A Survey of Human-Sensing: Methods for Detecting Presence, Count, Location, Track, and Identity

THIAGO TEIXEIRA
Yale University,
GERSHON DUBLON
Massachusetts Institute of Technology[1]
and
ANDREAS SAVVIDES
Yale University

An increasingly common requirement of computer systems is to extract information regarding the people present in an environment. In this article, we provide a comprehensive, multi-disciplinary survey of the existing literature, focusing mainly on the extraction of five commonly needed spatio-temporal properties: namely presence, count, location, track and identity. We discuss a new taxonomy of observable human properties and physical traits, along with the sensing modalities that can be used to extract them. We compare active vs. passive sensors, and single-modality vs. sensor fusion approaches, in instrumented vs. uninstrumented settings, surveying sensors as diverse as cameras, motion sensors, pressure pads, radars, electric field sensors, and wearable inertial sensors, among others. The goal of this work is to summarize the existing solutions from various disciplines, to guide the creation of new systems and point toward future research directions.

Categories and Subject Descriptors: I.2.9 [**Artificial Intelligence**]: Robotics—*Sensors*; I.4.8 [**Image Processing and Computer Vision**]: Scene Analysis—*Sensor Fusion*; *Tracking*

General Terms: Algorithms, Design

Additional Key Words and Phrases: Human counting, human detection, identification, localization, people counting, person detection, sensor fusion, tracking

1. INTRODUCTION

As the sensor network and ubiquitous computing communities increasingly focus on creating environments that are seamlessly aware of and responsive to the humans that inhabit them, the need to sense people will become ever more pressing. *Human-sensing* encompasses issues from the lowest level instantaneous sensing challenges

[1]Work done while a member of ENALAB at Yale University.

Author's address: Yale University, 10 Hillhouse Ave, New Haven, CT 06511.

all the way to large-scale data mining. Several questions circumscribe the problem. For example, we might ask of our sensors: Is there a person in this room? How many people are in this room? What is each person doing? What does each person need? Can we predict what they are going to do next?

The simplest applications of human sensing make direct use of such information to, for instance, open a door as people pass, turn lights on/off when a room is occupied/empty, or lock a computer when the user moves away. However, looking further ahead into the future, a medical application may ask "Which person in this room is John, and what is his body temperature and heart rate?". And, further, if John is found to be sick and contagious, it may wish to know "Who has he been in contact with in the past $24h$?" In addition, computing applications of the future will likely infer people's moods from the analysis of their speech, posture, and behavior, to make better decisions not only about the people themselves but concerning a variety of seemingly-unrelated subjects as well (i.e. affective computing [Picard 2000]). Going even further, such information can be gathered about *groups* of people, and *groups of groups* people, and so on, to make increasingly higher-level decisions. And so, the sheer breadth of these requirements make it clear that human-sensing is an inherently multi-faceted problem. Major contributions have traditionally arisen from the Radar and Computer Vision communities, while more recently Robotics and Sensor Networks researchers have proposed a variety of creative solutions based on multiple-sensor and multiple-modality systems. To expose the progress that has been made in each direction and to identify new opportunities, this paper provides a comprehensive overview of the solutions that exist today, using a unified vocabulary that clearly expresses the advantages and disadvantages of each, to serve as a guide for the design of future systems.

Given the immenseness of the field, the scope of this survey is restricted to sensor systems that detect a well-defined set of five low-level *spatio-temporal properties*, namely: presence, count, location, track, and identity. These properties can be observed by measuring specific *human traits* through a number of *sensing modalities*. We review solutions where people are uninstrumented and possibly adversarial, as well as those where people carry sensors, such as GPS. In our discussion, we find that although different sensors obviously have distinct advantages and failure modes, some emerge as clear winners in specific scenarios. Other approaches, we argue, may be employed in resource-constrained environments, or leveraged in sensor fusion. In our discussion we also point toward an open problem that we believe will steer human-sensing research in the near future, namely that of seamlessly coordinating massive amounts of sensors.

The rest of this paper is organized as follows. In Section 2, we discuss the major obstacles and noise sources that make human-sensing such a challenging task. We, then, introduce a taxonomy of human-sensing in Section 3, where we also discuss physical human traits and the sensing modalities to detect them. Afterwards, a detailed review of existing approaches is provided in Section 4, which is subdivided into uninstrumented and instrumented approaches, single-modality versus sensor fusion. A summary of our findings and a discussion of the open research directions are given in Section 5, and Section 6 concludes the paper.

2. CHALLENGES

More so than most other object-detection and sensing tasks, human-sensing is a challenging endeavor for a variety of reasons. Common obstacles, irrespective of sensing modality, can be grouped into six broad classes:

— **Sensing noise:** Sensors that rely on a very small number of particles (i.e. photons in an image sensor, or electrons in an ultra-low current circuit) are prone to shot noise due to statistical fluctuations in the particle arrival rates. Other types of sensing noise include thermal noise, $1/f$ noise (i.e. pink noise), and avalanche noise, as well as aliasing and quantization noise. These types of noise have been abundantly studied, and may be alleviated through well-known hardware- and sensor-design considerations. Thus we will not consider them any further in this paper.

— **Environmental variations:** Unexpected or sudden changes in environmental conditions are some of the most common sources of errors that occur in real-world scenarios. Radar signals, for instance, can be dampened by rain or fog, and PIR sensors are often triggered by heat currents flowing from HVAC (Heating, Ventilating, and Air Conditioning) systems. Furthermore, a large portion of the computer vision literature is aimed at dealing with moving foliage, lighting variations, shadows, and so on.

— **Similarity to background signal:** Clearly, separating a person from the background signal is a core requirement for human-sensing. However, this is often not possible outside a laboratory setting, as background signals in the real world can grow arbitrarily complex. The most obvious instances of such sensing failures come from the computer vision domain, where background-modeling is still a wide-open problem. In other domains, such as with ranging sensors (radars, ladars, sonars), the presence of unwanted signals with the correct frequency spectrum or timing characteristics (due to multipath, for instance) can often fool the system into producing phantom detections.

— **Appearance variability and unpredictability:** People sport non-rigid bodies which can be arranged in any number of poses, along at least 244 degrees of freedom [Zatsiorsky 1997]. To make matters worse, this appearance-space greatly increases as we consider different types of clothing, hats, backpacks, purses, and other carried objects. Finally, people can also behave unpredictably, moving in paths that may change on a whim, and thus present an enormous challenge to localization and tracking systems.

— **Similarity to other people:** In some applications, such as tracking or person identification, the main challenge to be overcome is the high degree of similarity amongst people. Moreover, physical limitations of the sensors themselves often lead to a further loss of personally-identifying information in the acquired signal — and likewise with environmental factors such as poor lighting, or interference sources. This is further aggravated in some situations such as corporate and military scenarios, where people wear similar-appearance uniforms.

— **Active deception:** In adversarial scenarios, it is important to consider possible attack vectors, through which a human-sensing system may be either fooled or debilitated. Jamming signals, for instance, are often used in military scenarios to

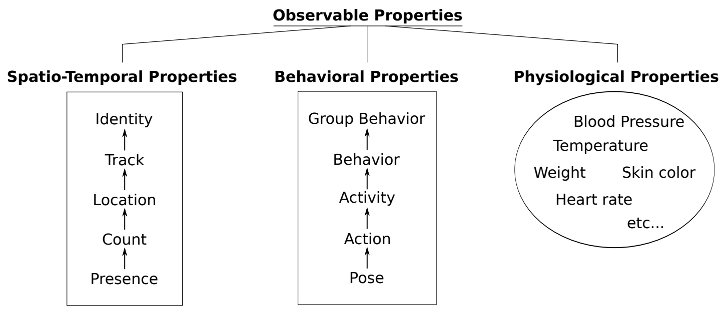

Fig. 1. Taxonomy of the human properties that are involved in the human-sensing problem. Arrows indicate an inner hierarchy of properties. For instance, knowledge about "count" implies knowledge of "presence", and "action" often implies knowledge of "pose".

disable the enemy's radars and communication systems. Other deceptive techniques may be as simple as turning off the lights in an area covered by cameras, or walking slowly to fool motion sensors.

3. HUMAN-SENSING TAXONOMY

We classify under the large umbrella of "human-sensing" the process of extracting *any information* regarding the people in some environment. Such information, as summarized in Figure 1, can be classified into three observable categories: spatio-temporal properties, behavioral properties, and physiological properties. In this survey we focus on the inference of spatio-temporal properties (STPs) only. These consist of low-level components regarding the position and history of people in an environment. More specifically:

(1) **Presence** — *Is there at least one person present?*
Presence is arguably the property that is most commonly sought-after in existing real-world applications, the most popular presence-sensor being motion sensors (PIR) and proximity sensors (scalar infrared range-finders). In cooperative scenarios, though, where people can be instrumented with portable or wearable devices, solutions such as RFID (radio-frequency identification) are becoming increasingly common.

(2) **Count** — *How many people are present?*
The number of people in an environment can be inferred by either employing a person-counting sensor (or sensors) that covers the entire area of interest, or by counting people at all the entry and exit points. Commercial people-counting solutions range from thermal imagers [SenSource] and break-beams, to simple mechanical barriers such as turnstiles.

(3) **Location** — *Where is each person?*
Location-detection, or "localization", consists of obtaining the spatial coordinates of a person's center of mass. Localization can be achieved using instrumented (such as GPS) or fully uninstrumented solutions (such as cameras). In

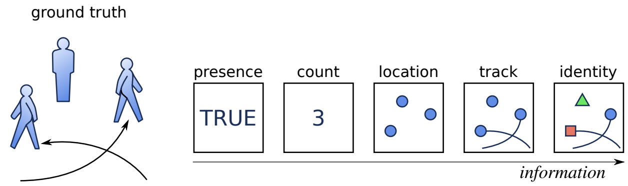

Fig. 2. The five spatio-temporal properties considered in this survey.

addition, since a grid of presence sensors can also be used to localize people, localization can be considered a higher-resolution generalization of presence-detection.

(4) **Track** — *Where was this person before?*
Tracking is the process of solving the correspondence problem, that is, extracting the spatio-temporal history of each person in a scene. Equivalently, tracking may be described as recovering a person's relative identity[2]. For example, if upon detection a person is labeled with a temporary ID (e.g. "person 6") then tracking is the problem of specifying at each subsequent sampling of the scene which detection is the same "person 6". This temporary ID is typically lost in the presence of sensing gaps, such as when the person leaves the scene and returns on the next day. At that point, yesterday's "person 6" will be given a new ID when re-detected. Situations that lead to the loss of a person's relative ID are often called *ambiguities*. In the remainder of this text, we will use the term *piecewise tracking* to qualify a tracker that is not capable of adequately handling ambiguities.

(5) **Identity** — *Who is each person? Is this person John?*
At a first glance it may seem odd to group "identity" into the category of *spatio-temporal properties*. However, identification is nothing more than a natural extension of tracking where each person is always assigned the same globally unique ID rather than solely relative IDs. Therefore, identity-detection extends tracking so that it becomes possible to recover a person's spatio-temporal history even across sensing gaps.

The five spatio-temporal properties are depicted in Figure 2. As described in the figure in the form of the "information" arrow, these properties present the following cumulative quality: the true value of property n for an environment contains all information about the true value for property $n-1$. For instance, "counting" contains all information from "presence", since "presence" is simply the condition $count > 0$. Similarly, if "location" is known for all people in an environment then the number of people ("count") must also be known, and so on.

Of course, numerous applications also require knowledge of human properties other than the STPs. Some are physiological properties (such as weight, temperature, heart rate, blood pressure, or skin/hair/eye color) while others are behavioral

[2]Given the above definition, the frequently-used term "single-target tracking" does not make logical sense, as there cannot be any ID ambiguities when it is known there is only one target present. What generally is meant by "single-target tracking" we here call by the name of *localization*.

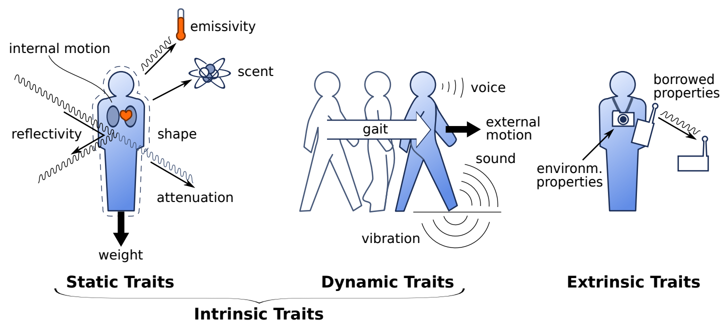

Fig. 3. Physical traits that may be used to measure the five spatio-temporal properties.

properties (pose, action, activity, and so on)[3]. Clearly, the subject of human-sensing is a broad and over-reaching one. Hence, for practical reasons we must limit the scope of this survey to the research problems that we consider are the most pervasive problems. In our experience, these tend to be exactly the detection of presence, count, location, track, and ID, which are at the core of a majority of human-sensing applications. In the discussion that follows, we analyze the physical traits from which these five spatio-temporal properties can be inferred, and the sensing modalities that can be used to measure them.

3.1 Human Traits Classification

At the lowest level, human-sensing is equivalent to measuring, either directly or indirectly, one or more of the myriad ways humans impact their environments. These are *human traits* that are either related to human presence or to human actions (Figure 3). People may also carry objects, such as mobile phones and RFID, which lend their signals and sensing capabilities to the person who is carrying it. We use the term *sensing modality* to denote in general terms the means by which any of these traits can be measured. In this definition, multiple different *sensors* can belong to the same sensing modality. Some notable examples of this are given on Table I.

Below we discuss the human traits that can be exploited in human-sensing systems, briefly describing the existing sensing modalities that can be used to measure them. A summary of our taxonomy of human traits can be seen in Figure 4.

— **Static, Intrinsic Traits:** Static traits stem from the physiological properties from Section 3, and are produced whenever a person is present, irrespective of what he or she is doing. Common static traits are **weight** and **shape**. While weight is typically measured *directly* through simple piezoresistive or piezoelectric sensors, shape is measured *indirectly*: shape detectors operate by intersecting a person's shape with geometric lines which are either actively produced by the sensor itself (in the case of radars, for example) or passively appropriated from the environment

[3]Note that like spatio-temporal properties, the behavioral properties can also be organized in a hierarchy, as shown through the use of arrows in Figure 1.

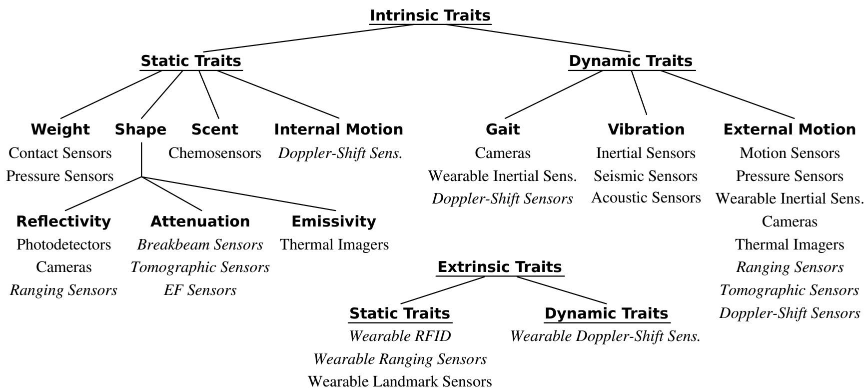

Fig. 4. Taxonomy of measurable human traits, listing the sensing modalities that can detect them. Italics are used to denote *active signaling* sensors, and the word *wearable* indicates instrumented approaches.

(e.g. cameras). Therefore, shape is a trait that must be extracted from one of three other traits: reflectivity (with cameras or radars, for example), attenuation (tomographic sensors), or emissivity (thermal imagers).

Another static trait is the involuntary **motion of internal organs**, such as the heart and lungs. This can be measured through skin-penetrating radio and ultrasound signals. Finally, a relatively new avenue for human-sensing lies in **scent** detection [Pearce et al. 2006]. However, although chemosensors have been developed for a wide variety of compounds (used, for instance, in bomb-sniffing [Yinon 2003] or detection of spoiled food), it is still not well-known which molecules and chemical compounds present in the human scents are best suited for person-detection. Recent studies with gas chromatography-mass spectrometry have shown it is possible to personally-identify people from their sweat, as well as detecting their gender [Penn et al. 2007]. Furthermore, CO_2 levels have also been used to detect the presence of people, albeit with slow response times [De Cubber and Marton 2009]. Other than these intial explorations, scent-based systems are highly uncommon and thus not further investigated in this survey.

— **Dynamic, Intrinsic Traits:** Dynamic traits are those that arise from human activity. They are only present when people move, and are not detectable for reasonably stationary persons. We divide these into three categories: **external motion**, **gait**, and **vibrations**. External motion is defined as any change in a person's pose or in the position of their BCOM (body center of mass). This, of course, includes all external motion due to walking. However, we single out a person's *gait*[4] as a special case of external motion, as it has been shown to possess personally-identifying information that other examples of external motion do not. As for *vibrations*, these are the pressure waves that people produce either involuntarily (in the form of sounds and vibrations from footsteps, for example)

[4]I.e. the characteristic motion pattern displayed by people's limbs, torso, and head during a walking or running activity.

sensing modality	example sensors
Binary sensors	Contact sensors, Breakbeams, PIRs, Ultrasound motion sensors
Pressure sensors	Piezo-resistors, Piezo-electric materials
Chemosensors	CO_2 sensors, Humidity sensors
Doppler-shift sensors	Radios, Ultrasound transducers
Photodetectors	Phototransistors, Photodiodes
Cameras	CMOS image sensors, CCD image sensors, Specialized motion- or edge-detecting imagers
Thermal imagers	Microbolometer arrays, PVDF (Polyvinylidene Fluoride) arrays
Ranging sensors	Ultrasonic range-finders
Scanning range sensors	Radars, Ladars, Sonars
Tomographic sensors	Radio pairs
Electric field sensors	Capacitors
Inertial sensors	Accelerometers, Gyroscopes, Magnetometers
Vibration sensors	Seismic sensors, Accelerometers, Piezoelectric sensors, Electrostatic microphones, Laser microphones
Motion sensors	PIRs, Motion cameras, Binary Doppler-shift sensors
ID sensors	RFID, plus any communication system
Envir. recog. sensors	WiFi fingerprinting, Wearable microphones, Wearable cameras

Table I. Examples of different sensors belonging to each sensing modality from our taxonomy.

or voluntarily (in the form of speech), which can be measured with accelerometers and microphones, respectively.

— **Extrinsic Traits:** Extrinsic traits are those that stem from objects or devices carried by a person. Approaches based on extrinsic traits are most commonly found in the Robotics and Sensor Networks literatures. We subdivide these into two groups. The first group, **borrowed traits**, represent the characteristics that in reality belong to devices placed on the person or people of interest. The second, **environmental traits**, are physical characteristics of the environment, which are sensed by wearable devices on the person's body to provide location measurements. Hence, as shown in Figure 3, the main distinction between environmental and borrowed traits lies in the direction of the information flow (the arrow, in the figure). Most borrowed and environmental traits are *static*, that is, they do not require the person to be moving. The main exceptions are Doppler-shift based device-to-device approaches [Kusy et al. 2007][Chang et al. 2008].

3.2 Sensor classification

As a result of the enormous arsenal of sensing modalities available, each of which can be used to leverage a great many different human traits, the sheer number of approaches to human-sensing that have been proposed in the literature is immense. To describe all of them is an impossible task. In the following sections, we limit ourselves to a selection of approaches which are either the most useful, the most ubiquitous, or the most ingenious. We discuss these in the context of similar solutions to illustrate the advantages and disadvantages of each. For this, let us define the following terminology:

— **Setting:** Approaches are classified as **instrumented** if they require each person to carry a device on them. In contrast, **uninstrumented** approaches are those that do not rely on any carried device, and can thus be used in adversarial scenarios where people may be actively trying to fool the system.

— **Signaling:** The term **passive** signaling is used to refer to approaches which measure signals that are readily available from the environment. Meanwhile, **active** signaling denotes those approaches which transmit their own carefully-formatted signals and measures the properties of the responses.

— **Network density:** Sensors have different capabilities when used by themselves versus when employed in a dense network. We quantify the network density (ND) using the order of magnitude (in base 2) of the number of sensors required to provide some specific service in an area A. For example, if a single camera can localize a person in A, then the density of this solution is $log_2(1) = 0$. If, instead of cameras, the same area A is instrumented with a network of 36 pressure-sensitive floor tiles to a similar effect, then the density increases to $log_2(36) = 5.17$. Since ND is logarithmic, the difference between density values should remain constant as the sensing area A increases (so long as the number of sensors scales linearly with A). For instance, if the sensing area triples to $3A$, requiring 3 cameras or $3 \times 36 = 108$ floor tiles, the density difference will remain the same: $log_2(108) - log_2(3) = 5.17$. Of course, exact values for ND are application- and implementation-dependent, and the numbers given in this paper should serve merely as a rough guide for comparison.

4. SURVEY OF EXISTING APPROACHES

Before we can finally delve into our survey of human sensing approaches per se, we must take a moment to make a few clarifications. First, that our goal in this section is to expose and organize the existing literature, rather than to detail the exact algorithms that they employ. Common trends in the algorithmics of the described solutions are faint across different sensing modalities. Presence, Count, Localization, and ID detectors may rely on traditional pattern matching techniques such as principal component analysis (PCA), but are usually heavily tweaked to the specific modality in an ad-hoc fashion. Where there is more uniformity among modalities is in Tracking, as the correspondence problem is itself greatly abstracted from specific sensors. As such, although traditional tools such as multiple-hypothesis tracking (MHT) [Reid 1979] and joint-probabilistic data association (JPDA) [Bar-Shalom and Tse 1975] are still largely effective, the emerging tool being utilized for tracking in the past decade has been without doubt the particle filter (PF) [Arulampalam et al. 2002][Isard and Blake 1998]. Furthermore, for sensor-fusion scenarios the particle filter is increasingly relevant, as we describe in Section 5.1. For more details about algorithms, throughout the text we point the reader to consider specialized surveys.

The second clarification we must make is that since the authors of the solutions reviewed in this section often do not agree on common performance metrics or even experimental scenarios, we are forced to compare different approaches in rather qualitative terms. Thus, we use words such as “accuracy” and “precision” loosely to denote a measure of the average error (e.g. the mean error of a person-localization approach) and a measure of classification correctness (e.g. in a person-detection approach, the ratio of true positives divided by all classifications), respectively. The exact meaning of these will vary from modality to modality, and is explained inline with the text if necessary. Other metrics, such as latency and algorithmic complexity are, more often than not, entirely missing from the surveyed papers, and thus cannot be consistently reported here.

The final clarification is with regards to application scenarios. In this work we make use of two diametrically opposite scenarios to guide the discussion: *resource-constrained* vs. *performance-driven*. In the former, issues such as localization accuracy take a secondary role to constraints such as cost, privacy, or energy expenditure. Examples include person-count estimation in public spaces and customer-tracking in supermarkets. On the other hand, in performance-driven scenarios the most pressing demand is for high-accuracy, high-precision data, typically for use in control systems, medical diagnosis, entertainment, security, or surveillance. An example is a hypothetical medical application that is meant to assess the patient's response to a new treatment for which close monitoring is vital.

4.1 Uninstrumented, Single-Modality Approaches

We start this discussion with a collection of uninstrumented, single-modality sensing approaches. These are characterized by sensors placed on the environment, and are the most commonly-found solutions in existing real-world deployments. However, existing deployments are typically characterized by simple usage of raw data without any high-level processing, such as with motion-sensitive lighting or CCTV (closed-circuit television) networks. In contrast, below we survey the use of such sensors for "smart" applications.

4.1.1 *Binary sensors.*

A variety of sensing modalities can be grouped into the broad category of "binary sensors". In the context of human-sensing, binary sensors are those that return a logic 1 if human presence is detected within a certain sensing area, otherwise returning a logic 0. The modality of binary sensors includes sensors such as break-beams, contact sensors, PIRs, and binary Doppler-shift sensors, all of which are currently used in resource-constrained scenarios.

In recent years there has been a growing tendency to research algorithms that operate on a purely abstract model of a binary sensor rather than on specific sensors such as PIRs [Aslam et al. 2003][Oh and Sastry 2005][Kim et al. 2005][Xiangqian et al. 2008]. The main disadvantage of using an abstract "binary sensor" model is that it can overlook some inherent differences between sensing modalities. For instance, binary sensors that rely on human motion (e.g. PIRs) tend to produce bursty positive detections and a large number of false negative detections.

In single-node configuration, binary sensors can only be used to detect presence, and nothing more. In contrast, when in a high-density network these sensors become capable of counting, localizing and partially tracking. Localization accuracy (as well as the maximum number of people that can be counted) depends both on the number of sensors and on the dimensions of the sensing areas of individual sensors. This is quantified in [Shrivastava et al. 2006]. Binary sensing approaches can only provide piecewise tracking, since they suffer from tracking ambiguities that are unsolvable from the binary information alone. That is, if "person 1" and "person 2" cross paths their relative identities will be lost and cannot be recovered due to the lack of personally-identifiable data in the binary signal. A popular way to bypass this problem is to make (often unrealistic) smoothness assumptions about people's paths. Below, we separately consider three binary sensor approaches: PIR, pressure-sensitive floor tiles, and electric field (EF) sensors.

— **Passive Infrared:** Most of the PIR-based methods follow a strictly geometric formulation, where the path of each person is calculated deterministically from intersecting sensing areas as in [Shrivastava et al. 2006]. More and more, however, PIR tracking approaches have been using data-inference tools such as Kalman or particle filtering [Schiff and Goldberg 2006]. Shankar et al. construct spherical sensor nodes composed of multiple PIR sensors pointed radially away from the sphere's surface [Shankar et al. 2006]. This allows the bearing of a person to be estimated from the direction of the PIR sensor that detected them. Using several of these multi-PIR sensor nodes placed on walls, the authors show it is possible to detect and localize a moving person. Of course, occlusions become a dominating issue as the number of people in the environment increases.

The main disadvantages of PIR sensors are: (1) they cannot detect people who are stationary, thus leading to a large number of false negatives; (2) their output is highly bursty[5]. These issues are largely ignored by the vast majority of PIR-based research by limiting their system to single-person scenarios and/or assuming people are always moving.

— **Pressure-Sensitive Tiles:** Most, although not all, solutions based on the installation of special-purpose floor tiles rely on pressure measurements. Research dating back to 1994 [Pinkston 1994] used force sensing resistors to measure the location and foot pressure of a single person. Even as early as 1988, similar technology was already commercially available in the form of Nintendo's Power Pad. More recently, Murakita et al. used a Markov chain monte carlo (MCMC) particle filter to track people based on a sequence of footsteps [Murakita et al. 2004]. The main challenge that they tackle is that people have two contact points with the floor. This leads to an additional type of correspondence problem, where the objective is to select the two contact points that belong to the same person. The authors report a mean localization accuracy of $0.21m$ in the direction of motion, and $0.01m$ in the perpendicular direction. Their system can robustly disambiguate between people who are separated by at least $1.1m$, performing poorly, however, if the separation was $0.5m$ or less.

More surprisingly, it has been demonstrated that floor tiles can also be used to identify people from force profile of their footsteps [Orr and Abowd 2000][Middleton et al. 2005]. For this, Orr et al. considered the time series of the pressure exerted by a person's entire foot. They were able to achieve 93% precision using a 15-person test sample that included multiple different footwear configurations. They also report that footwear does not greatly affect the precision of their identification approach. Middleton et al., on the other hand, have used arrays of binary contact sensors to measure the time spent at different areas of a person's feet. They measure the stride length, stride cadence and heel-to-toe ratio to identify 12 out of 15 test subjects (80% precision). It is possible that a much higher identification precision may be achievable using a high-resolution floor tile system such as the one presented

[5]Some commercial off-the-shelf sensors use a heuristic solution to make up for this, by ignoring detections that fall within a "refractory period" of an earlier event.

by Morishita et al. [Morishita et al. 2002], although we express some doubt as to whether this ID inference could resist larger databases, say with more than 20 subjects.

— **Electric Field Sensors:** Capacitors can be used to detect people's presence and to measure their distance with good accuracy. The basic operating principle of EF sensors is that an AC signal applied to a capacitor plate will induce a similar signal in a receiving plate. The effect of human presence between the transmitter and receiver can, then, be measured as changes in the received current. The specifics vary depending on three possible configurations (transmit mode, shunt mode, and loading mode), which are somewhat analogous to the emissivity, attenuation, and reflectivity traits discussed in Section 3.1. See [Smith et al. 1998] for an in-depth discussion. Electric field sensors are often used as binary proximity sensors that are placed either on a wall and as floor tiles [Henry et al. 2008]. In both cases, commercial off-the-shelf EF sensors are already available in the market today [Future-Shape]. Valtonen et al. take a hybrid approach by combining floor tiles and antennae on the walls to track a moving person with an accuracy of $41cm$ [Valtonen et al. 2009]. The main advantage of electric field sensors lies in their simplicity, as they consist simply of an oscillator and either one or two capacitor plates. However, these plates are generally much larger than other sensors that we review in this survey, such as cameras, radars, and PIRs, which can be cumbersome. Furthermore, like other binary sensors, EF sensors require a high network density to provide accurate locations.

4.1.2 *Vibration Sensors.*

Vibration-sensing devices placed on the floor can measure from a distance the signals produced by a person's footsteps. In outdoor applications, where these sensors are typically called "seismic sensors", or "geophones", Pakhomov et al. report footstep-based person detection at distances of up to 20 meters [Pakhomov et al. 2003] while, more recently, Audette et al. have achieved 80% detection rates at up to $38m$ even in the presence of noise from nearby vehicles [Audette et al. 2009]. Indoors, Diermaier et al. have shown a similar system using MEMS (microelectromechanical systems) accelerometers to detect room-level locations [Diermaier et al. 2008]. In both of these scenarios, it may be possible to localize people through a geometric localization method, as is often done for acoustic source localization [Potamitis et al. 2004][Cauwenberghs et al. 2005]. Potamitis et al., for instance, localize speakers in a room based on the time delay of arrival of the acoustic signal at different microphones [Potamitis et al. 2004]. These noisy location estimates are processed with a Kalman filter, leading to a localization error between $10cm$ and $40cm$ in their simulations.

The great selling-point of vibration and acoustic sensors (when compared to radio, ultrasound, cameras, etc.) is the simplicity of the signal processing steps. For example, Cauwenberghs et al. have developed a 4-microphone acoustic localization sensor node that can perform bearing angle calculation in hardware with a standard error of less than 2 degrees [Cauwenberghs et al. 2005]. Still, although all of these approaches are useful in relatively quiet scenarios, in busier environments the signals produced by multiple people interfere with one another, leading to a problem that in the audio domain is known as the "cocktail party problem". In the case of microphones, higher resolution location measurements can be obtained from directional sensors that provide bearing information [Moore and McCowan 2003].

4.1.3 *Radio, Ultrasound, Laser.*

— **Scanning Range-Finders:** Range-finders are devices that are used mainly for performance-driven applications. They transmit a signal and measure either the timing or energy of the response echo to calculate distance. The transmitted signal may consist of short series of pulses (ultra wideband) or a modulated carrier wave. Then, to obtain a 2D or 3D image of the environment, range-finders are often aimed at different bearings, in a process called "scanning". This can done by (1) physically rotating the transmitter, receiver, or a reflector, or (2) using multiple transmitters at different locations and phases (known as a *phased array*) to produce constructive and destructive interference at known spatial coordinates. Alternatively, it is also possible to extract this type of spatial information through geometric reasoning (i.e. triangulation, trilateration, or multilateration) using multiple receiving antennas. Depending on the medium used, scanning range-finders have been traditionally called by different names: radar (radio waves), sonar (sound or ultrasound), lidar (visible light), ladar (laser).

Although these sensors can, outdoors, easily extract 2D or 3D snapshots of the environment, in indoor environments the effects of multipath and scattering on clutter add considerable noise to their range and bearing measurements. This makes it difficult to detect people based on their shape alone. Zetik et al. make up for this by taking an approach that is often followed in computer vision: background subtraction. In their 2006 paper [Zetik et al. 2006], the authors describe a method to adaptively model the background signals obtained from an ultra-wideband (UWB) radar. This, they write, allows them to localize people with an accuracy of around $40cm$. In an unusual approach to the detection problem, Chang et al. have used UWB radars to detect people outdoors by modeling their scatter signature [Chang et al. 2009], rather than relying on shape. They show experimentally that this signature acts as a point process, where the time-of-arrival of the signals scattering off a person was found to follow a Gamma distribution, with its mode at the person's location. With this insight, they were able to segment people outdoors by leveraging solely their scatter signature. They extend their approach to detect and track multiple people using a multiple-hypothesis tracker [Chang et al. 2009]. The authors experimentally compare their ranging and velocity inferences to those of ladars, with very positive results.

Compared to radio and ultrasound approaches, laser-based ranging is relatively immune to multipath and clutter. As such, two-dimensional laser range-finders, have been utilized to detect people in a number of different ways. Often, people standing near the sensor are detected by searching for the double-minima pattern of a person's legs. More recently, a few researchers have proposed additional features for person-detection using ladars [Premebida et al. 2009][Arras et al. 2007]. However, due to the difficulty in reaching acceptable false-negative rates, it is more common to pair ladars with traditional camera approaches such as in [Bellotto and Hu 2007][Scheutz et al. 2004][Premebida et al. 2009]. Although less used, three-dimensional ladars are also commercially available [Mesa Imaging]. In theory, any human-sensing algorithm that is designed for stereo imaging should also work with a 3D range-finder, hence these sensors may potentially leverage the large body of research literature on that subject.

— **Doppler-Shift Sensors:** Doppler-shift sensors operate on the principle that waves reflected from a moving object will suffer a frequency shift that is related to the radial component of the object's velocity (i.e. the component toward or away from the sensor's transducer). The simplest Doppler-shift sensors are scalar and often serve as motion sensors, similar to PIRs. Where these differ from PIRs is that Doppler-shift sensors can also provide speed measurements.

Scalar Doppler sensors have found much use in human gait identification. This is often called by the name "micro Doppler", as it relies on the lower-amplitude signals that make up a person's Doppler signature. Significant work has been done to characterize the micro Doppler signature of human gait. For instance, Geisheimer et al. have used high resolution motion capture data to simulate micro Doppler signatures [Geisheimer et al. 2002]. Their simulation shows the contributions of different body parts to the Doppler signature. This closely matches the results found by Gürbüz et al., in their experiments with Doppler radars [Gürbüz et al. 2007]. One-dimensional Doppler radars have also been shown to detect *stationary* people from the motion of their breathing lungs. In [Falconer et al. 2000], for instance, Falconer et al. accomplish this by performing simple statistical analysis on the received Doppler signal: if the kurtosis of the measured samples resembles that of an exponential distribution, then a person is detected. Likewise, heartbeats have also been detected with Doppler radars. In [Zhou et al. 2006], Zhou et al. use a model of the heartbeat signal to devise a likelihood ratio test that can differentiate between scenes with 0 people, 1 person, and more than one. Their system is also able to, under special situations, obtain a reading of the person's heartbeat similar to an electrocardiogram. This could, in the future, prove extremely useful in medium-distance medical applications.

Of course, using similar principles as their radar siblings, micro Doppler *sonars* have also been developed. Kalgaonkar and Raj explore a low-cost acoustic Doppler sonar for gait-based person-identification in [Kalgaonkar and Raj 2007]. In their system, the spectral signatures of individual walkers are learned and used to uniquely identify them using vectors composed of Fourier spectrum slices. A Bayesian classifier is used to identify the individuals. For the laboratory scenario described in the paper, 30 subjects are identified correctly 90% of the time. Similar results are reported in [Zhang and Andreou 2008]. Note, however, that these tests were conducted only for a single walker at a time, moving directly towards or away from the sensor; other motion patterns may not be as easily classifiable. In addition, the subjects' clothing and gait type were consistent across testing and training, which were conducted in a single session in a well-controlled laboratory environment. In light of these concerns, the authors suggest that their system might be best suited in conjunction with existing vision-based solutions. The prime limitation of scalar Doppler sensors, however, is that if multiple people are walking with similar speeds their Doppler signatures will interfere with one another. For this reason, the use of scalar sensors is more fit for applications that require solely person-detection, such as search and rescue operations or border patrol, rather than counting or identification.

Of course, for the purposes of localization and identification Doppler sensors can, like the ranging sensors of the previous section, make use of scanning and/or triangulation. Lin and Ling have reported on Doppler radars that localize multiple moving targets with a narrowband radar using only three antennas, connected to a total of four receivers [Lin and Ling 2006][Lin and Ling 2007]. From the phase difference of the received signals, the authors are able to extract the bearing, elevation, and range, thus localizing moving objects in 3 dimensions. Their solution, however, can only localize multiple people if they are moving at sufficiently distinct speeds — or their Doppler signatures will interfere.

More commonly than the narrowband approaches mentioned above, UWB Doppler radars have been especially favored in the research community for their excellent spatial resolution, the ability to pass through numerous obstacles, and relative immunity to multipath interference [Yarovoy et al. 2006]. A number of commercial solutions for uninstrumented person localization and even through-the-wall (TTW) imaging are based on UWB Doppler signals [Cambridge Consultants][Time Domain b][Camero Tech]. Camero Tech's Xaver 800 radar, for instance, is capable of TTW detection and localization of moving objects (as close as $20cm$ apart) in 3D.

Although the current results with Doppler radars are extremely promising, there are some clear omissions. For instance, authors do not adequately report on their systems' precision / accuracy (i.e. using established, quantitative metrics). Instead, they are mainly interested in simply demonstrating the feasibility of person-detection and localization as a proof-of-concept. As a result there is little information regarding of the accuracy of the localization estimates obtained with these systems, nor on the maximum proximity between two targets that can be disambiguated. At a coarse analysis from published plots, it is clear that noise is still a primary issue with both ranging and Doppler sensors. This needs to be resolved before use in real-world indoor environments, especially as the number of people in the environment increases.

— **Tomographic Sensors:** Tomography has long been used for medical and other specialized applications. More recently, RF tomography has emerged as an area of active research into people detection, counting, localization, and tracking [Wicks et al. 2005][Coetzee et al. 2006]. In the latter work, Coetzee et al. demonstrate the use of narrow band radars for tomographic imaging, demonstrating a resolution of $15.8cm$ with their experiments. More importantly, they derive equations governing the resolution limits of narrowband tomography, and therefore paving the way for future improvements. In [Wilson and Patwari 2009], Wilson and Patwari have shown that tomography can be performed using commodity radio hardware with no modifications. They place a network of radios around the perimeter of the area of interest, and detect objects within the area by the attenuation of the messages transmitted between each pair of nodes.

With a large enough infrastructure of tomography nodes, these approaches can potentially achieve a good level of spatial resolution — albeit requiring a considerable investment in equipment and setup. In addition, it is not clear how phantom detections that occur when multiple people are in the sensing area can be resolved or, at least, minimized. These constraints limit the feasibility of deployments of tomographic systems on a large scale.

4.1.4 *Cameras, Other Imagers.*

Compared to other sensors, cameras are affordable, offer high spatial resolution, and provide a large gamut of information regarding objects in a scene, including size, shape, color, texture, and so on. Perhaps for this reason, the field of computer vision has traditionally been a hotbed for human-sensing research. This high-dimensionality of images and videos also makes them much harder to parse than signals from most other modalities. As such, cameras are often suited for performance-driven scenarios. The greatest challenge in vision-based human-sensing lies in person-detection — and, very often, ways to contain the overwhelming number of false positives produced by most approaches. Once a person is detected, though, *counting* and high-resolution *localization* are trivial. As such, in this section we concentrate solely on presence, track, and identity detection.

— *Presence:* The vast majority of person-detection approaches currently deployed (typically for security scenarios) rely on background subtraction. Examples of such systems includes [Snidaro et al. 2005][Shu et al. 2005]. Under the assumption that a background scene is either static or slowly changing, the main advantage of background subtraction is that it allows quick detection of non-background objects. Although numerous background subtraction methods have been proposed, such as [Barnich and Van Droogenbroeck 2009], [Li et al. 2003], and [Javed et al. 2002], in scenarios where the background varies, these methods tend to fail, or adapt much too slowly. For instance, in office or meeting-room situations, background objects such as chairs are moved quite frequently, leading to a large number of false positives. Other approaches may instead employ object segmentation or pattern matching. Object segmentation is the extraction of the person's shape from the image directly, without requiring a background subtraction preprocessing step, such as with Rother et al.'s GrabCut algorithm [Rother et al. 2004]. Meanwhile, pattern matching can be as simple as convolving the input image with sample images of the object to be detected, although most often this comparison is done in other feature spaces such as SIFT (scale-invariant feature transform) [Lowe 2004] and HoG (histograms of oriented gradients) [Dalai et al. 2005]. Then, objects are typically classified from a number mathematical tools such as PCA [Turk and Pentland 1991], support vector machines [Dalai et al. 2005], AdaBoost classifiers [Viola and Jones 2002], and neural networks [Rowley et al. 1996].

To aid in person-detection, it is often advantageous to explore alternative imaging hardware. A common technique is to use depth information from stereo cameras as an additional cue to differentiate people from the background scenery. This is done, for instance, in [Harville and Li 2004] (which employs simple template-matching on the depth images for use in a person-following robot) and in [Ess et al. 2009] for pedestrian detection from moving vehicles. More interestingly, Bertozzi et al. describe a person detection system that employs a stereo pair of thermal imagers in [Bertozzi et al. 2007]. Thermal imagers are able to differentiate people from background objects through their temperature. As such, they have an enormous potential for use in people sensing systems. Although commercially available for some time [FLIR], these sensors have traditionally been too expensive to allow for widespread use, with even a 32×31-element array costing over a thousand dollars [Heimann Sensors]. However, given recent advances in microbolometer technology and the impending expiration of key patents, there may be a surge in thermal-based human detection. Furthermore, P. Hobbs from IBM has successfully demonstrated a 96-pixel thermal imager technology that is orders of magnitude cheaper to manufacture than previous hardware [Hobbs 2001]. Such low-resolution sensors have been successfully shown to detect, count and localize people from top-view cameras in [Stogdale et al. 2003].

Of course, as is the case with the Doppler-shift sensors of the previous section, a simple and efficient method to detect people is to leverage motion information. In computer vision, this translates to either frame-differencing (i.e. subtracting consecutive frames pixelwise) or optical flow (i.e. measuring the motion vector of each pixel over a number of frames). Some advantages of using motion include an immunity to long-lived misdetections when compared to background subtraction or pattern matching approaches, and low processing requirements. A person-localization wireless camera network that operates on frame-differencing has been demonstrated by Teixeira et al. to execute in real-time on low-end hardware through the use of a density estimation technique called averaged shifted histograms [Teixeira and Savvides 2008]. Besides, a growing body of research is being dedicated to "smart cameras" that extract motion information at the hardware level [Lichtsteiner et al. 2008][Lichtsteiner et al. 2004][Fu and Culurciello 2008], making motion an evermore attractive feature for fast, low-power scene understanding. The main disadvantage of motion-based imaging, however, is that people "disappear" when they stop moving, requiring further processing in higher-level layers.

— *Tracking:* Where cameras and imaging sensors are farthest ahead from other uninstrumented modalities is in tracking and identification. This is not because the tracking algorithms themselves are fundamentally different from those in other modalities — they are not —, but due to the large breadth of information that cameras can capture to solve the correspondence problem. Some examples are height, width, shape, colors, speed, etc., and several specialized image features such as SIFT and HoG. Like other sensing modalities, most camera-based trackers operate on a Bayesian principle of using transition and emission probabilities to calculate the *a posteriori* probabilities of all plausible tracks. Classical approaches to this include multiple hypotheses tracking [Reid 1979] and joint-probabilistic data association [Bar-Shalom and Tse 1975], while more recently Monte-Carlo approaches have been favored (i.e. particle filtering) [Isard and Blake 1998]. The core differences between most trackers, though, lie in the appearance models that they employ and the different methods to handle the combinatorial explosion of the track space. Even so, obtaining correct tracks in crowded scenarios is still an open research problem, especially in the presence of clutter and occlusions. For further discussion of camera-based tracking see, for instance, [Enzweiler and Gavrila 2008] or [Yilmaz et al. 2006].

— *Identification:* Finally, cameras have been used to identify people using both face- and gait-recognition. Although almost 20 years old, one of the most widely used approaches is Turk and Pentland's eigenfaces-based method [Turk and Pentland 1991]. In their 1991 paper, they show it is possible to identify people with the vector coefficients of the person's face when represented in the space spanned by the eigenvector basis extracted by PCA. This is an example of a holistic approach (i.e. searches for entire faces) and thus is typically not robust to occlusions or unexpected variations in facial expressions. Depending on the number of same-person images in the training set, and on the similarity between the training and testing sets, PCA-based methods have been shown to achieve an precision of 99% [Wiskott et al. 1997]. However, this number falls dramatically as facial expressions change, rotate, or the lighting varies. For training sets with only a single image per person, other approaches have been proposed. For instance, one option is to consider a face as a group of fiducial points (eyes, nose, mouth, etc.), as done by Wiskott et al. [Wiskott et al. 1997]. Their approach, elastic bunch graph mapping (EBGM), consists of building a novel graph-like structure (called a *bunch graph*) where each edge corresponds to a fiducial point. Each vertex of the graph contains a "bunch" composed of Gabor wavelet coefficients of possible states of the fiducial point. For example the states of the "eye" node may be "open", "closed", "male", "female", and so on. People are, then, recognized by using a graph similarity measure. Despite the high complexity of this and other single-training-image approaches, the reported precision values vary widely (between 9% to 98%), with an average of 84% for non-rotated images and 39% for rotated. Note that, as opposed to the person-identification results given for other sensing modalities, these numbers come from datasets consisting of hundreds of people, and so the recognition rates must invariably suffer. More information on face-recognition approaches can be found in [Tan et al. 2006].

While face recognition approaches to person-identification saw their first spike in activity during the 80s, gait recognition only started to attract such levels of attention about a decade later. Most gait recognition methods are strongly dependent on the person's exact silhouette, and fail when people wear different clothing, carry silhouette-altering objects such as backpacks, or when the environment is highly cluttered (due to increased segmentation errors). One of the simplest approaches, discussed in [Kale et al. 2003], is to compare each silhouette's y-histogram to a database using time series correlation methods such as dynamic time warping. In [Wang et al. 2003], each person's silhouette was "unwrapped" into a 1-dimensional array which is then matched against a database using the largest PCA components. They report an precision of 70.42% across different views of the same person, and as low as 34.33% for different walking surfaces (grass vs. concrete). These numbers fall dramatically to 14.29% when all three tested conditions are varied (view angles, shoe types, and surface types). Similarly, of all methods surveyed by Sarkar et al., the best values for those precision rates were found to be 99%, 36%, and 23% respectively [Sarkar et al. 2005]. More recently, rates of 93%, 88% and 33% were obtained by [Tao et al. 2007] using averaged gait energy images and linear discriminant analysis along with a novel preprocessing method (general tensor discriminant analysis) for dimensionality reduction.

4.2 Instrumented, Single-Modality Approaches

Instrumented approaches have the unique advantage that they can leverage wearable devices that openly announce their presence. The result is that these approaches can attain near-perfect person-detection and counting and, since this local information can contain a unique identifier, they also achieve near-perfect identification and tracking. Thus, the greatest research problem in the category instrumented people-sensors lies in the 3^{rd} STP: that is, localization.

4.2.1 *Device-to-Device Ranging.*

For high-accuracy localization, it is possible to improve upon the signaling properties of range-finders reviewed earlier in this paper by taking range measurements between devices on the people and devices on the external infrastructure. This device-to-device approach to ranging, which emerged from robot and sensor node localization, has, as of late, been increasingly applied for human-sensing through the use of mobile phones. The most known example in this class is, of course, the global positioning system (GPS). In GPS, satellites belonging to a large supporting infrastructure transmit beacon packets carrying precise timestamps as well as their location at that time. Distances are, then, calculated from the propagation time of the radio packet (which, at GPS-like spatial scales is non-negligible) and the speed of light. This is known as the time of arrival (TOA) method. However, since the aging GPS satellites transmit their beacon in $30s$ intervals, it takes a receiver several minutes to obtain enough information to self-localize from a cold boot. Nowadays this is handled using a number of techniques, such as almanac memorization and AGPS (assisted GPS), which can to speed up a first-order location estimate considerably. Still, due to a number of sources of noise in the internal clocks and the signal propagation time, these location estimates are limited to an accuracy of around $10m$ — and often much worse. What is more, GPS does not function in most indoor environments, as the beacons don't generally propagate through walls.

In light of these shortcomings, a number of alternative approaches to localization have been proposed to achieve centimeter-scale accuracy in indoor environments. These approaches may, like GPS, leverage the signal time of arrival, or other properties such as time difference of arrival (TDOA) [Priyantha et al. 2000][Savvides et al. 2001][Harter et al. 2002], signal strength (SS) [Ni et al. 2004][Krumm et al. 2002], and angle of arrival (AOA) [Nasipuri and Li 2002][Rong and Sichitiu 2006]. Signal strength approaches such as RFID are typically highly prone to noise from interference and the sensitivity patterns of anisotropic antennas [Lymberopoulos et al. 2006]. AOA approaches must also handle antenna-related distortions, which can lead to large positional errors as the target distance increases and must be addressed with additional processing, such as the maximum likelihood algorithm in [Rappaport et al. 1996]. For this reason, TOA and TDOA have seen the most success, being limited mainly by clock synchronization errors. For a full treatment of the different localization methods see, for instance, [Mao et al. 2007] or [Srinivasan and Wu 2007]. Regarding accuracy, early efforts have reported localization errors under $20cm$ for a person traveling at $1m/s$ [Smith et al. 2004], and less than $9cm$ when using a high-density network of beacon nodes (100 nodes for 2 rooms) [Harter et al. 2002]. This latter system has also been commercialized in a version supporting both AOA and TDOA, plus local inertial sensors for better power savings [UbiSense]. Current systems using UWB radios have further improved their accuracy to provide centimeter-accuracy even in cluttered indoor environments [Alsindi et al. 2009]. For more information, a detailed theoretical analysis of the fundamental limits of UWB localization is given in [Gezici et al. 2005], while [Alsindi et al. 2009] provides extensive experimental characterization. Following the example of other range-finders reviewed earlier in this paper, device-to-device ranging may also make use of Doppler-shift effects. This has been investigated by Kusy et al. for moving targets [Kusy et al. 2007], and subsequently extended by Chang et al. to localize stationary targets by using spinning sensors [Chang et al. 2008]. These solutions, however offer a relatively poor spatial accuracy, in the order of one meter.

Nonetheless, device-to-device ranging is an incredibly promising sensor configuration for localization in human-sensing applications. Their main disadvantage lies is the requirement for a complex infrastructure of beacon nodes, which can be expensive and cumbersome to install and manage. This would seem to make them suitable solely for performance-driven scenarios, but in fact deployments can be easily adjustable to resource-constrained applications by simply reducing the infrastructure (e.g. RFID). Unfortunately, reported results are often obtained in the ideal conditions of a lab setup, where devices are placed on special supports that greatly reduce multipath and do not absorb RF signals as human bodies do. Such an analysis is notably missing from the existing literature. Finally, commercial solutions are currently available, aimed at tracking people or packages in large stores, warehouses, office buildings, and hospitals [Time Domain a], [UbiSense].

4.2.2 *Environment Recognition.*

As described in Section 3.1, it is possible to take advantage of both natural and artificial properties of the external environment in order to localize a person. This is the basic premise of environment-recognition sensors, which listen to signals from the environment and compares them to pre-acquired signatures in a local database. The main challenge with this method is handling changes in the environment, such as different lighting conditions or radio fingerprint variations. The most common example of environment-sensing is radio signal strength fingerprinting, which has been widely employed in mobile phones for the past few years. This method stems from the work by Castro et al. in which a database of WiFi signal strength signatures was used to infer the room in which a WiFi client was placed [Castro et al. 2001]. Since then, other researchers have used improved statistical models to lower the mean localization error from the room-level to under $1.5m$ [Ladd et al. 2005][Roos et al. 2002], and even to the sub-meter range [Youssef and Agrawala 2008]. Of course, the same techniques can be applied to other types of radio signals, such as GSM (Global System for Mobile Communications) [Otsason et al. 2005][Varshavsky et al. 2006] for up to a few meters of accuracy. To further improve the localization error, Youssef and Agrawala used signal modeling techniques to account for not only for small-scale spatial variations in the received signal strength, but also temporal variations [Youssef and Agrawala 2008]. They report average distance errors of under $60cm$ in scenarios with a high concentration of WiFi base-stations and where the offline database construction process was performed for a dense set of locations. It is unclear, however, whether their system can achieve such low errors for targets that move, since multiple samples are required to filter out temporal signal strength variations. Furthermore, the standard deviation of the errors in all of these systems is relatively large, typically near the $1m$ range. As a consequence, current RF fingerprinting methods are, in reality, limited to a relatively coarse localization accuracy.

Although less commonly used for human localization, other environment recognition methods that have been considered in the literature include camera-based [Se et al. 2005][Schindler et al. 2007], ladar-based [Zlot and Bosse 2009], and microphone-based [Korpipää et al. 2003] approaches. The former two types are often used for vehicle localization, but the same systems should be directly applicable to personal localization using wearable cameras. In a car localization application, Schindler et al. report that over 40% of their location estimates had errors greater than $10m$ [Schindler et al. 2007]. Perhaps due to these large errors, in indoor environments most systems of this kind are geared toward room-recognition applications rather than accurate localization. Pronobis et al. have recently built a large database of indoor images with three different robots, two different buildings, and three lighting conditions to serve as a benchmark for other researchers in the field [Pronobis et al. 2009]. They also propose a system to be used as a baseline in that benchmark, which is able to correctly recognize different rooms at rates between 74.5% and 87.3% in the most challenging case (where different lighting is present during training and evaluation). Since the precision/accuracy of all environment recognition methods is highly dependent on the quality of their database, it is likely that with more detailed databases (where each location contains a large number of exemplars in varying environmental conditions) these approaches may achieve a localization accuracy in the order of 1 meter.

4.2.3 *Dead-Reckoning.*

Dead-reckoning is the process of inferring the path of a moving body from inertial measurements, such as speed or acceleration. The sensors that are most widely used for this purpose are inertial measurement units (IMUs) containing accelerometers (acceleration sensors), gyroscopes (angular velocity sensors), and/or magnetometers (magnetic field sensors, used as a compass). The premise of dead-reckoning is that if a person's location at time t is known, then their location at $t + \delta t$ can be found by simply integrating their known velocity, or twice-integrating their acceleration, during the time interval δt. However, a number of sources of error accumulate during this integration, causing the location estimate to quickly diverge, often within a few seconds. The most prominent sources of error in dead-reckoning are calibration errors, quantization errors, the effect of gravity on the accelerometer, the effect of external magnetic fields and metals on the compass, and crosstalk between orthogonal signal components, to name a few. As such, the novelty in any dead-reckoning method lies in the different ways to mitigate these uncertainties.

One often-employed solution is to place the IMU on one of the person's shoes, rather than on the body, which allows for so-called zero-velocity updates (ZUPTs)[Dorota et al. 2002]: whenever the IMU detects that the shoe is touching the ground, it is safe to assume that the true velocity and acceleration of that foot is zero. Therefore, if at that moment the velocity inference is set to $0m/s$, then the errors accumulated from the integration of the acceleration component will be effectively discarded. Using this method, Ojeda and Borenstein have been able to infer a person's path in 3 dimensions with errors as little as 2% of the distance traveled [Ojeda and Borenstein 2007]. I.e., for a distance of $100m$, the localization error is expected to be as little as $2m$. Quite impressively, Foxlin has shown with his NavShoe system that errors of 0.2% are achievable by entering the ZUPT information as pseudo-measurements in an extended Kalman filter (EKF), rather than simply setting the velocity to zero [Foxlin 2005]. Another approach is to use the accelerometer as a step counter (pedometer) and to calculate the length of the person's step on the fly using an empirically-obtained equation. This, he writes, has been shown to lead to distance errors of 8% of the distance walked. Interestingly enough, in a recent paper Jimenez et al. have compared ZUPT against Weinberg's equation, with perhaps unexpected results: ZUPT errors were found to be in the range of 0.62%–1.15%, while a much lower error of 0.30%–0.78% was obtained for Weinberg [Jiménez et al. 2009]. Either way, the fact is that dead-reckoning with shoe-mounted IMUs is quickly becoming a viable method for motion path inference. For sensors mounted in other locations (such as mobile phones inside a person's pocket), however, the dead-reckoning problem is still largely unsolved, on account of integration errors. The solution for this will likely arrive in the form of more accurate inertial sensors.

4.3 Sensor Fusion Approaches

Sensor fusion approaches build upon the use of multiple sensors or sensing modalities in an attempt to combine their advantages while cancelling out their disadvantages as much as possible. In this section we quickly review a small number of sensor fusion examples to illustrate some of the benefits of multi-modality sensing. The capabilities of each are summarized in Table III.

Note that, both in the table and in the paragraphs below, the capabilities that we report are an account of the capabilities of the specific sensor fusion systems that are cited — that is, they *should not* be taken as a broad-reaching assessment of the combination of those sensing modalities.

4.3.1 *Cameras & Microphones.*

The idea of "sensor fusion" comes naturally in some applications. Consider, for example, a fully-automated video conference system where it is desired that anyone currently speaking be placed within the field-of-view of the camera by actuating pan-tilt-zoom motors. In such a case, it is only natural to conclude that the solution must involve the use of both microphone arrays (for sound source localization) and cameras (for the actual filming). And upon further investigation, it becomes clear that the speakers localized by the microphone arrays can be more precisely detected by fusing face-recognition information from the camera. For this, Shen and Rui propose the use of a two-level particle filter where the first level computes separate track hypotheses for each face seen by the camera and each speaker located with the microphones, while the second level joins the hypotheses from all modalities [Chen and Rui 2004]. Although they do not provide numerical results, they report that speakers are tracked more precisely/accurately than by sound alone, and that, in some instances, visual ambiguities (when a person moves too fast, for instance) are resolved from the audio fusion. Furthermore, Gatica-Perez et al. show that Markov chain Monte Carlo (MCMC) techniques applied to a particle filter lead to an improvement of close to 0.5 points to the tracker's F-measure (average of precision and recall) in complex scenarios [Gatica-Perez et al. 2007].

4.3.2 *Camera & Laser Range-Finder.*

In the same vein as the speaker localization approach described above, where face detection results from a camera were enhanced through an additional sensing modality (microphones, in that case), several researchers have developed robots that detect people by fusing a face detection system with a laser range finder [Bellotto and Hu 2007][Brooks and Williams 2003][Kleinehagenbrock et al. 2002]. By coupling face detection algorithms (using vision) with leg detection methods (using ladars), these authors are able to localize people around their robots even when their faces are not visible. Belloto and Hu's system uses a simple flowchart to fuse the two sensors, while Brooks and Williams use a more standard (and probably more robust) Kalman filter. Sadly, neither group provides quantitative metrics for the detection precision nor localization accuracy.

4.3.3 *Dead-Reckoning & Device-to-Device Ranging.*

As described in Section 4.2.3, dead-reckoning is by itself prone to cumulative errors which can quickly become unmanageable. An common solution to this issue is to periodically correct the person's absolute location using a separate sensor such as a GPS [Judd 1997][Beauregard and Haas 2006]. This is typically done by incorporating both the inertial measurements and the absolute locations from the GPS into a single filter, usually a Kalman or particle filter. This approach is followed, for instance, by Klingbeil et al. for indoor localization. The novelty in their case is that, in place of GPS measurements, they utilize a supporting network of infrastructure nodes that is able to coarsely localize a person using signal-strength-based binary proximity measurements [Klingbeil and Wark 2008]. Using a particle filter, they report a mean error rate of $2m$ in their experiments where the infrastructure nodes were placed 5 to $10m$ apart. With the addition of knowledge about the building's floorplan (which allows them to prune particles where people move through walls), they show that the mean error can be reduced to $1.2m$. Clearly, further accuracy can be directly obtained by using the full signal strength measurements rather than thresholding them, or by utilizing a TOA or TDOA approach instead.

4.3.4 *Dead Reckoning & Environment Recognition with Wearable Camera.*

Yet another variation on error correction for dead-reckoning is given in [Kourogi and Kurata 2003]. In that work, Kourogi and Kurata describe a system comprised of a wearable inertial measurement unit and a head-mounted camera. The intuition is that the dead-reckoning errors can be corrected whenever the camera recognizes the surrounding environment and provides an absolute localization estimate. Using the inertial sensors alone, their system employs a number of techniques to keep the dead-reckoning error at around 3.66% of the distance traveled. With the addition of the camera, the authors report being are able to periodically correct the dead-reckoning errors at all locations present in their image database, although they do not provide a measure of the overall localization accuracy of their system.

4.3.5 *Infrastructure Cameras & Wearable IMU.*

Another variation on the topic of inertial sensors plus external localization device is given in the work of Teixeira et al. [Teixeira et al. 2009b][Teixeira et al. 2009a][Teixeira et al. 2010]. In order to eschew the well-known cumulative errors of other approaches, the authors avoid performing dead-reckoning altogether. That is, they do not attempt to estimate the person's motion path from the inertial measurements, but, rather, utilize other properties of the inertial data. In their proposed systems, a camera network in the environment detects and localizes people while wearable sensors are leveraged to provide IDs to those detections. Their intuition is that the acceleration measured by a wearable accelerometer should match the acceleration of the person's image in the video. The challenge, then, is to find the best-matching acceleration pairs. The problem was defined as a bipartite graph matching where one set of vertices represents the different accelerometers in the scene, and the other set all current track hypotheses from the video. The authors approached the edge weights of the bipartite graph in several ways, including using Pearson's correlation coefficient [Teixeira et al. 2009b], a gait-synchronization metric [Teixeira et al. 2009a], and finally the maximum a-posteriori (MAP) likelihood of each two signals originating from the same person. The latter yielded the best results, with an precision above 90% in uncrowded scenarios.

4.3.6 *Laser Range-Finders & ID badges (infrared and ultrasound).*

Schulz et al. have presented a system to detect, count, localize, track, and identify people in an office environment using laser range-finders and wearable ID badges [Schulz et al. 2003]. In their system, the laser range-finders are used to anonymously detect and localize people in the environment, while the wearable ID badges provide sparse identity observations as people approach ID readers in the infrastructure. In their paper, they propose a Rao-Blackwellized particle filter that builds tracks from the laser measurements while simultaneously making ID inferences. The authors report a success rate of 10 out of 10 experiments, where a "success" was defined as the correct hypothesis being present within all hypotheses generated by the particle filter.

4.3.7 *Camera and RFID.*

In order to build a robot that is able to follow a person in a crowded environment, Germa et al. have recently explored the fusion of cameras with RFID sensors [Germa et al. 2010]. In that work, the authors equip a robot with a camera and an RFID reader connected to an array of 8 antennas aimed radially at different angles from its center. A particle filter is used to fuse the azimuth measurements from the antenna array with the detections from the camera by simply rejecting all particles that do not fall within the detected azimuth range. The authors, then, show that the fusion approach significantly outperforms the vision-only solution: using solely the camera, their system is able to track an given person only 21% of the times, while with the addition of the RFID cues this number increases to 86%.

5. DISCUSSION

Table II summarizes the capabilities of all sensing modalities surveyed in this paper, particularly emphasizing their detection performance for the 5 STPs, as well as network density. Although Table II necessarily abstracts away vital details discussed in Sections 4.1 and 4.2, it does make a few fundamental tendencies stand out. For one, the table clearly shows that instrumented approaches, on average, perform better than uninstrumented ones, especially for the purpose of identity-detection. The trade-off, of course, lies on a requirement for extraneous communication devices (in the case of passive sensors) and a large increase in network density (in the case

sensing modalities	signaling	presence	count	location	track	identity	ND
Uninstrumented							
Contact Sensors	passive	◯	∘	∘	∘		4
Pressure Sensors	passive	◯	∘	∘	∘	·	4
Chemosensors	passive	–	–	–	–	–	?
Photodetectors	passive	·	·	·	·		4
Cameras	either	◯	◯	◯	◯	∘	0
Thermal Imagers	passive	◯	◯	◯	∘	·	0
Breakbeam Sensors	active	∘	–	–	–		4
Scalar Range-Finders	active	∘	–	–	–		0
Scanning Range-Finders	active	∘	∘	◯	∘	·	0
Tomographic Sensors	active	◯	∘	∘	∘	–	5
EF Sensors	active	◯	∘	∘	∘		4
Doppler-Shift Sensors	active	∘	∘	∘	∘	·	0
Motion Sensors	either	∘	·	·	·		2
Seismic and Inertial Sensors	passive	∘	·	·	·		3
Microphones	passive	∘	·	·	·	·	1
Instrumented							
Wearable Inertial Sensors	passive	◯↯	◯↯	∘↯	◯↯	◯↯	×
Wearable Environment Recognition	passive	◯↯	◯↯	∘↯	◯↯	◯↯	×
Wearable SS Device-to-Device Rangers	either	◯	◯	·	◯	◯	2
Wearable AA Device-to-Device Rangers	active	◯	◯	∘	◯	◯	2
Wearable TOA/TDOA Dev.-to-Dev. Rang.	active	◯	◯	◯	◯	◯	2
Wearable Doppler-Shift Sensors	active	◯	◯	∘	◯	◯	2

◯ = good performance ∘ = medium performance · = low performance
– = plausible, but no detailed literature ? = no literature
↯ = requires communications (i.e. depends on the addition of a radio)
× = not applicable: this is solely a self-sensing method, so no network is involved.

Table II. Summary of the capabilities of each sensing modality. The network density (ND) is described in Section 3.2. Lower ND values are typically preferable to higher ones. Since we did not establish the size of the sensing area, the numeral value of the network density is meaningless by itself. The important value to note is the difference between NDs for two competing modalities.

of active sensors). For instance, a comparison between the best-performing instrumented and uninstrumented modalities shows that the former requires a network with approximately 4 times as many sensors as the latter ($ND = 2$ vs. $ND = 0$).

The overall best modality for instrumented scenarios is TOA/TDOA device-to-device ranging [Mao et al. 2007][Srinivasan and Wu 2007], especially those approaches using UWB [Alsindi et al. 2009]. These are able to attain good localization accuracy both outdoors and indoors (and are even available commercially [Time Domain a]) albeit requiring the installation of a complex infrastructure. For self-localization without the burden of additional infrastructure, GSM- [Otsason et al. 2005][Varshavsky et al. 2006] and WiFi-based [Ladd et al. 2005][Roos et al. 2002][Castro et al. 2001] environment sensing [Youssef and Agrawala 2008] is a good compromise with an accuracy of a couple of meters, which is acceptable in many use-cases. What is critically absent in the device-to-device ranging literature at this time is an in-depth characterization of the effects of different real-world factors on system performance, such as the body's RF absorption properties given different poses, antenna orientations, device placement locations, clothing, and so on. Without this, the results reported in Section 4.3.3 should be interpreted with caution.

For uninstrumented scenarios, the best modality across the board is vision (i.e. cameras and other imagers). Computer vision is far ahead from other instrumented modalities not only with respect to spatial-resolution and precision metrics, but also in terms of having the most field-tested solutions. For instance, background subtraction [Barnich and Van Droogenbroeck 2009][Li et al. 2003][Javed et al. 2002] and motion differencing [Teixeira and Savvides 2008] are often a "good enough" solution for quick-and-easy deployments. However, for reasons listed in Section 4.1.4, these solutions have a number of disadvantages. To bypass them, the ideal person detector should be able to discover a person given only a single frame and with no prior knowledge about the scene. This is only attainable using more complex pattern-matching approaches based on learned appearance models in smart feature spaces [Turk and Pentland 1991][Viola and Jones 2002][Lowe 2004][Dalai et al. 2005]. These specialized feature spaces, which make use of the abundance of personally-identifying features that are available in an image, also make cameras the best uninstrumented sensors for detecting the two higher spatio-temporal properties (i.e. tracking and identification). Although at present time these models are still lacking in precision, it is known that pattern-matching person detection, tracking, and identification problems are solvable, from the simple fact that our brains are able to do so astoundingly well.

Scanning range-finders [Zetik et al. 2006][Chang et al. 2009][Chang et al. 2009] and Doppler-shift sensors [Lin and Ling 2006][Lin and Ling 2007][Yarovoy et al. 2006] hold a somewhat distant second place in all of these regards. Where they *do* displace cameras is in their relatively low computational overhead, resistance to occlusions, and indifference to illumination. They are not, however, able to robustly detect static people, nor can they resolve tracking ambiguities with high precision. Yet, we expect such sensors to advance quickly in the next few years, especially using off-the-shelf radios as they become more and more ubiquitous.

For resource-constrained scenarios, the preferred solution is to employ simple binary sensors. These can be used as cost-effective occupancy sensors (usually in bathrooms, corridors, etc.) that, when smartly networked, allow for localization and piecewise tracking as well [Aslam et al. 2003][Oh and Sastry 2005][Kim et al. 2005][Xiangqian et al. 2008]. However, due to a number of issues with most existing binary sensors (for example, PIR cannot sense people who are standing still; floor tiles and EF sensors are difficult to install and interpret), there is a distinct research opportunity here to develop a true binary human-presence-detector. The solution will likely take the form of scalar Doppler-shift sensors that are not only used for large-scale motion [Gürbüz et al. 2007][Geisheimer et al. 2002] but also to detect breathing and heartbeat motions when a person is otherwise completely still [Falconer et al. 2000][Zhou et al. 2006].

5.1 Opportunity: Sensor Fusion at Massive Scales

Despite the progress, a number of classic sensing problems are not only still largely unsolved, but also *amplified* when applied to the domain of human-sensing as opposed to rigid objects. For instance, no sensing modality or sensor fusion approach can robustly[6] perform even presence detection — the lowest-level spatio-temporal

[6]With less than 1% false positives, and less than 1% false negatives

sensor fusion approaches	signaling	presence	count	location	track	identity	ND
Uninstrumented							
Camera & Microphones	passive	◯	◯	◯	∘		0
Camera & Laser Range-Finder	active	◯	◯	◯	∘		0
Instrumented							
Dead-Reck. & Dev.-to-Dev. Ranging	active			∘↯	◯↯	◯↯	×
Dead-Reck. & Env. Recog. w. Wear. Cam.	active			∘↯	∘↯	◯↯	×
Infrastructure Cameras & Wearable IMU	passive	◯	◯	◯	◯↯	◯↯	0
Laser Range-Finders & Wearable ID Badges	active	◯	◯	◯	◯	◯	2
Camera & RFID	active	∘	∘	∘	∘	◯	0

◯ = good performance   ∘ = medium performance   · = low performance
↯ = requires communications (e.g. self-localization followed by broadcasting)
× = not applicable: this is solely a self-sensing method, so no network is involved.

Table III. Summary of the capabilities of existing sensor fusion approaches.

property! In fact, the false-positive and false-negative rates of the best approaches typically lie near the 10% mark in uncontrolled environments. Likewise, multiple-person tracking is still a clear challenge in real-world, medium–crowd-density environments such as office buildings and airports. People are easily lost, and tracks are often terminated or, even worse, incorrectly extended in the face of ambiguities. Therefore, in spite of advances in the field, truly robust human-sensing is still by and large an unrealized goal.

As discussed in Section 4.3, and as has been long advocated in the pertinent research communities, the solution to these problems is expected to come from the fusion of multiple sensors or sensing modalities. Still, comparing each row in Table III with the respective modalities listed in Table II, it becomes clear that the crop of current sensor fusion research do not leverage the full potential of their specific sensor combinations. We believe a primary reason for this lies on the difficulty of designing and fine-tuning current fusion systems, since the entire design process must be performed by hand for each new problem instance.

More importantly, looking into a future where human-sensing networks will consist of massive numbers of highly heterogeneous sensors, hand-designing a fusion system for each problem instance will simply no longer be feasible. The number of parameters involved will be too numerous. Due to cost considerations, new sensing hardware will often not replace older generations in already-deployed networks — rather, several generations of sensors will operate alongside one another. Likewise, it is probable that the private sensor networks which are nowadays being deployed by distinct entities will, at some point, become interconnected into a great sensor internet. This new structure will certainly contain sensors from an assortment of vendors, with highly varying sensing characteristics (error distributions, sampling rates, spatial resolution, etc.). **As a result, we foresee a pressing demand for automated sensor fusion frameworks**, which will estimate the parameters of each particular instance of the human sensing problem on-the-fly through new unsupervised learning techniques.

Let us consider, as a possible starting-point, the sensor fusion systems surveyed in Section 4.3. It should be clear from that discussion that a mathematical tool that has emerged as an almost universally-accepted foundation for sensor fusion is the particle filter (PF) [Arulampalam et al. 2002]. The main reason for this is that PFs excel in handling complex probability distributions, such as those that may arise in fusion scenarios, by inherently representing them within a set of "particles". In essence, particle filters can be summarized in the following manner: (1) At each timestep k, the new measurement from each sensor goes through a data alignment step. (2) A density function is computed for each sensor's measurement, by taking into consideration the known error characteristics of that sensor. This represents the likelihood of the state given the measurement from that modality alone. (3) The probability density that had been computed at time $k-1$ is propagated into time k. (4) The densities from steps 2 and 3 are fused to obtain the density for timestep k, from which a state inference can be made.

Therefore, the designer of a PF-based sensor fusion system must currently enter the following information into the filter: the data alignment equations from step 1, the measurement likelihoods that are used in step 2, and the state propagation equations from step 3. All of these depend on details that concern the specific instance of the problem, such as the specific sensors being used, the expected behavior of the people in the scene, and the expected characteristics of the scene itself. In a truly plug-and-play fusion framework, though, these would not be available *a priori*. The problem that we are posing, then, is **to estimate these three pieces of information online in an unsupervised fashion**.

More concretely, consider the following example scenario. A researcher is given access to data from a large network of floor tiles, cameras, and ladars, which are densely placed over an entire office building with often-overlapping sensing areas. He is told that the cameras are mounted on the ceilings, pointing diagonally down, and that the ladars are on the walls, scanning horizontally to produce a 2D slice of the environment. But he does not know the precise sensor placement, nor does he know the exact sensing characteristics of the different pieces of hardware, which may have originated from different vendors. The researcher also has access to the tracks that were locally computed by each camera and ladar — however, due to sensing overlaps, the same person is often observed simultaneously by multiple tracks. Given this data, can a sensor fusion framework be built to "stitch" the tracks and floor tile observations together, so that (1) each person is described by a single unified track across the entire building, and (2) each person's location is more accurately measured than with any single sensing modality?

6. CONCLUSION

As computer systems transition from people's desks to their pockets and the world around them, there will be an increasing demand for person-centric information. In this paper we have surveyed the existing methods to acquire such information, and classified them according to a taxonomy of human-sensing. By analyzing the existing sensing modalities and sensor fusion approaches within the framework of our taxonomy, we anticipate that future human-sensing systems will likely consist of an amalgamation of three types of sensors:

(1) **Massive numbers of low-cost binary sensors** (usually motion sensors) to provide somewhat coarse information regarding the 5 STPs. Although coarse, this information will be appropriate for resource-constrained applications —

especially as binary-sensor fusion algorithms [Aslam et al. 2003][Oh and Sastry 2005][Kim et al. 2005][Xiangqian et al. 2008] are further improved.

(2) **A relatively smaller number of cameras placed at key locations**, wherever it is desirable to extract people's poses and gestures, and to obtain more fine-grained estimates of people's locations and tracks, including *some* idea of their ID.

(3) **Opportunistic use of sensors on mobile phones as they become available in an environment**, gracefully degrading the quality of the provided services for phone-less users or users with different privacy settings.

Of course, this setup will certainly suffer some modifications in a few specific scenarios, such as for long-distance outdoors situations where the cameras may be replaced by ranging [Zetik et al. 2006][Chang et al. 2009][Chang et al. 2009] or Doppler devices [Lin and Ling 2006][Lin and Ling 2007][Yarovoy et al. 2006], and the binary motion sensors by binary seismic sensors [Audette et al. 2009][Pakhomov et al. 2003]. In addition, wherever high-accuracy localization and precise identification are the main design constraint, device-to-device ranging [Mao et al. 2007][Srinivasan and Wu 2007][Alsindi et al. 2009] will continue to be the dominant solution in the years to come.

To further increase the sensing performance of the setup described above, we believe some design changes will necessarily take place within the sensor hardware itself. For one, the limitations of the current crop of binary sensors could well be bypassed through the use of scalar micro Doppler sensors that are able to detect breathing [Falconer et al. 2000] and heart motions [Zhou et al. 2006]. However, cheap micro Doppler sensors are not currently available. Similarly, the relative difficulty in detecting and segmenting people using vision alone would be greatly alleviated if a multi-modal camera were created containing a regular imager, a thermal imager, and a ladar. Since these three modalities are structurally similar (they all consist of 2D pixel arrays), the data produced by such a trimodal imager could be easily fused through well-known methods that have been developed for stereo imaging.

The multiple facets of human-sensing will no doubt become a hotbed of innovative research in the coming years. The great potential of this field lies in the fact that the more research results are obtained, the greater and the more complex will the datasets grow, thus leading to further questions to be asked — and the need for more specialized sensors to answer them.

Acknowledgements

This work was partially funded by the National Science Foundation under projects ECCS 0622133, IIS 0715180, CNS 0721632, and CNS 0448082. Any opinions, findings and conclusions or recommendation expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation (NSF).

REFERENCES

Alsindi, N., Alavi, B., and Pahlavan, K. 2009. Measurement and modeling of ultrawideband toa-based ranging in indoor multipath environments. *Vehicular Technology, IEEE Transactions on 58,* 3 (march), 1046 –1058.

Arras, K., Mozos, O., and Burgard, W. 2007. Using boosted features for the detection of people in 2d range data. In *Proc. of the int. conf. on robotics & automation*.

Arulampalam, M., Maskell, S., Gordon, N., and Clapp, T. 2002. A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *Signal Processing, IEEE Transactions on 50,* 2 (feb.), 174 –188.

Aslam, J., Butler, Z., Constantin, F., Crespi, V., Cybenko, G., and Rus, D. 2003. Tracking a moving object with a binary sensor network. In *Proceedings of the 1st international conference on Embedded networked sensor systems*. ACM New York, NY, USA, 150–161.

Audette, W., Kynor, D., Wilbur, J., Gagne, J., and Peck, L. 2009. Improved Intruder Detection Using Seismic Sensors and Adaptive Noise Cancellation.

Bar-Shalom, Y. and Tse, E. 1975. Tracking in a cluttered environment with probabilistic data association. *Automatica 11,* 5, 451–460.

Barnich, O. and Van Droogenbroeck, M. 2009. Vibe: A powerful random technique to estimate the background in video sequences. *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, 945 –948.

Beauregard, S. and Haas, H. 2006. Pedestrian dead reckoning: A basis for personal positioning. In *Proceedings of the 3rd Workshop on Positioning, Navigation and Communication*. 27–35.

Bellotto, N. and Hu, H. 2007. Multisensor data fusion for joint people tracking and identification with a service robot. *Robotics and Biomimetics, 2007. ROBIO 2007. IEEE International Conference on*, 1494 –1499.

Bertozzi, M., Broggi, A., Caraffi, C., Del Rose, M., Felisa, M., and Vezzoni, G. 2007. Pedestrian detection by means of far-infrared stereo vision. *Computer vision and image understanding 106,* 2-3, 194–204.

Brooks, A. and Williams, S. 2003. Tracking people with networks of heterogeneous sensors. In *Proceedings of the Australasian Conference on Robotics and Automation*. Citeseer, 1–7.

Cambridge Consultants. Prism 200. `http://www.cambridgeconsultants.com/prism_200.html`.

Camero Tech. Xaver 800. `http://www.camero-tech.com/xaver800.shtml`.

Castro, P., Chiu, P., Kremenek, T., and Muntz, R. 2001. A probabilistic room location service for wireless networked environments. *Lecture Notes in Computer Science*, 18–34.

Cauwenberghs, G., Andreou, A., West, J., Stanacevic, M., Celik, A., Julian, P., Teixeira, T., Diehl, C., and Riddle, L. 2005. A miniature low-power intelligent sensor node for persistent acoustic surveillance. In *Proc. SPIE*. Vol. 5796. Citeseer, 294–305.

Chang, H. et al. 2008. Spinning beacons for precise indoor localization. In *Proceedings of the 6th ACM conference on Embedded network sensor systems*. ACM, 127–140.

Chang, S., Wolf, M., and Burdick, J. 2009. An mht algorithm for uwb radar-based multiple human target tracking. *Ultra-Wideband, 2009. ICUWB 2009. IEEE International Conference on*, 459 –463.

Chang, S. H., Sharan, R., Wolf, M., Mitsumoto, N., and Burdick, J. 2009. Uwb radar-based human target tracking. *Radar Conference, 2009 IEEE*.

Chen, Y. and Rui, Y. 2004. Real-time speaker tracking using particle filter sensor fusion. *Proceedings of the IEEE 92,* 3 (mar), 485 – 494.

Coetzee, S., Baker, C., and Griffiths, H. 2006. Narrow band high resolution radar imaging. *Radar, 2006 IEEE Conference on*.

Dalai, N., Triggs, B., Rhone-Alps, I., and Montbonnot, F. 2005. Histograms of oriented gradients for human detection. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005*. Vol. 1.

De Cubber, G. and Marton, G. 2009. Human Victim Detection. In *Third International Workshop on Robotics for risky interventions and Environmental Surveillance-Maintenance, RISE*.

Diermaier, J., Neyder, K., Werner, F., and Zagler, W. 2008. Distributed Accelerometers as a Main Component in Detecting Activities of Daily Living. In *Proceedings of ICCHP*. Springer.

Dorota, A., Yudan, Y., and Charles, K. 2002. Bridging GPS Gaps in Urban Canyons: The Benefits of ZUPTs. *Navigation Journal 48,* 4.

Enzweiler, M. and Gavrila, D. 2008. Monocular pedestrian detection: Survey and experiments. *IEEE transactions on pattern analysis and machine intelligence*, 2179–2195.

Ess, A., Leibe, B., Schindler, K., and Van Gool, L. 2009. Moving obstacle detection in highly dynamic scenes. In *Proceedings of the 2009 IEEE international conference on Robotics and Automation*. Institute of Electrical and Electronics Engineers Inc., The, 4451–4458.

Falconer, D., Ficklin, R., Konolige, K., Int, S., and Park, M. 2000. Robot-mounted through-wall radar for detecting, locating, and identifying building occupants. In *IEEE International Conference on Robotics and Automation, 2000. Proceedings. ICRA'00*. Vol. 2.

FLIR. Tau. `http://www.flir.com`.

Foxlin, E. 2005. Pedestrian tracking with shoe-mounted inertial sensors. *Computer Graphics and Applications, IEEE 25,* 6 (nov.-dec.), 38 –46.

Fu, Z. and Culurciello, E. 2008. A 1.2 mW CMOS temporal-difference image sensor for sensor networks. In *IEEE International Symposium on Circuits and Systems, 2008. ISCAS 2008*. 1064–1067.

Future-Shape. Sensfloor. `http://www.future-shape.com/sensfloor.html`.

Gatica-Perez, D., Lathoud, G., Odobez, J.-M., and McCowan, I. 2007. Audiovisual probabilistic tracking of multiple speakers in meetings. *Audio, Speech, and Language Processing, IEEE Transactions on 15,* 2 (feb.), 601 –616.

Geisheimer, J. L., III, E. F. G., and Marshall, W. S. 2002. High-resolution doppler model of the human gait. *Radar Sensor Technology and Data Visualization 4744,* 1, 8–18.

Germa, T., Lerasle, F., Ouadah, N., and Cadenat, V. 2010. Vision and rfid data fusion for tracking people in crowds by a mobile robot. *Computer Vision and Image Understanding*.

Gezici, S., Tian, Z., Biannakis, G. B., Kobayashi, H., Molisch, A. F., Poor, H. V., Sahinoglu, Z., Gezici, S., Tian, Z., Giannakis, G. B., Kobayashi, H., Molisch, A. F., Poor, H. V., and Sahinoglu, Z. 2005. Localization via ultra-wideband radios. *IEEE Signal Processing Magazine*.

Gürbüz, S., Melvin, W., and Williams, D. 2007. Detection and identification of human targets in radar data. In *Proceedings of SPIE*.

Harter, A., Hopper, A., Steggles, P., Ward, A., and Webster, P. 2002. The anatomy of a context-aware application. *Wireless Networks 8,* 2, 187–197.

Harville, M. and Li, D. 2004. Fast, integrated person tracking and activity recognition with plan-view templates from a single stereo camera. *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on 2*, 398–405.

Heimann Sensors. Htpa32x31. `http://www.heimannsensor.com/products_imaging.php`.

Henry, R., Matti, L., and Raimo, S. 2008. Human tracking using near field imaging. In *Pervasive Computing Technologies for Healthcare, 2008. PervasiveHealth 2008. Second International Conference on*. 148 –151.

Hobbs, P. 2001. A $10 thermal infrared imager. In *Proceedings of the SPIE*. Vol. 4563. 42–51.

Isard, M. and Blake, A. 1998. Condensation: conditional density propagation for visual tracking. *International journal of computer vision 29,* 1, 5–28.

Javed, O., Shafique, K., and Shah, M. 2002. A hierarchical approach to robust background subtraction using color and gradient information.

Jiménez, A., Seco, F., Prieto, C., and Guevara, J. 2009. A Comparison of Pedestrian Dead-Reckoning Algorithms using a Low-Cost MEMS IMU. In *6th IEEE International Symposium on Intelligent Signal Processing, 26-28 August, Budapest, Hungary*. 37–42.

Judd, T. 1997. A personal dead reckoning module. In *ION GPS*. Vol. 97. 1–5.

Kale, A., Cuntoor, N., Yegnanarayana, B., Rajagopalan, A., and Chellappa, R. 2003. Gait analysis for human identification. In *Audio-and Video-Based Biometric Person Authentication*. Springer, 1058–1058.

Kalgaonkar, K. and Raj, B. 2007. Acoustic Doppler sonar for gait recogination. In *Proceedings of the 2007 IEEE Conference on Advanced Video and Signal Based Surveillance-Volume 00*. IEEE Computer Society Washington, DC, USA, 27–32.

Kim, W., Mechitov, K., Choi, J., and Ham, S. 2005. On target tracking with binary proximity sensors. In *Proceedings of the 4th international symposium on Information processing in sensor networks*. IEEE Press, 40.

Kleinehagenbrock, M., Lang, S., Fritsch, J., Lomker, F., Fink, G., and Sagerer, G. 2002. Person tracking with a mobile robot based on multi-modal anchoring. In *Proc. IEEE Int. Workshop on Robot and Human Interactive Communication (ROMAN)*. Citeseer, 423–429.

Klingbeil, L. and Wark, T. 2008. A wireless sensor network for real-time indoor localisation and motion monitoring. In *IPSN '08: Proceedings of the 2008 International Conference on Information Processing in Sensor Networks (ipsn 2008)*. IEEE Computer Society, Washington, DC, USA, 39–50.

Korpipää, P., Koskinen, M., Peltola, J., Mäkelä, S., and Seppänen, T. 2003. Bayesian approach to sensor-based context awareness. *Personal and Ubiquitous Computing 7,* 2, 113–124.

Kourogi, M. and Kurata, T. 2003. Personal positioning based on walking locomotion analysis with self-contained sensors and a wearable camera. *Mixed and Augmented Reality, 2003. Proceedings. The Second IEEE and ACM International Symposium on*, 103–112.

Krumm, J., Williams, L., and Smith, G. 2002. SmartMoveX on a graph-an inexpensive active badge tracker. *Ubiquitous Computing*, 343–350.

Kusy, B., Ledeczi, A., and Koutsoukos, X. 2007. Tracking mobile nodes using rf doppler shifts. In *SenSys '07: Proceedings of the 5th international conference on Embedded networked sensor systems*. ACM, New York, NY, USA, 29–42.

Ladd, A., Bekris, K., Rudys, A., Kavraki, L., and Wallach, D. 2005. Robotics-based location sensing using wireless ethernet. *Wireless Networks 11,* 1, 189–204.

Li, L., Huang, W., Gu, I., and Tian, Q. 2003. Foreground object detection from videos containing complex background. In *Proceedings of the eleventh ACM international conference on Multimedia*. ACM, 10.

Lichtsteiner, P., Kramer, J., and Delbruck, T. 2004. Improved on/off temporally differentiating address-event imager. In *IEEE International Conference on Electronics, Circuits and Systems, ICECS*. Vol. 4. 211 – 214.

Lichtsteiner, P., Posch, C., and Delbruck, T. 2008. A 128× 128 120dB 15us latency asynchronous temporal contrast vision sensor. *IEEE Journal of Solid State Circuits 43,* 2, 566–576.

Lin, A. and Ling, H. 2006. Three-dimensional tracking of humans using very low-complexity radar. *Electronics Letters 42,* 18, 1062–1063.

Lin, A. and Ling, H. 2007. Doppler and direction-of-arrival (DDOA) radar for multiple-mover sensing. *IEEE Transactions on Aerospace and Electronic Systems 43,* 4, 1496–1509.

Lowe, D. G. 2004. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision 60,* 2, 91–110.

Lymberopoulos, D., Lindsey, Q., and Savvides, A. 2006. An Empirical Analysis of Radio Signal Strength Variability in IEEE 802.15.4 Networks using Monopole Antennas. In *Lecture Notes in Computer Science*. Vol. 3868. 326.

Mao, G., Fidan, B., and Anderson, B. 2007. Wireless sensor network localization techniques. *Computer Networks 51,* 10, 2529–2553.

Mesa Imaging. Swiss Ranger SR-4000. `http://www.mesa-imaging.ch/prodview4k.php`.

Middleton, L., Buss, A., Bazin, A., and Nixon, M. 2005. A floor sensor system for gait recognition. In *Proceedings of the Fourth IEEE Workshop on Automatic Identification Advanced Technologies*. Citeseer, 171–176.

Moore, D. and McCowan, I. 2003. Microphone array speech recognition: Experiments on overlapping speech in meetings. In *Proc. ICASSP*. Citeseer, 497–500.

Morishita, H., Fukui, R., and Sato, T. 2002. High resolution pressure sensor distributed floor for future human-robot symbiosis environments. In *IEEE/RSJ International Conference on Intelligent Robots and System, 2002*. Vol. 2.

Murakita, T., Ikeda, T., and Ishiguro, H. 2004. Human tracking using floor sensors based on the Markov chain Monte Carlo method. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*. Vol. 4.

NASIPURI, A. AND LI, K. 2002. A directionality based location discovery scheme for wireless sensor networks. In *Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications*. ACM, 111.

NI, L., LIU, Y., LAU, Y., AND PATIL, A. 2004. LANDMARC: indoor location sensing using active RFID. *Wireless Networks 10,* 6, 701–710.

OH, S. AND SASTRY, S. 2005. Tracking on a graph. In *Proceedings of the 4th international symposium on Information processing in sensor networks*. IEEE Press, 26.

OJEDA, L. AND BORENSTEIN, J. 2007. Personal dead-reckoning system for gps-denied environments. *Safety, Security and Rescue Robotics, 2007. SSRR 2007. IEEE International Workshop on*, 1–6.

ORR, R. AND ABOWD, G. 2000. The smart floor: A mechanism for natural user identification and tracking. In *Conference on Human Factors in Computing Systems*. ACM New York, NY, USA, 275–276.

OTSASON, V., VARSHAVSKY, A., LAMARCA, A., AND DE LARA, E. 2005. Accurate GSM indoor localization. *UbiComp 2005: Ubiquitous Computing*, 141–158.

PAKHOMOV, A., SICIGNANO, A., SANDY, M., AND GOLDBURT, T. 2003. Seismic footstep signal characterization. In *Proceedings of SPIE*. Vol. 5071. 297–304.

PEARCE, T., SCHIFFMAN, S., NAGLE, H., AND GARDNER, J. 2006. *Handbook of machine olfaction: electronic nose technology.* Wiley-Vch.

PENN, D., OBERZAUCHER, E., GRAMMER, K., FISCHER, G., SOINI, H., WIESLER, D., NOVOTNY, M., DIXON, S., XU, Y., AND BRERETON, R. 2007. Individual and gender fingerprints in human body odour. *Journal of The Royal Society Interface 4,* 13, 331.

PICARD, R. 2000. *Affective computing.* The MIT Press.

PINKSTON, R. 1994. A touch sensitive dance floor/MIDI controller. *The Journal of the Acoustical Society of America 96*, 3302.

POTAMITIS, I., CHEN, H., AND TREMOULIS, G. 2004. Tracking of multiple moving speakers with multiple microphone arrays. *IEEE Transactions on Speech and Audio Processing 12,* 5, 520–529.

PREMEBIDA, C., LUDWIG, O., MATSUURA, J., AND NUNES, U. 2009. Exploring sensor fusion schemes for pedestrian detection in urban scenarios. In *Proceeding of the IEEE/RSJ Worshop on: Safe Navigation in Open and Dynamic Environments, held at the IROS2009.*

PREMEBIDA, C., LUDWIG, O., AND NUNES, U. 2009. Exploiting LIDAR-based Features on Pedestrian Detection in Urban Scenarios. In *Intelligent Transportation Systems, ITSC. IEEE Int. Conference on.*

PRIYANTHA, N., CHAKRABORTY, A., AND BALAKRISHNAN, H. 2000. The cricket location-support system. In *Proceedings of the 6th annual international conference on Mobile computing and networking.* ACM New York, NY, USA, 32–43.

PRONOBIS, A., CAPUTO, B., JENSFELT, P., AND CHRISTENSEN, H. 2009. A realistic benchmark for visual indoor place recognition. *Robotics and Autonomous Systems*.

RAPPAPORT, T., REED, J., AND WOERNER, B. 1996. Position location using wireless communications on highways of the future. *Communications Magazine, IEEE 34,* 10 (oct), 33 –41.

REID, D. B. 1979. An algorithm for tracking multiple targets. In *IEEE Transactions on Automatic Control.* Vol. 24. 843–854.

RONG, P. AND SICHITIU, M. 2006. Angle of arrival localization for wireless sensor networks. *Sensor and Ad Hoc Communications and Networks, 2006. SECON'06. 2006 3rd Annual IEEE Communications Society on 1.*

ROOS, T., MYLLYMÄKI, P., TIRRI, H., MISIKANGAS, P., AND SIEVÄNEN, J. 2002. A probabilistic approach to WLAN user location estimation. *International Journal of Wireless Information Networks 9,* 3, 155–164.

ROTHER, C., KOLMOGOROV, V., AND BLAKE, A. 2004. Grabcut: Interactive foreground extraction using iterated graph cuts. In *ACM SIGGRAPH 2004 Papers.* ACM, 314.

ROWLEY, H., BALUJA, S., AND KANADE, T. 1996. Neural network-based face detection. In *1996 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1996. Proceedings CVPR'96.* 203–208.

Sarkar, S., Phillips, P., Liu, Z., Vega, I., Grother, P., and Bowyer, K. 2005. The humanID gait challenge problem: data sets, performance, and analysis. *IEEE transactions on pattern analysis and machine intelligence*, 162–177.

Savvides, A., Han, C., and Srivastava, M. B. 2001. Dynamic fine grained localization in ad-hoc sensor networks. In *Proceedings of the Fifth International Conference on Mobile Computing and Networking, Mobicom 2001, Rome, Italy.* pp. 166–179.

Scheutz, M., McRaven, J., and Cserey, G. 2004. Fast, reliable, adaptive, bimodal people tracking for indoor environments. In *IEEE/RSJ international conference on intelligent robots and systems (IROS)*. 1340–1352.

Schiff, J. and Goldberg, K. 2006. Automated intruder tracking using particle filtering and a network of binary motion sensors. In *IEEE International Conference on Automation Science and Engineering (CASE06), Shanghai, China*. Citeseer, 1–2.

Schindler, G., Brown, M., and Szeliski, R. 2007. City-scale location recognition. In *IEEE Conference on Computer Vision and Pattern Recognition, 2007. CVPR'07*. 1–7.

Schulz, D., Fox, D., and Hightower, J. 2003. People tracking with anonymous and id-sensors using rao-blackwellised particle filters. In *Proceedings of the International Joint Conference on Artificial Intelligence*.

Se, S., Lowe, D., and Little, J. 2005. Vision-based global localization and mapping for mobile robots. *IEEE Transactions on Robotics 21,* 3 (june), 364 – 375.

SenSource. Thermal imaging directional people counter. `http://www.sensourceinc.com/wired-people-counters.htm`.

Shankar, M., Burchett, J., Hao, Q., Guenther, B., Brady, D., et al. 2006. Human-tracking systems using pyroelectric infrared detectors. *Optical Engineering 45*, 106401.

Shrivastava, N., Madhow, R., and Suri, S. 2006. Target tracking with binary proximity sensors: fundamental limits, minimal descriptions, and algorithms. In *Proceedings of the 4th international conference on Embedded networked sensor systems*. ACM, 264.

Shu, C.-F., Hampapur, A., Lu, M., Brown, L., Connell, J., Senior, A., and Tian, Y. 2005. Ibm smart surveillance system (s3): a open and extensible framework for event based surveillance. *Advanced Video and Signal Based Surveillance, 2005. AVSS 2005. IEEE Conference on*, 318 – 323.

Smith, A., Balakrishnan, H., Goraczko, M., and Priyantha, N. 2004. Tracking moving devices with the cricket location system. In *Proceedings of the 2nd international conference on Mobile systems, applications, and services*. ACM, 190–202.

Smith, J., White, T., Dodge, C., Paradiso, J., Gershenfeld, N., and Allport, D. 1998. Electric Field Sensing For Graphical Interfaces. *IEEE Computer Graphics Applications 18,* 3, 54–60.

Snidaro, L., Micheloni, C., and Chiavedale, C. 2005. Video security for ambient intelligence. In *IEEE transactions on systems, man, and cybernetics*. Vol. 35. 133–144.

Srinivasan, A. and Wu, J. 2007. A survey on secure localization in wireless sensor networks. *Encyclopedia of Wireless and Mobile Communications*.

Stogdale, N., Hollock, S., Johnson, N., and Sumpter, N. 2003. Array-based infra-red detection: an enabling technology for people counting, sensing, tracking, and intelligent detection. In *Proceedings of SPIE*. Vol. 5071. 465.

Tan, X., Chen, S., Zhou, Z., and Zhang, F. 2006. Face recognition from a single image per person: A survey. *Pattern Recognition 39,* 9, 1725–1745.

Tao, D., Li, X., Wu, X., and Maybank, S. 2007. General Tensor Discriminant Analysis and Gabor Features for Gait Recognition. *IEEE Transactions On Pattern Analysis and Machine Intelligence*, 1700–1715.

Teixeira, T., Jung, D., Dublon, G., and Savvides, A. 2009a. Identifying people by gait-matching using cameras and wearable accelerometers. In *ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC)*.

Teixeira, T., Jung, D., Dublon, G., and Savvides, A. 2009b. Identifying people in camera networks using wearable accelerometers. In *Pervasive Technologies Related to Assistive Environments (PETRA)*.

TEIXEIRA, T., JUNG, D., AND SAVVIDES, A. 2010. Tasking Networked CCTV Cameras and Mobile Phones to Identify and Localize Multiple People.

TEIXEIRA, T. AND SAVVIDES, A. 2008. Lightweight people counting and localizing for easily-deployable indoors WSNs. In *IEEE Journal of Selected Topics in Signal Processing*.

TIME DOMAIN. PLUS Precision-Location UWB System. `http://www.timedomain.com/plus.php`.

TIME DOMAIN. Pulson 220 Tagless tracking. `http://www.timedomain.com/pulson.php`.

TURK, M. AND PENTLAND, A. 1991. Face recognition using eigenfaces. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*. Vol. 591. 586–591.

UBISENSE. Series 7000 compact tag. `http://www.ubisense.net/en/resources/factsheets/series-7000-compact-tag.html`.

VALTONEN, M., MÄENTAUSTA, J., AND VANHALA, J. 2009. TileTrack: Capacitive human tracking using floor tiles. In *Proceedings of the 2009 IEEE International Conference on Pervasive Computing and Communications-Volume 00*. IEEE Computer Society, 1–10.

VARSHAVSKY, A., CHEN, M., DE LARA, E., FROEHLICH, J., HAEHNEL, D., HIGHTOWER, J., LAMARCA, A., POTTER, F., SOHN, T., TANG, K., AND SMITH, I. 2006. Are GSM phones THE solution for localization? In *Proceedings of the 7th IEEE Workshop on Mobile Computing Systems and Applications, 2006*. 34–42.

VIOLA, P. AND JONES, M. 2002. Robust real-time object detection. *International Journal of Computer Vision 57,* 2, 137–154.

WANG, L., TAN, T., NING, H., AND HU, W. 2003. Silhouette analysis-based gait recognition for human identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence 25,* 12, 1505–1518.

WICKS, M., HIMED, B., BRACKEN, J., BASCOM, H., AND CLANCY, J. 2005. Ultra narrow band adaptive tomographic radar. *Computational Advances in Multi-Sensor Adaptive Processing, 2005 1st IEEE International Workshop on*, 36 – 39.

WILSON, J. AND PATWARI, N. 2009. Radio tomographic imaging with wireless networks. *IEEE Trans. Mobile Computing*.

WISKOTT, L., FELLOUS, J., KRÜGER, N., AND VON DER MALSBURG, C. 1997. Face Recognition by Elastic Bunch Graph Matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence 19,* 7, 775.

XIANGQIAN, L., GANG, Z., AND XIAOLI, M. 2008. Target localization and tracking in noisy binary sensor networks with known spatial topology. *Wireless Communications and Mobile Computing*.

YAROVOY, A., LIGTHART, L., MATUZAS, J., AND LEVITAS, B. 2006. UWB radar for human being detection. *IEEE Aerospace and Electronic Systems Magazine 21,* 3, 10–14.

YILMAZ, A., JAVED, O., AND SHAH, M. 2006. Object tracking: A survey. *ACM Computing Surveys (CSUR) 38,* 4, 13.

YINON, J. 2003. Peer Reviewed: Detection of Explosives by Electronic Noses. *Analytical Chemistry 75,* 5, 98–105.

YOUSSEF, M. AND AGRAWALA, A. 2008. The Horus location determination system. *Wireless Networks 14,* 3, 357–374.

ZATSIORSKY, V. M. 1997. *Kinematics of Human Motion.* Human Kinetics Publishers.

ZETIK, R., CRABBE, S., KRAJNAK, J., PEYERL, P., SACHS, J., AND THOMÄ, R. 2006. Detection and localization of persons behind obstacles using M-sequence through-the-wall radar. In *Proceedings of the SPIE*. Vol. 6201.

ZHANG, Z. AND ANDREOU, A. 2008. Human identification experiments using acoustic micro-doppler signatures. *Micro-Nanoelectronics, Technology and Applications, 2008. EAMTA 2008. Argentine School of*, 81–86.

ZHOU, Q., LIU, J., HOST-MADSEN, A., BORIC-LUBECKE, O., AND LUBECKE, V. 2006. Detection of multiple heartbeats using Doppler radar. In *2006 IEEE International Conference on Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings*. Vol. 2.

ZLOT, R. AND BOSSE, M. 2009. Place Recognition Using Keypoint Similarities in 2D Lidar Maps. In *Experimental Robotics: The Eleventh International Symposium*. Springer Verlag, 363.