

The Conference on Pedestrian and Evacuation Dynamics 2014 (PED2014)

## Taking the temperature of pedestrian movement in public spaces

Søren Z. Nielsen <sup>a,\*</sup>, Rikke Gade <sup>b</sup>, Thomas B. Moeslund <sup>b</sup>, Hans Skov-Petersen <sup>a</sup>

<sup>a</sup>Department of Geosciences and Natural Resource Management, University of Copenhagen, Rolighedsvej 23, Frederiksberg C 1958, Denmark

<sup>b</sup>Visual Analysis of People Lab, Aalborg University, Rendsburggade 14, Aalborg 9000, Denmark

---

### Abstract

Cities require data on pedestrian movement to evaluate the use of public spaces. We propose a system using thermal cameras and Computer Vision (CV) combined with Geographical Information Systems (GIS) to track and assess pedestrian dynamics and behaviors in urban plazas. Thermal cameras operate independent of light and the technique is non-intrusive and preserves privacy. The approach extends the analysis to the GIS domain by capturing georeferenced tracks. We present a pilot study conducted in Copenhagen in 2013. The tracks retrieved by CV are compared to manually annotated ground truth tracks, and an example of pedestrian behavior is analyzed.

© 2014 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

Peer-review under responsibility of Department of Transport & Planning Faculty of Civil Engineering and Geosciences Delft University of Technology

**Keywords:** tracking; thermal camera; computer vision; GIS; smart cities; pedestrian behaviors; movement pattern analysis; trajectories; visual analysis; space-time cube; urban plazas

---

### 1. Introduction

Planners designing well-functioning liveable cities for people need to know how streets and public spaces are being used and how pedestrians move. The classic approach to collect such data is to make sample counts of people at points of interest a few times a year and conduct qualitative urban analysis (Bauer et al. (2009), Gehl and Svarre (2013)). With the rapid development of computing and networking technologies, the miniaturization of sensors, and the introduction of smartphones, a range of new ways to capture data on people's movement have become available in recent years with potential to supplement and extend the classic methods.

Several studies that track people by using data from smartphones and their signals and sensors, such as Bluetooth, Wi-Fi and Global Navigation Satellite System (GNSS), have been made (Delafontaine et al. (2012), Giannotti et al. (2011), Shoval (2008); van Schaick and van der Spek (2008), Zandbergen (2009)). These studies are interesting on a city wide scale to understand macro movement patterns of samples of people, but the spatial accuracy of data from smartphones is not good enough to study detailed pedestrian movement patterns and behaviors in urban streets and plazas. This instead requires accurate and simultaneous tracking of several individuals who may move close together, and where the movement of each individual depends upon interactions with others as well as on the physical layout

---

\* Corresponding author. Tel.: +45-3533-6773

E-mail address: [szn@ign.ku.dk](mailto:szn@ign.ku.dk)

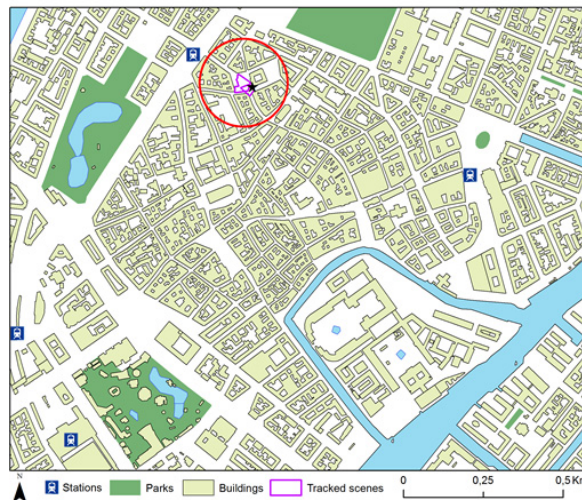


Fig. 1. Overview map of central Copenhagen with the tracked area at 'Kultorget' highlighted. The background data for this map and the buildings in Fig. 5, 6, and 7 are taken from the open public geographic data courtesy of the Danish Geodata Agency.

of the place and attractors in the space traversed (Moussaïd et al. (2010), Timmermans (2009)). For such micro scale pedestrian studies Computer Vision (CV) based tracking technology is more appropriate to use as it is able to passively register the activity in a plaza without affecting the behavior of people in the space. However there are several challenges to capture reliable data this way. This paper presents results of capturing and exploring data on pedestrian movement with CV tracking technology from a pilot study we conducted in the summer of 2013 in the urban plaza 'Kultorget' in central Copenhagen (see Fig. 1).

## 2. Methods

As Computer Vision technology has made rapid progress in recent years (Gowsikhaa et al. (2012), Ko (2011), Moeslund et al. (2011)) we wanted to test it in studies of pedestrian movement patterns and behaviors in everyday traffic in public spaces to assess its potential as a tool to capture such data and aid planners in future Smart Cities (Batty et al. (2012)). Camera surveillance of public spaces in the form of CCTV systems is already installed in many cities, but these systems are often calibrated to aid the police in crime fighting or as traffic cameras to identify vehicles and report on the traffic situation. Police cameras are often Pan-Tilt-Zoom (PTZ) cameras which can be used to zoom in on situations and identify suspects and follow them within a network of cameras around the city, and traffic cameras are focused on vehicle traffic on the road network. To conduct pedestrian movement studies with CV tracking it is necessary to have dedicated cameras with a fixed Field of View (FOV) that can be set up to constantly monitor an Area of Interest (AOI), such as an urban plaza, preferably from an elevated position. This in order to get close to a nadir looking position to avoid the occlusions that occur when people pass each other in front of the camera as CV algorithms can have difficulties to handle occlusions and to distinguish individuals that move close together. The height for optimal camera installation is also a balance between how large a FOV needs to be surveyed versus the level of detail that can be seen in the image. The fact that the further away objects are in a camera's FOV the smaller they appear in the image needs to be taken into account since it is more difficult for CV algorithms to detect and distinguish individuals if they only take up few pixels in the image (Ko (2008)).

Computer Vision algorithms can be applied on video from both normal RGB cameras and thermal cameras. In terms of performance of CV algorithms there are advantages and disadvantages in both technologies that need to be considered (Davis and Sharma (2004), Gade and Moeslund (2014)). Normal RGB cameras record reflected light in three channels, one for the red, green, and blue color respectively, and therefore these cameras depend on sufficient light to operate. RGB sensors are cheap, but they are also somewhat complicated to use for CV tracking of humans in urban scenes. Light conditions change between day and night, and they can change fast in different weather situations

or when the scene is illuminated from headlights on a moving vehicle etc. This is a challenge for CV algorithms that depend on a reliable background model to segment moving objects from the background by the use of background subtraction. Segmentation using background subtraction assumes that it is possible to obtain a reliable background model and that only people, or other objects of interest, are moving. A dynamic background model can be adjusted and updated during run-time of the algorithm, but it is often not enough to achieve good results with RGB cameras. Another well-known problem for CV in the RGB-domain is the occurrence of shadows which often cause false detections, as the shadows move just like people.

In the thermal domain there are no shadows and no issues with fast changing lighting conditions as thermal cameras record the long-wave infrared radiation (8-15 $\mu$ m) emitted by all objects with a temperature above absolute zero. Thermal cameras can thus detect people and objects with a temperature different from the surroundings both day and night independent on the light in the scene. The concept of establishing a background model and use background subtraction to segment moving objects is the same in the thermal domain, and it is easier here to model the background and update it dynamically as temperature conditions in a scene often change more slowly compared to lighting conditions. Dynamic updating of the background model is necessary in outdoor scenes as the surrounding temperature change throughout the day as well as the sun can heat dark pavements to temperatures hotter than the human body temperature. Therefore it cannot just be assumed that people are constantly hotter or colder than the background and thus the CV algorithm has to adjust accordingly. Still people can often more reliably be segmented with background subtraction using thermal cameras in comparison with RGB cameras, and the method is computationally fast, making it well suited for real-time applications. In terms of false detections in thermal images there can be issues for glossy surfaces as thermal radiation can be reflected from these. This is however considered a rare problem in Smart Cities applications as the surfaces are often not reflective and those that are, such as windows, are permanent installations so data from these areas in a scene can easily be filtered out.

When it comes to resolving occlusions and ambiguous situations the RGB cameras have an advantage over the thermal cameras. The three channel RGB images contain more information compared to the one channel of thermal cameras. Colors and textures extracted from RGB images can be applied to re-identify individuals after occlusions, which is not possible for thermal images. However, tracking can often be simplified to the detection of an object, and then assigning the detection to a path established in previous frames by considering the velocity and direction of the object. Where perfect re-identification of objects is not required tracking in thermal imaging can perform well. At the same time the inability to identify individuals in thermal imaging is also one of its greatest assets because privacy issues related to recording video in public places can be neglected as there is no risk of revealing individuals identity in the thermal images. Privacy by design is thus ensured when using thermal cameras for tracking. The pros and cons of RGB versus thermal cameras are summarized in table 1 below.

Table 1. Overview of pros and cons of RGB versus thermal camera for computer vision tracking applications in urban outdoor scenes.

Camera type	Pros	Cons
Thermal	Easier segmentation Independent of light No privacy issues	Re-identification difficult More expensive Fair resolution
RGB	Re-identification possible Cheap sensors High resolution	Sensitive to light Privacy issues Shadows

Even though thermal cameras are more expensive than RGB cameras, especially because of the germanium metalloid needed for the lenses, they have some clear advantages in CV tracking applications (Gade and Moeslund (2014)). With the development in thermal camera technology the resolution of thermal cameras is slowly increasing as the technology evolves and new materials are explored. The cost is also lowering and will probably continue to do so and make them more cost effective to deploy. Considering the capabilities of thermal cameras we wanted to explore their potentials and pitfalls, as well as the quality of the data they create, in a real life CV tracking experiment on pedestrian movement patterns and behavior in urban public spaces. Related studies on pedestrian tracking using thermal cameras are Padole and Alexandre (2010) and Skoglar et al. (2012). However these studies use slightly different Computer Vision techniques and are applied in a different context.

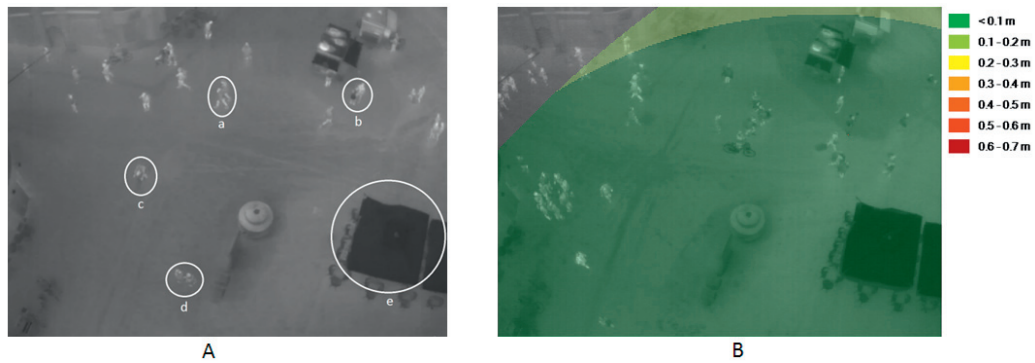


Fig. 2. The near nadir view. The areas highlighted in A refer to examples of behaviors and situations. The image in B is generated in T-analyst and shows the uncertainty for the pixels in the image in relation the real world coordinates.

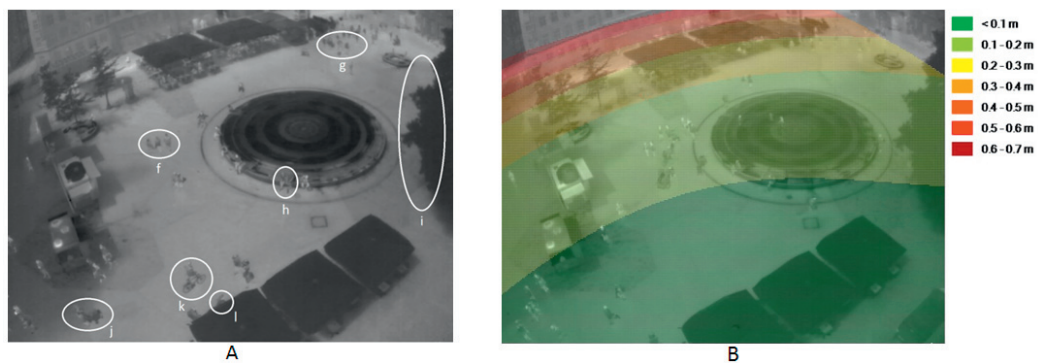


Fig. 3. The view overlooking the plaza. The areas highlighted in A refers to examples of behaviors and situations. The image in B is generated in T-analyst and shows the uncertainty for the pixels in the image in relation the real world coordinates.

In our pilot study we used one state-of-the-art uncooled passive thermal camera with a resolution of 640x480 pixels (Axis Q1922), a lens with a focal length of 10 mm, a viewing angle of 57°, and 30 fps camera frame rate. Background subtraction was applied to detect people. A background model was obtained by calculating the median value for each pixel over a 30 second initializing period. The background was updated during run-time, using a selective update method, meaning that only pixels segmented as background contribute to the updated background. The foreground objects were filtered by size, in order to remove noise. In order to solve partial occlusions, we were able to split BLOBs both vertically and horizontally (Gade et al. (2012)). Binary Large Objects (BLOB) refers to a group of connected pixels in a binary image (Moeslund (2012)). After converting the position of the remaining objects to world coordinates, they were tracked using Kalman filtering (Kalman (1960)). The processing time of the computer vision tracking algorithm was 20ms per frame on an Intel Core i7-3770K 3.5 GHz CPU with 8GB RAM. This is fast enough to provide the raw tracking data in real-time, even though this capability was not tested in our pilot study.

A coordinate system is needed in order to relate the captured tracks to each other and the surrounding space. In a fixed FOV the pixel coordinates in the camera image will remain pointed at the same locations in the scene. By measuring control points in the scene homography can be used to relate the pixel coordinates in the image plane to the real world coordinates in the ground plane by use of a transformation matrix (Criminisi (1997)). To georeference the scene we measured the control points with high precision GPS equipment. With a georeferenced scene the movement data can be related to the surroundings and other geospatial data layers in Geographic Information Systems (GIS), which also allow for spatio-temporal analysis of the data. Depending on people's position in the FOV the spatial accuracy for each pixel is between 0.1 and 1 meter (see Fig. 2B and Fig. 3B).

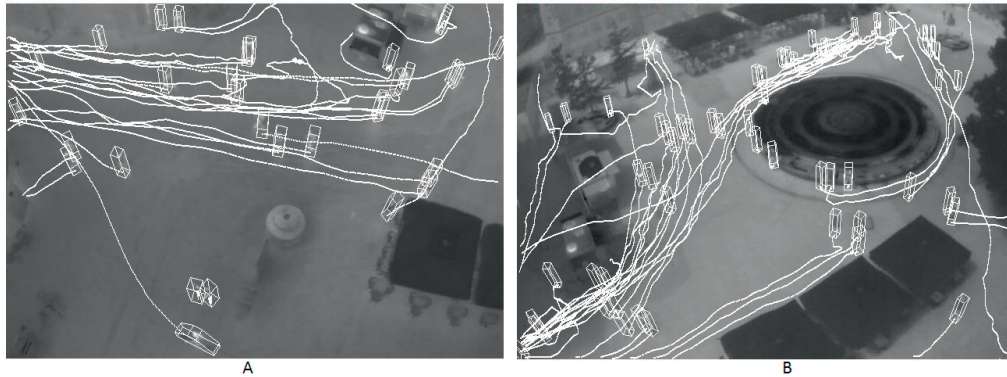


Fig. 4. The two views with the GT tracks digitized manually in T-Analyst overlaid.

For each frame, the CV 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 text files from the CV software were passed to 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 five track points per second than the original rate of 30 points per second. 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 were generated for each individual track, including distance, duration, Euclidean distance, average speed, number of points etc.

To assess the quality of the CV trajectories in terms of their completeness and accuracy the CV tracks needed to be able to be evaluated against Ground Truth (GT) trajectories. To do so we have manually digitized the GT trajectories of all individuals in the video recordings presented in this study. This has been done in T-Analyst developed at Lund University (Lauresbyn (2013)). In this software pedestrians are modeled as 3D rectangles with the dimension of 0.5x0.5x1.8 meters (see Fig. 4). The user can then manually digitize the position of a pedestrian frame by frame. The GT tracks digitized in T-Analyst were also transformed from pixel to real world coordinates in a similar manner as for the CV tracks and imported into GIS.

### 3. Scene description

The plaza 'Kulturvet' was used as a test scene. It is in a pedestrian zone in central Copenhagen with occasional bicycle traffic and goods delivery by vehicles. The part of the scene closest to the camera was situated where one of the city's major shopping streets meets a perpendicular street at the entrance to the open plaza. The street at the far end of the plaza leads directly to the subway station with the most traffic in the city. The scene had a continuous flow of pedestrians (50-100 per minute) coming from several directions that needed to negotiate and avoid each. The thermal camera was placed on the roof top terrace of a five story building. Two views were recorded. The first one, which we label the near-nadir view, was a straight down view to get as close to the nadir position as possible in order to minimize the number of people occluding each other in the camera FOV (Fig. 2A). Consecutively the second view, labeled the overlooking view, was taken from the same spot but now instead overlooking the entire plaza from an oblique angle (Fig. 3A). The scenes were recorded around noon on a Friday. The weather was overcast with occasional showers of rain.

Tracks of people walking alone or in social groups of different sizes were recorded (Fig. 2A a and Fig. 3A f), as well as people sitting or waiting (Fig. 3A h), people having a conversation (Fig. 2A d and b), and people dragging their bikes (Fig. 2A k) or pushing a pram or stroller (Fig. 3A j). The tracks of 'facers' working for a charity organization trying to stop people in the street to make them donate to the cause were also recorded in the scene (Fig. 2A b). Occasionally cyclists riding through the scene despite the legislation were observed (Fig. 2A c).

Following the video recordings the permanent objects in the scene such as hotdog stands, sun shades, benches, a fountain etc. were digitized as polygons in GIS from the most recent orthophotos of the place and cross checked with

the thermal video recordings to confirm their actual positions (grey objects in Fig. 5, Fig. 6, and Fig. 7). In this way these objects could be drawn and georeferenced in GIS and the tracks thus analyzed in relation these objects. A layer with the surrounding building footprints were added and drawn as well (green polygons in Fig. 5, Fig. 6 and Fig. 7).

#### 4. Analysis and results

Half an hour of thermal video was recorded for both views. It was decided to only process five minutes of the video from the first view and one minute from the second view. The reasons for processing only these sections was mainly due to long time it took to manually digitize the GT tracks for assessment of the CV tracks. However it was considered that this was enough to get data on the general movement patterns as well as to identify areas and standard situations in the scenes in which the CV software was challenged.

The data generated for the CV and GT tracks were imported to a geodatabase in the ArcGIS software as a point feature class. For each point record the basic data was in the format (*FrameNumber*, *ID*, *Xcoordinate*, *Ycoordinate*, *TimeStamp*). Furthermore the derived variables in terms of the accumulated time, distance and average speed for the track were written for each point record. The Python script passing the data also generated a table summarizing statistics such as total distance, total duration, average speed, start and end frame number, and number of track points for each *ID*. The track point table and the statistics table were joined on the *ID*. To enable visualization of the temporal dimension the data was converted to 3D by using the *FrameNumber* attribute as the Z coordinate. Since the geodatabase date-time format did not natively support milliseconds the *TimeStamp* field could not be used. To connect track points to lines a standard point to line tool based on the *ID* and sorted by the *FrameNumber* was applied.

Inspired by the works on visual analytics of movement by Andrienko et al. (2013) and Keim et al. (2008) the data was explored visually in 2D and 3D. The tracks obtained for the near nadir view and the overlooking view are shown in Fig. 5 and Fig. 6 respectively. The A parts in the figures depict the tracks obtained from the CV algorithm in a 2D map, the B parts show the manually digitized GT tracks for the same scene, and the C parts show the data from A and B plotted in a Space-Time Cube (STC) for visualization of the temporal dimension. For Fig. 5C only the first minute of data is shown in the STC, but data for all five minutes are shown in Fig. 5A and Fig. 5B. A visual exploration of the data is naturally best undertaken in a 3D GIS program where one can work with the STC to rotate, pan and zoom in on interesting areas of the cube. It is much more challenging to communicate the results visually on paper. The following describes our visual inspection and interpretation of the data to identify areas and situations in the scenes where the CV software had difficulties in tracking individuals correctly.

To filter out noise in the CV data it was decided to accept only CV tracks with duration of three or more seconds. This approach identified most of the ambiguous situations in which people had occluded each other from view which confused the CV software. The situations occurred most often in the area with the main flow of people i.e. the highest density of tracks as seen on Fig. 5A, Fig. 5B and Fig. 6A, Fig. 6B. The short detections are plotted as red dots. A prominent red area is spotted near the camera in Fig. 5A. This was caused by the two people marked with d on Fig. 2A, who stood close on the same spot and talked for all five minutes of video, which gave the CV algorithm a challenge to obtain a lock on them as two individuals. Instead it saw them as part of the background in most frames and assigned them new track IDs each time they moved slightly. The same was the case for the sitting people marked with h in Fig. 3A. These two are thus examples of false negative detections.

It was evident from the data that there were areas with a large amount of detections that could not be moving people, but instead movement of objects such as tree branches (area i on Fig. 3A) or the canvas on sun shades (areas e and l on Fig. 2A and 3A) swaying in the wind. These were thus all tagged as red short detections as seen in the upper grey area depicting the sun shades in Fig. 5A, and in the long dense red area in Fig. 6A that is due to the tree branches. These areas are thus examples of false positive detections.

Since it was difficult for the CV algorithm as well as manually to track individuals in the far end of the plaza in the overlooking view (area g in Fig. 3A), it was decided to draw a polygon in which at least unambiguous manual GT tracks could be digitized (see Fig. 6). All CV detections outside this area were tagged as short detections and clipped from the good CV tracks inside the polygon.

In our manually digitized GT we found 297 tracks crossing the near nadir scene in the 5 minutes of data. The CV algorithm found 460 tracks that could be classified as good tracks. 1475 track IDs were classified as short detections.

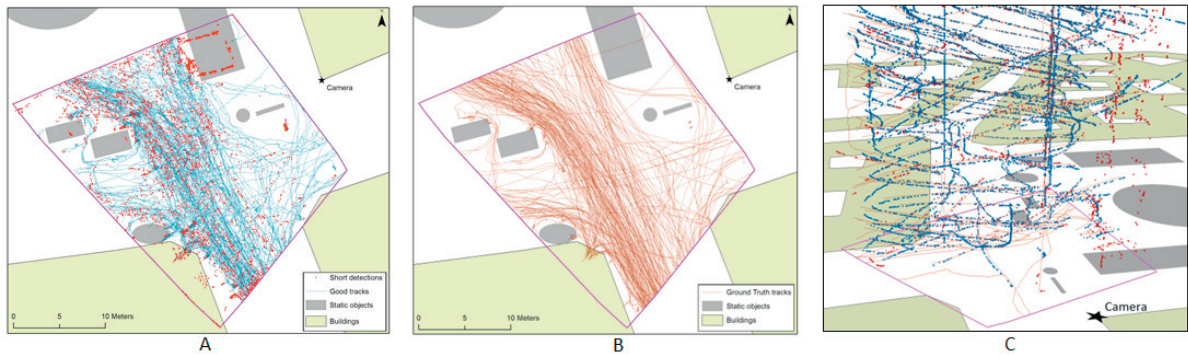


Fig. 5. The tracks from the near nadir view. A shows the good CV tracks as well as the short CV detections. B shows the manually digitized GT tracks for comparison to A. C shows a Space-Time Cube of the first minute of data from A and B.

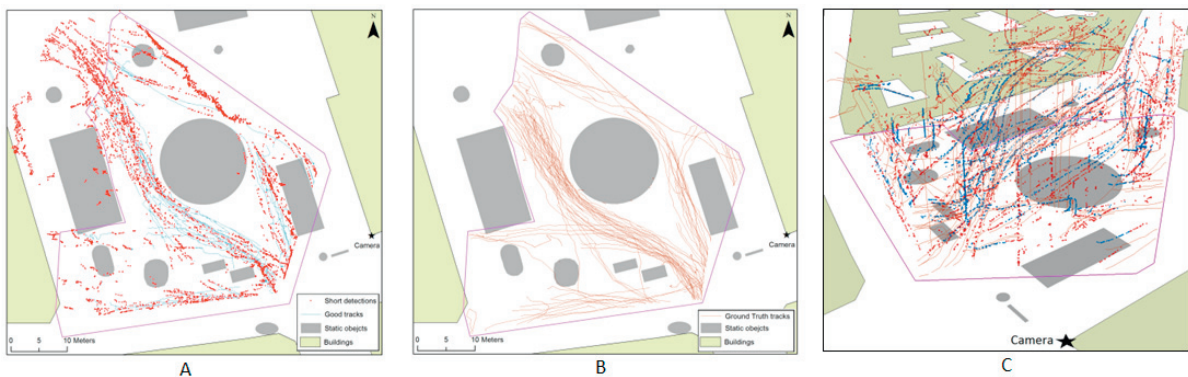


Fig. 6. The tracks from the view overlooking the plaza. A shows the good CV tracks as well as the short CV detections. B shows the manually digitized GT tracks for comparison to A. C shows a Space-Time Cube of the data from A and B.

Close inspection of the STC showed that several of the good CV tracks could be referred to as being parts of one corresponding GT track. This indicates the IDs of good CV tracks have been changed after ambiguous situations, but that the same individuals have been tracked all through the scene. In the one minute data from the overlooking view we found 124 tracks by the manual GT method and the CV method found 146 tracks. There were 977 IDs classified as short detections.

The overall movement patterns detected by the CV fit well with the GT tracks, and in the STC in Fig. 5C and Fig. 6C it can be seen that individual GT and CV tracks are fairly good aligned. A quick inspection (not shown) of the speeds of the good CV tracks yielded a Gaussian distribution around 5km/h for pedestrian tracks. A small spike around 13 km/h was clearly identified as cyclists in the scene.

The data also allowed for extraction of individual tracks for analysis of movement behaviors seen in the video. An example of this is shown in Fig. 7 where the two tracks of a couple are shown and where three situations from their tracks are highlighted. In situation 1 they walk together, in 2 they stop to discuss and one points at the fashion shop on the street corner, in 3 the two tracks split up as the woman (verified by live observation when the scene was recorded) decides to enter the shop while the man waits outside. The thermal images of the situations are shown in Fig. 7C. In Fig. 7A the GT tracks are shown in a 3D plot with the two tracks colored in red when they move more than 2 km/h and green when they move slow or stand still. The tracks are projected on the 2D plane for reference. In Fig. 7B the CV track points detected within a radius of one meter and one second, in the spatial and temporal dimensions respectively, of the corresponding GT tracks are displayed. The different CV tracks are colored according to their IDs. It is evident that the IDs are changed quite a few times for the CV track points, even though the two persons remained

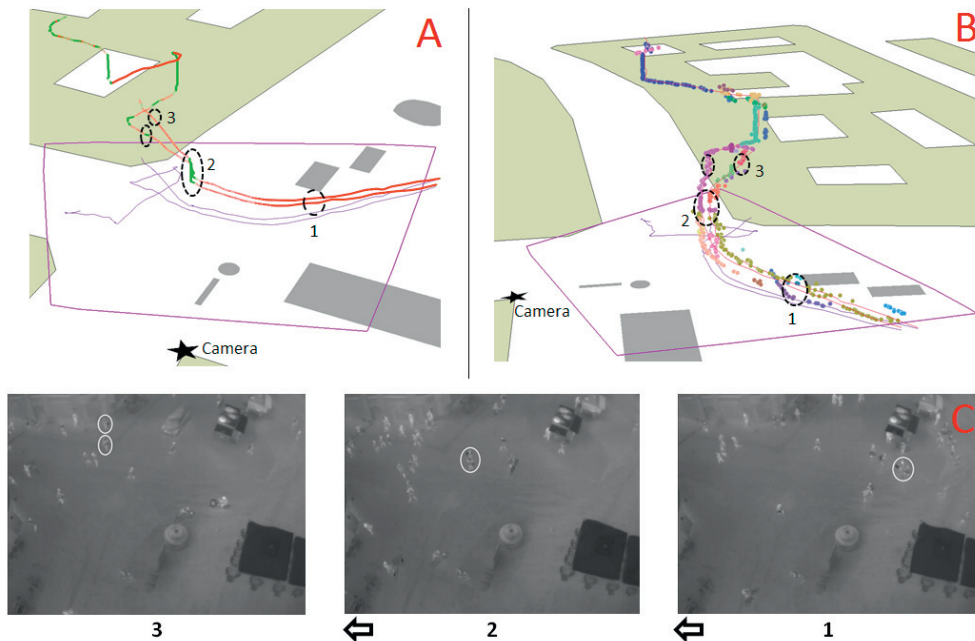


Fig. 7. Example of extraction of the two tracks of a couple. The GT tracks are shown in A, the CV track points are shown along with the GT tracks in B, and the thermal images for three step sequence is shown in C.

tracked throughout the sequence. The CV method is thus not robust enough yet for tracking individuals consistently in advanced and ambiguous situations, but it performs fairly well to extract overall movement patterns.

## 5. Discussion and conclusion

In this paper we have presented a method using thermal cameras and state-of-the-art Computer Vision technology to track pedestrians in public plazas. The tracks are georeferenced to enable analysis in GIS. Furthermore we have compared the tracks obtained from Computer Vision with manually annotated ground truth tracks by visual analysis of the data on 2D maps and in 3D Space-Time Cubes. The analysis showed examples of situations in which the Computer Visions tracking was challenged, but indicated also that the method has potential to provide reliable data on general movement patterns in a plaza with a similar density of people. An example of extraction and analysis of two tracks of a couple in a standard situation was shown. From this it was clear that the Computer Vision software was not able to keep the individuals tracked consistently, however it was able to detect the persons present all the time. The changing of track IDs was most likely caused by the two people walking and standing close together as well as by others passing close by. It could be useful to test our method in other types of plazas and with higher densities, and to extract and analyze tracks from different situations to compare data.

Further development of the method could be to enable computation of statistics for the relation between Computer Visions and ground truth tracks in terms of accuracy and completeness. An idea could also be to enable automatic classification of moving individuals into types such as pedestrians, cyclist, people pushing prams etc. The dataset obtained in this pilot study has enabled a better understanding of the difficulties for further automation. While going through the videos to digitize ground truth tracks we identified characteristic movement behaviors such as meeting, flocking, avoidance, and following a leader (Gudmundsson et al. (2012)). Further research could hence be in the field of Computational Movement Analysis to automate search for characteristic movement patterns and behaviors in this kind of dataset.

We are working on plans to carry out a full scale study with multiple thermal cameras over a sustained period of time. Inspired by the works of Gehl and Svarre (2013) and Whyte (1980) the ambition is to contribute with new



digital methods and tools in the field of urban analysis by linking the field of Computer Vision to that of GIS. The long term goal is to further investigate the applicability and suitability of this type of studies to provide data for models of everyday pedestrian behavior in urban public spaces.

## Acknowledgements

We would like to thank Aliaksei Lareshyn from Lund University for permission to use the T-Analyst software and assistance with setting it up. We would also like to thank the administration of the building Knud Hjgaards Hus in Copenhagen and the company Fokustranslatrerne for allowing access to their roof top terrace for us to install the thermal camera for the pilot study. Thanks to Louise Schnegell for help with setting up the manuscript in LaTeX.

## References

- Andrienko, G., Andrienko, N., Bak, P., Keim, D., Wrobel, S., 2013. *Visual Analytics of Movement*. Springer Berlin Heidelberg, Berlin, Heidelberg. doi:10.1007/978-3-642-37583-5.
- Batty, M., Axhausen, K.W., Giannotti, F., Pozdnoukhov, A., Bazzani, A., Wachowicz, M., Ouzounis, G., Portugali, Y., 2012. Smart cities of the future. *The European Physical Journal Special Topics* 214, 481–518. doi:10.1140/epjst/e2012-01703-3.
- Bauer, D., Brändle, N., Seer, S., Ray, M., Kitazawa, K., 2009. Measurement of Pedestrian Movements: A Comparative Study on Various Existing Systems, in: Timmermans, H. (Ed.), *Pedestrian Behavior*. Emerald Group Publishing Limited. chapter 15, pp. 325–344.
- Criminisi, A., 1997. *Computing Plane to Plane Homography*.
- Davis, J., Sharma, V., 2004. Robust detection of people in thermal imagery. *Proceedings of the 17th International Conference on Pattern Recognition*, 2004. ICPR 2004. , 713–716 Vol.4doi:10.1109/ICPR.2004.1333872.
- Delafontaine, M., Versichele, M., Neutens, T., Van de Weghe, N., 2012. Analysing spatiotemporal sequences in Bluetooth tracking data. *Applied Geography* 34, 659–668. doi:10.1016/j.apgeog.2012.04.003.
- Gade, R., Jørgensen, A., Moeslund, T.B., 2012. Occupancy analysis of sports arenas using thermal imaging. *Proceedings of the International Conference on Computer Vision and Applications*.
- Gade, R., Moeslund, T.B., 2014. Thermal cameras and applications : a survey. *Machine Vision and Applications* , 245–262doi:10.1007/s00138-013-0570-5.
- Gehl, J., Svarre, B., 2013. *How to Study Public Life*. Island Press.
- Giannotti, F., Nanni, M., Pedreschi, D., Pinelli, F., Renso, C., Rinzivillo, S., Trasarti, R., 2011. Unveiling the complexity of human mobility by querying and mining massive trajectory data. *The VLDB Journal* 20, 695–719. doi:10.1007/s00778-011-0244-8.
- Gowsikhaa, D., Abirami, S., Baskaran, R., 2012. Automated human behavior analysis from surveillance videos: a survey. *Artificial Intelligence Review* doi:10.1007/s10462-012-9341-3.
- Gudmundsson, J., Laube, P., Wolle, T., 2012. *Computational Movement Analysis*, in: *Springer Handbook of Geographic Information*, pp. 423 – 438.
- Kalman, R.E., 1960. A New Approach to Linear Filtering and Prediction Problems 1 82, 35–45.
- Keim, D., Andrienko, G., Fekete, J.d., Carsten, G., Kohlhammer, J., Melancon, G., 2008. *Visual Analytics : Definition , Process , and Challenges*, in: Kerren, A. (Ed.), *Information Visualization*. Springer, Berlin, Heidelberg, pp. 154–175.
- Ko, T., 2008. A survey on behavior analysis in video surveillance for homeland security applications. 2008 37th IEEE Applied Imagery Pattern Recognition Workshop , 1–8doi:10.1109/AIPR.2008.4906450.
- Ko, T., 2011. A Survey on Behaviour Analysis in Video Surveillance Applications, in: Lin, W. (Ed.), *Video Surveillance*. InTech.
- Lareshyn, A., 2013. T-Analyst software. URL: [http://www.tft.lth.se/video/co\\_operation/software/](http://www.tft.lth.se/video/co_operation/software/).
- Moeslund, T.B., 2012. *Introduction to Video and Image Processing. Building Real Systems and Applications*. Springer. doi:10.1007/978-1-4471-2503-7.
- Moeslund, T.B., Hilton, A., Krüger, V., Sigal, L. (Eds.), 2011. *Visual Analysis of Humans: Looking at People*. Springer. doi:10.1007/978-0-85729-997-0.
- Moussaïd, M., Perozo, N., Garnier, S., Helbing, D., Theraulaz, G., 2010. The walking behaviour of pedestrian social groups and its impact on crowd dynamics. *PloS one* 5, e10047. doi:10.1371/journal.pone.0010047.
- Padole, C.N., Alexandre, L.a., 2010. Motion Based Particle Filter for Human Tracking with Thermal Imaging. 2010 3rd International Conference on Emerging Trends in Engineering and Technology , 158–162doi:10.1109/ICETET.2010.120.
- van Schaick, J., van der Spek, S. (Eds.), 2008. *Urbanism on Track*. IOS Press, Amsterdam.
- Shoval, N., 2008. Tracking technologies and urban analysis. *Cities* 25, 21–28. doi:10.1016/j.cities.2007.07.005.
- Skoglar, P., Örguner, U., Törnqvist, D., Gustafsson, F., 2012. Pedestrian tracking with an infrared sensor using road network information. *EURASIP Journal on Advances in Signal Processing* 2012, 26. doi:10.1186/1687-6180-2012-26.
- Timmermans, H. (Ed.), 2009. *Pedestrian Behavior*. Emerald Group Publishing Limited.
- Whyte, W.H., 1980. *The Social Life of Small Urban Spaces*. Project for Public Spaces Inc, New York.
- Zandbergen, P.a., 2009. Accuracy of iPhone Locations: A Comparison of Assisted GPS, WiFi and Cellular Positioning. *Transactions in GIS* 13, 5–25. doi:10.1111/j.1467-9671.2009.01152.x.