Chapter 2
Sensors for Human Behavior Analysis

This chapter proposes a brief description of the most common sensors used in literature to perform the automatic analysis of the human behavior. For each kind of sensor, at least a scientific work using it for human behavior analysis is presented. This chapter helps the readers to understand because the author has focused his attention on the analysis of human behavior using video streaming.

2.1 Motivation

Automatic human behavior analysis and recognition are complex tasks that have attracted a lot of researchers in the latest years. One of the primary tasks to perform implementing such analysis is to define a representation of the real world using the data sampled by some sensors.

Nowadays the most frequently used sensors are camera devices, but in literature there are also approaches based on the employment of other kinds of sensor. These approaches achieve good results in some specific domains but often they cannot be generalized to other contexts. In literature, there are many works applying data fusion techniques to build complex systems based on using different type of sensors [20–22].

In this chapter a brief overview about the following sensors and of the works using them is presented: Radio Frequencies Identifier (RFID), pressure sensors, Micro-Electro-Mechanical Systems (MEMS) and image sensors.

The aim of this brief overview is to show the most important used sensors in human behavior analysis, their applications and their limits. This discussion should help the readers to understand because the author has focused his attention on the analysis of human behavior using video streaming. The main goal of the book is the semantic analysis of video streaming. From this point of view, since the sensors play a secondary role in this book, a critical overview of the literature about sensors is beyond the scope of this book.

A. Amato et al., *Semantic Analysis and Understanding of Human Behavior in Video Streaming*, DOI: 10.1007/978-1-4614-5486-1_2,

2.2 Radio Frequencies Identifier Technology

RFID is the acronym of Radio Frequencies Identifier. This technology is based on four key elements: the RFID tags themselves, the RFID readers, the antennas and choice of radio characteristics, and the computer network (if any) that is used to connect the readers.

RFID tags are devices composed of an antenna and a small silicon chip containing a radio receiver, a radio modulator for sending a response back to the reader, control logic, some amount of memory, and a power system. A RFID tag transmits the data stored inside its memory module when it is exposed to radio waves of the correct frequency sent by the reader.

According to the used power system there are two kinds of tags: passive tags where the power system can be completely powered by the incoming RF signal and active tags where the tag's power system has a battery. Passive tags are cheaper than active tags but they have a shorter range of action (the reader must be positioned very close to the tag).

Figure 2.1 shows a diagram of the power system for a passive inductively coupled transponder-RFID tag. An inductively coupled transponder comprises an electronic data-carrying device, usually a single microchip, and a large area coil that functions as an antenna. The reader's antenna coil generates a strong, high frequency electromagnetic field, which penetrates the cross-section of the coil area and the area around the coil. The antenna coil of the transponder and the capacitor C_1 form a resonant circuit tuned to the transmission frequency of the reader. The voltage U at the transponder coil reaches a maximum due to resonance step-up in the parallel resonant circuit. The layout of the two coils can also be interpreted as a transformer (transformer coupling), in which case there is only a very weak coupling between the two windings.

The simplest RFID chips contain only a serial number. This serial number is written into the chip by the manufacturer but there are also tags where this code can be written by the end user. An example of this code is the EPC (Electronic Product Code) that is a number composed of 96 bits. This number is the kernel of an international standard making RFID technology a pillar element of the international logistic chain. This standard is under the oversight of EPCglobal Inc™ a not-for-profit joint venture between GS1 (formerly EAN International) and GS1 US (formerly the Uniform Code Council).

More sophisticated RFID chips can contain read-write memory that can be programmed by a reader

Among the various applications where RFID technology is used, there is also the human activity detection and monitoring, an example of such applications is [23]. In this work the authors propose a system to detect the activity of daily living performed at home. This is a challenging task for various reasons such as: each one can perform the same action in different ways, there are many possible activities that a system should model with minimum human effort, etc. The key observation done by the authors is that the sequence of objects that a person uses performing a

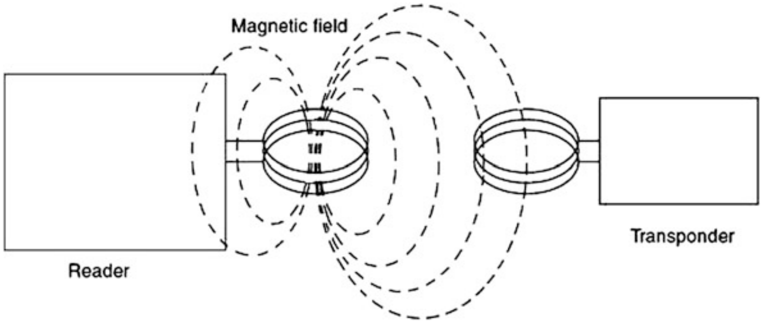

Fig. 2.1 A block diagram of a RFID tag

given activity is a good marker both of the action type and quality. Starting from this observation they propose a system composed of three modules: specialized sensors to detect object interactions, a probabilistic engine that infers activities given observations from sensors, and a model creator to create probabilistic models of activities. In this work the used sensors are RFID tags attached on each object of interest. The RFID reader is built inside a glove that the user should wear while performing his activities of daily living. The system was tested on 14 predefined activities and the obtained results showed good performance both in terms of precision (88 %) and recall (73 %).

Despite these good results the system presents various limits. From a technological point of view, water and metal absorb the radio waves that most RFID tags use; metal can also short-circuit the tag antenna. This fact limits the number of correctly observable actions. But the most important limitation of this approach is that it is too invasive. Indeed it requires that the user must wear a glove. This can be a serious problem due to the fact that many people are not too attracted by using gloves while performing activities of daily living.

2.3 Pressure Sensors

Pressure transducers are very common and cheap. They are used in various applications and they work using various principles (variation of capacity, variation of resistance, piezoelectric, etc.).

In [24] the authors propose a human behavior recognition system using a set of pressure sensors based on the variation of resistance. For this kind of sensors the conversion of pressure into an electrical signal is achieved by the physical deformation of strain gages which are bonded into the diaphragm of the pressure transducer and wired into a Wheatstone bridge configuration. Pressure applied to the pressure transducer produces a deflection of the diaphragm which introduces strain to the gages. The strain will produce an electrical resistance change proportional to the pressure.

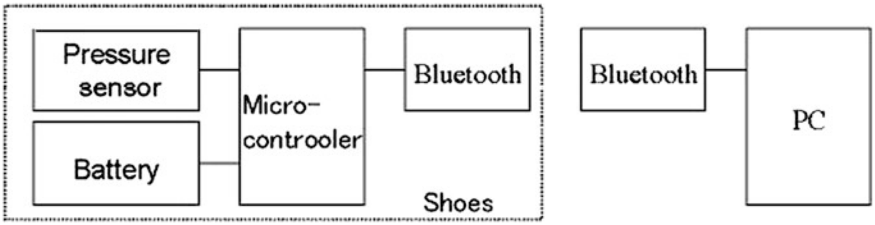

Fig. 2.2 A block diagram of the pressure sensor proposed in [24]. (reproduced by permission of IEEE)

Figure 2.2 shows a block diagram of the whole pressure measurement system proposed in [24]. The transducer is connected to a microcontroller that is able to communicate with a standard PC using a Bluetooth link. A serious constrain for this system is its power supply module that is composed of a 9 V alkaline battery. This fact introduces the well known limitations due to the lifecycle of the battery, maintenance, etc.

The main idea at the base of this work is that measuring the plantar pressure distribution it is possible to infer information about the actions performed by the user. Four sensors were installed in each shoe. According to the authors, it is possible to classify fifteen different behaviors: walking (slow, normal, fast), running (slow, normal, fast) standing (leaning forward, load on tiptoe, upright), leaning standing to one foot (leaning forward, normal), sitting (bending forward, normal), floating, and no wearing. The parameters used to classify the various actions are pressure values and length of time where a given pressure is measured. These parameters are compared to a fixed set of thresholds to identify the various actions.

According to the authors, the system achieves good classification rate (about 90 % of successfully classifications) but the experiment settings are not well described. The main limits of this approach are the necessity of a calibration stage for each person to define the various thresholds and its invasivity (indeed the measuring system is installed in the shoes and it is quite visible). These considerations make the system not suitable for applications in everyday life situations.

2.4 Micro Electro-Mechanical Systems Sensors

Micro-Electro-Mechanical Systems (MEMS) are devices built by means of the integration of mechanical elements, sensors, actuators, and electronics on a common silicon substrate through micro-fabrication technology. The current leaders in commercially successful MEMS technology are accelerometers. These devices are used in a large number of applications such as: automotive industry, inertial navigation systems, cellular phones, etc.

The physical mechanisms underlying MEMS accelerometers include capacitive, electromagnetic, piezoelectric, ferroelectric, optical, etc. The most successful types are based on capacitive transduction; the reasons are the simplicity of the sensor element itself, no requirement for exotic materials, low power consumption, and good stability over temperature.

These sensors are used in many applications involving the so-called "wearable-sensors". For example, in [25] the authors propose a system to classify the human pose in "sitting", "standing" and "walking" using a bi-axial accelerometer attached to the user's thigh. The sensor is positioned in order to measure the gravity acceleration using the Y-axis when the user is in "standing" position and the X-axis when she/he is in "sitting" position. The "walking" action is detected when the variance of acceleration is greater than a predefined threshold.

Reference [26] presents a method to detect physical activities from data acquired using five small biaxial accelerometers worn simultaneously on different parts of the body. In particular the sensors are placed on each subject's right hip, dominant wrist, non-dominant upper arm, dominant ankle, and non-dominant thigh to recognize ambulation, posture, and other everyday activities.

The experiments carried-out aim at identifying twenty different actions of everyday life (see [26] for further details). The system was tested on twenty subjects from the academic community volunteered. Data was collected from 13 males and 7 females. Data were classified using various methods but decision tree classifiers showed the best performance recognizing everyday activities with an overall accuracy rate of 84 %. Interestingly, the obtained results show that some activities are recognized well with subject-independent training data while others appear to require subject-specific training data.

2.5 Image Sensors

This kind of sensors are used in all the cameras and camcorders used to create still images and video streaming. The kernel of this kind of sensors consists of an array of tiny pixels (Picture Elements). Sensor pixels are composed of photodiodes.

During the imaging process, the light starts to fall on photodiodes, and they convert photons into electric charge. Photodiodes are not sensible to color, so digital cameras use different color filters to transmit light through. The most common ones are filters for three basic colors: red, green and blue. So, the camera is able to calculate the number of photons of three basic colors that fell on each photodiode while the camera shutter was open. To calculate all color components around each photodiode, red, green and blue filters should be situated adjacently. This makes it possible to convert raw image data into a full-color image in RGB (Red Green Blue) space.

The most common type of color filter array is called a "Bayer array". This filter gives priority to the green mimicking the behavior of human eyes. A schematic overview of the imaging process is shown in Fig. 2.3. The light is filtered by an infra red filter (Fig. 2.3). The filtered light goes through a Bayer array filter and finally hits the pixels of the sensor.

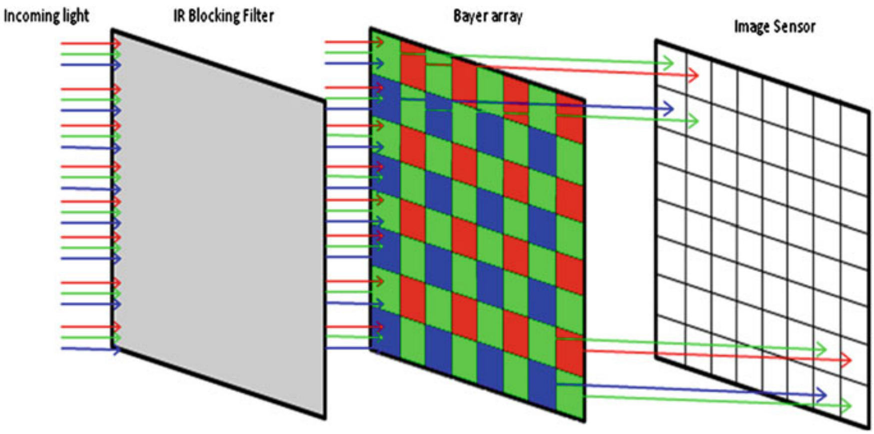

Fig. 2.3 A schematic view of an image sensor. **a**—IR-Blocking Filter, **b**—Color Filters, **c**—Color blind sensors, **d**—The image sensor composed of millions of light sensors

Nowadays, the image sensors are built using two technologies: CCD (Charge-Coupled Device) and CMOS (Complementary Metal–Oxide–Semiconductor). Both the technologies use photodiodes to convert light in electrons, but they use two different methods to read these values from each pixel of the sensor. In a CCD device, the charge is actually transported across the chip and read at one corner of the array. A special manufacturing process is used to create the ability of transporting charges across the chip without distortion. This process leads to very high-quality sensors in terms of fidelity and light sensitivity. In most CMOS devices, there are several transistors at each pixel that amplify and move the charge using more traditional wires. The CMOS approach is more flexible because each pixel can be read individually. CCD sensors give better images than CMOS sensors but the latter are cheaper than the former. Furthermore, in the last years the quality difference between the images sampled by these family of sensors are becoming smaller indeed, ever more often CMOS sensors are used in good quality cameras.

Most of the works in literature on human behavior analyze video streaming sampled using these image sensors. The next chapter proposes a review of such works.

2.6 Summary

This chapter has presented a brief and not exhaustive overview of the sensors used in some applications of human behavior analysis. Some sensors have been omitted because they are often used in conjunction with video analysis by means of data fusion techniques. Two relevant examples of such sensors are: audio and multi-spectral sensors.

Audio sensors are becoming ever more interesting due to progress in speech recognition. In literature it is possible to find their applications for the recognition of specific human behaviors. For example, in [27] a system to detect aggressions in trains is proposed while in [28] a method for action detection in action movies is described. Furthermore, audio-vision data are used to detect human emotions as shown in [29] where the system is able to detect 4 cognitive states (interest, boredom, frustration and puzzlement) and 7 prototypical emotions (neural, happiness, sadness, anger, disgust, fear and surprise).

Multi-spectral sensors are used in many remote sensing applications. In human behavior analysis there is a strong interest to infrared sensors because they can operate in total darkness allowing for the person detection during the night. In [30] a data fusion approach working on infrared images and classical CCD camera images is presented.

The sensors presented in this chapter are used in systems called wearable sensors. These kind of systems are intrusive (because the user are required to wear the sensors) and so they can be applied only in some restricted applications (such as human–machine interface applications).

Multi-spectral sensors can give excellent results but they are too expensive to be used in real world applications.

Audio sensors are cheap but they are not able to give good results in terms of human activity detection without using some data fusion techniques with video streaming. But these systems have a high computational cost due to the complexity of the used algorithm.

For these reasons, in this book, only approaches based on video streaming analysis will be considered.

http://www.springer.com/978-1-4614-5485-4

Semantic Analysis and Understanding of Human
Behavior in Video Streaming
Amato, A.; Di Lecce, V.; Piuri, V.
2013, XII, 108 p., Hardcover
ISBN: 978-1-4614-5485-4