D2.1 – System Requirements Report

Due date of deliverable: 30/11/2021

Responsible partner: FBK

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement **No 965486**

1. OVERVIEW OF THE DELIVERABLE

Scope

The goals of this document are twofold. 1) To detail our investigations into emerging technologies and products in the domains of object recognition, people smart sensing, etc. in smart-supermarkets, and, 2) to outline the hardware and software architectural design requirements for a MIMEX installation, matching functional requirements highlighted in T2.2. We have highlight necessary optimisations to avoid system bottlenecks (e.g., sensors, data processing and user interface), and introduced sensitive issues that will need to be thought about for a practical solution, so that we are compliant with privacy protection legislation.

Audience

This deliverable has been created for a PUBLIC audience.

Summary

This report summarises work carried out in the period M1-M12, in the scope of T2.1. Outcomes from this deliverable have been fed into system implementation tasks: T3.1, T3.2 and T3.3.

Structure

In Chapter 2, we present the findings from our study into appropriate technologies for MIMEX. In addition to looking at what technologies our competitors are exploiting, we also go into detail about the hardware and software options available for MIMEX's components and discuss their advantages and disadvantages. In Chapter 3, we introduce the architectural design of MIMEX, in terms of its constituent components and practical interconnections between them. In Chapter 4, we draw some conclusions.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement **No 965486**

2. STATE OF THE ART

For automated product tracking and shopper check-out monitoring, any unmanned micro-market system needs to be able to automatically understand which products a shopper is picking up and placing into their baskets. This implies that 'pre-sales' product selection technologies must be able to understand exactly what items each unique shopper is picking up. A person's interactions with a shop's inventory must be monitored at all times, tracking physical interactions such as: what do the objects 'look like' that a shopper is handling, have there been any correlating 'weight' changes on smart shelves, or have any radio-based marker/tags that are attached to products moved their locations. For automated micro-markets to be attractive to shoppers, the process of product selection and the subsequent process of automated payment must be simple, intuitive, and very accurate.

Shopper activity tracking technologies need to be very robust to interference such as occlusions, variations in illumination, the redressing of a shop's aesthetics, etc.., otherwise detection errors will decrease shopper trust, and re-visits would diminish.

Shop retailers/managers also require a very high level of technological robustness and performance. The retailer must be able to trust automated product tracking and auto-purchasing technologies so that they can accurately manage their stock supply scheduling, and of course detect possible thefts.

In this section, we present findings from our study into current state-of-the-art technologies in this area, introducing the reader to a wide variety of appropriate components. The technologies that we investigated fall into the classes of: (AI Deep Learning) Vision, 3D Stereo Video Analytics, 2D Monocular & Fisheye Video Analytics, Thermal Imaging, Time of Flight (ToF), Lidar 3D Laser Scanning, UWB (Ultra-Wide Band) Radar Imaging, RFID (Radio Frequency Identification) Tags & Tracking and NFC (Near Field Communication). For each, we discuss their advantages and disadvantages for micro-markets and especially for MIMEX.

Non-vision tracking solutions

In this section, we summarise some of the non-vision solutions available for smart product-tracking and person-tracking inside shops, such as those that exploit radio frequencies. We will look at factors such as the number and types of sensors that are (or would be) needed for a reasonably sized micro-market deployment, and their suitability for different types of products on sale (considering factors such as: materials of packaging, size, weight, appearance, etc.), how they would need to be placed/arranged on shelves (i.e. might there be detection shadows in some areas - like refrigerators), how a sensor's range would affect installations, and other limitations there might be and what would be needed to get around these caveats. Stakeholder impact factors relating to these factors have also been discussed during our market analysis task in T2.2.

Radio Frequency solutions

There are four main categories of suitable radio frequency technologies for micro-markets - RFID, NFC, UWB and Bluetooth - each exhibiting different range and cost attributes, each having its own advantages and disadvantages. Factors such as unit cost, maintenance cycles, performance, robustness, and scalability vary for each solution, influence the choices of technology used, matching the type of product for sale and the environment.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement **No 965486**

RFID: Radio-Frequency IDentification uses electromagnetic fields to automatically identify, and track tags attached to shop products. An RFID system consists of a tiny radio transponder, a radio receiver and transmitter. This requires the attaching of passive tags to each product on a shelf with data incorporated into each tag. Each RFID tag can hold several data fields that include product numbers and information about each product/SKU (Stock Keeping Unit). In retail, RFID is primarily used for supply chain and loss prevention applications through the tagging and tracking of the movement of products inside a monitored space. As RFID adoption rate gains momentum, retailers are deploying more advanced RFID solutions to enhance their in-store customer experiences[1] - tracking more complex customer behaviours which can lead to valuable insights into shopping habits. The pairing of SKU interactions with customer data in real-time has many benefits for micro-markets. For example, in a checkout zone you can record product types as someone exits a store, which when combined with customer data like payment methods can determine whether buyers are returning shoppers. RFID tags can also be exploited to maintain and act on how customers move through a store. Decisions can be made by store owners about how to arrange store layouts and sharpen marketing tactics based on long-term movement analytics.

- Advantages:
  - *Range*: ~1m;
  - *Cost*: Basic passive RFID tags cost ~10cents, and can be used for paper, non-metal, and liquid materials.
- Disadvantages
  - *Information capacity:* Tags do not carry much data;
  - *Usage:* Basic passive RFID tags cannot be used on metallic products (e.g.: cans). Errors can occur when there are too many tags in the same area or with different types of packaging;
  - *Cost* - Metal passive RFID tags are larger and can be used on metal assets, costing ~1€. Active RFID tags are completely automated, but cost ~15€. Antennas in a shape format suitable for smart shelves are expensive, ~100€/shelf;
  - *Sustainability* - difficult to recycle, resulting in higher sales costs and sustainability issues.

NFC (Near-Field Communication): NFC is a set of protocols for communications between two electronic devices over a distance of 4cm or less. In a retail space, it would require the attaching of passive tags to each product on a shelf with stock data incorporated into it. One or multiple readers on the surface of a shelf then detect movements.

- Advantages:
  - *Dimensions*: small tags (less intrusive for products);
  - *Positional accuracy*: close range therefore location is known;
  - *Information capacity:* More data with respect to RFID, especially the rewritable memory;
  - *Usage*: Tags are small, so can easily be attached to products.
- Disadvantages:
  - *Range*: Very close (almost contact), so more readers per square meter are required.
  - *Cost*: Tag is slightly more expensive than basic UHF RFID (~10 cent), but readers are one order of magnitude cheaper than UHF RFID. On-metal tags cost ~1€.

[1] https://www.shopify.com/retail/5-examples-of-innovative-uses-for-rfid-technology-in-retail

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement **No 965486**

UWB (Ultrawide band): UWB is a radio technology that can use a very low energy level for short-range, high-bandwidth communications over a large portion of the radio spectrum. UWB has traditional applications in non-cooperative radar imaging and has been exploited for data collection, precision locating and tracking applications in the retail space.

- Advantages:
  - *Positional accuracy*: 5 to 10cm accuracy;
  - *Range*: ~100m line of sight, indoor ~25m;
  - *Availability*: already present on high-end Samsung and Apple smartphones, likely arriving on more phones in the near future;
  - Information capacity: none, only distance.
- Disadvantages:
  - *Cost*: ~80€ per device, anchors (on the shelf) and beacons (users);
  - *Usage*: Must attach the device to the user, by putting it in a pocket, on a lanyard, or on a shopping bag/cart;
  - *Dimensions*: Large ~15x10x5cm.

Bluetooth: Bluetooth (BLE) is a radio technology used for the transmission of data. Many devices like smart watches and wireless keyboards work with Bluetooth. A Bluetooth device can be used as a beacon to show its presence in a certain area. Existing applications have shown that it is possible to roughly measure the distance between devices by measuring the strength of the signal at the receiver side. A new Bluetooth version (BT5) is also emerging, with increased capabilities both in terms of localization (angle of arrival (AoA) measurements) and localization accuracy. As this technology is still relatively new, it is not prevalent on all consumer smartphones and devices that can exploit the benefits, especially of AoA are both expensive and only prevalent in development kits.

- Advantages:
  - *Adoption potential:* Bluetooth is included in most smartphones. Beacons are readily available so user tracking is possible;
  - *Range*: ~10m;
  - *Information capacity:* Huge amounts of data can be carried.
- Disadvantages:
  - *Positional accuracy*: ~50cm with costly infrastructure, otherwise ~2m.
  - *Usage:* it requires anchors to be installed in supermarkets - e.g. in the ceiling or in shelves.
  - *Cost*: the cost of anchors and development kit ~3000€ for a small shop. Low-cost beacons (~15€) are also available, but the cost is prohibitive for tagging individual objects.

Other sensors

In addition to exploiting radio technologies for tracking products and shoppers, the following approaches can also be exploited:

Pressure Sensors: They use resistance variations of an elastic material that has been doped with conductive particles (e.g., carbon/nanotubes). Resistance variations are one order of magnitude worse than traditional strain gauges. Currently, these types of pressure sensors are being exploited in a wide variety of retail applications. However, to measure small variation in weight for small items, the precision needed might make them prohibitively expensive.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement **No 965486**

- Advantages:
  - *Cost*: cheap and easy to fabricate;
  - *Form factor:* Very thin layout;
  - *Usability:* Output signal easy to read.
- Disadvantages:
  - *Usage:* Products of a similar dimension and weight from different prices/brands can be mixed up. Extra care needs to be taken about product placement;
  - *Precision:* Weight accuracy is very poor and hence location estimates from product movement events is very poor. 'Creep effect' of the materials can make it difficult to quantify static loads.

Load-cells/weight-sensors/scales: These rely on strain gauges, which exploit resistance variations in conductive wires when they get deformed by objects being place on top. Technically, they are actually deformation sensors, but when these devices are coupled with a support, where the relationship between force and deformation is known, a strain gauge can be used as a force sensor.

- Advantages:
  - *Precision*: Accuracy ~5g;
  - *Location*: An object's location can be determined by observing weight variations by the use of a single cell that measure an object's weight.
- Disadvantages:
  - *Usage*: Similar products of similar dimensions and weight from different prices/brands can be mixed up. Care needs to be taken about product placement;
  - *Cost*: ~40€/cell, but performance can suffer with low-cost devices. Signals from cells need to be increased using a differential low-noise amplifier which adds an additional ~10€/cell.

Distance Sensors: Sensors, such as ultrasound, infrared, photo, radar or laser, can be exploited to measure the distance between a device and an obstacle (such as a person or a product on a shelf). In the case where many products are piled up on a shelf, and a sensor is placed on the bottom, or above a shelf, such sensors can be exploited to estimate how many products (i.e. an approximate volume) are left in a location.

Radar: Motion and structural sensing can be achieved using new types of radar sensors. Radar technology functions by transmitting microwave or millimetre wave Radio Frequency (RF) signals to a target and then analysing the backscattering reflections to extract a target's movements.

- Advantages:
  - *Power*: Ultra-low power consumption. The radiation is only one-tenth of Bluetooth;
  - *Privacy:* No camera means that there are no privacy or other sensitive issues;
  - *Range:* Long detection distance is also possible, which makes them suitable for various installation heights.
  - *Robustness:* Not affected by environmental obstacles, such as smoke, dirt, low-light, heat sources, etc.
- Disadvantages:
  - *Precision*: Spatial resolution decreases with distance. Multipath reflections in complex environments (e.g. a supermarket shelf) can create confusing and indiscernible signals;
  - *Materials*: Different reflective (e.g. glass, metals, etc) or absorbent surfaces/materials (e.g. fabrics) can drastically degrade performance.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement **No 965486**

Ultrasonic: These sensors measure the time it takes for ultrasonic waves to travel to, and then be reflected from an object/person. The sensor is composed of a transmitter and a receiver. When the transmitter sends the ultrasonic wave, it starts to count the time until the reception of the same wave. Knowing in advance the speed of the ultrasonic wave, the sensor knows the distance between the sensor and the object.

- Advantages:
  - *Cost*: The cost of a single sensor ~1€;
  - *Precision*: Measurement accurate ~1cm for nearby objects.
- Disadvantages:
  - *Variability*: Some materials absorb ultrasonic waves, making accurate measurements impossible;
  - *Dimensions*: The angle and shape of a product can change the response of the sensor.

Infra-red: These sensors measure the time elapsed between the transmission and reception of an infrared beam. They can measure the distance between a sensor and the first obstacle the beam reaches. Infrared ranging sensors are generally exploited to determine whether a person is present.

- Advantages:
  - *Precision*: More accurate than ultrasonic sensors at short distances, ~1cm precision;
  - *Cost*: Cheap, ~10€/unit.
- Disadvantages:
  - *Variability*: Some materials do not reflect infrared radiation, so the signal will not come back to the receiver. Cannot detect stationary people/objects. Large atmospheric changes, e.g. air conditioner, ventilation fan, temperature differences between indoor and outdoor rooms, can cause false alarms;
  - *Range*: Can only operate under 2m distances;
  - *Accuracy*: The recognition of objects is not accurate. Prone to false alarms.

Photo sensor: This type of sensor generates an output proportional to the amount of light received. They are very common in mobile phones, where the phone understands the light condition to automatically increase the brightness or to lock the screen when we hold the phone close to the ear.

- Advantages:
  - Cost: <1€ per sensor.
- Disadvantages:
  - *Scalability*: A matrix of sensors are needed to understand if a product is taken;
  - *Robustness:* Light conditions must be very stable.

Lasers: Single-point ranging lasers, generally called LiDAR (Light Detection and Ranging), can be considered as a laser distance sensor. They measure the distance to targets through light waves emitted from a laser instead of radio or sound waves, and are a relatively novel sensing method.

- Advantages:
  - *Cost*: ~25€;
  - *Accuracy*: Distance estimates are extremely accurate, <0.03% error;
  - *Range*: ~10m.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement **No 965486**

- Disadvantages:
  - *Coverage*: A single-point ranging laser sensor can only cover a scanning cone of a few degrees of radius;
  - *Robustness:* Can be affected by light sources, exhaust fans, etc..

Gesture Sensors: These sensors are optic matrices chips and they can directly recognize 9 basic gestures: Up, Down, Left, Right, Front, Back, Clockwise, Counter clockwise, and Swing. They have a built-in infrared LED and optical lens that can work in low light and dark environment. They are suitable for smart homes, robot interactions, gesture detections.

- Advantages:
  - *Positional accuracy*: Depends on density of the sensors on a shelf and distance to target, typically ~1cm;
  - *Usage*: Can differentiate fine shopper interactions like the difference between grabbing a product from a shelf and/or leaving it on a shelf;
  - *Dimensions*: Small package size, ~3cmx3cm.
- Disadvantages:
  - *Cost*: ~20€.

Self-checkout Technologies

Some of the technologies described above can also be exploited to automatically verify the content of a shopping bag, when a shopper enters the checkout area. This feature can be used to confirm that the shopping list computed by the automated object tracking technologies tallies with what is actually in the bag.

UHF RFID: details of the technology is described in section 2.1.1.

- Advantages:
  - *Transparent interactions:* the user does not have to "push any button", they either place the bag on a checkout surface, or just walk through some smart gates;
  - *Reliable*: RFID technology is mature and has a low misdetection rate.
- Disadvantages:
  - *Range*: The detection range of the technology is not limited to near-field, due to the frequency band used; hence the problem is that multiple paths of radiation can go for distances of up to 10m. This can lead to a detection of shelf objects in the store, items in other shoppers' backets, etc., hence not only those in the shopper's bag.

We have conducted experiment to assess the practical use of this technology in MIMEX using metal and liquid products and 4 antennas in a gate-like configuration. Details of the experiments will be presented in a WP4 deliverable, but the conclusions are as following:

1. With a proper power configuration, all of the items in a shop can be detected. It is unlikely to miss items in a normal shopping bag.
2. The main limitation is that it can detect the presence of other tags in the local surroundings, and therefore leading to a wrong shopping list. To remedy this, the other items in the surroundings should be >4m from the check-out antennas.

Barcode based checkout: Barcode is an optical system that has been exploited in assisted checkouts since the 80s. A bar code is included on all products, with a neglectable cost incurred during the

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement **No 965486**

production phase. At the checkout, a reader then scans this code using a laser beam that hits the barcode and a sensor reads the reflected light.

- Advantages:
  - *Cost:* The barcode is already embedded on all the consumers products and therefore it has no cost added in the process;
  - Reader Cost: ~100€.
- Disadvantages:
  - *Direct visibility:* The reader and the barcode must be in line-of-sight and reasonably close (~1m), products must be placed at the correct angle in front of the reader, thus interfering with MIMEX's invisible-technology paradigm.

Millimeterwave ID (MMID): instead of using frequencies in the range of 800 MHz, Millimeterwave devices use frequencies between 30 and 60 GHz. It needs a reader and a transponder.

- Advantages:
  - *Bandwidth:* As it operates at a higher frequency, it has a higher bandwidth potential data-rate;
  - *High channel attenuation*: Millimeterwaves have a high atmospheric attenuation. This can be an advantage over RFID, as it only reads nearby objects;
  - *Small transponder size:* The size of transponders is reduced thanks to the smaller wavelengths. As a consequence, tags are easier to embed in products.
- Disadvantages:
  - *Cost:* The reader requires a high frequency and relatively high-power oscillator and therefore it costs more than UHF solutions;
  - *Not mature technology:* The field of MMID is still in a preliminary exploration and there are only a few papers presenting such systems. Therefore, there is no "out of the box" product available.

UWB-RFID: This is a combination of RFID technology and UWB technology that enables the detection and localization of tagged objects. The system can understand whether a tagged-object or device is inside a shopping bag or in the surroundings.

- Advantages:
  - *Precise location:* An UWB tag provides information about time of flight, therefore a tag's position can also be estimated if there are at least 3 readers to triangulate the signal.
- Disadvantages:
  - *Not mature technology:* This is a recent research field and literature is very limited and based often on simulations. The challenging part is the construction of a passive UWB tag that can provide information about the time-of-flight (i.e. distance).

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement **No 965486**

AI Vision solutions

In terms of customer engagement analysis, vision AI is becoming a game-changer for in-store customer tracking. Some commercial examples of what AI vision can do are: Modcam[2] (Sweden), Aura Vision Labs3 (UK), and C2RO4 (Canada). AI vision is opening up the option for the cost-effective and highly accuracy capturing of customer behaviours throughout stores. The technology can generate location and time metrics, facial demographic attributes, product selection estimates, shopper engagement data, and much more. Shopper and product tracking technologies can generate a wide variety of actionable insights that can increase the profits of retail stores. The tracking of people as they shop can generate quantifiable behaviour motion patterns (location and time), activities (such as gaze tilting and eye-tracking), and attributes (such as facial demographics).

Vision-based solutions, however, are often sensitive to illumination, and camera placement in an unobtrusive way is not always straightforward. Visual tracking networks involve the installation of multiple types of cameras in various locations inside shops, that then need to be calibrated to a common reference system so that the information they gather can be integrated to make global decisions. AI Vision exploits a set of Deep Learning algorithms. Because the technology requires the training of the algorithms to solve specific problems, solutions tend to specialize on particular aspects of retail.

People tracking

Government regulations (e.g. GDPR) and consumer awareness of privacy mean that any visual tracking system that exploit the deployment of visual sensors in public spaces must take care of ethics, privacy and data security. For this reason, MIMEX exploits @Edge technologies that store raw camera-data locally as much as possible and then transmits only some anonymous metadata to central servers on the cloud. There is a big difference between tracking technologies that identify shoppers by name and those that track people anonymously. For solutions such as MIMEX, it is important to ensure that a person picking up products and putting them into their bags remain consistent, i.e. a shopper's backet is associated to a single shopper. The identity (i.e. the real name of the shopper) is not important, but the person tracking system must maintain the same machine generated ID throughout the store, especially when passing from one tracking cell with shelves to another, as is the case in larger micro-markets. Vision technology can exploit parameters derived from clothing colours, person height, facial features, etc, to consistently track a person's unique identifier and their location. The choice of cameras exploited depends upon the task required. In the following bullet points we explore briefly some camera solutions available:

- ***Stereo cameras***, often these cameras have embedded processing, and they can capture the world in three-dimensions. 3D sensing capabilities can help to compensate for occlusions and shadows by exploiting depth information. Stereo cameras can produce a high-level of tracking accuracy with only a single unit, facilitating the tracking of complex behaviours such as high-volume traffic, queue management, and customer engagement. Premium 3D stereo sensors can go beyond people counting and path analytics, and they can also provide information on gender estimation, group

[2] https://www.modcam.com/
[3] https://auravision.ai/
[4] https://c2ro.com/

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement **No 965486**

counting, and gaze direction. Commercial examples of 3D Video Analytics providers include TDI[5] (Singapore), Hella[6] (Germany), Eurocam[7] (France), and Xovis[8] (Switzerland).

- ***Monocular sensors*** can also be exploited for people tracking, either individually or configured into overlapping visual networks. A monocular fisheye lens is ideal for confined spaces, like those experienced in micro-markets as they offer an ultra-wide-angle panoramic picture, albeit distorted. Because monocular sensors are cost-effective and Fisheye cameras are a standard in security and surveillance scenarios, 2D people tracking solutions are widespread, especially for counting. Axis[9], Bosch[10], and Panasonic[11] sell embed video analytics solutions inside a range of their smart cameras.
- ***Thermal cameras*** can also be exploited to track people. Since thermal technology is less sensitive to ambient light changes, it can function in any physical space almost independent of illumination. The challenge is the identification of a person's heat signature from the surrounding environment, especially in areas with solar illumination or other warm objects. Thermal sensors are easy to install and calibrate. Solution providers include Delopt[12] and Irisys[13].
- ***Time of Flight cameras*** can also be exploited for tracking people in micro-markets. They exploit a type of sensor that emits and then measures the time it takes for light to travel between the camera and an object. By sending the laser beams in many directions, the sensor can track the exact positioning of objects. BEA Helma[14], have developed an embedded people counting camera system.

Object detection

Object (i.e. shop product) detection using computer vision requires systems to automatically recognize and track objects as they are taken off, or replaced, on supermarket shelves. Some of the more important aspects to be considered in this challenge are:

1. **Adding new products** to the detection inventory in a systematic and minimal time-consuming way. *Approach: Exploit recent advances in machine learning to facilitate the faster training and adding of new objects to a system's database;*
2. **Low hardware requirements.** *Approach: Exploit next-generation edge hardware to run more complex neural network models in parallel, on greater numbers of cores making it possible to share hardware across multiple shelves;*
3. Obtaining a **high-level of detection accuracy**. *Approach: increase the number of low-cost cameras to better handle customer self-occlusions.*

[5] https://www.tdintelligence.com/
[6] http://www.hella.com/microsite-electronics/en/Sensors-94.html
[7] https://3dpeoplecounting.com/
[8] https://www.xovis.com/en/xovis/
[9] https://www.axis.com/ro/en/solutions-by-application/people-counting
[10] https://youtu.be/OXazoQF6eMI
[11] http://security.panasonic.com/products/functions/business_intelligent/
[12] https://www.delopt.co.in/people-counting-system.html
[13] https://www.irisys.net/
[14] https://www.beainc.com/en/technologies/

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement **No 965486**

Some examples of commercial solutions that track objects using AI-vision are: Amazon Go[15], and Walmart's Intelligent Retail Lab[16]. Intelligent Retail Lab exploits deep learning and cameras to automatically detect the out-of-stock products and alert staff members when to restock. There are a number of intelligent retail facilities, such as automatic vending machines and self-serve scales, that have emerged recently. A Chinese company, DeepBlue Technology[17], has developed automatic vending machines and self-checkout counters based on deep learning algorithms, which can accurately recognize commodities using cameras. Malong Technologies18 is another well-known business in China that aims to provide deep learning solutions for the retail industry. The solutions from Malong Technologies include AI Cabinets that perform automatic product recognition using computer vision and AI Fresh that enables the identification of fresh products on a self-service scale in an autonomous way. This interest from big players indicates that there is interest and great potential in unmanned retail solutions using deep-learning/AI computer vision. In the following paragraphs, we explore the more critical aspects relating to the exploitation of AI vision for product tracking and we offer a list of state-of-the-art approaches whose application can improve upon the current market approaches.

Adding new products to a shop's inventory. This procedure must be cheap, fast and simple, as a reasonable number of objects on sale in a typical micro-market's product range will be anything from fifty to several hundred. To make things more complicated, the packaging of some products changes often, and micro-market managers frequently change product display patterns and area displays. This implies that algorithms used for object recognition need to handle lots of products in different lighting conditions, and that they must be easily extendable to recognize new ones. The most effective approaches to solve this problem are supervised-algorithm based, i.e. new objects are learnt through a process of a person showing different views of a product to a camera and then telling the algorithm what it is seeing. In this challenging domain, MIMEX will attempt to reduce the number of images needed to be manually collected, thus creating a more efficient training process. We will achieve this target through the following means:

- ***Exploiting synthetic datasets.*** We will exploit particularly promising synthetic-dataset creation approaches that take 3D models of an object/product and then generate many new 2D views of the same object with simulated lighting conditions.
- ***Using algorithms that need small datasets to be trained:*** Many efforts are being carried out by the scientific community to develop new algorithms that can learn to recognize objects starting from small numbers of samples. An active subfield of this research in machine learning is focussed on the so called 'few-shot learning' (see recent survey[19]). The goal of this area goal is to exploit prior knowledge about the domain to artificially enrich the set of available samples for poorly represented classes, and/or restrict the class of models of where to search for an optimal classifier in order to increase the reliability of the results even in presence of few sample classes. Recent proposed approaches include data-augmentation, embedded learning, transfer learning, generative modelling and meta-learning.
- ***Automatically collect training data using a sensor fusion approach***

[15] https://www.amazon.com/b?ie=UTF8&node=16008589011
[16] https://www.intelligentretaillab.com/
[17] https://en.deepblueai.com/
[18] https://www.malong.com/en/home.html
[19] Yaqing Wang, Quanming Yao, James Kwok, Lionel M. Ni, "Generalizing from a Few Examples: A Survey on Few-Shot Learning", 2019 (arXiv:1904.05046v3)

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement **No 965486**

Maintain **low-cost hardware requirements:** deep-learning approaches have been proved to be a huge improvement in visual processing tasks. However, this improvement has come at a cost: usually the hardware necessary to run them is much more expensive than what was previously required. This is a problem in many applications where the business model does not have a sufficient financial margin to absorb the cost of the hardware necessary to run them. MIMEX is one of those cases, and therefore we are investigating new solutions that will enable us to reduce the hardware requirements necessary to run machine learning based, object detection algorithms. To obtain this expected result, we are investigating some of these promising approaches:

- **Use ad-hoc hardware:** AI accelerators are a class of specialized hardware accelerators designed to drastically improve the execution of ML and AI applications, by leveraging dedicated designs with many-core processors and low-precision arithmetic. This approach limits costs, as hardware requirements are trimmed to essential components only to perform specialized tasks. The common solutions come in the form of a single chip, or system on chip (SoC). The amount of energy consumed by these chips tends to be constant, plus it is lower in comparison to general-purpose architectures that are more energy-hungry when facing computational-heavy loads. In recent years, companies have focused on the design of AI accelerators (mostly based on ASIC and FPGA architectures) to respond to an increasing request from the market of dedicated computing capabilities for algorithms applied to IoT, robotics and other data-intensive tasks. In this context, it is not only a chip's capabilities that make the difference, but also the ecosystem that supports the technology. The software supporting the accelerator must be iteratively updated to reflect a smooth integration into development workflow, and thus it must be compatible/ portable with existing efforts in solution development. Software platform updating makes their use sustainable, limiting risks that the chips will become a bottleneck that hampers performance or forces the discontinuation of products. In MIMEX, we will explore the potential of such technologies from a pool of top AI accelerator providers, including Google, Xilinx, Intel, and Qualcomm, as replacements for the current GPU setup.
- **Reduce the dimension of the deep learning network:** TinyML is a field of study in Machine Learning and Embedded Systems that explores a subset of models that can run on small, low-powered devices like microcontrollers. It enables low-latency, low-power and low-bandwidth model inference on edge devices. The scientific community is actively working on novel methods to automate the compression of convolutional neural networks, while preserving elevated levels of accuracy in the target application. In MIMEX we aim to apply these methods to the network used in our object detection algorithm to understand if we can reduce the hardware requirements with a minimal impact on performance.

Covid restriction monitoring solutions

In the past year a relevant number of systems have been proposed on the market for the (semi-) automated control of different aspects that contribute to the prevention of diffusion of the Covid virus. Most of them are integrated system that are composed of a smartphone-like unit that includes cameras, speakers, processing unit, output port, etc.. They are proposed for access control applications: checking the temperature of people, verifying the correct wearing of a protective mask that often includes a face recognition/verification function. The following a list is a brief summary of such systems:

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement **No 965486**

ZKTeco Access Control with Mask Detection[20] (from Spain) provides mask wearing control, temperature control and face identification. The main characteristics are:

1. Face detection and identification (with or without wearing a face mask) vs face-data stored in a database;
2. Mask detection that supports a tolerance angle of 30-degree for facial recognition at distance up 2.5 meters;
3. Response time (from screening to display) is typically about 1.5 secs;
4. Infrared temperature detection to provide accurate and fast temperature screening during identity verification within 0.3/0.5ºC deviation range from 0.5 m away;
5. Installation on wall or desktop, integrated into turnstiles and speed gates or simply with a pedestal on floor.

COMMENDED - Automated Temperature and Face Mask Control[21] (from Austria) is a package consisting of a freely placeable pillar with built-in touch screen display unit, complete with two cameras (thermal and video). The solution interfaces seamlessly with existing door opener mechanisms to provide automated control access. The scanning process takes a few seconds, and the result is shown on the screen. Persons who are not wearing a face mask are requested to do so by acoustic and visual prompts. In case the scanned temperature is in the fever range, the device automatically triggers a call to service staff. In case of problems, users can also initiate a support call via the terminal. Main characteristics are:

1. Easy plug-and-play installation using regular power and network connections;
2. All components are IP capable;
3. Voice control with additional visual prompts to ensure correct face position;
4. Face mask recognition can be disabled as needed;
5. Integrated control contact for automated actuation of turnstile gates, doors, barriers, etc.;
6. IP-based query terminal with TFT touch screen for user interaction and video display.

KTC Technology Solution[22] (from Italy) proposes a system for body temperature measuring, mask wearing control, and facial identification. Main features: temperature accuracy ±0.3℃, display 7” touchscreen IPS, TCP/IP communication, a steel support included. The current price is about 1.200,00€.

TM18F00 Facial temperature and mask detection[23] (from Italy) is a biometric access control system based on the control of face, palm and fingerprint. It includes facial temperature measuring (from 34°C to 45°C +/- 0,3°C) and mask detection functionalities. Other features: Wi-Fi, Ethernet and RS232 connections, colour and infrared sensors, 5” colour display touch screen TFT, LED illumination, wall mounting bracket. The current price is about 1.000,00€.

THERMALFACE[24] (from Singapore) system has an in-built temperature measurement device, which is capable of measuring up to an accuracy of ±0.3°C. Even when individuals are wearing masks, ThermalFace is capable

[20] https://zkteco.eu/news-center/news/zkteco-access-control-mask-detection
[21] https://www.commend.com/applications/automated-temperature-and-face-mask-control.html
[22] https://ktcshop.it/termoscanner/4225834-dispositivo-rilevazione-temperatura-corporea-riconoscimento-facciale-supporto-verticale-face-temp-kit-.html
[23] https://www.idcolor.it/controllo-facciale-temperatura-e-mascherina.html
[24] https://www.intercorpsolutions.com/thermalface-thermal-scanner/

This project has received funding from the European Union’s Horizon 2020 research and innovation program under grant agreement **No 965486**

of recognizing them. Mask wearing can be set as a mandatory requirement when entering the premises, or it can be set as a reminder should individuals forget to wear one. CDS[25] (from USA) proposes a stand-alone Kiosk/Access Control Unit with the following features: face recognition access control terminal with digital detection module and no-contact wrist scanner for temperature measurement. Main characteristics:

1. Face recognition accuracy >99% (detection range from 8" to 9.5', detection time ~0.2s);
2. Face mask detection;
3. Non-contact wrist scanning temperature measurement (1–3 cm);
4. Door control options with video recording and two-way audio;
5. 7" Touch-capable screen.

Mooving[26] del Gruppo Ambita (from Italy) proposes a set of integrated systems for access control to measure facial or wrist temperature, check mask wearing and face recognition. SecurItaly srl (from Italy) proposes iAccess ScanFACE V2[27] a touchless thermo-scanner with face mask recognition. It supports:

1. Body Temperature Check;
2. Checking the use of the mask;
3. Face recognition (optional);
4. Door open relay;
5. Multilingual speech synthesis.

The current cost should be about 1.000€. Other companies propose apps to be installed on standard mobile devices or integrated in standard surveillance systems providing mask wearing control, social distancing control and, if a thermal camera is available, temperature measuring. Asura Technologies' mask detection software[28] (from Hungary) is a video analytics solution which detects in real-time whether people are wearing masks even in crowded environments. For the operator's convenience an automatic notification can be associated with any of the event types, e.g. the absence of a mask on a person. The TRIDENT Face Mask Detection System[29] (from India) is composed of a standard surveillance camera combined with Trident Computer Vision platform to detect people without masks. It can be integrated with existing surveillance setups and extended to monitor social distancing. It provides real-time reports and produces statistical data. The a2-VCA Mask Detection[30] (from Turkey) system detects people who do not wear masks in open or closed areas where people come together. Cameras are placed in the area to be checked, a2-VCA starts the camera stream and analyse the images. For the non-masked person, a2-VCA captures a picture and video, generates an alarm, and reports to the manager or security staff. The DeepSight SDK[31] (from the Netherlands) is a face analysis library based on advanced proprietary deep-learning models with best-in-class accuracy and performance. DeepSight can detect faces from a large variety of angles and distances. It includes also a mask detection function. Mobotix proposes AI-FaceDetect Deep (Protect Areas and Detect Masks[32]) a system for

[25] https://www.cdsofficetech.com/fixed-video-surveillance/thermal-imaging-temperature-scanning/
[26] https://www.mooving.eu/speciale-covid19-dispositivi-sicurezza/
[27] https://www.iaccess.eu/termoscanner-e-igienizzatori/145-termoscanner-iaccess-scanface.html
[28] https://asuratechnologies.com/products-and-solutions/video-analytics-solutions/mask-detection-software/
[29] https://www.tridentinfo.com/face-mask-detection-systems/
[30] http://a2technology.co/en/products/covid-19-en/a2-vca-mask-detection-system/
[31] https://sightcorp.com/deepsight-sdk/
[32] https://www.mobotix.com/en/protect-areas-and-detect-masks

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement **No 965486**

people counting based on face detection and mask detection control, that can work both in indoor and outdoor environments.

3. MIMEX's Architecture Design and Requirements

In this section, we will detail our chosen design for the system architecture, going into details about all of their functional requirements. As we are building upon existing components (primarily from the SpinRetail Intelligent shelf and the SmarTrack person tracker) that have been developed over the past few years by three of the five partners: SPXL, CEFLA and FBK, we are able to bootstrap MIMEX with a tried and tested (albeit currently limited) solution. These components will form the core technological backbone which we will then make more robust through the integration of additional sensing technologies explored in Section 2.

Technological Components

The MIMEX solution is composed of three technological components:

1. Smart racks;
2. A people tracker;
3. Covid-19 module.

Here we briefly present the role of each component.

Smart Racks

A smart rack houses and tracks the movement of products to be sold. It continuously monitors the products being taken and released on its shelves using data generated by its inbuilt weighing scales and cameras.

Figure 1 – Representation of MIMEX's Smart Racks in action

The smart rack can detect which objects, from a pre-trained set, (i.e. a known shop inventory) have been removed or placed. It relies on two different kinds of sensors working in harmony: ***weight sensors*** (one per shelf) installed underneath, and an ***RGB camera*** (one per shelf) that is installed above.

The ***camera*** continuously streams images of the area in front of the rack to a computational unit. Events generated by the scales and the vision module are merged by a sensor fusion module. In this way events can be enriched with information coming from both kinds of sensors.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement **No 965486**

People tracker

The people tracker is a multi-camera system. It computes the location of people via a particle probability approach, and is designed for scenarios where frequent and persistent occlusions occur, e.g. inside a micro-market with more than one person shopping. The system tracks each person during their entire visit, enabling the system to match picking-up and releasing product events (generated by the smart racks) with a user's identity. We have adopted a vision-based approach for this, which utilises multiple low-cost monocular cameras. We have investigated two possible solutions:

1) SmarTrack: FBK tracking technology

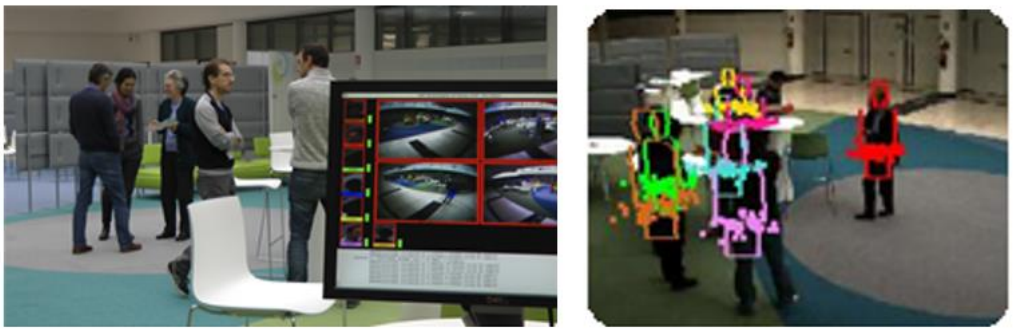

Figure 2 – SmarTrack, people tracking module in action

SmarTrack (patented by FBK) is a video tracking technology that provides accurate real-time information about the spatial location of people using a number of time-stamped image streams. Patents:

- Method and apparatus for tracking a number of objects or object parts in image sequences (EP1879149B1, US7965867B2)
- Method for efficient target detection from images robust to occlusion (EP2302589B1, US8436913B2)

- For each target entering a monitored scene, an appearance model is generated that is comprising of shape and colour information. This model is then tracked across the monitored scene using probabilistic representations and context information. An outstanding feature of SmarTrack is that it can handle large, persistent occlusions among several targets and known obstacles at an affordable computational cost. The probabilistic nature of the tracker mean that it can deal with ambiguity and uncertainty in a computationally efficient way. The resolution of probabilistic estimates is dynamically, and is automatically adapted by the system during use. The appearance models are descriptive hence they preserve a target's identity over time, a feature that is essential in many final applications.

2) Deep learning-based people tracking

Deep-learning people tracking modules can be divided into two categories: human-pose estimation and multi-object graph-based tracking. These types of modules perform detection and tracking on individual

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement **No 965486**

image planes (i.e. for each camera independently) and the result from each of the multiple viewpoints is then fused to obtain pose in 3D. The tracking of people and their joints in public places is challenging from viewpoints. The figure below shows a person that has been detected from different viewpoints and their pose has been reconstructed in 3D through the fusing of information using multi-view geometry techniques.

Figure 3 – Multi-view, deep-learning, people tracking in action

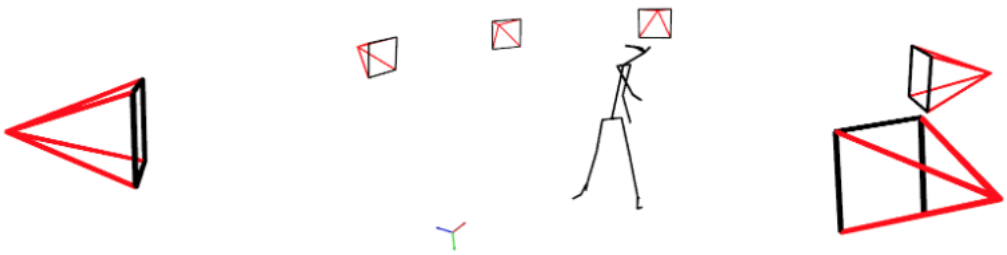

Figure 4- Results from the fusing of the several different human pose estimations

In MIMEX we will explore the use such an approach, but targeting crowded scenarios (i.e. people shoppers occluding other shoppers), thus the system will have to handle the challenge of discriminating target identities over time.

Covid-19 safety measure detection

Covid-19 safety measure require users to match some constraints and perform some actions before entering a store: e.g. to wear a mask, to have a normal body temperature, to sanitize their hands, etc..

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement **No 965486**

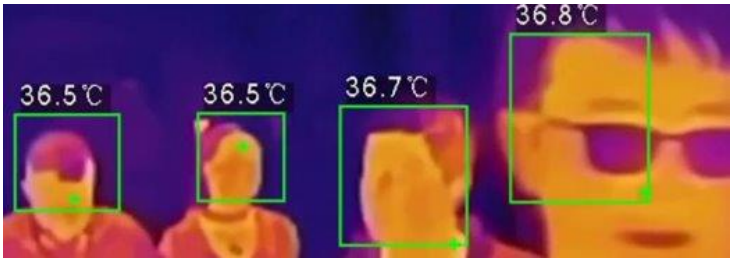

Figure 5- Example of face tracking and thermal temperature sensing

For the monitoring of Covid-19 safety measures, we do not have an in-house solution and thus we have looked for a third-party one the satisfies the following requirements:

Sanitizing hands

We will use an IoT sanitizer at the entrance of the mini-market. The sanitizer sends a message to the system at each usage. If the user does not perform this action, the mini-market door will not open until the system receives a message from the sanitizer that it has been used.

Body temperature and mask

At the entrance we will install a system with a thermal camera that will acquire images to detect body temperature and to understand if a person is wearing (or not) a mask (e.g. ProFace X from ZKTeco[33]). Only if the person has a temperature within an acceptable range, and if they are wearing a mask, will the door to the market open.

Installation and Maintenance Requirements

For any complex installation, like we are proposing in MIMEX, there is a need to not only design the technologies, but also to install and maintain them. Listed below are some of the major skills and competencies that need to be considered:

- Installation expertise;
- Configuration and reconfiguration of the shopping space to maintain performance (i.e. not cramming too many products into one area, or creating black-spots due to poor lighting);
- Maintenance of a computer network infrastructure (routers, switches, cables, Wi-Fi);
- An ability to troubleshoot network problems on computers running Linux;
- Understanding software configuration procedures;
- Installation, testing and maintenance of RFID antennas;
- Installation, calibration and maintenance of USB or wireless cameras (e.g. RGB, ToF, infrared, etc)
- A maintenance stock-taking system that can track/store spare-parts for quick maintenance of the market in case of failure;
- Installation and reconfiguration of shelfs;
- The adding of new products to the trained list of saleable stock in a store;
- Creation of the physical mini-market structure from prefabricated elements/modules.

[33] https://zkteco.eu/products/body-temperature-measurement/proface-x-ti

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement **No 965486**

Identified Bottlenecks and Optimisations

In this section we briefly highlight some of the necessary optimisations formulated to avoid bottlenecks, based on current observations relating to sensors, data processing and user interfaces:

- ***Person density*** – people in the shop will need to be regulated to a person/area limit (exact number to be decided during the pilot phases).
- ***Data processing for deep learning CPU/GPU requirements for person/object tracking*** – Scalability will be handled by dividing the entire micro-market shopping space into cells/zones. Inside a cell, people will be tracked. When a person leaves a tracking area, a hand-over procedure will be instigated and the tracking will then be conducted by the processors in the new cell. In this way, tracking processing will be less complex and more robust.
- ***Object tracking –*** The number and type of objects to be track will be balanced with tracking robustness and complexity. Real life experiments will be conducted to understand this trade off. Additionally, RFID tags will be exploited if the weight and visual systems suffer a decrease in performance due to object similarity, variability, or if there are a huge number of items to track.

Potential Issues and Compliance with Privacy and Ethics

The process of defining system hardware components and integrated software modules, and monitoring their operation requires us to bear in mind stakeholder needs too (this info will come from T2.2). This relates to issues like usability and ergonomics, as well as privacy and ethics issues (this will come from WP7). Moreover, to create a scalable solution, we also had to consider modularity & scalability issues (explored in T2.3) and cost effectiveness factors (explored in WP5).

4. Conclusions

In this deliverable we summarised our work carried out in the period M1-M12 (in the scope of T2.1) and detailed our investigations of emerging technologies in the domain of smart supermarkets (e.g. object recognition, people smart sensing, etc.). We looked into emerging commercial solutions and also at the wide spectrum of potential options coming from advanced research (e.g. software/hardware sensors, communication protocols, components, etc.). We looked at what technologies our competitors are exploiting and went into detail about the hardware and software options available for MIMEX's components, discussing their advantages and disadvantages.

We detailed our design for MIMEX's system architecture (based on the outcomes of our study) in terms of constituent components and practical interconnections between them, with an eye on user functionality requirements identified in T2.2 (where we gathered opinions and thoughts from potential stakeholders).

We also highlighted necessary optimisations to avoid bottlenecks (e.g., sensors, data processing and user interface) and introduced sensitive issues that will need to be thought about for a practical solution, so that we are compliant with privacy protection legislation.

Outcomes from this deliverable have been fed into system implementation tasks: T3.1, T3.2 and T3.3.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement **No 965486**