Non-contact Wideband Sonar for Human Activity Detection and Classification

Gaddi Blumrosen[α], Ben Fishman[α], Yossi Yovel

***Abstract*—This paper suggests using a wideband sonar system to detect and classify human activity in indoor environment. While most existing sonar systems used for assessment of human activity are based on a narrowband Doppler based technology, this paper suggest using wideband sonar. It enables precise tracking of body parts, and its enhanced correlation properties can be used to distinguish between human and non-human objects. Maximal Likelihood (ML) criterions to derive kinematic features and analytical methods to estimate the subject activity level and activity type were derived and tailored to the wideband sonar. For tracking and association of the echoes reflected from the different body part, we developed an efficient approximation of the sequential ML estimator. The algorithm works in the natural time-space domain, which eases the exploitation of the a-priori knowledge about the human subject target. For classification of the activity, a weighted two level nested k-Nearest Neighbor classifier was applied on only four kinematic features. A set of experiments with five subjects, performing three different activity types of standing, walking, and swinging upper limbs, was carried out in a typical indoor environment. The proposed technology has managed to classify well the different activity types and demonstrated the potential of this technology for continuous assessment of various kinematic features of humans in indoor environment with reduced costs, under any light, smoke, or humidity conditions. This can be useful for instance for monitoring patients at home, and for detecting intruders.**

***Index Terms*— classification, k-NN classifier, human kinematics, sonar, and tracking.**

I. Introduction

IDENTIFICATION of an human subject's kinematics and characterization of his activity in different environments over time plays an important role in security [1] and medicine [2]. Human motion monitoring can help in detection of intruders and abnormal activities and send an alert, assist in the process of rehabilitation, design of treatment plans and follow-up monitoring [3], enable diagnosis and treatment of numerous neurological disorders [4], detect risk situations like falls in elderly people homes [5], and in hospitals, assisting the medical staff to monitor patients, in particular at night time.

Human subject kinematic assessment includes estimation of different body part position, velocity, and acceleration. The type of action is more abstract, has a temporal characteristic and can be divided into different classes such as standing, moving from sitting to standing, walking, falling, and jumping. Systems designed for human motion acquisition, can be classified by their technology, measurements, and processing methods. Methods based on Inertial Navigation System (INS) like in [6], or marker based optical systems like [7], require the sensor, or the marker to be attached to the body, which is sometimes not comfortable and furthermore, often requires battery replacement every few days. Among the non-contact methods for motion acquisition, the most common ones are based on optical, electromagnetic, and ultrasonic technologies.

Optical technology is commonly used in gait analysis laboratories [8]. It is usually implemented by a video recording system. It can enable estimation of a 3D pose and shape of human targets from multiple, synchronized video streams using an a-priori physical attributes of the targets, e.g., [9]. The estimation quality of optical methods is limited in range, requires calibration [10] and a-priori knowledge about the human subject, heavy data streams, and computational resources for enhanced resolution. Optical systems also cannot work in some conditions that are crucial for security like low light and smoke conditions.

Electromagnetic based technologies can be based on narrow band, or wideband signals. A narrow band radar has been used [11] for the detection and classification of patients' movements and location based on the Doppler effect. Reference [12] demonstrate how gait signature can be effectively captured by feature vectors based on Doppler effect. Reference [13] uses a classifier on the human body radar signature to characterize gait, in particular step rate and mean velocity. The Microsoft Kinect™ (Kinect), an active system, was recently developed for the game industry, and becomes more popular for applications for human activity acquisition [15]. It radiates infra-red radiation, and from its line of sight reflections, constructs an image [16]. The validity of Kinect to assess human kinematic data compared with an

[α]Equal contribution

Manuscript received July 5, 2013.

G.Blumrosen is with the Department of Zoology, Tel Aviv University, Tel Aviv, Israel (phone: 972-54-2286420; fax: 972-74-7249397; e-mail: gaddi.b@gmail.com).

B. Fishnam., was with the Tel Aviv University, Tel Aviv, Israel. (e-mail: benf22@gmail.com).

Y. Yovel is with the Department of Zoology and the sagol school of neuroscience, Tel Aviv University, Tel Aviv, Israel. (e-mail: yossiyovel@hotmail.com).

optical marker-based 3D motion analysis was recently performed in [17]. Still, the Kinect is restricted in range of around 2 meters from the system, and relatively a narrow location. An Ultra Wide-Band (UWB) radar, which uses a large portion of the radio spectrum, has recently been suggested for acquisition of body part displacement and motion kinematics [18]. The high EM transmission bandwidth yields accurate position location and possible material penetration. An algorithm for UWB radar-based human detection in urban environments was presented in [19]. These technologies emit EM radiation to the environment, suffer from multi-path fading, and are mostly limited in range.

A sonar system consists of an ultrasonic transmitter and receiver. It transmits a pulse to the medium of interest, and from its echoes, it can construct an image of the objects and activities in the medium. When a human walks, the motion of various components of the body including the head, torso, arms, legs, and feet produce an acoustic signature. A main degree of freedom in sonar systems is their pulse design. The two basic ultrasound pulses include Frequency Modulated (FM) wideband chirps, where the pulse starts with one frequency and changes gradually during the transmission to another frequency, and of Constant Frequency (CF) pulses [20]. The FM chirp enables high localization resolution and exploitation of properties of the bandwidth to detect different objects, while the CF pulse relies on detection of changes in the received frequency caused by the Doppler shift, and thus obtains information about the kinematics of objects in the medium. In some sonar systems, as well as in animal's bio-sonar CF and FM pulses are used simultaneously [20]. Echo processing techniques used by frequency-modulated bats that exploit both CF and FM pulses are shown in [21].

The Doppler modulations contain unique target features and thus can be used to characterize and classify human activity. A portable acoustic micro-Doppler system operating in a 40 kHz acoustic frequency range, has managed to identify different gait cycles based on the sonar signature [22]. Average speed of walking, torso velocity, walk cycle time, and peak leg velocity, can be extracted by the micro Doppler sonograms [23]. The performance of a range of classifiers and feature extraction algorithms were presented in [24]. Doppler signature can be used to distinguish between three human gait classes: one-arm swing, two-arms swing, and no-arms swing [25]. Reference [26] derived human kinematic features based on a model containing 12 body parts. Doppler based sonar system can assess the pattern of movement well, and can decompose the different body part movements, but cannot give precise information about the absolute location of different body parts in time and space due to their relatively long pulse duration and low correlation properties. Furthermore, due to its narrow band-width, it lacks the capability to distinguish between different static objects like stationary humans or walls. The absolute location of body parts and the ability to detect static humans is important for many applications in security and bio-medicine. While most existing Doppler based motion acquisition systems utilize a low frequency spectrum, there is much potential in exploitation of the richness of a wide spectrum [18]. A compressed chirp, like FM chirps, can give a precise localization of the object and contain spectral information in a large bandwidth that can be used to detect object structure and sometimes composition.

This paper presents a new wide-band sonar system based on Linear FM (LFM) chirp for human activity classification. A Maximal Likelihood (ML) criterion to derive target displacements over time based on the echo properties of delay, intensity, and the correlation of the echoes over space and time was derived. For tracking the echoes and the acoustic objects they represent, an approximation of the Sequential ML estimator was derived based on echo parameters. The tracker and the association of acoustic object to groups (clutters) that represent real targets in the environments are performed in the space-time domain. In the space-time domains, the different reflections from objects are described by their location over time, which enable exploitation of the a-priori knowledge about the human subject target in a relatively simple manner. Furthermore, it enables direct extraction of kinematic features, like target velocity, and target body parts' displacement variability, that have clear physical meaning. As a classifier, a weighted two level nested k-Nearest Neighbor classifier was applied on only four kinematic features of the clusters. The technology was verified by a set of experiments with five subjects, 3 males, and two females, performing three different activity types of standing, walking, and swinging upper limbs in a typical indoor environment.

The paper has three main contributions. A first contribution lies in advantages of the suggested technology in tracking human targets in indoor environment, in compare to the common optical technologies: noncontact (does not require attaching markers to different body parts), can work under any light conditions, and in the presence of smoke during a fire, or high humidity conditions in bathroom, unlike marker based optical technologies, and maintains the privacy of the subject in situations where is needed, like in bathroom, where the risk of falling is very high. A second contribution is in comparison to sonar systems based on Doppler technology. The wideband sonar, can give accurate information about human body parts location over time. This enables an enhanced classification of motion activity and eases the exploitation of the a-priori knowledge about the human subject target, compared to the commonly used frequency-time domains used in Doppler based methods. The high bandwidth also enables using the enhanced correlation properties of the wide bandwidth signal to classify between human and non-human objects. The third contribution lies in the original processing techniques that were tailored to the high bandwidth pulse characteristics. This includes the tracking of multiple targets in the room with low complexity and minimal a-priory assumptions, extraction of informative feature set, and the classification methods. The processing methods exploit the high accuracy distance estimations, the enhanced correlation properties of the wide-band signal, and integrate available a-priori knowledge about the human kinematics to the solution. The classification stage uses these features directly, and is divided to different

informative controllable stages that give maximal information about human kinematics.

This paper is organized as follows. Section II describes the active sonar modeling. Section III, describes the human kinematic modeling. Section IV, describes the data analysis methods for tracking, associating the multi-paths, and classifying the subject activity. Section V describes the experimental set-up for evaluation of the new technology. In section VI the experimental results are given and discussed. Section VII summarizes the results and suggests directions for future research.

II. Active Sonar Modeling

An active sonar node is composed of an acoustic transmitted (speaker), an acoustic receiver (microphone), and a processing and storage unit. A pulse is transmitted into the medium where the object of interest is located. The sonar receiver receives acoustical reflections of the transmitted pulse from the medium. The reflections convey information about object location, structure, and sometimes composition [27]. First we define the signal and propagation model. To adopt the sonar to tracking subjects in an indoor environment we give the basic sonar design considerations. Then we define the echoes' properties we use in our data analysis.

A. Sonar Signal and Propagation modeling

A received echo in the sonar is characterized by attenuation and delay [28]. The received signal for multiple pulse transmission, at time instance *t* is:

$$\mathrm{r(t)} = \mathrm{A_0} \sum_{\mathrm{m,k}} \beta_{\mathrm{m,k}} \mathrm{(t\text{-}mT)p(t\text{-}mT\text{-}}\tau_{\mathrm{m,k}}\mathrm{)} + \mathrm{n(t)}, \qquad (1)$$

where $p(t)$ is a transmitted pulse implemented by a LFM (Linear FM) chirp with a pulse width T_p, a bandwidth B, and a a peak energy E; m is the pulse index and T is the pulse repetition time, i.e., the interval between 2 pulses; $\tau_{m,k}$ is the k'th echo delay in the m'th pulse; $\beta_{m,k}$ is its related attenuation factor which is commonly assumed constant during the observation time and is affected by geometrical factors and atmospheric attenuation factors, which depend on temperature, humidity and frequency; A_0 is a gain factor determined by the sonar bearing angle and the sonar received and transmitted radiation pattern; and $n(t)$ is an additive noise component. The noise includes thermal and system noise which can be modeled by white Gaussian processes, and distortion from non-linearity of the speaker membrane.

The sonar system can be extended to a set of multiple sensor nodes. Each sonar node is capable of sensing motion features in one dimension (1-D). To assess motion in three dimensions (3-D) at least three senor nodes employed in different locations are needed [29]. Each object in the environment reflects the signal according to its cross-section. The cross section depends on the object's material, surface, size, and on the transmitted pulse's frequency range. Figure 1. shows a 1-D sonar system in a scene with a person, wall, and a chair. From each object there are different reflections of the transmitted wave sound, observed in the receiver.

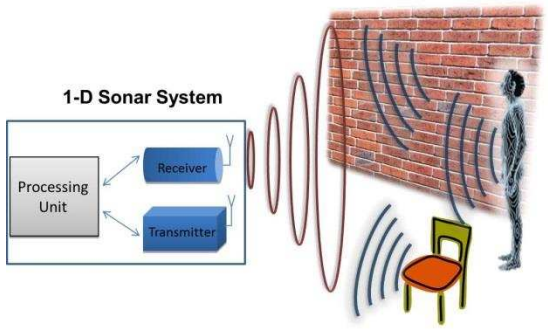


Fig 1. 1-D sonar based motion acquisition system in an indoor environment. The red circles represent the transmitted spherical wave, and the blue curves, the returning waves from the objects.

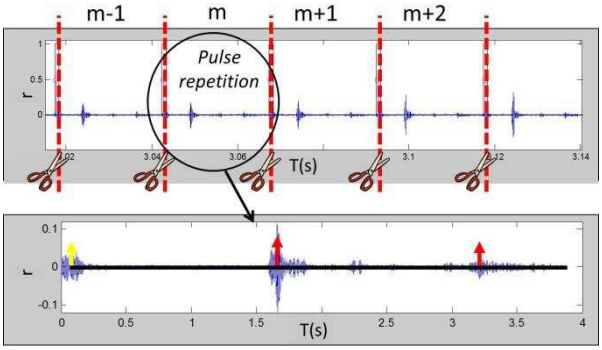


Fig 2. Extraction of echoes from the continuous received signal. The figure in the bottom, includes two main echoes in the period of the *m*'th pulse repetition. Each section of the signal depicted between two red lines represents one row in r_0.

The received signal in (1) is sampled every T_s seconds. The samples of M consecutive pulse repetitions are stored in an observation matrix $\boldsymbol{r_0}$ of size $M \times N_h$, where $N_h = T/T_s$ is the number of samples for each pulse repetition. A row of $\mathrm{r_0}$ includes the received signal samples which represent the different echoes of one transmitted pulse (Fig. 2). These echoes are related to different reflection of the pulse at different locations in the medium and therefore are related to spatial dimension. The pulse repetition period is defined such that the last echo from the scanned space returns before the next sonar emission is emitted.

B. Indoor Tracking Sonar Design Considerations

The sonar system design for tracking human subject should be tailored to the target and environment. There are several fundamental parameters that are assessed in most sonar systems: target range, target location and orientation, target size, target velocity, and target spatial-temporal pattern [21].

The pulse energy and bandwidth need to be high enough to enable tracking the target in the indoor environment. The spatial resolution increases with frequency. Thus, the higher the frequencies are, the smaller objects will be detected. Bats use frequencies of up to 150 kHz, which enable them detecting objects of size of less than a centimeter from a distance of a meter [20]. A pulse composed of high frequencies, suffers from distortion of the signal due to increase attenuation in the high frequencies [21]. of the signal. For tracking human subjects' body parts, a spatial resolution of few square

centimeters will be adequate. Thus a high frequency of 60 kHz, which represent a wavelenght of around a centmieter, will be adequate for tracking human subject in an indoor enviroement, and yet will not suffer from signal distorrtion due to atteenuation. For the lower freqeucny we can choose freqency of over 20 kHz, which is the upper hearing range of humans. With pulse bandwidth between 20-60 kHz, body parts and large scatterers in the medium, like walls, will reflect most of the transmitted pulse with minimal distortion. Reflections from small body parts with surfaces of a few centimeters, or a textured surface with different distances from the sonar will have a varying pattern, which can be significantly different from wide reflectors like walls [30].

The maximum range at which a target can be located has to guarantee that the leading edge of the received backscatter from that target is received before transmission begins for the next pulse. This range is called maximum unambiguous range and is given by:

$$R_E = \frac{v(T-T_p)}{2}, \quad (2)$$

where v, is the speed of sound and is around 343 m/s.

From the unambiguous range, the maximal Pulse Repetition Frequency (PRF) can be obtained, $v/(2R_E + vT_p)$. For instance, for indoor environment with distances of up to 4 meters, and of pulse duration of around 2.5 ms, the maximal PRF is 48 Hz. The minimal PRF, is obtained from the minimal range, which is determined mainly by the transmission delay, and in case of short FM chirp, is usually very low.

The Doppler shift is proportional to the target velocity relative to the sonar, and inverse proportional to the transmission wavelength. A LFM pulse is tolerant to Doppler shift of up to 10 percent of its bandwidth, $B/10$ [31]. For example, a 40 KHz Linear FM chirp would be tolerant to Doppler shift of up to 4 kHz, which is significantly more than the typical Doppler shift range, which is up to 1 kHz. Therefore, for human detection, the Doppler effect of the LFM based sonar can be neglected.

The achievable range resolution of a sonar system depends on the range of the transmission bandwidth [31], and is for example, for a 40 kHz transmission, in the range of few centimeters. Therefore, for tracking coarse human movements, a bandwidth around 40 kHz, can be adequate.

III. Human Movement Modeling

A human body can be described as a combination of body parts (BPs). While the subject performs different kind of activities, each of his/her body parts has a typical kinematic pattern, which can be captured by the body part displacement over time [13]. The kinematic features of the human can be derived from the group of discrete numbers of body part displacements [17]. Each body part can be divided into a relatively static component (e.g. torso, head), and to dynamic (e.g. upper and lower limbs' movement while walking) components. The displacement of the *l*'th body part in a Cartesian coordinate axis, in reference to the sonar location at instance time *t* is:

$$d_l^B(t) = d_{l,s}^B(t) + d_{l,d}^B(t), \quad (3)$$

where $d_{l,s}^B(t)$ is the *l*'th BP's absolute displacement component of very slow movements at instance time *t*, $(x_{l,s}(t), y_{l,s}(t), z_{l,s}(t))$, which reflects relatively static displacements with low velocity and standard deviation, usually with frequency content of less than 0.5 Hz, e.g. torso movement; $d_{l,d}^B$ is the *l*'th echoes' absolute displacement component of higher velocities and standard deviation at instance time *t*, $(x_{l,d}(t), y_{l,d}(t), z_{l,d}(t))$, with frequency content in range of 0.5-3 Hz, and is related to motion of different body parts like legs [32] and arms during performing daily life activities like lifting a bag, or walking [33]

It can be more informative to use the displacements relative to the torso, instead of the absolute body part displacement. In walking, where the whole body moves, some of the body parts, like the head, will have a relatively constant displacement from the torso, while the upper and lower limbs, will have a periodic displacement pattern.

IV. Data Analysis

The echoes' properties can be used to assess the human kinematics features by using advanced signal processing methods. The analysis is applied for one sensor in a single axis, which can be described as the radial axis (from the microphone outside).

Human activity classification in our system can be divided into four main stages. In the first stage, properties of received echoes like range, intensity and correlations are used to detect acoustic objects and track their location over time using a variant of sequential MLE object tracking [34]. These acoustic objects, referred in the paper just as objects, are sub-objects of a target, or of a real object in the medium. They are related to one or more echoes that shares similar echo properties and are

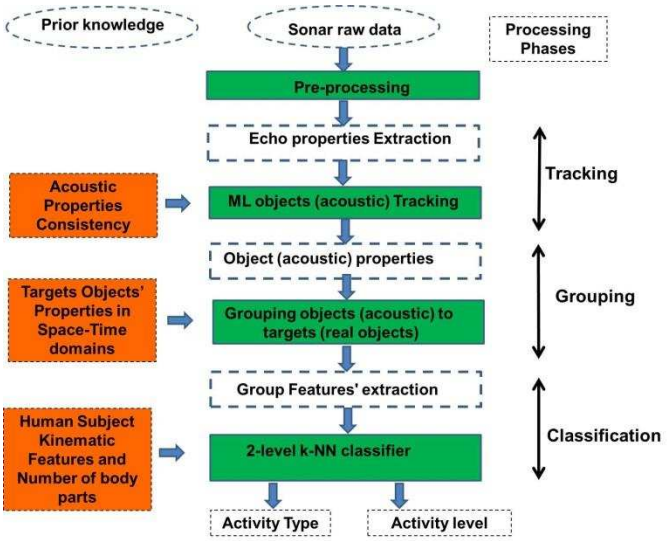


Fig 3. Data processing flow consists of three main phases: tracking objects (acoustic), associating these objects to groups that represent targets or real objects, and in a final stage, classifying the human activity type and level. At each phase, a typical prior knowledge is used: at the acoustic object tracking stage, the continuity of the echoes; at the grouping phase, the properties of target's dimensions, velocity, and standard deviation of its body parts in space and time; and at the classification phase, the kinematic features, and the number of body parts are used

reflected from the same location [35]. In the second stage, the objects (acoustic) are mapped to different groups (clusters) that represent the real target objects. In the third stage, features of the different object groups, like group average velocity, are derived. In the fourth stage, the features are used to distinguish between human and non-human object groups, and to estimate the activity level and activity type of object groups that relate to humans. Figure 3 shows the data processing flow.

A. Acoustic Objects Detection and Tracking

All acoustic objects' (referred as objects, in compare to target or real object) displacements in the medium over the observation time are estimated by using Multi-Target Dynamic Sequential MLE Tracking technique [36], which is a variant of a Sequential MLE Tracking technique. Before the tracking, a pre-processing stage on the raw data is performed, and an echo processing stage in which the echoes' properties of range, intensity and correlations, are estimated.

Pre-processing stage

Pre-processing of the measurements include a Band Pass Filter (BPF), match-filtering, frame synchronization, and echo selection. A BPF on the received samples removes frequencies that are out of the transmission band, and are related to interferers and other noise sources. The match-filtering operation detects the received echoes delays from The observation matrix $\boldsymbol{r_0}$. The matched filtered outputs during the observation time are stored in the matrix $\boldsymbol{r}$ of size $M \times N_h$. To estimate the set of delays in the *m*'th frame, $\{\tau_{m,k}\}$, a peak detection is used on the on the match filter output. The delays are relative to the start of the frame. Frame synchronization is performed to estimate the start of frame. To reduce the computational resources, and exclude some of the noise, high Signal to Noise Ratio (SNR) echoes are selected using a Detection Threshold (DT), which is related to the size of the detected object and the noise tolerance of the system.

Echoes' properties Extraction

The basic echoes' properties of delay, intensity, and echoes' spatial-temporal pattern can be used to detect and classify the different targets (objects) of interest.

The distance between the *k*'th object and the sonar system at instance time *m*, can be denoted by d_k^m, and is the round trip time divided by a factor of two. Object position can be obtained by the range, azimuth and elevation angles from the sonar to the object, and by using multiple sonar sensors located at different locations, using statistical or geometrical methods. In Doppler based systems object velocity can be assessed by the measure of Doppler shift. In a wideband sonar with a high Signal to Noise Ratio (SNR), the velocity can be estimated by deviation of the object location estimations over time [31].

Object dimensions can be estimated by analysis of the number of echoes reflected from an object, their spatial spread, and by their energy [21]. One indication about object's dimension is its related echoes' intensity [37]. The intensity is normalized by a factor of the range, to enable tracking objects in different locations. The normalized intensity of the k'th acoustic object, at instance time m, is given by:

$$I_k^m = \frac{A_0 \beta_{m,k} \sqrt{E}}{(a_k^m)^{NF}}, \quad (4)$$

where $\beta_{m,k}$, $\sqrt{E}$, and A_0, are the channel attenuation, peak amplitude that is the square root of the peak energy, and the gain factor, as defined in (1); k is the echoe index that relates to the *k*'th acoustic object, and NF is a normalization factor, which for a 2D object is around 2 [31].

The spatial-temporal correlations of the different received echoes indicate on the object characteristics. In particular, it indicates on the feasibility of the spatial-temporal pattern to be used for association of different echoes to different objects. The spatial temporal correlations can be split to spatial correlation, of echoes reflected from different objects in the same time instance, and to temporal correlations, of echoes related to the same objects over consecutive time instances. Let us denote $\rho_{k,l}^m$, and $\rho_k^{m,m+1}$, as the spatial correlation coefficient between the *k*'th and the *l*'th echoes at time instance *m*, and the temporal correlation coefficient of the echoes that are related to the *k*'th object at time instance *m* and *m*+1.

Tracking Criterion

The Maximum Likelihood (ML) criterion to extract from the match filter output $\boldsymbol{r}$ the set of *L* body part displacements that belong to a human subject is:

$$\widehat{\boldsymbol{d^B}} = \text{argmax}_{\boldsymbol{d^B}} P\{\boldsymbol{r}|\boldsymbol{d^B}\}, \quad (5)$$

where $\boldsymbol{d^B} = [d_1^B, \ldots d_l^B \ldots d_L^B]$ is a matrix of the *L* body part displacements in 1-D over observation time *M*; and $P(\cdot)$ is a probability distribution function.

The received signal of one pulse repetition includes different echoes with a profile that change over time as the target and other objects in the medium move. Since the channel and target distribution functions are not linear, solving the non-linear likelihood function in (5) for multiple objects is cumbersome. The solution for two targets is given as a function of posterior likelihood probability densities of each target, and target detection probability [38]. The posterior likelihood probability densities are hard to estimate, depend on many parameters, and require extensive a-priori knowledge, which is commonly not fully available in targets like humans that change their position, activity, and shape continuously.

A simplified solution to (5) can be obtained by splitting the solution into two stages. In a first stage, all reflections from objects in the medium, including reflections from different targets like human and other scatterers like walls or chairs, are tracked using an acoustic tracking algorithm. The algorithm exploit constraints that reflect the difference in different reflections' characteristics, like location, and correlation, and the continuity of motion in space and time. In a second stage, the different acoustic objects are mapped to different groups that relate to the different targets. At this stage, an a-priori knowledge about the targets, in space-time domains can be used, instead of the cumbersome a-priori distribution functions that is needed for solving the general MLE.

For the acoustic object tracking, a constrained MLE criterion can be defined as:

$$\widehat{\boldsymbol{d}} = argmax_{\boldsymbol{d}} P\{\boldsymbol{r}|\boldsymbol{d}\}, \quad (6)$$

s. t.

$$|d_l^m - d_k^m| < \delta_D^S, |I_l^m - I_k^m| < \delta_I^S, \rho_{l,k}^m < \delta_\rho^S$$
$$|d_k^{m+1} - d_k^m| < \delta_D^T, |I_k^{m+1} - I_k^m| < \delta_I^T, \rho_k^{m,m+1} < \delta_\rho^T$$

where $\boldsymbol{d} = [\boldsymbol{d_1}, \ldots \boldsymbol{d_k} \ldots \boldsymbol{d_K}]$ is the set of K objects' displacements vectors over time in the medium that are associated with Q groups (one group can be the target human), d_l^m, and d_k^m, I_l^m, and I_k^m, and $\rho_{l,k}^m$ and $\rho_k^{m,m+1}$, are the *l*'th and *k*'th objects displacement, intensity, and the spatial and temporal correlation coefficient as defined above. The coefficients $\delta_D^S, \delta_I^S, \delta_\rho^S, \delta_D^T, \delta_I^T, \delta_\rho^T$ are the corresponding spatial and temporal thresholds on the parameters that depend on the medium and on the pulse properties, and can be determined experimentally.

The constraints are based on spatio-temporal correlations of the different echoes, and can be separated to temporal constraints between successive pulse repetitions, and spatial constraints between objects in different ranges. Echoes close in time and space are more likely to be related to the same object, and to have close delays, intensity, and to have more correlation to each other. In case the constraints are tailored only to a specific to the target, the solution to (6) will coincide with the solution to (5). The specific target constraints can be based on the a-priori knowledge about a target [39], like target size or shape. Still, these constraints are not always available, in particular when tracking dynamic scenes with targets that change their shape.

Sequential MLE approximation

The tracking problem (6) can be solved using the sequential MLE [34] for all (acoustic) objects in the medium. This can be implemented with a Viterbi algorithm [40] with a distance metric error that maximizes the constrained probability function in (6).

A trellis diagram is used to represent the different objects' possible locations over time. For *M* discrete locations, a state at time interval *m*, S_k^m and S_l^m, represent the location of the *k*'th and *l*'th objects. A path in the diagram is a transition between states at consecutive discrete time intervals. Each possible transition represents a possible motion of the object from one position to another. The transition between the states depends on the PRF, and on the motion. Slow motion with high PRF will have fewer transitions in the trellis diagram. Each "legal" transition between states at time instance *m*, can be defined as a branch with a branch metric $M_{k,l}^m$,which is a function of the similarity between consecutive states. The branch matrix between objects *k*, and *l*, at instance time *m* can be defined similar to [41], to maximize the likelihood ratio of:

$$M_{k,l}^m = \frac{p(r|d_k^m)}{p(r|d_l^m)}, \quad (7)$$

This branch metric is a function of the distance between two states, the pattern of the echo, and the echo intensity [34]. The branch metric can be normalized to values between 0 and 1 to represent a probability function. The branch metric between objects *k*, and *l* at time instance *m*, can be approximated by the following analytical function:

$$M_{k,l}^m = e^{-\alpha \Delta D_{k,l}^m} \left(I_{k,l}^m\right)^\beta \left(\rho_{k,l}^m\right)^\gamma, \quad (8)$$

where, $\Delta d_{k,l}^m = |d_k^m - d_l^{m-1}|, I_{k,l}^m = \frac{min(I_k^m, I_l^{m-1})}{max(I_k^m, I_l^{m-1})}$, $\rho_{k,l}^m$, are measures of distance, intensity, and cross-correlation between the *k*'th and the *l*'th objects at time instance *m*, and α, β, γ, are constants that are determined experimentally and reflect the reliability and significance of the distance, intensity, and correlation measures to the detection probability respectively. The metric in (8), approximate the metric in (7), and maintain the asymptotic properties of the optimal solution of (7) [42].

An object *i*'th path metric is the sum of the branch metrics that are related to the objects in the time interval W:

$$C_i^m = \sum_{m'=m-W+1}^{m'=m} M_{i,j}^{m'} . \quad (9)$$

The time interval is called a constraint length and must be big enough to reflect enough statistics to detect the object. But too long constraint length, will result in accumulation of estimation errors, and will affect the tracking of fast movements. An object *j* at instance time *m*, is selected to be related to an object *i*, according to the following criterion [15]:

$$\hat{\jmath} = argmax_{j'}(C_i^{m-1} + M_{i,j'}^m). \quad (10)$$

To enable flexibility of the tracking scheme and for tracking dynamic objects that can enter or exit the range of the sonar and change their object properties in time and space, the implementation of the solution to (6) in the trellis diagram includes object creation and deletion, splitting from one object to two objects, and merging with existing one.

In case a new object enters the sonar coverage range, a creation process of a new object will start. If there is no other object in the trellis diagram with a close enough metric to the new one, and the object exists over certain threshold duration, which is usually in range of the constraint length, then the a new object is created in the diagram. In case an object leaves the sonar coverage range, or gets too far from the sonar and the intensity goes below the detection threshold, the object path is cut in the trellis diagram. Objects can be split one from another. For example, in case of a movement of a body part away from the torso, like lifting an arm, where the object exceeds the constraint length, a new object that relates to the torso will be created. In the trellis diagram, the new object is seen as a split of the branch metric of the previous object. In case an object, like an arm, returns back to the body, two or more reflections from one object can be merged. A multi-path combining algorithm can be then applied. A post processing stage, to filter-out discontinuities and mitigate over missing estimations of an object due to noise, or scatterers similar, is performed.

Figure 4 shows the trellis diagram during tracking of three objects: A, B and C. In Figure 4.a, the states are the set of the distances between the object and the sonar of the detected echoes marked by ellipses. The green color ellipses are states that are not chosen. The pink ones are the location of the three objects over time. The arrows in Figure 4.b represent the chosen branch metric magnitudes. The sum of branch matrices with the lowest value over the constraint length is chosen and

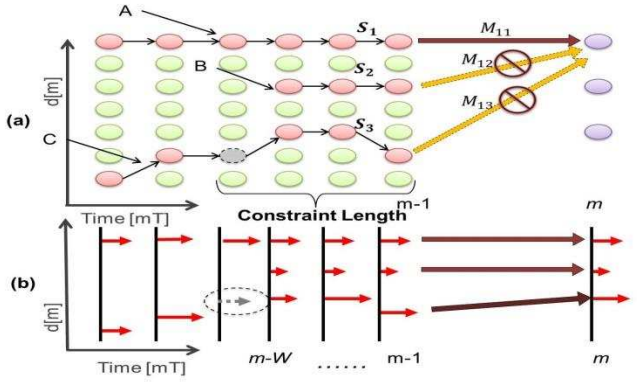


Fig 4. A trellis diagram that implement the tracking algorithm in (10) for three objects: A, B, and C. In panel a, the states are the set of the distances between the object and the sonar of the detected echoes marked by ellipses. The arrows in panel b represent the chosen branch metric magnitude corresponding to the states in panel a. The sum of branch matrices with the lowest value over the constraint length (4 pulse repetitions in the example) is chosen and maximizes the ML probability criterion. Object A, represents an object (optical) of a static body. The algorithm is capable of mitigating for misdetection of object C (shown in blue color) by interpolation, and to dynamically delete and create new objects without prior assumptions, like the creation of object B.

maximizes the ML probability criterion in (10). Object A is a static object. Object B will be created if it lasts for more than a pre-determined time, in the constraint length. Misdetection in object C, marked by a light blue color, is mitigated by interpolation, and becomes part of object C path over space and time.

B. Grouping to real Objects

The different objects (optical) are mapped to groups in a segmentation process, in which objects (optical) are assigned to different groups (clusters) that statistically relate to real objects (targets) in the environment [35]. The assigned groups and their related object properties are used for targets' classification.

Many segmentation and grouping processes are supervised, and a training set is needed. Methods without or with minimal training usually require a-priori assumptions [43]. Reference [43] carry out a detection phase using an unsupervised Markov random field (MRF) model and assuming a-priori spatial information on the physical size and geometric signature of the objects.

For tracking human subjects, after the stage of tracking the objects (acoustic), a-priory assumptions about the subject features can be utilized in a simple manner, compared to the common assumptions about the a-prior distribution of the reflections from the body [43]. Such assumptions can include the subject body size, the maximum spread of his body parts, and his kinematic properties, like trajectory velocity. This a-priori knowledge can be translated to constraints on echoes intensity, correlation, and spatial-temporal distribution on the space-time diagram.

Let us define ${I_G}^m$ as object mapping index vector at instance time m in the length of number of objects, K. Let us denote by $\boldsymbol{O^K}$, the set of objects properties that relate to K objects. An object k is mapped to the j'th group G_j, if the output of the mapping is j ${I_G}^m(q) = j$. A criterion to the grouping of multiple objects can be defined as:

$$\hat{\boldsymbol{I}}_{\boldsymbol{G}}{}^{\boldsymbol{m}} = argmax_{{I_G}^m} P\{\boldsymbol{O^K}|{\boldsymbol{I}_{\boldsymbol{G}}}^{\boldsymbol{m}}\} \quad (11)$$

The criterion in (11) applies that an object will be related to a group, if its probability is higher than the one that the object is related to the other group. Solution to the criterion in (11) is complex [41]. A simpler criterion can be to associate to each target its object separately based on it's a-priori knowledge, and then, in a later processing stage, exclude estimation noise from the targets estimations.

The set of indexes of objects mapped to the j'th group G_j at instance time m is denoted by ${I_{G_j}}^m$. A MLE criterion for mapping objects to the j'th group can be defined as:

$$\hat{\boldsymbol{I}}_{\boldsymbol{G_j}}{}^{\boldsymbol{m}} = argmax_j P\left\{\boldsymbol{O^K}|{\boldsymbol{I}_{\boldsymbol{G_j}}}^{\boldsymbol{m}}\right\} \quad (12)$$

s.t. (O_j) ,

where $\hat{\boldsymbol{O}}^{\boldsymbol{K}}$ is the set of objects properties related to the approximated displacement vector, $\hat{\boldsymbol{d}}^{\boldsymbol{Gj}}$, and $\boldsymbol{F}(O_j)$ is the set of attributes of the j'th target, which are assumed to be known a-priori.

To solve (12) with the constraint the probability function of the object properties can be used [44]. The probability function of the targets is not deterministic, and hard to evaluate. A simpler approach, can search in the space-time defined by the trellis, set of objects in the area defined by the target size [43]. The area boundaries can be set around the center location of the probability, which is usually denser or with higher intensity. The central object, which is usually more consistent and strong over time, is denoted as the main object, and the other objects in the group, as sub-objects.

A simple realization of the algorithm can first detect a main object along a moving window of size $\boldsymbol{W}$, and in boundaries determined by the maximal spread of body parts in the space-time diagram, ΔD. The main object displacement can be chosen according to the maximal intensity, in a recursive manner, according to:

$$\boldsymbol{d_j^{m+1}} = \boldsymbol{max}\{\boldsymbol{I}(\boldsymbol{O^K})\}_{\boldsymbol{d_j^m-\Delta D/2,m-W}}^{\boldsymbol{d_j^m+\Delta D/2,m}} \quad (13)$$

where $\boldsymbol{I}(\boldsymbol{O^K})$ is matrix of intensities of the K objects.

The criterion in (13) can be extended to include other attributes of the constraints $\boldsymbol{F}(O_j)$ in (12). Such attributes can be group velocity, constituency over time, and distribution of the target group objects [44].

After determination of the main object, all other objects in the space-time are assigned as sub-objects. As a post processing stage, the groups can be merged or split based on MLE similarity measure of the different features of the groups. This can also enable merging groups that correspond to more than one different acoustic objects' groups, which are reflected from massive reflectors like walls in NLOS environment. For example, reflection from the floor of the subject echoes.

For determination of human motion kinematics, the objects can be divided to additional categories, according to their size, location, and kinematics statistics. A fundamental category is of dynamic and static objects. Dynamic objects are sub-objects that fluctuates more than a certain threshold, usually around

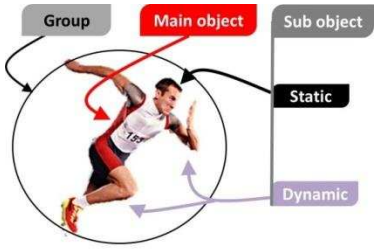


Fig 5. Determination of object type in a group. The strongest reflection is usually marked as the main object in a group, and the other are called sub objects. The sub-objects are divided to static and dynamic ones, and can be used to form a feature set.

the main body of the group, e.g. lower and upper limbs are dynamic parts, while walking. Static objects are sub-objects that are relatively static in relation to the main-body (torso), e.g. head. A threshold on the standard deviation of the object location from the main object location can be used to determine if an object is dynamic or static.

For example, three objects of torso, head, arms, and legs in the human scheme in Figure 5, will be part of one group, that relate to human. The torso will be the main object and the head and hand will be sub objects. The arms and legs will be dynamic objects, and the torso, and head, which do not move relative to the main body, will be considered as static objects.

Similar to the object tracking stage, a post processing is performed on the results of the grouping. It includes exclusion of objects that appears for only short instance time, or its intensity is below a threshold, by filtering, that apply the range, intensity and correlation group constraints. A low pass filtering can mitigate over inaccuracies in the interpolation and merging operations.

C. *Features' Extraction*

To classify groups related to humans, as opposed to groups related to non-human objects, significant features should be extracted. This enables tracking only the desired object (human), and excludes clutters, like walls, based on the statistical characteristics. The second goal is to extract features that can be used to classify activity level and type. The features can be divided into kinematic features that relate to the distribution of displacement in the group, and to more static features that are based on correlation properties [45].

The dynamic features at instance time m can be: 1) the average velocity of the main object, v_m^{Gi}; 2) average standard deviation of the group, σ_m^{Gi}; or 3) number of dynamic objects in the group, N_m^{Gi}, over a time window.

Static features depend more on the object's size and pattern. In radar systems, where a large portion of the radiation can penetrate through the body, the reflection can be changed in relation to tissue composition [18]. In sonar systems, the reflection is mostly from the body surface [27] and therefore, human detection can be obtained by its distinct body surface structure. These features can be based on echoes' pulse autocorrelation properties. For example, $\bar{\rho}^{Gi}, \sigma_\rho^{Gi}$, mean and standard deviation of the auto-correlation function of the main object's echoes over a temporal window, can be used to distinguish between different objects, in particular between echoes that return from a human and ones that are reflected from other objects.

D. *Human Motion Classification*

Human motion classification is related to its sonar signature. Different objects, different people, and people doing different activity types, have different sonar signatures that characterize their body properties and kinematics [45] . In addition, the kinematic features that are unique to humans [45], can be aggregated [32]. To classify human activity level or activity type, kinematic features of the body that form the sonar signature need to be derived. The classification can be obtained by using the group features together with the features of each object in the group like object velocity, range, dimensions, and pattern of change in time and space [46]. The classification can be formed in one stage or different stages. Dividing the classification process into multiple stages makes the analysis more observable and helps in controlling its parameters.

Activity level

The measure of human activity level is important indication for patient monitoring. The features $v_m^{Gi}, \sigma_{m,}^{Gi} N_m^{Gi}$, of group velocity, sub-objects standard deviation, and number of dynamic objects, respectively, can be used to assess physical activity. The velocity can be an indication if the human walks. When a human is doing an action that involves the arm and hands movements, while staying in one location, dynamic body parts will change their location, while the main body will remain in one place. A simple classifier can be linear sum of the features, where each feature is weighted in the summation according to its significance to the desired goal using a feature selection algorithm [12].

Activity type

To classify simple activity types, the dynamic features of an average window can be used [32]. These features will enable classifying between fundamental activity types like standing, walking, and doing physical activity that include movement of body parts like hands or legs. To classify more complex activity types like sitting or lifting a bag, the classifier must use a short observation window and a state diagram over time. More features should be added, including body part size and acceleration. A wide enough training set should be used. Since the sonar system can also detect and classify other static or dynamic objects in the room the context information can be added. For instance, if the human subject approaches an area that is known to be a library, and then lifts his hands, we can assume that the subject is taking a book from a shelf.

E. *Classifier implementation*

A simple and effective classifier commonly used in sonar systems is the k-Nearest Neighbor (k-NN) classifier [24]. An object is classified by a majority vote of its neighbors in the feature space, with the object being assigned to the class most common amongst its k nearest neighbors in the feature space

(*k* is a positive integer, typically small). The basic high dimension feature space k-NN classification suffers from over-sensitivity problems due to irrelevant and noisy features. The k-NN classification accuracy can be improved by selecting relevant features, assigning a weight to each one [47], combining multiple classifiers using decision trees, and by selecting the best *k* value in a dynamic way [48] .

A k-NN classifier with 2-level decision trees, optimal feature weighting and k values can therefore be a good compromise between efficiency and performance. One implementation can use a first decision level classifier that distinguishes between human and non-human groups. Second decision level classifiers can then be used to classify activity types and levels. This implementation can suffer from trailing classification error from the first level, but can minimize these errors by a learning algorithm over time.

Another implementation suitable for environments where the non-human objects are static includes two decision level classifiers. The first classification can be between static or dynamic objects, using a set of features that are related to motion kinematics. The second level classification can be divided to two classifiers: 1) classification of the static object between non-human object and humans; 2) classification of the dynamic objects by different activity types and levels using the dynamic features.

For the dynamic classifier, motion kinematic features like main body velocity, number of dynamic body parts, and the average standard deviation of the sub-object in the group can be used. For the static classifier, correlation measurements can be used to distinguished between static object and humans as they have different reflection surfaces and properties [49]. The output of the static classifier can be feedback to the activity classifier, decrease trailing error, and include a static position of the human in the activity classification, and thus enhance the overall classification accuracy. The features can be weighted at each level according to their relevancy to the specific classification, and the *k* value can be higher in the first level, and lower in the second decision level in order to reduce trailing error. The decision tree, when applied for short time instances, and with a state diagram, can be applied to classify more settle activity types, like lifting a bag. Figure 6 describes the decision tree for the classifier.

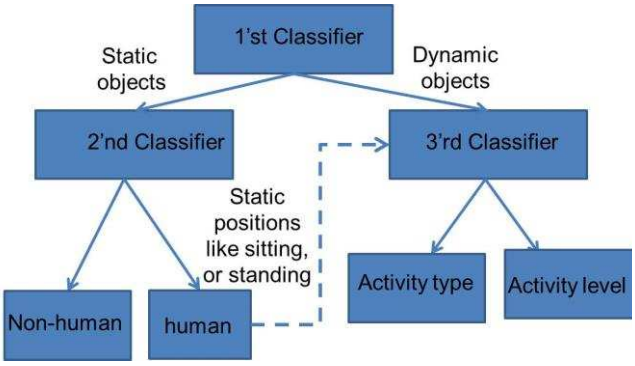


Fig 6. Two level decision tree k-NN classifier for activity type and activity level estimation. The non-human objects are assumed to be static. The static positions of the human can be aggregated by the third classifier, in case complex activities are classified.

V. Experimental Setup

The experimental setup is designed to produce a first order feasibility test for the technology and to evaluate the system's performance for classification of different fundamental activity types. The experimental setup includes a sonar system, a reference video system, and a processing unit. The experiment was performed in a 4x3x2.5 m^3 non-acoustic room, in Tel-Aviv University, in Tel-Aviv, Israel.

The sonar system is shown in Figure 7. The processing unit was a laptop (Dell, Vostro), and the reference video system was a webcam (Logitech, HD 720p) with 30 frames per second. The sonar system was composed of an ultrasonic dynamic transmitter (speakerphone) and receiver (Avisoft INC), and a synchronization cable. The speaker and the microphone were connected through a D\A converter (Avisoft INC, UltraSoundGate Player 116) and an A\D converter (Avisoft INC, UltraSoundGate 116Hm) to the laptop. The sampling receiver rate was 500 kHz. Both ultrasonic transmitter and receiver were directional, with beam width of around 30 degrees [50]. Frequency responses and beams of the speaker and microphones can be found in [50]. The directionality of the sonar spatially filtered clutters and scenes that were out of the area of observation. The pulse was a linear upsweep FM chirp with frequency range that moved from 20 to 60, kHz, which resides in the frequency range defined in II.C. This frequency range is above human hearing, enables range resolution in scale of a centimeter, and has a neglected Doppler shift. The chirp was windowed by a Hanning window to avoid clipping. The peak and Root Mean Square (rms) values of the transmitted pulse were 102, and 92.5 dB SPL (at 1 meter), respectively. These values were adequate for tracking targets larger than of a few square centimeters, which is the size of a typical body part. The pulse repetition rate was 40 Hz, which was assumed to be adequate for tracking normal motion patterns. The experimental setup is shown in Figure 8.

For first order feasibility of the tracking and grouping algorithm described in IV.A, and IV.B, 8 different experiments with multiple people performing different activity types were performed in an indoor environment. In each experiment, the position of one or two subjects was estimated along 8 seconds. The human subjects were located inside the range captured by the sonar system, and were either standing, walking, or standing with arbitrary arm swing. A reference video system was activated simultaneously with the sonar system. The video frames were synchronized to the sonar transmissions by synchronization pulse, at start of each recording. Snap shots of the different experiment sets as captured by the video reference system are shown in Figure 9.

For the evaluation of the classification algorithm described in IV.C and IV.D, a set of 150 experiments with three different activity types was performed on five different people (3 males, 2 females) with 10 repeats for each of the three activity types. In each experiment a human located inside the range captured by the sonar system performed different activity types for 8 seconds: undirected walking, standing still in different positions, and standing with arbitrary hand swing. These activity types were chosen because they are

fundamental in human motion. In addition, these activity types can be used to assess also activity level. For each experiment, the sonar system was activated simultaneously with the video reference system.

The data analysis methods in section IV were applied to the raw data for the tracking and classification experiments. From the experiments, 70 percent of the data (105 experiments) were used as a training set, and the other 30 percent (45 experiments), were used as a test set. To increase classification reliability, the subsets of training and test sets were resampled randomly 10 times [51]. Artifacts were removed from the training set for each feature by using a median filter. Each experiment included only one known activity: walking, standing, or standing with swinging hand(s). The different groups were tagged according to different classes according to the video reference: non-human static targets and subject humans standing, walking or swinging hands. These tags were used by the classifiers for the training set.

Three k-NN classifiers were used in a similar manner to the one shown in Figure 5. For the first level classifier (between static and dynamic targets), the features of the human related group (assumed as the *i*'th group) were the average group standard deviation, σ^{Gi}, and the average main body velocity v^{Gi}. The velocity had double weight compared to the standard deviation. Artifacts were removed using a median filter with two standard deviations, and the value of k was 5. This high k value was chosen to produce a coarse clustering to reduce trailing errors. For the second human detection classifier, the features were group average standard deviation, $\sigma_{,}^{Gi}$, and average and standard deviation of main body correlation over time, $\bar{\rho}^{Gi}, \sigma_{\rho}^{Gi}$. The median filter standard deviation was 1.1, and the *k* value was 2, to produce fine classification between static targets like walls and humans. For the human activity classifier, the features were in addition to the group average standard deviation and velocity, along with the average number of dynamic objects, N^{Gi}. The classifier artifact filter was finer than the first classifier with threshold of 1.4 standard deviation, and the *k* value was 2. Table 1 summarizes the k-NN classifier configuration.

VI. Results and Discussion

A. Pre-Processing

The received signal in frequency-time domains for one pulse repetition is shown in the spectrogram in Figure 10. The experiment includes a human subject standing at a distance of around 1.5 meters from the sonar, and a wall in the background. The signal bandwidth is between 20-60 kHz, and the peak frequency is in the middle. The first echo is the one

TABLE I
Classifier Configuration

Classifier type	Features	k	Artifact filter (standard deviation)
Dynamic/Static	$2v^{Gi}, \sigma_{,}^{Gi}$	5	2
Non human static object /Human Standing	$\sigma_{,}^{Gi}, \bar{\rho}^{Gi}$ σ_{ρ}^{Gi}	2	1.1
Swinging Hands/walking	$v^{Gi}, \sigma_{,}^{Gi} N^{Gi}$	2	1.4

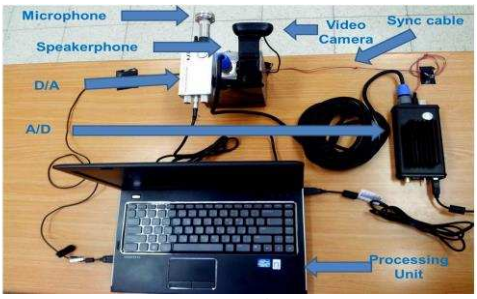


Fig. 7. The sonar system. It is composed of an ultrasonic speakerphone, the A/D and D/A converters, a cable to synchronize the transmission and reception, and a computer which functions as a processing unit.

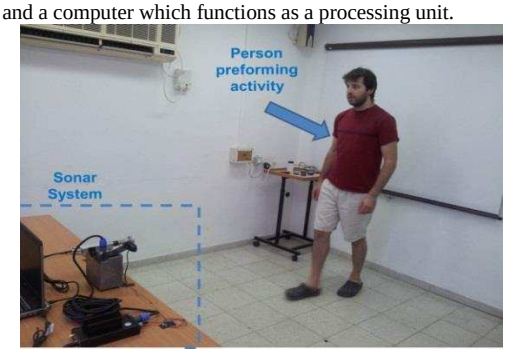


Fig. 8. Experimental setup which include a subject perfoming an activity, and s sonar system. The range between the subject and the sonar vary from 0.5 meter to 3 meters.

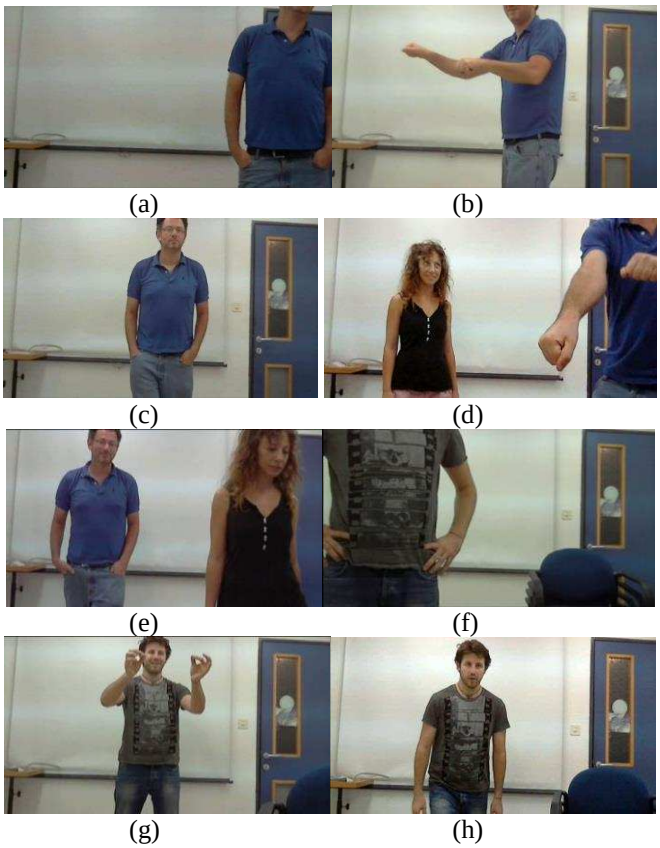


Fig. 9. Different human tracking experiment sets as captured by the reference video system. Figure 9.a, 9.b, and 9.c, demonstrate examples of a human subject standing, swinging upper limbs, and walking towards the sonar system and away, respectively. In experiments described in Figures 9.d, and 9.e, 9.f, and 9.g, and 9.h, two people are standing, one swinging hands and one subject standing still; two people are walking to and from the sonar system; a subject stands in front of a chair; a subject stands behind a chair swinging his hands, and a subject walks around a chair, respectively.

received at the moment of transmission directly from the speaker to the microphone, the second one, is composed mainly of two echoes related to the subject, and the third one to the echoes from the wall. The echoes from the wall are more spread out, due to the relatively large reflection surface.

The observation matrix was calculated at the beginning of each pulse repetition. Then, the received samples were match-filtered with the transmitted pulse shape. The results were threshold to exclude reflections from small objects and noise. Figure 11 shows the echoes' auto-correlation matrixes over time for the walking subject in Figure 9.c. The fluctuating stronger echo returns from the walking subject (Figure 9.a), and the later echoes return from the walls, and possibly from indirect reflections from other objects in the room (Figure 9.b). The wall echoes change less over time, and hence are correlated. The wall mean and standard deviation values were 0.709, and 0.116, respectively, compared to 0.617, and 0.1211, and one of the upper echoes related to the wall (Figure 9.b). The wall echoes change less over time, and hence are correlated. The wall mean and standard deviation values were 0.709, and 0.116, respectively, compared to 0.617, and 0.1211, of the human. The difference can be explained by the continuous change in the effective reflective surface, in the beam attenuation in different ranges and frequencies, and the cross section over time, caused by change in the subject's body parts' location and orientation while walking.

The metric parameter values α, β, γ, were chosen empirically in a way that minimizes the tracking error over a range of train experiments. The values were found to be $\alpha = 0.3, \beta = 0.2, \gamma = 0.03$. These values reflect the higher importance of the distance and the intensity, while the correlation value is less distinctive, and therefore has a low value. The correlation property, even with its relatively small weight, filters noise that has a very little correlation with the transmitted pulses.

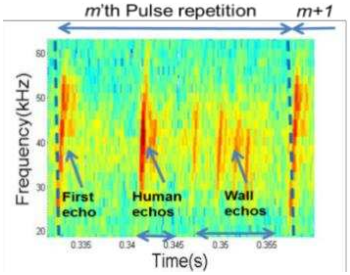
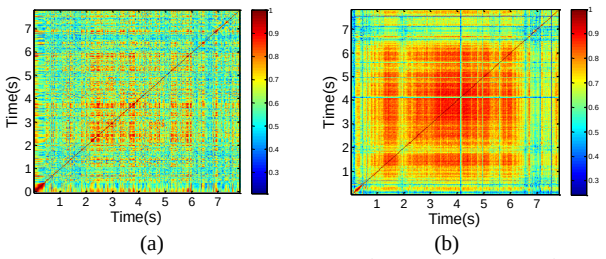


Fig. 10. Signal spectrogram for *m*'th pulse repetition for the experiment described in (Figure 13.a). The first signal is of the transmission, the second echo group is from the human, the third is from the wall.



(a) (b)

Fig. 11. Auto-correlation matrix over time of the main echoes of the wall (Figure 11.a), and the walking subject (Figure 11.b) for the experiment described in Figure 9.c. The wall is more correlated with a lower variance.

B. Object Tracking and Grouping

Figure 12 shows the processing stages of the object detection, tracking, and grouping algorithm as described in section IV.A. and IV.B in the space-time domains of the experiment shown in Figure 9.c. Figure 12.a shows the result of the match filter after the thresholding operation. The late echoes are related to the wall, and the two fluctuating ones are most likely the walking subject's body parts. According to the differential distance of around 30 cm, using the video reference, and geometric considerations, the sub-object that is nearer the sonar is the torso and the farther one is the upper body, near the head. Figure 12.b shows the results of the sequential MLE object tracking with more than 20 different objects. Figure 12.c shows the last stage in object tracking of post processing, in which missing estimations are mitigated for each object by interpolation. The next stage is the grouping performed on the objects in the space-time diagram. First, the main object is detected in the space-time diagram according to intensity and consistency over time according to (13). Then its related sub-objects, with lower intensity, are chosen. The results of grouping are shown in Figure 12.d, where each group has a different color.

Figure 12.e shows the results after a group merging and splitting algorithm based on similarity of the different group features. There are 4 main groups, where each group is marked with a different color. The blue group, with its continuously moving displacements, is related to the subject. The other three groups, marked by black, green, and red, are associated with static objects from the scene. The black group is associated with a nearby static reflector, probably from a surface near the sonar system, like the sonar itself. The green group can be associated with direct reflection from the wall, and the red group from wall sides, or from indirect reflections from the wall via the floor. Red and green are separated groups, since the criterion for grouping was based on detection of human motion and use the a-priori knowledge of human body parts' maximal span. When using only a-priori assumption about human target size, other groups will be "projected" to human groups, at least in the sense of group spatial dimensions. Note that the objects in the wall groups, mainly the red group, fluctuate over space and time. This can be explained by estimation errors, and by the effect of non-stationary human movement.

To derive the group features, the objects in each group are sorted into dynamic or static objects according to their standard deviation from the main object as defined in section IV.B. Figure 12.f shows the results of this process: main body (red) and its related static (black) and dynamic (purple) objects, which are in case of a human, his torso and body-parts.

Figure 13 shows the final grouping results for the set of experiments shown in Figure 9. The static walls in the background are well estimated in all experiments. There is more than one group that is related to the wall. This is due to the wide reflection surface of the wall, and the a-priory size of target which was defined as 1.2 meters to reflect the span of a subject's limbs, as this method is dedicated to tracking people.

In our method, there can be, multiple groups associated with a target. Applying in future, multi-target target approach can enable association of only one group to a target.

The main difference between standing still and swinging arms (Figures 13.a, and 13.b) is of the larger spread out of body parts that relate to the partial upper limb movement captured by the sonar. (The human group is colored blue; the walls are green and red). The change in the arm and hand orientation and effective surface, and in case of very rapid movement, insufficient pulse repetition rate has caused occasional discontinuity in location estimations over time. The experiment of walking without moving the upper limbs (Figure 13.c) shows two distinct body parts (human in red color). Comparing the results to the video reference shows that the main body part is related to the torso, and the upper body part is related to the reflections from near the head. In the experiment with two people shown in Figure 13.d, the subject far from the sonar, near the wall is well distinguished from the wall (green), due to different group properties. The subject close to the sonar swinging his upper limbs (blue), is successfully estimated by the blue group. In Figure 13.e, two people going in opposite directions, are well estimated, even when the radial distance of the two people is approximately the same. This separation is possible due to the usage of continuity of location in the tracking and in grouping methods, and continuity of the subject's velocity in the grouping.

The chair and person objects are successfully mapped to different groups in Figure 13.f. The standard deviation of the standing subject is slightly higher than the chair due to slight movements of the standing subject. In Figure 13.g, swinging hands (green color), the chair (blue color) and wall, are well distinguished according to the group spatial standard deviation, $\sigma_{,}^{Gi}$. The human group in Figure 13.h, in the experiment of walking toward the sonar and crossing a chair, is continuously successfully estimated (turquoise), not including some body parts near the wall that were included in the group related to the wall. The chair group is split into two groups (yellow and purple). This is reasonable, as it was almost fully covered by the man, when he crossed past the chair. In the future, information from additional sonar nodes deployed in different locations could be used to merge these two objects that relate to the chair.

The tracking and grouping results indicate on significant differences in object properties between groups that are related to humans and non-humans. In addition to correlation and intensity properties, group kinematic features like velocity and standard deviation of object locations, can help to classify human groups from non-human groups, and can be a basis for advanced classification of human activity level and type.

C. Object Classification

Figure 14 shows the training set in the feature space for the first, second, and third classifiers after artifact removal. For the first classifier, (static or dynamic objects classifier), the clusters appear well separated. The velocity feature can distinguish between a walking and non-walking human, but can hardly distinguish between standing and swinging hands. The standard deviation of the group can better distinguish between hand swinging and standing, and even the between the static object (wall). The second classifier between human and not human is used to distinguish between a human standing without movement, and other static objects. The subject does move slightly even when standing, and its surface has different echoes, therefore the standard deviation can also be used here. However, it does not classify some of the data, and the correlation properties of mean and standard deviation over time, which indicate the surface pattern, are informative.

For the third activity classifier, the velocity feature is the most significant in case of activity type that involve walking, while the feature of standard deviation, is more significant in distinguishing between swinging hands and standing still. The activity types in this experiment are separable in the different feature spaces for the three decision tree classifiers. This justifies the use of a relatively simple classifier like the k-NN classifier for an efficient performance, as it can operate well with separable distributions and a relatively low complexity.

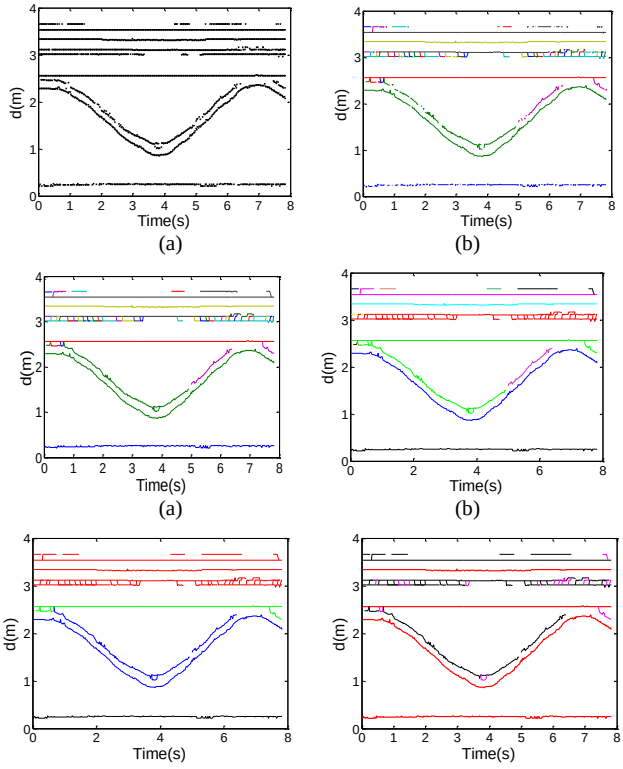


Fig. 12. Object detection, tracking, and grouping stages in space-time diagram. Figure 12.a shows the result of the match filter after the thresholding operation. Figure 12.b shows the results of the sequential MLE object tracking results. Figure 12.c shows the result of object tracking of post processing, in which missing estimations are mitigated for each object by interpolation. The results of grouping are shown in Figure 12.d, where each group has a different color. Figure 12.e shows the results after a group merging and splitting algorithm based on similarity of the different group features. Figure 12.f shows the sort results into dynamic or static objects according to their standard deviation from the main object.

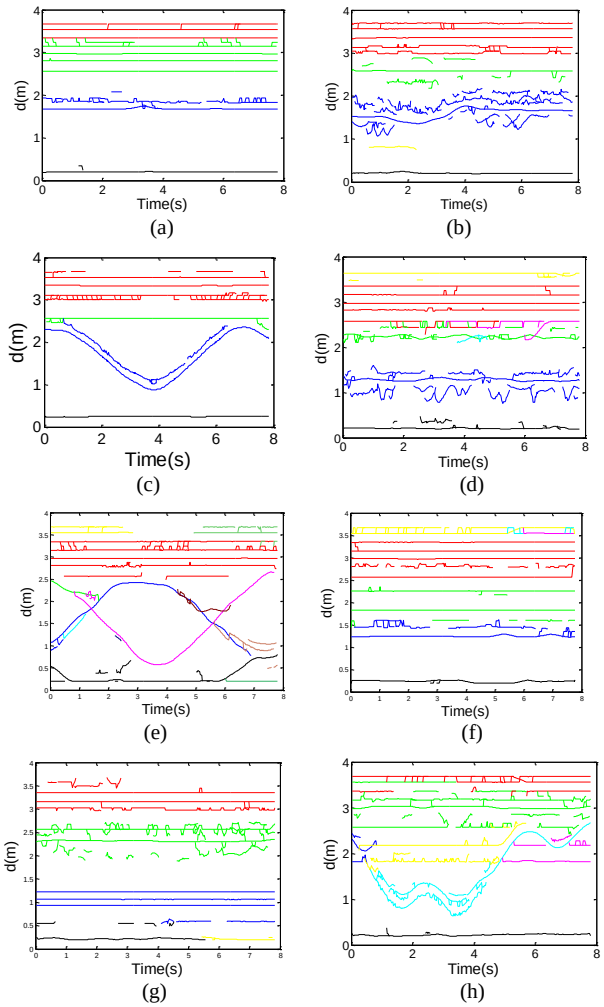


Figure 13 : Eight object grouping results for the experiments shown in Figure 9: a human subject standing, swinging upper limbs, walking towards the sonar system and away, two people are standing, one swinging hands and one subject standing still; two people walking to and from the sonar system; a subject stands in front of a chair; a subject stands behind a chair swinging his hands, and a subject walks around a chair, respectively.

The classification results are presented in Figure 15. The first classifier detected 2324, and 634, as static, and dynamic groups (cluster of objects), respectively. The second classifier identified the static groups as standing and walls. The third classifier distinguished between the two different activities, walking and swinging hands. The total number of groups in all the experiments were 2051, 273, 327, 307, for the classes of static groups (primarily the wall), standing, walking, and swinging arms and hands, respectively.

Figure 15.a, shows the results for each target (group). The classification reference is the tagging of the groups according to the video reference by an independent observer. The classification accuracy is around 97 percent for static objects, and 95 percent for dynamic objects, an average of 96.6 percent. The walking activity is classified in the 98th percentile. Swinging hands is occasionally misclassified as walking or standing, and only errors around 1% for static objects (wall). The wrong classification of the classes of walking or swinging arms, to standing or static object is partially explained by a towed-error from the first static-dynamic classifier. Standing and walls are more mixed, but can be separated, mainly using the correlation property, based on the information obtained using the high-bandwidth transmissions. Hence, the classification errors between static objects and standing human are a result of the second classifier.

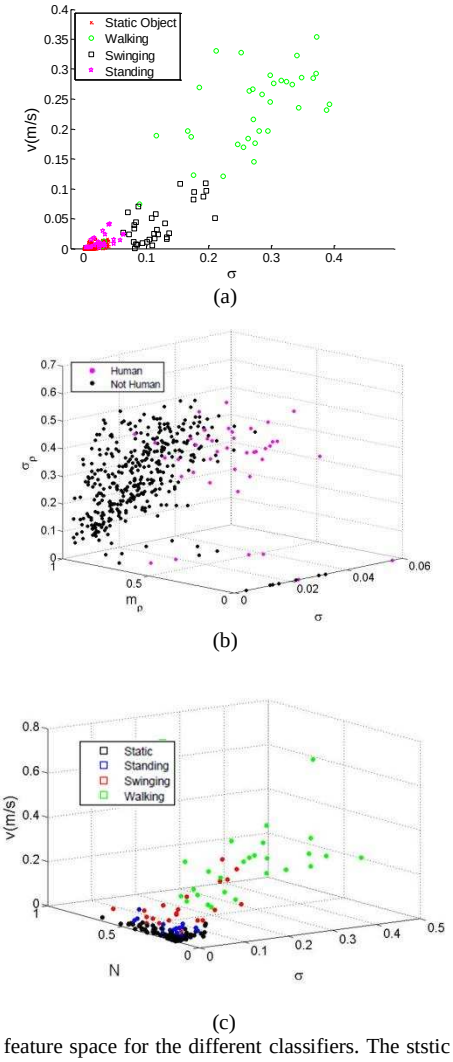


Fig. 14. The feature space for the different classifiers. The ststic or dynamic; human or non-human, and activity type classifiers are shown in figures 14.a, 14.b, and 14.c, respectively.

Figure 15.b describes the activity types classification performance. A success in classification is defined when there is at least one group that is classified to the correct activity. Similar to the classification for each object, walking is classified accurately (100 percent). The standing per experiment rate is lower than per group. In some experiments few groups were identified as standing, and in some none. The swinging of upper limbs is classified correctly in almost 90 percent, compared to 77 percent in classification per-object. The average higher percentage per experiment can be explained by the criterion of having at least one true object classification out of many. This non-rigid criterion is adequate for tracking patient a subject activity during daily life routines, where detection of even one activity over a short period of observation time is adequate. The activity type classification

results are summarized by Table II.

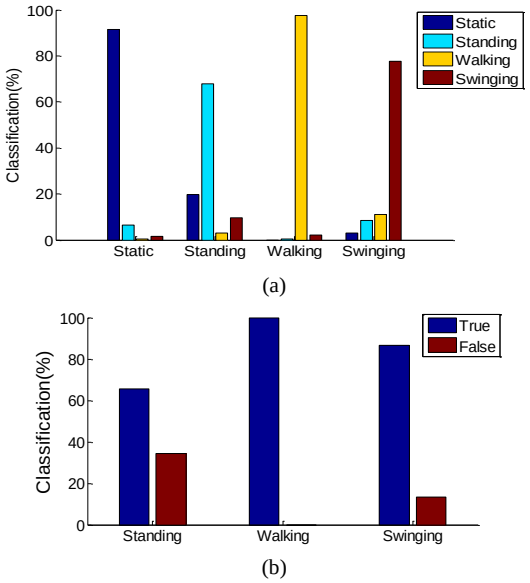


Fig. 15. The classification performance for classification. Figure 15.a describes the classification per group (object), while figure 15.b describes the classification per experiment.

TABLE II
CLASSIFIER PERFORMANCE

Classifier	Object Success rate (percent)	Experiment success rate (percent)
Standing	67.8	65.6
Walking	97.6	100
Swinging Arms	77.5	86.7

VII. CONCLUSIONS AND FUTURE RESEARCH

In this work we derive a new method to detect and classify human activity level and type, in a contactless manner, and affordably, using a simple wideband sonar system with one speaker and one microphone. We have developed analytical methods, and evaluated the technology in assessment of different activity types. The results show that with only one sonar sensor node, simple activity types, like standing still, standing with moving hands, or walking can be well classified.

The high bandwidth, gives additional information about subject location, and can be used to detect a human in a regular indoor cluttered environment. The high SNR, which is a reasonable assumption in indoor environment, enable to derive from the location estimations, the kinematic information of the different targets. The location information enables use in the spatial domain and enables context aware applications.

The suggested system enables accurate range and enhanced correlation properties, which cannot be achieved by the Continuous Wave (CW) Doppler based systems. Such a sonar based system has an affordable price, does not require attaching active markers or inertial sensors to the body, an does not use additional electro-magnetic radiation like radar systems. This system can work under any light conditions and in other risky circumstances such as in the presence of smoke during a fire, or high humidity conditions, as in a bath room, while maintaining the privacy of the patients.

Future experiments can use shorter observation periods to classify more subtle activity types, can use multiple sensor nodes to give 2-D and 3-D tracking for enhanced classification accuracy and can classify more complex activity like sitting down, falling, or carrying a bag. Future aggregation of low bandwidth Doppler pulses with the FM-chirp, inspired by bio-sonar, is expected to further improve the classification accuracy.

In the future this technology is expected to enable continuous assessment of various kinematic features of humans with reduced costs, under any light conditions in various environments. Unlike optical cameras this system can detect risk circumstances even in the presence of smoke during a fire, or high humidity conditions, as in a bathroom, where the risk of falling is very high, and still maintain personal privacy.

ACKNOWLEDGMENT

We would like to thank Valachi Pikovsky Foundation for the post-doctoral fellow scholarship which helped in financing the research, to Dr. Abraham Freedman and Paul Berkowitz for their helpful comments, to the Bio-medical and zoological departments in Tel Aviv University where the research took place, to Mrs. Raya Zeltser and to Mrs. Miri Zilka for participating in experiments, to Mr. Shimon Ohaion, from Coffe Rohale, for the support during the very long mornings and nights of writing the paper, and last to the anonymous reviewers that with their good comments contributed to improve the paper quality and readability.

REFERENCES

[1] G. Bocchetti, F. Flammini, C. Pragliola, and A. Pappalardo, "Dependable integrated surveillance systems for the physical security of metro railways," in *Distributed Smart Cameras, 2009. ICDSC 2009. Third ACM/IEEE International Conference on*, 2009, pp. 1-7.

[2] Z. Chaczko, A. Kale, and C. Chiu, "Intelligent health care; A Motion Analysis system for health practitioners," in *Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), 2010 Sixth International Conference on*, 2010, pp. 303-308.

[3] S. Beynon, J. L. McGinley, F. Dobson, and R. Baker, "Correlations of the Gait Profile Score and the Movement Analysis Profile relative to clinical judgments," *Gait & Posture*, vol. 32, pp. 129-132.

[4] M. Sekine, T. Tamura, M. Akay, T. Fujimoto, T. Togawa, and Y. Fukui, "Discrimination of walking patterns using wavelet-based fractal analysis," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 10, pp. 188-196, 2002.

[5] E. Campo, S. Bonhomme, M. Chan, and D. Esteve, "Remote tracking patients in retirement home using wireless multisensor system," in *e-Health Networking Applications and Services (Healthcom), 2010 12th IEEE International Conference on*, pp. 226-230.

[6] N. Barbour and G. Schmidt, "Inertial sensor technology trends," *Sensors Journal, IEEE*, vol. 1, pp. 332-339, 2001.

[7] A. Jobbagy and G. Hamar, "PAM: passive marker-based analyzer to test patients with neural diseases," in *Engineering in Medicine and Biology Society, 2004. IEMBS '04. 26th Annual International Conference of the IEEE*, 2004, pp. 4751-4754.

[8] C. B. Alan, *Handbook of Image and Video Processing* Academic Press, Inc., 2005.

[9] A. Kanaujia, N. Haering, G. Taylor, and C. Bregler, "3D Human pose and shape estimation from multi-view imagery," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on*, 2011, pp. 49-56.

[10] G. Blumrosen, B. Hod, T. Anker, D. Dolev, and B. Rubinsky, "Enhanced calibration technique for RSSI-based ranging in body area networks," *Ad Hoc Networks*, vol. 11, pp. 555-569, 2013.

[11] S. S. Ram, Y. Li, A. Lin, and H. Ling, "Doppler-based detection and tracking of humans in indoor environments," *Journal of the Franklin Institute*, vol. 345, pp. 679-699, 2008.

[12] B. G. Mobasseri and M. G. Amin, "A time-frequency classifier for human gait recognition," in *Proc. of SPIE Vol*, 2009, pp. 730628-1.

[13] C. Hornsteiner and J. Detlefsen, "Characterisation of human gait using a continuous-wave radar at 24 GHz," *Adv. Radio Sci.*, vol. 6, pp. 67-70, 2008.

[14] "http://www.microsoft.com/en-us/kinectforwindows/."

[15] X. Lu, C. Chia-Chih, and J. K. Aggarwal, "Human detection using depth information by Kinect," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on*, 2011, pp. 15-22.

[16] J. Sung, C. Ponce, B. Selman, and A. Saxena, "Unstructured human activity detection from RGBD images," in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, 2012, pp. 842-849.

[17] R. A. Clark, Y. H. Pua, A. L. Bryant, and M. A. Hunt, "Validity of the Microsoft Kinect for providing lateral trunk lean feedback during gait retraining," *Gait Posture*, vol. 2, pp. 00188-4, 2013.

[18] G. Blumrosen, M. Uziel, B. Rubinsky, and D. Porrat, "Noncontact tremor characterization using low-power wideband radar technology," *IEEE Trans Biomed Eng*, vol. 59, pp. 674-86, 2012.

[19] S. Chang, N. Mitsumoto, and J. W. Burdick, "An algorithm for UWB radar-based human detection," in *Radar Conference, 2009 IEEE*, 2009, pp. 1-6.

[20] Y. Yovel, M. Geva-Sagiv, and N. Ulanovsky, "Click-based echolocation in bats: not so primitive after all," *J Comp Physiol A Neuroethol Sens Neural Behav Physiol*, vol. 197, pp. 515-30, 2011.

[21] J. Simmons and J. Gaudette, "Biosonar echo processing by frequency-modulated bats," *IET Radar, Sonar & Navigation*, vol. 6, pp. 556-565, 2012.

[22] Z. Zhang, P. O. Pouliquen, A. Waxman, and A. G. Andreou, *Acoustic micro-Doppler radar for human gait imaging*: J Acoust Soc Am. 2007 Mar;121(3):EL110-3.

[23] M. Bradley and J. M. Sabatier, "Acoustically-observable properties of adult gait," *The Journal of the Acoustical Society of America*, vol. 131, pp. EL210-EL215, 2012.

[24] A. Balleri, K. Chetty, and K. Woodbridge, "Classification of personnel targets by acoustic micro-doppler signatures," *Radar, Sonar & Navigation, IET*, vol. 5, pp. 943-951, 2011.

[25] B. Lyonnet, C. Ioana, and M. G. Amin, "Human gait classification using microDoppler time-frequency signal representations," in *Radar Conference, 2010 IEEE*, 2010, pp. 915-919.

[26] S. Yang and L. Kong, "Research on Characteristic Extraction of Human Gait," in *Bioinformatics and Biomedical Engineering, 2009. ICBBE 2009. 3rd International Conference on*, 2009, pp. 1-4.

[27] Y. Yovel and W. W. L. Au, "How Can Dolphins Recognize Fish According to Their Echoes? A Statistical Analysis of Fish Echoes," *PLoS ONE*, vol. 5, p. e14054, 2010.

[28] A. F. Molisch, J. R. Foerster, and M. Pendergrass, "Channel models for ultrawideband personal area networks," *Wireless Communications, IEEE*, vol. 10, pp. 14-21, 2003.

[29] G. Blumrosen, M. Uziel, B. Rubinsky, and D. Porrat, "Tremor acquisition system based on UWB Wireless Sensor Network," *BSN2010 conference on Body Sensor Network Proceedings*, June 7-9 2010.

[30] Y. Yovel, M. O. Franz, P. Stilz, and H. U. Schnitzler, "Complex echo classification by echo-locating bats: a review," *J Comp Physiol A Neuroethol Sens Neural Behav Physiol*, vol. 197, pp. 475-90, 2011.

[31] M. I. Skolnik, "Radar handbook, 1990," ed: McGraw-Hill.

[32] G. Blumrosen and A. Luttwak, "Human Body Parts Tracking and Kinematic Features Assessment Based on RSSI and Inertial Sensor Measurements," *Sensors*, vol. 13, pp. 11289-11313, 2013.

[33] J. A. Gallego, E. Rocon, J. O. Roa, J. C. Moreno, A. D. Koutsou, and J. L. Pons, "On the use of inertial measurement units for real-time quantification of pathological tremor amplitude and frequency," *Procedia Chemistry*, vol. 1, pp. 1219-1222, 2009.

[34] W. Blanding, P. Willett, and S. Coraluppi, "Sequential ML for Multistatic Sonar Tracking," in *OCEANS 2007 - Europe*, 2007, pp. 1-6.

[35] D. Von Helversen and O. von Helversen, "Object recognition by echolocation: a nectar-feeding bat exploiting the flowers of a rain forest vine," *Journal of Comparative Physiology A*, vol. 189, pp. 327-336, 2003.

[36] G. blumrosen, B. Fishman, and Y. Yovel, "Enhanced Indoor Sonar Based Human Segmentation and Tracking Technique Based on FM-chirp," *Technical Report*, 2013.

[37] M. I. Skolnik, "Introduction to radar systems / Merrill I. Skolnik," ed: New York : McGraw-Hill, 1962.

[38] W. Blanding, P. Willett, and Y. Bar-Shalom, "ML-PDA: Advances and a new multitarget approach," *EURASIP J. Adv. Signal Process*, vol. 2008, p. 38, 2008.

[39] G. Blumrosen, B. Hod, T. Anker, B. Rubinsky, and D. Dolev, "Enhancing RSSI-based Tracking Accuracy in Wireless Sensor Networks," *ACM Transactions on Sensor Networks*, to appear in Nov 2013.

[40] G. D. Forney, Jr., "The viterbi algorithm," *Proceedings of the IEEE*, vol. 61, pp. 268-278, 1973.

[41] W. R. Blanding, P. K. Willett, Y. Bar-Shalom, and R. Lynch, "Directed subspace search ML-PDA with application to active sonar tracking," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 44, pp. 201-216, 2008.

[42] G. Blumrosen and A. Freedman, "Sensitivity study and a practical algorithm for ml ostbc and beam forming combination," in *The 23rd IEEE Convention of Electrical and Electronics Engineers in Israel*, 2004.

[43] S. Reed, Y. Petillot, and J. Bell, "An automatic approach to the detection and extraction of mine features in sidescan sonar," *Oceanic Engineering, IEEE Journal of*, vol. 28, pp. 90-105, 2003.

[44] J. M. Aughenbaugh and B. R. La Cour, "Use of prior information in active sonar tracking," in *Information Fusion, 2009. FUSION '09. 12th International Conference on*, 2009, pp. 1584-1591.

[45] Z. Zhang, P. Pouliquen, A. Waxman, and A. G. Andreou, "Acoustic Micro-Doppler Gait Signatures of Humans and Animals," in *Information Sciences and Systems, 2007. CISS '07. 41st Annual Conference on*, 2007, pp. 627-630.

[46] A. Kundu and G. C. Chen, "An integrated hybrid neural network and hidden Markov model classifier for sonar signals," *Signal Processing, IEEE Transactions on*, vol. 45, pp. 2566-2570, 1997.

[47] Y. Bao, X. Du, and N. Ishii, "Combining Feature Selection with Feature Weighting for k-NN Classifier Intelligent Data Engineering and Automated Learning." vol. 2412, ed: Springer, 2002, pp. 137-142.

[48] L. Jiang, Z. Cai, D. Wang, and S. Jiang, "Survey of Improving K-Nearest-Neighbor for Classification," in *Fuzzy Systems and Knowledge Discovery, 2007. FSKD 2007. Fourth International Conference on*, 2007, pp. 679-683.

[49] R. Simon, M. W. Holderied, C. U. Koch, and O. von Helversen, "Floral Acoustics: Conspicuous Echoes of a Dish-Shaped Leaf Attract Bat Pollinators," *Science*, vol. 333, pp. 631-633, 2011.

[50] http://www.avisoft.com/.

[51] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *The Journal of Machine Learning Research*, vol. 7, pp. 1-30, 2006.

Gaddi Blumrosen was born in Jerusalem, Israel. He received the B.Sc. degree in electrical engineering from the Technion, Israeli Institute of Technology, Israel, the M.S. degree in electrical engineering from Tel Aviv University, Israel, and the PhD degree in Bio-Medical engineering from the Hebrew University of Jerusalem, Israel, in 2011. He is currently Valachi Pikovsky Foundation post-doctoral fellow in Tel Aviv University. His current research interests include wireless communication, radar and sonar systems, tracking systems, gait analysis, and biomedical signal modeling and processing.

Ben Fishman was born in Rishon Le Zion, Israel. Received his B.Sc degree in Bio medical engineering from Tel-Aviv university, Israel in 2012.

Yossi Yovel Was born in Beer Sheva, Israel. He completed his B.Sc. degree in physics and biology in the Tel-Aviv University, his MSc in neuroscience in Tel-Aviv University and his PhD in Tuebingen University, Germany. He is an assistant professor in the department of zoology and the Sagol School of neuroscience in Tel-Aviv University since 2011, and is an expert on bat cognition, bat SONAR and bat-SONAR bio-mimicry.