

HUMAN BODY DETECTION METHODS
A literature review

Supervisors:
Jos van den Heuvel and Matthias Rauterberg

May/June 2005

Bram van der Vlist (student)
Department of Industrial Design
Technical University Eindhoven (The Netherlands)

1. INTRODUCTION

The problem my Individual Bachelor project addresses is the risk of getting injured by a running boat propeller. In order to solve this safety problem I am investigating possibilities to detect humans which are close to the propeller and make a system perform a certain action to prevent injury.
The reason for this literature review is to assess if there are any human body detection methods available which could be used to detect the presence of human beings in above and underwater situations. The situation is a very complex one, a constantly changing environment and passing objects with quite high speeds (top speeds up to 90 km/h approx).

2. SEARCH QUESTIONS, STRATEGY AND DATABASE CHOICE

A literature search is done in the digital databases Compendex and Inspec.

The following search question(s) was used:

(Object recognition OR object detection OR target tracking)

AND

(Human body OR human bodies OR person OR people OR marine animal*)*

Restrictions for underwater situations are left out because these search parameters did not achieve the desired results.

The building block strategy was used, searching for part 1 and 2 of the search question and combine with AND. The results were filtered using exclusions of face and facial recognition, person identification, static situations and static images using the NOT operator.

After scanning the search results, the most relevant references were selected. From the abstracts the articles were categorised in the following categories:

Pedestrian detection
Human body detection in backgrounds
Victim detection
Human posture detection/recognition
(Other) Human body detection methods

3. DETECTING PEDESTRIANS

As accidents with pedestrians in traffic increases, there is a growing need for pedestrian detection systems which make sure that pedestrian crossings are empty before the traffic-lights for other traffic turns green. To solve this issue several techniques are proposed. Tani, H. et al [1] proposes to use space-time images made with a CCD camera to control the time green for the handicapped and the aged. “First, measurement areas are defined along the white line of the crosswalk. Next, space-time images are made such that a slip of image on the measurement area is arranged. Finally, pedestrians are detected by processing the image”. [1]

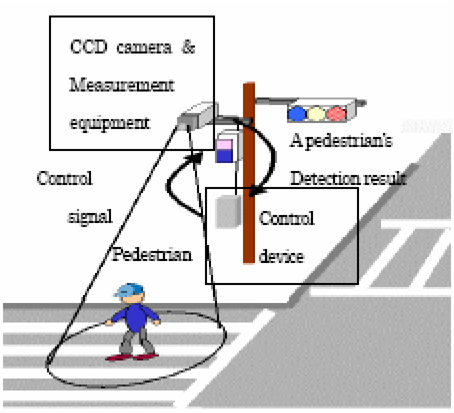
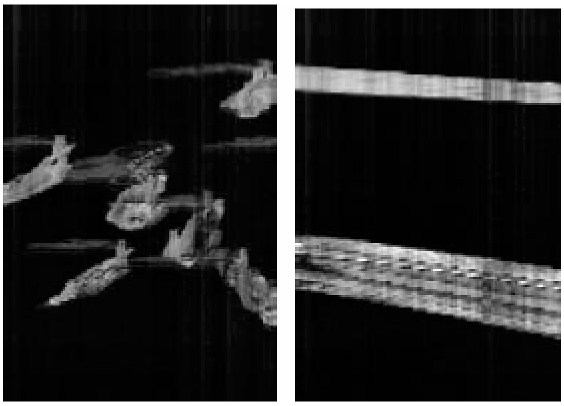


Fig.1. *overview of the system [1]*



pedestrian vehicle

Fig.2. *difference in space-time image [1]*

“The overview of our system is shown in figure 1. A CCD camera should be installed near a traffic light in order to overlook the whole crosswalk, and images can be taken per 100ms. Two measurement areas should be set on crosswalk in order to make our system robust. One should be set on dark lines, the other should be set on white lines. By using them, we can make space-time images. In addition, we can make use of difference between background images and input images, and can detect only passing objects. Space-time images of pedestrians and vehicles are shown in figure 2. We can extract some features, such as size, shape and texture pattern, from space-time images and detect pedestrians.
When something passes over a crosswalk, the colour of some blocks will change. We can detect it by size of blocks. It is also necessary to recognize passing objects as pedestrians or vehicles. In order to identify them, we use shape of blocks and the difference of edge. Usually, the shape of pedestrians on a space-time image is complex. While, the shape of vehicles on a space-time image is like a rectangle. In addition, pedestrians walk perpendicularly against white lines of crosswalk. On the other, vehicles move along white lines of crosswalk. Therefore, we can decide whether passing ones is pedestrians or not by the concavo-convex degree and texture pattern. Feature parameters are calculated every 1 second, pedestrians are detected. In our experiment, we took a picture of crosswalk for 1 hour.. As a result, the precision of detecting pedestrians was more than 99%. We demonstrated that our system is effective”. [1]

Another way to detect pedestrian is proposed by Muranaka, N. et al [2] "In recent years, traffic accident occurs frequently with the explosion of traffic density. Therefore, we think that the safe and comfortable transportation system to defend the pedestrian is necessary. Firstly, we detect and recognize the pedestrian (the crossing person) by the image processing. Next, we inform all the drivers of the right or left turn that the pedestrian exists by the sound and the image and so on. By prompting a driver to do safe driving in this way, the accident to the pedestrian can decrease."[2] In his paper a background subtraction method is proposed for the movement detection of the moving object.

The main reasons for errors in pedestrian counting according to Brauer-Burchardt, C. et al [3] is the counting of shadows and non separated persons. To prevent these errors he proposes an new robust way to detect walking persons in busy public areas by using vertically fixed monochromatic cameras. The main focus lies on shadow detection and object separation. "The main novelty of the method is a new approach to shadow detection. The information about existence and kind of shadows from the former images of the sequence is used to predict the properties of the shadows in the current image. Two different methods to detect weak and strong shadows are presented." [3]

4. Person detection in backgrounds

One of the techniques to detect human bodies is background subtraction. This often causes problems like with shadows and no separated persons (like Brauer-Burchardt describes in his paper). In order to detect persons in complex scenes, backgrounds, Maojun, Z et al [4] proposes a method which creates a panoramic image from the background before a person enters. Once someone comes in front of the background a detection process is started. "We propose an algorithm for determining the camera motion parameters, which is used to get the background image hidden by the persons from the panoramic mage. The current captured image and the background image are compared to detect the persons using the background subtraction algorithm based on logarithmic intensities. Experiments show that the proposed method can be real-time run on a high-performance personal computer."[4]

Baisheng, C. et al. [5] proposes a background model initiation and maintenance algorithm for video surveillance. "In order to detect foreground objects, firstly, the initial background scene is statically learned using the frequency of the pixel intensity values during training period. The frequency ratios of the intensity values for each pixel at the same position in the frames are calculated; the intensity values with the biggest ratios are incorporated to model the background scene. Secondly, a background maintenance model is also proposed to adapt to the scene changes, such as illumination changes (the sun being blocked by clouds, or illumination time-varying), extraneous events (a person stops walking and stay motionless, people getting out of a parked car, etc.). Finally, a three-stage method is performed to detect the foreground objects: thresholding, noise clearing and shadow removal. The experimental results demonstrate robustness and real-time performance of our algorithm". [5]

Another approach is proposed by Leo, M et al. [6] he focuses on the problem of people detection in outdoor environments with a static TV camera. He proposes a video surveillance system for outdoor environments, in this case the context of a park, which can highlight objects of interest and recognise behaviour like theft and violence. "An example-based learning technique to detect people in dynamic scenes has been proposed. The classification is purely based on the people shape and not on the image content. Adaptive background subtraction has been used for detecting the objects of interest, then geometric and statistical information extracted from the horizontal and vertical projections are used to represent people shape and, finally, a supervised three layer neural network has been used to classify the extracted patterns. The experiments show that both a single person and a group of people are correctly detected also when other moving objects are in the scene. People are not detected only when their blob is connected with the blob of a moving car that modifies widely the whole binary shape. In this case people detection from the binary shape is hard even for human eyes. In conclusion it possible to assert that the proposed method is robust, reliable, fast and it can be easily adapted for the detection of other moving objects in the scene. Future works will deal with the problem of gesture recognition of the detected people in order to individuate illegal behaviours such as thefts or damaging."[6]

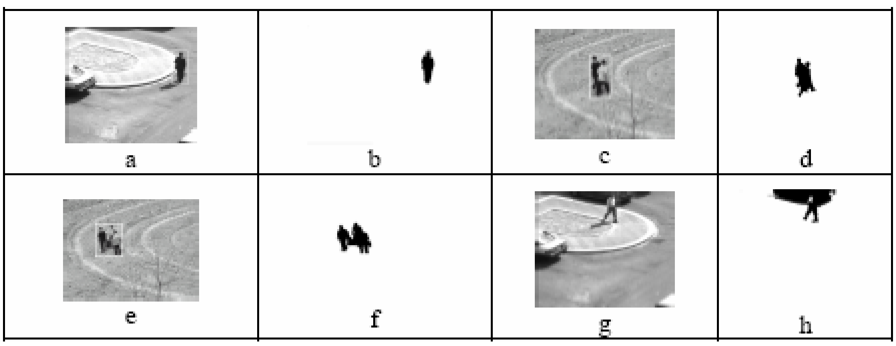


Fig.3. *a-f) Three examples of test images correctly processed; g-h) An example of test images wrongly processed. The people and the car move together and their binary shape are spoiled [6]*

5. Human posture detection / recognition

As M. Leo et al states in his paper: “Detecting people in images is more challenging than detecting other objects because people can be articulated in their shape and can assume a variety of postures, so it is nontrivial to define a single model that captures all these possibilities. Moreover, people wear different dresses of different colours and then the interclass variation in the people class can be very high, making difficult the recognition by using colour-based and fine edge-based techniques.” A simple and reliable approach for the estimation of human body postures is proposed by Spagnolo, P. et al [7] The applicative context is the visual surveillance of an archaeological site. “Motion detection and object recognition subsystems process image sequences coming from a still camera. Whenever a human is detected, his postures are characterized by the proposed pose estimation module. Then the results are fed to a HMM subsystem that identifies the current activity of the examined subject The proposed algorithm is based on an unsupervised clustering approach that makes the system substantially independent from any a-priori assumption about the possible output postures. The features selected for posture estimation are the horizontal and vertical histograms of binary shapes. A modified version of the Manhattan distance is used for both cluster identification and for run-time classification. After extensive experimental tests with different clustering schema, the BCLS algorithm (basic competitive learning scheme) has been selected. The proposed approach makes possible to change the number of classes, during the classification phase, without repeating the training step. Moreover it provides a measure of the reliability of its results. The proposed method has been verified on sequences acquired while typical illegal activities involved in stealing were simulated in a real archaeological site”. [7]

An method to solve the human silhouette tracking problem using 18 major human points is proposed by Panagiotakis, C. et al. [8] “We used: a simple 2D model for the human silhouette, a linear prediction technique for initializing major points search, geometry anthropometric constraints for determining the search area and colour measures for matching human body parts. In addition, we propose a method to solve the problem of human members recognition and 18 major human points detection using silhouette. This result can be used to initialize a human tracking algorithm for real time applications. Our main purpose is to develop a low computation cost algorithm, which can be used independently of camera motion. The output of the tracking algorithm is the position of 18 major human points and a 2D human body extraction. In cases of low quality imaging conditions or low background contrast, the result may be worst. For these cases we defined an appropriate criterion concerning tracking ability.” [8]

In the context of human-behaviour analysis Cucchiara, R et al. [9] proposes to analyze human behaviours by classifying the posture of the monitored person and, consequently, detect corresponding events and alarm situations, like a fall. “Our approach can be divided in two phases: for each frame, the projection histograms (Haritaoglu et al., 1998) of each person are computed and compared with the probabilistic projection maps stored for each posture during the training phase; then, the obtained posture is further validated exploiting the information extracted by a tracking module in order to take into account the reliability of the classification of the first phase. Moreover, the tracking algorithm is used to handle occlusions, making the system particularly robust even in indoors environments. Extensive experimental results demonstrate a promising average accuracy of more than 95% in correctly classifying human postures, even in the case of challenging conditions”. [9]

In order to solve the problem with detecting people in arbitrary poses Mori, G et al. [10] proposes a technique to detect a human figure image and localize his joints and limbs. "In this work we attempt to tackle this problem in a general setting. The dataset we use is a collection of sports news photographs of baseball players, varying dramatically in pose and clothing. The approach that we take is to use segmentation to guide our recognition algorithm to salient bits of the image. We use this segmentation approach to build limb and torso detectors, the outputs of which are assembled into human figures. We present quantitative results on torso localization, in addition to shortlisted full body configurations". [10]

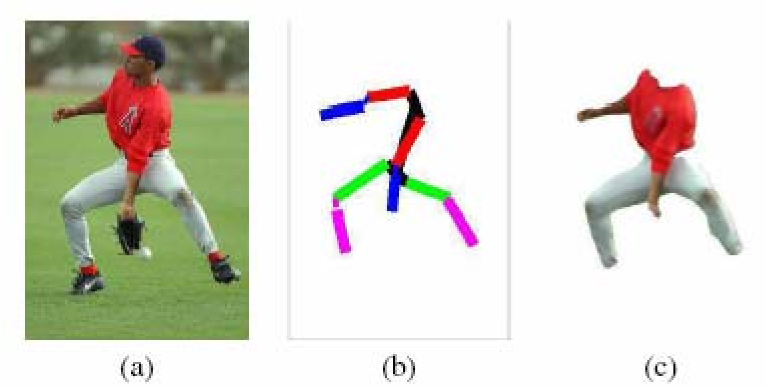


Fig.4. *The problem: (a) Input image. (b) Extracted skeleton of Localized joints and limbs. (c) Segmentation mask associated with human figure. [10]*

6. VICTIM DETECTION

Detecting victims buried under rubble is a lifesaving application. Banerjee, P.K. et al [11] proposes a design for a remote victim locator. It has been designed for locating remotely, any animate subject buried under rubble or behind a barrier. It works on the principle of Doppler effect at microwave frequencies. Electromagnetic signal travels down the debris from a transmit/receive antenna, illuminates the animate subject under the rubble and returns to the top of the debris having phase modulated with body movements. The antenna picks up the phase-modulated signal, which is processed by a microwave phase synchronous receiver. The detected signal is displayed on an oscilloscope or any other suitable display device. The display is either a regular breathing cycle or any sudden shift in its steady level, resulting from even a slight movement of any part of the body. An important part of the system is the clutter (unwanted reflections) canceller. Presently, the system can penetrate a dry concrete debris depth of , about 2 m. Detection depth may be greater if there are more air voids in the debris. Efforts are on to improve the depth of penetration in the rubble. The same system can detect presence of any person behind a barrier, which has security applications.

Skin detection or segmentation is employed in many tasks related to the detection and tracking of humans and human-body parts. However, skin detection is not robust enough for dealing with some real-world conditions, such as changing lighting conditions and complex backgrounds containing surfaces and objects with skin-like colours. This situation can be improved by incorporating context information in the skin detection process. For this reason, a skin detection approach that uses neighbourhood information is proposed. This idea is implemented through a diffusion process that allows a robust segmentation of skin regions at a high processing speed. Ruiz Del Solar, J. et al [12]

7. HUMAN BODY DETECTION

Beside the human body detection and recognition methods for specific situations or applications described above there is a wide variety of other possible approaches. These more general human body detection approaches are not further classified.

A problem with the use of a monaural camera is the obtained 2D data. This makes it difficult to detect a person's position, direction and shape. To solve this problem Terada, K. et al [13] proposes a method of observing the position, direction and shape of people by using 3D information acquired using FG vision sensor, which has been developed by the authors. Position information, direction information and height information are detected by extracted a cylinder from 3D data. In this process, Hough transform is used to detect position, X-axis and Y-axis of ellipse of a cylinder. The tracking is carried out using the position and the height information in real-time.

Orrite-Urunuela, C. et al [14] proposes a statistical model for detection and tracking of human silhouette and the corresponding 3D skeletal structure in gait sequences. We follow a point distribution model (PDM) approach using a principal component analysis (PCA). The problem of non-linear PCA is partially resolved by applying a different PDM depending of pose estimation; frontal, lateral and diagonal, estimated by Fisher's linear discriminant. Additionally, the fitting is carried out by selecting the closest allowable shape from the training set by means of a nearest neighbor classifier. To improve the performance of the model we develop a human gait analysis to take into account temporal dynamic to track the human body. The incorporation of temporal constraints on the model helps increase the reliability and robustness.

A low cost 16*16 un-cooled pyroelectric detector array, allied with advanced tracking and detection algorithms, has enabled the development of a universal detector with a wide range of applications in people monitoring and homeland security. Violation of access control systems, whether controlled by proximity card, biometrics, swipe card or similar, may occur by 'tailgating' or 'piggybacking' where an 'approved' entrant with a valid entry card is accompanied by a closely spaced 'non-approved' entrant. The violation may be under duress, where the accompanying person is attempting to enter a secure facility by force or threat. Alternatively, the violation may be benign where staff members collude either through habit or lassitude, either with each other or with third parties, without considering the security consequences. Examples of the latter could include schools, hospitals or maternity homes. The 16*16 pyroelectric array is integrated into a detector or imaging system which incorporates data processing, target extraction and decision making algorithms. The algorithms apply interpolation to the array output, allowing a higher level of resolution than might otherwise be expected from such a low resolution array. The pyroelectric detection principle means that the detection works in variable light conditions and even in complete darkness, if required. The algorithms can monitor the shape, form, temperature and number of persons in the scene and utilise this information to determine whether a violation has occurred or not. As people are seen as 'hot blobs' and are not individually recognisable, civil liberties are not infringed in the detection process. The output from the detector is a simple alarm signal which may act as input to the access control system as an alert or to trigger CCTV image display and storage. The applications for a tailgate detector can be demonstrated across many medium security applications where there are no physical means to prevent this type of security breach. Rimmer, A.N. et al [15]

We design and calibrate an efficient human detection system, capable of detecting and tracking a single person while minimizing the number of required sensors. Our infrared detector has a 2 meter range and a 250 ms rise/fall time. We sample their signal using an MSP430F149 microprocessor, correlate the results to reference patterns, and collect the binary decisions of each sensor on a workstation wirelessly. We modulate the optical field of the detectors by introducing selectively opaque reference structures into their optical path. Segmenting the physical space into signature cells enables direct measurement of source configuration. Brute force determination of the mapping is prohibitively expensive; we propose a method to estimate this mapping and predict signatures for every source position by observing signatures along prescribed tracks through the physical space. Burchett, J.B. et al [16]

Kettnaker, V. et al [17] present a system that detects people in indoor scenes by modelling the motion history of foreground blobs, rather than their shape or appearance. The system tracks all foreground blobs over time with a multi-hypothesis tracker, and considers a blob to be a person if it exhibited sufficient autonomous movement in the course of its tracking history. This way, people can be correctly classified even if they are seen in a wide range of body poses, if they remain still for a long time, or if they change appearance by taking off a coat. Evaluation on over 1h of video demonstrated good performance for both heuristic and decision tree based classification.

Observation of long sequences of video images in surveillance applications may encounter several problems due to camera motion or rotation, unexpected size and speed of objects, variation of colour due to sunshine and shadowy areas. Robust tracking algorithms are needed to compensate for the variations of different recording conditions. In this paper we evaluate the detection probability of our tracking algorithm with ROC curves and with synthetic degradation methods. Recorded experimental multi-sensor data is used to compare the accuracy in different spectral bands. Moving object detection in a guarded area can produce many false alarms due to the moving environment such as trees and bushes, birds and animals. By applying tracking and classification, false alarms can be reduced avoiding unnecessary recordings and preventing the displacement of guards. Track speed, size, direction and range (distance to camera) are calculated. The objects are classified roughly into classes as person, vehicle, and fast moving object or simply, as moving object. The results of the algorithm applied to the experimental data and the algorithm evaluation are presented. Romeo, K. et al [18]

8. Conclusions

The literature described above provides different ways of human body detection and recognition for several applications. Most approaches are video based which means that they work analysing video streams. This approach immediately causes a problem in the application I am dealing with. In this situation persons are most of the time half underwater. These camera based systems will not work underwater because of high water turbidity. Cameras might be replaced by infrared cameras and use similar image processing software as used in above water situations. Because humans are only partly underwater, human body part detecting and reconstructing systems as described by Mori, G [21], Panagiotakis, C [13] and Ruiz Del Solar [12] are most suitable. Other suitable possibilities are the use of pyroelectric sensors [15] and electromagnetic micro waves travelling trough the water and change in reflection characteristics when a human body is detected. [11]

9. References

[1] *Tani, H. et al. Surveillance of pedestrians on crosswalk by camera.* ***In*** *Transactions of the Institute of Electrical Engineers of Japan, Part-C. March 2004. p. 798-804.*

[2] *Muranaka, N. et al. The pedestrian detection method using an extension background subtraction about the driving safety support systems.* ***In*** *Transactions of the Institute of Electrical Engineers of Japan, Part C. Aug. 2004.*

[3] *Brauer-Burchardt, C. Pedestrian identification, tracking, and counting with shadow detection in monochromatic video streams.* ***In*** *IASTED International Conference on Modelling Identification and Control. 2003. P. 506-511. Available from World Wide Web: http://www.sciencedirect.com/science/article/B6WPV-4C2R1X4-JB/2/7931e1a1766cb05ecbe4c8e117a4ceba. Subscription needed.*

[4] *Maojun, Z. et al. A visual method for real-time detecting persons in complex scenes.* ***In*** *Proceedings of the 21st IEEE Instrumentation and Measurement Technology Conference. IEEE, Piscataway, NJ, USA. 2004. 438-40 Vol.1*

[5] *Baisheng, C. et al. A novel background model for real-time vehicle detection.* ***In*** *Yuan, B. et al (eds). 7th International Conference on Signal Processing Proceedings IEEE. IEEE, Piscataway, NJ, USA. 2004. 1276-9 vol.2*

[6] *Leo, M. et al. Shape based people detection for visual surveillance systems.* ***In*** *Audio and Video Based Biometric Person Authentication. 4th International Conference, AVBPA 2003. Proceedings Lecture Notes in Computer Science. Springer-Verlag, Berlin, Germany. 2003. Vol.2688*

[7] *Spagnolo, P. et al. Posture recognition in visual surveillance of archaeological sites.* ***In*** *Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems IROS. IEEE, Piscataway, NJ, USA. 2003. 1542-7 vol.2*

[8] *Panagiotakis, C. Tziritas, G. Recognition and tracking of the members of a moving human body.* ***In*** *Perales, F. J. Draper, B.A. (eds) Articulated Motion and Deformable Objects. Third International Workshop, AMDO 2004. Proceedings Lecture Notes in Computer. Sci. Springer-Verlag, Berlin, Germany. 2004. Vol.3179. p. 86-98.*

[9] *Cucchiara, R. Probabilistic posture classification for Human-behavior analysis.* ***In*** *IEEE Transactions on Systems, Man & Cybernetics, Part A Systems & Humans. Jan. 2005. p. 42-54.*

[10] *Mori, G. et al. Recovering human body configurations: combining segmentation and recognition.* ***In*** *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer. Soc, Los Alamitos, CA, USA. 2004. Vol.2*

[11] *Banerjee, P.K. Sengupta, A. Doppler technique for detection of victims buried under earthquake rubble.* ***In*** *Indian Journal of Pure and Applied Physics. CSIR. Dec. 2003.*

[12] *Ruiz Del Solar, J. Verschae, R. Robust skin segmentation using neighbourhood information.* ***In*** *International Conference on Image Processing ICIP2004. IEEE, Piscataway, NJ, USA. 2004. 207-10 Vol. 1*

[13] *Terada, K. Umemoto, T. Observing passing people by using fiber grating vision sensor.* ***In*** *Proceedings of the 2004 IEEE International Conference on Control Applications. IEEE, Piscataway, NJ, USA. 2004. 1112-17 Vol.2*

[14] *Orrite-Urunuela, C. et al. 2D silhouette and 3D skeletal models for human detection and tracking.* ***In*** *Proceedings of the 17th International Conference on Pattern Recognition. IEEE Comput. Soc, Los Alamitos, CA, USA. 2004. 244-7 Vol.4*

[15] *Rimmer, A.N. Access control violation prevention by low cost infrared detection.* ***In*** *Proceedings of the SPIE The International Society for Optical Engineering. SPIE-Int. Soc. Opt. Eng. April 2004.*

[16] *Burchett, J.B. et al. Efficient human pose tracking with multiplex sensing.* ***In*** *Proceedings of the SPIE The International Society for Optical Engineering. SPIE-Int. Soc. Opt. Eng. April 2004*

[17] *Kettnaker, V. Gahm, J.k. Closed-loop person tracking and detection.* ***In*** *Proceedings - 1st Canadian Conference on Computer and Robot Vision. 2004. P. 314-319. Available from World Wide Web: http://www.sciencedirect.com/science/article/B6WPV-4DG5VSD-YY/2/44703481e05857bd7f5ae03d109ef76e. Subscription needed.*

[18] *Romeo, K. et al. Evaluation of tracking in video sequences.* ***In*** *Proceedings of SPIE - The International Society for Optical Engineering. 2001. Vol. 4232. P. 464-471. Available from World Wide Web: http://www.sciencedirect.com/science/article/B6WPV-4436XCJ-GG/2/8a07724671a3cb1e2ae189aac098cd0b. Subscription needed.*