Strojniški vestnik - Journal of Mechanical Engineering 63(2017)5, 331-348

DOI:10.5545/sv-jme.2017.4324
Review Scientific Paper
Received for review: 2017-01-13
Received revised form: 2017-03-24
Accepted for publication: 2017-04-05

Machine-vision-based human-oriented mobile robots: A review

Miha Finžgar* – Primož Podržaj
University of Ljubljana, Faculty of Mechanical Engineering, Slovenia

In this paper we present a study of vision-based, human-recognition solutions in human-oriented, mobile-robot applications. Human recognition is composed of detection, tracking and identification. Here, we provide an analysis of each step. The applied vision systems can be conventional 2D, stereo or omnidirectional. The camera sensor can be designed to detect light in the visible or infrared parts of the electromagnetic spectrum. Regardless of the method or the type of sensor chosen, the best results in human recognition can be obtained by using a multimodal solution. In this case, the vision system is enhanced with other forms of sensory information. The most common sensors are laser range finders, microphones and sonars. As medicine is expected to be one of the main fields of application for mobile robots, we give it special emphasis. An overview of current applications and proposal of potential future applications are given. Without doubt, properly controlled mobile robots will play an ever-increasing role in the future of medicine.
Keywords: mobile robot, machine vision, human recognition, image-processing algorithms, medical applications, review, overview

Highlights
- *An overview of vision-based, human recognition solutions in human-oriented, mobile-robot applications is presented.*
- *A comparison of conventional 2D, stereo and omnidirectional vision systems is made.*
- *The advantages of multimodal systems over vision-only systems are described.*
- *A table of overviewed hardware and software solutions together with their performance is provided in order to make a comparison between various systems easier.*
- *The most promising and relevant mobile-robot applications from the field of medicine are discussed.*

0 INTRODUCTION

Mobile robots are receiving more and more attention because of their ability to move around in a real-world environment and to perform various tasks. These two characteristics make mobile robots suitable for use in numerous industrial and domestic applications (e.g., personal assistance, guidance, surveillance, transportation, cleaning). Furthermore, they can operate in environments that are hostile or even inaccessible to humans. Human-oriented mobile robots are becoming increasingly important since the need for health-based assistance is increasing for the growing number of elderly and/or chronically ill people. Mobile robots could offer assistance and reliable health monitoring and therefore improve people's quality of life. These robots could also be used in telemedicine, which would not only reduce the costs needed to travel to outpatient-based doctors and the number of missed working days, but also save patients' time [1]. On the other hand, even the mere presence of a mobile robot can have a positive effect on people's well being [1].

In general, robotic applications in healthcare and social care can be classified into two main groups [2]: traditional robots intended for (telerobotic) surgery and rehabilitation and robots supporting "softer" human-robot interaction tasks such as logistics, telepresence, companionship, education of children with special needs and motivational coaching. Mobile examples include HelpMate [3] used in the transportation of supplies to healthcare staff and PR7 [4] used for telecommunication. More recent mobile-robot applications are dealing with the assistance of elderly people [5], support for autism diagnosis and intervention [6] and assessments of people's physiological state [7]. However, the challenges remain. One of the most important is human-robot interaction (HRI). In order to make it as natural as possible and to ensure reliable execution of the mobile robots' tasks these systems should be autonomous, robust, fast, non-contact and, most importantly, safe. These characteristics are needed for unbiased, real-time measurements in different situations (occlusions, varying illumination, etc.). Moreover, it is necessary for the mobile robots to provide only the tasks for which they were built and not to keep people under surveillance and/or disturb their privacy [8]. Mobile robots' functions in general include 1) reaching the goal and 2) performing certain tasks. In order to reach the goal, quickly acquired, low-resolution data about the mobile robot's environment is needed, while for the task execution, more detailed data is required [9]. For mobile robots to be efficient during their interactions with humans, the implementation of machine vision is essential. Additionally, visual systems have an important role in medicine for diagnosing [10], screening and monitoring [11]. This

*Corr. Author's Address: University of Ljubljana, Faculty of Mechanical Engineering, Aškerčeva c. 6, SI-1000 Ljubljana, Slovenia, miha.finzgar@fs.uni-lj.si

allows mobile robots to not only recognize humans and avoid obstacles but also to perform certain healthcare tasks.

In order to assess the feasibility of implementation with respect to different visual systems for assessments of human physiology on mobile robots it is of great importance to review the proposed vision-based, human-recognition solutions in human-oriented mobile robots. In the first section, each step of the human recognition in mobile robots is addressed. In the second section an overview of the applied sensor modalities that offer human recognition are presented. This section is divided into two parts: the first one presents vision-only-based solutions, while the second one presents multimodal solutions. The third section presents the proposed applications of human-oriented mobile robots. The fourth section includes a short discussion, with the authors' views on the challenges and available solutions regarding the implementation of mobile robots in any real-world environment, with an emphasis on performing health-care tasks.

1 HUMAN RECOGNITION

The crucial characteristic of mobile robots needing to work in a human environment is the ability to recognise people. This is important for safety reasons, the successful performance of the mobile robots' tasks and a natural HRI. Human recognition consists of 3 basic steps [12]: detection, tracking (localisation) and identification.

Humans can be detected with vision-based, invisible-band, sensor-based and sensor fusion-based approaches [13]. It can involve image-background subtraction in cases when the mobile robot (together with the camera(s) mounted on it) is not moving [14], but most commonly it employs colour-based feature detection, shape- or model-based approaches and machine-learning-based approaches. Colour-based approaches use the predefined colour models of human skin and fit the pixels to these models. These approaches are fast, but are at the same time susceptible to illumination variation and changes to a person's position relative to the mobile robot's vision system [15]. Additionally, they provide false-positive detections due to skin-coloured, static background regions or objects [16]. In order to differentiate between real skin and a skin-like coloured object, the size of the detected object and the object's width-to-height ratio can be used [14]. Model-based approaches use various parameters that describe the shape and/or motion of the target. They are usually computationally more demanding and require constraining the dynamics of the system, but provide additional pieces of information regarding the tracked object's position and the correspondence of specific parts of the tracked objects with the actual image. The speed and accuracy of these approaches are affected by the initial conditions and any feature variances [15]. Additionally, the detection of facial features strongly depends on the size of the entire face blob [17]. When it comes to face detection algorithms, interested reader can refer to the article written by Zafeiriou et al. [18]. It offers a thorough description of face detection algorithms together with their comparison, it presents benchmarks and evaluation metrics and it also discusses future challenges in the field. Machine-learning-based approaches require a predefined set of images and their computational cost increases with increasing image/video resolution [15].

Various approaches can be used for the tracking: mean-shift and its variant the continuously adaptive mean-shift algorithm (CamShift), optical flow (e.g., Lucas-Kanade method), particle filters, Kalman filters, multiple hypothesis tracking, etc. The performance of tracking approaches depends on the environment. For example, CamShift can provide false tracking if the background colour is similar to the target's colour [19] or when the illumination is too intense [20]. The Lucas-Kanade approach has difficulties in cases where there is a lot of movement near the target or when the target moves too far from its initial position [20]. Additionally, if the mobile robot is not moving there are no cues for optical flow computation. This disadvantage can be overcome using a motion-dependent approach [21] in which background subtraction is performed when the robot is not moving, with an optical-flow-based approach being used otherwise (the switching is based on the processing load and the mobile robot's movement). A particle filter might offer the best results in comparison to the Lucas-Kanade and CamShift approaches due to its non-parametricity and robustness to background colour distribution and movements in the background [20]. In the case of multiple humans a set of independent particle filters can be used (each human appearing in the scene for the first time is detected, while previously detected ones continue to be tracked) [22]. An important characteristic of independent tracking filters is their computational efficiency, which allows real-time tracking, but their performance deteriorates if the tracked objects are too close to each other [22]. The cost effectiveness of the Kalman filter can be increased by applying the filter only to the region of interest (ROI) and not over the entire image [23]. Kalman filters are constrained by linear or

Gaussian assumptions, whereas particle filters are not. However they are, in general, computationally more demanding [24]. The ground truth for tracking can be obtained by using ceiling-mounted cameras ([24] and [25]) or by manual annotation [26]. The tracking is more successful if the motion of the moving target is predicted [27]; the state-estimation problem can be solved in a 2D image plane [28]. Classic approaches to human recognition deal separately with detection and tracking, which results in the potential loss of information as well as in an increased computational load [29]. The other approaches, known as track-before-detect or unified tracking [30], deal with detection and tracking simultaneously.

In general, human identification can be most successfully achieved using biometrics, which is based on measurements of physiological and behavioural characteristics [31]. Human identification by mobile robots is, however, based mostly on colour features [12], texture features or combinations of both [32]. The latter offers high human-recognition rates as well as real-time performance in crowded environments; a combination of features offers better results than feature used on its own [32]. An example of the colour-feature approach is a comparison of the colour histograms of people's clothes [24]. However, this approach does not work in environments in which human apparel is identical (e.g., in industrial and healthcare environments) and it does not offer an instant daily identification of a single person, since human apparel is usually changing daily. On the other hand, this approach can be useful for emergency personnel following applications [16].

Human recognition in real-world applications is very challenging due to varying illumination, varying appearances, directions and behaviours of humans, background variations, limited time for computation [13] and vibrations of the mobile robot [33], which are due to uneven floors or mobile-robot construction deficiencies. In the outdoor environment, there are additional challenges, such as weather and terrain diversity [34]. Often, the robust solutions for the aforementioned challenges are not applicable to mobile robots. For example, occlusions can be easily solved by implementing a camera mounted above the observed environment [35] or by using multiple cameras [36], which is not always possible and/or desirable. The effect of occlusions on the correct tracking rate in mobile robots is therefore handled by applying kinematic models to each tracked object (implicit solution) [37] or by predicting human behaviour in detected occlusions (explicit solution) [22]. In order to achieve the best human recognition possible it is important to choose an appropriate number of features, while at the same time achieving real-time execution of the algorithms programmed in the mobile robots.

Table 1. *Reviewed literature*

Research	Hardware	Studied environment	Applied algorithms	Algorithms' performance
Böhme et al. [14]	Extended B21 RW1, IS Robotics with IR layer, 2 sonar sensors, omnidirectional camera, two frontally aligned colour cameras, binaural auditory system $f_{system} = 0.5$ Hz	indoor environment (home store)	• updated motion-based foreground-background segmentation [37] • binaural sound localization based on inter-aural time differences and spikes [38] and [39] • upper body contour modelling • skin colour (dichromatic r-g) detection [40] • CCNNW-based face detection [41] using public data set [42] • dynamic neural field for final selection • tracking: condensation algorithm [43]	no quantitative evaluation
Fritsch et al. [44]	Bielefeld robot companion – BIRON (based on ActiveMedia Pioneer PeopleBot) with Sony EVI-D31 PT camera, two AKG far-field microphones and a SICK LRF $f_{system} = 5$ Hz $f_{microphone} = 5.5$ Hz $f_{LRF} = 4.7$ Hz	indoor environment (office); robot is tracking a target human, who at one point is turned away from the robot, is not speaking and his legs are occluded	• face detection (Viola Jones) [45] • mixture of Gaussians-based colour (LUV) representation for torso recognition • Cross-Power Spectrum Phase Analysis based sound source localization [46] • leg detection [47] • custom simple cue fusion • multi-modal anchoring for data fusion extended by supervising module [47] • attention system for focusing the mobile robot on target human [48]	tracking: $SR \approx 80$ %

Cielniak and Duckett [12]	Pan-tilt colour camera Canon VC-C4R, IR camera NEC Thermal Tracer TS7302 $f_{camera} = 5$ Hz	indoor environment (office corridor); mobile robot was following the corridor with 10 walking humans;	• thresholding and connectivity- plus size-based segmentation for human detection (thermal images) • temperature and colour (HSV) statistics (first two moments) • human identification: k-NN classifier, Bayes' classifier and dynamic identification	classification (dynamic version of Bayes' classifier): $SR_{thermal} = 69.84$ % $SR_{colour} = 89.42$ % $SR_{combination} = 94.04$ %
Wilhelm et al. [49]	mobile robot B21 RWI, IS Robotics with omnidirectional camera Sony DWW VL500, 24 sonar sensors in 2 layers, two frontal cameras on PT unit	indoor environment (home store)	• skin colour (dichromatic r-g) detection using look up table with manually classified colour pixels [50] • automatic white-balance algorithm for colour calibration • sonar-based distance measurements • face detection (Viola-Jones) [45] • tracking: condensation algorithm [43]	no quantitative evaluation
Treptow et al. [51]	ActivMedia PeopleBot mobile robot with NEC Thermal Tracer TS730 $f_{camera} = 15$ Hz	indoor environment (unconstrained corridor and laboratory room) with 18 different humans; 1) person following; 2) corridor following; 3) stationary robot	• elliptic contour model for human detection similar to [43] • integral image features model based on Viola-Jones approach [45] • cascaded model evaluation for combining both models • tracking: set of independent particle filters (multiple humans)	tracking (single human): $ACC_{object\ count} \approx 92$ % $ACC_{object\ area} \approx 78$ % tracking (multiple humans): $ACC_{object\ count} \approx 84$ % $ACC_{object\ area} \approx 64$ % All results are for the combination of contour and feature-based model
Martin et al. [25]	Home Robot System – HOROS with fish-eye omnidirectional camera, SICK LRF and 16 sonar sensors Pentium M 1.6 GHz $f_{sonar} = 10$ Hz $f_{LRF} = 10$ Hz $f_{camera} = 7$ Hz	indoor environment (hallway); people walking past the mobile robot performing survey task	• heuristic method for detection of leg-pairs using LRF scans [47] • distance measurements on sonar scans of leg profiles • skin-colour (dichromatic r–g) detection [49] • tracking: condensation algorithm [43]	tracking: $SR = 93$ % $FPR = 25$ % $CPU\ load =$ $=$(40 to 50) %
Kim and Suga [33]	wheelchair mobile robot with omnidirectional camera $f_{camera} = 15$ fps	indoor environment (undefined place with undefined moving object and humans)	• expansion of grayscale omnidirectional image into panoramic • Lucas-Kanade optical flow method [52] • estimation of FOE and FOC • detection of moving objects using the evaluation value	tracking: $ERR_{OF\ min} = 2.15$ % (rotation) $ERR_{OF\ max} = 3.86$ % (right turn)
Chang et al. [53]	Kondo KHR-1 with webcam $f_{camera} = 3$ fps	indoor environment (office?); 4 individual humans	• skin-colour (dichromatic r-g) detection • hand-shape recognition (Hu moment invariants [54]) • tracking: active contour model with mean-shift, active contour model only	hand shape recognition: $SR = 96.8$ %
Vadakkepat et al. [55]	Magellan Pro with Sony EVI-D30 pan-tilt camera, 16 sonar and tactile sensors Pentium II	indoor environment (office?); 6 individual humans	• skin-colour (YCbCr and YUV) features and geometry features for face detection • tracking: CamShift [56] (HSV colour space)	tracking: $SR = 84.2$ % (YCbCr colour space) $SR = 89.8$ % (UV colour space)
Bellotto and Hu [24]	Pioneer mobile robot with PTZ colour camera and SICK LRF Pentium III 800 MHz, 128 MB RAM $f_{LRF} = 5$ Hz $f_{camera} = 10$ Hz $f_{system} = 5$ Hz	indoor environment (laboratory, corridor, office); 1) human following through different rooms, 2) 3 humans walking in front of the mobile robot or hiding	• leg detection based on the recognition of their typical patterns • Viola-Jones' face detection [57] • state prediction model [58] based on CV model • tracking: unscented Kalman filter [59] • human identification: comparison of colour histograms of human clothes [60] and NN data (multiple humans)	no quantitative evaluation

Fernandez-Caballero et al. [21]	MoviRobotics mSecurit mobile robot with thermal IR camera, PTZ dome camera, ultrasound sensors Intel Celeron M 600 MHz, 1 GB RAM $f_{IR\ camera}$ = (5 to 6) fps f_{system} = 6 Hz	indoor (undefined) and outdoor (?) environment plus chosen images from OTCBVS data set [61]	• normalization and thresholding for human candidates' blob detection • image subtraction or Lucas-Kanade optical flow method [52] (depending on mobile robot's motion)	human detection (image subtraction approach): TPR = (83.09 to 90.94) %; PR = (98.62 to 100) % (robot acquired images) TPR = (72.52 to 82.54) %; PR = (98.62 to 100) % (OTCBVS data set) human detection (optical flow approach): TPR = (79.62 to 98.57) %; PR = (94.59 to 100) % (robot acquired images)
Cielniak et al. [22]	ActivMedia PeopleBot mobile robot with PTZ camera Canon VC-C4R, NEC Thermal Tracer TS7302 0.85 GHz (robot); 2.00 GHz (PC) f_{camera} = 15 Hz (both cameras)	indoor environment (a corridor and a laboratory room); 1) humans walking in front of the 1.1) non-moving robot and 2.1) moving robot	• elliptic contour model adaptive colour (RGB) model based on the first three moments of colour distribution [62] • classification algorithm occlusion handling: combination of thermal and colour features determined by AdaBoost [63] • (used in occlusion handling) • tracking: particle filter/set of independent particle filters	occlusion classification: $SR_{thermal}$ = (76.4±4.5) % SR_{colour} = (69.0±1.9) % $SR_{combination}$ = (89.4±2.5) % t_{p_robot} = 68.8 ms t_{p_PC} = 25.9 ms (for 1000 samples using colour representation using first three moments)
Kang et al. [13]	Dasa Robot Tetra-DS with Point Grey LadyBug2 camera Intel Core2 Quad Q9400 2.66 GHz, 4 GB RAM, Nvidia GTX460 f_{camera} = 30 fps f_{system} = (9.3 to 12.8) fps	outdoor environment; moving and non-moving humans (occlusions, different positions relative to the robot), cars and other objects	• combined local-global optical flow method (based on global Horn's approach [64] and Lucas-Kanade approach [52]) computed on GPU • detection of ROIs using parallax flow • shape-based human detection (based on parallax flow estimation, Chamfer distance and HOG-based SVM classifier)	human detection: $TPR_{mid\ range}$ = 69.14 % $TPR_{near\ range}$ = 98.36 % t_p = (78 to 108) ms
Alvarez-Santos [32]	Pioneer P3DX mobile robot, SICK-LMS200 laser scanner, PointGrey Chameleon CMLN-1352C with Fujinon Fujifilm Vari-focal CCTV lens Intel Core2 Duo P8600 (2.4 GHz), 4 GB RAM	indoor environment (office); 1) varying lightning conditions (20-400 lx, shadows, reflections), 2) walking humans trying to distract the mobile robot system; indoor environment (museum): crowded, various light sources, reflective and uneven floor	• human detector based on HOG [65] • torso detection based on the extensive initial pool of colour (H1L1S1, L2AB, YCbCr, H2S2V, greyscale) and texture features (local binary patterns: classic [66], census [67], centre-symmetric [68] and semantic [69], edge density based on Canny edge detector [70], HOG [65], MPEG-7 edge histogram inspired descriptor [71]) • leg detection using laser scanner • sensor fusion at the tracker stage	F_{max} = 0.987 (combinations of 8 features) t_p = 20 ms
Baltzakis et al. [15]	no information 2.8 GHz, 4 GB RAM f_{system} = 16 Hz (maximum value)	indoor environment (office); 1) a single human and 2) multiple humans; indoor environment (exhibition centre): human and tour-guide robot interaction scenario	• background subtraction (28) and skin-colour (YUV) detection (face, hand) based on [72] and [73] • tracking: propagated pixel hypotheses algorithm [74] extended by an incremental probabilistic classifier • Viola-Jones' boosted cascade detector [57] (facial features) with anthropometric constraints • tracking: feature-based;(eyes, mouth) using normalized cross-correlation as similarity measure	tracking: TPR_{max} = 95.09 %; FPR = 0.22 % (mouth) TPR_{max} = 95.58 %; FPR = 0 % (left eye) TPR_{max} = 93.30 %; FPR = 0.22 % (righty eye)

Fotiadis et al. [34]	Robotnik Summit XL with a pan-tilt zoom camera Point Grey Firefly MV (60 Hz) and a LRF Hokuyo UTM-30LX-EW f_{LRF} = 40 Hz f_{camera} = 60 fps f_{system} = 8 Hz	indoor environment (gymnasium); 1) one human passed by non-moving mobile robot 2) random number of humans passed by moving mobile robot; outdoor environment: humans walking in front of moving mobile robot	• jumping distance segmentation, novel feature set for feature extraction and real AdaBoost classifier (LRF data) • HOG descriptor [65] with SVM classifier [75] (visual data)	indoor: ACC_{max} = 99.88 % TPR_{max} = 95.20 % TNR_{max} = 99.76 % (all Bayesian/mean fusion + adaptive projection) outdoor: ACC_{max} = 99.63 % (Bayesian/mean fusion + fixed-size projection) TPR_{max} = 93.01 % (maximum fusion; adaptive projection) TNR_{max} = 99.99 % (Bayesian/mean fusion + fixed-size projection)
Zhang et al. [26]	Pioneer 3-DX mobile robot with ASUS Xtion Pro Live RGB-D camera Intel Core i7 2.0 GHz (quad core) and 4 GB RAM (DDR3) f_{system} = (7 to 15) fps	indoor environment (laboratory); 1) humans walking with simple trajectories, 2) humans lifting humanoid robots and putting them away, 3) humans picking up objects, exchanging them and delivering them to other rooms	• separation of candidate point clusters using RANSAC [76] guided with prior-knowledge • candidate detection based on DOI • cascade of detectors [77] using height-, size-, surface- and HOG-based detector • DAG-based framework (human object classification [78], data association, matching, tracking: extended Kalman filter [79])	multiple object tracking: ACC_{max} = 95.39 % FNR_{min} = 2.77 % FPR_{min} = 1.10 %
Susperregi et al. [16]	RMP Segway mobile platform with Kinect (4 Hz), Heimann HTPA thermal sensor and Hokuyo UTM-30LX laser f_{Kinect} = 30 fps f_{system} = 4 Hz	indoor environment (museum); various lightning conditions, people naturally walking in front of the camera	• leg detection [80] • colour-based (RGB) vest detection • temperature-based human detection (thermal vision) • tracking: SIR particle filter [81]	$ERR_{estimation\ min}$ = (17.44 ± 22.54)° $ERR_{estimation\ min}$ = (0.10 ± 0.31) m (both results are obtained with following weighted combination of sensory data: 0.15× leg detection, 0.7× vest detection, 0.15× thermal detection)
Susperregi et al. [82]	RMP Segway mobile platform with Kinect, Heimann HTPA thermal sensor and Hokuyo UTM-30LX laser f_{Kinect} = 30 fps f_{system} = 1 Hz	indoor environment (manufacturing shop floor and museum); varying illumination conditions and human-like objects	• 23 different image transformations • 5 supervised machine-learning approaches (IB1 [83], Naïve-Bayes [84], Bayesian network [85], C4.5 [86], SVM) with hierarchical classifier [87]	human detection: ACC = 96.74 % FPR = 4.64 % FNR = 1.88 % PR= 95.36 % TPR = 98.07 %
Petrović et al. [23]	unknown mobile robot with Point Grey Bumblebee XB3 f_{camera} = 12 Hz f_{system} = 4 Hz	indoor (office) and outdoor (meadow) environment; a single human walking in front of the robot	• disparity map segmentation (connected pixel labelling [88]) • feature-based (2D - Hu moment invariants [54], 3D – object's height and width) object classification • tracking: modified Kalman filter	no quantitative data t_p = 81 ms t_{lat} = 100 ms (both results for case of sequential processing + distributed computing)
Mehdi et al. [8]	Autonomous Robot for Transport and Service - ARTOS, with LRF range finder, RFID reader, PTZ camera, sonar and tactile sensors	simulated indoor environment (apartment); a single human	• MDP-based human search similar to [89] and [90] • face detection using Haar cascade classifier [91] standing posture detection based on HOG [65])	no relevant quantitative data
Ćirić et al. [92]	DaNi mobile robot with FLIR E50 thermal camera 400 MHz, 128 MB RAM (embedded operation) + 256 MB RAM (storage) f_{camera} = 60 Hz	indoor environment (an unconstrained corridor and a hall); humans walking in front of the robot during 1) corridor following, 2) person following, 3) non-moving robot	• thermal image threshold segmentation optimized with genetic algorithm • feature detection (Hu moment invariants [54]) • SVM classification [75]	classification: SR = 97.3 %

Hu et al. [93]	Pioneer3DX mobile robot with RGB-D camera, LRF and sonar sensor f_{camera} = 30 fps	indoor environment (T-shaped corridor); 8 different humans leading the robot from starting to the end position	• human leg tracking using adaptive breakpoint detector [94] • estimation of human walk model • tracking: modified mean-shift [95] with depth information	anticipative front following: t = (54.2 ± 19.6) s d_{robot} = 16.94 ± 3.80 m d_{human} = (21.29 ± 6.83) m passive front following: t = (49.1 ± 26.0) s d_{robot} = (16.41 ± 3.81) m d_{human} = (19.49 ± 5.71) m combined strategy: t = (33.7 ± 4.5) s d_{robot} = (15.06 ± 1.51) m d_{human} = (17.40 ± 0.80) m (no significant differences between anticipative and passive back and side following)
Ali et al. [20]	Pioneer 3AT mobile robot with stereo camera and 16 sonar sensors Intel Core i3 2.4GHz, 4 GB RAM	indoor (laboratory) and outdoor environment (corridor, natural environment); multiple humans	• Haar-based human upper-body and face detection [91] • manual selection of target person • tracking: CamShift [56], Lucas-Kanade [52], particle filter [96] • Kalman-filter-based estimation and correction • stereo correspondence and linear triangulation (for target positioning)	$t_{p\_meanshift}$ = 0.02649 s t_{p_LK} = 0.02712 s t_{p_PT} = 0.02891 s
Bayram et al. [19]	modified Turtlebot II mobile robot with two Kinect modules, microphone array 2 netbooks with Intel Celeron 1.5GHz (dual core), 4 GB RAM f_{Kinect} = 10 fps $f_{microphone}$ = 100 Hz	indoor environment (laboratory); 1) moving humans with changing face direction towards the mobile robot; various lightning conditions and background variation (vision only) 2) one human, moving and speaking simultaneously 2) two humans speaking 2.1) with each other and 2.2) independently (vision and audio) 3) speaking human not present in robot's eye-sight (audio only)	• GEVD-MUSIC [97] • face detection • eyes detection based on Haar cascade classifier • skin-color (YCbCr) detection [98] • particle-filter based sensor fusion • tracking: CamShift [99]	face detection: PR = 98 % TPR = 94 % audio-visual human tracking: $ERR_{localization}$ = 1.86° (one human) $ERR_{localization}$ = 1.40° (two humans)

Legend: ACC – accuracy [%], ACC_{max} – maximum accuracy [%], $ACC_{object\ area}$ – accuracy of object area [%], $ACC_{object\ count}$ – accuracy of object count [%], AdaBoost – adaptive boosting, CCNNW – cascade-correlation neural network, CPU – central processing unit, CV – constant velocity, DAG – directed acyclic graph, DOI – depth of interest, d_{human} – total human displacement [m], d_{robot} – total robot displacement [m], $ERR_{localization}$ – target localization error, $ERR_{OF\ max}$ – maximum error percentage (ratio of true and detected optical flow), $ERR_{OF\ min}$ – minimum error percentage (ratio of true and detected optical flow), f_{camera} – camera frame rate, f_{Kinect} – sampling rate Kinect sensor(s), f_{LRF} – sampling rate of LRFs, F_{max} – maximum F measure, $f_{microphone}$ – sampling rate of microphone(s), FNR – false negative rate [%], FNR_{min} – minimum false negative rate [%], FPR – false positive rate [%], FPR_{min} – minimum false positive rate [%], f_{sonar} – sampling rate of sonar sensor(s), f_{system} – frequency of the entire system, GPU – graphics processing unit, HOG – histogram of oriented gradients, IB1 – instance based algorithm 1, IR – infrared, k-NN – k-nearest neighbour; LRF – laser range finder, MDP – Markov decision process, NN – nearest neighbour, OTCBVS – Object tracking and classification beyond the visible spectrum, PR – precision [%], PT – pan-tilt, PTZ – pan-tilt-zoom, RANSAC - random sample consensus, RFID - radio frequency identification, SN – sensitivity [%], SR – success rate [%], SR_{colour} – success rate using colour features [%], $SR_{combination}$ – success rate using combination of thermal and colour features [%], $SR_{thermal}$ – success rate using thermal features [%], SVM – support vector machine, t – time [s], t_{lat} – latency [ms], TNR_{max} – maximum true negative rate [%], t_p – processing time [ms], t_{p_LK} – processing time for Lucas-Kanade algorithm [ms], $t_{p\_meanshift}$ – processing time for mean-shift algorithm [ms], t_{p_PC} – processing time on personal computer [ms], t_{p_PT} – processing time for particle filter algorithm [ms], t_{p_robot} – processing time on mobile robot computer [ms], TPR – true positive rate [%], TPR_{max} – maximum true positive rate [%]

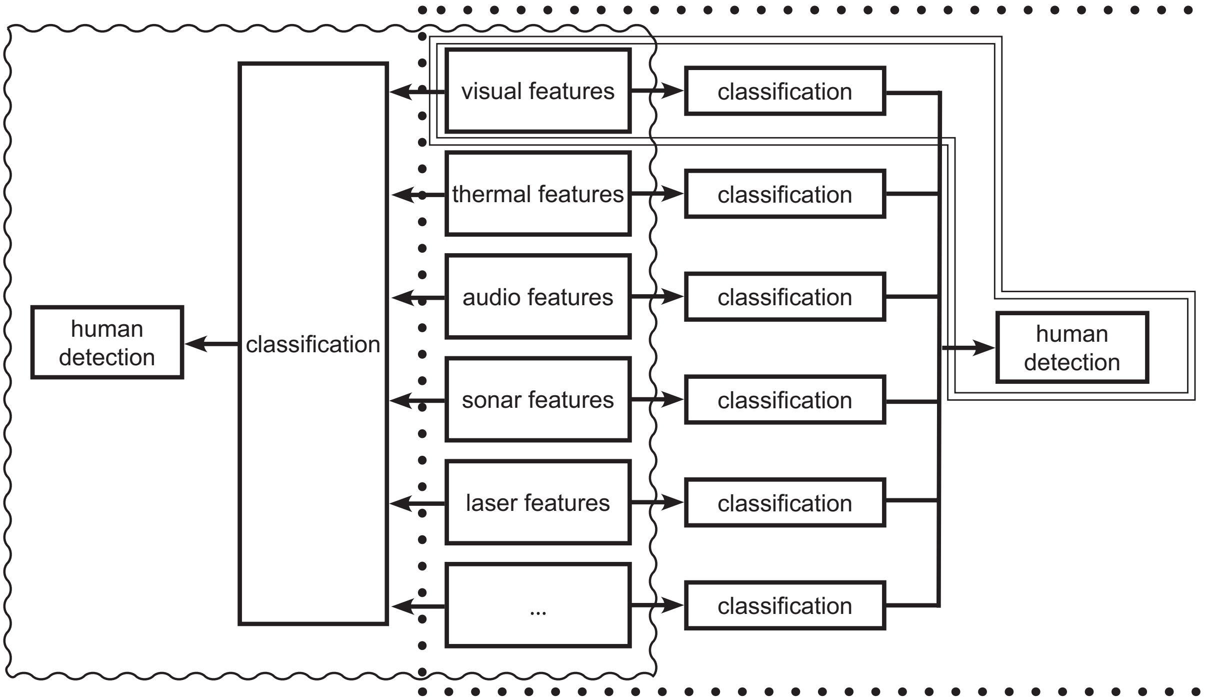

Fig. 1. *A block diagram representing basic difference between vision-only-based system (double-thin-lined shape) and multimodal systems (wavy and dotted shapes correspond to two different kinds of fusion level) shown on the example of human detection task*

2 SENSOR MODALITIES IN HUMAN-ORIENTED MOBILE ROBOTS

Machine-vision-based, human-oriented mobile robots can be either vision-only or multimodal. Table 1 offers information about the hardware and software solutions of these systems together with their performance and environmental settings.

2.1 Vision-Only-Based Systems

Vision-only-based mobile-robot systems are composed of colour vision, thermal vision or a combination of the two. In colour-vision systems, the following approaches have been implemented: conventional 2D vision, stereo vision and omnidirectional vision.

Colour vision offers robustness to geometric distortions [12], but it is susceptible to lighting variations [12] and resolution [32]. Solutions for reducing the sensitivity to illumination variations include the application of alternative colour spaces: HSV, dichromatic r-g, YUV, YCbCr and LUV. In general, any colour space that offers separate brightness and colour information can be used [49]. However, the employment of an alternative colour space may still not guarantee successful human detection. Therefore, Baltzakis et al. [15] applied prior-probability-based skin-colour detection [73], which adapts to the illumination changes. On the other hand, Wilhelm et al. [49] dealt with varying illumination by applying an automatic white-balance algorithm to the captured images in YUV colour space (a coated aluminium ring was used as a white reference). Additionally, the mean Y value was used for maintaining a brightness value of about 80 % of the maximum by controlling the camera's iris.

Conventional 2D vision lacks information about the object's location and/or its size. On the other hand, stereo systems offer additional depth information, which, however, raises the computational load due to the need for an accurate stereo correspondence. Stereo systems are less susceptible to different positions of people relative to the cameras and work even in short occlusions [100]. Depth information also offers smoother tracking by adjusting the mobile robot's speed to keep the distance to the target fixed [23]. Omnidirectional vision can be performed using various lenses, which define the characteristics of the acquired images. For example, images acquired using an optical system with a fish-eye lens have poorer resolution in the peripheral region when compared to the centre region and a perspective image [9]. Images taken with omnidirectional cameras also do not provide accurate distances between the target objects and the mobile robot, but only the angle of detection [25]. Furthermore, the optical flow pattern in omnidirectional images differs from the pattern in perspective images. This can be solved by transforming an omnidirectional image into a panoramic image [33]. An important advantage of omnidirectional vision is its ability to visualize a broader field of view for the mobile robot.

Thermal vision is based on thermal infrared (IR) video cameras, which detect emitted thermal energy in IR spectrum. Therefore, the pixels in the thermal images correspond to the temperature values. Due to the distinctive thermal profile of humans, their detection is simplified (no need for environment mapping and/or creating background models [12]). Additionally, temperature features are not susceptible to lighting variations, which offers visualization even in darkness and robustness to the direction of the human relative to the mobile robot. Important drawbacks include phantom detections, hard differentiation between humans, varying thermal characteristics of the airflow and the dependency of multiple-person tracking on their mutual position [51]. Human detection in thermal images can be performed by simple thresholding [21], whether by using a single threshold value or defining the optimal value using a genetic algorithm as in [101]. Treptow et al. [51] improved thermal-vision-based human detection by proposing an elliptical contour model (one ellipse for body position and one for head position).

Vision systems can also be mounted on the ceiling and not on the mobile robots. This approach can be used in the vision-only-based control of mobile robots [102], when obtaining the ground truth for tracking (as mentioned in Section 1) or in numerous Intelligent Space (iSpace) applications. iSpace is a term that refers to a space equipped with sensors and actuators, which provide an understanding of people's behaviour as well as providing them with information [103]. iSpace additionally controls electrically connected systems and robots in order to provide a particular service for people [103]. It also enables mobile robots to perform human-oriented operations, without having their own sensors and intelligence. Research regarding mobile robots in iSpace is not discussed in this manuscript.

2.2 Multimodal Systems

A single-sensor system cannot usually offer robust human tracking. It has been suggested that the most complete system for human recognition should be multimodal [22], since the integration of multiple sensory channels can improve a mobile robot's performance [19]. This is mostly achieved by overcoming the limitations of each individual sensor. For example, in leg detection using laser range finders (LRF) false positives due to leg-like-shaped objects (e.g., table or chair legs) often occur. On the other hand, false negatives appear if a human stands sideways relative to the robot's position, is wearing clothes that hide their legs [25] or his/her legs are occluded. Sonar sensors are usually noisy, inaccurate and unreliable (highly dependent on the distance between the mobile robot and the target human) [25]. However, in sudden illumination changes sonar sensors might be able to detect a human, in contrast to the colour-vision-based approaches. They can also aid colour vision in differentiating between human faces and potential skin-coloured objects positioned behind the human faces (data from sonar sensors is used to modify the weights of the skin-colour detector) [49]. Auditory modality is very susceptible to noise [19], which can disturb the detection and localization of sound sources of interest, but in combination with visual information human identification can be improved (if the person is not in the mobile robot's field of view [19]). The presence of an additional camera in an omnidirectional-based system offers verification of whether the detected object in the image obtained with the omnidirectional camera is really a human [49]. Fig. 1 shows basic difference between vision-only-based and general multimodal systems.

Since multimodality can improve a mobile robot's performance, the majority of the reviewed literature proposed multimodal systems (see Table 1). They include different combinations of visual systems, LRF, sonar sensors and/or microphones. Some of the research even used Kinect (Microsoft Corporation, USA), since it consists of multiple sensing devices. These include RGB cameras, 3D depth sensors (IR laser and monochrome CMOS sensor) and multi-array microphones, all mounted on a motorized tilt. Kinect has some limitations when it comes to its implementation on mobile robots. For a valid depth the mapping distance between the device and the object has to be more than 0.8 m, whereas from the resolution and noise points of view this distance should be even larger [104]. Besides that the sunlight influences the measurements with an embedded IR camera [105] and Kinect's software is adopted to images captured with a static camera. Multimodal systems are also associated with some other problems. The common ones include the computational load and the increased costs of a mobile robot. An example of a particular technical challenge includes the occurrence of misalignments due to a vertical projection from the laser to the image plane. Therefore, Fotiadis et al. [34] developed an adaptive ROI technique that compensates for these misalignments and offers a greater detection range.

The sensory data can be integrated using sequential integration (one data type is used for human detection and a reduction of the search space for other data, which is used for the verification). The outcome of this approach greatly depends on the human-detection phase. If it fails, the entire approach fails. A solution to this problem is the concurrent/parallel processing of sensory data and the integration (the entire space is tracked and the data is fused) [34]. An example of parallel processing is the generation of a Gaussian probability-based hypothesis for each data type and a combination of all the distributions by covariance intersection [25]. Jin et al. [106] proposed a fusion technique that also takes the temporal information of the measured data into account (previously acquired sensor data is used for a better measurement accuracy).

An example of a multimodal system was proposed by Fritsch et al. [44], who used auditory, vision and LRF channels. In the situations in which the target human is not facing the robot (which prevents face detection), is not speaking and the LRF fails to detect the legs the authors used colour-based torso detection (by applying a mixture of Gaussians). This approach, however, requires that each human wears different, uniformly coloured clothes. Another interesting multimodal approach was proposed by Wilhelm et al. [49], who used a two-component system for human detection. The first one helps positioning the potential human target by means of skin-colour detection and sonar data, while the second uses face detection on a high-resolution image. This component is used for verification and also offers the potential for extracting useful information about the state of a human in order for a robot to adapt to it.

3 PROPOSED APPLICATIONS OF HUMAN-ORIENTED MOBILE ROBOTS

Some of the proposed applications for the reviewed robot systems include: tracking a pre-registered person [13], an autonomous search for a single elderly person in an unstructured indoor environment [8], support for emergency personnel [16], additional support for autonomous guidance of humans in museums [15], visual guidance of mobile robots [53] and [107], providing information for staff and visitors of a specific public building [24], survey tasks [25], interactive shopping assistance [14], surveillance of large outdoor infrastructures [34], human-robot cooperation in transportation and investigations of hazardous environments [23]. Please note that all the above-mentioned proposed application have not yet been realised.

Another useful characteristic of mobile robots is human following, which is important in applications in which a proper interaction of a mobile robot with a walking human is essential (e.g., in rehabilitation [108]). Human following can be passive or anticipative. In the former, the mobile robot's motion is defined only by the position of the target human. This is useful in situations in which a person wants to control the movement of the mobile robot. On the other hand, an anticipative approach is based on predicting a person's trajectory by observing his or her walking mode. This is useful in front human following (useful in leading people through healthcare facilities). However, it has been reported that only using an anticipative approach in front following is not successful, because people change their behaviour unintentionally in the presence of a mobile robot and try to lead it [93]. A short overview of recent human-following robot applications was published in [93].

4 FUTURE CHALLENGES IN THE FIELD OF HUMAN-ORIENTED MOBILE ROBOTS

In this section we provide our view of the challenges that need to be considered when implementing mobile robots in real-world environments.

4.1 HRI and Perception of Mobile Robots by the Elderly People and Chronically Ill People

Human-oriented mobile robots with healthcare tasks should be modelled as social robots. This means that they should be able to interact with humans in various situations [109]. For a social interaction, people need to treat the robots as social beings [109]. In order to achieve as successful HRI as possible it is important to understand the perception of mobile robots by the elderly and chronically ill people.

It is hard to generalize this perception, since the age is not the only factor influencing it. In order to accept any robot, its user needs to be motivated for using the robot, which 1) has to be easy to use and 2) has to allow its user to feel comfortable (physical-, cognitive- and emotional-wise) [110]. There are numerous factors influencing the aforementioned criteria, namely individual and robot factors. Besides the age, the former include needs, gender, experience with technology, cognitive ability, education, culture, anxiety and attitudes towards robots [110]. Robot factors include appearance (humanness, size, facial expression, gender), personality and adaptability [110].

In general, with increasing age the willingness of the people to use robots decreases. However, there are reports that elderly are more likely to accept robots in order to gain back independence in handling everyday tasks upon losing it [111]. Furthermore, lack of familiarity with technology, which is often present in the elderly people, can result in uncertainty toward the robots [112]. Additionally, due to various attitudes toward aging in different cultures the cross-cultural differences exist [113]. When it comes to the robot factors, the elderly people do not want to be accompanied by the robots, which would make them look weak or dependent [114]. Additionally, they prefer smaller sized robots [115]. One of the most important characteristics is also robots' ability to adapt to the particular elderly user, since the elderly differ between each other in e.g. eyesight, movement abilities and hearing capability [110]. When it comes to chronically ill patients, the same factors need to be considered. Another important characteristic, which influences HRI in chronically ill and/or elderly people is the fact that interaction between this people is not short-termed or even single-termed, so the long-term interaction studies are highly important [116]. Unfortunately robots' appearance and behaviour cannot be adopted entirely to human expectations, so it has been suggested that human expectations can be modified in order to achieve better perception of robots [110]. More advanced view on this topic is out of scope of this article. The interested reader can refer to the following review articles: [110], [116] and [117].

4.2 Human Identification Challenges

In general, human identification can be performed using hard biometrics (iris, fingerprint and face). These trails are unique to the individuals, but the identification accuracy strongly depends on the data quality [118]. Distance from the sensor to the target, noise and user's willingness to cooperate are some of the factors that influence this quality [118]. Multimodal hard biometrics systems can offer improved performance, but can be time consuming and require even more human cooperation. As an alternative approach, one can use soft biometrics. These are composed of global (age, gender, skin colour, etc.) and local traits (eyes, eyebrows, nose, mouth, etc.) [118]. They are not unique to the individuals but can as a whole enhance the identification performance of hard biometrics and can be extracted even from lower quality data or from semantic descriptions [119]. Another popular way to identify humans (mainly in surveillance applications) is gait analysis [120], but this approach lacks applicability in healthcare, mobile-robot applications (bed bound patients, gait disorders, etc.). Lastly, human identification should be as pleasant as possible. The user should not be agitated in any way, since that could influence user's (patho)physiological state.

4.3 Possible Extensions of Multimodal Systems

Besides vision and auditory sense, there are three additional senses: touch, olfaction (airborne chemical sensing) and taste. Tactile sensors in mobile robotics are useful in applications in which physical contact with humans is required. Examples include lifting up a dummy human [121] and assisting elderly or disabled people by moving heavy objects instead of them [122]. In the latter case, its user is guiding the robot by the means of tactile communication. In these applications stability of the mobile robots [123] needs to be carefully addressed. In contrast to visual, auditory and tactile signals, which are based on single physical quantities, taste and olfaction only have a meaning when humans interpret them, since taste and odour are not properties of chemical substances [124]. This makes the implementation of olfaction and taste in any real mobile robotic applications very challenging.

Olfaction has been however already implemented in mobile robots for the purpose of gas distribution mapping, trail guidance and gas source localization [125]. On contrary, sense of taste has been implemented in the form of electronic tongues, which have their potential in use in food and pharmaceutical industries for objective and reproducible assessment of taste of foods and drugs [124].

Since healthcare mobile robots are mainly used in indoor environments, the airborne chemical sensing would be very useful in carbon monoxide detection and in prevention of sick building syndrome [126]. Furthermore, some medical conditions have characteristic odours [127] and therefore e.g., the analysis of exhaled breath could be used as a supportive diagnostic tool [128]). On the other hand different odour sensations can also cause symptoms in humans (e.g., headache, nausea, cough, stress) [129]. Identification of these odours could therefore be helpful in preventing mistakes by attributing symptoms to the wrong causes.

4.4 Other Challenges

In order to implement mobile robots in many uncontrolled environments (useful in outdoor healthcare tasks) it is also necessary to ensure the continuous mobility of these mobile robots. In [130], an approach capable of detecting failure at any wheel and optimizing traction is proposed. Additionally, this solution does not increase the hardware complexity of the mobile robot, nor the control system. The next desired characteristic of mobile robots with human-oriented tasks is real-time performance with as little energy consumption as possible. For example, the ideal system would be smooth, rapid, accurate and energy efficient. This can be achieved by mimicking animal-like coordination of the head, neck and eyes. For example, in [131], an approach using chameleon-inspired binocular vision for a swift search of a mobile robot's surroundings and a two-step aim at the target (rough and accurate) is proposed.

When it comes to real-time performance, it is suggested to first define this term. In human-oriented mobile robots the real-time performance can be defined on the basis of a person's reaction time (RT) [26]. The RT for the detection of a visual stimulus is (180 to 200) ms [132] (which is longer than for tactile and auditory RT). This means that frame rates higher than 5 fps offer real-time performance. Similar criterion was proposed in [26]. The real-time performance can be achieved using distributed computing [23] or by using computationally more demanding solutions only when needed [26].

Next, mobile robots implemented in public spaces are likely to attract people, which can result in the narrowing of passageways due to a large number of people surrounding the robot. The former can make it difficult for humans to avoid the crowd, while the latter can aggravate the performance of the mobile robot. Both situations are highly undesirable, e.g., in emergency situations. One of the proposed solutions (based on pedestrian-behaviour simulation) is to anticipate the crowding and try to avoid congestion, while at the same time respecting the surrounding humans' walking comfort and the performance of tasks, for which the mobile robot was built [133]. In cases of people gathering around the robot, an obstacle-avoiding behaviour based on a human-behaviour model can be applied [134]. For an even more successful implementation of mobile robots into public places, a long-term study in terms of their usage [135] would most likely be highly beneficial.

Other challenges include the secure transmission and collection of the measured personal data [136], which needs to be collected ethically.

4.5 A Possible Role of Vision-Based Mobile Robots in Healthcare Measurements

From the perspective of measuring clinically relevant parameters using vision systems we see the following implementations. Mobile robots with thermal vision could be used in fever screening [137], but, in general, also in thermoregulation studies, the detection of breast cancer, diagnosing diabetic neuropathy and vascular disorders, dermatology, etc. [10]. Many of these fields could significantly improve the patient's well being by regular home monitoring of a disease, which could reduce its burden (e.g., by monitoring diabetic patients with thermo vision it could be possible to prevent diabetic foot ulcers [138]). Colour-vision-based mobile robots could provide some physiological data by means of remote photoplethysmography (remote PPG) measurements, whether in reflection [11] or transmittance mode [139]. It can also be used in telemedicine in the form of simple online consultations or as a tool for diagnosing/monitoring diseases (teledermatology) [140]. In the near future, we aim to develop a human-oriented, colour-vision-based mobile robot performing certain healthcare tasks.

4.6 Techonological Trends

Current trends in robotics are focused on soft robotics. This term primarily covers implementation of soft materials, actuators and sensors in different machine application. Soft robots are expected to offer softness and safety (by the means of more natural physical HRI), which are highly desirable characteristics for the use in healthcare applications (lifting of the patients, minimally invasive surgeries, various wearable and implantable devices) [141]. From the perspective of elderly people these robots could be used as an adaptive exercisers for cognition and daily activities [141].

5 CONLCUSIONS

The reviewed literature reveals that there is no universal solution for a human-oriented mobile robot. Different hardware and software solutions have their pros and cons in different environmental settings and situations, which makes us believe that mobile robots with multiple sensor modalities will be the most studied in the near future. An interesting solution for indoor environments, such as households or clinics, is iSpace, which could offer the implementation of mobile robots with different tasks in the same environment. From the perspective of human recognition, the human-identification step seems to be the most challenging. We believe that a lot of effort will be put into it, since correct identification is crucial in healthcare, where misidentification could result in the wrong treatment, having potentially fatal consequences. Because vision systems are already being widely used in medicine, the use of thermal vision and colour vision in mobile robots for diagnostic/screening purposes is promising. In the future, to implement mobile robots in as many healthcare applications as possible, the focus will need to be put on HRI, human identification, robustness of the mobile robots' task performance and quality, together with security of the measured data.

6 ACKNOWLEDGEMENT

The presented research of the Young Researcher Miha Finžgar was supported by the Slovenian Research Agency (ARRS).

7 REFERENCES

[1] Wada, K., Shibata, T. (2007). Living with seal robots - its sociopsychological and physiological influences on the elderly at a care house. *IEEE Transactions on Robotics*, vol. 23, no. 5, p. 972-980, **DOI:10.1109/Tro.2007.906261**.

[2] Dahl, T., Boulos, M. (2014). Robots in health and social care: A complementary technology to home care and telehealthcare? *Robotics*, vol. 3, no. 1, p. 1-21, **DOI:10.3390/robotics3010001**.

[3] Evans, J.M. (1994). Helpmate: An autonomous mobile robot courier for hospitals. *Proceedings of the IEEE/RSJ/GI International Conference on Intelligent Robots and Systems*, p. 1695-1700, **DOI:10.1109/iros.1994.407629**.

[4] Sucher, J.F., Todd, S.R., Jones, S.L., Throckmorton, T., Turner, K.L., Moore, F.A. (2011). Robotic telepresence: A helpful adjunct that is viewed favorably by critically ill surgical patients. *The American Journal of Surgery*, vol. 202, no. 6, p. 843-847, **DOI:10.1016/j.amjsurg.2011.08.001**.

[5] Pripfl, J., Körtner, T., Batko-Klein, D., Hebesberger, D., Weninger, M., Gisinger, C., Frennert, S., Eftring, H., Antona, M., Adami, I., Weiss, A., Bajones, M., Vincze, M. (2016). Results of a real world trial with a mobile social service robot for older adults. *11th ACM/IEEE International Conference on Human-Robot Interaction*, p. 497-498, **DOI:10.1109/hri.2016.7451824**.

[6] Galán-Mena, J., Ávila, G., Pauta-Pintado, J., Lima-Juma, D., Robles-Bykbaev, V., Quisi-Peralta, D. (2016). An intelligent system based on ontologies and ICT tools to support the diagnosis and intervention of children with autism. IEEE Biennial Congress of Argentina, p. 1-5, **DOI:10.1109/ARGENCON.2016.7585361**.

[7] Ma, Y., Xiao, D., Li, R., Ruan, H., Shan, Z., Junlong, Z., Zhang, Y. (2015). Android-based intelligent mobile robot for indoor healthcare. *17th International Conference on E-health Networking, Application & Services (HealthCom)*, p. 472-474, **DOI:10.1109/HealthCom.2015.7454548**.

[8] Mehdi, S.A., Berns, K. (2014). Behavior-based search of human by an autonomous indoor mobile robot in simulation. *Universal Access in the Information Society*, vol. 13, no. 1, p. 45-58, **DOI:10.1007/s10209-013-0301-8**.

[9] Yagi, Y. (1999). Omnidirectional sensing and its applications. *IEICE Transactions on Information and Systems*, vol. 82, no. 3, p. 568-579.

[10] Lahiri, B.B., Bagavathiappan, S., Jayakumar, T., Philip, J. (2012). Medical applications of infrared thermography: A review. *Infrared Physics & Technology*, vol. 55, no. 4, p. 221-235, **DOI:10.1016/j.infrared.2012.03.007**.

[11] Aarts, L.A., Jeanne, V., Cleary, J.P., Lieber, C., Nelson, J.S., Bambang Oetomo, S., Verkruysse, W. (2013). Non-contact heart rate monitoring utilizing camera photoplethysmography in the neonatal intensive care unit - a pilot study. *Early Human Development*, vol. 89, no. 12, p. 943-948, **DOI:10.1016/j.earlhumdev.2013.09.016**.

[12] Cielniak, G., Duckett, T. (2004). People recognition by mobile robots. *Journal of Intelligent & Fuzzy Systems*, vol. 15, no. 1, p. 21-27.

[13] Kang, S., Roh, A., Nam, B., Hong, H. (2011). People detection method using graphics processing units for a mobile robot with an omnidirectional camera. *Optical Engineering*, vol. 50, no. 12, p. 1-8, **DOI:10.1117/1.3660573**.

[14] Böhme, H.-J., Wilhelm, T., Key, J., Schauer, C., Schröter, C., Groß, H.-M., Hempel, T. (2003). An approach to multi-modal human–machine interaction for intelligent service robots. *Robotics and Autonomous Systems*, vol. 44, no. 1, p. 83-96, **DOI:10.1016/S0921-8890(03)00012-5**.

[15] Baltzakis, H., Pateraki, M., Trahanias, P. (2012). Visual tracking of hands, faces and facial features of multiple persons. *Machine Vision and Applications*, vol. 23, no. 6, p. 1141-1157, **DOI:10.1007/s00138-012-0409-5**.

[16] Susperregi, L., Martinez-Otzeta, J.M., Ansuategui, A., Ibarguren, A., Sierra, B. (2013). RGB-D, laser and thermal sensor fusion for people following in a mobile robot. *International Journal of Advanced Robotic Systems*, vol. 10, p. 1-9, **DOI:10.5772/56123**.

[17] Ying-li, T. (2004). Evaluation of face resolution for expression analysis. *Conference on Computer Vision and Pattern Recognition Workshop*, p. 82-82, **DOI:10.1109/cvpr.2004.334**.

[18] Zafeiriou, S., Zhang, C., Zhang, Z. (2015). A survey on face detection in the wild: Past, present and future. *Computer Vision and Image Understanding*, vol. 138, p. 1-24, **DOI:10.1016/j.cviu.2015.03.015**.

[19] Bayram, B., Ince, G. (2015). Audio-visual multi-person tracking for active robot perception. *IEEE/SICE International Symposium on System Integration*, p. 575-580, **DOI:10.1109/sii.2015.7405043**.

[20] Ali, B., Ayaz, Y., Jamil, M., Gilani, S.O., Muhammad, N. (2015). Improved method for stereo vision-based human detection for a mobile robot following a target person. *South African Journal of Industrial Engineering*, vol. 26, no. 1, p. 102-119, DOI:10.7166/26-1-891.

[21] Fernandez-Caballero, A., Castillo, J.C., Martinez-Cantos, J., Martinez-Tomas, R. (2010). Optical flow or image subtraction in human detection from infrared camera on mobile robot. *Robotics and Autonomous Systems*, vol. 58, no. 12, p. 1273-1281, DOI:10.1016/j.robot.2010.06.002.

[22] Cielniak, G., Duckett, T., Lilienthal, A.J. (2010). Data association and occlusion handling for vision-based people tracking by mobile robots. *Robotics and Autonomous Systems*, vol. 58, no. 5, p. 435-443, DOI:10.1016/j.robot.2010.02.004.

[23] Petrović, E., Leu, A., Ristić-Durrant, D., Nikolić, V. (2013). Stereo vision-based human tracking for robotic follower. *International Journal of Advanced Robotic Systems*, vol. 10, no. 5, DOI:10.5772/56124.

[24] Bellotto, N., Hu, H. (2008). Multimodal people tracking and identification for service robots. *International Journal of Information Acquisition*, vol. 5, no. 3, p. 209-221, DOI:10.1142/S0219878908001612.

[25] Martin, C., Schaffernicht, E., Scheidig, A., Gross, H.M. (2006). Multi-modal sensor fusion using a probabilistic aggregation scheme for people detection and tracking. *Robotics and Autonomous Systems*, vol. 54, no. 9, p. 721-728, DOI:10.1016/j.robot.2006.04.012.

[26] Zhang, H., Reardon, C., Parker, L.E. (2013). Real-time multiple human perception with color-depth cameras on a mobile robot. *IEEE Transactions on Cybernetics*, vol. 43, no. 5, p. 1429-1441, DOI:10.1109/TCYB.2013.2275291.

[27] Xu, L.-Q., C. Hogg, D. (1997). Neural networks in human motion tracking — an experimental study. *Image and Vision Computing*, vol. 15, no. 8, p. 607-615, DOI:10.1016/S0262-8856(97)00007-3.

[28] Tsai, C.-Y., Song, K.-T. (2009). Dynamic visual tracking control of a mobile robot with image noise and occlusion robustness. *Image and Vision Computing*, vol. 27, no. 8, p. 1007-1022, DOI:10.1016/j.imavis.2008.08.011.

[29] Okuma, K., Taleghani, A., de Freitas, N., Little, J.J., Lowe, D.G. (2004). A boosted particle filter: Multitarget detection and tracking. Pajdla, T., Matas, J. (eds.), *Computer Vision - ECCV,* Springer-Verlag Berlin Heidelberg, Berlin, Heidelberg, p. 28-39, DOI:10.1007/978-3-540-24670-1_3.

[30] Stone, L.D., Streit, R.L., Corwin, T.L., Bell, K.L. (2013). *Bayesian multiple target tracking*. Artech House, Norwood, MA.

[31] Ahmad, S.M.S., Ali, B.M., Adnan, W.A.W. (2012). Technical issues and challenges of biometric applications as access control tools of information security. *International Journal of Innovative Computing Information and Control*, vol. 8, no. 11, p. 7983-7999.

[32] Alvarez-Santos, V., Pardo, X.M., Iglesias, R., Canedo-Rodriguez, A., Regueiro, C.V. (2012). Feature analysis for human recognition and discrimination: Application to a person-following behaviour in a mobile robot. *Robotics and Autonomous Systems*, vol. 60, no. 8, p. 1021-1036, DOI:10.1016/j.robot.2012.05.014.

[33] Kim, J., Suga, Y. (2007). An omnidirectional vision-based moving obstacle detection in mobile robot. *International Journal of Control Automation and Systems*, vol. 5, no. 6, p. 663-673.

[34] Fotiadis, E.P., Garzon, M., Barrientos, A. (2013). Human detection from a mobile robot using fusion of laser and vision information. *Sensors* (Basel), vol. 13, no. 9, p. 11603-11635, DOI:10.3390/s130911603.

[35] Intille, S.S., Davis, J.W., Bobick, A.F. (1997). Real-time closed-world tracking. *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, p. 697-703, DOI:10.1109/CVPR.1997.609402.

[36] Mittal, A., Davis, L.S. (2003). M2tracker: A multi-view approach to segmenting and tracking people in a cluttered scene. *International Journal of Computer Vision*, vol. 51, no. 3, p. 189-203, DOI:10.1023/A:1021849801764.

[37] Wren, C.R., Azarbayejani, A., Darrell, T., Pentland, A.P. (1997). Pfinder: Real-time tracking of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, p. 780-785, DOI:10.1109/AFGR.1996.557243.

[38] Schauer, C., Zahn, T., Paschke, P., Gross, H.-M. (2000). Binaural sound localization in an artificial neural network. *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, p. 865-868, DOI:10.1109/icassp.2000.859097.

[39] Schauer, C., Paschke, P. (1999). A spike-based model of binaural sound localization. *International Journal of Neural Systems*, vol. 9, no. 05, p. 447-452, DOI:10.1142/S0129065799000460.

[40] Yang, J., Lu, W., Waibel, A. (1998). Skin-color modeling and adaptation. *Asian Conference on Computer Vision*, p. 687-694.

[41] Fahlman, S.E., Lebiere, C. (1990). The cascade-correlation learning architecture. Touretzky, D.S. (ed.), *Advances in neural information processing systems 2*. Morgan Kaufmann Publishers, San Francisco, p. 524-532.

[42] AT&T Laboratories Cambridge. The database of faces, from *http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html*, accessed on 2016-12-16.

[43] Isard, M., Blake, A. (1998). Condensation—conditional density propagation for visual tracking. *International Journal of Computer Vision*, vol. 29, no. 1, p. 5-28, DOI:10.1023/A:1008078328650.

[44] Fritsch, J., Kleinehagenbrock, M., Lang, S., Fink, G.A., Sagerer, G. (2004). Audiovisual person tracking with a mobile robot. *Proceedings of International Conference on Intelligent Autonomous Systems*, p. 898-906.

[45] Viola, P., Jones, M. (2001). Robust real-time object detection. *International Journal of Computer Vision*, vol. 4, p. 34-47.

[46] Giuliani, D., Omologo, M., Svaizer, P. (1994). Talker localization and speech recognition using a microphone array and a cross-powerspectrum phase analysis. *International Conference on Spoken Language Processing.*

[47] Fritsch, J., Kleinehagenbrock, M., Lang, S., Plötz, T., Fink, G.A., Sagerer, G. (2003). Multi-modal anchoring for human–robot interaction. *Robotics and Autonomous Systems*, vol. 43, no. 2, p. 133-147, **DOI:10.1016/S0921-8890(02)00355-X**.

[48] Lang, S., Kleinehagenbrock, M., Hohenner, S., Fritsch, J., Fink, G.A., Sagerer, G. (2003). Providing the basis for human-robot-interaction: A multi-modal attention system for a mobile robot. *Proceedings of the 5th International Conference on Multimodal Interfaces*, p. 28-35, **DOI:10.1145/958432.958441**.

[49] Wilhelm, T., Bohme, H., Gross, H.M. (2004). A multi-modal system for tracking and analyzing faces on a mobile robot. *Robotics and Autonomous Systems*, vol. 48, no. 1, p. 31-40, **DOI:10.1016/j.robot.2004.05.004**.

[50] Feyrer, S., Zell, A. (1999). Detection, tracking, and pursuit of humans with an autonomous mobile robot. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, p. 864-869, **DOI:10.1109/iros.1999.812788**.

[51] Treptow, A., Cielniak, G., Duckett, T. (2006). Real-time people tracking for mobile robots using thermal vision. *Robotics and Autonomous Systems*, vol. 54, no. 9, p. 729-739, **DOI:10.1016/j.robot.2006.04.013**.

[52] Lucas, B.D., Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. *International Joint Conference on Artificial Intelligence*, p. 674-679.

[53] Chang, J.S., Kim, E.Y., Kim, H.J. (2008). Mobile robot control using hand-shape recognition. *Transactions of the Institute of Measurement and Control*, vol. 30, no. 2, p. 143-152, **DOI:10.1177/0142331207080144**.

[54] Hu, M.-K. (1962). Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory*, vol. 8, no. 2, p. 179-187, **DOI:10.1109/TIT.1962.1057692**.

[55] Vadakkepat, P., Lim, P., De Silva, L.C., Jing, L., Ling, L.L. (2008). Multimodal approach to human-face detection and tracking. *IEEE Transactions on Industrial Electronics*, vol. 55, no. 3, p. 1385-1393, **DOI:10.1109/Tie.2007.903993**.

[56] Bradski, G.R. (1998). Computer vision face tracking for use in a perceptual user interface. *Intel Technology Journal Q2*, p. 1-15.

[57] Viola, P., Jones, M.J. (2004). Robust real-time face detection. *International Journal of Computer Vision*, vol. 57, no. 2, p. 137-154, **DOI:10.1023/B:VISI.0000013087.49260.fb**.

[58] Bellotto, N., Hu, H. (2006). Vision and laser data fusion for tracking people with a mobile robot. *IEEE International Conference on Robotics and Biomimetics*, p. 7-12, **DOI:10.1109/ROBIO.2006.340251**.

[59] Julier, S.J., Uhlmann, J.K. (2004). Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, vol. 92, no. 3, p. 401-422, **DOI:10.1109/jproc.2003.823141**.

[60] Comaniciu, D., Ramesh, V., Meer, P. (2000). Real-time tracking of non-rigid objects using mean shift. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, p. 142-149, **DOI:10.1109/cvpr.2000.854761**.

[61] Oklahoma State University. OTCBVS benchmark dataset collection, from *http://vcipl-okstate.org/pbvs/bench/,* accessed on 2016-12-16.

[62] Stricker, M.A., Orengo, M. (1995). Similarity of color images. *IS&T/SPIE's Symposium on Electronic Imaging: Science & Technology*, p. 381-392.

[63] Freund, Y., Schapire, R.E. (1995). A desicion-theoretic generalization of on-line learning and an application to boosting. *European Conference on Computational Learning Theory*, p. 23-37, **DOI:10.1007/3-540-59119-2_166**.

[64] Horn, B.K., Schunck, B.G. (1981). Determining optical flow. Artificial Intelligence, vol. 17, no. 1-3, p. 185-203, **DOI:10.1016/0004-3702(81)90024-2**.

[65] Dalal, N., Triggs, B. (2005). Histograms of oriented gradients for human detection. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, p. 886-893, **DOI:10.1109/cvpr.2005.177**.

[66] Ojala, T., Pietikäinen, M., Harwood, D. (1996). A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, vol. 29, no. 1, p. 51-59, **DOI:10.1016/0031-3203(95)00067-4**.

[67] Zabih, R., Woodfill, J. (1994). Non-parametric local transforms for computing visual correspondence. *European Conference on Computer Vision*, p. 151-158, **DOI:10.1007/bfb0028345**.

[68] Heikkilä, M., Pietikäinen, M., Schmid, C. (2006). Description of interest regions with center-symmetric local binary patterns. Kalra, P.K., Peleg, S. (eds.), *Computer Vision, Graphics and Image Processing. Lecture Notes in Computer Science,* Springer, p. 58-69, **DOI:10.1007/11949619_6**.

[69] Mu, Y., Yan, S., Liu, Y., Huang, T., Zhou, B. (2008). Discriminative local binary patterns for human detection in personal album. *IEEE Conference on Computer Vision and Pattern Recognition*, p. 1-8.

[70] Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 6, p. 679-698, **DOI:10.1109/TPAMI.1986.4767851**.

[71] Won, C.S., Park, D.K., Park, S.-J. (2002). Efficient use of MPEG-7 edge histogram descriptor. *Etri Journal*, vol. 24, no. 1, p. 23-30, **DOI:10.4218/etrij.02.0102.0103**.

[72] Baltzakis, H., Argyros, A., Lourakis, M., Trahanias, P. (2008). Tracking of human hands and faces through probabilistic fusion of multiple visual cues. *International Conference on Computer Vision Systems*, p. 33-42, **DOI:10.1007/978-3-540-79547-6_4**.

[73] Argyros, A.A., Lourakis, M.I. (2004). Real-time tracking of multiple skin-colored objects with a possibly moving camera. *European Conference on Computer Vision*, p. 368-379, **DOI:10.1007/978-3-540-24672-5_29**.

[74] Baltzakis, H., Argyros, A.A. (2009). Propagation of pixel hypotheses for multiple objects tracking. *International Symposium on Visual Computing*, p. 140-149, **DOI:10.1007/978-3-642-10520-3_13**.

[75] Cortes, C., Vapnik, V. (1995). Support-vector networks. *Machine Learning*, vol. 20, no. 3, p. 273-297, **DOI:10.1007/BF00994018**.

[76] Fischler, M.A., Bolles, R.C. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, vol. 24, no. 6, p. 381-395, **DOI:10.1145/358669.358692**.

[77] Viola, P., Jones, M.J., Snow, D. (2005). Detecting pedestrians using patterns of motion and appearance. *International*

Journal of Computer Vision, vol. 63, no. 2, p. 153-161, **DOI:10.1007/s11263-005-6644-8**.
[78] Bourdev, L., Malik, J. (2009). Poselets: Body part detectors trained using 3D human pose annotations. *IEEE 12th International Conference on Computer Vision*, p. 1365-1372, **DOI:10.1109/iccv.2009.5459303**.
[79] Einicke, G.A., White, L.B. (1999). Robust extended kalman filtering. *IEEE Transactions on Signal Processing*, vol. 47, no. 9, p. 2596-2599, **DOI:10.1109/78.782219**.
[80] Martínez-Otzeta, J.M., Ibarguren, A., Ansuategi, A., Susperregi, L. (2009). Laser based people following behaviour in an emergency environment. *International Conference on Intelligent Robotics and Applications*, p. 33-42, **DOI:10.1007/978-3-642-10817-4_4**.
[81] Arulampalam, M.S., Maskell, S., Gordon, N., Clapp, T. (2002). A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, vol. 50, no. 2, p. 174-188, **DOI:10.1109/78.978374**.
[82] Susperregi, L., Sierra, B., Castrillon, M., Lorenzo, J., Martinez-Otzeta, J.M., Lazkano, E. (2013). On the use of a low-cost thermal sensor to improve Kinect people detection in a mobile robot. *Sensors* (Basel), vol. 13, no. 11, p. 14687-14713, **DOI:10.3390/s131114687**.
[83] Aha, D.W., Kibler, D., Albert, M.K. (1991). Instance-based learning algorithms. *Machine Learning*, vol. 6, no. 1, p. 37-66, **DOI:10.1007/BF00153759**.
[84] Cestnik, B. (1990). Estimating probabilities: A crucial task in machine learning. *Proceedings of the 9th European Conference on Artificial Intelligence*, p. 147-149.
[85] Sierra, B., Lazkano, E., Jauregi, E., Irigoien, I. (2009). Histogram distance-based Bayesian network structure learning: A supervised classification specific approach. *Decision Support Systems*, vol. 48, no. 1, p. 180-190, **DOI:10.1016/j.dss.2009.07.010**.
[86] Quinlan, J.R. (2014). C4. 5: *Programs for Machine Learning*. Elsevier.
[87] Martínez-Otzeta, J.M., Sierra, B., Lazkano, E., Astigarraga, A. (2006). Classifier hierarchy learning by means of genetic algorithms. *Pattern Recognition Letters*, vol. 27, no. 16, p. 1998-2004, **DOI:10.1016/j.patrec.2006.06.001**.
[88] Natarajan, S.K., Ristic-Durrant, D., Leu, A., Gräser, A. (2011). Robust stereo-vision based 3D modelling of real-world objects for assistive robotic applications. *IEEE/RSJ International Conference on Intelligent Robots and Systems*, p. 786-792, **DOI:10.1109/iros.2011.6094716**.
[89] Beetz, M., Arbuckle, T., Belker, T., Cremers, A.B., Schulz, D., Bennewitz, M., Burgard, W., Hahnel, D., Fox, D., Grosskreutz, H. (2001). Integrated, plan-based control of autonomous robots in human environments. *IEEE Intelligent Systems*, vol. 16, no. 5, p. 56-65, **DOI:10.1109/mis.2001.956082**.
[90] Foka, A.F., Trahanias, P.E. (2010). Probabilistic autonomous robot navigation in dynamic environments with human motion prediction. *International Journal of Social Robotics*, vol. 2, no. 1, p. 79-94, **DOI:10.1007/s12369-009-0037-z**.
[91] Viola, P., Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, p. I-511-I-518 vol. 511, **DOI:10.1109/cvpr.2001.990517**.
[92] Ciric, I., Cojbasic, Z., Nikolic, V., Antic, D. (2013). Computationally intelligent system for thermal vision people detection and tracking in robotic applications. *11th International Conference on Telecommunication in Modern Satellite, Cable and Broadcasting Services*, p. 587-590., **DOI:10.1109/telsks.2013.6704447**.
[93] Hu, J.S., Wang, J.J., Ho, D.M. (2014). Design of sensing system and anticipative behavior for human following of mobile robots. *IEEE Transactions on Industrial Electronics*, vol. 61, no. 4, p. 1916-1927, **DOI:10.1109/Tie.2013.2262758**.
[94] Borges, G.A., Aldon, M.-J. (2004). Line extraction in 2D range images for mobile robotics. *Journal of Intelligent and Robotic Systems*, vol. 40, no. 3, p. 267-297, **DOI:10.1023/B:JINT.0000038945.55712.65**.
[95] Comaniciu, D., Ramesh, V., Meer, P. (2003). Kernel-based object tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 5, p. 564-577, **DOI:10.1109/TPAMI.2003.1195991**.
[96] Li, P., Zhang, T., Pece, A.E. (2003). Visual contour tracking based on particle filters. *Image and Vision Computing*, vol. 21, no. 1, p. 111-123, **DOI:10.1016/S0262-8856(02)00133-6**.
[97] Nakamura, K., Nakadai, K., Asano, F., Ince, G. (2011). Intelligent sound source localization and its application to multimodal human tracking. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems,* p. 143-148, **DOI:10.1109/iros.2011.6094558**.
[98] Singh, S.K., Chauhan, D., Vatsa, M., Singh, R. (2003). A robust skin color based face detection algorithm. *Tamkang Journal of Science and Engineering*, vol. 6, no. 4, p. 227-234.
[99] Bradski, G.R. (1998). Real time face and object tracking as a component of a perceptual user interface. *Proceedings of the 4th IEEE Workshop on Applications of Computer Vision*, p. 214-219, **DOI:10.1109/acv.1998.732882**.
[100] Ess, A., Leibe, B., Schindler, K., Gool, L.V. (2008). A mobile vision system for robust multi-person tracking. *IEEE Conference on Computer Vision and Pattern Recognition*, p. 1-8, **DOI:10.1109/cvpr.2008.4587581**.
[101] Ciric, I.T., Cojbasic, Z.M., Nikolic, V.D., Igic, T.S., Tursnek, B.A.J. (2014). Intelligent optimal control of thermal vision-based person-following robot platform. *Thermal Science*, vol. 18, no. 3, p. 957-966, **DOI:10.2298/Tsci1403957c**.
[102] Simončič, S., Podržaj, P. (2014). Vision-based control of a line-tracing mobile robot. *Computer Applications in Engineering Education*, vol. 22, no. 3, p. 474-480, **DOI:10.1002/cae.20573**.
[103] Lee, J.-H., Hashimoto, H. (2002). Intelligent space—concept and contents. *Advanced Robotics*, vol. 16, no. 3, p. 265-280, **DOI:10.1163/156855302760121936**.
[104] Khoshelham, K., Elberink, S.O. (2012). Accuracy and resolution of Kinect depth data for indoor mapping applications. *Sensors* (Basel), vol. 12, no. 2, p. 1437-1454, **DOI:10.3390/s120201437**.
[105] Suarez, J., Murphy, R.R. (2012). Using the Kinect for search and rescue robotics. *IEEE International Symposium on*

Safety, Security, and Rescue Robotics, p. 1-2, DOI:10.1109/ssrr.2012.6523918.

[106] Jin, T.-S., Lee, J.M., Tso, S. (2004). A new approach using sensor data fusion for mobile robot navigation. *Robotica*, vol. 22, no. 01, p. 51-59, DOI:10.1017/S0263574703005381.

[107] Huang, S.J., Liu, S., Wu, C.H. (2015). Intelligent humanoid mobile robot with embedded control and stereo visual feedback. *Journal of Mechanical Science and Technology*, vol. 29, no. 9, p. 3919-3931, DOI:10.1007/s12206-015-0838-y.

[108] Gross, H.-M., Scheidig, A., Debes, K., Einhorn, E., Eisenbach, M., Mueller, S., Schmiedel, T., Trinh, T.Q., Weinrich, C., Wengefeld, T., Bley, A., Martin, C. (2016). ROREAS: Robot coach for walking and orientation training in clinical post-stroke rehabilitation—prototype implementation and evaluation in field trials. *Autonomous Robots*, p. 1-20, DOI:10.1007/s10514-016-9552-6.

[109] Breazeal, C., Scassellati, B. (1999). How to build robots that make friends and influence people. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, p. 858-863, DOI:10.1109/iros.1999.812787.

[110] Broadbent, E., Stafford, R., MacDonald, B. (2009). Acceptance of healthcare robots for the older population: Review and future directions. *International Journal of Social Robotics*, vol. 1, no. 4, p. 319-330, DOI:10.1007/s12369-009-0030-6.

[111] Arras, K.O., Cerqui, D. (2005). *Do we want to share our lives and bodies with robots? A 2000 people survey*. ETH-Zürich, Zürich.

[112] Dijkers, M.P., deBear, P.C., Erlandson, R.F., Kristy, K., Geer, D.M., Nichols, A. (1991). Patient and staff acceptance of robotic technology in occupational therapy: A pilot study. *Journal of Rehabilitation Research & Development*, vol. 28, no. 2, p. 33-44, DOI:10.1682/JRRD.1991.04.0033.

[113] Shibata, T., Wada, K., Ikeda, Y., Sabanovic, S. (2009). Cross-cultural studies on subjective evaluation of a seal robot. *Advanced Robotics*, vol. 23, no. 4, p. 443-458, DOI:10.1163/156855309X408826.

[114] Hirsch, T., Forlizzi, J., Hyder, E., Goetz, J., Kurtz, C., Stroback, J. (2000). The ELDer project: Social, emotional, and environmental factors in the design of eldercare technologies. *Proceedings on the Conference on Universal Usability*, p. 72-79, DOI:10.1145/355460.355476.

[115] Giuliani, M.V., Scopelliti, M., Fornara, F. (2005). Elderly people at home: Technological help in everyday activities. *IEEE International Workshop on Robot and Human Interactive Communication*, p. 365-370, DOI:10.1109/roman.2005.1513806.

[116] Leite, I., Martinho, C., Paiva, A. (2013). Social robots for long-term interaction: A survey. *International Journal of Social Robotics*, vol. 5, no. 2, p. 291-308, DOI:10.1007/s12369-013-0178-y.

[117] Broekens, J., Heerink, M., Rosendal, H. (2009). Assistive social robots in elderly care: A review. *Gerontechnology*, vol. 8, no. 2, p. 94-103, DOI:10.4017/gt.2009.08.02.002.00.

[118] Reid, D., Samangooei, S., Chen, C., Nixon, M., Ross, A. (2013). Soft biometrics for surveillance: An overview. Rao, C.R., Govindaraju, V. (eds.), *Handbook of statistics: Machine learning: Theory and applications*, Elsevier, p. 327-352, DOI:10.1016/B978-0-444-53859-8.00013-8.

[119] Almudhahka, N.Y., Nixon, M.S., Hare, J.S. (2016). Unconstrained human identification using comparative facial soft biometrics. *IEEE International Conference on Biometrics Theory, Applications and Systems*, p. 1-6, DOI:10.1109/btas.2016.7791206.

[120] Lee, T.K.M., Belkhatir, M., Sanei, S. (2014). A comprehensive review of past and present vision-based techniques for gait recognition. *Multimedia Tools and Applications*, vol. 72, no. 3, p. 2833-2869, DOI:10.1007/s11042-013-1574-x.

[121] Mukai, T., Onishi, M., Odashima, T., Hirano, S., Luo, Z. (2008). Development of the tactile sensor system of a human-interactive robot. *IEEE Transactions on Robotics*, vol. 24, no. 2, p. 505-512, DOI:10.1109/TRO.2008.917006.

[122] Wang, H., Liu, X.P. (2011). Haptic interaction for mobile assistive robots. *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 11, p. 3501-3509, DOI:10.1109/TIM.2011.2161141.

[123] Moosavian, S.A.A., Alipour, K. (2007). On the dynamic tip-over stability of wheeled mobile manipulators. *International Journal of Robotics & Automation*, vol. 22, no. 4, p. 322, DOI:10.2316/Journal.206.2007.4.206-3036.

[124] Toko, K., Tahara, Y., Habara, M., Kobayashi, Y., Ikezaki, H. (2016). Taste sensor: Electronic tongue with global selectivity. Nakamoto, T. (ed.), *Essentials of Machine Olfaction and Taste*, p. 87-174, DOI:10.1002/9781118768495.ch4.

[125] Lilienthal, A.J., Loutfi, A., Duckett, T. (2006). Airborne chemical sensing with mobile robots. *Sensors*, vol. 6, no. 11, p. 1616-1678, DOI:10.3390/s6111616.

[126] Redlich, C.A., Sparer, J., Cullen, M.R. (1997). Sick-building syndrome. *The Lancet*, vol. 349, no. 9057, p. 1013-1016, DOI:10.1016/S0140-6736(96)07220-0.

[127] Liddell, K. (1976). Smell as a diagnostic marker. *Postgraduate Medical Journal*, vol. 52, no. 605, p. 136-138, DOI:10.1136/pgmj.52.605.136.

[128] Whittle, C.L., Fakharzadeh, S., Eades, J., Preti, G. (2007). Human breath odors and their use in diagnosis. *Annals of the New York Academy of Sciences*, vol. 1098, p. 252-266, DOI:10.1196/annals.1384.011.

[129] Schiffman, S.S., Williams, C.M. (2005). Science of odor as a potential health issue. *Journal of Environmental Quality*, vol. 34, no. 1, p. 129-138.

[130] Xu, H., Gao, X.Z., Xu, Y., Wang, K., Yu, H., Li, Z., Alipour, K., Ani, O.A. (2016). Continuous mobility of mobile robots with a special ability for overcoming driving failure on rough terrain. *Robotica*, p. 1-21, DOI:10.1017/S0263574716000606.

[131] Xu, H., Xu, Y., Fu, H., Gao, X.Z., Alipour, K. (2014). Coordinated movement of biomimetic dual PTZ visual system and wheeled mobile robot. *Industrial Robot: An International Journal*, vol. 41, no. 6, p. 557-566, DOI:10.1108/ir-05-2014-0345.

[132] Thompson, P., Colebatch, J., Brown, P., Rothwell, J., Day, B., Obeso, J., Marsden, C. (1992). Voluntary stimulus-sensitive jerks and jumps mimicking myoclonus or pathological startle syndromes. *Movement Disorders*, vol. 7, no. 3, p. 257-262, DOI:10.1002/mds.870070312.

[133] Kidokoro, H., Kanda, T., Brščić, D., Shiomi, M. (2015). Simulation-based behavior planning to prevent congestion

of pedestrians around a robot. *IEEE Transactions on Robotics*, vol. 31, no. 6, p. 1419-1431, **DOI:10.1109/TRO.2015.2492862**.

[134] Szemes, P.T., Hashimoto, H., Korondi, P. (2005). Pedestrian-behavior-based mobile agent control in intelligent space. *IEEE Transactions on Instrumentation and Measurement*, vol. 54, no. 6, p. 2250-2257, **DOI:10.1109/TIM.2005.858824**.

[135] Brščić, D., Kanda, T. (2015). Changes in usage of an indoor public space: Analysis of one year of person tracking. *IEEE Transactions on Human-Machine Systems*, vol. 45, no. 2, p. 228-237, **DOI:10.1109/THMS.2014.2374172**.

[136] Wu, P., Yi, W.J., Saniie, J. (2016). Security assessment for personal health data management system. *IEEE International Conference on Electro Information Technology*, p. 0422-0427, **DOI:10.1109/eit.2016.7535277**.

[137] Ng, E.Y.K., Kawb, G.J.L., Chang, W.M. (2004). Analysis of IR thermal imager for mass blind fever screening. *Microvascular Research*, vol. 68, no. 2, p. 104-109, **DOI:10.1016/j.mvr.2004.05.003**.

[138] Sousa, P., Felizardo, V., Oliveira, D., Couto, R., Garcia, N.M. (2015). A review of thermal methods and technologies for diabetic foot assessment. *Expert Review of Medical Devices*, vol. 12, no. 4, p. 439-448, **DOI:10.1586/17434440.2015.1032251**.

[139] Amelard, R., Scharfenberger, C., Kazemzadeh, F., Pfisterer, K.J., Lin, B.S., Clausi, D.A., Wong, A. (2015). Feasibility of long-distance heart rate monitoring using transmittance photoplethysmographic imaging (PPGI). *Scientific Reports*, vol. 5, p. 1-11, **DOI:10.1038/srep14637**.

[140] Sharma, P., Kovarik, C.L., Lipoff, J.B. (2016). Teledermatology as a means to improve access to inpatient dermatology care. *Journal of Telemedicine and Telecare*, vol. 22, no. 5, p. 304-310, **DOI:10.1177/1357633X15603298**.

[141] Rossiter, J., Hauser, H. (2016). Soft robotics - the next industrial revolution? [industrial activities]. *IEEE Robotics & Automation Magazine*, vol. 23, no. 3, p. 17-20, **DOI:10.1109/MRA.2016.2588018**.