Long-Term Occupancy Analysis using Graph-Based Optimisation in Thermal Imagery

Gade, Rikke; Jørgensen, Anders; Moeslund, Thomas B.

Published in:
IEEE conference on Computer Vision and Pattern Recognition

DOI (link to publication from Publisher):
10.1109/CVPR.2013.474

Publication date:
2013

Document Version
Early version, also known as pre-print

Long-term Occupancy Analysis using Graph-Based Optimisation in Thermal Imagery

Rikke Gade, Anders Jørgensen and Thomas B. Moeslund
Visual Analysis of People Lab
Aalborg University, Denmark
{rg, andjor, tbm}@create.aau.dk

Abstract

This paper presents a robust occupancy analysis system for thermal imaging. Reliable detection of people is very hard in crowded scenes, due to occlusions and segmentation problems. We therefore propose a framework that optimises the occupancy analysis over long periods by including information on the transition in occupancy, when people enter or leave the monitored area. In stable periods, with no activity close to the borders, people are detected and counted which contributes to a weighted histogram. When activity close to the border is detected, local tracking is applied in order to identify a crossing. After a full sequence, the number of people during all periods are estimated using a probabilistic graph search optimisation. The system is tested on a total of 51,000 frames, captured in sports arenas. The mean error for a 30-minute period containing 3-13 people is 4.44 %, which is a half of the error percentage optained by detection only, and better than the results of comparable work. The framework is also tested on a public available dataset from an outdoor scene, which proves the generality of the method.

1. Introduction

Measuring the occupancy maps from people has become an essential step towards an intelligent and efficient society [21, 33]. A well-known example of this is that the whereabouts of people in shopping malls provides valuable information for the managers. The same goes for sports arenas. These facilities are in high demand, but very expensive to build, so focus of the political systems has shifted towards optimising the use of the existing arenas. The first step in this analysis is to monitor the occupancy of such facilities. As this analysis should run for several weeks in each arena, manual observations would be expensive and cumbersome, and an automatic system based on computer vision is therefore suggested. While RGB-based systems are normally used in previous research in sports analysis [1, 22, 12, 40], a general public acceptance of more permanent installations in such facilities are harder to come by due to privacy issues. We therefore apply thermal imagery, which captures the infrared radiation instead of visible light, and creates an image whose pixel values represent temperature. People can not be identified in thermal images, thereby eliminating the privacy issues. A positive side effect of thermal imaging is that detection can often be reduced to a trivial task. However, thermal imaging also introduces new problems, as people are often fragmented into small parts, and reflections can be seen in the floor. Moreover, the challenges of occlusions remain in thermal images, see figure 1.

Figure 1. Examples of the challenges for detection of people.

The contribution of this work is a reliable method for occupancy analysis in thermal video. The method does not assume a perfect detection in each frame, but handles the detection challenges by including temporal information. The main focus is not short lab sequences, but rather long, real-life sequences. Here we use data from sports arenas, which are very challenging, due to the natural physical interaction in sport.

The main idea is to split the video sequences into two types of periods. The first type is the stable periods, where no people exit or enter the court. In these periods, the number of people on the court must be the same, which in turn introduces a constraint on the problem. The second type defines unstable periods, where the occupancy is likely to change. Combining these two types of information to model the periods and transitions between them provides a unified framework to optimise over a long period of time.

1.1. Thermal radiation

Thermal imaging is still a relatively new modality in computer vision applications, and the theory behind it is relatively unknown in the computer vision society. This section will therefore provide information on the physical foundation of thermal radiation and cameras.

All objects with a temperature above the absolute zero emit infrared radiation, mainly in the mid-wavelength infrared spectrum (MWIR, 3-5 μm) and long-wavelength infrared spectrum (LWIR, 8-15 μm). This is often referred to as thermal radiation. The intensity of the radiation from an object with temperature T is described by Planck's Law as a function of the wavelength λ:

$$I(\lambda, T) = \frac{2\pi hc^2}{\lambda^5 \left(e^{hc/\lambda k_B T} - 1\right)} \tag{1}$$

where h is Planck's constant ($6.626 \times 10^{-34} Js$), c the speed of light ($299,792,458 m/s$) and k_B Boltzmann's constant ($1.3806503 \times 10^{-23} J/K$). From this expression, it can be seen that the intensity peak shifts to shorter wavelenghts as the temperature increases. For extremely hot objects, the radiation extends into the visible spectrum.

The thermal radiation originates from energy in the molecules of an object. The energy can be expressed as a sum of four contributions [36]:

$$E = E_{electronic} + E_{vibration} + E_{rotation} + E_{translation} \tag{2}$$

Only the energy caused by translation, rotation and vibration in a molecule contributes to the temperature of an object.

It is well-known from quantum physics, that visible light consists of photons that causes electron transitions when they are absorbed or emitted from a molecule. The same principle applies to infrared light, with the difference that the photons contain less energy and cause transitions in the vibrational and rotational energy levels instead. The electromagnetic radiation can be absorbed or emitted by the molecule, then the incident radiation causes the molecule to rise to an excited energy state, and when it falls back to ground state a photon is released. Only photons with specific energies, equal to the difference between two energy states, can be absorbed and emitted.

If more radiation is absorbed than emitted, the temperature of the molecule will rise until equilibrium is re-established. Likewise, the temperature will fall if more radiation is emitted than absorbed, until equilibrium is re-established.

1.2. Thermal cameras

Generally two types of detectors exist for thermal cameras: photon detectors and thermal detectors. Photon detectors convert the absorbed electromagnetic radiation directly into a change of the electronic energy distribution in a semiconductor by the change of the free charge carrier concentration. This type of detector typically works in the MWIR spectrum, where the thermal contrast is high, making it very sensitive to small differences in the scene temperature. The main drawback is the need for cooling of the detector, making it more expensive and with a higher need for maintenance. The thermal detector converts the absorbed electromagnetic radiation into thermal energy causing a rise in the detector temperature. Then, the electrical output of the thermal sensor is produced by a corresponding change in some physical property of material, e.g., the temperature-dependent electrical resistance in a bolometer. This type of detector measures radiation in the LWIR spectrum. They are uncooled and have been developed with two different types of sensors: ferroelectric detectors and microbolometers, where today the microbolometer has shown to have more advantages.

1.3. Related work

Detection of people is the first step in many applications, e.g. surveillance, tracking, or activity analysis. General purpose detection systems should be robust and independent of the environment. The thermal cameras can here often be a better choice than a normal visual camera.

The methods applied to thermal imaging span from simple thresholding and shape analysis [43, 17, 39, 15, 7] to more complex, but well-known methods such as HOG and SVM [42, 37, 41, 31, 26] as well as contour analysis [10, 9, 27, 38]. Using simple methods allows for fast real-time processing, and combined with the illumination independency, the thermal sensor is very well suited for detecting humans in real-life applications.

An obvious application area for thermal imaging is pedestrian detection systems for vehicles, due to the cameras' ability to "see" during the night. These systems are being developed both as assistance for drivers in low visibility, and as a navigation tool for the future automatic vehicles. One of the car-based detection systems is proposed in [4], where they present a tracking system for pedestrians. It works well with both still and moving vehicles, but some problems still remain when a pedestrian enters the scene running. [13] proposes a shape-independent pedestrian detection method. Using a thermal sensor with low spatial resolution, [28] builds a robust pedestrian detector by combining three different methods. [19] also proposes a low resolution system for pedestrian detection from vehicles. [32] proposes a pedestrian detection system that detects people based on their temperature and dimensions, and tracks them using a Kalman filter. In [2] a stereo-vision system has been tested, detecting warm areas and classifying if they are humans, based on distance estimation, size, aspect ratio, and head shape localisation.

A more general interest in pedestrian detection based on thermal imaging can also be seen in surveillance or for analysis of pedestrian flow in cities. A general purpose pedestrian detection system is proposed in [8]. The foreground is separated from the background, after that shape cues are used to eliminate non-pedestrian objects and appearance cues help to locate the exact position of pedestrians. A tracking algorithm is also implemented. [3] uses probabilistic template models of four different poses for detection. [30] also uses probabilistic template models, here they use three models representing different scales. [29] uses a statistical approach for head detection as the first step in the pedestrian detection.
The previously described methods use thermal sensors only. Combining different types of sensors could, however, eliminate some of the disadvantages from both sensors. Examples of systems combining thermal and RGB cameras are given by Davis et al. [9, 11] and Leykin et al. [23, 24]. Other sensors like laser scanners and near-infrared cameras, have also been combined with thermal sensors [14, 35].
Due to privacy issues, this work will concentrate on thermal cameras only. We will also take advantage of the easy foreground segmentation, but as shown in figure 1, challenges still remain. As opposed to most existing work, it will be tested on long sequences of real data with high complexity.

2. Approach

As described in the introduction, precisely counting people in single frames can be a nearly impossible task, due to occlusions and segmentation errors. Therefore, it is suggested to include temporal information, and estimate the occupancy over longer periods. The idea is to automatically split a video sequence into stable periods, with no activities near the border of the court, and transition periods with activity near the border. During the stable periods, the detected number of people in each frame contributes to a distribution of observations for that period. For the transition periods, local tracking of the blobs in the border area is applied, in order to estimate the likelihood of crossings. The two types of data and their uncertainties are combined in a graph, where the nodes represent the number of people, and the edges represent the change in number between two periods. A dynamic programming approach is applied to find the optimal path of the graph.
The remaining part of section 2 describes the details of the people detection and the monitoring of transitions. In section 3 the graph optimisation is described, and in section 4 the system is evaluated. The conclusion is found in section 5.

2.1. People detection

The first step towards detecting people is to separate foreground from background. Using thermal imagery in an indoor environment simplifies this task, as the surrounding temperature is normally stable and colder than the human temperature. There can, however, be observed warm spots, e.g, from heaters, hot water pipes, and doors or windows heated by the sun. A background subtraction method is used to remove static objects from the foreground. Since the image depicts the temperature of an indoor scene, it can be assumed that only slow changes will occur in the background. Therefore, the background image simply consists of the average of the previous n frames, but only pixels that are classified as background will contribute to the new background estimate.
Even though the foreground is now found, pixel noise should be removed. Moreover, due to the camera having automatic gain adjustment, the level of pixel values can suddenly change, without any temperature changes in the scene. To overcome these challenges, an automatic threshold method based on maximum entropy is used to calculate the threshold value for each frame [20]. From this point the image is binary, and all blobs found are considered potential persons. The next part, section 2.2 and 2.3 will deal with the splitting and sorting of blobs into single persons.

2.2. Groupings

Since a side-view of the scene is obtained, see section 4.1, it is necessary to be able to handle occlusions. Generally, two types of occlusions are seen: people standing behind each other, seen from the camera's point of view (”tall blobs”) and people standing close together in a group (”wide blobs”).

2.2.1 Split tall blobs

In order to split people that form one blob by standing behind each other, it must be detected when the blob is too tall to contain only one person. We here adapt the method from [17]. If the blob has a pixel height that corresponds to more than a maximum height at the given position, see section 4.1, the algorithm should try to split the blob horizontally. The point to split from is found by analysing the convex hull and finding the convexity defects of the blob. Of all the defect points, the point with the largest depth and a given maximum absolute gradient should be selected, meaning that only defects coming from the side will be considered, discarding e.g. a point between the legs. See examples in figure 2.

2.2.2 Split wide blobs

People standing close to each other, e.g., in a group, will often be found as one large blob. To identify which blobs contain more than one person, the height/width ratio and the perimeter are considered, as done in [17]. If the criteria are satisfied, the algorithm should try to split the blob. For

this type of occlusion, it is often possible to see the head of each person, and split the blob based on the head positions. Since the head is more narrow than the body, people can be separated by splitting vertically from the minimum points of the upper edge of a blob. These points can be found by analysing the convex hull and finding the convexity defects of the blob. See examples in figure 2.

Figure 2. Examples of wide and tall blobs that have been split.

2.3. Sorting people candidates

In addition to occlusions, other problems like reflections from people in the floor, or one person split into many blobs can be observed. This means that blobs can not always be mapped into individual people. In order to solve these challenges, the idea of generating a probabilistic occupancy map [16, 6] is adapted to find the probability that a person is observed at a given location. The original ideas were applied for multi-camera tracking, where it is possible to observe the 3D location of the scene. For this work, part of the idea is adapted to work on binary objects, captured from a single view. The algorithm will take all the bottom points of the blobs as person location candidates, and calculate the probability for each of them being a true position. A rectangle is generated from each candidate point, with a height corresponding to a given average height of people and the width being one third of the height. Two parameters are used for evaluating the probability of the rectangle containing a person: the ratio of white pixels inside the rectangle and the ratio of the rectangle perimeter that is white. Figure 3 shows two histograms of the ratio of white pixels inside the rectangles for true candidates (blue) and false candidates (red). The histograms are built from manual annotation of 340 positive samples and 250 negative samples.

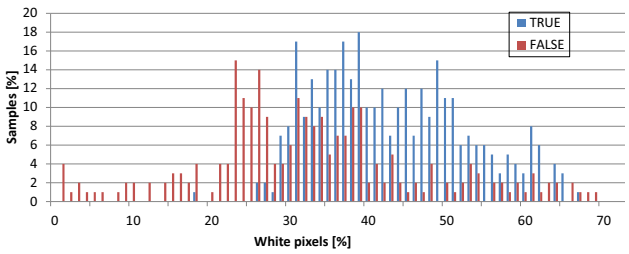

Figure 3. Histograms of the percentage of white pixels in each candidate rectangle. The blue histogram is for true candidates and red is for false candidates. No samples are found above 70 %.

From figure 3 it is seen that only 1 % of the true candidates have a white ratio less than 25 %, while a large part of the false candidates are found here, and no true candidates are above 70 %.
For the rectangle perimeter it is found that the lower the ratio of the rectangle perimeter that is white, the better is the fit of the rectangle to the person. The weighting of a person, $w_p(i)$, is described in equation 3 from the ratio of white pixels in the rectangle, r_r, and the ratio of white pixels on the perimeter, r_p:

$$w_p(i) = \begin{cases} 0, & \text{if} r_p > 50\% \parallel r_r < 20\% \\ 0.8, & \text{if} r_r > 70\% \\ 0.9, & \text{if} r_r < 30\% \parallel r_r > 60\% \\ 1, & \text{otherwise} \end{cases} \quad (3)$$

Candidates with $w_p(i) = 0$ are deleted.
There are still a lot of false candidates that will not be affected by these criteria. Many of them contain part of a person, and overlap in the image with a true candidate. Due to the possibility of several candidates belonging to the same person, the overlapping rectangles must be considered. By tests from different locations and different camera placements, it is found that if two rectangles overlap by more than 60 %, they probably originate from the same person, or from reflections of that person. As only one position should be accepted per person, only one of the overlapping rectangles should be chosen. Due to low resolution images compared to the scene depth, cluttered scenes, and no restrictions on the posture of a person, the feet of a person can not be recognised from the blobs. Furthermore, due to the possibility of reflections below a person in the image, it can not be assumed that the feet are the lowest point of the overlapping candidates. Instead, the best candidate will be selected on the highest ratio of white pixels, as it is seen from figure 3, that the probability of false candidates are lower here.

2.4. Identification of people entering and leaving

During the periods with activities detected at the border of the court, it is very likely that a change will happen. For these periods, the people near the border are monitored in order to detect crossings. The people are detected as described in section 2.3, but will not be counted during these unstable periods. Instead, the position of each person near the border is tracked, and if the border is crossed, it is registered along with the direction. Until a new stable period is observed, the number of people entering or leaving the court will contribute to the total transition in number.

3. Graph search optimisation

Two types of data exist now, the number of detected persons during the stable periods, and the number of entering

or leaving persons during periods with activity at the border. The last step is now to combine these estimates in a graph for the total observed period and estimate the most probable number during all periods. The graph will consist of nodes, representing the number of people in the stable periods and edges, representing the change in number between two periods. Figure 4 is a simple example of a graph with three stable periods. Edges exist between all nodes in two consecutive periods, but to simplify the illustration they are not drawn.

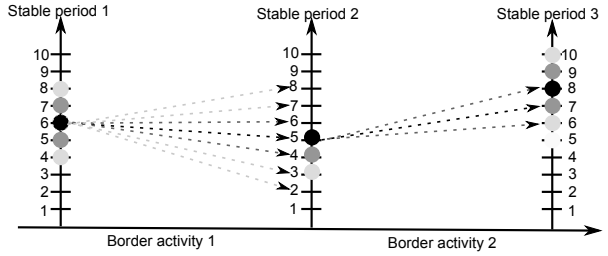

Figure 4. Example of a simple graph. Dark nodes and edges have the highest weight. Edges exist between all nodes in two consecutive periods, but to simplify the illustration they are not drawn.

A dynamic programming approach is taken to calculate the optimal path. The problem is solved by a version of Dijkstra's Algorithm modified to calculate the path with the highest votes, instead of the traditional minimum cost. The probabilistic weighting of nodes and edges will be described in the next section.

3.1. Weighting

Each node and edge must be weighted in order to calculate the best path. We define the weights as positive, meaning that a higher weight is a better path. As described, each node in the graph represents a possible number of persons in a given period. The weights for the nodes will be distributed according to the weighted histogram of the number of detected people in all frames during the stable period. The histogram is constructed from the detected people in each frame, with a weight describing the probability of each detection being true, and a weight describing the uncertainty of the frame, caused by occlusions and clutter. Each frame counting is weighted like this:

$$w_f = \alpha \cdot \prod_{i=1}^{n} w_p(i) + \beta \cdot w_s \tag{4}$$

where n is the number of people, $w_p(i)$ is the probability of people i being a true detection (see equation 3), and w_s is a weight that decreases with the number of splits performed (described in section 2.2), indicating how cluttered the scene is. α and β are the weighting of each part and should sum to one. The observed number in a frame will be added to the histogram with the weight w_f, and after a stable period has ended, the histogram will be scaled to an accumulated sum of 1. The circles in figure 4 illustrate the weighted histogram for each period.

The weighting of edges depends on the total number of crossings during the period of border activity, as well as the weighting of the individual people crossing the border. The probability of change x in number of people ($+n$ for people entering the court and $-n$ for people leaving) is modelled as a Gaussian distribution, with the mean value μ being the calculated number, and the variance σ proportional to the total number of crossings. The probability is described as $w_b(x)$:

$$w_b(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)} \times w_p \tag{5}$$

$$w_p = \frac{1}{m} \sum_{i=1}^{m} w_p(i) \tag{6}$$

where m is the number of people crossing the border. Each dashed line in figure 4 illustrates the edges weighted with $w_b(x)$. In the example the variance σ is high for the first period of border activity and low for the second period of border activity.

4. Experimental results

Comparing our results with others is difficult, because as far as we know, only [17] has focused on occupancy analysis of thermal video. We therefore compare our work to related work based on RGB cameras. Moreover, no public datasets with long thermal videos containing more than a few people exist. We therefore capture a new dataset, that will be available for download after publication [1]. The data contained in this video is from six different arenas, in order to be able test the robustness of the algorithms in different environments and set-ups. Several different activities are captured as well as both children and adults. We test on a 5-minute sequence from each of the five arenas for the evaluation of the detection algorithm and the tracking algorithm for the border areas. The full system with graph search optimisation should benefit from a longer video sequence, and will therefore be tested on a 30 minute video from a sixth arena. Thereby, the system has been tested on a total of 51,000 frames, which are manually annotated to provide ground truth. This data contains between 3-16 people in each frame. The processing time is approximately 0.125 seconds per frame on an Intel Core 2 Duo 3 GHz CPU, without any optimisation of the software.

To prove the generality of our framework, a final test has also been conducted on a public dataset of a totally different scenario, which is an outdoor scene from the OSU Color/Thermal database [11]. This test will be described in

[1] Available for download at www.vap.aau.dk

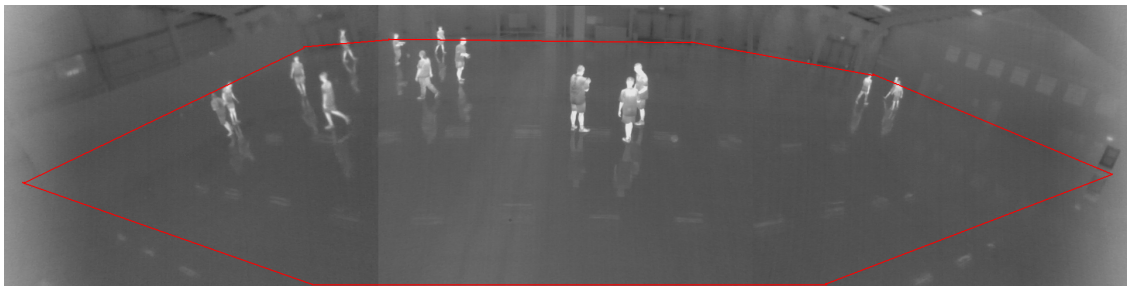

Figure 5. Example of the thermal image, with the outline of the court drawn as a red line.

section 4.5.
The remaining part of this section will describe the calibration and initialisation needed for the system, before results for each test are presented.

4.1. Camera calibration and initialisation

Installing a camera in the ceiling above most courts is very cumbersome and expensive and therefore not realistic in general. Therefore, it must be installed on one of the walls or rafters around the court. A standard arena has a court of 40 $\times$ 20 metres, corresponding to a handball field, indoor soccer field, etc. As the lenses of commercial thermal cameras today have a maximum field-of-view of around 60°, more than one camera must be used to cover the entire court. The camera set-up used in this work consists of three thermal cameras placed at the same location, and adjusted to have adjacent fields-of-view. Each camera is of the type Axis Q1922, which uses an uncooled microbolometer for detection. The resolution is 640 $\times$ 480 pixels per camera and the horizontal field-of-view is 57° per camera. To make the system invariant to the cameras' set-up, the images are stitched together before processing. This requires the cameras to be perfectly aligned and undistorted in order to secure smooth "crossings" between two cameras. Calibration of thermal cameras is not a trivial task, as they can not see the contrast differences of a typical chessboard used in most applications. Therefore, a special calibration board is made with 5x4 small incandescent light bulbs. With this it is possible to adapt the traditional method to estimate the intrinsic parameters of the cameras. The cameras are manually aligned horizontally so that their pitch and roll are the same. This mimics the well-known panorama effect, but with three cameras capturing at the same time. An example of the resulting image is shown in figure 5. When the cameras are put up in an arena, an initialisation is made. This consists of finding the mapping between image and world coordinates, as well as finding the correlation between peoples' real height and their height in the images, corresponding to their distance to the camera.
As the cameras are fixed relative to each other and then tilted downwards when recording in arenas, the result is that people in the image are more tilted the further they get from the image centre along the x-axis. This means that a person's pixel height can not always be measured vertically in the image. Therefore, the calibration must include both the height and the angle of a person standing upright at predefined positions on the court. For this work we used points on a grid of 5 $\times$ 5 metres on the court resulting in 45 different calibration images. In each image the world coordinates, image coordinates, pixel height and angle are learned as well as the person's real height in metres. The four corners are used to calculate a homography for each square, making it possible to map image coordinates to world coordinates. Using interpolation, an angle and maximum height are calculated for each position.

4.2. Detection of people

The first test evaluates the detection algorithm described in section 2.1. The number of detected people is registered as well as the manually counted number. This is done for 5 videos of 5 minutes each, captured with 10 fps, altogether 15,000 frames.
The mean error for each video is found to be between 8.5 % and 22.0 %. The errors are independent of the arena and seems primarily to depend on the level of occlusions seen in the scene. Periods with large groupings have a higher detection error than periods with people separated from each other. This is also expected, as the detection algorithm works on each frame independently, and people that are fully or mostly occluded can not be detected. Apart from the initialisation described in section 4.1, nothing has been done to fit the system to the specific arena, and it is concluded that it is independent of the arena.

4.3. Transition recognition

For the five videos of five minutes, it is registered each time a person crosses a specified border in order to evaluate the tracking algorithm. A total of 154 crossings are detected manually, and 168 crossings are detected automatically. 108 of the crossing are detected at the exact time, which is considered within $\pm$ 2 frames of the manual detection. Most of the false crossings detected are compensated with a cross-

ing in the opposite direction within a few frames. These will therefore not affect the global estimation of the number.

4.4. Full system test

The full system is tested on a 30 minute video, captured with 20 fps. Calculating the error for each frame gives an average error of 0.38 persons, corresponding to 4.44 %. For comparison, the result using detection only is also found, the error here is twice as high, 8.87 %. The number of detections is very unstable, and could suggest to do a simple low pass filtering, to overcome what looks like high frequency noise in the measurements. Low pass filtering the detection data reduces the error to 7.70 %. This indicates that a simple filtering of the data will not reduce the error as efficiently as the graph optimisation method. In table 1 our results are compared to related work, based on both thermal and RGB images.

	Reported error
Gade et al. [17]	7.35-11.76 %
Rabaud and Belongie [34] *	6.3-10 %
Hou and Pang [18] *	10 %
Celik et al [5] */**	8 - 14 %
Our method	**4.44** %

Table 1. Reported error percentage from related work compared to our result. * uses RGB images. ** calculates the error as percentage of frames with an error larger than one person.

4.5. Test on OSU dataset

To show the generality of our framework, we tested the system on the thermal video from the OSU Color-Thermal database [11], which is dataset three from the OTCBVS Benchmark Dataset Collection. We used sequences 4, 5 and 6, which are videos of approximately one minute each. They contain between zero and four people in each frame. Due to the low number of people in this dataset, instead of error we calculated the precision, being the number of frames with the correct number of people estimated. The results are presented in table 2 and compared to the results of detection alone, as well as the results of [25], which were provided with the dataset. However, it should be noted that the results of [25] are obtained by fusing the thermal and visible modalities and are intended for people tracking.

	Seq. 4	Seq. 5	Seq. 6
Detection only	86.72 %	83.11 %	77.72 %
Leykin et al. [25]	85.52 %	88.77 %	64.89 %
Our full method	**87.12** %	**93.70** %	**87.89** %

Table 2. Counting precision on the OSU dataset.

It is seen that the results of our full method are better than both the results from [25] and from detection alone.

5. Conclusion

In this work we have presented a unified framework for occupancy analysis. This method includes temporal information in the estimation by measuring the transition in numbers, and using that together with the detection of people in the global optimisation. The application of this work is the analysis of a given facility over days, weeks or even months. The need for real-time analysis is minor, and offline processing therefore allows for a more global approach. The main focus was on sports arenas, but we also proved that it works well in a general outdoor scene. We have shown that including the transition information improves the precision significantly, compared to using detection alone; even if the detection results are filtered afterwards. The mean error for the 30-minute test is 4.44 %, compared to 8.87 % if only the detection method was used.
The occupancy analysis is the foundation in many applications and can be continued to further activity analysis.

References

[1] R. M. Barros, R. P. Menezes, T. G. Russomanno, M. S. Misuta, B. C. Brandao, P. J. Figueroa, N. J. Leite, and S. K. Goldenstein. Measuring handball players trajectories using an automatically trained boosting algorithm. *Computer Methods in Biomechanics and Biomedical Engineering*, 14(1):53 – 63, 2011.

[2] M. Bertozzi, A. Broggi, C. Caraffi, M. D. Rose, M. Felisa, and G. Vezzoni. Pedestrian detection by means of far-infrared stereo vision. *CVIU*, 106(23):194 – 204, 2007.

[3] M. Bertozzi, A. Broggi, C. H. Gomez, R. I. Fedriga, G. Vezzoni, and M. Del Rose. Pedestrian detection in far infrared images based on the use of probabilistic templates. In *IEEE Intelligent Vehicles Symposium*, june 2007.

[4] E. Binelli, A. Broggi, A. Fascioli, S. Ghidoni, P. Grisleri, T. Graf, and M. Meinecke. A modular tracking system for far infrared pedestrian recognition. In *IEEE Intelligent Vehicles Symposium*, june 2005.

[5] H. Celik, A. Hanjalic, and E. Hendriks. Towards a robust solution to people counting. In *IEEE International Conference on Image Processing*, pages 2401 –2404, oct. 2006.

[6] Y. Cho, Y. Choi, S. Bae, S. Lim, and H. Yang. Multi-camera occupancy reasoning with a height probability map for efficient shape modeling. In *16th International Conference on Virtual Systems and Multimedia*, oct. 2010.

[7] C. Dai, Y. Zheng, and X. Li. Layered representation for pedestrian detection and tracking in infrared imagery. In *CVPR Workshops*, june 2005.

[8] C. Dai, Y. Zheng, and X. Li. Pedestrian detection and tracking in infrared imagery using shape and appearance. *CVIU*, 106(2-3):288–299, May 2007.

[9] J. W. Davis and M. A. Keck. A two-stage template approach to person detection in thermal imagery. In *Seventh IEEE Workshops on Application of Computer Vision*, 2005.

[10] J. W. Davis and V. Sharma. Robust detection of people in thermal imagery. In *ICPR*, 2004.

[11] J. W. Davis and V. Sharma. Background-subtraction using contour-based fusion of thermal and visible imagery. *CVIU*, 106(23):162 – 182, 2007.

[12] E. F. de Morais, S. Goldenstein, and A. Rocha. Automatic localization of indoor soccer players from multiple cameras. In *Proceedings of the International Conference on Computer Vision Theory and Applications*, feb. 2012.

[13] Y. Fang, K. Yamada, Y. Ninomiya, B. K. P. Horn, and I. Masaki. A shape-independent method for pedestrian detection with far-infrared images. *IEEE Transactions on Vehicular Technology*, 53(6):1679 – 1697, nov. 2004.

[14] B. Fardi, U. Schuenert, and G. Wanielik. Shape and motion-based pedestrian detection in infrared images: a multi sensor approach. In *IEEE Intelligent Vehicles Symposium*, 2005.

[15] A. Fernández-Caballero, J. C. Castillo, J. Serrano-Cuerda, and S. Maldonado-Bascón. Real-time human segmentation in infrared videos. *Expert Systems with Applications*, 38(3):2577 – 2584, 2011.

[16] F. Fleuret, J. Berclaz, R. Lengagne, and P. Fua. Multicamera people tracking with a probabilistic occupancy map. *PAMI*, 30(2):267 –282, feb. 2008.

[17] R. Gade, A. Jørgensen, and T. B. Moeslund. Occupancy analysis of sports arenas using thermal imaging. In *Proceedings of the International Conference on Computer Vision and Applications*, feb. 2012.

[18] Y.-L. Hou and G. Pang. People counting and human detection in a challenging situation. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 41(1):24 –33, jan. 2011.

[19] J.-E. Kallhammer, D. Eriksson, G. Granlund, M. Felsberg, A. Moe, B. Johansson, J. Wiklund, and P.-E. Forssen. Near Zone Pedestrian Detection using a Low-Resolution FIR Sensor. In *IEEE Intelligent Vehicles Symposium*, 2007.

[20] J. Kapur, P. Sahoo, and A. Wong. A new method for gray-level picture thresholding using the entropy of the histogram. *Computer Vision, Graphics, and Image Processing*, 29(3):273 – 285, 1985.

[21] R. Kitchin and M. Dodge. *Code/Space: Software and Everyday Life*. MIT Press, 2011.

[22] S. Kopf, B. Guthier, D. Farin, and J. Han. Analysis and retargeting of ball sports video. In *IEEE Workshop on Applications of Computer Vision*, jan. 2011.

[23] A. Leykin and R. Hammoud. Robust multi-pedestrian tracking in thermal-visible surveillance videos. In *CVPR Workshop*, 2006.

[24] A. Leykin and R. Hammoud. Pedestrian tracking by fusion of thermal-visible surveillance videos. *Machine Vision and Applications*, 21:587–595, 2010.

[25] A. Leykin, Y. Ran, and R. Hammoud. Thermal-visible video fusion for moving target tracking and pedestrian classification. In *CVPR*, 2007.

[26] W. Li, D. Zheng, T. Zhao, and M. Yang. An effective approach to pedestrian detection in thermal imagery. In *Eighth International Conference on Natural Computation*, 2012.

[27] Z. Li, J. Zhang, Q. Wu, and G. Geers. Feature enhancement using gradient salience on thermal image. In *International Conference on Digital Image Computing: Techniques and Applications*, 2010.

[28] M. Mahlisch, M. Oberlander, O. Lohlein, D. Gavrila, and W. Ritter. A multiple detector approach to low-resolution FIR pedestrian recognition. In *IEEE Intelligent Vehicles Symposium*, 2005.

[29] U. Meis, M. Oberlander, and W. Ritter. Reinforcing the reliability of pedestrian detection in far-infrared sensing. In *IEEE Intelligent Vehicles Symposium*, 2004.

[30] H. Nanda and L. Davis. Probabilistic template based pedestrian detection in infrared videos. In *IEEE Intelligent Vehicle Symposium*, 2002.

[31] D. Olmeda, A. de la Escalera, and J. Armingol. Contrast invariant features for human detection in far infrared images. In *IEEE Intelligent Vehicles Symposium*, 2012.

[32] D. Olmeda, A. de la Escalera, and J. M. Armingol. Detection and tracking of pedestrians in infrared images. In *Int'l Conference on Signals, Circuits and Systems*, 2009.

[33] E. Poulsen, H. Andersen, R. Gade, O. Jensen, and T. Moeslund. Using human motion intensity as input for urban design. In *Constructing Ambient Intelligence*, 2012.

[34] V. Rabaud and S. Belongie. Counting crowded moving objects. In *CVPR*, june 2006.

[35] R. Schweiger, S. Franz, O. Lohlein, W. Ritter, J.-E. Kallhammer, J. Franks, and T. Krekels. Sensor fusion to enable next generation low cost night vision systems. *Optical Sensing and Detection*, 7726(1), 2010.

[36] R. A. Serway and J. W. Jewett. *Physics for Scientists and Engineers with Modern Physics*. Brooks/Cole—Thomson Learning, sixth edition, 2004.

[37] F. Suard, A. Rakotomamonjy, A. Bensrhair, and A. Broggi. Pedestrian detection using infrared images and histograms of oriented gradients. In *IEEE Intelligent Vehicles Symposium*, 2006.

[38] W. Wang, J. Zhang, and C. Shen. Improved human detection and classification in thermal images. In *17th IEEE International Conference on Image Processing*, 2010.

[39] W. K. Wong, Z. Y. Chew, C. K. Loo, and W. S. Lim. An effective trespasser detection system using thermal camera. In *Second International Conference on Computer Research and Development*, 2010.

[40] J. Xing, H. Ai, L. Liu, and S. Lao. Multiple player tracking in sports video: A dual-mode two-way bayesian inference approach with progressive observation modeling. *IEEE Transactions on Image Processing*, 20(6):1652 –1667, june 2011.

[41] F. Xu, X. Liu, and K. Fujimura. Pedestrian detection and tracking with night vision. *IEEE Transactions on Intelligent Transportation Systems*, 6(1):63 – 71, march 2005.

[42] L. Zhang, B. Wu, and R. Nevatia. Pedestrian detection in infrared images based on local shape features. In *CVPR*, 2007.

[43] T. T. Zin, H. Takahashi, and H. Hama. Robust person detection using far infrared camera for image fusion. In *Second International Conference on Innovative Computing, Information and Control*, 2007.