

ANALYSIS OF COMBINED UAV-BASED RGB AND THERMAL REMOTE SENSING DATA: A NEW APPROACH TO CROWD MONITORING

S. Schulte ^{a*}, F. Hillen ^a, T. Prinz ^a

^a Institute for Geoinformatics, University of Münster, 48149 Münster, Germany - (s_sch032, florian.hillen, prinz)@uni-muenster.de

KEY WORDS: Video Processing, Data Fusion, UAV, Crowd Detection, RGB Imagery, Thermal Imagery

ABSTRACT:

Collecting vast amount of data does not solely help to fulfil information needs related to crowd monitoring, it is rather important to collect data that is suitable to meet specific information requirements. In order to address this issue, a prototype is developed to facilitate the combination of UAV-based RGB and thermal remote sensing datasets. In an experimental approach, image sensors were mounted on a remotely piloted aircraft and captured two video datasets over a crowd. A group of volunteers performed diverse movements that depict real world scenarios. The prototype is deriving the movement on the ground and is programmed in MATLAB. This novel detection approach using combined data is afterwards evaluated against detection algorithms that only use a single data source. Our tests show that the combination of RGB and thermal remote sensing data is beneficial for the field of crowd monitoring regarding the detection of crowd movement.

1 INTRODUCTION

During major events (like music festivals, demonstrations or shopping events) the gathering of large crowds might cause dangerous situations if the density of persons in an area reaches a critical level. The mass panic during the Love Parade 2010 in Duisburg is an example for this. Twenty-one festival visitors lost their lives. According to Lynskey (2011), G Keith Still states that "Accidents don't just happen, they're caused, every disaster we've researched was avoidable". It is often assumed that it is the peoples fault when a mass panic occurs, but in fact the reality is just the opposite. Organisers of large scale events have the responsibility to assure the safety of their guests. They need to pre-plan for specific situations, organise exit routes to quickly relief areas. According to Still (2014; 2017), most of the disasters were a result of overcrowding. One main problem is that the behaviour of crowds has to be assessed correctly in order to prevent potential disasters. Currently, the common techniques to monitor a crowd are, for example, security staff on the ground at specific vantage points, patrols or closed-circuit television (CCTV) that consists of multiple cameras controlled by an operator. The latter generates a lot of data but ultimately it is the operator who creates the information that is required to plan actions to prevent damage to humans or property. Pedestrian monitoring is also a form of crowd monitoring, it is of interest for a variety of reasons. For example, knowledge about the flow of pedestrian traffic is important for city officials, as they have to manage the available resources. One approach might be to alarm the operator when suspicious behaviour occurs in the video (Boghossian and Velastin, 1999).

In a time of rising processing power, the question is not any longer what can a computer achieve but how a human can use this processing power effectively. Gathering large amounts of data is easy and already done. CCTV cameras are creating large datasets, even though they are somewhat limited in their perspective and orientation. Therefore, the aim is not to collect more data, but better suitable data. Above that, the data have to

be processed in order to derive real world decisions from it. Thus, the operators of crowd monitoring systems do not have to work with raw data (i.e. CCTV camera footage) anymore. Instead it would be more efficient assist them in their decision making process with pre-processed and accumulated information. This paper focuses on a crowd monitoring approach that converts data recorded by unmanned aerial vehicles (UAVs) into usable information. Such information assist operators with assessing the situation at hand in a more effective manner. The first data sources we are using for this approach is thermal imagery, as it allows for an efficient detection of humans due to their heat signatures. We add RGB imagery to address the lack of information about the surroundings of thermal images. Above that, the amount of details that are visible in the RGB image data increased the visual impression for the operators. Our approach might help to detect dangerous situations within the crowd in an early stage and might thus increase the security of major events.

2 STATE OF THE ART

2.1 Related Work in Crowd Monitoring

Crowd monitoring is a field of research that directly correlates to the advancements in computer vision. Vannoorenberghe et al. (1997) published a paper about crowd monitoring using image sequence processing. In their research, they focused on a "video-based system used to collect pedestrian traffic data". They analysed angled shots at pedestrian crossings. First, they established a reference image with all the static edges in an image, then they detected all the moving deformable bodies in an image. This system was used as a smart sensor for the project Intelligent Crossroads developed by the French National Research Institute of Transportation and Security. The German aerospace centre (DLR) also conducted research in this field. They facilitated an approach to use high-resolution images from the Worldview-2 satellite. Their approach was to focus on

* Corresponding author

eliminating redundant features, like buildings, through an automatic generation of a digital elevation model. They processed the remaining features to detect individual persons using feature extraction (Sirmacek and Reinartz, 2011). In an effort to estimate the crowd density more accurately, Abuarafah et al. (2012) used a thermal camera in their research. They analysed a large crowd during the pilgrimage event Haji, an event where almost three million Muslims gather in Makkah. Estimations were based on how much heat is present in the angled shot of a crowd. Burkert and Fraundorfer (2013) presented an approach to monitor pedestrian groups. Their ultimate goal was to detect complex events in public scenes, an experimental approach was used to create different scenarios of crowd movements. An UAV was used to create a dataset of nadir (i.e. direction pointing directly below a particular location) imagery that depicts crowd scenarios. The volunteers were marked with coloured hats, to detect them in the imagery.

2.2 Research Question

In this paper a novel UAV-based approach to crowd monitoring is studied. The central research question is, whether the combination of RGB and thermal remote sensing data, captured with an unmanned aerial vehicle, is able to provide advantages in detecting and analysing the movement of human crowds? Previous studies extract face and body features, this is not possible with nadir imagery. To extract information out of such images, new approaches are necessary (Sirmacek and Reinartz, 2011; Meynberg and Kuschik, 2013). RGB imagery typically has a high resolution and contains many details about the study area, but the detection of people is no trivial task. As previously mentioned Burkert and Fraundorfer (2013) used to mark the volunteers in their research to detect them. Thermal imagery, however, provides additional attributes that could (in combination with RGB imagery) eliminate the need to mark the persons in a crowd to detect them properly and decrease the false-positives. Yet, thermal imagery typically has a limited spatial resolution. Thus, we try to utilise the benefits of both datasets by fusing them. We assume that this generates more relevant information about crowd in the context of automatic crowd monitoring.

3 METHODOLOGY

3.1 Experimental Setup

The focus for the initial test was not to use the most advanced thermal and RGB camera but an affordable solution, as thermal cameras quickly cost more than thousand euros. If the results prove fruitful on affordable hardware, they can always produce better information with more specialised sensors.

In the context of continuous crowd monitoring, it makes sense to use a vertical take-off and landing aircraft (VTOL) because they are able to hover at a fixed position. This enables the usage in limited space and allows to position the vehicle over a crowd. The payload capacity of such vehicles has to be rather high in order to carry two cameras at the same time and to still get reasonable flight times. The md4-1000 of the German manufacturer microdrones is a reliable aircraft that fulfils the requirements of the task (Figure 1).



Figure 1. Md4-1000 from microdrones, equipped with thermal and RGB sensors.

The forward-looking infrared (FLIR) camera Scout TK was chosen as a suitable camera for the thermal imagery. It measures relative heat differences instead of absolute values. Thus it is not possible to calculate the actual temperature with this camera. It is a low-cost solution with a relatively small form factor and moderate weight of 170 gram, while supporting picture and video recording with a maximum detection range of 90 metres. The field of view (FOV) is rather limited with 20° horizontal and 16° vertical opening angle. One drawback of this camera is the lack of line-out capabilities. Thus, no direct image downlink is possible to correctly position the camera over the crowd. The video mode records at about nine frames per second and the resolution is fixed at 320 x 240 pixels. For the RGB imagery the Sony A5100 with a resolution of 1920 x 1080 pixels and a weight of 224 grams was selected.

For the data recording, a crowd of 12 adult volunteers, one child, and a child in a stroller, was gathered in a marked field of 20 m x 16 m, that represented the area of interest (AOI) for the study. In order to keep the participants in the FOV of the thermal camera, the AOI was calculated using the maximum flight height and the specifications of the thermal camera. Aluminium plates were used as ground control points (GCPs), as they reflect 90% of thermal radiation in the atmosphere resulting a black signature in the thermal video (Hartmann et al., 2012). This helps to improve the matching of both images and to create a georeferenced dataset for the upcoming processing. The flight conditions of approx. 10° C with wind speeds up to 10.9 km/h (measured at 34 m above ground level) allowed for a stable positioning of the UAV in the air. The mechanical vibrations of the drone emerged as a problem for the thermal camera due to its narrow FOV. The resulting shift in the image increased the mismatch between the video sources.

Different crowd scenarios were performed by the group of volunteers illustrated in Figure 2. Those scenarios should reflect the real-world behaviour of crowds during major events. The main stage scenario (Figure 2a) is intended to depict a situation at a concert, where a crowd is forming in front of a stage and people are moving between standing people. The scene 'Big Group Pass-through' (Figure 2b) depicts a situation where one person walks in and out of a large group of people. Other scenarios depict the group quickly disperse (Figure 2f), moving through a narrow passage (Figure 2d), doing irregular movements (Figure 2e) and two small groups where one member of each group exchange places (Figure 2c).

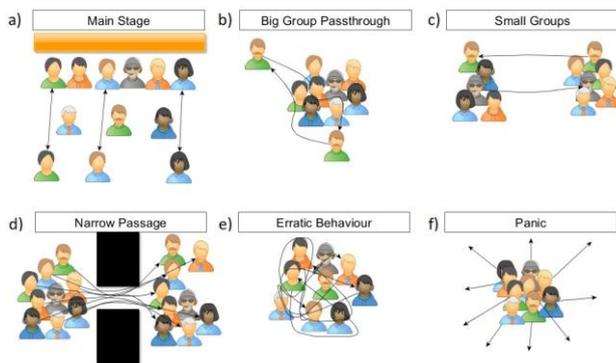


Figure 2. Different crowd scenarios performed by the group of volunteers during the flight campaign.

3.2 Data Pre-processing

As mentioned in the previous section, both datasets have different specifications. The thermal camera had to be levelled due to its round shape. However, it was not possible to perfectly level the camera compared to the RGB camera. Moreover, it was required to enlarge the thermal image by a small amount to have the same topography dimensions. Otherwise, the persons and distances would have been smaller than in the RGB image. The framerate for the RGB video had to be rendered down to nine frames per second to match the framerate of the thermal video. The videos were also matched with more than one keyframe to reduce differences caused by mechanical vibrations.

3.3 People Detection Algorithms

The workflow to detect people in the images is split into two parts, the RGB detection algorithm and the thermal detection algorithm. In the following, both algorithms are described in detail.

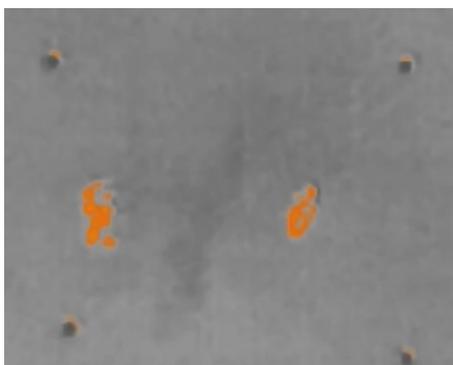


Figure 3. Cropped thermal imagery at an altitude of 70 metres.

The thermal detection algorithm is designed to extract the connected components of the local maxima, i.e. the highest temperatures recorded at specific locations. The algorithm to detect people and their centroid in an image was adopted from the approach of Anand Bhargav (2010). The script presents a solution to detect red objects in an RGB webcam stream and draw bounding boxes around them. As the thermal imagery is represented as RGB datasets as well, this approach can be adapted. Subtracting the red channel (Figure 4, left) from the combined grayscale representation of all three channels of the

thermal image (Figure 4, middle) results in pixel brightness values of zero for regions containing no red colour. Both images only differ when red is present, as shown by the contrived digital number (DN) values beneath.

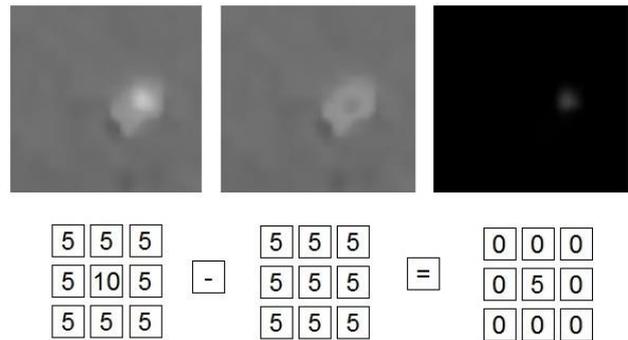


Figure 4. Subtraction of the red channel (left) from all three channels as one grayscale image (middle), resulting in only red parts of the image (right); with corresponding artificial digital number (DN) values for every image.

The resulting image (cf. Figure 4, right) is then binarised with a determined threshold. All values above the threshold are converted to 1, while all values below the threshold are set to 0. At this point one option can be to remove large objects when they are too large to represent a person. However, this is not practical as people standing in a uniform group are represented as one large heat signature in the thermal imagery. Structures with only a few pixels are removed as they seem to represent false positive matches (no person). The resulting objects have different properties which can be extracted out of binary images with MATLAB's built-in function *regionprops*. This function can be set to solely extract specific properties. Of interest are the coordinates for the centre and the smallest rectangle containing the heat signature referred to as centroid and bounding box respectively. The centroid has two values, the image space coordinates *x* and *y*. However, the bounding box has to be stored with four values, the coordinates and the dimensions (i.e. width and height). All this information for each object is stored in an array. MATLAB labels these objects with numbers from lower-left to upper-right. The information about the people is now extracted out of the thermal imagery and the array is further called: 'thermalprops'.



Figure 5. RGB imagery at an altitude of 70 metres.

The detection of people in the RGB imagery (Figure 5) is based on geometric processing. We utilise MATLAB's Sobel edge detector to determine edges in the image. This neighbourhood operation requires a grayscale image, so the first step is to

convert the RGB video frame to grayscale. The built-in MATLAB function *rgb2gray* is used for this purpose. Afterwards, a Gaussian filter is applied to the grayscale image to filter the image noise and to increase the edge detail. This also removes smaller structures while larger ones are preserved (Priese, 2015). The first iteration of the edge detection algorithm is used to determine the threshold factor. This factor determines which edges are holding a major transition in brightness. This threshold is then multiplied with a so-called ‘fudge factor’ to reduce the amount of edges leaving only the edges with a larger brightness difference. Further improvement is achieved by removing edges that have too few pixels to represent a person. Afterwards, the built-in function *regionprops* is used to calculate the centroid and bounding box coordinates of every edge. The result of this operation is stored in an array which is further called: ‘*rgbprops*’.

3.4 Near Real-time Data Fusion

As previously mentioned the results of the thermal detection have to be brought into relation with their surroundings so that they can be interpreted. The algorithms for the thermal and RGB imagery are executed simultaneously. The complete process is visualised in Figure 6.

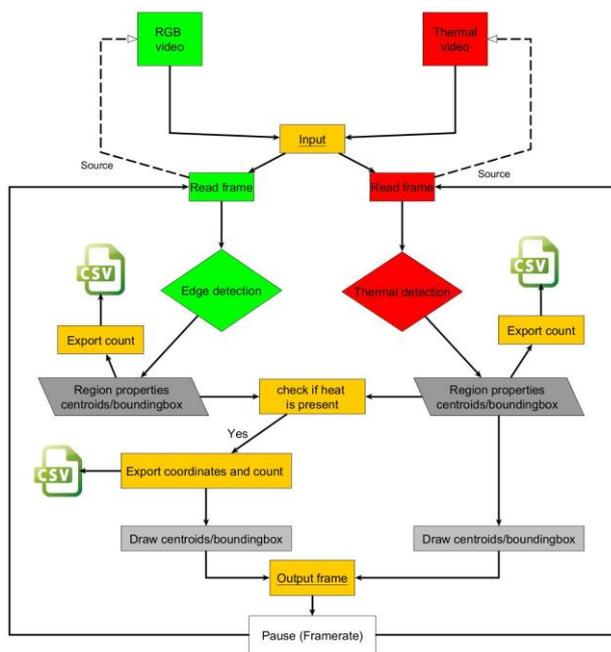


Figure 6. Schematic illustration of the near real-time fusion process of RGB and thermal video data.

Both algorithms result in arrays that store the region properties of the centroids and bounding boxes of connected components. The length (i.e. the number of objects) of the arrays is exported into two separate comma-separated values (CSV) documents. In order to separate false positives, every coordinate from the *rgbprops* array is checked whether it overlaps with a bounding box generated by the objects of *thermalprops*. This verification process uses the centroid coordinates for the RGB objects and the bounding box dimensions of the thermal objects, as both detection methods do not result in the same centroid coordinates. This is also the most important step with respect to data fusion as it combines information from both data sources. Additionally, these values are appended to a CSV document

that is created for every object. The corresponding bounding box and centroid of the object is drawn over the displayed RGB frame. Depending on the source, the boundaries have different colours: red for objects that are only detected in the thermal image and white for objects that are concurrent in the RGB image and hold a corresponding heat signature in the thermal image. In order to achieve an animated visualisation of the combined data, the calculated frame has to be displayed at a specific framerate. To ensure a reasonable framerate a pause was implemented between every frame. This procedure is repeated until both videos have been evaluated completely.

3.5 Crowd Analysis

In order to evaluate the different detection methods and to analyse the trend of the crowd and the individuals contained, the results of every frame are evaluated over a certain time. During the near real-time analysis of the remote sensing data locations of every detected person and the headcount of each detection method are stored in CSV documents. To visualise the results the first frame of the corresponding RGB video is displayed as a reference image. Different visualisation styles can be used to visualise the people and their movement. These coordinates can be visualised either with points or by lines (Figure 7).

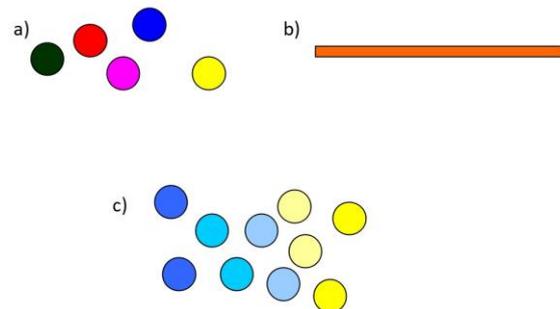


Figure 7. Available visualisation styles: Point style (a) and line style (b) with individual colour for each detected person; (c) point style with gradient colour scheme based on time for whole crowd

The path is plotted out of these coordinate lists, resulting in a coloured line for every detected object. The high sampling rate is a problem with the line visualisation because this results in Zig-Zag lines, as the image space coordinates are relative and thus change with vibrations. Zig-zag lines are the result of the frame-by-frame analysis, as the objects of one frame do not correlate to the next frame. To counter this, the coordinate sampling rate can be lowered to a suitable value. When a rate of ‘18’ is chosen, coordinates are sampled in an interval of two seconds as every second of video consists of 9 frames. This reduces the chance of falsely assigned coordinates as well as the oscillations from the sensor.

The point style more appropriate to the line style for our task as every recognised location of our participants is displayed over the reference image. This eliminates the risk of coupling the wrong coordinate pairs in a line while still being able to grasp the covered area by the crowd. Another visualisation option that we implemented, is to colour the points with a gradient colour scheme varying over time. This emphasises the change of the crowd over time, making it possible to see how the crowd moved. A blue to yellow gradient is implemented, blue represent the beginning of the clip, yellow the end. Above that,

we calculated the travelled distances of the people as it is an indicator to determine whether a person is hectic, calm or even lost. For this, the Pythagorean distance formula is used.

4 RESULTS

4.1 Prototype Implementation

For a more convenient testing, a prototype implementation with a graphical user interface (GUI) for the combination of RGB and thermal video datasets has been conducted. Figure 8 shows a screenshot of the developed GUI which allows for adjusting the parameters of the processing algorithms (e.g. to address differences in lighting condition of scenes). This gives the user the possibility to test different setups of the detection methods.



Figure 8. Graphical user interface of the prototype implementation for the combined processing of RGB and thermal video datasets.

The visualisation is reduced to depict only the AOI. The processing is not optimised regarding speed, therefore the current processing status is displayed right under the frame in order to keep the user informed. It is possible to switch to a script that exploits the processing power of the graphics processing unit (GPU). However, this feature is still under development and does not provide a significant speed increase at the moment.

The analysis features of the GUI become visible when the processing is finished. The user is able to set the required settings (e.g. sampling rate or visualisation style) and is able to choose between the arithmetic mean, median or maximum values of the headcount. The GUI allows to switch between implemented visualisation styles quickly. The three styles offer different kinds of information (Figure 7). The outcome of the processing of the 'Narrow Passage' scenario, is displayed in Figure 9. Visualising the results with a gradient colour scheme allows the user to see where the crowd was at the beginning of the clip (blue) and where at the end (yellow), everything in between is marked by the colour gradient. In the context of crowd monitoring, this visualisation style is preferred.



Figure 9. Visualisation styles side by side: Gradient (left), points (middle), lines (right), on the result of the 'Narrow Passage' scene

The performance of the frame-by-frame analysis of 9 frame per second within the prototype is appropriate compared to the approach of Burkert and Fraundorfer (2013), with only 1 frame per second. The calculations were performed on a laptop (quad-core CPU with 2.7 GHz, 32 GB RAM, NVIDIA GTX 1070 graphics card) and usually took around 3.5 times longer than the actual runtime of the video. It is possible to visualise the results of the detection algorithms in near real time. However, a lot of processing effort is needed to provide the visualisation for the user in near real time. In preliminary tests, the processing time could be reduced to one half, when the visual feedback is not shown and only the results are stored for the analysis.

4.2 Differences Between Detection Algorithms

Figure 10 shows a sample frame and the corresponding result (i.e. possible people locations) of the detection algorithms. The thermal detection algorithm shows difficulties separating the different heat signatures in a dense crowd and shows some errors on the upper right of the image. The edge detection algorithm applied to the RGB image on the other hand, falsely recognises our GCPs and the AOI markings, as possible persons. These errors are reduced using the fused information from both data sources, as shown in the right image of Figure 10.

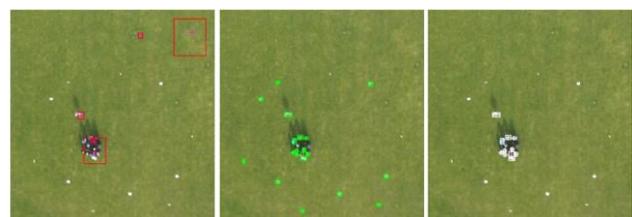


Figure 10. Results of the detection algorithms for the same video frame: thermal image only (left), RGB image only (middle) and fused information (right).

While the combined approach recognises 10 people, the thermal detection algorithm recognises only 3 and also has 2 false detections. The RGB detection algorithm returns 10 people with 9 false-positives. The actual real world crowd consisted of 12 adults and 2 children as mentioned in Section 3.1. This indicates that the combined method is the most accurate and has the least amount of errors, while still not detecting every person in the scene. The mean headcount for each individual scenario is shown in Figure 11. The errors of the detection processes are not removed because it is impossible to do so for every frame. Therefore, the good performance of the thermal detection algorithm seems to be misleading as it more error-prone than the combined approach. If errors would be subtracted the actual headcount and performance of thermal detection algorithm would be lower. However, sometimes the combined detection is producing false positives too. The thermal camera recorded a

heat signature around the GCPs resulting in a detected object in RGB and thermal images. The performance of the RGB detection algorithm is non-satisfying with an average headcount of 70 for every scenario. However, this might be a result of the larger extent of the analysed image.

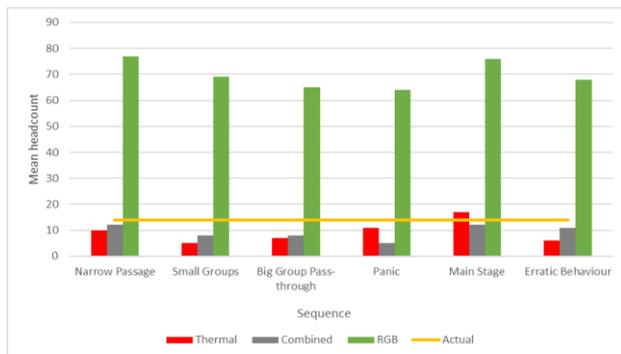


Figure 11. Diagram of the mean headcount for every crowd scenario and every detection method with the actual real world headcount.

A better understanding of why the combined detection is superior is provided with Figure 12 visualising the result of the “Big Group Pass-through” scenario. Both detection methods, the thermal detection algorithm on the left and the combined detection approach on the right, record almost the same amount of people. However, due to the large errors at the border of the thermal dataset in this clip, the thermal detection algorithms is not capable of detecting the large crowd in the middle of the scene. The combined approach clearly depicts one large group with one person walking into the group on the upper left, coming from the lower left.



Figure 12. Frame of the thermal detection (left) and of the combined detection (right), taken from the result of the “Big Group Pass-through” scenario.

The weaknesses of the thermal detection algorithm are apparent in the “Small Groups” scenario. This depicts two people exchanging positions between opposing groups. The thermal detection algorithm recognises the two groups as two persons as shown by the red rectangles in the right frame of Figure 13. Even with the human vision, it is not possible to accurately distinguish between the people in the thermal frame (Figure 13, left).



Figure 13. Cropped thermal video frame on the left and the same frame during the processing on the right. The red bounding boxes show detected heat signatures; the white show detected people from the combined approach.

In the “Panic” scenario the performance of the combined approach is lower than in the other scenarios, with a mean headcount of 5 people. It struggles to recognise every person in the frame. This is the result of the mechanical vibrations during the video recording. The thermal camera has a smaller FOV, thus motion is affecting the captured frame more. The heat signatures differ by a small amount to the real-world position of the person (Figure 14, right). The algorithm for the combined approach is not able to find a corresponding heat signature at the people’s locations.



Figure 14. Result of the “Panic” scenario visualised with a point style on the left and a red frame during the processing on the right.

However it is still possible to summarise what happened in this scenario with just one picture. This is depicted in the left image of Figure 14 and with the gradient point visualisation style in Figure 15. The crowd was confined to a small area at the beginning (blue points in Figure 15) and then quickly dispersed in all directions (yellow points in Figure 15). Thus, the mean headcount just means that not every person was detected in every frame of the sequence. This is not that problematic, as the maximum headcount for the combined detection is 15. This connotes that the general movement of the crowd is still captured. Above that, the dynamics of the crowd in this particular scenario can clearly be depicted from the visualisation with a gradient colour scheme presented in Figure 15.

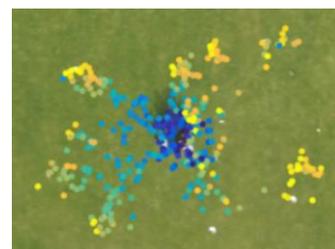


Figure 15. Result of the “Panic” scenario visualised as points with a gradient colour scheme.

5 CONCLUSION

In prior research, crowd detection methods usually exploit only one source of information. However, the presented approach combines two sources to provide a more accurate and more reliable method. For this, video datasets acquired from RGB and thermal sensors are analysed. The datasets were taken with a UAV that had both sensors mounted to the gimbal at the same time. A group of 14 participants performed six diverse pre-planned scenarios that were designed in the context of crowd movement during major events. Data pre-processing has been conducted on the thermal video source to fit in the dimensions of the RGB video source. The frame-by-frame detection algorithms were implemented in MATLAB, as was the presented analysis process. The results of the combined detection of people are promising and show advantages over the analysis of the two image sources separately. A better and more error-proof detection is feasible. Even in a small AOI, that does not contain many different objects, it is necessary to remove the false-positives of the RGB edge detection. The results indicate that it is possible to detect crowd movements on an appropriate level, i.e. on a level that might potentially be suitable for integrating in modelling approaches or crowd simulations. It can be seen that even our combined approach produces less errors regarding the detection and runs in near real time.

The laws in Germany prohibit the hovering over individuals or groups with a UAV. Thus, it was not possible to gather datasets of real events for this study. The results of this study rely on the fact that our testing environment approximates the reality in order to give an accurate answer to the research question. The maximum amount of people depicted in the video frames is only 14. It has to be noted that this is not a large crowd. However, due to them being confined in a smaller AOI, they can still provide an outlook on the possibilities and advantages of the developed approach. The RGB detection algorithm is only based on an edge detection filter, however, there are different approaches that could improve the detection rates. These approaches usually do not use nadir imagery as they often classify persons by their visual appearance which is not possible with the low number of characteristics in aerial images. Moreover, the imaging devices are not calibrated and do thus not account for the different distortions that might occur. This can result in errors in the detection and misplaced alignment of thermal attributes to the RGB imagery. One additional limitation at the moment is that the developed approach produces image coordinates only. A completely stable setup is necessary to work within a geographical coordinate system. The thermal detection algorithms inherits another limitation, as every detected heat signature is recognised as a human body. This might be problematic under real-world conditions as there are usually a lot of different heat sources apparent, like street lights, heated tarmac or solar panels. Future research has to evaluate how to handle those heat sources in the developed detection method. If the regulations allow for it in the near future, it is necessary to validate our approach with real-world data during a major event. These scenarios hold more crowd dynamics and are most likely different from the simulated scenarios used in this study. An interesting aspect in this regard would also be the processing time needed to analyse one second of video data using a larger dataset. Above that, it would also be interesting to see how our approach performs in comparison with a more sophisticated airborne crowd detection algorithm (e.g. Sirmacek and Reinartz, 2011). The thermal camera used for this study is potentially not sufficient for applications in real-world scenarios of crowd monitoring. The lack of line-out

capabilities prevents the a real-time downlink to a ground station. However, such link is required for a real-time crowd monitoring test. Above that, information from a more sophisticated thermal camera, that shows absolute temperature values, would also increase the possibilities and the certainty of people detection. The detection algorithm could be adjusted specifically for the body temperature of humans. Detecting and maintaining the same object (here a single person) through every frame is also a very important task in the context of crowd monitoring. In this context, the advantage of UAV-based imagery is that there are almost no occlusions meaning that a person could always be tracked. This potentially allows for calculating the speed, travelled distance and repetition of movement. On-the-fly georectification of video frames would allow for reducing the jagged lines in the post-process analysis. This would also make it possible to incorporate a navigation layer that opens up routing possibilities, for example, the calculation of the least-crowded route to a point of interest or the routing for emergency teams to an injured person. Another interesting aspect would be to estimate the imminent crowd behaviour. This would allow for warning the operator before a critical event occurs, e.g. when a large group approaches a narrow passage similar to the complex event detector proposed by Burkert (2013).

ACKNOWLEDGEMENTS

We thank Jan Lehmann of the Department for Applied Landscape Ecology, University of Münster for his technical assistance and for flying the md4-1000.

REFERENCES

- Bhargav Anand, A., 2010. Tracking red color objects using matlab, <https://de.mathworks.com/matlabcentral/fileexchange/28757-tracking-red-color-objects-using-matlab>. (Accessed 9 March, 2017).
- Boghossian, B.A., Velastin, S.A., 1999. Image Processing System for Pedestrian Monitoring Using Neural Classification of Normal Motion Patterns. *Measurement and Control* 32 (9), 261–264.
- Burkert, F., Fraundorfer, F., 2013. UAV-based monitoring of pedestrian groups. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XL-1/W2*, 67–72.
- G. Keith Still, 2017. Crowd risks, <http://www.gkstill.com/ExpertWitness/CrowdRisks.html>. (Accessed 1 June, 2017).
- Hartmann, W., Tilch, S., Eisenbeiss, H., Schindler Konrad, 2012. Determination of the UAV position by automatic processing of thermal images.
- Lynskey, D., 2011. Danger in numbers – can mega-gigs ever truly be safe?, <https://www.theguardian.com/music/2011/jul/20/love-parade-crowd-safety-crush>. (Accessed 1 March, 2017).
- Meynberg, O., Kuschik, G., 2013. Airborne Crowd Density Estimation. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences II-3/W3*, 49–54.

Priese, L., 2015. Computer Vision. Einführung in die Verarbeitung und Analyse digitaler Bilder, 1. Aufl. 2015. Springer Berlin Heidelberg, Berlin, Heidelberg.

Sirmacek, B., Reinartz, P., 2011. Automatic crowd analysis from very high resolution satellite images. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XXXVIII-3/W22, 221–226.

Still, G.K., 2014. Introduction to crowd science. CRC Press, Boca Raton.

Vannoorenberghe, P., Motamed, C., Blosseville, J.M., Postaire, J.G., 1997. Crowd monitoring using image sequence processing, 197-206.