MASTER

Distributed Human Posture Recognition using Thermopile Array Sensors

Schipper, J.M.

Award date:
2019

Technische Universiteit
Eindhoven
University of Technology

Department of Mathematics and Computer Science
System Architecture and Networking Research Group

Distributed Human Posture Recognition using Thermopile Array Sensors

Master thesis

Jeroen Schipper

Supervisors:
dr. Tanir Ozcelebi
dr. Qingzhi Liu
dr. Nicola Zannone

version 1.1

Eindhoven, November 2019

Abstract

Human posture recognition is an attractive and challenging topic with a wide range of applications. Current solutions rely on high-resolution cameras, depth cameras, or body sensor network to provide accurate posture recognition. The main drawback of cameras is the privacy concerns related to their use and body sensor networks are not suited for long term use due to the high number of sensors used. Emerging technologies for privacy-preserving posture recognition show promising results.

One of these technologies is the thermopile array sensor that provides low-resolution thermal images. The goal of this research was to develop a privacy-preserving room-wide posture recognition system using thermopile array sensors. The challenge with using low-resolution thermopile array sensors is that as the distance to the sensor increases, the noisier and less detailed the images become.

Three methods of using localization for posture recognition were explored to improve the accuracy of the system. The localization uses the positions and angles of arrival of the sensors for its estimations. The thermal images are used to determine the angle of arrival. Eight postures were used to compare the methods. The proposed system achieved an accuracy of 93.3%, which shows that privacy-preserving room-wide posture recognition can achieve high accuracy using thermopile array sensors.

Acknowledgements

To start, I would like to express my gratitude to my supervisor Dr. Tanir Ozcelebi for his encouragement and support during my master project and for helping me find a project that suited my interests. I would like to thank my daily supervisor Dr. Qingzhi Liu for his insights and continuous support. Next, I would like to thank Dr. Nicola Zannone for being on my assessment committee. Finally, I want to thank my parents and friend for the support and encouragements during my master.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Human posture recognition is an attractive and challenging topic with a wide range of applications in the areas of personal health care, human-computer interaction (HCI), environmental awareness, and surveillance systems. The application of human posture recognition in assisted living is gaining increased interest as the number of elderly is continuing to grow with the estimate that globally there will be 2.1 billion elderly in 2050 [27]. Coinciding with the increase in elderly people, there is a strong desire from the elderly to continue living independently at home. It is known that the elderly have a higher risk of falling and sustaining an injury due to falling. The risk of falling increase with their age and so does the mortality rate of fall-related injuries [21]. Indoor posture recognition systems can be used to monitor the activities of the elderly and detect accidental falls.

Another application for human posture recognition is smart buildings that are becoming more prevalent in recent years with the advancement of Internet of Things (IoT) and the integration of building automation systems. One of the challenges in smart buildings is adapting room conditions, such as lighting, heating, ventilation and air condition, to the preference of the users and the activities that take place within the room. Human posture recognition is the basis of human action recognition that offers methods to determine the activities within a given area.

There are generally two approaches when it comes to posture recognition systems. The camera-based approach uses the images captured by different types of cameras to determine the posture, while the sensor-based approach uses inertial-based sensors as a basis for posture recognition. Both approaches are briefly introduced in the next section.

1.1 Existing solutions

Most early human posture recognition solutions use RGB cameras to extract the silhouette of the human body from the captured images [9, 11, 12]. Cameras have the problem of being perceived as privacy-invasive due to their size and purpose among other disadvantages, such as high computational cost, and sensitive to lighting conditions. With the introduction of cost-effective depth cameras, such as the Microsoft Kinect, many researchers have used these sensors in their posture recognition systems [1, 4, 23, 24, 28] as they are able to provide real-time 3D data and insensitive to lighting changes [3]. The Kinect gained more popularity as it became able to create human skeletons constructed by the estimation of 3D joint positions from its depth sensor [25]. Recently, time-of-flight (TOF) sensors have been used for posture recognition [5]. These sensors use a different approach from depth cameras to construct a depth image of its surroundings at high frame rates.

Wearable sensors or inertial sensors are used in inertial-based human posture recognition systems. A combination of accelerometers and gyroscopes are used to form a wireless body area network that is worn by the user [13, 15, 29]. Wearable sensors have been used to solve some of the problems of the camera systems as they are not sensitive to lighting conditions, can freely move with the user, and are not considered privacy invasive, but they have their own problems with sensor drift over time and require the users to correctly place and maintain the sensors making them not suitable for long term applications.

1.2 Proposed solution

In the last couple of years, thermopile array sensors have come down in price making them a suitable alternative for the above-mentioned sensors for the applications of presence detection, fall detection, and posture recognition systems [2, 7, 10, 16, 26]. These sensors measure the temperature in the area by using a number of thermopiles that detect an object's infrared energy. These sensors are ideal for embedded applications in assisted living and smart buildings due to their small size, low energy consumption, and ability to detect stationary objects. The privacy of the users is preserved by the low resolution of these sensors compared to cameras.

The main goal of this research is to develop a privacy-preserving room-wide human posture recognition system using low-resolution thermopile array sensors and distributed neural networks. The whole system forms a wireless sensor network. The posture data is captured by the sensors at different angles and this information is processed to be used in the posture recognition. A localization algorithm is used to estimate the location of the person based on the positions of the sensors and the posture data. This location is used to perform the recognition locally if the person is close to a sensor. Otherwise, global posture recognition is used to recognize the posture. The localization is also used to deal with the fuzzy images that are captured when the person stands further away from the sensors.

1.3 Report layout

The remainder of this report is organised as follows. Chapter 2 describes the shortcomings of the current solutions, intrduces the goal of this research and the challanges. The related work is presented in Chapter 3. In Chapter 4, the proposed solution is explained. The experiments that were performed and their results are described in Chapter 5. Chapter 6 consists of the conclusions.

Chapter 2

Problem statement

2.1 Research motivation

Human posture recognition is the basis for human action recognition and both have a wide range of applications in multiple areas. Improving posture recognition will have the potential to enhance the functionalities of these systems in all application areas, such as assisted living [7, 24, 10], smart buildings [2] and action recognition [1, 4, 11].

As mentioned in Chapter 1, the majority of the posture recognition system are camera-based due to their successful results, ease of use, and availability. While camera-based solutions continue to improve, they are becoming a less appealing option for certain applications due to privacy concerns. Cameras are relatively large and their shapes are recognisable making them be considered privacy invasive. The high resolutions of cameras are another reasons for the privacy concerns as their images contain detailed information about the people that are in the frame. Most research on camera-based solutions focuses on single subject posture recognition within a limited area, often a few meters from the centre of the FOV of the camera(s) used [3]. Many posture recognition applications would benefit from room-wide recognition (e.g. assisted living, smart buildings, etc.).

Sensor-based solutions are not considered privacy-invasive. One of the benefits of sensor-based solutions is that they are not limited to a given area by using wearable sensors; however, their main drawback is the high number of sensors used [29, 15, 13]. These sensors have to be positioned correctly and maintained, which is often the responsibility of the user. This makes sensor-based solutions ill-suited for long term use.

Thermopile array sensors are an alternative for posture recognition solutions as they solve some of the issues of the other sensors and enable new application scenarios. The privacy concerns with camera-based solutions is not an issue with thermopile array sensors due to their small size and low resolution. An example of a thermal image captured by these sensors is shown in Figure 2.1. This figure shows a person that spread both arms out wide. Due to the low resolution, no discernible details of the person, such as gender and age, can be extracted from this image.

While thermopile posture recognition solutions show promising solutions there are some limitations with these sensors:

- Using the thermal images from these sensors makes posture recognition more challenging.
- Postures with slight differences are difficult the distinguish due to the low resolution of the thermal images.
- The thermal images have a higher noise ratio compared to cameras.
- Details in the thermal images are lost at smaller distances compared to cameras.

Some of the limitations can be overcome by using multiple sensors to provide additional details.

The solution proposed by Burns et al. [2] has trouble recognizing certain postures, which is likely due to the sensor configuration as it only uses two sensors on perpendicular axes. Another

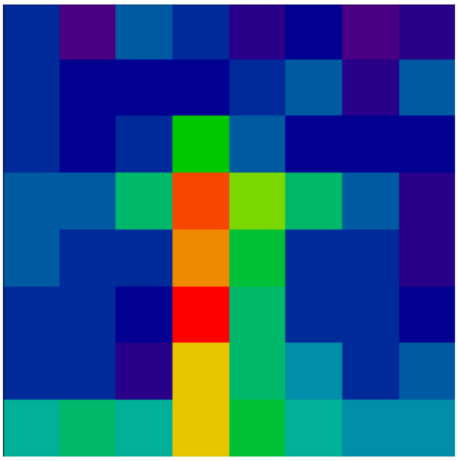


Figure 2.1: Example of a thermal image captured using a thermopile array sensor

limitation resulting from the low resolution is that as the distance to the sensor increases, it becomes more challenging to distinguish the posture. Gochoo et al. [7] proposed a solution using three sensors at different axes with promising results at a distance of 1.5 meters from the sensors, where the postures are still easily recognizable.

The aim of this research is to develop a proof of concept for a privacy-preserving room-wide posture recognition system. The system uses thermopile array sensors to capture the posture data and performs the recognition in a distributed way. Interest for this research was initially sparked from the results of the proposed system of Gochoo et al. [7], which uses multiple thermopile array sensors to perform posture recognition at a fixed position. What this concept misses is the ability to recognize postures at larger distances such that it can be used for room wide recognition. An initial test shows that posture recognition can be performed at larger distances using a single low resolution thermopile array sensor on simple but distinct postures. What the test also shows is that the accuracy of the recognition drops as postures are captured at increasing distances. The details of this test are described in section 5.3.1.

The main research question that is answered in this research is formulated as follows.

RQ. How to improve the accuracy of a room-wide posture recognition system using thermopile array sensors?

2.2 Research challenges

Existing solutions have shown that posture recognition using thermopile array sensors is possible with high accuracy when the recognition is performed at a fixed position. The goal of the research is to recognize the postures of people at any position in the area. At larger distances, the noise in the thermal images increases while the level of details decreases. Posture recognition of more complex postures becomes challenging at these distances. To achieve these this goal, using only a single thermopile array sensor is not enough, because of its low resolution. Therefore, a distributed posture recognition system is proposed.

Distributed posture recognition

Yang et al. [29] explored offloading some of the computation by having the sensors determine whether the data should be sent to the central computing unit, reducing the workload of the central unit and the used bandwidth. The research of Bi et al. [1] shows that distributed posture recognition is possible using the PCANN architecture; however, they only distributed the image from a single camera over multiple nodes in the WSN. Before the distributed posture recognition can be performed, the images have to be obtained by the classification device. This architecture reduces the workload and the memory consumption of the sensor nodes, but increases the bandwidth used by the system as the images are transmitted twice over the network. The goal of this research is to perform posture recognition on images produces by multiple sensors locally. By performing the local posture recognition on the sensors reduces the used bandwidth, and the workload and memory consumption on the central unit. Through this reduction, it is possible to deploy the central unit on an embedded platform, making remote deployments easier.

Long distance recognition

The radiation received by the thermopile array sensor is reduced at greater distances. Combined with the fact that the individual thermopile elements cover a larger area at these distances results in fuzzy images. Posture recognition on these fuzzy images is challenging as it is difficult to distinguish between the different postures. The theory is that location data can be used to improve the recognition. This additional data could be extracted from the thermal information of the thermopile array sensors or from additional sensors added to the system. Another option is to pre-process the obtained images to remove some of the noise from the image.

Chapter 3

Related work

3.1 Camera-based solutions

3.1.1 Regular camera-based solutions

Kyaw K. Htike et al. [11] proposed a real-time human posture recognition system for video surveillance using a single static camera to detect four postures (*standing, sitting, lying down*, and *crawling*). The video feed from the camera was sampled at 6 frames per second (fps) as many frames at higher frame rates are redundant or very similar. In the data pre-processing, the body silhouette was extracted using several algorithms such as background extraction, Otsu thresholding, and median filtering were used. Four classifiers (K-means, Fuzzy C-means, Multi-Layer Perceptron (MLP) Neural Network, and Self Organizing Maps) were evaluated and compared. The MLP neural network achieved the highest accuracy of 96% while the other classifiers performed worse with K-means, Fuzzy C-means, and Self Organizing Maps achieving accuracies of 31%, 33%, and 86% respectively.

Ninghang Hu et al. [12] describe a posture recognition system using a top view camera that is able to recognize six postures (*standing, sitting, bending, pointing, stretching*, and *walking*). The posture recognition is performed using a Posture Descriptor that assigns a matching score to an image for each posture category that it was trained with. These scores are considered the feature vector for the standard Support Vector Machine (SVM). The average accuracy of the system is 79.75%, which is an improvement of 23% over similar solutions.

Kamal Sehairi et al. [24] propose an elderly fall detection system based on multiple shape features and motion analysis. Background subtraction is used to determine the silhouette from which the head position is estimated. The vertical velocity of the head was calculated using a finite state machine that determined the direction of movement. This velocity and other features extracted from the camera feed was used in three classifiers (Radial basis function SVM, K-Nearest Neighbours, and Backpropagation Neural Network) to detect whether the subject has fallen. The highest accuracy of 99.61% was achieved with BPNN and a global error rate of 1%.

Rafik Gouiaa [9] developed a human posture recognition system by combining a single camera with four IR lights. The camera and the lights were mounted on the ceiling with the lights in the corners. The lights cast an IR shadow on the ground in succession that is used in combination with the silhouette, extracted through background subtraction, for the recognition. The silhouette and cast shadow are than rescaled and a Euclidean distance transform is applied to obtain the feature vector. A weighted KNN was used for the classification for each cast shadow and a weighted majority vote scheme determines the final posture prediction of the nine postures. The average accuracy of the system was 94.22% when picking K=3.

The main issue with these works is that they require extensive pre-processing algorithms, which require a lot of computational power, to perform posture recognition. Privacy is also a concern with these solutions as they capture identifiable images of everyone in view.

3.1.2 Kinect-based solutions

Chen Chen et al. [4] propose a real-time human action recognition system based on depth motion maps (DMMs). The DMMs are created using the depth images of a Kinect over a period of time to capture the motion energy from an action. The system uses the DMMs from the x, y, and z-axis concatenated and normalized as the feature vector for the l_2-regularized collaborative representation classifier. The average accuracy of the classifier was 90.5%.

Orasa Patsadu et al. [23] compared four classification methods: BPNN, SVM, decision tree, and naive Bayes on their performance in human posture recognition. The classification methods were tested on three postures (*standing, sitting*, and *lying down*) with 1200 samples for each of the postures that consisted of the 3D body joint positions obtained by the Kinect SDK. Of the four tested classification methods, BPNN has the highest accuracy of 100%.

Wen-June Wang et al. [28] proposed a human posture recognition system based on the depth images from the Kinect sensor. The system is designed to recognize five postures: sitting, standing, stooping, kneeling, and lying. The system is divided into four steps. First, the system extracted the silhouette from the depth image. Secondly, A horizontal projection is made of the silhouette to determine if the subject is in the kneeling posture. Thirdly, a star skeleton is created by determining the upper and lower body ratios, and the centre of gravity. A learning vector quantization (LVQ) neural network is used for the classification of the sitting, stooping, and lying postures. In the final step, a vertical projection of the silhouette is created and used with the horizontal projection to distinguish between the standing and non-forward sitting posture. The average accuracy of this system is higher than 99% and is able to achieve this accuracy while recognizing the postures at multiple angles from the Kinect sensor.

Tianyu Bi et al. [1] have proposed a distributed ANN architecture, Parallel Channel Artificial Neural Network (PCANN), for image recognition on resource-constrained IoT devices. They showcased a system that used the depth images for a Kinect camera for human posture and action recognition. The system separates a single machine learning model into several small connected modules that were deployed on the IoT devices. The system was used to classify four postures (*right arm raised, left arm raised, one person standing*, and *two persons standing*) and five actions (*moving left, moving right, waving up, waving down*, and *standing*) and the data-set consisted of 1800 samples for posture recognition and 1000 samples for action recognition collected from a large number of volunteers, both male and female. The depth images were pre-processed to extract the area of the human from the depth images after which an optional processed would use projection-based feature extraction to reduce the size of the feature set. These feature-sets were then split across the different modules that were part of the input layer. The output of each module is sent to the next module until the final module, that houses the output layer of the mode, receives all its inputs and performs the final classification. The system has an average accuracy of 93.83% for both the human posture and human action recognition.

The recognized postures of these systems are relatively simple compared to the other related works. The Kinect camera is a large piece of equipment that needs to be connected, by cable, to a computer and power source. This makes these systems impractical to be used in certain scenarios. Privacy is still a concern with the Kinect as it is easily spotted and recognized as a camera and it has a high resolution.

3.1.3 Time-of-Flight-based solutions

Giovanni Diraco et al. [5] investigated the use of Time of Flight (ToF) cameras as a privacy-preserving solution for human posture recognition. A ToF camera was wall-mounted near the ceiling in two rooms to detect four postures in four levels of detail starting from posture recognition on simple postures, such as standing and sitting. To more complex postures such as sitting on the chair with arms along the body. The system was able to recognize up to 19 postures on the highest detail level. For the classification of the postures an SVM with a Radial Basis Function (RBF) was used and two descriptors, a volumetric-based and a topological-based, were used for the feature extraction. The system has an average accuracy of 96% for the first two detail levels, 92% for the third level, and 96% for the highest detail level at short distances ($\leq$ 3m) using the volumetric descriptor.

Though the Diraco claims that this solution is privacy-preserving, a clear silhouette of the subject can be seen in the images. The used cameras required a wired connection to the central computing unit that did all the data processing and posture recognition.

3.2 Sensor-based solutions

Allen Y. Yang et al. [29] propose a human action recognition system using wearable motion sensor networks and a distributed sparsity classifier. The sensors network consists of five sensor nodes containing an accelerometer and gyroscope were placed on the waist and limbs. The system deploys a local classifier on the sensors and a global classifier on the server. The local classifiers determine whether the measured data is part of an action and transmit the data to the server, which uses the received data from all sensors to perform the action recognition. Thirteen actions can be recognized with an average accuracy of 93%.

Ye Liu et al. [15] developed an efficient algorithm for temporal patterns recognition for sensor-based activity recognition. The algorithm uses low-level action and the relations between them to convert them into temporal patterns of high-level actions. These temporal patterns are then used as the feature vectors for the classification methods used (SVM, NB, and kNN). The algorithm was tested on a data set containing body sensor data of five activities of daily living: *relaxing, early morning, coffee time, sandwich time,* and *clean up.* The output of the algorithm was used in three classification methods: SVM, NB, and KNN. With a pattern dimension ≥ 3, all classification methods have an accuracy of higher than 90% and NB has the best performance with an average accuracy of 98%.

Jian Huang et al. [13] propose a human posture recognition and indoor localization system using a wireless wearable sensor system. The posture recognition system consists of a wireless sensor network (WSN) containing five sensors placed on the waist and the lower limbs. The sensor nodes are built with an accelerometer, gyroscope and magnetometer. The features chosen for the posture recognition are the pitch angles of the waist and tight sensor node using these features the system was able to recognize five postures, namely *standing, sitting, squatting, supine*, and *prone*, with an average accuracy of 100%.

These solutions use multiple sensors that are placed at specific points on the human body. When these sensors are not accuracy placed, the pattern recognition algorithms can produce incorrect results. The user of these solutions has to maintain and correctly place these sensors making them inconvenient to use.

3.2.1 Thermopile-based solutions

Munkhjargal Gochoo et al. [7] have developed a multi-sensor human posture recognition system that uses three grid-EYE sensors on different axes. The sensors on the x-axis and y-axis were placed 0.9m above the ground while the z-axis sensors were places 2.7m above ground. Four volunteers (two males, two females) in their mid-twenties to mid-thirties were used to create the data set of eight postures (*stand, hand raised, akimbo, wide open arms, squat, toe touch, crawl* and *Lie*) resulting in 15063 samples for each sensor. 46x16 heat maps were created by stitching the samples of the sensors together and enlarging them. The deep convolutional neural network (DCNN) used these heat maps as input. The experiment achieved an average accuracy of 99.95% over all eight postures.

Matthew Burns et al. [2] proposed a single component activity recognition system using two thermopile array sensors. These sensors have a resolution of 32x31 pixels with one mounted on the ceiling and the other in the corner of the room. The activities that were recognized are *opening and closing fridges*, *using fridge*, *using coffee cupboard* and *sitting at the table.* The activity recognition was divided into the posture recognition and the nearest object recognition. For the posture recognition, the feature sets both sensors were combined into a single feature set. The nearest object recognition determined the shortest distance between the subject and the objects and whether that distance was small enough to the object such that it should be considered for the activity. A single subject was used to create the data-set of 586 feature-sets. The Random forest algorithm was used for the recognition of the single component activities and achieved an accuracy of 91.47%.

The thermopile array sensors used by Burns have a high resolution compared to those used by Gochoo. Both works use a central computation unit with high computational power to process the data and perform the posture recognition. The work of Burns also has a problem with recognizing certain postures due to occlusion, while the work of Gochoo has not been tested with occlusion or other factors that introduce noise into the thermal images of the sensors.

Chapter 4

Posture recognition system

The proposed room-wide posture recognition is described in this chapter. This first few sections described the common posture recognition solutions and compare them with the proposed system. The rest of the sections describe the proposed system in more detail.

4.1 System Analysis

The most basic form of posture recognition uses a single camera to recognize the posture at a fixed position. The camera is generally placed on a wall facing the position such that the full posture is captured. An illustration of a typical posture recognition setup is shown in Figure 4.1. There have been several research studies [3, 11, 12, 24] exploring posture recognition using a single high resolution camera. Generally, there are four steps in the posture recognition. The captured data is pre-processed to extract the silhouette using a number of algorithms. A feature set is created from this silhouette which is then used in a machine learning algorithm, such as Support Vector Machine (SVM), Neural network (NN), K-Nearest Neightbours (KNN) to perform the posture recognition. These systems have high accuracy when recognizing simple postures with *standing, sitting, lying down, pointing* and *arms wide* as common postures used for the recognition. The disadvantage of these systems is that the performance is highly dependent on the lighting conditions and the algorithms used to extract the silhouette.

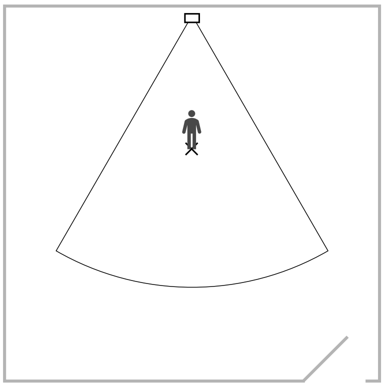
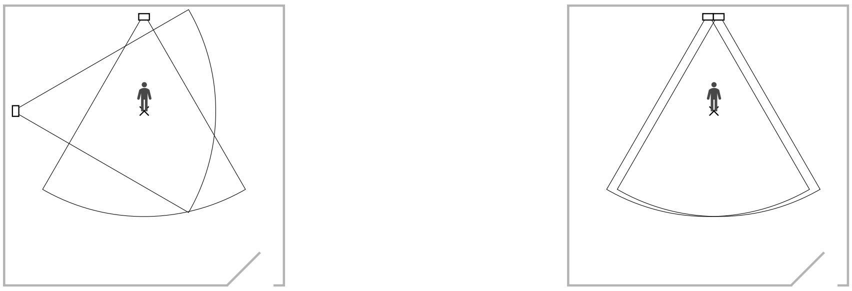


Figure 4.1: Example of a single sensor setup for posture recognition

One of the solutions to recognize more complex postures is to use multiple sensors to increase the information that captures the postures. These sensors can be configured to capture the posture from multiple angles or they can be placed next to each other, as can be seen in Figure 4.2. The information of the multiple sensors is combined differently depending on the sensors used. With the second configuration, it is often the case that different types of sensors are used. A popular setup for posture recognition is the Microsoft Kinect that consists of a high-resolution camera and a depth sensor [1, 3, 9, 28].

With the increased focus on privacy-preserving methods of posture recognitions more low-resolution sensors are used. These sensors do not have the details needed to recognize more complex postures on their own, but by using them in the first configuration, as shown in Figure 4.2a, they can overcome their lack of detail. The research of Burns et al. [2] and Gochoo et al. [7] are examples of how the first configuration can be used for posture recognition. Other reasons to capture the postures from multiple angles, even with high-resolution sensors, is to capture postures of which certain details are occluded from the view of a single sensor.

A derivative of the second configuration is shown in the research of Gouiaa [9] which uses a high resolution camera sensor in combination with four IR lights to perform posture recognition at a fixed position. The IR lights are mounted in the corners of the room against the ceiling and the camera is mounted in the centre of the ceiling. Posture recognition was performed by combining the silhouette and the IR shadows cast on the floor by the different IR lights in succession. The main challenge of these systems is how to combine the outputs of the different sensor and their placement to obtain the desired input for the ML algorithm.



(a) Multi-angle configuration

(b) Parallel sensor configuration

Figure 4.2: Examples of sensor configurations for multi-sensor posture recognition

Posture recognition has come a long way in the last few years and is used in an increasing number of use cases. Several of the more recent use cases, such as smart buildings, assisted living, require that the postures are recognized in larger areas. As mentioned in the previous chapter, posture recognition systems generally use one of two approaches. The first approach uses camera sensors to capture the posture data and the second approach uses inertial sensors. Most of the current solutions that use the first approach have focused on posture recognition at a fixed position or within a small area. While most solutions that use the second approach can be used in any area as long as the user stays within the range of the communication method used to communicate with the sensors, they have the disadvantages that a large number of sensors have to be worn by the users. The second approach is not considered privacy-invasive, but they do require a daily routine from the users to function.

Localization

There are several reasons to use localization for posture recognition:

- The location can be used in the training of the ML model and in recognizing the posture by using it as input for the ML model.
- In situations where multiple users are partially occluded, the localization can be used in the pre-processing of the sensor data to separate them.
- Localization can be used to train ML models specialized for different locations and the model for the recognition can be selected based on the location.

The main envisioned use case for this system is to control the HVAC system of a smart building per room by recognizing the postures of the people present in the different rooms. The first step in doing this is pre-processing the captured data to separate each occupant from the data. Low-resolution sensors are used to make the system privacy-preserving increasing the chance that when several occupants are standing in the room and occluding parts of each other in the images. This increases the difficulty of the separation process. When knowing the locations of the occupants of the room, the separation process can use this information to make the process more accurate. An example of this situation is illustrated in Figure 4.3.

Another reason for using the location is that when the distance to the low-resolution sensors increases, details of the posture is lost and a large portion of the images only contains noise which either has to be filtered out through pre-processing or by using a more complex and larger neural network model. The location can be used to train more specialized models at certain positions to improve the accuracy of the recognition and less complex model can be used making the system more efficient and use fewer resources.

There are a significant number of methods to estimate the location and most of them require additional sensors to be added to the system. Kemper et al. [14] propose a localization method that uses thermopile array sensors to estimate the location. The method uses multiple sensors for the localization and estimates the location using the position of the sensors and the angle at which the person is detected. This angle consists of the angle of the sensors in relation to the edges of the room and the Angle of Arrival (AoA), which is calculated from the data of the sensors. The accuracy of the proposed method has room for improvement as the measured maximum error was 80cm with an average error of 34cm. While this localization method is not as accurate as other methods it does not require additional sensors.

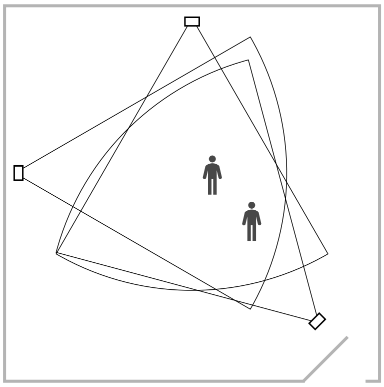


Figure 4.3: Example of an occlusion situation. From the viewpoint of the sensors closest to the door, a single shape is shown, the other sensors have a clearer view of the situation.

Adaptive system

One of the ways that a posture recognition system can be made adaptive is in the number of sensors used for the recognition. Multiple machine learning models are used to perform the posture recognition in the proposed system. This choice was based on one of the experiments that compared the accuracy of a system that used multiple models, each trained on the data set at a given position, against the performance of a single model trained on the data from all positions. Details of this experiment are given in section 5.7.5.

The noise in the data captured by the thermopile array sensors increases as the distance between the sensor and the person grows larger. The reasons for the increase in noise are:

- The IR radiation emitted by the human body is captured at a reduced intensity by the sensors. Making it harder to differentiate the human shape from the background.

- Only a small portion of the data captured by the sensor represent the shape of the human body.

A large portion of the posture recognition is performed on noisy data. To deal with the amount of noise, either the noise should be filtered out (i.e. by background extraction algorithms) or larger ML models are needed to accurately recognize the postures. Another method of reducing the noise is performing posture recognition using only the sensors that are close to the person.

In an experiment, it was observed that when a person was close to one of the sensors that the performance of using a single sensor was similar to that of using all sensors for the recognition. Machine learning models trained for these scenarios have a smaller input that contains less noise resulting in smaller models. The adaptive system chooses the best ML model based on the location, making the system more efficient in terms of the resources used. This allows for distributed posture recognition by performing the recognition on the sensor node when the person is close to it. An example of an adaptive system is shown in Figure 4.4.

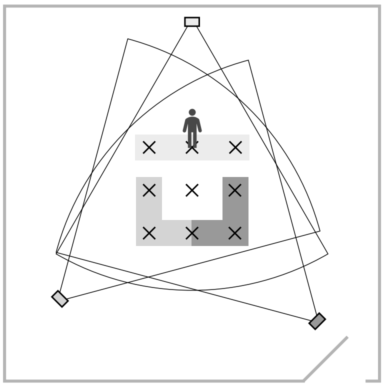
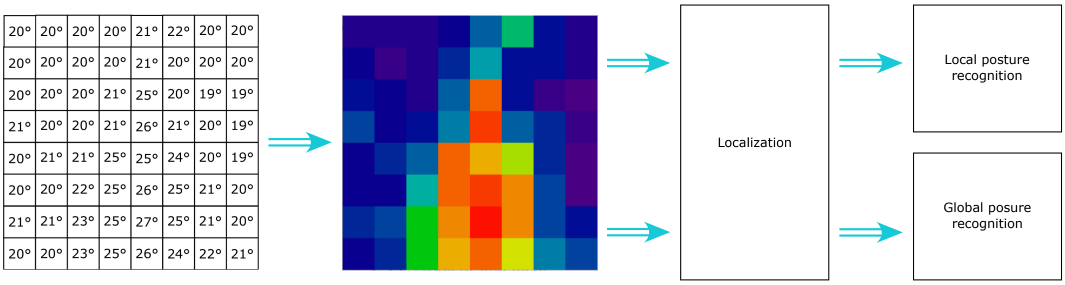


Figure 4.4: Example of a deployed adaptive system. The grey areas denote where the only the closest sensor is used for the posture recognition and the crosses represent the positions at which posture recognition is performed.

4.2 System design

The posture recognition system consists of multiple thermopile array sensor based wireless sensor nodes that form a wireless network with a central computing unit to perform posture recognition within a specified recognition area. The sensor nodes are placed such that all of them cover the recognition area in its entirety, effectively capturing the posture from multiple angles. There are several positions within the recognition area at which the postures are recognized. To aid with the recognition at these positions, the location of the users is estimated using an algorithm based on the work of Kemper et al. [14]. This algorithm uses the information collected from the thermopile array sensors and therefore does not require additional sensors. Posture recognition is performed using multiple Convolutional Neural Networks (CNN). Selecting which CNN to use depends on the location of the person and the distance to the closest sensor node. When the person is close to a sensor node, the recognition is performed using only the heatmap collected from that sensor and at larger distances, the recognition is performed using the heatmaps from all sensors. The framework for the posture recognition of the proposed solution is shown in Figure 4.5.

The posture recognition process consists of four steps. In the first step, the sensors data is collected and pre-processed to create the heatmap and the data for the next step. In the second step, the information is used to perform the localization. The third step consists of selecting the correct CNN based on the estimated location and the final step uses this CNN to perform the posture recognition.



20°	20°	20°	20°	21°	22°	20°	20°
20°	20°	20°	20°	21°	20°	20°	20°
20°	20°	20°	21°	25°	20°	19°	19°
21°	20°	20°	21°	26°	21°	20°	19°
20°	21°	21°	25°	25°	24°	20°	19°
20°	20°	22°	25°	26°	25°	21°	20°
21°	21°	23°	25°	27°	25°	21°	20°
20°	20°	23°	25°	26°	24°	22°	21°

Figure 4.5: Posture recognition framework of the system

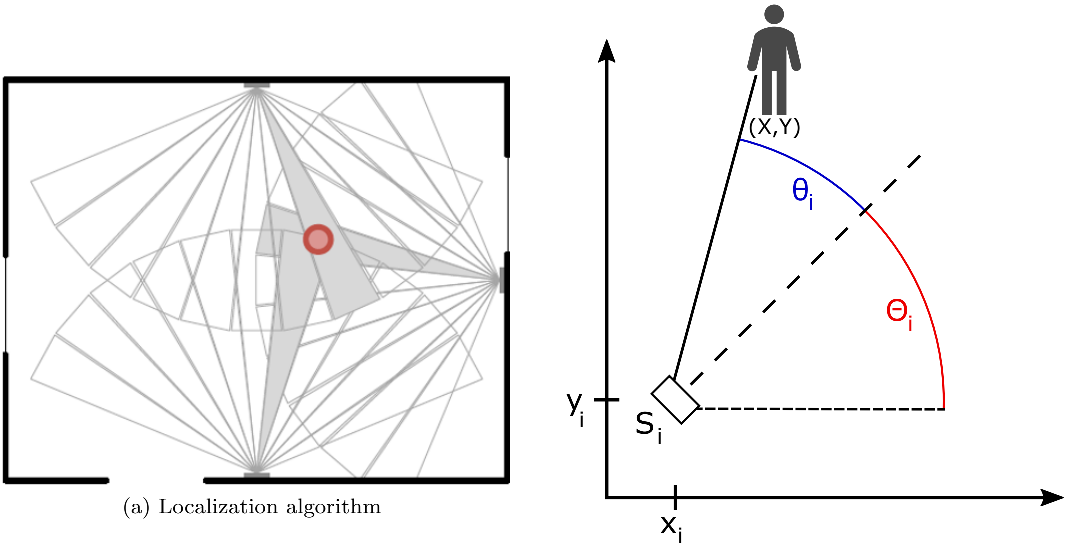
4.2.1 Data pre-processing

The first step in processing the data is to convert the sensor data of each sensor into a heatmap. Thermopile array sensors consist of a number of sensing elements. These sensing elements, called thermopiles, capture the infrared (IR) radiation emitted in its Field-of-View (FoV). The sensor translates the captured radiation into temperature readings. Each of the thermopiles in the sensors has its own FoV which combined form the FoV of the sensor. The temperature readings are used to create a heatmap where the colours red and blue represent the highest and coldest temperature. A minimum and maximum temperature are used when translating the temperature readings into pixel values. When this minimum and maximum temperature is static, the images change significantly when the conditions within the room or the distance between the subject and the sensor changes. To create more consistent images regardless of the conditions in the room and the distance to the sensor, the minimum and maximum temperatures are equal to the minimum and maximum temperature in the sensor readings for each image. As the distance to a sensor increases the details of the posture are less distinct with unscaled images, the scaling helps to bring out those details and creates more uniform images across all positions. This method does increase the intensity of the noise in the temperature data. The size of the heatmaps depends on the number of elements in the sensor and their configuration.

The second step is to use the created heatmap to determine the angle of arrival for each of the sensors and estimate the location using the angles and positions of the sensors. The final step uses the heatmaps to recognize the posture. A more detailed description of the second step can be read in section 4.3 and the details of selecting the correct CNN is described in section 4.5.

4.3 Localization algorithm

The performance of the posture recognition can be improved when the location of the person is known. By using the location to select the CNN that can is best suited for the recognition. The localization algorithm is based on the proposed algorithm of Kemper et al. [14] as this algorithm is based on thermopile array sensors. Therefore, no additional sensors are needed to add localization to the system and the pre-processing steps are not computationally intensive. This algorithm uses the location and positions of multiple sensors to triangulate the location of the users, as shown in Figure 4.6a. The localization is based on an arbitrary origin point and definition of the x-axis and y-axis. The variables that describe the location of a sensor are x_i and y_i, while θ_i describes the Angle of Arrival. The angle of the sensor, Θ_i, is determined from the x-axis , which always is positioned on the right side of the sensor, and is physically limited between $0°$ and $180°$. Since the angle of the sensor is determined from the right side of the sensors results in a negative AoA if the person is positioned within the FoV of the sensor on the right side. A visualization of these parameters is shown in Figure 4.6b. The location of the person is captured in the variables X and Y. The system of equations described in the paper of the algorithm can be converted into the following minimization problem:



(a) Localization algorithm

(b) Parameters

Figure 4.6: Visualizations of the localization algorithm and its parameters

$$\min \sum_{i=1}^{N} (Y - X \cdot \tan(\Theta_i + \theta_i) - y_i + x_i \cdot \tan(\Theta_i + \theta_i))^2 \tag{4.1}$$

One of the limitations of this algorithm is the use of the tangent function, which is undefined when the input is 90° and results in an unusable estimation of the location. When the combined angle of the sensor is 90° then it is save to assume that the person is standing at the same position as the sensor on the x-axis. So to cope with the limitation of the equation, the initial guess of X is set to the location of the sensor on the x-axis x_i and the information of that sensor is removed from the minimization problem.

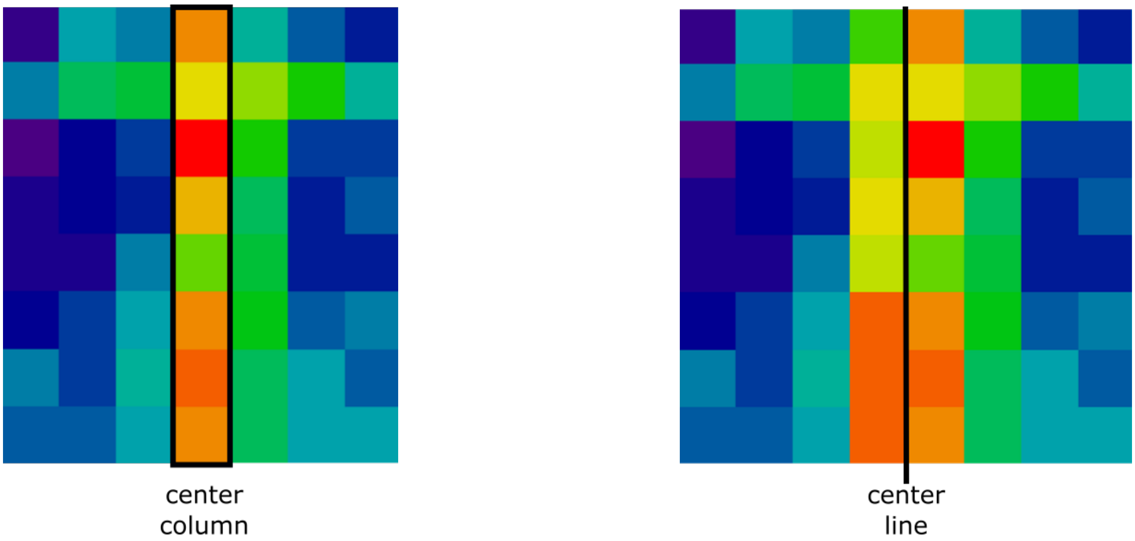
4.3.1 Determining the Angle of Arrival

What the algorithm of Kemper et al. [14] does not specify is the method of obtaining the Angle of Arrival (AoA) from the images generated from the sensors, as this is very specific to the sensors used and the implementation of the algorithm. The determination of the AoA is closely tied to the Field-of-View (FoV) of the sensor and in the case of thermopile array sensors, the FoV of each sensing element. Other factors that have to be taken into account are the number of thermopile elements and their configuration on the sensor. The configuration is commonly represented by a $m \times n$ grid.

The main factor that influences the AoA calculation is whether the number of columns, n, is even or odd. When n is odd then there is a column of elements that represents the centre of the sensor and when n is even then there is no such column. A centre line can be appointed to the heatmaps of thermopile sensors where n is even. Both cases are visualized in Figure 4.7.

The Angle of Arrival is obtained from these images and should be close to the centre of the silhouette to get the most accurate result. Using the hottest pixel to determine the AoA is the most intuitive method. The hottest pixels resides in a column of pixels and the index of this column is used to determine the AoA using the following formula:

$$\theta_{raw_{cc}}(i) = \begin{cases} f \cdot (c - i), & \text{if } i < c \\ -f \cdot (i - c), & \text{if } i > c \text{ where } i \text{ represents the index of the column} \\ 0, & \text{if } i = c \end{cases} \tag{4.2}$$



(a) odd number of pixels with centre pixels

(b) even number of pixels with centre line

Figure 4.7: Sensor image illustrating the centre line within the image.

The variables c and f represent the centre column in the heatmap and the FoV of each thermopile element respectively. The angle of the sensor is calculated from the right side of the sensor, so the determined AoA has to be negative if the hottest pixel is on the right side of the centre column, to get the correct angle for the localization.

A different formula is used when the heatmaps contain a centre line instead of a centre column. The AoA is this case is based on the distance between the centre line and the centre of the column, this adds half of the FoV to the distance compared to the calculation with a centre column. The columns adjacent to the centre line have an AoA of half the FoV (f). To capture this in a formula, the variable c is equal to the index of the column to the right of the centre line, resulting in the following formula:

$$\theta_{raw_{cl}}(i) = \begin{cases} f \cdot ((c-1)-i) + \frac{f}{2}, & \text{if } i < c \\ -f \cdot (i-c) - \frac{f}{2}, & \text{if } i \geq c \end{cases} \text{where } i \text{ represents the index of the column} \tag{4.3}$$

Basing the AoA solely on the hottest pixel has some disadvantages as the hottest pixel can be located anywhere on the silhouette and is sensitive to noise, which results in inaccurate estimations. The alternative is to use the hottest column to calculate the AoA as the hottest column will be closer to the centre of the silhouette and is less sensitive to noise. While this improves the accuracy of the estimations, they can be further improved by including the adjacent columns.

Due to the low pixels count, it is possible that the heat from a human body is captured by adjacent pixels with a reduced temperature. Therefore, the accuracy of the AoA can be improved by including the adjacent columns of the hottest column in the determination of the AoA. These columns can then be used to estimate the centre of the silhouette by determining the concentration of heat between these columns. The centre of heat can be determined in a similar fashion as the centre of gravity is determined of an object consisting of multiple parts.

Determining which column is the hottest column is be based on the pixels values of the heatmap, which uses the RGB standard for those values. Since high temperatures are represented with a reddish colour in the heatmaps, the hottest column is determined based on the sum of the R values of its pixels is the highest. The sums of R values of the hottest column and its neighbours are stored in the variables h_{ln}, h_{hc}, h_{rn}. The distance of the neighbouring columns is equal to the

FoV of the thermopiles. These sums are used to calculate the concentration of heat between these columns. The formula used for this is derived from the formula that is used to calculate the centre of gravity with the distances taken as the FoV of the thermopiles and the sums replace the weights resulting in the following formula:

$$\theta_{detail}(h_{ln}, h_{hc}, h_{rn}) = \frac{f \cdot h_{n1} - f \cdot h_{n2}}{h_{n1} + h_{hc} + h_{n2}} \tag{4.4}$$

The final Angle of Arrival is the sum of the previous formulas:

$$\theta = \theta_{raw} + \theta_{detail} \tag{4.5}$$

4.4 Convolutional Neural Network architecture

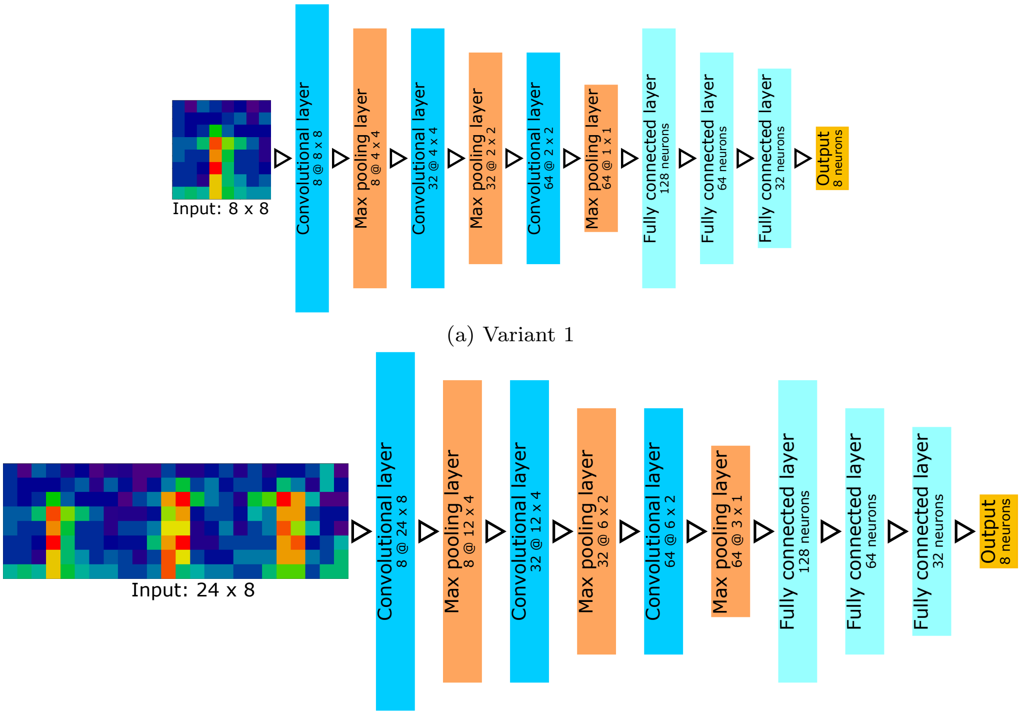
The posture recognition is based on the heatmap created from a single sensor or the concatenation of the heatmaps for all sensors. The size of the heatmaps is dependent on the number of sensors used in the system and even though low-resolution sensors are used, the pre-processing of these heatmaps is complex when using regular neural networks. There is a type of neural networks that is commonly used for analyzing visual imagery called Convolutional neural networks (CNN) that requires far less pre-processing. That is the main reason for this system to use convolutional neural networks for posture recognition. Compared to deep neural networks, a CNN has two additional types of layers, the convolutional layer and the pooling layer. These two types of layers are used to reduce the input image into a feature set that is smaller and captures the spatial and temporal dependencies within the image. The aim of these layers is to reduce the memory footprint and computational complexity.

Convolutional layers consist of a number of kernels and an activation function. Kernels are $m \times n \times d$ matrices that move over the image from left to right and top to bottom. While the size of the kernel can be specified through the parameters of the layer, the depth of the kernel is always equal to the depth of the input. The values of the kernels are learned during the training of the CNN. The movement is controlled by the *stride* parameter which controls how many pixels the kernels move between each convolution step. In each step, the dot product is calculated between each kernel and input. It is possible that the size of the kernels do not perfectly fit the input in which case padding can be applied. When no padding is applied then the parts of the image were the kernel does not fit is discarded and this is called Valid Padding. Same padding applies zero-padding to the images to ensure that the kernels perfectly fit the image. After all convolutions steps have been taken and the outcomes of the kernels have been stacked along the depth dimension, creating the activation map of the layer, the activation function of the layer is applied.

Pooling layers consist of a filter that is used to down-sample the input. The size of the filter is a parameter of these layers and determine the factor at which the input is down-sampled. The filter moves over the input from in the same manner as the kernels do in the convolutional layers and the same parameter is used to control the movement. There are three types of pooling layers, the max, average, sum pooling. Max-pooling layers take the maximum values within the filter area and projects it onto the output, while average-pooling layers compute the average and sum-pooling layers simply use the sum of the values. Max-pooling layers are the most common pooling layers in convolutional neural networks.

The architecture of CNNs used in this system was based on the one presented by Gochoo et al [7]. This architecture was chosen as it was developed for posture recognition using thermopile sensors and could easily be adapted to fit our use case. The architecture uses three convolutional layers that are each followed by a max-pooling layer. After the last max-pooling layer, there are three dense layers and an output layer. The input used by their system is larger than the input that this system uses. Therefore, the feature extraction elements in the network should be changed to better should the input resulting in halving the number of kernels for all convolutional layers.

To make the system adaptive, two variations of the chosen architecture are needed to accept both types of input. When the person is close to one sensor, only that heatmap is used, otherwise the concatinated heatmap of the heatmaps of all sensors is used. The test setup used for the verification of the system uses three thermopile array sensors that use a grid of 8 × 8 thermopiles. This results in the first variant of the architecture based on an input of 8 × 8 pixels, while the second variant processes heatmaps of 24 × 8 pixels. The variations of the architecture are shown in Figure 4.8.



(a) Variant 1

(b) Variant 2

Figure 4.8: The two variants of the CNN architecture

4.5 Adaptive ML Model Selection

For each of the positions in the recognition area, a separate CNN model is trained with training data from that position. This makes the models more specialized and improves the accuracy of the system. The location information is used to determine which trained CNN model to use for the recognition. Directly using the location estimations to determine which CNN to use is not possible as the estimations contain a small error. By dividing the recognition area into a grid that separates the recognition positions and using the location estimations to determine in which grid position they are, mitigating most errors in the estimations. A visualization of the grid is shown in Figure 4.9. The CNNs are placed in the same grid to easily determine which CNN to use. This results in the following formula so that the following formula can be used for determining the CNN to use for the recognition based on the estimation location:

$$x = \min(N_x, \max(0, round(\frac{x_{estimate} - X}{\delta_x}))) \tag{4.6}$$

$$y = \min(N_y, \max(0, round(\frac{y_{estimate} - Y}{\delta_y}))) \tag{4.7}$$

where X,Y are coordinates of the first point in the recognition area, N is the number of points on a given axis, and δ is the distance between the recognition points. This allows for an error in the estimation without affecting the accuracy of selecting the correct model. The min and max functions ensure that a grid location is always chosen correctly.

The disadvantage of the specialized models is that they require more resources, which for a centralized recognition is not an issue but for a distributed system is a disadvantage. Experiments have shown that performing posture recognition with a single sensor has similar performance compared to using all sensors when the person is close to that single sensor. This can be used to reduce the resources required on the sensor nodes by reducing the number of models stored on each node and reducing the size of those models. Figure 4.9 shows an example of a deployment of the system were the coloured rectangles represent the positions where the posture recognition only uses the image of the closest sensor. This example also shows that multiple sensors are considered close to a single position. In these cases, the estimated location is used to determine which sensor is closer to the person and that sensor is used for the posture recognition.

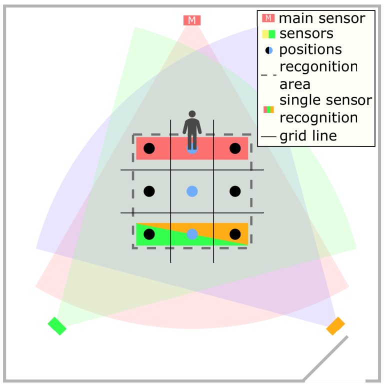


Figure 4.9: Example of an adaptive model selection setup

4.6 Integrated System

The deployment of the system is shown in Figure 4.10. It shows the topology of the software components and their communication. The system is divided into a central computing unit and multiple wireless sensor nodes.

Central computing unit functionality

The convolutional neural networks for the posture recognition are created and trained on the central computing unit. The distribution of the trained models to the sensor nodes is also the responsibility of the central computing unit. The other function of the central computing unit is to perform global posture recognition. The posture recognition uses the heatmaps obtained from all sensors and the location of the subject as input. Depending on the communication method used, as described below, the central computing unit might also be responsible for performing the localization estimations and determines whether the posture recognition is performed locally on the sensor nodes or on the central computing unit. When a sensor node joins the network, it receives its assigned convolutional neural networks and a configuration file, which contains information about when local posture recognition is performed and at what rate the sensor information is processed.

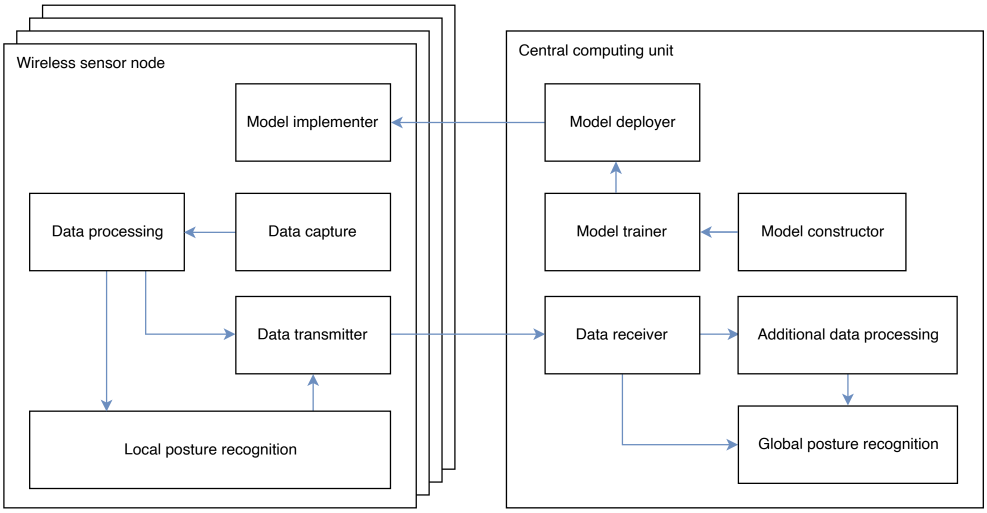


Figure 4.10: Deployment view of the system. The wireless sensor node converts the temperature data into a heatmap that is used to determine the AoA. Local posture recognition is performed when the person is close to that sensor node. The central computing unit receives the result of the local posture recognition. If the person is far from all sensors, the global posture recognition is performed using the information from all sensor nodes.

Wireless sensor node functionality

The sensor node converts the raw thermal data of the sensor into an 8x8 heatmap where red indicates the highest temperature and blue the lowest. This heatmap is used to determine the angle of arrival, which is used in the localization estimation of the subject. Depending on the chosen communication options, described below, the sensor node is responsible for determining whether the posture recognition is performed locally.

Communication options

Communication with sensors is required to perform the recognition locally on the sensors. This communication can be centralized by only allowing the sensors to communicate with the central computing unit. This reduces the frequency of the communication but increases the amount of data the needs to be transmitted to the central computing unit. Unnecessary communication also takes place when the posture recognition is performed on locally at a sensor node. An alternative is to have the sensors nodes communicated between themselves to determine the location and where the posture recognition is performed. This reduces the amount of data the is transmitted over the network as no unnecessary data is communicated.

This heatmap, as shown in Figure 4.5, is used as input for the local posture recognition. The data processing is responsible for the creation of the additional data used in the global posture recognition, which initially be the angle of arrival of the subject.

Chapter 5

Experimental Setup and Results

This chapter describes the experimental setup, the experiments that were performed and their results. In Section 5.1, the selection process of the thermopile array sensor is explained. The experimental setup used for the experiments is described in section 5.2. Experiments to explore the performance of the chosen sensor are explained in section 5.3 and their results are presented. The next section (5.4) describes the localization experiments and their results. The sensor configuration used for further experiments is described in section 5.5. The test setups used for the posture recognition experiments are described in section 5.6. The results of the posture recognition experiments are shown in section 5.6.

The experiments described in section 5.7 have lead to the adaptive system. To test the performance of this system, eight postures are selected and shown in section 5.8.1. The benchmarks that the adaptive system is compared against are described in section 5.8.2. For these benchmarks, a number of data sets are created from a sample set that contains all selected postures, the creation of which is described in section 5.8.3. The final section 5.8 reveals the results of the benchmarks and the adaptive system.

5.1 Thermopile array sensors selection

A literature study was performed to determine which thermopile array sensors are suited for human posture detection. The sensors should either capable of human detection or human posture detection, have a small footprint in terms of power consumption and size, and preserve the privacy of the users. The sensors that were chosen are the FLIR Lepton [6], Melexis MLX90621 [17] and MLX90640 [18], Panasonic Grid-EYE [22], and the Omron D6T [20]. The specifications of the sensors are presented in Table 5.1. While the FLIR Lepton sensor has the highest resolution and is able to detect humans at long distances, it is not suited for human posture recognition due to privacy concerns related to its high resolution combined with the high price and power consumption. The Omron D6T has the lowest resolution and the second-lowest power consumption; however, it is unable to detect human postures due to its low-resolution. The other sensors have solid specifications that meet the requirements and are capable of human posture detection. The Melexis MLX90621 might have a hard time discerning postures that differ on the vertical axis because it has a small vertical resolution and it has the lowest temperature accuracy of the three. The images produced by the Melexis MLX90640 sensor have a higher resolution and temperature accuracy compared to the Panasonic Grid-EYE but also has higher power consumption and price. Of these two sensors, the Grid-EYE is the only one that has been used for posture recognition in the work of Gochoo et al. [7]. The ambition of this research is to perform posture recognition with a low-resolution sensor as this area of posture recognition has not been extensively researched. This has resulted in the Grid-EYE sensor to be chosen for this research.

Manufacturer	FLIR	Melexis	Melexis	Panasonic	Omron
Sensor name	Lepton	MLX90640	MLX90621	Grid-EYE	D6T
Resolution (pixels)	80x60 160x120	32x24	16x4	8x8	1x8 4x4
Field of view (HxV)	25° 50° 51°x38° 56°x42°	55°x35° 110°x75°	40°x10° 60°x15° 25°x120°	60°x60°	54°x5° 44°x45°
Frame rate	8.7Hz	0.5Hz - 64Hz	0.5Hz - 512Hz	1Hz or 10Hz	0.5 - 4Hz
Current consumption	50mA 200mA*	25mA	9mA	4.5mA	5mA
Temperature accuracy	+/-5°C	+/-2°C	+/-3°C	+/-2.5°C	+/-1.5°C
Sensing range	-10°C - 400°C	-40°C - 400°C	-70°C - 380°C	0°C - 80°C	5°C - 50°C
Size (mm)	10.5x12.7x7.14	9.3x9.3x11.25 9.3x9.3x5.70	9.15x9.15x4.85 9.15x9.15x11.15 9.15x9.15x14.15	11.6x8x5.3	18x14x10.7 18x141x8.8
Price	\$175 - \$259	\$47 - \$51	\$29 - \$50	\$22 - \$32	\$17 - \$40

Table 5.1: Thermopile array sensor comparison (*with shutter active)

5.2 System setup

The scope of the prototype of the proposed system is limited to a single subject within a clean rectangular room with 3 sensors. The central computing unit of the proposed system is implemented on a personal computer that has an Intel P7350 CPU @ 2.0 GHz with 4 GB RAM. The sensor nodes, as shown in Figure 5.1, consist of a Panasonic Grid-EYE sensor and Raspberry Pi model 3B where Grid-EYE sensor collects the thermal data for the heatmaps and the Raspberry Pi is responsible for the data processing, and communication with the central computing unit over Wi-Fi. The central computing unit processes the heatmaps and location information of the sensor nodes, perform the localization estimations, and the posture classification.

For the posture classification, multiple Convolutional neural networks (CNNs) are used. These CNNs are modelled using the TensorFlow [8] platform, which is a machine learning platform that provides Python APIs for various supervised learning methods and the same programming language is used for the data processing. For communication between the central computing unit and sensor nodes, the Message Queuing Telemetry Transport (MQTT) protocol [19] is used. While this general setup is used for all experiments, the specific setup for the experiments differ on a number of parameters. These parameters can be split into two categories. The first category is hardware related and the second category is related to the CNN used and its training parameters. The parameters that are hardware related are shown in Table 5.2 and the CNN related parameters are shown in Table 5.3. Some of these parameters are used in the formulas described in the system design and are included in the tables.

The chosen thermopile array sensor, the Panasonic Grid-EYE, consists of 64 thermopiles, each having an FoV of 8° and they are configured in an 8×8 grid. This means that the heatmaps generated from this sensor contain a centre line between the 4^{th} and 5^{th} column of pixels. This information can be used to simplify the formulas in the localization algorithm. Since the heatmaps contain a centre line, the formulas 4.3 and 4.4 are used to determine the Angle of Arrival. This centre line is captured as the parameter c and is equal to 5 and the parameter f represents the FoV of the thermopile elements, which is equal to 8. The formulas used in the localization algorithm can be simplified using these parameter values to:

Parameters	**Description**
Number of sensors	The number of sensors used in the test setup
Recognition area dimensions	The dimensions of the recognition area, expressed in meters. The dimensions are the combination of the grid dimensions and the spacings
Number of recognition positons	The number of positions in the recognition area.
Recognition position grid	The dimensions of the grid in which the positions are placed. The parameters N_x and N_y represent these dimensions
Recognition position spacing	The distance between two positions both in the horizontal and vertical direction, expressed in meters and are represented by the parameters δ_x and δ_y

Table 5.2: Hardware related parameters of the experimental setup

Parameters	**Description**
Number of CNNs	The number of CNNs used to perform posture recognition.
CNN architecture variants	Specifies which variants of the CNN architecture are used
Learning rate	Determines the rate at which the weights in the layers are adjusted during training. This parameter affects the training time and performance of the network
Batch size	Determines the size of data samples to use in a single training step. This parameters affects how many samples are used to update the weights in the layers.
Epochs	Determines the number of steps that are taken during a training phase. This parameter affect the training time and performance of the network
Training steps	A step in the training phase that uses a batch of samples to updated the weight of the network based on the learning rate and can be expressed in the multiplication of the batch size and epochs

Table 5.3: CNN related parameters of the experimental setup

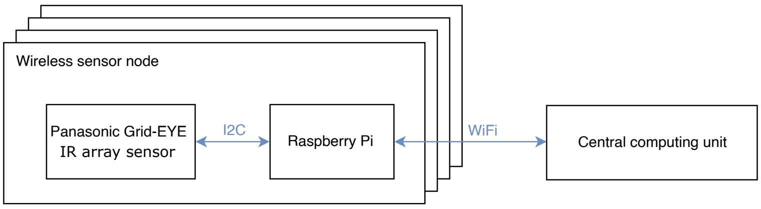


Figure 5.1: Block diagram of the wireless sensor node

$$\theta_{raw}(i) = \begin{cases} 8 \cdot (4 - i) + 4, & \text{if } i < 5 \\ -8 \cdot (i - 5) - 4, & \text{if } i \geq 5 \end{cases} \text{where } i \text{ represents the index of the column} \tag{5.1}$$

$$\theta_{detail}(h_{ln}, h_{hc}, h_{rn}) = \frac{8 \cdot h_{n1} - 8 \cdot h_{n2}}{h_{n1} + h_{hc} + h_{n2}} \tag{5.2}$$

5.2.1 Calibration

Another important aspect of the system is the calibration of the sensors, as the localization algorithm is very sensitive to the location and angle of the sensors. The calibration is both a physical and digital process. When the sensor physical position is determined, the horizontal angle (Θ) of the sensor has to be adjusted such that it covers the whole recognition area. Determining this physical angle precisely is a difficult and tedious task, which can be simplified by estimating the physical angle and digitally tuning it. The process of digitally tuning the angle is explained in the next paragraph.

At the start-up of the system, the locations of the sensors and their estimated angles are used to digitally calibrate the system. A high temperature object is placed in the centre of the recognition area to serve as a known reference point for the localization algorithm. A number of localizations are performed to obtain the error distance between the known location and the location obtained through the algorithm. Then the (digital) angles of the sensors should be changed in the opposite direction of the estimated location towards the known location. This process is repeated until the smallest error is found. This process can easily be automated to improve the calibration speed and accuracy.

5.3 Experiments to understand the performance of Thermopile Array sensors

5.3.1 Effective sensing distance of the Grid-EYE sensor

This first experiment was performed to test what the largest distance was to accurately perform posture recognition with a single thermopile array sensor. A test setup was made with a single wireless sensors node as described above is placed 90cm above the ground with a laptop for the central computing unit. Two postures (*standing* and *arms wide*) were captured at six positions in front of the sensor starting from 0.5m increasing by 0.5m up to 3 meters, as shown in Figure 5.2. Initially, the heatmaps created from the sensor data was created using a constant minimum and maximum temperature and results in the heatmaps created from the data at positions over 1.5 meters were the shape of the posture is barely distinguishable from the background noise. Using the data process described section 4.2.1, the postures are distinguishable from the background, but at distances larger than 1.5 meter from the sensor it become hard to distinguish the postures from each other. Starting from 2.5 meter from the sensor the noise in the images seem to increase significantly. The images taken at 0.5m from the sensor do not contain the full posture, most of the legs are missing for all postures and ends of the arms of the postures arms wide and arms up are not captured. So only the images from 1m onward were used for the posture recognition.

For each of the distances and postures, 82 samples were taken creating a sample set of 820 samples. A data set was created using half of the sample set and still containing all postures and distances. This data set was used to train a CNN using the second variant of the CNN architecture. This CNN was trained once with 8000 training steps using batches of 100 data samples and once using 20.000 steps using batches of 20. Four test sets were created from the same sample set. The first two test sets contain the samples taken at 1 meter with the first set having only the standing posture and the second set only the arms wide posture. The next two sets are similar to the last two only this time they contain the respective samples taken at 1 and 1.5 meter. The final test set contained both postures at all distances. The results of the posture recognition of both CNNs and the different test sets are shown in Table 5.4. These experiments shows that it is possible to do posture recognition with a single sensor when the person is standing close to the sensor. A final training of the CNN was performed consisting of 820.000 training steps using a batch size of 20. The results of the trained CNN show that with more training it would be possible to do posture recognition at large distances.

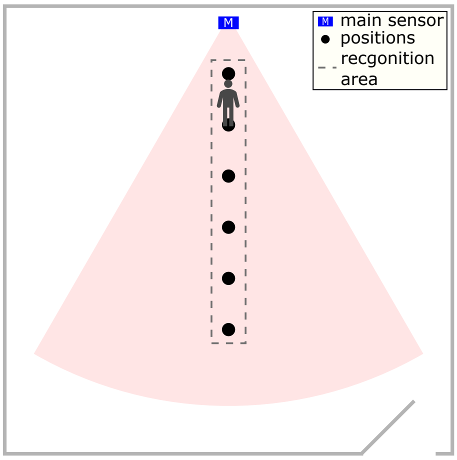


Figure 5.2: Test setup for sensing distance Grid-EYE

	8000 training steps	20.000 training steps	820.000 training steps
standing at 1m	100%	100%	100%
arms wide at 1m	86.6%	100%	100%
standing at 1-1.5m	50.0%	50.0%	88.7%
arms wide at 1-1.5m	95.2%	100%	82.1%
all postures at 1-3m	57.9%	59.9%	65.0%

Table 5.4: Posture recognition results for sensing distance Grid-EYE experiment

5.3.2 Angled posture recognition

The use-cases of the posture recognition system in smart buildings or assisted living would benefit if the posture can be recognized when the person is not directly facing any sensors. Two experiments were performed to test whether this was possible. The first experiment was performed to test if the front and back of the person could be detected, which should be possible in theory as the front side of the person often shows more skin and therefore emits more IR radiation. The test setup, as shown in Figure 5.4a, was used to capture a person facing the left-most sensor with all sensors having a distance of 1m to the person. Instead of determining the minimum and

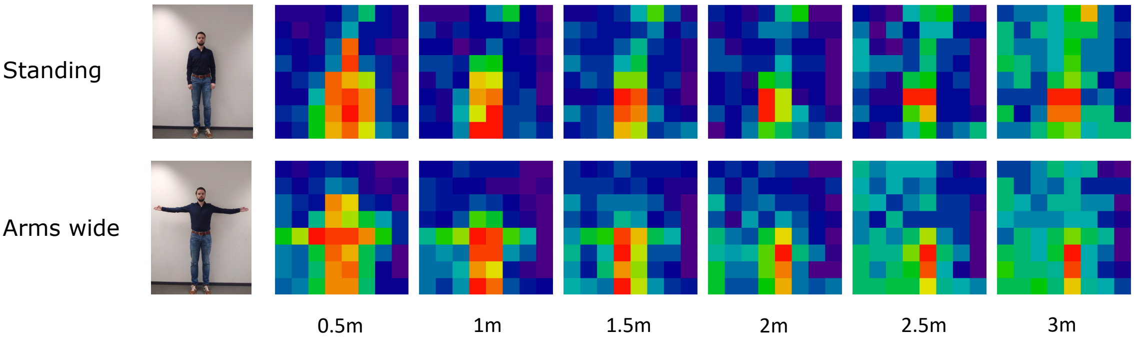


Figure 5.3: Examples of the heatmaps of both postures at the different distances.

maximum temperature for the translation to the heatmaps, the min and max temperature were determined by the min and max temperature found in the sensor data from all sensor nodes. With this translation method, it was possible to see a difference between the front and back side of the person. It would become more difficult to make the distinction between the front and back view as the distance increases. A second experiment was devised to determine whether it was possible to recognize the angle of the posture without using additional information. The postures (standing and arms wide) were captured at a single position using the setup shown in Figure 5.4b. The person would start with facing to the main sensor and then change his direction clockwise by 45° increments until the person was facing the other sensor. The data set consists of 47 samples for each posture and angle with a total size of 376 and a test set of the same size. The data samples contain the angle at which they were taken and two CNN models were trained on this data, one network is based on the second variant of the architecture, while the second network is based on the fourth variant that uses the known angle as input instead of the location, as shown in Figure 4.8b and 5.11b respectably. Both networks were trained with two sets or training parameters, training for 20000 steps with a batch size of 20 and a learning rate of 0.001, and training for 7500 steps with the same batch size and learning rate. The accuracy of the networks and training parameters are shown in Table 5.5 and they show that it is possible to recognize postures at multiple angles. However, knowing the angle needs more training as some angles, in this experiment the angles 45° and 135° show a very similar image making them difficult to distinguish. Recognizing postures at an angle was not further pursued in this research and the same holds for estimating the angle of the posture from the obtained heatmaps.

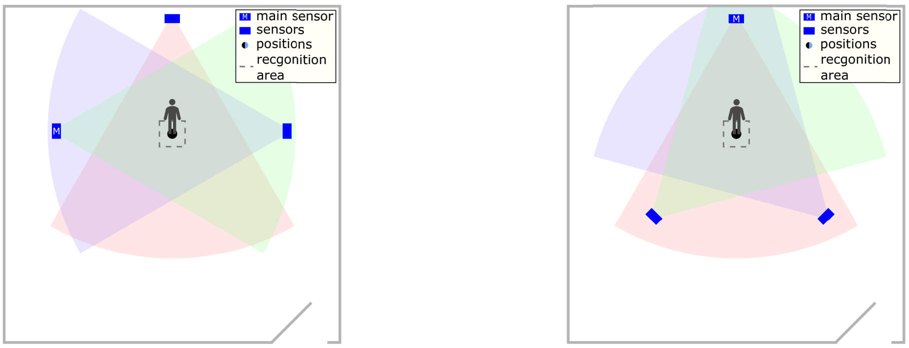
	7500 training steps	20000 training steps
main sensor	65.4%	100%
all sensors	100%	100%

(a) posture recognition using angle as input

	7500 training steps	20000 training steps
main sensor	100%	100%
all sensors	100%	100%

(b) posture recognition without using angle as input

Table 5.5: The accuracy of the posture recognition at an angle from the main sensor. All angles (0°, 45°, 90°, 135°) were used to get these results.



(a) Test setup 1

(b) Test setup 2

Figure 5.4: Test setups for recognizing posture at various angles

5.4 Localization experiments

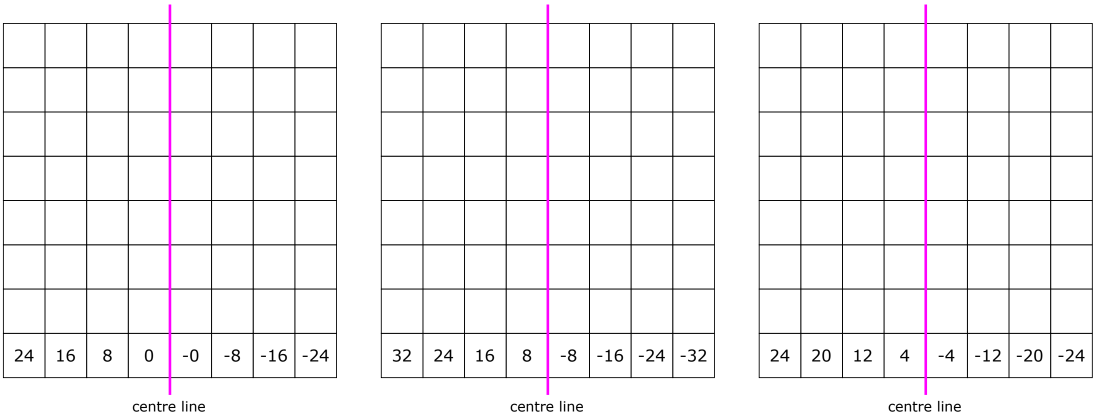
In this section, the experiments related to the localization algorithm are described. In the first subsection, the experiments related to using the hottest pixel to determine the Angle of Arrival are explained. The second section describes the experiments that use the hottest column to determine the AoA. Section 4.3 explains in detail how these two values are used in the determination of the AoA.

5.4.1 Determining Angle of Arrival based on the hottest pixel

The initial idea was to use the hottest pixel to determine the AoA since this was the method used in Kemper's paper. The heatmaps obtained by the thermopile array sensors are 8x8 pixels corresponding to the 8x8 sensing elements that have an FoV of 8°. The angle of the sensor is calculated from the right side and this translates to the AoA being negative on the right side of the image and positive on the left side. The image does not have a centre pixel so the split is made between the 4th and 5th column of pixels. The AoA has to be able to be zero when the subject is standing in the centre of the image to get accurate results. Each pixel has an field of view that is used to translate the position of the hottest pixel to an angle from the centre of the image. Three options of translating the position and FoV to the angle of arrival were tested.

The first two options base the translation on the edges of the hottest pixel while the third option uses the centre of the pixel. The first option sets the AoA to zero when the hottest pixel is adjacent to the centre line, located between the 4th and 5th column, and increases the angle by 8° for each column of pixels to the side of these two columns. The second option increases the angle by 8° for each column, including the 4th and 5th columns. The third option is silimar to the second option only the angle for the fourth and fifth column is set to four degrees. Figure 5.5 shows the AoA for each column of pixels obtained by these three options.

Two experiments were performed to evaluate the different options of translation from pixel to AoA. In the first experiment three sensors were used to estimate the location of the subject at two positions, the configuration and resulting estimations are shown in Figure 5.6. The second experiment used a similar setup to perform location estimation at 4 positions. The errors measured in the first experiment were between 12cm and 50cm, with the maximum error for position 1 was lower at 35cm. Option 3 showed the highest accuracy in the first experiment with the first option having the worst accuracy. For the second experiment, the measured errors also showed a large variance between the different positions and the used AoA method. This time the measured errors were between 18cm and 92cm over all positions, but at position 2 they were between 18cm and 26cm, while at position 4 the minimum error was 84cm. This experiment showed a different result with the first AoA option being the most accurate overall and the third option having the highest overall error. Figures 5.6a and 5.6b give a visual representation of the setups and the results of the first and second experiment respectively. The second AoA option was chosen to be implemented

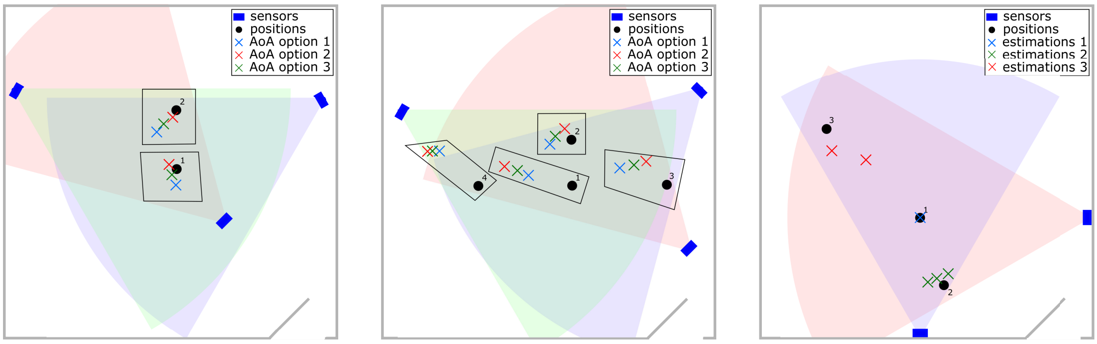


(a) option 1 (b) option 2 (c) option 3

Figure 5.5: The three translation options for the AoA. Each image represents the pixels in an 8 × 8 heatmap. The number at the bottom of each column represents the distance, in degrees, between the centre line of the image and the column, which translates to the value used for the AoA.

as it performed better on average in both tests with the first option being the second best.

After these experiments, a third one was performed using only two sensors to see whether the performance of the localization would be worse. The sensors were placed 90° from each other to estimate the location at three points and for each of these points, 10 measurements were taken. This experiment uses the first AoA option since the results of the previous experiments were not available yet. The results are shown in Figure 5.6c and they are better compared to the other two tests with no errors at the first position. The errors for the second position were between 10cm and 17cm, while for the third position a maximum error of 53cm was measured with a minimum of 24cm. The largest factor that contributed to the good performance of this test was that the system was better calibrated and one of the positions was placed in the centre of both sensors which resulted in no error in the estimations of that position. This would indicate that the actual sensor angles of the previous sensors differed from the angles used in the algorithm. The larger errors at the third position indicate that the distance of the positions in relation to the sensors also have an impact on the accuracy of the estimations. This can be explained by the fact that the hottest pixel would move within the shape of the posture resulting in a change in the AoA for one or more sensors. At short distances, this changed AoA would have a small impact on the location estimation.



(a) results experiment 1 (b) results experiment 2 (c) results experiment 3

Figure 5.6: Visual representations of the three localization experiments.

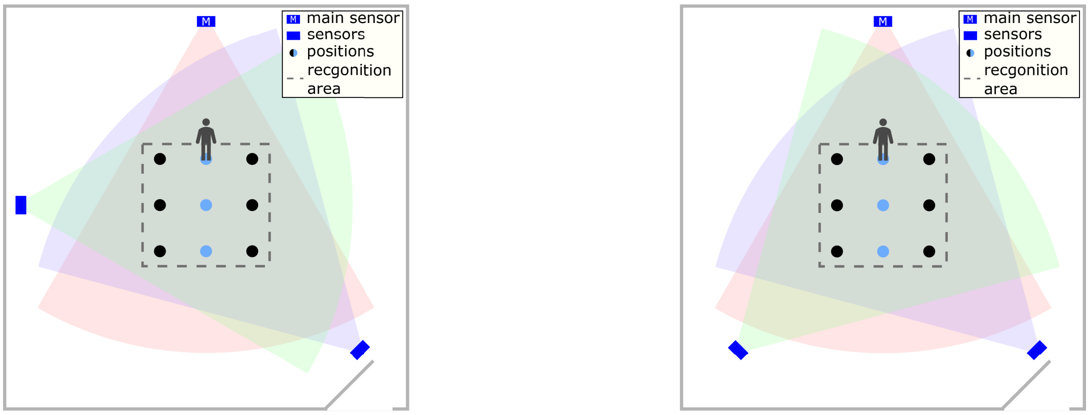
5.4.2 Determining Angle of Arrival based on the hottest column

After calibrating the angle of the sensors, the estimated locations still have a maximum error of 70cm. This prompted the search for alternative methods of determining the AoA to further improve the localization. The first option that was explored was the inclusion of neighbouring pixels of the hottest pixel in the determination. A more accurate AoA would be determined by using the hottest pixels for the initial AoA guess and then improve the guess with the heat information in the neighbouring pixels. With the exploration of how to use the neighbouring pixels to improve the AoA it became clear that using the hottest pixel for the initial guess did not have the desired accuracy and the neighbouring pixels did not have a large impact on the accuracy. The second option that was explored was using the hottest column for the initial guess and improve the guess with the heat information in the neighbouring columns.

The hottest column is determined by the highest sum of the red values of pixels in a column. Determining the AoA using the same formula but basing it on the hottest column significantly improved to the accuracy of the localization algorithm. Two methods of using the neighbouring columns were tested based on the sum of red values in the pixels, the heat sum. The first method would offset the initial guess by a maximum of 4° towards the neighbouring column that has the highest heat sum. The offset value is proportional to the difference in heat sums between the hottest column and its hottest neighbour. The second method is based on the idea using in the calculation of the centre of gravity, by shifting the AoA in the direction where the heat is concentrated. Multiple data sets of the posture recognition experiments were used to test both methods. The results varied between the data sets, but the results show that the first method in certain scenarios has a smaller maximum error while the second method is almost always more accurate on average. The second method was chosen to be implemented as it has a higher average accuracy.

5.5 Sensor configuration experiment

One of the design choices when there are multiple sensors in a system, is how to place them. Two configurations were tested, one where a front, side and angled view of the posture was captured and the other configuration captured the front view and two angled views of the posture, both configurations can be seen in Figure 5.7. For each of these configurations a data set of two postures was captured at each position in the recognition area. A single CNN was trained on the data sets for all positions for each configuration using the heatmaps of all sensors. The results show that the difference in performance between these is small with an accuracy of 97.5% for the first configuration and 98.3% for the second configuration. The second sensor configuration was chosen because of its slightly higher accuracy.



(a) configuration 1

(b) configuration 2

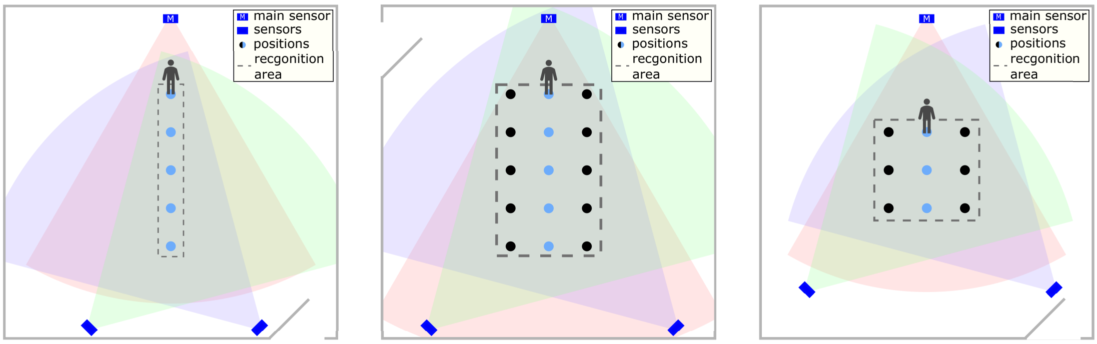
Figure 5.7: Sensor configurations

5.6 Test setups for the integrated system

There are multiple test setups that were used for the posture recognition experiments. These setups differ in their sensor configuration and recognition area. A sensor configuration defines the number of sensor nodes used and their positions in relation to the recognition area. The recognition areas define an area in which an number of positions are defined including the spacing between them. The multiple test setups use two sensor configurations and t5ree recognition areas.

An area of 0.5 by 3 meter forms the first recognition area where five positions in a line are defined, spaced 0.5m apart. The second recognition area is 1.5 by 2 meter and contains 15 positions that form a 3×5 grid. The positions are spaced 0.5m apart. The final recognition area is 1.5 by 1.5 meter and contains nine positions. These positions are placed in a 3×3 grid and the distance between them is 0.5 meter. For the recognition areas that contain multiple lines of positions, two distinctions are made. The centre line of positions in the area are considered the one-dimensional (1D) positions, while all positions are considered the two-dimensional (2D) positions.

The two sensor configurations differ in the number of sensor nodes that are used. The first configuration uses a single sensor node, while the second configuration uses three sensor nodes which on most experiments were as described in the next few sentences. The *main* sensor node was placed in the centre of the recognition area facing the area such that all positions were in the FoV. The other two sensor nodes were placed at the corners of the area on the opposite side. The distance between the sensors and the recognition area can differ per experiment, but most of the experiments use a distance of 1.5 meter for all sensors. A sensor node is depicted in Figure 5.10.



(a) Setup 1 with a recognition area of 0.5m by 2m and a sensor distance of 1m.

(b) Setup 2 with a recognition area of 1.5m by 2m and a sensor distance of 1m.

(c) Setup 3 with a recognition area of 1.5m by 1.5m and a sensor distance of 1.5m.

Figure 5.8: Overview of the different test setups

The test setup used for the final posture recognition experiments use the final recognition area and the second sensor configuration, a physical version of this setup is shown in Figure 5.9. An overview of the all test setups is shown in Figure 5.8 where the 1D positions are marked blue.

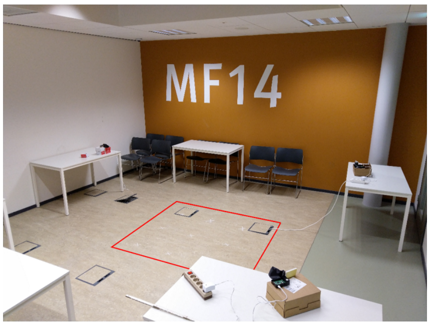


Figure 5.9: Physical testbed using 3 sensor nodes and a laptop for the central node

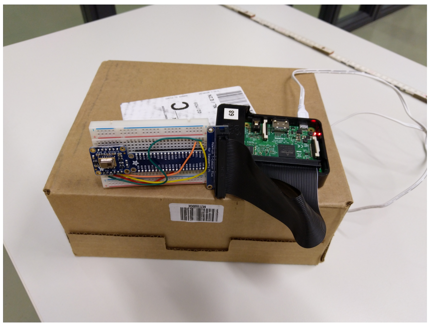


Figure 5.10: Wireless sensor node consisting of a Raspberry Pi 3 model B and the AMG8833 breakout board.

5.7 Posture recognition experiments using the integrated system

The first round of posture recognition experiments consists of six experiments. These experiments test different aspects of the recognition system using two postures (standing and arms wide) as test data. The following list gives an overview of these six experiments.

- Experiment 1 - Use multiple sensors to recognize postures at multiple distances from the main sensor.
- Experiment 2 - Uses multiple sensors to perform posture recognition at a larger set of positions. These positions are at different distances from the main sensor in both the horizontal and vertical direction.
- Experiment 3 - Uses the same setup as the second experiment and introduces localization as input for the CNN.
- Experiment 4 - For each position in the recognition area, a CNN is trained and the CNN used for posture recognition is selected based on the estimated location.
- Experiment 5 - Improves the system for the previous experiment by using an improved localization method and a different recognition area.
- Experiment 6 - Simulates a system that uses a smaller CNN for the positions that are close to a sensor.

To show the performance of the posture recognition system, the number of postures that can be recognized by the system was gradually increased from two to eight. The performance of each set of postures is compared using the benchmarks described in section 5.8.2.

5.7.1 Posture recognition at multiple distances results

The setup as shown in Figure 5.8a was used to capture two postures (*standing* and *arms wide*) at the position closest to the main sensor. While capturing the posture data the volunteer was facing the main sensor. Two data sets were created using this data, one set only contained the heatmaps from the main sensors and the other set contained the heatmaps from all sensors. A CNN based on the second variant of the architecture was trained on the first data set and one based on the first variant was trained on the second data set. For the training of both networks a learning rate of 0.001, batch size of 20, and the models were trained for 20.000 steps. The accuracy of both networks is based on the confusion matrix that shows what the recognized postures are and whether the postures were correctly recognized. The results were as expected with both models having an accuracy of 100%, as shown in Table 5.6.

	Accuracy
Main sensor	100%
All sensors	100%

Table 5.6: Accuracy's of the second posture recognition experiment

5.7.2 Posture recognition at multiple positions results

This experiment uses the setup shown in Figure 5.8b, which is an evolution on the setup used for the initial test. The sensors were placed 1m from the closest recognition position and the samples were taken with the volunteer facing the main sensors at all positions. In this experiment 4260 samples were captured, 142 for each posture at each position. From this sample set, four data sets were created, which can be split into two groups. The first group only contains the samples that were taken at the 1D positions and the second group contains all samples. For each group the sets similar to the previous experiments with one set containing only the heatmaps of the main sensors. This time four CNNs were trained using the same variants and training parameters as the previous experiment. The results of all are shown in Figure 5.7.

	Main sensor	All sensors
1D positions	99.2%	90.0%
2D positions	89.6%	92.0%

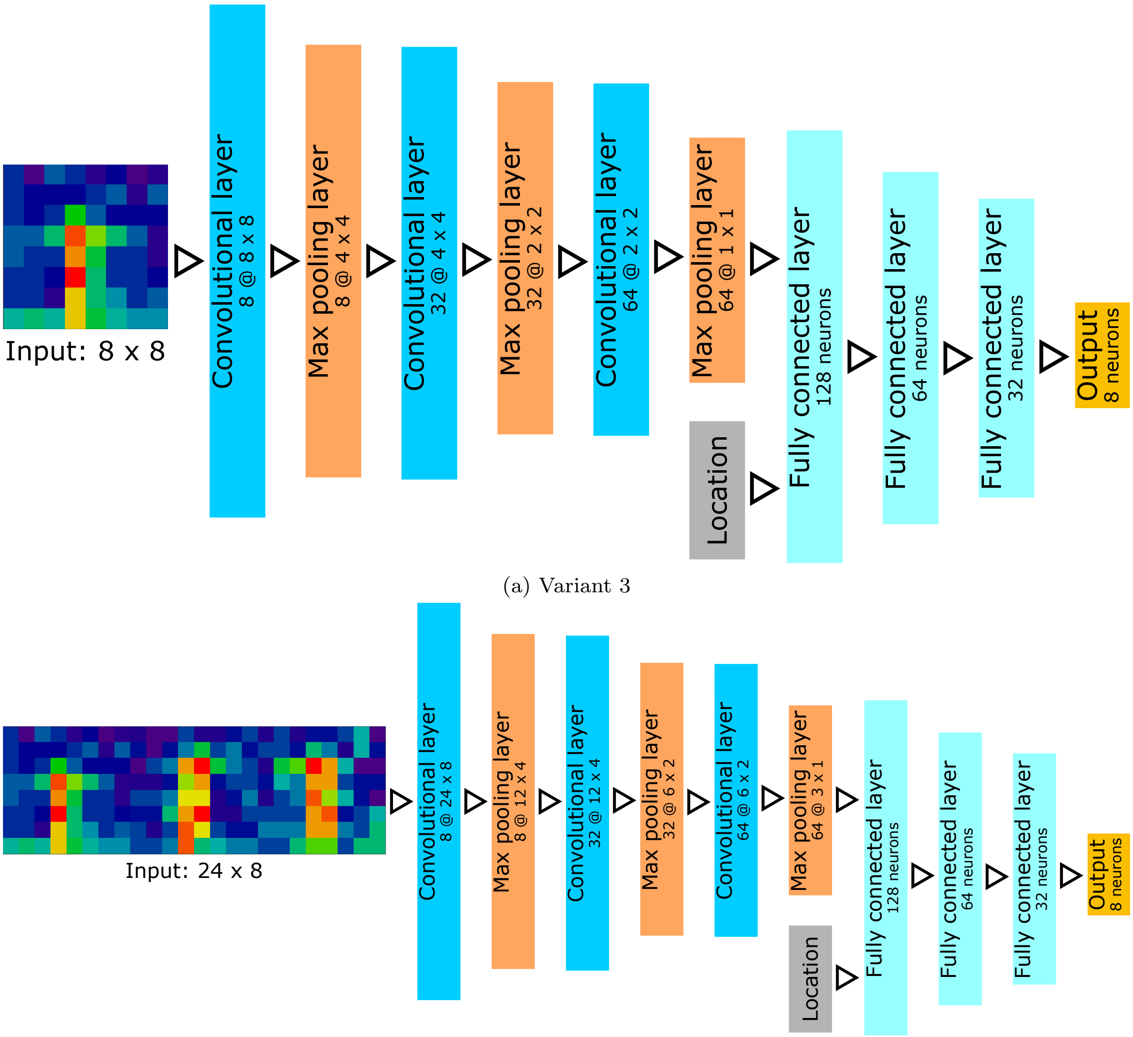
Table 5.7: Accuracy's of the second posture recognition experiment

The expectation was that the accuracy would drop as the initial experiment showed similar results. Another interesting comparison with the initial experiment is that positions of that experiment are the same as the 1D positions for this experiment. The results at the 1D positions are significantly high than the results of the initial test which can be attributed to different training parameters, the batch size in particular. Other contributing factors are a larger number of samples to train with and cleaner samples. In the initial experiments, an additional heat source, a lamp, was captured in the samples, as shown in Figure 5.3 on page 27.

5.7.3 Posture recognition using location as input results

The next step was to introduce the location into the posture recognition process. This information needed to be stored in such a way that they could easily be retrieved when training the CNNs. The localization was performed after the samples for a posture at a position were captured and the AoA was solely based on the hottest pixel. An additional row of pixels was added to the bottom of the heatmaps to store the estimated location in. The estimation location was stored as two separate values, one for the location on the x-axis and one for the location at the y-axis. When training and validating the CNNs, this additional row would be separated from the heatmaps and the location was extracted.

Similar to the previous experiments, four data sets were created using the sample set, but this time the known location would be stored in the additional row of pixels for the training and validation sets. The architectures of the CNNs used for this experiment were different than the previous experiments. These architectures, the third and fourth variant, add the (estimated) location to the feature set after the last max-pooling layer, as shown in Figure 5.11. The training parameters for the CNNs were the same as in the previous experiment The expected accuracy of this experiment should be high than that of the previous experiment; however, the accuracy of using the location as input for the CNN, as shown in Table 5.8 was slightly worse than without using that information. The main reason for the decreased accuracy can be found in the low accuracy of the estimated locations. While the estimations of the positions ≤ 2 meters were accurate with a small error margin, beyond this distance the accuracy dropped significantly with a maximum error of 1 meter.



(a) Variant 3

(b) Variant 4

Figure 5.11: The two variants of the CNN architecture

	Main sensor	All sensors
1D positions	91.3%	89.0%
2D positions	89.9%	86.9%

Table 5.8: Accuracy's of using location as input for posture recognition

5.7.4 Posture recognition using location to select the ML model results

To confirm whether the postures could be accurately recognized at all distances, multiple CNNs were trained on the data captured at each location. The accuracy's of these networks are shown in Table 5.9 and they show that without using the location as input and training the networks of each location is more accurate than training a single model for all locations. Making the recognition use these networks to recognize the postures in a single data set was a bit of a challenge of its own. The estimated locations are not very accurate so the models should be correctly selected even with an error in the localization. Resolving the estimated location to an identifier that allows for some error in the locations was performed by converting the estimated location into a grid location. This grid was defined using the formulas 4.6 and 4.7 which in this setup allows the estimated location within a minimum area of 0.25m by 0.25m around the actual location to be recognized by the model trained for the actual location. To ensure that all locations are recognized the areas for the positions at the edge of the recognition area also cover the areas outside of the recognition area. A visualization of the recognition area of this setup and the areas for the different models is shown in Figure 5.12 and the formula used is shown in 5.3 and 5.4. This method is based on the actual location of the first position, defined as X and Y in the formulas, which has to be known on the central computing unit and the sensors nodes.

$$x = \min(3, \max(0, round(\frac{x_{estimate} - X}{0.5}))) \tag{5.3}$$

$$y = \min(5, \max(0, round(\frac{y_{estimate} - Y}{0.5}))) \tag{5.4}$$

When simply selecting the correct model for each sample in the data set requires a long time to recognize the whole set as the models need to be reloaded into the system for each sample. So the samples in the set were first sorted per model and then the recognition was performed on the samples for each model in succession. The results were combined afterwards to produces the final result of the input data set. The accuracy of using this method to select the CNN model based on the location is shown in Table 5.9 and is worse than that of the manual selection, as expected since the estimated locations were not accurate enough.

	Manual model selection		Automated model selection	
	Main sensor	All sensors	Main sensor	All sensors
1D positions	100%	100%	79.9%	75.8%
2D positions	99.7%	99.8%	66.0%	72.9%

Table 5.9: Accuracy's of using multiple models for posture recognition

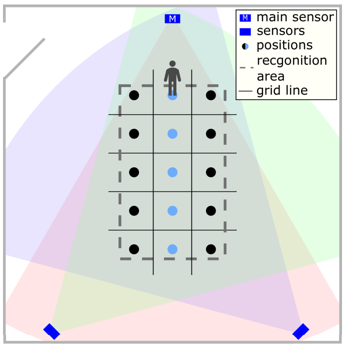


Figure 5.12: Recognition area used for the fourth experiment overlaid with the automated model selection areas

5.7.5 Posture recognition using the final recognition area results

This experiment improves on the results of the previous experiments by decreasing the recognition area and using the improved localization algorithm based on the hottest column and concentration of heat. Another improvement over the previous experiments is the use of the calibration method described in section 5.2.1. The setup used in this experiment is shown in Figure 5.8c and a sample set of 3510 samples was captured. The recognition area was decreased as the localization algorithm performed worse at the positions furthest away from the main sensor and the posture recognition at the same positions was less accurate. Another reason for reducing the recognition area is that at a distance of 1 meter between the sensors and their closest position is a too short of a distance to capture the postures in their full detail. Most noticeable, at the positions to the left and right of the position in front of the main sensor, the samples for the arms wide posture captured at these two positions were missing one of the arms. The resulting recognition area is the previous area minus the front and back row of positions reducing the 3×5 area into a 3×3 area. The sensors configuration remains the same but the distance is increased to 1.5 meter. The maximum measured location estimation error was reduced to 17cm with this setup and the improved localization algorithms.

data sets		**training set**	**test set**	**validation set**
front row positions	1D positions	352	40	388
	2D positions	1058	118	1164
All positions	1D positions	352	40	388
	2D positions	1058	118	1164
Single positions	1D positions	352[1]	40[1]	1164
	2D positions	352[1]	40[1]	3492

[1] for each model/position

Table 5.10: The training, test, and validation set sizes for the sample set of the fifth experiment

This sample set was used to recreate data sets that follow the line of progression of the experiments. The sample split is made between the 1D and 2D positions. So one set contains the postures captured at the positions in the front row, closest to the main sensor. Another set contains the posture data from all positions and a copy of this data set is created. Finally, a data set for each position is created. All data sets are split into training, test, and validation sets with a 45%, 5%, 50% split. For the training and test sets of the copy of the data set containing all position, the actual location is used as the estimated location which is stored in the additional row of pixels. An exception is made for the validation sets of the single position sets, these are combined into a single test set. The sizes of these sets are given in Table 5.10.

In total 30 CNN models are trained, four for the front row positions, four for all positions, four for all positions with localization, and 18 total for the individual positions. The training parameters for all CNN are a learning rate of 0.001, batch size of 20, and the number of epochs was chosen to get the training steps closest to 9000. The results of this experiment are shown in Figure 5.13

The results show an increase in accuracy for all data sets. The conclusion of this experiment is that using the location in the posture recognition does improve the accuracy. The highest gain in performance is achieved by using multiple models in the recognition. Another interesting finding of previous experiments is that in certain cases it is beneficial to only use the heatmap from the main sensor instead of using all sensors.

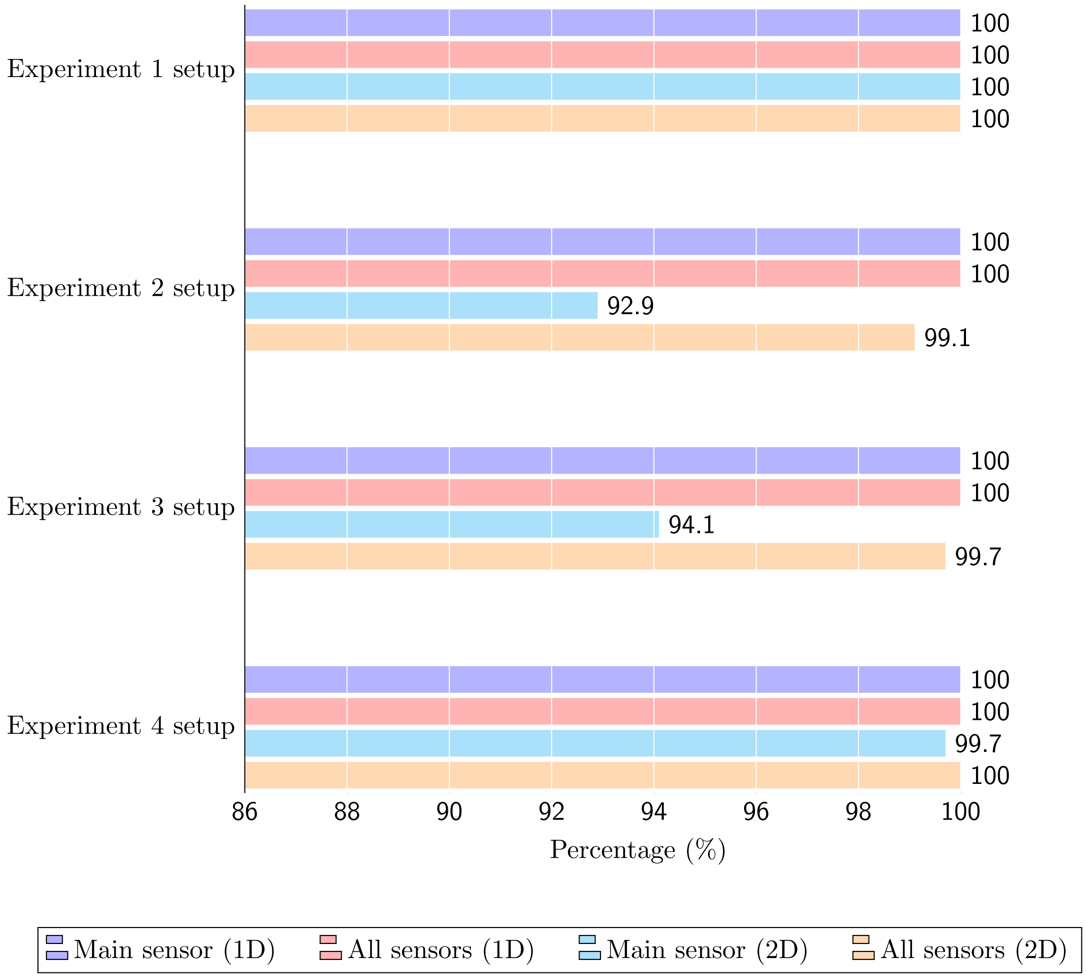


Figure 5.13: Accuracy's of the fifth experiment

5.7.6 Posture recognition using adaptive ML model selection results

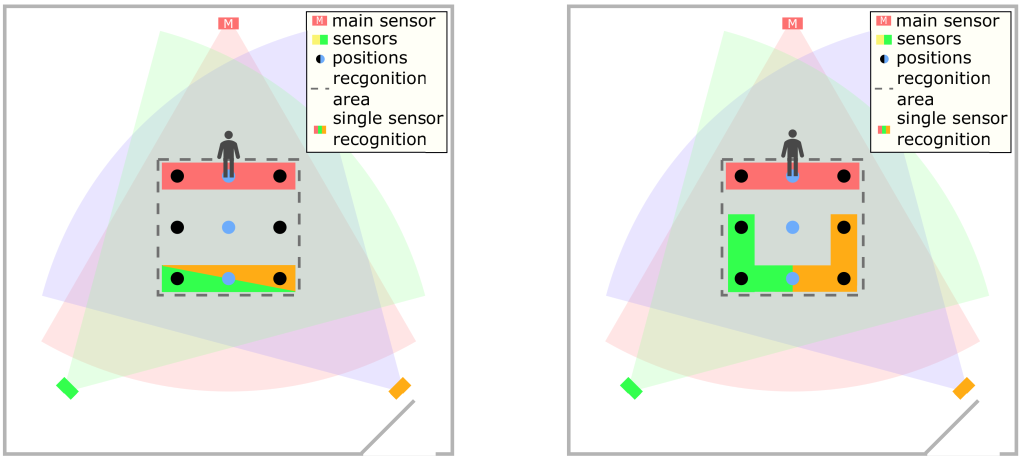
The purpose of this experiment was to explore using only the heatmaps of a single sensor for the recognition. In the previous experiments, in particular experiments 2 and 3, show that the performance of the recognition in some cases is better when using only a single sensor compared to using all sensors. This was most noticeable when the person is close to a sensor. The time needed to properly design a system that detects when a person is close to a sensor and adapt the recognition process to that information was not available. Therefore a simulation of such a system was designed. The sample set of the previous experiment was used to create a new data set. Figure 5.14b shown the separation of the recognition area when the closest positions to the sensors are separated. This separation only has a single position that uses the heatmaps of all sensors for the posture recognition. For the simulation, another separation was used that divides the recognition area into equal parts. This separation is shown in Figure 5.14a and one section consists of the row of position closest to the main sensor. The second section consists of the positions closest to the other two sensors. The remaining section consists of the middle positions of all columns. The choice for this separation was made because all sections are of equal size and would give a better indication of the overall performance of such a system. This separation results into four data sets, one for the section for each sensor and the section in the middle. The data set for each of the sensors only contain the heatmaps from that sensor and the other data set contains the heatmaps from all sensors and the sizes of these data sets is given in Table 5.11. These data set contain the

2D positions, no 1D version was made.

Data sets	training set	test set	validation set
sensor 1	264	30	291
sensor 2	264	30	291
sensor 3	264	30	291
centre positions	264	30	291

Table 5.11: The training, test, and validation set sizes for experiment 6

A CNN based on the second variant of the architecture is trained on the data sets for each of the sensors and for the remaining data set a CNN based on the first variant of the architecture is trained. The accuracy of the simulated adaptive system is based on the average accuracy of the different CNNs, the results of which are shown in Table 5.12.



(a) Separation based on distance to the sensors and equally size sections

(b) Separation based on distance to the sensors

Figure 5.14: The two options for separating the recognition area for the adaptive system

This simulation shows that the adaptive system can achieve a similar performance as a system that uses multiple models. The benefit of the adaptive system is that smaller models are used for when the person is closer to a sensor and these models can be placed on their respective sensors nodes to perform the recognition locally.

	sensor 1	sensor 2	sensor 3	centre
individual elements	100%	100%	100%	100%
whole system	100%			

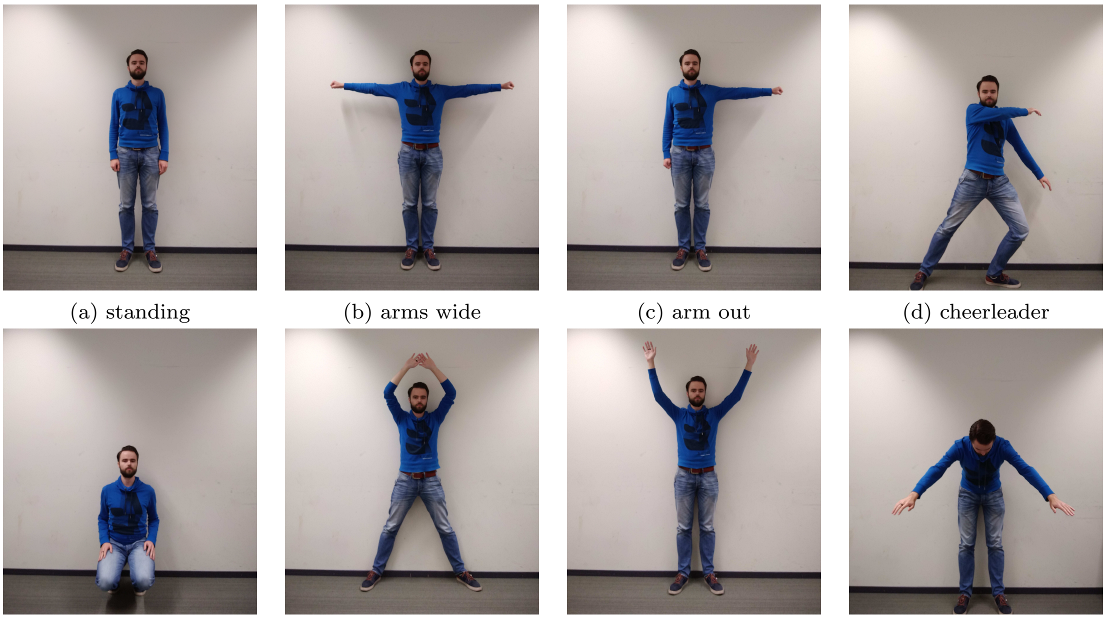
Table 5.12: Adaptive system simulation results

5.8 Final Experiments and their results

With the system design finished and tested, the performance of the system with more postures needs to be validated. More postures were added in increments of two postures resulting in three validation posture sets consisting of four, six, and eight postures. The postures were added in increments to observe the changes in performance. Section 5.8.1 describes which postures are selected for each of the posture sets. Four benchmarks are used to compare the performance of the adaptive system to existing solutions and other solutions that use localization for posture recognition. Further details of these benchmarks can be found in section 5.8.2. For these benchmarks, a sample set of 14040 samples, containing all eight postures, was created using the setup shown in Figure 5.17a. The process of creating the data sets for the different benchmarks and sets of postures is given in section 5.8.3. The results of all benchmarks are described in section 5.8.4.

5.8.1 Postures

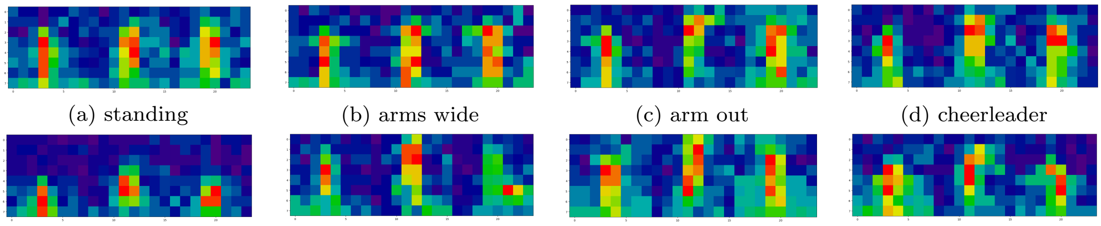
To test the performance of the system eight postures were selected to be used in the posture recognition. The selected postures are: *standing, arms wide, arm out, cheerleader, squat, tree, arms up,* and *leaning forward.* An overview of the different postures is shown in Figure 5.15 and their heatmap representations are shown in Figure 5.16.



(a) standing (b) arms wide (c) arm out (d) cheerleader

(e) squat (f) tree (g) arms up (h) leaning forward

Figure 5.15: Overview of the postures



(a) standing (b) arms wide (c) arm out (d) cheerleader

(e) squat (f) tree (g) arms up (h) leaning forward

Figure 5.16: Examples of the heatmaps of each posture

The postures that are included in the three validation sets are given in the following list:

- Four postures - *standing, arms wide, arm out,* and *cheerleader*
- Six postures - *standing, arms wide, arm out, cheerleader, squat,* and *tree*
- Eight postures - *standing, arms wide, arm out, cheerleader, squat, tree, arms up,* and *leaning forward*

5.8.2 Benchmarks

Four benchmarks were designed to compare the posture recognition system against. These benchmarks simulate existing solutions and different approaches to improve the performance by using localization. A list of the benchmarks is shown below:

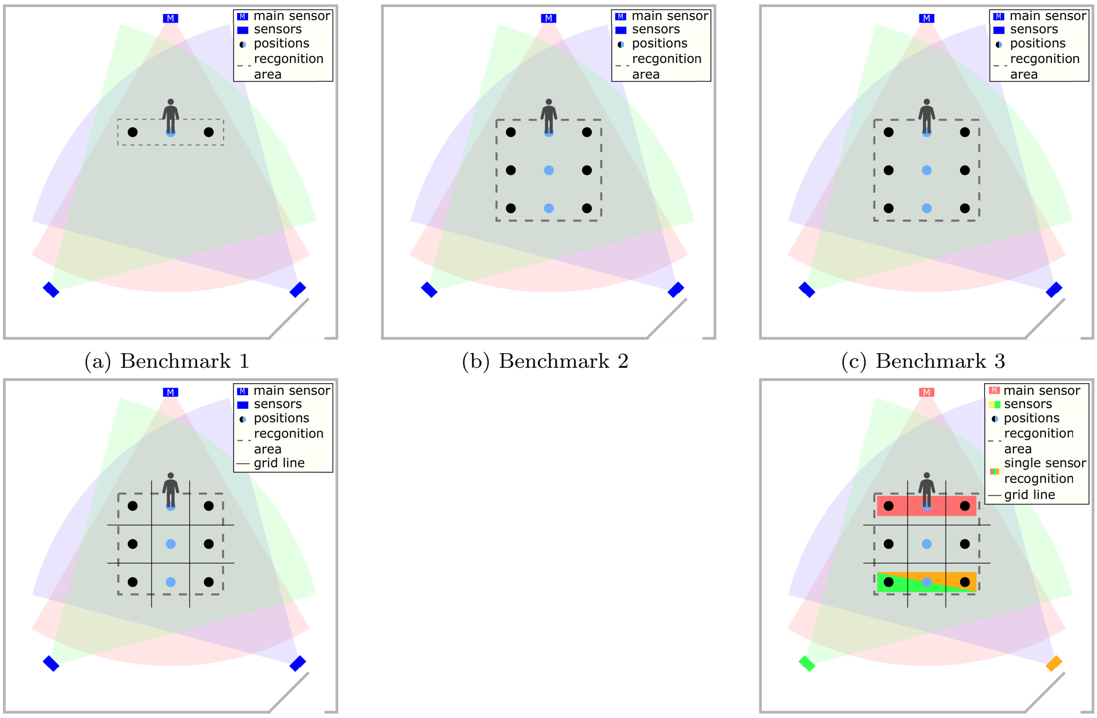
- Benchmark 1 - Simulates a system that recognizes the postures at a fixed position using a single CNN. For this benchmark, there are two 2D positions that are to the left and right of the fixed position.
- Benchmark 2 - Simulates a system that uses a single CNN to recognize the postures at multiple positions.
- Benchmark 3 - Simulates a system that uses a single CNN to recognize the postures at multiple positions. This CNN uses the estimated locations as input
- Benchmark 4 - Simulates a system that uses a CNN for each position to recognize the postures. The estimated location is used to select which CNN is used for the recognition.
- Adaptive system - Simulates the proposed system that uses two types of CNN for posture recognition. For positions close to a sensor a CNN is used that is trained on the data for that sensor and for other position a CNN is used that is trained on the data of all sensors. The estimated location is used to select which CNN is used for the recognition.

For each of these benchmarks, there are four results which are a combination of two parameters. The parameters are which sensors are used for the posture recognition and which positions are included in the data set. There are two options for the sensors used, which are the *main sensor* and *all sensors*. Only the main sensor is used in the first option and all sensors are used in the second option. The positions in a recognition area are split into two categories, *1D positions* and *2D positions*. The 1D positions are shown as blue in Figure 5.17 and represent the positions that vary in one dimension. The 2D positions vary in two dimensions and are equal to all positions in the recognition area, shown as black in Figure 5.17. The same figure shows the setups that are simulated in the different benchmarks.

5.8.3 Data sets

For the training and validation of the system, a sample set was captured using the setup shown in Figure 5.17a. A single volunteer was used to capture the posture data of all positions and postures. The postures were captured with the volunteer facing the main sensor at all positions. At each position, the volunteer was asked stand a the position and hold each posture for 20 seconds while occasionally making small movements. The small movements were used to introduce noise into the images. The movements for postures were the arms are not next to the body consisted of moving the arms up and down, and leaning from side to side. For the other postures, the only leaning from side to side was performed. The sample set captured using this method consists of 14040 heatmaps, 195 samples for each position and posture. These 24×8 heatmaps are the concatenation of the 8×8 heatmaps of the sensors based on their number in the sensor configuration.

This sample set was used to create the data sets for the adaptive system and the benchmarks used to compare its performance. The data set for each benchmark is split into a one-dimensional and two-dimensional version. For each of the data sets, a training, test and validation set was created with a split of 0.45, 0.05, and 0.5. These sets were created for each set of postures and an overview of them can be found in their respective tables, Table 5.13, Table 5.14, and Table 5.15.



(a) Benchmark 1

(b) Benchmark 2

(c) Benchmark 3

(d) Benchmark 4

(e) Adaptive system

Figure 5.17: Visual representations of the five configurations for the final posture recognition experiments. The different positions are shown as circles (blue 1d, black 2d)

benchmarks		**training set**	**test set**	**validation set**
Benchmark 1	1D positions	352	40	388
	2D positions	1058	118	1164
Benchmark 2	1D positions	1058	118	1164
	2D positions	3175	353	3492
Benchmark 3	1D positions	1058	118	1164
	2D positions	3175	353	3492
Benchmark 4	1D positions	352[1]	40[1]	1164
	2D positions	352[1]	40[1]	3492
Adaptive system	1D positions	1058[2]	118[2]	1164[2]
	2D positions	1058[2]	118[2]	1164[2]

[1] for each model/position

[2] for each sensor and the centre positions

Table 5.13: The training, test, and validation set sizes for all benchmarks with four postures

benchmarks		**training set**	**test set**	**validation set**
Benchmark 1	1D positions	529	59	582
	2D positions	1587	177	1746
Benchmark 2	1D positions	1587	177	1746
	2D positions	4762	530	5238
Benchmark 3	1D positions	1587	177	1746
	2D positions	4762	530	5238
Benchmark 4	1D positions	529[1]	59[1]	1746
	2D positions	529[1]	59[1]	5238
Adaptive system	1D positions	1587[2]	177[2]	1746[2]
	2D positions	1587[2]	177[2]	1746[2]

[1] for each model/position

[2] for each sensor and the centre positions

Table 5.14: The training, test, and validation set sizes for all benchmarks with six postures

benchmarks		**training set**	**test set**	**validation set**
Benchmark 1	1D positions	705	79	776
	2D positions	2116	236	2328
Benchmark 2	1D positions	2116	236	2328
	2D positions	6350	706	6984
Benchmark 3	1D positions	2116	236	2328
	2D positions	6350	706	6984
Benchmark 4	1D positions	705[1]	79[1]	2328
	2D positions	705[1]	79[1]	6984
Adaptive system	1D positions	2116[2]	236[2]	2328[2]
	2D positions	2116[2]	236[2]	2328[2]

[1] for each model/position

[2] for each sensor and the centre positions

Table 5.15: The training, test, and validation set sizes for all benchmarks with eight postures

5.8.4 Results

During the training of the CNNs for the benchmarks using the first validation posture set, it became apparent that the training parameters from the previous experiments did not result in the desired performance. The biggest drop in performance was seen in benchmark 3, having only an accuracy of 75%. Training this model with a higher number of epochs did improve the performance. So to keep the training steps equal to the previous experiments, a learning rate of 0.1 was chosen for all further training. This resulted in a performance that was more in line with the expected performance compared to the previous experiments, described in section 5.7. The results for each validation set are shown in Figure 5.18 for the validation set with four postures, Figure 5.19 for the validation set with six postures, and Figure 5.20 for the final validation set.

The accuracy when increasing the number of postures does decrease, but not in the manner that was expected. Recognizing the posture at the position in front of the main sensor remains the same, at 100%. It was assumed that the accuracy would decrease slightly because some of the postures were harder to recognize. It is odd that the accuracy of the front row of positions are close to that of the accuracy of all positions even though it contains fewer positions. With the eight posture version, it would seem that only using the main sensor for posture recognition has a larger negative effect on the accuracy of the recognitions which might be caused by an increase in noise with the data of the final to postures. Another possibility is that some postures are more difficult to recognize as they are similar to the other postures. What does remain true over all increments of postures is that a higher accuracy is achieved when using multiple models over only using the location as input for a single model. The finding of the sixth experiments, that using the adaptive system to achieve similar accuracy's as the multi-model, does not hold when adding more postures to the system. The accuracy of the adaptive system seems more in line with the single model system. Improving the performance of the adaptive system lies in the separation of the recognition area and the training of the models and the architectures used for those modes. It is likely that using the separation shown in Figure 5.14b would have a high accuracy as the separated positions are closer to the sensors giving the models more details to work with. Further improvements could be achieved by improving the architecture of the CNN for single sensor recognition and improving the training parameters for all CNN models.

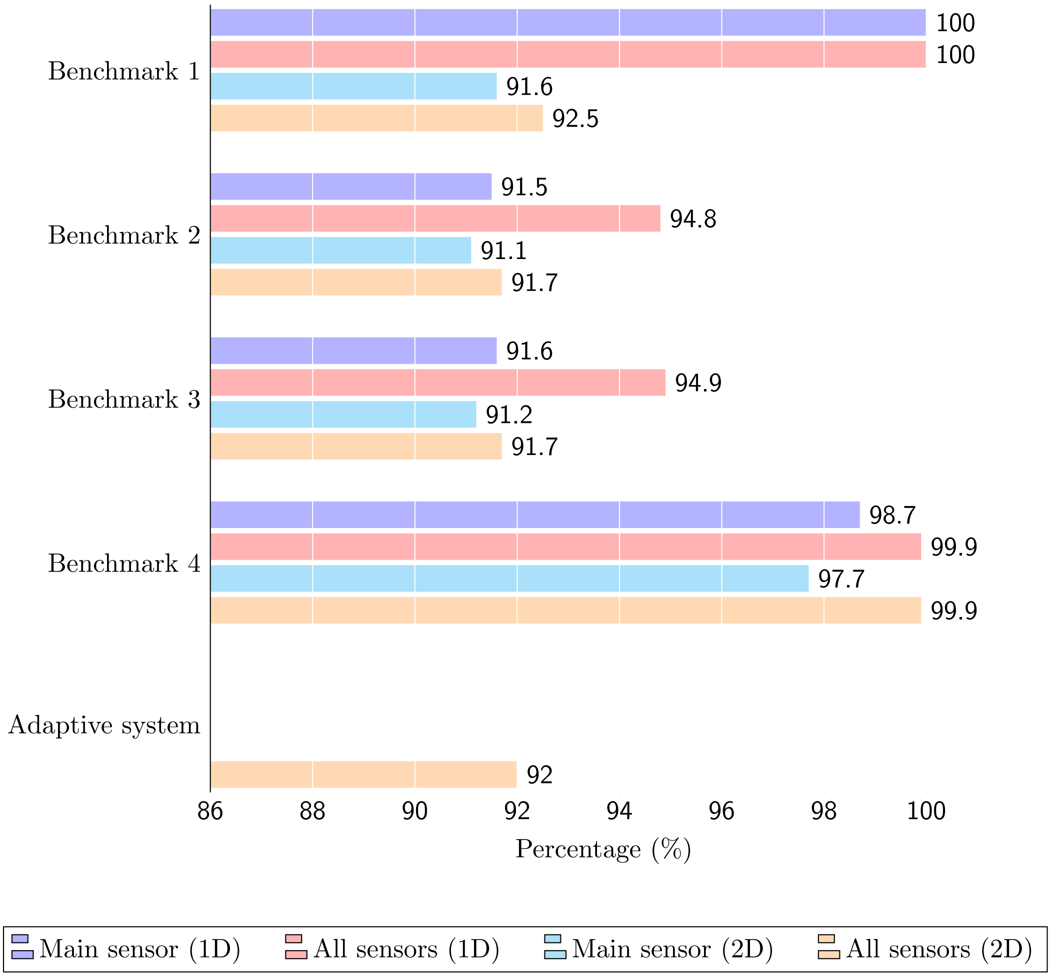


Figure 5.18: Recognition accuracy of the benchmarks for the four postures

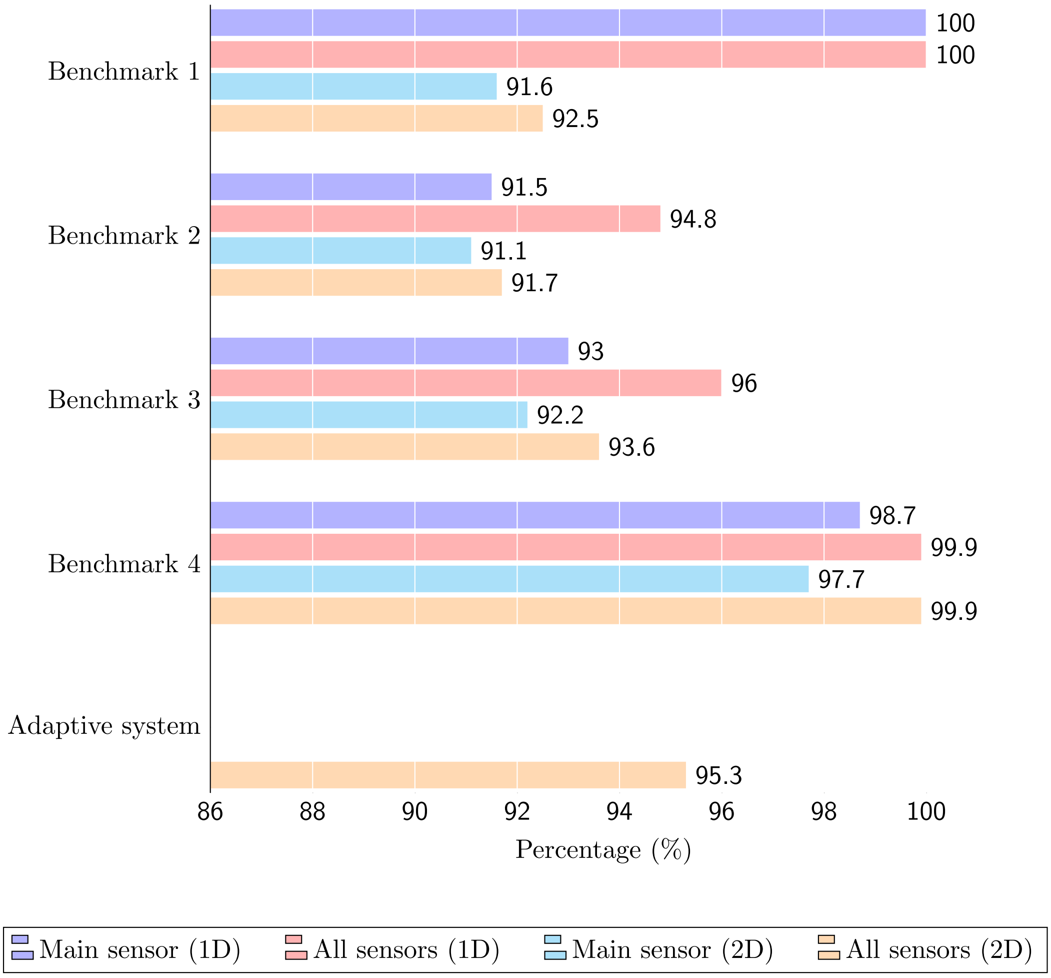


Figure 5.19: Recognition accuracy of the benchmarks for the six postures

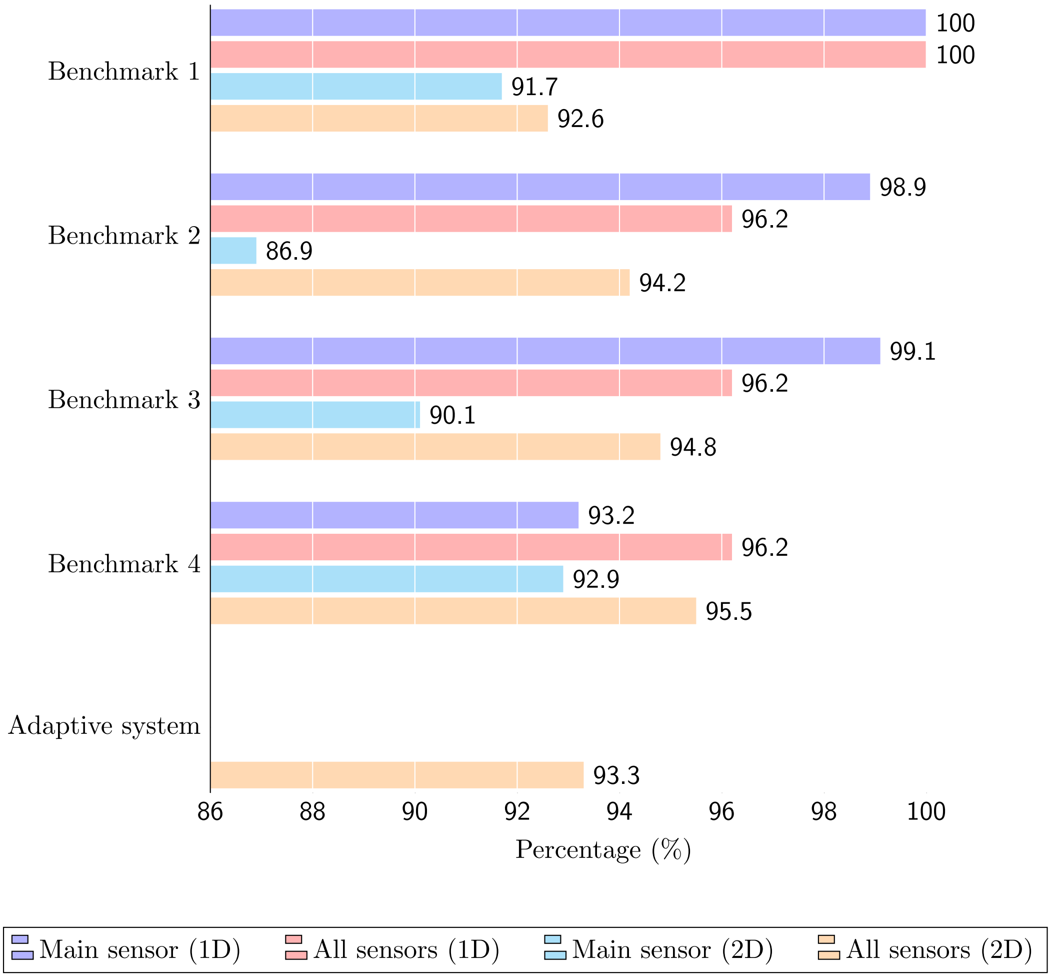


Figure 5.20: Recognition accuracy of the benchmarks for the eight postures

Chapter 6

Conclusions

In this chapter the work is evaluated, the observations are discussed and some recommendations are given for future work.

The aim of this research was to investigate the possibilities to use the location of the subject in combination with thermopile array sensors to provide accurate posture recognition within a room. The first task was finding the largest distance at which thermopile array sensor can be used for posture recognition. This distance was found to be 3 meters at which the noise and details of the posture were insufficient for posture recognition. Secondly, a localization algorithm was implemented that uses the heatmaps of multiple thermopile array sensors to estimate the location of the subject. This algorithm was improved by basing the Angle of Arrival on the hottest column and calculating the concentration of heat surrounding that column. Two system were designed and implemented that use the localization algorithm to improve the accuracy of the posture recognition. One system uses the estimated location as input for the ML model and the other system uses a ML model for each location and uses the estimated location to select which ML model to use for the recognition. To compare the performance of these two systems, two benchmark systems were designed and implemented that did not use the localization.

The final aim was to design a adaptive system were the posture recognition of positions close a particular sensor is performed on the resource-constrained devices housing that sensor. This system uses smaller models for these positions such that they can fit on these constrained devices. The performance of all systems are evaluated and compared leading to the conclusion that using a CNN model for each positions and selecting the correct one based on the estimated locations achieved the highest accuracy. Another important conclusion is that recognizing postures at an angle is possible, but further research is needed to be able to recognize these angled postures at any position in a room. The noise in the heatmaps of the thermopile array sensor at distances over 2 meter is limiting the performance of the posture recognition and research into reducing the noise or extracting the shape of the posture needs to be done. This should also improve the performance of the proposed system and should allow for larger recognition areas. The adaptive system has a lower accuracy than the system that uses all sensors for all positions, but it allows for distributed posture recognition. Further research into the adaptive system

The main contributions of this thesis are:

- Provides evidence that posture recognition at various positions in two dimensions is possible when using thermopile array sensors.
- Provide an implementation for a localization algorithm based on the output heatmaps of thermopile array sensors.
- Implement a system that uses the location and Convolutional Neural Networks for room-wide posture recognition.
- Show that using the location improves the accuracy of the posture recognitions.

- Design an adaptive system to use the location to reduce the complexity of the CNN models.
- Show that high accuracy can be achieved for posture recognition system using thermopile array sensors.

Bibliography

[1] Tianyu Bi, Tanir Ozcelebi, Dmitri Jarnikov, and Dragon Sekulovski. PCANN: Distributed ANN Architecture for Image Recognition in Resource-Constrained IoT Devices. In *2019 The 15th International Conference on Intelligent Environments*, 2019. 2, 4, 5, 8, 12

[2] Matthew Burns, Philip Morrow, Chris Nugent, and Sally McClean. Fusing thermopile infrared sensor data for single component activity recognition within a smart environment. *Journal of Sensor and Actuator Networks*, 8(1):10, 2019. 3, 4, 9, 12

[3] Chen Chen, Roozbeh Jafari, and Nasser Kehtarnavaz. A survey of depth and inertial sensor fusion for human action recognition. *Multimedia Tools and Applications*, 76(3):4405–4425, 2017. 2, 4, 11, 12

[4] Chen Chen, Kui Liu, and Nasser Kehtarnavaz. Real-time human action recognition based on depth motion maps. *Journal of real-time image processing*, 12(1):155–163, 2016. 2, 4, 8

[5] Giovanni Diraco, Alessandro Leone, and Pietro Siciliano. Human posture recognition with a time-of-flight 3d sensor for in-home applications. *Expert Systems with Applications*, 40(2):744–751, 2013. 2, 8

[6] FLIR. FLIR LEPTON Engineering Datasheet. Available at `https://www.flir.com/products/lepton/`, 2018. 22

[7] Munkhjargal Gochoo, Tan-Hsu Tan, Tsedevdorj Batjargal, Oleg Seredin, and Shih-Chia Huang. Device-free non-privacy invasive indoor human posture recognition using low-resolution infrared sensor-based wireless sensor networks and dcnn. In *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 2311–2316. IEEE, 2018. 3, 4, 5, 9, 12, 18, 22

[8] Google. Tensorflow. Available at `https://www.tensorflow.org/`. 23

[9] Rafik Gouiaa and Jean Meunier. Human posture recognition by combining silhouette and infrared cast shadows. In *2015 International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pages 49–54. IEEE, 2015. 2, 7, 12

[10] Akira Hayashida, Vasily Moshnyaga, and Koji Hashimoto. The use of thermal ir array sensor for indoor fall detection. In *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 594–599. IEEE, 2017. 3, 4

[11] Kyaw Kyaw Htike, Othman O Khalifa, Huda Adibah Mohd Ramli, and Mohammad AM Abushariah. Human activity recognition for video surveillance using sequences of postures. In *The Third International Conference on e-Technologies and Networks for Development (ICeND2014)*, pages 79–82. IEEE, 2014. 2, 4, 7, 11

[12] Ninghang Hu, Gwenn Englebienne, and Ben Kröse. Posture recognition with a top-view camera. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2152–2157. IEEE, 2013. 2, 7, 11

[13] Jian Huang, Xiaoqiang Yu, Yuan Wang, and Xiling Xiao. An integrated wireless wearable sensor system for posture recognition and indoor localization. *Sensors*, 16(11):1825, 2016. 2, 4, 9

[14] Jurgen Kemper and Holger Linde. Challenges of passive infrared indoor localization. In *2008 5th Workshop on Positioning, Navigation and Communication*, pages 63–70. IEEE, 2008. 13, 14, 15, 16

[15] Ye Liu, Liqiang Nie, Li Liu, and David S Rosenblum. From action to activity: sensor-based activity recognition. *Neurocomputing*, 181:108–115, 2016. 2, 4, 9

[16] Shota Mashiyama, Jihoon Hong, and Tomoaki Ohtsuki. A fall detection system using low resolution infrared array sensor. In *2014 IEEE 25th Annual International Symposium on Personal, Indoor, and Mobile Radio Communication (PIMRC)*, pages 2109–2113. IEEE, 2014. 3

[17] Melexis. MLX90621 16x4 IR array datasheet. Available at `https://www.melexis.com/en/product/MLX90621/Far-Infrared-Sensor-Array-High-Speed-Low-Noise`, 2016. 22

[18] Melexis. MLX90640 32x24 IR array datasheet. Available at `https://www.melexis.com/en/product/MLX90640/Far-Infrared-Thermal-Sensor-Array`, 2018. 22

[19] MQTT. Message queuing telemetry protocol. Available at `https://www.mqtt.org/`. 23

[20] Omron. Usage of D6T Thermal sensor. Available at `https://www.omron.com/`, 2013. 22

[21] World Health Organization, World Health Organization. Ageing, and Life Course Unit. *WHO global report on falls prevention in older age*. World Health Organization, 2008. 2

[22] Panasonic. Grid-EYE: state of the art thermal imaging solution. Available at `https://eu.industrial.panasonic.com/`, 2016. 22

[23] Orasa Patsadu, Chakarida Nukoolkit, and Bunthit Watanapa. Human gesture recognition using kinect camera. In *2012 Ninth International Conference on Computer Science and Software Engineering (JCSSE)*, pages 28–32. IEEE, 2012. 2, 8

[24] Kamal Sehairi, Fatima Chouireb, and Jean Meunier. Elderly fall detection system based on multiple shape features and motion analysis. In *2018 International Conference on Intelligent Systems and Computer Vision (ISCV)*, pages 1–8. IEEE, 2018. 2, 4, 7, 11

[25] Jamie Shotton, Toby Sharp, Alex Kipman, Andrew Fitzgibbon, Mark Finocchio, Andrew Blake, Mat Cook, and Richard Moore. Real-time human pose recognition in parts from single depth images. *Communications of the ACM*, 56(1):116–124, 2013. 2

[26] Anna A Trofimova, Andrea Masciadri, Fabio Veronese, and Fabio Salice. Indoor human detection based on thermal array sensor data and adaptive background estimation. 2017. 3

[27] United Nations, Department of Economic and Social Affairs, Population Division. World population prospects the 2017 revision, key findings and adance tables, 2017. 2

[28] Wen-June Wang, Jun-Wei Chang, Shih-Fu Haung, and Rong-Jyue Wang. Human posture recognition based on images captured by the kinect sensor. *International Journal of Advanced Robotic Systems*, 13(2):54, 2016. 2, 8, 12

[29] Allen Y Yang, Roozbeh Jafari, S Shankar Sastry, and Ruzena Bajcsy. Distributed recognition of human actions using wearable motion sensor networks. *Journal of Ambient Intelligence and Smart Environments*, 1(2):103–115, 2009. 2, 4, 5, 9