Automatic pedestrian trajectory detection to support planning

Barsi1, Arpad – Lovas1, Tamas – Ignazvolgyi2, Zsuzsanna – Radoczy1, Karoly

1 BME Department of Photogrammetry and Geoinformatics
2 BME Department of Highway and Railway Engineering
[barsi.arpad; lov.as.tamas; igazvolgyi.zsuzsanna]@epito.bme.hu

Abstract
Planning pedestrian corridors requires various information on the pedestrian flow. Prior to renovating, rebuilding areas frequently used by pedestrians, the preferred paths, the walking speeds and the critical conflict zones can be detected. The goal of this paper is twofold. First, it discusses an automatic detection method that enables deriving pedestrian trajectories based on video footage. Second, besides numerical analysis, the infrastructure planning is supported by generating dynamic heat maps.

The proposed data acquisition method is inexpensive and rapid; footage is captured by amateur video camera, the image series are processed using MatLab built-in functions. After detecting a pedestrian, its trajectory is computed and recorded while waking in the camera field-of-view. Dividing the pedestrian corridor to virtual cells (which is not necessary a regular grid), an occupancy value/level can be calculated to each cell that is the base of the heat map. Since data is coming from time-stamped image series, the walking speed can also be calculated, and the velocity distribution map of the corridor can be created.

By combining these information with static data, i.e. mapping road furniture or any kinds of obstacles on the ground, the critical conflict zones can be detected. Conflict zones can also be weighted by the walking speed; crossing high speed trajectories means higher risk. By inverting the problem, locating “calm zones” enables supporting the planning and deployment of service facilities or green zones.

Introduction
Estimating indoor and outdoor pedestrian traffic flow parameters is very important in crowded areas. The volume, speed, and directions of pedestrian traffic in a shopping mall, in airports or around other sizeable buildings, or in the streets have great value for planning, but also for maintaining the infrastructure. This demand can be fulfilled by several techniques, which are able to derive the necessary data. The sensing techniques have great variability; Cessford and Muhar have classified the methods of field traffic counting [1]:

- Counting personnel – staff counting traffic on particular locations,
- Video recording – on-site data acquisition, postprocessing afterwards,
- Remote sensing – post-processing of remotely sensed data,
- Mechanical equipment – e.g. counting gate,
- Pressure sensors – devices counting based on pressure change (e.g. by steps),
- Seismic sensors – sensors integrated in the road pavement sensing the surface vibrations,
- Active optical sensors – sensing by blocking of the light beam,
- Passive optical sensors – sensing by changes of the emitted infrared pattern,
- Magnetic sensors – the passing metal object changes the magnetic field,
- Radar – sensing the changes of reflected radio waves.

The data capture method of the current paper is based on optical imagery. The improvement of image capture (both video and still images), as well as the processing methodology in the last years is easy to be noticed. The new similarity based features, blob analyses, statistical and artificial intelligence algorithms, or the techniques based on dynamic system description have broadened the possibilities to acquire the needed information. The real time detection of pedestrians is also a valuable input feature for the modern automotive industry.

Tracking of individual pedestrians is not only a timeline extension of independent moving object detection, but is also a continuous identification, movement estimation and data logging. Video footage capturing techniques are therefore amazing data sources, where the automation of accurate object extraction is one of the biggest challenges.

Tyagi et al. (2008) used multiple cameras to track people in crowded outdoor environment. Dehgan et al. (2012) showed how video footages can be used to detect individuals in low and high crowd density. Louis and Nishino (2012) used video analysis to estimate pedestrian efficiency. Motions of individuals often deviate from the crowd’s flow and decreasing the pedestrian’s efficiency. The authors demonstrated the use of this efficiency to detect unusual events and to track individuals in crowded scenes. Nielsen et al. (2014) proposed to apply a system using thermal cameras and computer vision combined with Geographical Information System (GIS) to track and assess pedestrian dynamics and behavior in urban areas. Aniket et al. (2015) extracted pedestrian trajectories and experienced problems in processing images without camera calibration and handling distortions.
During the research our goal was to develop an automatic solution for capturing and evaluating video footages, where the results can effectively support infrastructure planning and design purposes. The applied equipment was an action cam (described in chapter 2), the processing of the image sequences was done in Matlab (see chapter 3). The next chapters bring the derived results and conclusion.

**Image capture by action cam**

The required images can be gathered by different cameras: from cheap web cams through security cameras to the professional devices. The selected equipment was an action cam from the lower price category, but can offer even full HD or 2K image quality. Because the primary purpose was to collect information about pedestrian flow in a well-defined space, the lens of the camera could be fixed or fix-focus lens can also be used. Considering price factors, an ActionCam SJ400 camera was applied with full HD (1920 × 1080 pixel) image capture at a speed of 30 fps (Fig 1a). The field of view of the camera is 170°, the focal length is 2.99 mm. The videos and images are stored on a 32 GB microSD memory card.

Images were captured in the campus of the Budapest University of Technology and Economics on 17th of February, 2015, starting at 8:12 – just before the morning peak hour in the university life. The logged footage was about 8 minutes long; the 14513 frames required ~964 MB storage. The first and last several frames were blurred because of the small movement of pushing the record button, so the original sequence was cut.

The study area is a usually crowded foot-path between the campus buildings; it was observed from the 4th floor window of a neighboring building. There is no vehicle traffic on the study area, only pedestrians (maybe cyclists) can be detected here.

Since the camera has visible radial (barrel) distortion, prior camera calibration is necessary. The Matlab Camera Calibrator tool contains a calibration checkerboard with 10 × 7 black and white squares of side length of 25 mm. The calibration image was shown on a display, while the rotating camera captured the video. The sections of all black and white squares (9 × 6 = 54) are the image points which have to be measured automatically using image processing interest operators. The best 15 frames, where the checkerboard mostly filled the image areas were extracted from the video for calibration. The Camera Calibrator tool measured the image points, and computed the camera calibration coefficients as follows (Fig 1b):

- focal length in x and y direction,
- principal point coordinates (x and y),
- radial distortion coefficients (3),
- skew factor,
- tangential distortion coefficients (3).

![Figure 1. The ActionCam SJ400 and its calibration in Matlab Camera Calibrator tool](image1)

The mean reprojection error was 0.6909 pixels, which was acceptable, thus the whole video can be undistorted with the derived camera parameters. An example of the original and undistorted image can be seen in Fig. 2.

![Figure 2. Original barrel distorted and corrected image frame](image2)
The image correction for the whole video took about 50 minutes, then the undistorted (corrected) image sequence was the input for pedestrian detection.

**Detecting pedestrians**

The pedestrian detection is a task that has to be executed on all frames; the algorithm was developed for a general sample frame, then the whole video was processed. As one can see in Fig. 2, the crowd move only in a part of the captured image, so the complementary part of this region can be masked out. Since the camera is fixed, the background doesn’t change, therefore only the differences have to be detected between each frame and the background. This leads to the first processing step: a general background image has to be derived. The first 40 frames were used to parameterize the background, then all differences in the foreground can be interpreted as moving objects. The built-in Matlab function is based on Gaussian Mixture Model [2]. The output of the foreground extraction is a binary image containing blobs. In order to reduce noises, binary image morphology operations were followed: opening and closing with $2 \times 2$ rectangular structural elements, then hole filling finishes the detection phase. The blobs prepared this way show the most pixels of the pedestrians.

![Figure 3. The masked input image and the detected noiseless blobs](image)

The image frames can be considered as observations for pedestrians in a dynamic system. The most widely used dynamic system estimation method is the Kalman filtering. In our case the objects have features like position and velocity, these are the elements of the state vector in a Kalman filter. The observations are the centroids of the noise filtered blobs. The advantage of using Kalman filters are twofold: (1) it can smoothen the trajectory and (2) can estimate the position for the future (even for case where the object is hidden).

The above mentioned processing steps (blob detection, noise elimination, Kalman filtering) can be executed in a loop; for the whole image sequence it required about 39 minutes on an average i7 processor 8 GB RAM laptop. The result is the trajectory points for all independent objects (i.e. for all detected pedestrians). The reference system is the same as that of the images, therefore the trajectories are given in a local coordinate system.

![Figure 4. The extracted pedestrian trajectories](image)
When the goal is to support the planning, the trajectories are to be converted into a global coordinate system; a step that requires ground control points (GCPs). In our solution we handled the pedestrian trajectories in a common system, which was a high definition Google Earth Pro image about the study area. The GCPs have to be measured both in the action cam images and in the reference image. The point coordinate measurement was also done in Matlab, as it can be seen in Fig. 5.

Figure 5. Ground control point coordinate measurement in the local (action cam - left) and global (Google Earth - right) images

The measurements of the GCP image coordinates were used to compute a projective 2D transformation between the two image sets. All trajectory points can be quickly transformed into the global coordinate system and the result is a footprint (ground view – see Fig. 6)

Figure 6. Ground view of the trajectories
After getting the horizontal coordinates of the trajectories, further accumulation can be executed. For this reason, a regular grid was laid over the ground view and all trajectory points were counted in each cell. The count gives the cumulative traffic in the affected region, which can be illustrated in form of traffic heat map, as it can be seen in Fig. 7.

Figure 7. Traffic heat map (pedestrian density chart) of the study area

Transportation experts are modelling pedestrian traffic in microsimulation environment (e.g. VISSIM); modelling trajectories around sharp turns is always complicated, hard to simulate. The trajectories and the heat map also enable to measure the across-track distances the pedestrians keep by passing an object, since this effect also reduces the capacity. Note how people pass the rest place (with benches and high-speed wi-fi access point) resulting an empty space in the middle of their main flow (Fig. 4, Fig 6 and Fig. 7.).

Conclusion
Modelling pedestrians’ trajectories is a key input parameter in planning. They walk with different speed, select optimal path that is not necessarily the shortest one and pass different obstacles in different way. Such obstacles are the road furniture (bench, fountain, garbage bin etc.). Capturing videos in peak traffic periods enables deriving the pedestrian trajectories that supports obtaining additional products, e.g. heat maps showing the frequently used areas. Evaluating the trajectories in highly occupied areas the method can be used for capacity planning. The main advantage of the method is its coverage; it can be used for tracking pedestrians for a long distance in open spaces.

Surveying and analyzing crowded areas and collaborating with architects, parks, squares with high walkability indices can be planned and created. The method effectively can be used for reconstruction, leaving sufficient space to pedestrians to pass an obstacle, setting appropriate safety distances from cyclists.

References


