Aalborg Universitet

AALBORG UNIVERSITY
DENMARK

Thermal Imaging Systems for Real-Time Applications in Smart Cities

Gade, Rikke; Moeslund, Thomas B.; Nielsen, Søren Zebitz; Skov-Petersen, Hans; Andersen, Hans Jørgen; Basselbjerg, Kent; Dam, Hans Thorhauge; Jensen, Ole B.; Jørgensen, Anders; Lahrmann, Harry; Madsen, Tanja Kidholm Osmann; Skouboe, Esben Bala; Povey, Bo Ø.

Published in:
International Journal of Computer Applications in Technology

DOI (link to publication from Publisher):
10.1504/IJCAT.2016.076790

Publication date:
2016

Document Version
Accepted author manuscript, peer reviewed version

Link to publication from Aalborg University

Citation for published version (APA):
Gade, R., Moeslund, T. B., Nielsen, S. Z., Skov-Petersen, H., Andersen, H. J., Basselbjerg, K., Dam, H. T., Jensen, O. B., Jørgensen, A., Lahrmann, H., Madsen, T. K. O., Skouboe, E. B., & Povey, B. Ø. (2016). Thermal Imaging Systems for Real-Time Applications in Smart Cities. *International Journal of Computer Applications in Technology*, *53*(4). https://doi.org/10.1504/IJCAT.2016.076790

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy
If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Thermal imaging systems for real-time applications in smart cities

Rikke Gade* and Thomas B. Moeslund

Visual Analysis of People Lab,
Aalborg University,
Aalborg, Denmark
Email: rg@create.aau.dk
Email: tbm@create.aau.dk
*Corresponding author

Søren Zebitz Nielsen and Hans Skov-Petersen

Department of Geosciences and Natural Resource Management,
University of Copenhagen,
København, Denmark
Email: szn@ign.ku.dk
Email: hsp@ign.ku.dk

Hans Jørgen Andersen

Department of Architecture, Design and Media Technology,
Aalborg University,
Aalborg, Denmark
Email: hja@create.aau.dk

Kent Basselbjerg and Hans Thorhauge Dam

Department of Electronic Systems,
Aalborg University,
Aalborg, Denmark
Email: kebakent@gmail.com
Email: hans.dam@gmail.com

Ole B. Jensen

Department of Architecture, Design and Media Technology,
Aalborg University,
Aalborg, Denmark
Email: obje@create.aau.dk

Anders Jørgensen

Visual Analysis of People Lab,
Aalborg University,
Aalborg, Denmark
Email: andjor@create.aau.dk

Harry Lahrmann and
Tanja Kidholm Osmann Madsen

Department of Development and Planning,
Aalborg University,
Aalborg, Denmark
Email: hsl@civil.aau.dk
Email: tkom@civil.aau.dk

Copyright © 2016 Inderscience Enterprises Ltd.

Esben Skouboe Bala

Department of Architecture, Design and Media Technology,
Aalborg University,
Aalborg, Denmark
Email: info@esbenskouboe.dk

Bo Ø. Povey

Department of Electronic Systems,
Aalborg University,
Aalborg, Denmark
Email: bopovey@gmail.com

Abstract: In the modern world, cities need to keep up with the demand for mobility, efficient infrastructure and environmental sustainability. The future smart cities use intelligent information and communication technologies to raise the quality of life. This includes computer vision as one of the main technologies. It can observe and analyse human activities from a distance in a non-invasive manner. Traditional computer vision utilises RGB cameras, but problems with this sensor include its light dependency, and the privacy issues that can be raised by people being observed. In this paper, we propose the use of thermal imaging in real-time smart city applications. Thermal cameras operate independently of light and measure the radiated infrared waves representing the temperature of the scene. In order to showcase the possibilities, we present five different applications which use thermal imaging only. These include both indoor and outdoor scenarios with the purposes of people detection, counting and tracking, as well as one application for traffic safety evaluation.

Keywords: thermal imaging; people tracking; people counting; real-time systems.

Reference to this paper should be made as follows: Gade, R., Moeslund, T.B., Nielsen, S.Z., Skov-Petersen, H., Andersen, H.J., Basselbjerg, K., Dam, H.T., Jensen, O.B., Jørgensen, A., Lahrmann, H., Madsen, T.K.O., Bala, E.S. and Povey, B.Ø. (2016) 'Thermal imaging systems for real-time applications in smart cities', *Int. J. Computer Applications in Technology*, Vol. 53, No. 4, pp.291–308.

Biographical notes: Rikke Gade is PhD student at Aalborg University with the Visual Analysis of People Lab.

Thomas B. Moeslund is Professor at Aalborg University and Head of the Visual Analysis of People Lab.

Søren Zebitz Nielsen is PhD student at Department of Geosciences and Natural Resource Management, University of Copenhagen, Denmark.

Hans Skov-Petersen is senior researcher at Department of Geosciences and Natural Resource Management, University of Copenhagen, Denmark.

Hans Jørgen Andersen is Head of Department of Architecture, Design and Media Technology at Aalborg University, Denmark.

Kent Basselbjerg was student of Electronics Engineering at Aalborg University, Denmark.

Hans Thorhauge Dam was student of Electronics Engineering at Aalborg University, Denmark.

Ole B. Jensen is Professor of Urban Theory at Aalborg University and Director of the Research Cluster for Mobility and Tracking Technologies (MoTT).

Anders Jørgensen is laboratory engineer at Aalborg University with the Visual Analysis of People Lab.

Harry Larhmann is Associate Professor at the Department of Development and Planning, Aalborg University.

Tanja Kidholm Osmann Madsen is research assistant at the Department of Development and Planning, Aalborg University.

Esben Skouboe Bala was PhD student at Department of Architecture, Design and Media Technology, Aalborg University.

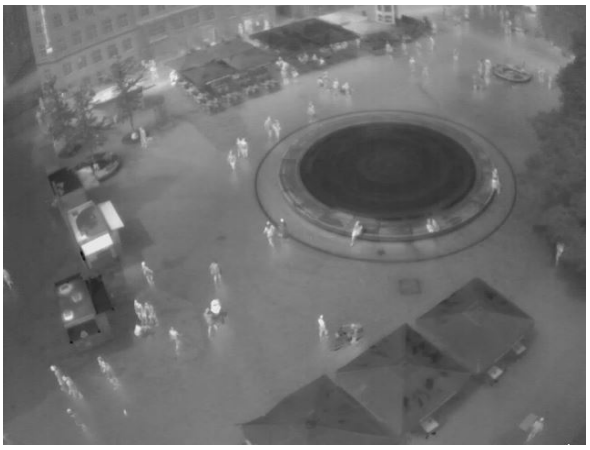
Bo Ø. Povey was student at Aalborg University, Electronics Engineering.

1 Introduction

Contemporary urbanisation on a global scale has led to a number of challenges in terms of environmental stress, continuing migration, demographic shifts and urban liveability. In the literature on the application of digital technologies to these vast urban challenges, the notion of a 'Smart City' is frequently seen (Batty et al., 2012; Copenhagen Cleantech Cluster, 2012). If one sets aside the policy and branding value of an overarching, but rather simplifying term, the notion of a 'Smart City' carries the potential for engaging with these urban challenges. The important feature to notice is that the contemporary city no longer can be comprehended as a bounded and mono-nuclear unit; rather, cities are connected (or dis-connected) in globally reaching networks that create new dynamics of mobility, regardless of whether these are of the nature of moving goods, ecological flows or in-migrating people (Jensen, 2013a). But next to these vast new material transformations, the 'network city' has also become over-layered with new types of digital technology, leading some scholars to speak of a 'sentient city' (Shepard, 2011) or a 'programmable city' (Kitchin, 2011). The essence of this development is that networked technologies are now able to track and detect different types of movement (container, information, cars, or people) to such an extent that we may speak of a new 'digital layer' to the city (Jensen, 2013a). Obviously, many scenarios may play out from this. Some would be focusing on the issue of surveillance and control, whereas others would put focus on the potential for these new technologies to optimise the flow of people, goods and information, as well as to engage citizens more in their cities and to create better environmental solutions. However, in order to explore the potential of these new networked technologies, and thus apply them in 'smart' ways, basic research needs to be undertaken. One must explore what the new networked technologies may afford in solutions to the grand challenges facing the contemporary cities globally. As we are beginning to see cities where real-time location aware data is being created on a continuous basis, we may start to ask how such data can inform better systems of traffic planning, city governance, energy provision and other types of urban utilities (Jensen, 2013b). But moreover, the new 'digital layer' opens up a discussion of whether it is possible to involve citizens more in the social life of public spaces, if the new technologies hold the potential to create new aesthetic experiences 'on top' of the existing ones; and if real-time data fed back to city inhabitants will create a whole new experience of the city. These issues are relevant dimensions of a new research agenda on the 'Smart City', and in this paper we will be focusing on one particular combination of technologies that we think need to be researched for their undiscovered potentials.

Computer vision technologies have great potential in smart city applications, owing to the non-intrusive nature of the sensor. With cameras, it is possible to record data at a distance and at real-time. There are, however, some problems with regular RGB cameras. First of all, the deterrent effects of surveillance and control caused by cameras are high, as it is possible to visually recognise and follow people. (In some surveillance situations, discovery of a person's identity is the ultimate goal; but in general, data acquisition is preferred to be anonymous.) Moreover, robust systems can be very hard to develop, as cameras capture reflected lighting. Thus, they depend on sufficient light in order to capture anything, and the visual perception of things changes with the lighting. In the context of smart cities, these problems are critical, as 24 hour operation is assumed in many applications. This is not always possible to achieve with RGB sensors, while the natural lighting changes with the weather and the nights are dark. An alternative sensor, which is still non-intrusive, but also independent of lighting, is the thermal camera; and with this sensor, privacy issues are also eliminated. Thermal cameras capture the long-wavelength infrared radiation, which is radiated from all objects with a temperature above absolute zero. The amount of radiation depicts the temperature of the object, resulting in an image that visualises the temperature of the scene. People and objects with a temperature different from the surroundings can therefore be detected both day and night. Figure 1 shows an example of a thermal image from an urban square.

Figure 1 Thermal image overlooking an urban square



We have in recent years done a large amount of research in thermal imaging for different applications related to the Smart City. The purpose of this paper is firstly, to introduce the thermal sensor and its potential to the computer vision research community. We will compare and discuss the use of thermal and RGB sensors for real-time people detection and tracking. Furthermore, we will showcase five successful use-cases where we have applied computer vision to thermal imagery in the context of Smart Cities.

The following section will discuss thermal technology and compare RGB and thermal cameras. After that, we present five different smart city applications where real-time thermal imaging is applied.

2 Thermal sensors

Originally developed for military purposes, thermal cameras have now reached a reasonable price and have become available for commercial use. The technology has therefore started to be deployed for surveillance purposes, where RGB cameras would typically have been used.

The resolution of thermal sensors is slowly increasing as the technology evolves and new materials are explored. Available today are sensors ranging from cheap 8×8 pixel arrays (Panasonic Electric Works, 2013), up to 1280×1024 pixel sensors (FLIR Systems, 2013a). The field-of-view ranges from very narrow (approx. 1°), to wide angle lenses up to 80° (FLIR Systems, 2013b). Some types of cameras have the possibility for optical zoom. The wide angle lenses are very useful in applications like surveillance of indoor rooms and urban spaces. However, the selection of wide angle lenses is small and the price is very high, owing to the expensive lens material germanium. With very narrow field-of-view lenses, it is possible to detect people several kilometres away. This is very useful, especially for border and coastal surveillance.

For commercial use, the best known types of thermal cameras are the handheld cameras for building inspections. However, thermal surveillance cameras are becoming very popular, owing to their independence of lighting. These cameras are very similar to regular RGB surveillance cameras in terms of size, look and interfaces (AXIS Communications, 2013b). RGB cameras are still significantly cheaper than thermal cameras, though.

Thermal network cameras with data communication over IP can be part of larger camera networks, and they enable easy data transfer to a computer (AXIS Communications, 2013a). Many cameras come with built-in memory or a slot for a memory card, but external computers, and thereby storage, can also be connected to the cameras. Some cameras have small built-in processing units, where simple image processing algorithms can be programmed; e.g., motion detection and cross line detection. Real-time online processing of a video with more demanding algorithms is possible when a computer is connected.

2.1 Segmentation of people

Our main purpose of using computer vision in Smart Cities is to automatically, and in real-time, detect and track the movements of humans (Al-Mutib et al., 2014; Andriluka et al., 2008). Therefore, we will focus on the possibility of people tracking and detection in each of the image modalities and start with a general overview of detection methods.

In almost any application of computer vision figure-ground segmentation is needed to locate the desired object(s) in the images. Mainly, two approaches are widely used for detecting humans: pixel-based detection and object-based detection (Moeslund et al., 2011). Pixel-based detection methods consider each pixel individually, e.g. by comparing to a background model. The basic idea is to compare each pixel to a background model, and if the difference exceeds a given threshold, the pixel is classified as foreground. However, especially in outdoor scenes, obtaining a valid background model is challenged by the shifting sun, clouds, moving branches of trees, etc. Different ways of modelling a changing background, as well as updating it appropriately, is still being researched. In applications with a moving camera, it might be impossible to model the background, though.

Image thresholding is applied after background subtraction, but can also be applied to original images of different modalities. Depending on the application, thresholding methods vary from the simplest constant threshold value over dynamic and automatic methods (Kapur et al., 1985) to bio-inspired algorithms (Ouadfel and Meshoul, 2014).

Object-based methods detect either entire objects or major parts in case of a part-based model. Often these methods run by translating a window over the image and calculate the likelihood of each window containing a human (Moeslund et al., 2011). The Histogram of Oriented Gradients (HOG) detector (Dalal and Triggs, 2005) is one example of a very popular object-based detector, searching for a learned object shape.

The pixel-based methods detect everything that is or has been moving. This approach is fast, but it will often require post processing to filter noise and unwanted objects from the detections.

Object-based methods on the other hand are designed to detect a specific type of object and normally need no post processing. However, the output is usually bounding box locations, compared to the more precise silhouettes produced by pixel-based methods. The detection rates for HOG degrade rapidly when the resolution decreases (people less than 80 pixels tall), or if people are partly occluded (Dollar et al., 2012). Furthermore, object-based methods are generally computationally expensive, and often require GPU-based implementations in order to perform in real-time.

These limitations of object-based methods affect its use in Smart City applications, since the appearance of people may vary a lot. People are often observed from a far view, resulting in low resolution and different viewing angles of the people. On the other hand, the camera is most often static, reducing the problems of modelling a background for pixel-based methods significantly.

For the applications at hand, with focus on robust real-time performance, pixel-based detection methods seem best

suited for detection. In the following section we will test and compare detection in RGB and thermal images.

2.2 Comparison of thermal vs. RGB cameras

For any kind of detection, thermal cameras have clear advantages compared to RGB cameras in dark conditions, due to the independence of lighting. Figure 2 compares the two modalities from a night scene. Even though parts of the scene are illuminated, the person is very hard to detect in the RGB image. In the thermal image, the person is fully visible.

Owing to the properties of thermal imaging, detection of people can be easy and fast in the situations where human temperature differs from the surroundings. Here, a thresholding of the images can be sufficient for segmentation. Figure 3 shows an example where an automatic threshold method based on Maximum Entropy (Kapur et al., 1985) is applied to a thermal image.

Figure 2 Example of a night scene



(a) RGB (b) Thermal

Figure 3 Segmentation of people using automatic thresholding



(a) Thermal input (b) Automatic thresholding

Figure 4 An urban scene captured with an RGB and a thermal camera. The cameras have slightly different views. The images are captured on an early summer day with temperatures around 15°C



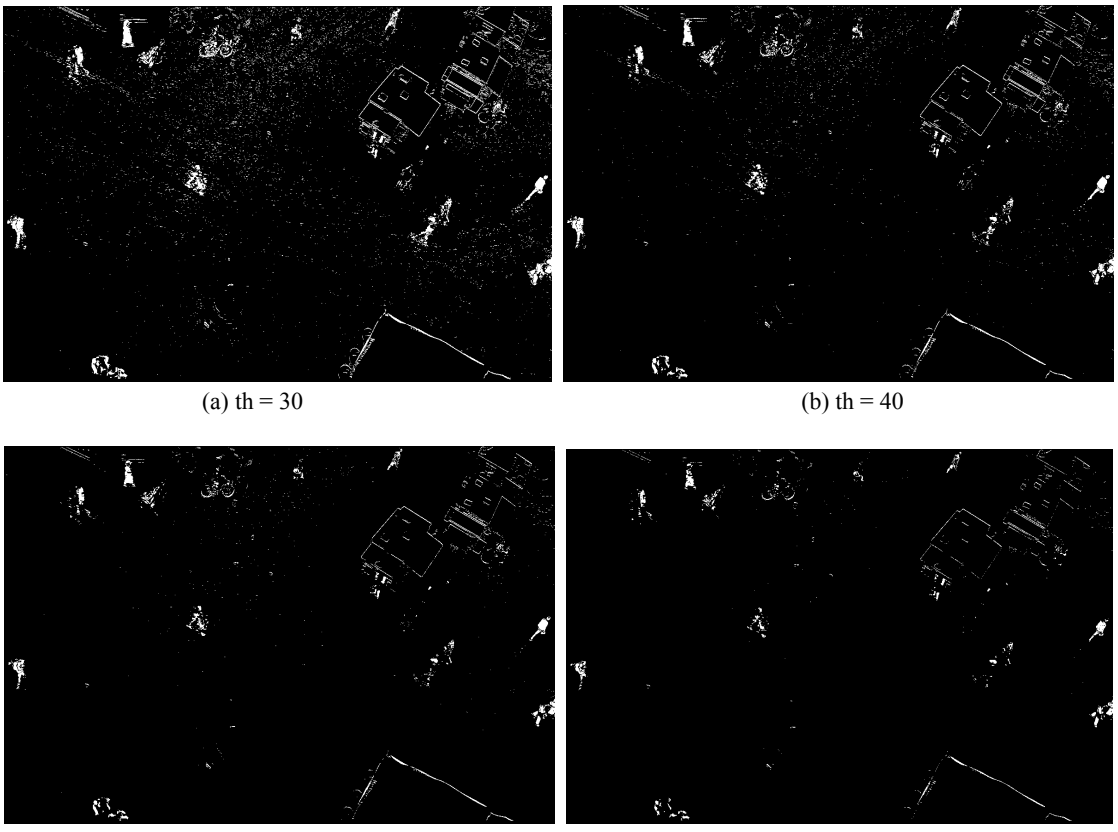
(a) RGB (b) Thermal

A similar fast approach which can be applied to both RGB and thermal images is background subtraction, which we described in Section 2.1. This approach assumes that only people, or other objects of interest, are moving. Otherwise filtering of the detections must be applied as post processing.

Figure 4 shows an example of an urban outdoor space captured simultaneously by an RGB and thermal camera (with slightly different views). Figures 5 and 6 show the results of background subtraction and binarisation with different threshold values. The threshold values are here chosen manually to illustrate the best possible segmentation and to show the effect of changing the values. In a real application an automatic thresholding method could be applied for choosing the best threshold value in any given frame.

Dark colours of clothes and the relatively small size of people make it very hard to detect people in the RGB image. Furthermore, the ground is partly wet, making it even harder to distinguish people. Using background subtraction, some people can be detected, but as Figure 5 shows, it is a trade-off between too much noise and missed detections. Using the thermal image, less noise is detected, even with a much lower threshold value of 5. Thus, as seen in Figure 6, the trade-off is less distinct, and people can more reliably be segmented with background subtraction using a thermal image.

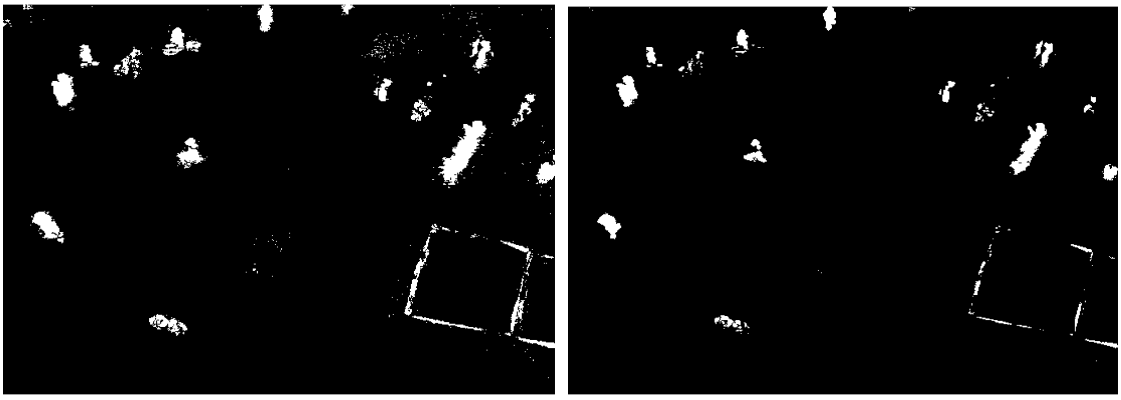
Figure 5 Background subtraction of the image in Figure 4(a), then binarisation with threshold values from 30 to 60. The edges of the carts in the upper right corner are detected due to small vibrations of the camera



(a) th = 30 (b) th = 40

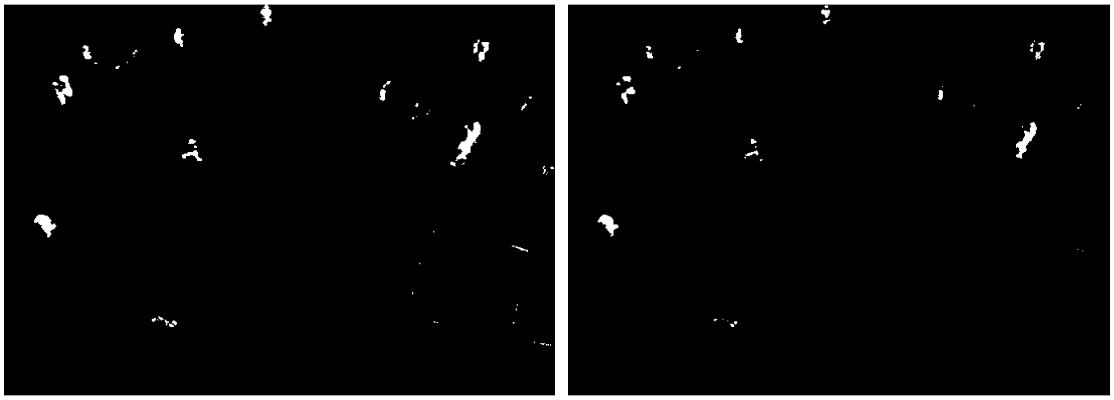
(c) th = 50 (d) th = 60

Figure 6 Background subtraction of the image in Figure 4(b), then binarisation with threshold values from 5 to 30



(a) th = 5 (b) th = 10

Figure 6 Background subtraction of the image in Figure 4(b), then binarisation with threshold values from 5 to 30 (continued)



(c) th = 20 (d) th = 30

Segmentation using background subtraction assumes that it is possible to obtain a reliable background model. Fast changing lighting/temperature conditions can cause problems, and a dynamic background model must be adjusted and updated during run-time (Shoushtarian and Bez, 2005). Thermal images have only one channel, compared to three channels of an RGB image. Furthermore, the temperature often changes more slowly than the lighting. This can make it easier to model the background in the thermal domain. Another well-known problem in the RGB domain is the occurrence of shadows. Shadows often cause false detections, as the shadows move just like people. Shadows are related to lighting, and thereby do not exist in the thermal domain. However, thermal radiation can be reflected in glossy surfaces and cause false detections that are similar to shadows. This is considered a rarer problem in Smart City applications though, as the surfaces are often not reflective.

After detection, some Smart City applications require the object to be tracked. Tracking humans is a complex problem, owing to the dynamically changing motion, and often occlusions must be resolved. In these situations, it can be necessary to re-identify people after an occlusion or an ambiguous situation. Recognising individual people is difficult in thermal images, owing to the lack of colour and texture information. Thereby, re-identification of people in thermal images is very difficult, making it very hard to solve these ambiguous situations in tracking. Thus, RGB cameras have advantages in complex tracking scenarios.

In Table 1, the pros and cons of using RGB and thermal cameras for people detection and tracking are summarised.

Table 1 Comparison of RGB and thermal cameras for people tracking and detection

	Pros	*Cons*
RGB	Cheap sensors Re-identification possible	Sensitive to light Privacy issues Shadows
Thermal	Easier segmentation Independent of light No privacy issues Single channel images	Re-identification difficult More expensive Reflections

3 Application of real-time thermal imaging

In the remaining part of this paper, we present five different Smart City applications. The purposes are all related to real-time monitoring of the city, in order to optimise mobility, accessibility and safety for the citizens. In all applications, we use one or more uncooled thermal cameras with a resolution of 384 × 288 pixels (AXIS Q1921) or 640 × 480 pixels (AXIS Q1922). Table 2 summarises the applications in terms of purpose, equipment, test time, type of scene, technique used and processing time. The test time stated here is the full run time of the set-up, the manually annotated period for quantitative evaluation may be shorter.

Table 2 Summary of the applications presented in this paper

Sec.	*Purpose*	*Camera(s)*	*Test time*	*Scene*	*Technique*	*Proc. time*
4	People counting	1 AXIS Q1921 10 mm	1 week	Outdoor pedestrians	Image differencing	1.2 ms/frame
5	People tracking	3 AXIS Q1921 19 mm	1 week	Outdoor pedestrians	Background subtraction and Kalman filtering	66 ms/frame
6	Car and bicycle detection	1 AXIS Q1922 10 mm	20 days	Outdoor traffic	Optical flow	41 ms/frame
7	People counting and activity recognition	3 AXIS Q1922 10 mm	1 month	Indoor sports	Background subtraction and Fisher faces	12.5-60 ms/frame
8	People tracking	1 AXIS Q1922 10 mm	2 hours	Outdoor pedestrians	Background subtraction and Kalman filtering	20 ms/frame

4 People counting in urban environments

In a Smart City, human movement is automatically registered and analysed. For both real-time and long-term perspectives, this knowledge can be beneficial in relation to urban planning and for shopkeepers in the city. Information in real-time can be used for analysing the current flow and occupancy of the city, while long-term analysis can reveal trends and patterns related to specific days, time or events in the city. In this work, we continuously counted people passing through a pedestrian zone in a city environment during one week. The location is illustrated in Figure 7, along with a picture of the camera's view.

4.1 Methods

When counting people in a pedestrian zone or a sidewalk, it is assumed that most people are moving across the scene. But being an outdoor scene, it cannot be assumed that people are constantly hotter or colder than the background. The surrounding temperature will change throughout the day, and the sun can heat dark pavement to temperatures hotter than the human body temperature. Therefore in this work, it has been chosen to do segmentation based on image differencing. The activities are then counted solely on the pixel response as opposed to connected silhouettes of people. Double differencing will be used in order to eliminate noise. Figure 8 explains the algorithm.

Figure 7 Illustration of the location and camera view



Figure 8 Double differencing algorithm

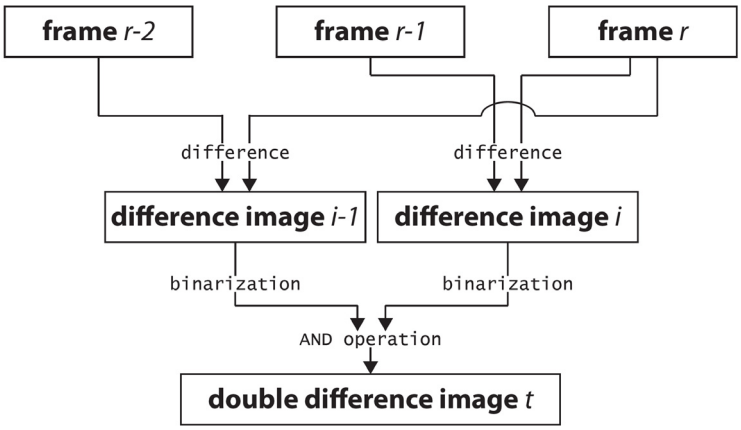
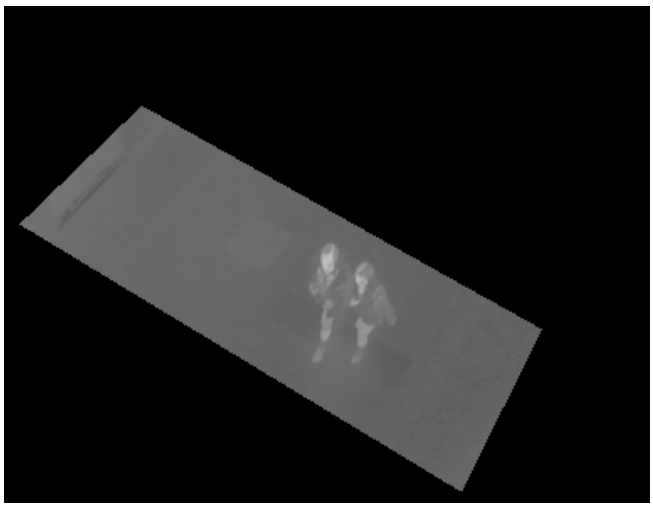


Figure 9 Left: Thermal input image, right: Result of double differencing and thresholding



A threshold value of 2 is applied to binarise the image. Figure 9 shows an example of an input frame, and the result of double differencing with the two previous frames.

Figure 10 The chosen region of interest



The activity, here represented by white pixels, must be converted into a number of people. From training data, the relation between the amount of activity and the number of people can be calculated. This factor depends on the specific set-up; camera specifications and mounting location. For each new location, a training phase is needed in order to learn this factor. However, it is tested to generalise appropriately over different days and activity levels at the same location. When the velocity of people, and thereby the activity in the frame, is lower, the person will instead stay longer within the region of interest, so that the accumulated activity over a time period will be equal to that of a person moving fast through the scene. The number of people should therefore be estimated for time windows, rather than momentarily. Choosing a smaller region of interest (ROI) in the image, as illustrated in Figure 10, will improve the results by reducing the perspective effects in the image. The region is chosen to represent a section of the street as close to the camera as possible. The top-view of the scene in combination with the chosen region of interest implies that occlusions can be neglected.

The ROI is applied by using an AND-operation between a binary image representing the ROI and the current frame.

4.2 Results

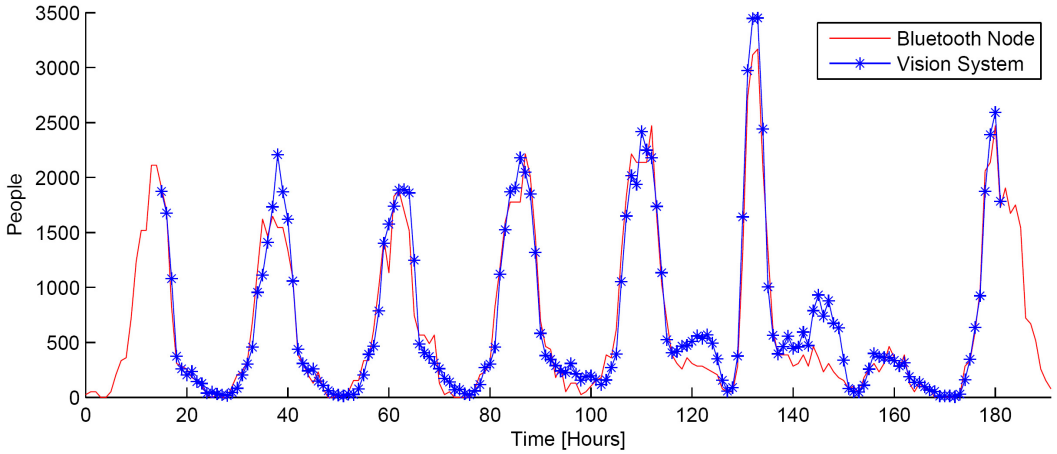
One full week of video has been captured. Owing to the high time consumption of manual video annotation, 13 video sequences of 20 minutes each have been chosen for tests. These video sequences cover different days, times of the day and the level of activity in the video, and they have been manually annotated. Using a leave-one-out cross-validation, with 12 video sequences used for training and one for testing in each iteration, the mean accuracy is 90.75%.

The full week of video has been processed and compared to the results of a commercial system (BLIP Systems, 2013). This system uses observed Bluetooth IDs to estimate the number of people passing a node. This does, however, only represent the percentage of people carrying a device with an open Bluetooth connection. We scale the Bluetooth results with a factor of 25.76, in order to compare the results as shown in Figure 11. The factor of 25.76 has been found experimentally by dividing the mean of our person detections with the mean of detected Bluetooth devices.

None of these results has been manually verified, making it unclear which is the most precise, but it is clear that they follow the same trends throughout the week. Two times, just after 120 hours and 140 hours, our vision system detects more people than the Bluetooth system. This is caused by police cars driving in the pedestrian zone on Friday and Saturday nights. This is not accounted for yet in the vision system. The advantage of a vision system in this application is that it counts the actual number of people, compared to counting a number of, e.g. Bluetooth devices, where the relation to the number of people is unknown.

The processing takes 1.2 ms per frame (384×288 pixels), which easily obeys real-time requirements. The test has been performed with a multi-threaded implementation of the system on a laptop with a 2.00GHz Intel Core i7-2630QM CPU and 8GB RAM.

Figure 11 Our system (blue graph) compared to a commercial system based on Bluetooth devices (red graph). The Bluetooth results have been multiplied with a factor 25.76



5 Interactive urban lighting

In the coming years, it is expected that the type of urban illumination will change to light emitting diodes (LED) to a greater extent. In addition to being less energy demanding, this type of light source enables digital control of both colour and intensity. That opens up new possibilities for designing interactive and adaptive lighting. Interactive environments can engage people to feel more connected to the urban spaces and encourage them to stay or to play in the environment. Intelligent lighting can also make the environment appear more comfortable and secure. In this full scale experiment, we put up three thermal cameras and 16 lamps in an urban space. A picture of the space is shown in Figure 12.

Figure 12 Overview of part of the observed urban space



Four different light scenarios were tested:

1 *Ambient Illumination*: Static white lighting.

2 *Glowing Light*: Each lamp individually fades up and down between 0% and 20% intensity.

3 *White Aura*: Illuminated circle with a diameter of min (10 metre around each person).

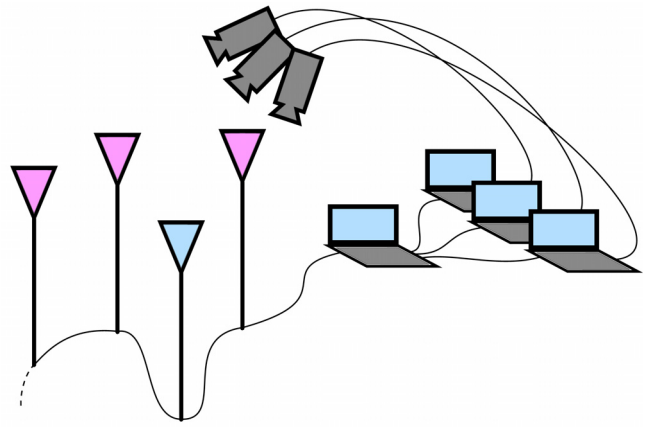
4 *Red Treasure Hunt*: When a person approaches a 'trigger light', a wave of light is sent out through the square.

The first two scenarios do not depend on the movement of people, while the last two are interactive scenarios, where thermal cameras are applied to estimate people's movements (Poulsen et al., 2012).

5.1 Hardware setup

Figure 13 illustrates the hardware of the system. Three thermal cameras are each connected to a computer, which processes the video and converts tracks of each person at the square into a common world coordinate system. This tracking data is sent to a computer that controls the lamps based on the live data and a chosen light scenario. A total of 16 RGB LED lamps are controlled by the system.

Figure 13 System overview (see online version for colours)

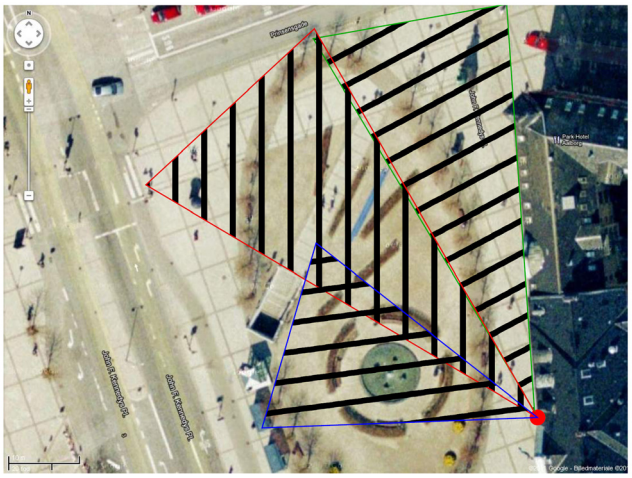


Inside each lamp is installed a DMX module, which controls each colour of the LED lamp in 255 brightness steps, making it possible to control both colour and brightness of each lamp individually. The area covered by each camera is illustrated in Figure 14.

The real-time aspects of this project are very important, because the lighting must react to people's concurrent movements and have an update rate that allows for smooth

control. Communications between both computers and lamps are handled using the Open Sound Control protocol (CNMAT, 2013).

Figure 14 Overview of the square with camera views illustrated



5.2 *Methods*

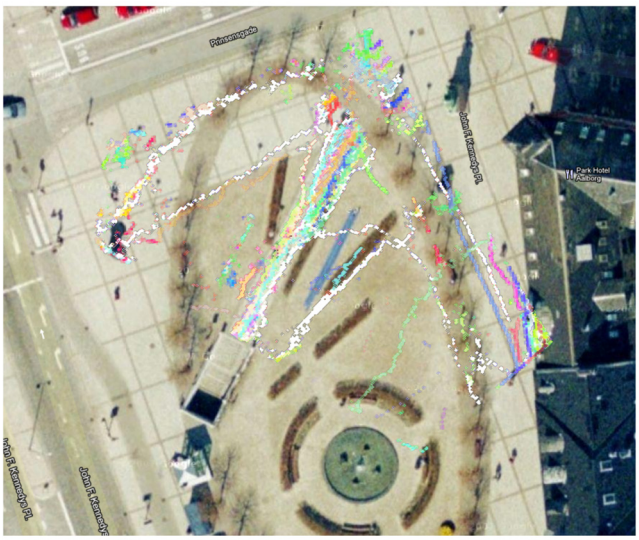
Input to the interactive lighting system is the real-time human movement. Since the observed area is an urban space with restricted car access, it is assumed that all observed activity is human activity of interest. People are detected by performing background subtraction, and then the objects are filtered by size. Since the background temperature naturally changes over time, the background must be updated. It is chosen to perform a running average background subtraction with a selective update, meaning that only pixels segmented as background will contribute to the new background. The detected image coordinates of people are transformed into world coordinates using a homography matrix calculated during an initialisation. Further description of the methods can be found in Poulsen et al. (2012). People walking closer than three metres to each other are treated as groups in this work. Each person or group object is tracked in world coordinates with a Kalman filter (Welch and Bishop, 1995; Ali and Terada, 2012). Treating groups of people as one object to track also helps to overcome the problems of occlusions between people, as it is not necessary to distinguish single persons.

The videos from each camera are processed individually, but after transforming the positions, all the tracking and lamp positions refer to the same world coordinate system. The map with overlayed tracking results for a 5-minute period is shown in Figure 15.

For each frame, the position and velocity vector for each person is registered from the tracking and sent to the light control system. In addition, merging and splitting of groups are registered as special events that could trigger light events.

The reactions from people are analysed both through interviews and from the recorded occupancy and movements.

Figure 15 Overview of the square with tracking results overlayed. Each person is assigned one colour



5.3 *Results*

The system has been tested for one week during the winter, when very cold weather could possibly affect the behaviour of people. Observations and interviews showed that people used the space mostly for transit, and often they did not notice the changed illumination. But people observing the square from outside noticed and appreciated the interactive behaviour (Poulsen et al., 2012). However, later experiments have shown that the lighting causes a range of abnormal behaviour, such as cats walking, dancing and light chasing.

The vision system is evaluated qualitatively by looking at the positions of visualised tracks on a digital map compared to the real position of the person, as well as the actuated lighting scenario compared to the expected scenario. With a precision of approximately one metre for the mapped position, the system works as expected, with correct and fast feedback to the lamps.

The full system runs real-time with a frame rate of 15 fps on 384 × 288 pixel images with a CPU implementation on a Intel Core i5-2430M 2.4GHz CPU. The real-time performance is crucial here, as the application does not allow delays in the feedback to the lamps. Testing the system for one full week without interruptions proves that the real-time performance is stable during changing conditions. For higher frame rate or larger images the performance could easily be improved using a faster CPU or possibly a multi-threaded implementation.

Figure 16 shows two frames from the 'White Aura' scenario, where an illuminated circle follows a person.

This study also showed that it is possible to save up to 90% of the energy for lighting, without people changing their behaviour (Poulsen et al., 2012).

The tracking information may be used for instant control of the lighting as described here, and it could also be used for later evaluation and design of urban spaces.

Figure 16 Two frames from the 'White Aura' scenario



6 Automatic near-collision detection

In urban environments, the high density of people leads to heavy traffic. Many people tend to use their private cars for comfortable transportation, but in order to reduce emission and congestion, a transition from the car to more eco-friendly means of transportation, such as the bicycle, is needed. To do this, an enhancement of the bicycle conditions must be done in order to increase the share of cyclists. Cities, such as Copenhagen (Nielsen et al., 2013), have put special focus on the cyclists, e.g., by designing special bike lanes on all the roads. However, studies have shown that bike lanes do not improve the safety of cyclists; even though there is a reduction in the number of crashes on road sections, where bike lanes are established, the number of crashes increases in the intersections; in particular at signalised intersections (Agerholm et al., 2006; Jensen, 2006).

In this work, thermal cameras are used to compare different types of intersections in terms of safety. The objective is to evaluate if some geometric designs of bike lanes through intersections are better than others. Since accidents are rare, the evaluation is based on the Swedish Traffic Conflict Technique (Hyden, 1987). The idea is that accidents and near-collisions are related. Thus, the near-collisions will be used as a measure of safety, since they occur more frequently than accidents. The detection of near-collisions can be done manually, but is of course very time consuming and ineffective owing to the low number of near-collisions. In the smart city, this can be done automatically using a camera system.

6.1 Methods

In this work, we will look for close interactions between cars and bikes, which are prerequisites for near-collisions to happen. Interactions are defined as the simultaneous motion of a car and a bicycle in a given zone of interest. Shown in Figure 17 is the car zone in blue and the bike zone in red. When both cars and bikes are present at the same time in their respective zones, there is a risk that the cyclists going straight in the red zone could be hit by a car from lane 1 or lane 2 turning towards lane 3. In both situations, the cars will pass the blue zone with a direction towards lane 3. An example image from one intersection is shown in Figure 18. Since both position of the camera and the layout of intersections and bike lanes changes between each scene captured, the car and bike zones must be defined manually for each intersection.

Figure 17 Illustration of an intersection with marked zones of interest. Blue is the car zone, and red is the bike zone. Blue arrows indicate cars' paths of interest, and the red arrow indicates bikes' path of interest

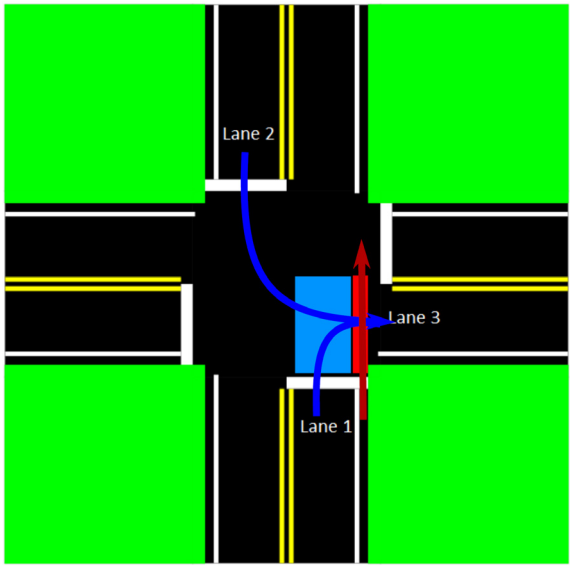


Figure 18 Example of an interaction between a car and a bicycle



In order to detect interactions, optical flow (Horn and Schunck, 1981) is applied to calculate direction and magnitude of the motion. For cyclists, the direction must correspond to going straight, and for cars, they must head towards the cyclist zone. An angle histogram based on the optical flow vectors will be used to decide whether the direction is correct. As shown in Figure 18 the camera angle is chosen in order to have a straight view on both bike and car lanes, but occlusion between cyclists can occur as they happen to drive close to each other. However, the optical flow method does not rely on detection of individuals, rather the overall motion in the chosen zone. The magnitude of motion must exceed a specified threshold in order to eliminate noise. This threshold depends on the camera location and angle, but due to a relatively small detection area, the effects from image perspective within each image are neglected.

The motion must be consistent for a short interval of time in order to be considered a real detection. Therefore, we implement a buffer that flags the frames in which either a cyclist or car's motion is observed. In order to register an interaction, both buffers must be flagged for a number of consecutive frames.

6.2 Results

One hour of video from each of the four intersections is used for testing the system. The videos are captured from 7am to 8am, corresponding to the morning rush hour. All interaction situations are manually labelled and compared to the output of the automatic system. The results show that we get approx. 33% false positives. This is of course a high rate, but as it is more critical to miss any detections, we allow more false positives. A manual verification process of the detections can be conducted afterwards. A few false negatives are observed. These situations occur when either a car or cyclist has a very low speed, resulting in a motion vector below the threshold.

The speed of the algorithm obeys real-time requirements with a total processing time of 41 ms per frame (640 × 480 pixels) on a 3.4GHz Intel Core i7-3770 CPU with 8 GB RAM. In idle situations, with no vehicles present, the processing time is 22 ms per frame.

7 Analysing the use of sports arenas

The interest in analysing and optimising the use of public facilities in cities has a large variety of applications in both indoor and outdoor spaces. While the previously described projects focus on outdoor spaces, this work copes with indoor scenes. Here we are focusing on sports arenas, but other possible applications could be libraries, museums, shopping malls, etc. We aim to estimate the occupancy of sports arenas in terms of the number of people and their positions in real time. Potential use of this information is both online booking systems, and post-processing of data for analysing the general use of the facilities. For the purpose of analysing the use of the facilities, we also try to estimate the type of sport observed based on people's positions.

7.1 Methods

In indoor spaces, the temperature is often kept constant and cooler than the human temperature. Foreground segmentation can therefore be accomplished by automatic thresholding the image. In some cases unwanted hot objects, such as hot water pipes and heaters, can appear in the scene. In these situations, background subtraction can be utilised. After obtaining a binary image, the foreground should be converted to a number of people. Each white blob is simply counted as a person, but in order to handle partial occlusions, the blobs can be split both vertically and horizontally before counting them (Gade et al., 2012). As described in Section 2, no very wide angle lenses exist for thermal cameras yet. To capture a wide area, more cameras can be put together to form a wide angle image. In Figure 19, three images have been stitched together to cover a 20 × 40 m arena. The lower image shows the segmentation of people by automatic thresholding and removing white pixels outside the region of interest, here the court area.

Figure 19 Upper: Thermal input image from a sports arena. Lower: Result after thresholding the image and removing white pixels outside the court area (ROI)

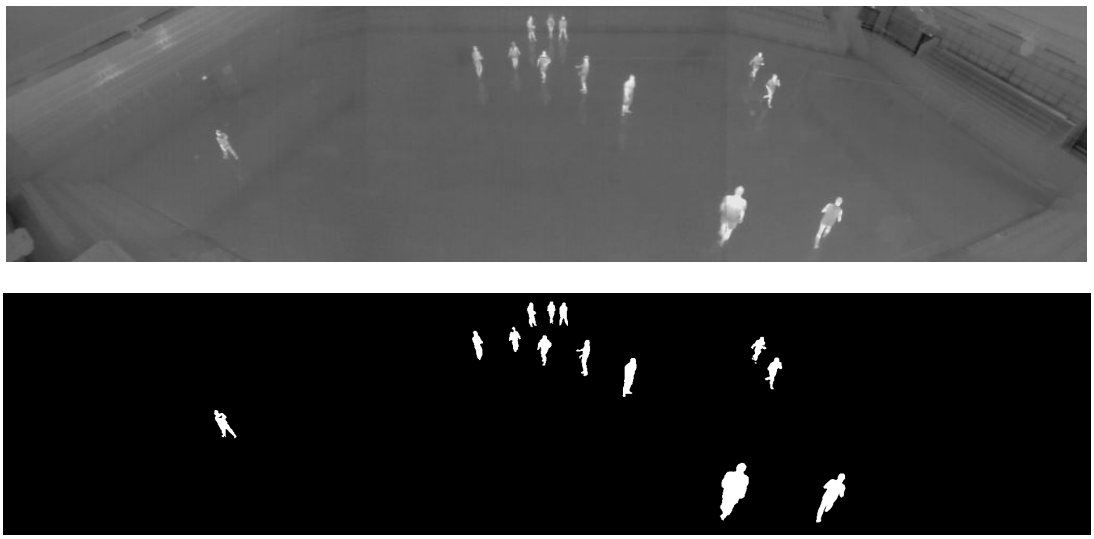
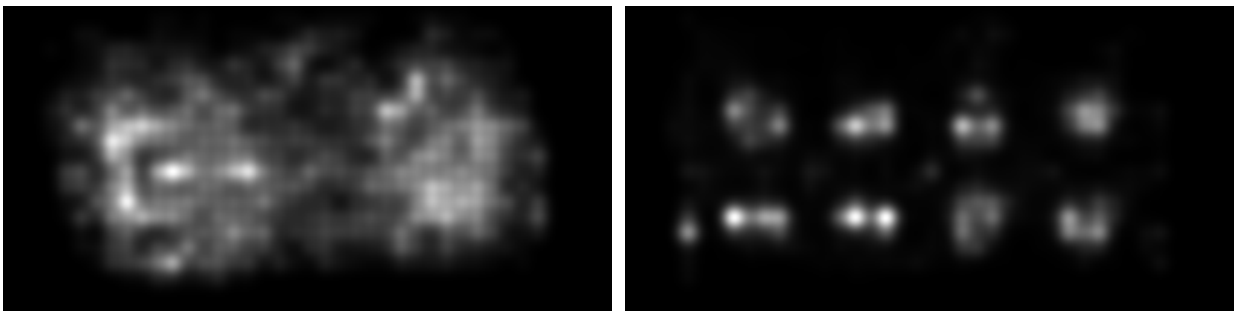


Figure 20 Heatmaps for 10-minute periods



(a) Basketball (b) Badminton

For crowded scenes, occlusions will result in missed detections, and reflections and non-human objects can result in false detections as well. Including temporal information may stabilise these measurements. Detecting when people leave or enter the scene gives information about the transitions in number; during periods with no activity in the border area, the number of people inside the monitored area must be the same. Using a dynamic programming approach, the results are optimised over long periods for more stable measurements (Gade et al., 2013).

From the detection of people, their activities can also be analysed. For sports arenas, we estimate the sports type from the occupancy patterns over 10-minute periods. Each registered position of people on the court contributes with a Gaussian distribution to a heatmap. An example of two heatmaps for basketball and badminton are shown in Figure 20.

A classifier is trained to detect five different sports types from the heatmaps only. From these images, the dimensionality is reduced with PCA and Fisher's Linear Discriminant (FLD), where FLD uses labelled training data and seeks the dimensions which discriminate the classes. After transforming each sample data to the new space, the nearest class is found using the Euclidean distance (Gade and Moeslund, 2013).

7.2 Results

The detecting method has an error rate between 7.35% and 11.76%, depending on the activity level in the videos. Adding the border activity detection and dynamic programming optimisation reduces the error rate to 4.44%. This full approach is tested on 30 minutes of annotated video, while the underlying detection method has been tested on a large annotated dataset from five different arenas. The detection algorithm easily runs real-time, with a processing time of 12.5 ms per frame for large images of 1920×480 pixels. The approach described in Gade et al. (2013), which includes tracking of people near the border takes 60ms per frame, without any optimisation of the software. This means that even this method will be able to perform in real-time. However, the dynamic programming optimisation is a post-processing method and must run after the full period of video has been processed. Both methods are tested on an Intel Core i7-3770K 3.5 GHz CPU with 8GB RAM, and processing 1920×480 pixel images.

The sports type classification has been tested on 30 heatmaps representing five different sports types, plus a category of miscellaneous activities. The recognition rate is 90.76%.

8 Mapping and modelling human movement and behaviour in public spaces

In matters of urban planning and management, it is essential to know how streets and public spaces are being used and how people move around. To quantify and eventually model human movements and patterns of use of a public space over time, which we refer to as Human Spatial Dynamics (HSD), it is thus necessary to track each individual crossing the space under scrutiny. Pivotal to this research is the use of Geographical Information Systems (GIS). The idea is to use computer vision technology to extract accurate geo-referenced tracks of people and use GIS based methods to store and analyse the HSD data created. The advantage of utilising GIS is that the HSD data captured can be easily related to other geospatial data layers and be directly available in the GIS workflow of professional planners and managers.

Along with the advances in software technology and computing power in the last decade, there has been a growing interest in modelling pedestrian and bicyclist behaviour based on a bottom-up approach of programming them as individual entities or agents, which can interact in simulations and yield emergent movement patterns and behaviours resembling those observed in the real world. Collectively, these micro models are referred to as Agent-Based Models (ABM) and there are several approaches to programme the underlying principles. Concepts such as Social Forces (Helbing et al., 2001; Moussad et al., 2010), Cellular Automata (Blue and Adler, 2001), Behavioural Heuristics (Moussad et al., 2009; Moussad et al., 2011), Discrete Choice (Antonini et al., 2006; Bierlaire and Robin, 2009; Robin et al., 2009) and Behavioural Geography and spatial cognition (Torrens, 2012) have been suggested. However, most models focus specifically on crowd or evacuation dynamics and not so much on modelling entire trips of pedestrians in regular traffic in public spaces. Despite advances in modelling techniques towards sophisticated ABMs, they still have challenges in reproducing real world

behaviours reliably in all situations (Castle and Crooks, 2006; Crooks et al., 2008; Papadimitriou et al., 2009). A main reason for that collectively mentioned in the literature is that there exist few empirical studies and verified standard HSD datasets from recordings of real life pedestrian and bicycle traffic to calibrate the models against. The thermal video tracking technology holds the potential for being a way to collect long time HSD datasets in various places, and thus contribute to improve the ABM models. This quantitative approach to tracking urban public life may also be able to supplement the traditional and intuitive manual approaches to HSD data collection used in the studies of urban public spaces and qualities. Inspired by the works of Whyte (1980) and Gehl (2010), a possible outcome of the project is also to contribute with new digital methods to this field.

8.1 Methods

A pilot study was made to prove the concept. A pedestrian zone in Copenhagen with occasional bicycle traffic and goods delivery by vehicles was used as a test scene. The scene was situated where one of the city's major shopping streets meets a perpendicular street and an open square on the way to a major subway station. The site therefore had a continuous flow of pedestrians from several directions that had to negotiate with others to make their way through the scene. At the same time, there were also people in the scene waiting, meeting and talking for longer periods of time. People dragging their bikes or pushing prams or wheelchairs were also observed, as well as groups of school children on excursions. Occasionally, cyclists were observed riding their bikes despite the legislation. Figure 4 shows the scene as both an RGB and thermal image, though with slightly different views. The camera was placed as high as possible on the roof top terrace of a five-storey building next to the scene to minimise the number of people occluding others in the camera FOV, while at the same time capturing the traffic of as large an area as possible. Control points in the scene used to calibrate the homography matrix transferring image pixel coordinates to real world coordinates were measured with high precision GPS equipment.

Computer vision software was applied to analyse 5 minutes of thermal video for which a simultaneous RGB video was also recorded for reference. Background subtraction was used in order to detect people. Since the observed scene was very busy, it was not possible to find an empty frame which could be used as background. Instead, a background model was obtained by calculating the median value for each pixel over a 30 second initialising period. The background was updated during run-time, using a selective update method equal to the one described in Section 5.2. The foreground objects were filtered by size, in order to remove noise. Even though the camera was mounted at a high position, occlusions still caused problems due to the high density of people. In order to solve partial occlusions, we are able to split blobs both vertically and horizontally (Gade et al., 2012). After converting the position of the remaining objects to world coordinates, they were tracked using Kalman filtering (Welch and Bishop, 1995). For each frame, the tracking software yields a list of ID numbers and positions of the detected persons in real-world coordinates. To read the data in a GIS, the raw files were passed by a Python script to render a list of locations for the tracking of each of the IDs. To reduce the amount of data processed while still maintaining sufficient location accuracy the points were down-sampled to fewer instances per second than the original frame rate (30 fps). During parsing of the raw files, a series of attributes were added to the individual points, including speed (in relation to the previous point, a given interval back in time, and accumulated for the track up to the given point) and incremental distance and time. Further metadata was generated for each individual track, including distance, duration, Euclidean distance, average speed, number of points etc.

8.2 Results

The attributes added allowed for various ways of sorting and visualising the tracks in GIS. First noise and false detection needed to be identified and removed. A threshold value of 3 seconds was chosen as the acceptable quality criterion for the minimum duration of a track. From the segmentation of the thermal video described in Figure 6 it was known that there were areas with a high degree of noise, particularly to the right in the FOV near the camera, caused by wind movement of the edges of some sun shades in the scene. Tracks detected along or intersecting the edges of this area were thus classified as unreliable and removed from the rest of the tracks. In Figure 21 the green lines depict the tracks that met the quality criteria (460 tracks) whereas the red crosses indicate detections that were too short in time or distance (1475 IDs). The green tracks clearly show the movement patterns of the area as well as the density of the tracks indicating which routes were the most used. The shaded areas depict the obstacles that the traffic had to evade. An interesting observation concerning the obstacles is the diverging traffic flow in the upper part of the figure, where it is clearly seen that tracks split into two routes, indicating that people are passing either to the left or right around the shaded area in the urban square seen outside the FOV.

Concerning the false detections, the ones caused by the sun shades are clearly seen in the FOV to the right of the camera position. A dense cluster of short detections is also seen near the camera. This was caused by two persons standing close together at that same spot talking for the entire period analysed, including the initialisation period. This caused the two persons to be part of the background model, but small movements and gestures of the people generate short detections. A dense cluster of short detections is also evident at the entrance to the building in the southern end of the FOV. This was caused by several persons passing the doorway simultaneously, thus creating heavy occlusions. Several short detections are also seen spread out across the area where the green tracks dominate. These are probably caused by people occluding each other when walking close together or

passing others in the scene. The number of occlusions would probably be able to be lowered substantially with a nadir looking camera to monitor the scene.

Figure 21 The green lines indicate all the tracks from the 5-minute period analysed, where an ID was followed for more than 3 seconds. The red crosses represent short detections that did not fit the quality criterion. The shaded areas represent obstacles that the traffic had to evade

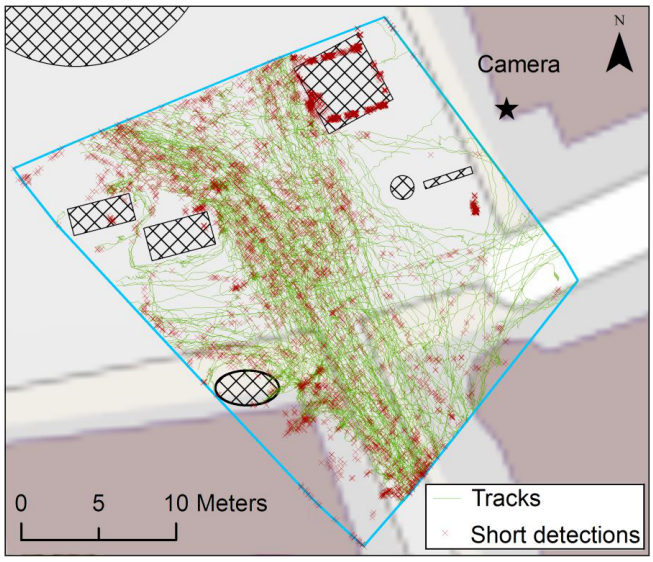
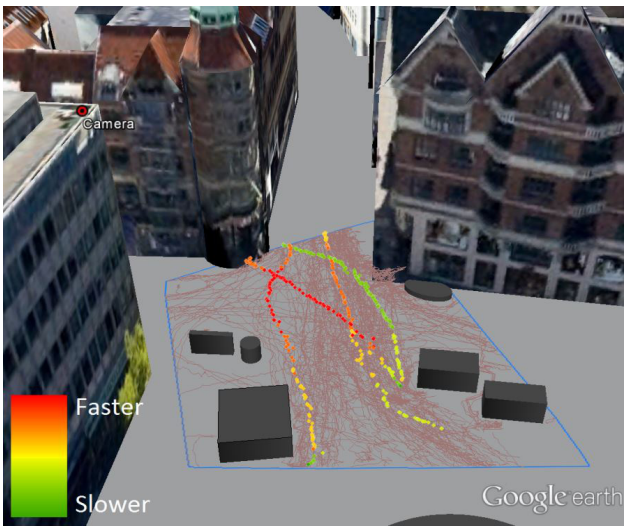


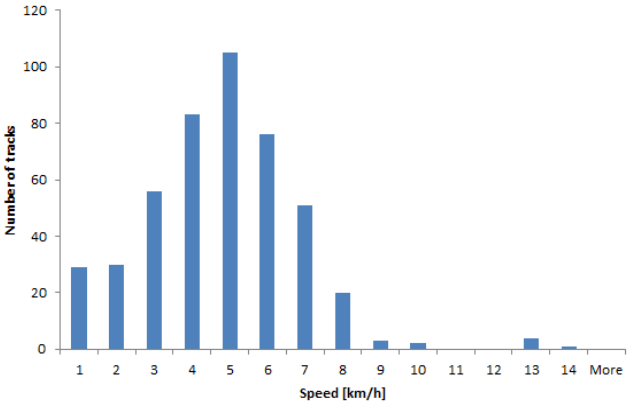
Figure 22 The site and the FOV illustrated in a 3D view with four selected tracks from the same 30 second period coloured according to their speed. All four tracks start in the bottom of the image in the open square and move up towards the street. The tracks indicate both increasing and decreasing speeds. All tracks from Figure 21 are displayed as background. The position of the camera is shown on the corner of the building to the left. The obstacles are indicated in grey



The GIS setup made it possible to make analysis of each individual track and to compare tracks spatio-temporally in order to assess behavioural patterns and the HSD seen in the scene. To show an example of this, four tracks of people moving in the same direction in the same 30 second period were selected and colourised according to their speed as shown in Figure 22. Both increasing and decreasing speeds are seen on the tracks.

To assess the method's overall ability to measure speed, the distribution of the average speeds of all tracks from Figure 21 were plotted in the histogram shown in Figure 23. The graph shows a normal distribution of speeds around 5 km/h, which is typical for pedestrian traffic. In the lower extreme around speeds of 1–2 km/h there are more tracks than in the higher end around 8–9 km/h. This is possibly due the fact that more people are stopping or waiting briefly while in the scene, which lowers their average speed, as opposed to fewer individuals that hurry through the scene. The few tracks in the upper extreme around 13–14 km/h were identified as bicyclist in the video.

Figure 23 The graph shows the distribution of average speeds of the green tracks displayed in Figure 21. The distribution is as expected for pedestrians with speeds normally distributed around 5 km/h. The few tracks with an average speed of around 13 km/h were identified as bicyclists in the video (see online version for colours)



Further research based on this project will aim to develop more advanced GIS methods to study behaviour, such as people's choices of direction and speed, and the interaction with others, in order to enable extraction of behavioural parameters that can be used in ABMs. The analyses shown here were all made as post-processing procedures, but there is nothing hindering the GIS analysis from being automated to generate near real-time online maps of the tracked scene. The processing time of the computer vision tracking algorithm was 20 ms per frame for 640 × 480 pixel images on an Intel Core i7-3770K 3.5 GHz CPU with 8GB RAM. Even without any optimisation or parallelisation of the algorithm this easily obeys real-time requirements, and could be used as input for any real-time analysis of the human behaviour in the public space. GIS methods could thus also be applied in conjunction with some of the other projects presented to make spatial analyses of the tracks generated.

9 Conclusion

The thermal camera is considered an important technology for use in the future Smart Cities. Being a non-intrusive passive sensor, which also preserves privacy, makes it very suitable for the purpose. Furthermore, the thermal camera is independent of light, and thereby operates equally well during day and night, compared to other sensors like the RGB camera; which is strongly dependent on sufficient and stable lighting. For the purposes of e.g. people counting and simple tracking, thermal imaging is highly suited. Some segmentation methods, such as thresholding and image differencing are extremely fast, and still accurate. For more complex tasks, such as tracking of individual people through the city, a different technology able to detect unique features or ID's must be applied.

This paper presented five different Smart City applications in which we applied thermal imaging. They cover both indoor and outdoor environments, monitoring the movements of people, cars and bikes. All systems have proven to be real-time compatible and are tested over very long time in real-world settings.

We have shown here, that by employing thermal cameras it is possible to measure the human use of a city, without violating the privacy of citizens. For the expected future scenario of large scale implementation of intelligent technology in smart cities, we find it crucial to consider sensors and methods that protect the privacy of people. Furthermore, being able to operate day and night without any manual involvement opens up a great number of new applications. Thus, the applications presented in this paper could easily be extended to other smart city applications based on detection and tracking of humans or vehicles.

Acknowledgements

The authors direct a special thanks to Nordea-fonden, The Danish Foundation for Culture and Sports Facilities, The Danish Road Directorate and Danish Lighting Innovation Network for economical support. We would also like to thank the following partners for an innovative collaboration: Teamtronic A/S, Riegens A/S, Alfred Priess A/S and Lund University, and we would like to thank the administration of the building Knud Højgaards Hus and the company Fokustranslatørerne for allowing access to their roof top terrace for the recording of data.

References

Agerholm, N., Caspersen, S., Madsen, J.C.O. and Lahrmann, H. (2006) 'Cykelstiers trafiksikkerhed; en før-efterundersøgelse af 46 nye cykelstiers sikkerhedsmæssige effekt [traffic safety of bike lanes; a before-and-after investigation of the safety effect of 46 new bike lanes]', *Dansk Vejtidsskrift*, Vol. 83, No. 12, pp.52–57.

Ali, A. and Terada, K. (2012) 'A fast approach for person detection and tracking', *International Journal of Computer Applications in Technology*, Vol. 44, No. 3, pp.210–216.

Al-Mutib, K., Emaduddin, M., AlSulaiman, M., Ramdane, H. and Mattar, E. (2014) 'Motion periodicity-based pedestrian detection and particle filter-based pedestrian tracking using stereo vision camera', *International Journal of Computer Applications in Technology*, Vol. 50, Nos. 1/2, pp.113–121.

Andriluka, M., Roth, S. and Schiele, B. (2008) 'People-tracking-by-detection and people-detection-by-tracking', *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.1–8.

Antonini, G., Bierlaire, M. and Weber, M. (2006) 'Discrete choice models of pedestrian walking behavior', *Transportation Research Part B: Methodological*, Vol. 40, No. 8, pp.667–687.

AXIS Communications (2013a) *Axis IP surveillance*, AXIS Communications.

AXIS Communications (2013b) *Axis network cameras*, AXIS Communications.

Batty, M., Axhausen, K.W., Gianotti, F., Pozdnoukhov, A., Bazzani, A., Wachowicz, M., Ouzounis, G. and Portugali, Y. (2012) 'Smart cities of the future', *European Physical Journal Special Topics*, Vol. 214, pp.481–518.

Bierlaire, M. and Robin, T. (2009) 'Pedestrians choices', in Timmermans, H. (Ed.): *Pedestrian Behavior. Models, Data Collection and Applications*, Emerald Group Publishing Limited, pp.1–26.

BLIP Systems (2013) *BlipTrack Smart City Solution*, BLIP Systems A/S.

Blue, V.J. and Adler, J.L. (2001) 'Cellular automata microsimulation for modeling bi-directional pedestrian walkways', *Transportation Research Part B: Methodological*, Vol. 35, No. 3, pp.293–312.

Castle, C.J.E. and Crooks, A.T. (2006) 'Principles and concepts of agent-based modelling for developing geospatial simulations', *CASA working papers*.

CNMAT (2013) *Open sound control*, The Center for New Music and Audio Technology, UC Berkeley.

Copenhagen Cleantech Cluster (2012) *Danish Smart Cities: Sustainable Living in an Urban World*, Copenhagen Capacity.

Crooks, A., Castle, C. and Batty, M. (2008) 'Key challenges in agent-based modelling for geo-spatial simulation', *Computers, Environment and Urban Systems*, Vol. 32, No. 6, pp.417–430.

Dalal, N. and Triggs, B. (2005) 'Histograms of oriented gradients for human detection', *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 1, pp.886–893.

Dollar, P., Wojek, C., Schiele, B. and Perona, P. (2012) 'Pedestrian detection: An evaluation of the state of the art', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 34, No. 4, pp.743–761.

FLIR Systems (2013a) *FLIR X8400 sc specifications*, FLIR Systems Inc.

FLIR Systems (2013b) *FLIR T620 & T640 datasheet*, FLIR Systems Inc.

Gade, R. and Moeslund, T.B. (2013) 'Sports type classification using signature heatmaps', *IEEE Conference on Computer Vision and Pattern Recognition Workshops*.

Gade, R., Jørgensen, A. and Moeslund, T.B. (2012) 'Occupancy analysis of sports arenas using thermal imaging', *Proceedings of the International Conference on Computer Vision and Applications*.

Gade, R., Jørgensen, A. and Moeslund, T.B. (2013) 'Long-term occupancy analysis using graph-based optimisation in thermal imagery', *IEEE Conference on Computer Vision and Pattern Recognition*.

Gehl, J. (2010) *Cities for People*, Island Press.

Helbing, D., Molnár, P., Farkas, I.J. and Bolay, K. (2001) 'Self-organizing pedestrian movement', *Environment and Planning B: Planning and Design*, Vol. 28, No. 3, pp.361–383.

Horn, B.K.P. and Schunck, B.G. (1981) 'Determining optical flow', *Artificial Intelligence*, Vol. 17, Nos. 1–3, pp.185–203.

Hyden, C. (1987) *The development of a method for traffic safety evaluation: The Swedish Traffic Conflicts Technique*, Department of Traffic Planning and Engineering, Lund Institute of Technology, Sweden.

Jensen, O.B. (2013a) *Staging Mobilities*, 1st ed., Routledge.

Jensen, O.B. (2013b) 'The will to connection – a research agenda for the "programmable city" and an ICT "toolbox" for urban planning', *Local and Mobile: Linking Mobilities, Mobile Communication, and Locative Media*, Routledge.

Jensen, S.U. (2006) *Effekter af Cykelstier og Cykelbaner: Før-og-efter evaluering af trafiksikkerhed og trafikmængder ved anlæg af ensrettede cykelstier og cykelbaner i Københavns Kommune [The effects of bike paths and bike lanes; Before-and-after evaluation of traffic safety and traffic volume when establishing one-way bike paths and bike lanes i Copenhagen Municipality]*, Trafitec.

Kapur, J.N., Sahoo, P.K. and Wong, A.K.C. (1985) 'A new method for gray-level picture thresholding using the entropy of the histogram', *Computer Vision, Graphics, and Image Processing*, Vol. 29, No. 3, pp.273–285.

Kitchin, R. (2011) 'The programmable city', *Environment and Planning B: Planning and Design*, Vol. 38, No. 6, pp.945–951.

Moeslund, T.B., Hilton, A., Krüger, V. and Sigal, L. (Eds) (2011) *Visual Analysis of Humans – Looking at People*, Springer.

Moussad, M., Helbing, D. and Theraulaz, G. (2011) 'How simple rules determine pedestrian behavior and crowd disasters', *Proceedings of the National Academy of Sciences*, Vol. 108, No. 17, pp.6884–6888.

Moussad, M., Helbing, D., Garnier, S., Johansson, A., Combe, M. and Theraulaz, G. (2009) 'Experimental study of the behavioural mechanisms underlying self-organization in human crowds', *Proceedings of the Royal Society B: Biological Sciences*, Vol. 276, No. 1668, pp.2755–2762.

Moussad, M., Perozo, N., Garnier, S., Helbing, D. and Theraulaz, G. (2010) 'The walking behaviour of pedestrian social groups and its impact on crowd dynamics', *PloS one*, Vol. 5, No. 4.

Nielsen, T.A.S., Skov-Petersen, H. and Carstensen, T.A. (2013) 'Urban planning practices for bikeable cities – the case of Copenhagen', *Urban Research & Practice*, Vol. 6, No. 1, pp.110–115.

Ouadfel, S. and Meshoul, S. (2014) 'Bio-inspired algorithms for multilevel image thresholding', *International Journal of Computer Applications in Technology*, Vol. 49, Nos. 3/4, pp.207–226.

Panasonic Electric Works (2013) *Infrared Array Sensor: Grid-EYE*, Panasonic Electric Works Corp.

Papadimitriou, E., Yannis, G. and Golias, J. (2009) 'A critical assessment of pedestrian behaviour models', *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 12, No. 3, pp.242–255.

Poulsen, E.S., Andersen, H.J., Jensen, O.B., Gade, R., Thyrrestrup, T. and Moeslund, T.B. (2012) 'Controlling urban lighting by human motion patterns - results from a full scale experiment', *ACM International Conference on Multimedia (MM)*.

Robin, T., Antonini, G., Bierlaire, M. and Cruz, J. (2009) 'Specification, estimation and validation of a pedestrian walking behavior model', *Transportation Research Part B: Methodological*, Vol. 43, No. 1, pp.36–56.

Shepard, M. (Ed.) (2011) *Sentient City: Ubiquitous Computing, Architecture, and the Future of Urban Space*, MIT Press.

Shoushtarian, B. and Bez, H.E. (2005) 'A practical adaptive approach for dynamic background subtraction using an invariant colour model and object tracking', *Pattern Recognition Letters*, Vol. 26, No. 1, pp.5–26.

Torrens, P.M. (2012) 'Moving agent pedestrians through space and time', *Annals of the Association of American Geographers*, Vol. 102, No. 1, pp.35–66.

Welch, G. and Bishop, G. (1995) *An introduction to the Kalman filter*, Technical report, Chapel Hill, NC, USA.

Whyte, W.H. (1980) *The Social Life of Small Urban Spaces*, Project for Public Spaces Inc.