likely that a large amount of training data is required for neural networks. Some neural networks some requiring upwards of 500 images to work reliably [11]. The proposed method looked to improve on this by only requiring three parameters equations to be changed for A_{max} , A_{min} and the circularity constant. These equations are easily adjustable and can be done through trial and error on a single image.

A. Future Research

The first step to further develop the proposed method would be to have a single set of equations for A_{max} , A_{min} and circularity that would work to detect people accurately in any lecture theatre. This would mean the system could be more easily implemented into a lecture theatre without needing a calibration stage to find optimal parameters. This could be achieved by using automatic detection of blobs and scaling according to largest blob present as implemented by P. Majer [18].

Further research to optimise the proposed method can still be made to reduce the number of false positives and people missed by the proposed method. A genetic algorithm to could be implemented to quickly find the optimal equations that could be used to minimise the number of false readings.

Moving the thermal camera to different positions within the lecture theatre could act to minimize blob merging. A top-down camera angle could help remove some of the blob merging as people will no longer be covered by people in front of them. Further research into different camera angles could help improve the occupancy measurement accuracy.

Research into the thermal image collection showed that the third-party software [15] was not required to read data from the Lepton 3. Using the Lepton 3 as a basic USB camera gave images with different colour schemes. Further work could be done to adapt the code to work for the different image type

Finally, the proposed method could be modified to run off a Raspberry Pi with a constant image feed. The method would need to be modified to take multiple images during a lecture time and compare it to the maximum capacity of the room to determine utilization. By taking multiple images of the room during the lecture, the room occupancy could be determined multiple times and averaged to give more accurate readings.

VI. REFERENCES

- [1] J. Intoy and E. Lam, "Room occupancy sensing using a thermal tripwire," Jul. 2017.
- [2] Axis Communications, "AXIS People Counter," *Axis Communications*, 01-Feb-2019. [Online]. Available: <https://www.axis.com/en-nz/products/axis-people-counter>. [Accessed: 20-Jun-2019].
- [3] R. Gade, T. B. Moeslund, S. Z. Nielsen, H. S. Petersen, H. J. Andersen, K. Basselbjerg, H. T. Dam, O. B. Jensen, A. Jørgensen, H. Lahmann, T. K. O. Madsen, E. S. Bala, and B. Ø. Povey, "Thermal imaging systems for real-time applications in smart cities," *International Journal of Computer Applications in Technology*, vol. 53, no. 4, p. 291, Nov. 2016.
- [4] N. Bhattarai, T. Nakamura, and C. Mozumder, "Real Time Human Detection and Localization Using Consumer Grade Camera and Commercial UAV," Nov. 2018.
- [5] S. Suzuki and K. Abe, "Topological structural analysis of digitized binary images by border following," *Computer Vision, Graphics, and Image Processing*, vol. 29, no. 3, p. 396, Apr. 1985.
- [6] M. Berger and A. Armitage, "Room occupancy measurement using low-resolution infrared cameras," *IET Irish Signals and Systems Conference (ISSC 2010)*, Jul. 2010.
- [7] M. Arif, S. Daud, and S. Basalamah, "Counting of People in the Extremely Dense Crowd using Genetic Algorithm and Blobs Counting," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 2, no. 2, Aug. 2013.
- [8] J. W. Davis and V. Sharma, "Background-subtraction using contour-based fusion of thermal and visible imagery," *Computer Vision and Image Understanding*, vol. 106, no. 2-3, pp. 162–182, Jun. 2007.
- [9] A.-W. Negied, E. B. Hemayed, and M. Fayek, "HSBS: A Human's Heat Signature and Background Subtraction Hybrid Approach for Crowd Counting and Analysis," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 30, no. 08, Mar. 2016.
- [10] I. Amin, A. Taylor, F. Junejo, A. Al-Habaibeh, and R. Parkin, "Automated people-counting by using low-resolution infrared and visual cameras," *Measurement*, vol. 41, no. 6, pp. 589–599, Feb. 2008.
- [11] Y. Xue and N. Ray, "Cell Detection in Microscopy Images with Deep Convolutional Neural Network and Compressed Sensing," Feb. 2018.
- [12] D. D. Oliveira and M. Wehrmeister, "Using Deep Learning and Low-Cost RGB and Thermal Cameras to Detect Pedestrians in Aerial Images Captured by Multirotor UAV," *Sensors*, vol. 18, no. 7, p. 2244, 2018.
- [13] J. Davis and V. Sharma, "Robust detection of people in thermal imagery," *Proceedings of the 17th International Conference on Pattern Recognition*, 2004. ICPR 2004., Aug. 2004.
- [14] X. Bai, C. Sun, and F. Zhou, "Touching Cells Splitting by Using Concave Points and Ellipse Fitting," *2008 Digital Image Computing: Techniques and Applications*, Nov. 2008.
- [15] KheirIb, "KheirIb/purethermal1-uv-capture," *GitHub*, 19-Jun-2019. [Online]. Available: <https://github.com/KheirIb/purethermal1-uv-capture>. [Accessed: 20-Jun-2019].
- [16] P. Ganesan, V. Rajini, B. Sathish, and K. B. Shaik, "HSV color space based segmentation of region of interest in satellite images," *2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICT)*, Jul. 2014.
- [17] "Occupancy Analysis Of Sports Arenas Using Thermal Imaging," *Proceedings of the International Conference on Computer Vision Theory and Applications*, May 2012.
- [18] P. Majer, "The Influence of the γ -Parameter on Feature Detection with Automatic Scale Selection," *Scale-Space and Morphology in Computer Vision*, pp. 245–254, Nov. 2001.