H2020 FRAMEWORK PROGRAMME
ICT-01-2014: Smart Cyber-Physical Systems

PROJECT NUMBER: 645496

Agile, eXtensible, fast I/O Module for the cyber-physical era

D3.1 – Scenarios and requirements Report

Due date of deliverable: 31st January2016
Actual Submission: 9th February 2016

Start date of the project: 1st February 2015 Duration: 36 months

Lead contractor for the deliverable: UNISI

Revision: See file name in document footer

Project co-founded by the European Commission within the HORIZON FRAMEWORK PROGRAMME (2020)	
Dissemination Level: PU	
PU	Public
PP	Restricted to other programs participant (including the Commission Services)
RE	Restricted to a group specified by the consortium (including the Commission Services)
CO	Confidential, only for members of the consortium (including the Commission Services)

Change Control

Version#	Date	Author	Organization	Change History
0.1	09.01.2016	Antonio Rizzo	UNISI	Initial version
0.2	15.01.2016	David Oro	HERTA	Added SVS
0.3	22.01.2016	Nicola Bettin	VIMAR	Added SHL
1.0	28.01.2016	Antonio Rizzo	UNISI	Final version
1.1	02.02.2016	Antonio Rizzo	UNISI	Appendixes A, B, C

Release Approval

Name	Role	Date
Antonio Rizzo	WP Leader	02.02.2016
Roberto Giorgi	Project Coordinator for formal deliverable	02.04.2016

The following list of authors will be updated to reflect the list of contributors to the document.

Antonio Rizzo, Francesco Montefoschi, Sara Ermini
Department of Social, Political and Cognitive Sciences
University of Siena – UNISI – AXIOM

David Oro
RD department
Herta Security – HERTA – AXIOM

Nicola Bettin
RD department
VIMAR spa – VIMAR – AXIOM

Printed in Siena, Italy, Europe.
Part number: *Please refer to the File name in the document footer*.

TABLE OF CONTENTS

Executive summary

This document reports the production of two scenarios, one for each domain, Smart Home Living (SHL) and Smart Video Surveillance (SVS), and the related requirements. The scenarios definition was the output of the following process: i) A review of trending applications/services in the domain of Smart Video Surveillance and Smart Home Living; ii) A collection of scenarios envisioned by the partners in the consortium; iii) A survey of current FPGA products and projects.

i) Application Review - In the first phase we produced a review of trending applications concerning SVS and SHL. The aim of this investigation is to individuate case studies.

ii) Scenario production - All the project members and stakeholders have been invited to freely propose envisioning scenarios for the application domains. With the contribution of stakeholders from the entire value chain, we have selected and refined twenty scenarios considered worthwhile for benchmarking our approach.

iii) FPGA Survey - All the project partners have been asked to fill a survey concerning recent FPGA boards proposals in order to support the decision process about the AXIOM architecture. All the partners expressed their opinion on the FPGA boards compiling the survey (about pros, cons, community, modularity, programmability and power consumption). Feedbacks have been then summarized, highlighting the main features and characteristics of the analyzed boards. Despite some board showed the salient AXIOM features, none of them combines all the features, proving that there is no board which has features that are overlapping with our feature set, in particular for what concerns easy programmability and scalability.

Finally, among all the proposed scenarios, we selected, refined and detailed in respect to the computational challenges two scenarios, one per domain. The selected scenarios exploit the AXIOM architecture from the computational point of view. These scenarios will represent also a common ground on which simulations and porting will be made. Moreover, these scenarios contribute to define a first precise hardware specification.

GLOSSARY

ADC – Analog to digital converter
ALU – Arithmetic logic unit
AXIOM-link – the interconnects that permits board-to-board communication in AXIOM
Beagle – Contraction of BeagleBone, a single board computer
Bitstream – the binary code used for configuring the PL
CNN – Convolutional neural network
DAC – Digital to analog converter
DoA – Description of Action (acronym set by the European Commission)
FPGA – Field Programmable Gate Array
GPU – Graphics processing unit
GUI – Graphical user interface
IDE – Integrated development environment
IP – Internet Protocol
LBP – Local binary patterns
Mercurium – the OmpSs compiler
NEON – A SIMD architecture extension for the ARM Cortex-A processor
OmpSs – Extension of OpenMP programming model to support task dataflow programming
OmpSs@FPGA – FPGA extension of OmpSs
OmpSs@Cluster – Cluster extension of OmpSs
Raspy – Contraction of Raspberry PI, a single board computer
RTSP – Real Time Streaming Protocol
SATA – Serial AT Attachment
SDK – Software development kit
SHL – Smart Home Living – one of the key AXIOM scenarios
SIMD – Single instruction multiple data
SPI – Serial Peripheral Interface
SVS – Smart Video Surveillance – one of the key AXIOM scenarios
SoC – System on Chip
UART – Universal asynchronous receiver/transmitter
VHDL – VHSIC Hardware Description Language
ZYNQ – A System-on-Chip commercialized by XILINX, which includes FPGA and CPUs

1 Introduction

1.1 Document structure

This Deliverable contains a report on the work about the envisioned scenarios and (related hardware requirements) for the AXIOM architecture. The work is reported in a single section, Section 2, organized as follows:

- Section 2.1, where trending application and services related to our domains are analyzed and reported
- Section 2.2, where all the envisioned scenarios are reported
- Section 2.3, where the result of the FPGA boards survey is summarized
- Section 2.4, where the two scenarios are refined and the related requirements are expressed
- Section 2.5, where an early benchmark set is proposed

1.2 Relation to other deliverables

This deliverable and the related Milestone 3.1 provide input to:

- Deliverable 6.1 (redacted from partner SECO) for hardware definition;
- Deliverable 7.1 (from UNISI) for testing and platform evaluation;
- Deliverable 4.1 (BSC) and 5.1 (EVI) for the application porting.

1.3 Tasks involved in this deliverable

This deliverable is the result of the work developed in tasks:

- **Task 3.1**: Definition of Scenarios for SHL and SVS Case Studies
  Scenario collection and production in the domain of Smart Home/Living (SHL) and Smart Video-Surveillance (SVS) carried out by Contextual Inquiries methods on the field with end Users and Stakeholders.

- **Task 3.2**: Proof of Concept and Porting of SHL and SVS Case Studies
  Selection, envisioning and refinement of Scenarios to be put in to scene by prototypes of AXIOM architecture in the domain of Smart Living/Home and Smart Video-Surveillance.

2 Scenario and requirement definition

At month m9, we collected and produced scenarios in the SHL/SVS domains. A prerequisite was to take in consideration the most promising SHL/SVS services and applications. Then all the partners in WP3 were invited to submit their own vision of forthcoming services and products. We also arranged a survey on the existing and envisioned FPGA boards in order to support the decision process about the AXIOM architecture.

At m12, we selected and refined the envisioned scenarios. These scenarios will represent a common ground on which both simulations and porting will be made. Computational challenges have been articulated in growing steps of complexity.

2.1 Trending applications/services in SVS and SHL

We took in consideration Smart Home Living (SHL) and Smart Video Surveillance (SVS) services/applications and reviewed some of the most promising case studies from a technological and business innovation point-of-view. Such case studies are related to the production of scientific papers and at the same time received attention from media.

Video surveillance and smart home sectors, consistently with other areas of research, deal with a hot topic in computer science: the algorithms and processes related to machine learning. A rough inquiry on Google Scholar, reported in Table 1, points out that more than half of the publications produced in the last two years about the smart home and smart video surveillance take care of the theme of learning.

Table 1. Number of publications related to SVS/SHL and "learning"

Google Scholar query	Number of results
"smart video surveillance" [1]	896
"smart video surveillance" learning [2]	469
"smart home" [3]	19600
"smart home" learning [4]	15000

Machine learning branches related to neural networks, like deep learning and convolutional neural networks, are currently actively employed in exploring new services and products. Those computer science techniques are not new, but dates back to the early 90's. However, only recently they have been considered for real world due to the proliferation of low-cost high performance computing hardware on the mass market.

Machine learning is opening up very interesting perspectives. For this reason, we devoted some attention on services and applications using it. However, it is important to point out that we are not interested in machine learning itself, but we are rather interested in its exploitation, for the customization of video surveillance and smart home applications. From the research point-of-view, our case studies are related to classification purposes, while from a product and market point-of-view, the goal is the customization and the user experience related to the context of use.

2.1.1 Assessed applications and services

We examined several services and applications based on machine learning techniques somehow related to SHL and SVS.

In Appendix A, we report a list of what we considered the most interesting ones, based on the interest in the specialist press, their impact and the availability on the market and the investments received.

2.2 Scenarios envisioning

In parallel with the survey of the state of art, all the partners in WP3 were invited to submit their own vision of forthcoming services and products without specific constraints. The partners produced the following envisioned scenarios in the domain of SVS and SHL.

The list of the envisioned scenarios is available on Appendix B.

2.3 FPGA boards survey

In order to support the decision process about the AXIOM architecture, we prepared a survey concerning recent proposals (successful or not) of FPGA boards. Their architecture and potential domain of applications represent a good indicator of the current zeitgeist.

We selected about twenty boards and invited all of AXIOM partners to express their opinion, feeling, intuition about the projects so to contribute to create a more solid common ground that will help the project design team to put forward the first envisioned solution.

All the partners expressed their opinion compiling a survey. Most of the boards presented in the survey are coming from crowdfunding initiatives, some of them succeeded some failed, but what is important is that all have been presented for their envisioned potentialities and received comments and remark by the potential users. Furthermore, we invited the partners to explore some sources related to these projects before expressing their opinion about these main aspects:

- Pros: describe one (or more) positive aspect of this FPGA project
- Cons: describe one (or more) negative aspect of this FPGA project
- Community: rate the support, documentation and community system
- Modularity and Scalability: evaluate the possibility to interconnect more boards
- Programmability: easiness and diffusion of programming environment
- Power: low energy and efficient solution

At the end of the survey, we resumed the feedbacks and comments expressed by all the partners in a summary. We report the main features and characteristics of the boards analyzed in Appendix C.

Table 2. A quick overview on the recent FPGA-based boards. AXIOM-related aspects are highlighted.

	FPGA / CPU	RAM	Stand alone	ATMEL based Arduino	Connettivity	Programmability
LOGi FPGA	Spartan6 LX9	256MB	No	Pinout	SATA, Raspy/Beagle, SPI	IDE, GUI
MiniSpartan6+	Spartan6 LX9/25	32MB	Yes	--	IO ports, DAC, ADC	IDE, GUI
Papilio DUO	Spartan6 LX9	2MB	Yes	ATmega32U4, Mega pinout	Wings	IDE, GUI
MOJO	Spartan6 LX9	--	Yes	ATmega32U4, custom pinout	--	IDE, GUI
SmartZynq	Zynq 7010/7020	8Gb	No	--	Fast network and board2board	Hard
Parallella	Zynq 7010/7020 16-core Ephiphany	1GB	Yes	--	Gigabit ethernet, 4 high speed connectors	Standard tools
aijuboard	Zynq 7015	1GB	Yes	--	SATA, gigabit ethernet	Standard tools
RED PITAYA	Zynq 7010	512MB	Yes	--	Gigabit eth, 4 fast analog in-puts	Standard tools
OHO	Spartan 3E	--	Yes	--	IO ports	Xilinx ISE only
RetroCade Synth	Spartan 3E or LX9	4MB	No	--	Analog and digital inputs, MIDI, audio jacks	?
PAPILIO	Spartan 3E	8MB	Yes	--	Wings	No SDK
TRIFDEV	Lattice MACHXO2-1200	--	Yes	Partial pinout, I2C	IO ports	No SDK
owlBoard	Spartan6 LX9	--	Yes	--	IO ports	No SDK
Alan	Spartan6 LX45	--	Yes	ATmega32U4 and pinout	IO ports	Arduino IDE, Xilinx ISE
4CH sig.gen.	Spartan6 LX9	--	Yes	--	4 DAC	None
Logitraxx	Spartan6 LX9	64MB	Yes	Compatible with shields	IO ports	No SDK
KromaLights	Spartan6 LX9 Cortex-M3	256MB	Yes	Arduino Due (SAM3X)	LVDS, CAN, USART	SDK (Ardunio IDE?)
CrystalBoard	Spartan6 LX9 4-core Cortex-A9	2GB	Yes	Atmega328, Uno pinout	Ethernet, wifi	?
PSHDL board	Actel A3PN250	--	Yes	Atmel XMega32 (to program FPGA)	UART, IO ports	Simplified VHDL
Helix-4	Altera Cyclone 4 (22k)	4MB	Yes	Arduino Uno shield	IO ports	Altera Quartus II IDE

2.3.1 Bottom line

A year after the presentation of the project, essentially all the elements of the objectives proposed by AXIOM (small flexible and energy efficient board, modularly scalable, easy programmable with an easy interconnection to the cyber physical world) are still targets that are not fully achieved by other systems.

In addition, the survey shows how the features related to AXIOM's objectives (e.g. programmable logic amount, RAM, easy programmability, board-to-board connectivity) are important objectives for some boards, although none of them combines all the features of AXIOM.

2.4 Scenario definition and requirements

Among all applications/services represented in scenarios, we considered the solutions on different aspects. The selected scenarios potentially should be able to exploit the AXIOM architecture from the computational point of view. They also must take the full advantage of the value added by the partner that will implement it, and they must be instrumental to the exploitation opportunities.

These scenarios represent a common ground on which both simulations and porting will be made. Moreover, these scenarios contribute to define a first precise hardware specification.

To achieve this, for each scenario the computational challenges have been articulated in growing steps of complexity, offering bases for the valuation and comparison of the software and hardware definition belonging to the other work packages.

As a continuous development process, the reference software (benchmarks and their reference inputs) is passed to Work-packages WP4-5-6-7 for properly developing the AXIOM board and its software stack.

2.4.1 Smart Video Surveillance

For the smart video surveillance (SVS) case study, we selected one scenario of automated smart marketing involving the computational tasks of face detection in crowds and demographics estimation. This is achieved by means of several computational challenges: high-definition video stream decoding and conversion (format parsing, demuxing and color space conversion). Due to the fact that our implementation is going to be based on an evaluation of sliding-window classifiers, each frame must be downscaled at several sizes in order to run the face detection kernel at maximum speed. This kind of task can be computed in parallel on several ALUs, thus exploiting the maximum computing capabilities of FPGAs. Once the generation of the abovementioned synthetic image pyramid is completed for a given frame, demographics estimation (gender, age and ethnicity) is carried out via a convolutional neural network analyzing faces detected and extracted on the previous step. All pixels constituting a given face will thus be analyzed by the convolutional neural network, which will internally process features and trigger the neuron activations that will estimate facial demographic attributes.

To sum up, all fundamental kernel operations of the SVS case are structured on a pipeline implementing the steps described below:

1. Frame retrieval from the input source (remote IP camera, webcam or video file)
2. Video format de-multiplexing and preprocessing

3. Video decoding (e.g. H.264 video decoding)
4. Face detection and image rescaling
5. Demographics estimation
6. Color conversion
7. Display

These operations are going to be implemented and tested on the AXIOM board prototype with the aim of proving the validity of the designed hardware architecture as a low-power/low-cost high-performing heterogeneous computing architecture. Since some steps does not benefit from massively parallel FPGA computing (e.g. frame retrieval), as they must be serially implemented on a single thread, it is expected that the ARM CPU cores will play a major role in the execution of parsing and de-multiplexing stages.

A good starting point that we agreed with other interested WP-leader is to use the ERA benchmark suite [66] [67] [68], which contains the following benchmarks: cjpeg/djpeg (compression/decompression using JPEG method), H.264 (video encoder/decoder), ECDS (cryptography), Tesserac (OCR recognition), YUV-regb24 (Conversion algorithmic from YUV to rgb24 format. Quite used in mobile environment), MPEG2 decoder (Multi-threaded (platform independent) MPEG2 video decoder), AC3 decoder (Multi-threaded (platform independent) AC3 audio decoder), susan (Image recognition application) or a subset of them.

Description of a potential SVS with smart marketing features

Potential target users of the technology developed on the SVS use case scenario are shopping malls and fashion retailers. A shopping mall is typically equipped with cameras distributed along corridors, entrances, checkout aisles and storefronts. AXIOM boards may be used to collect statistical information about visitors by following their path through the mall, from the entrance to the exit, in order to define marketing strategies, optimize supply chain management and improve the quality of service.

Visitors are tracked from one camera to another, so as to trace the main paths they take through the mall and how long they stay at different locations. Cameras are positioned at the cashier or embedded in objects. These smart objects are, for example, special mannequins which retailers can place inside windows to collect data about those looking at their products. Smart shelves with product weight sensors report if a customer picks something up from the shelf, thereby revealing interest level as it relates to sales or the efficacy of special offers, as well as potentially protecting from theft.

Demographic analysis is conducted on the captured facial snapshots, identifying customer profiles, the distribution of gender and age across segments, and even relevant emotional states. This information could be sent to the remote servers on the cloud. These servers could then post-process all customer profile statistics from multiple shops, conduct business intelligence tasks and optimize stocks through changes in the supply chain with final purpose of increase the productivity of the internal manufacturing and logistic processes of the company.

Implementation steps

The work to be carried out for the SVS scenario relies on the correct implementation of a highly parallel face analysis pipeline on the Xilinx Zynq SoC FPGA, which has been selected for the architec-

ture of the AXIOM board. The designed video processing pipeline should be designed for efficiently processing video frames broadcasted from remote IP surveillance cameras. As such, since these devices typically encode video using H.264 video codecs it is also required to perform video decoding on the AXIOM platform. With this requirement in mind, it is expected that video processing operations will be offloaded to hardware logic IP blocks on the FPGA to maximize data parallelism while reducing latency. Therefore, the proposed pipeline will be structured in the following steps to clearly separate fundamental operations that are going to be implemented in a collection of kernels: frame retrieval, video de-multiplexing and conversion, face detection, color space conversion and demographics estimation.

Frame retrieval

The first step of the pipeline is to perform the connection to the remote IP cameras in order to retrieve the video stream. The proposed implementation will rely on a software layer using the open source LibAv library [65] for performing the RTSP connection to the camera (more particularly, the *libavformat* library). This task involves getting the settings of the camera, the codec parameters (i.e. H.264 profile settings, bitrate and the color space used for the frames) and the container format. Once this has been done, it is required to retrieve and cache several video frames in a buffer stored in the RAM memory of the FPGA. This sub step is required because the rate in which the video frames are sent from the camera and the rate of video processing on the AXIOM board differ. The amount of cached frames is going to be determined experimentally and involves knowing both the latencies and throughput of the H.264 decoding engine and the face detection kernel.

Video conversion and de-multiplexing

Most of the hardwired H.264 decoder IP logic blocks available in the market are only capable of decoding video frames that are encoded using the Annex B H.264 video format. Unfortunately, most of the available cameras does not automatically generate multiplexed H.264 frames in the abovementioned format. This H.264 Annex B conversion can be efficiently performed in software on an ARM A9 CPU core by the *libavcodec* library. Since this operation is not computationally intensive, we expect to dedicate only one of the available ARM cores for performing this subtask.

Additionally, it is also required to parse the RTSP protocol to extract the H.264 slices that are going to be processed by the H.264 decoder. This de-multiplexing subtask is going to be implemented also by software means on a single ARM A9 core using the abovementioned *libavformat* library.

Face detection and image rescaling

Once the video input has been decoded, the face detection kernel must be executed to determine the size and locations of faces appearing on a given video frame. This kernel is going to be synthesized in the programmable logic (FPGA) with purpose of reducing the latency of operations as much as possible. Since the face detection kernel is trained to detect fixed-sized faces (typically, 24x24 or 48x48 pixels), it is required to downscale the input video frame several times to detect faces of various sizes. The downscaling kernel could be implemented either on the programming logic in hardware or on software on the ARM A9 cores. As this kernel is several orders of magnitude less time consuming than the face detection kernel, we have opted to implement it on software.

For the face detection kernel several alternatives are going to be considered. The first one proposes implementing a face detector based on boosting machine learning techniques and local binary patterns

(LBP). If there are not enough configurable logic blocks and look-up tables on the FPGA for implementing both the cascade of classifiers and the evaluation engine, we would then switch to a design based on a neural network face detector, which would rely on a smaller model.

Demographics estimation

Facial attributes such as gender, ethnicity or age are going to be determined using deep learning techniques based on convolutional neural networks. Thus these algorithms will be structured into a kernel that will implement a generic inference engine for convolutional neural networks. The execution of such kernel will be conducted by the programming logic of the FPGA, while the models are expected to be stored in the internal RAM memory available on the FPGA. This scheme ensures rapid prototyping, and the accuracy of the algorithms could be easily updated in the future just by replacing the models, thus not requiring to perform any modifications in the neural network inference engine.

Color space conversion

Since the face analysis kernels (i.e. face detection and demographics) works with luminance components or grayscale images, it is also required to perform multiple color space conversions from the input video to the internal buffers. Typically, IP cameras encode H.264 videos using YUV color space, while displays rely on RGB color components. These conversions are going to be implemented on software using the SIMD NEON instructions available on the ARM cores, as they are not as time consuming as the face analysis kernels.

Development strategy

With the abovementioned pipeline stages in mind, it is wise to adopt an iterative and incremental development model starting with tasks that are less complex than the final required ones. The first iteration will work with raw unencoded still images. The purpose of this early model will be to fine-tune the correctness and accuracy of the face detection kernel, as it is the most complex requirement for implementing the envisioned scenarios. Secondly, we will implement H.264 video decoding using a software engine on the ARM A9 CPU cores. The idea of this step would be to further test and benchmark both the throughput and latency of the face detection kernel. Once this task is done, we will offload the video decoding process to a hardware logic block on the FPGA, thus freeing CPU resources for other tasks such as the color conversion for displaying or encoding the results to end users. Then we will implement the neural network inference engine on software on the ARM A9 cores in order to test the validity and accuracy of the selected demographic models. Finally, we will synthesize the neural network inference engine on the programmable logic of the FPGA to speed up this kernel as maximum as possible.

Further steps will involve proving the validity of the AXIOM platform on multi-camera scenarios using several boards. For implementing this use case, we will leverage the high-performance interconnect developed during the project to distribute computations into several FPGAs. This task will thus involve heavy testing and dealing with the scalability issues that typically arise on such distributed architectures. Whenever possible these design option will be more precisely evaluated through the AXIOM Evaluation platform that is developed in WP7 (see Deliverable D7.1).

Computational challenges

- Efficient H.264 video decoding;

- Low-latency multi-face detection on high-definition videos;
- High-performance demographic estimation on all previously detected faces in parallel.

Hardware requirements

- ARMv7 or ARMv8 cores with NEON SIMD capabilities;
- Xilinx Zynq FPGA SoC;
- RJ-45 connector for retrieving frames from remote IP cameras and sharing information across multiple AXIOM boards (included already in the prototype specification in D6.1);
- HDMI connector for displaying the results to the end user (included already in the prototype specification in D6.1).

2.4.2 Smart Home Living

For the SHL case study, we selected a scenario on home security which aims to identify users through several points of access in the house, using both audio and video analysis. Recorded sounds coming from both vocal and non-vocal signals will be analyzed on the FPGA-based SoC detecting if the sound is produced by the voice of an authorized or an unauthorized user, or by a noise source. To increase the level of security the iris recognition is introduced. For this task frames must be decoded and analyzed on the FPGA-based SoC in order to extract the eye features required to identify the authorized users.

The AXIOM-based access control system is deployed in a residential home. Its purpose is to monitor the access to the building controlling gates or main door.

Physical access to the house is granted through the analysis of sound captured by microphones and through the analysis of images produced through the camera.

These two features allow more than one level of authentication. For instance, a task without implications with regard to safety (like activation of the security services) could be carried out just recognizing the vocal timbre. On the contrary, the opening of the main door could be done only after both timbre and iris recognition.

The AXIOM infrastructure that will result from the AXIOM Project could later be extended to perform audio analysis in other contexts of home automation. An example it would be the analysis of audio signals coming from microphones installed in different rooms, such as the bedroom, bathroom, living room or the kitchen, to understand when there is activity in a room or when people are asleep, and the system could dim the lights slowly before turning them off. Moreover, it will be also possible to treat the noise and background noise as signals, in order contextualize the user's main activity.

SHL scenario in the context of VIMAR market

The aim of this scenario in the context of the AXIOM Project is to develop a smart module that will be integrated into VIMAR home automation system, starting from access control and security of living units devices.

The project involves the development of a system capable of performing the extraction and analysis of biometric features of voice and iris using as input the recordings of the voice and the photos focused on the area of the face where the eye is.

The audio and video data used in the analysis will be provided to the smart module by external devices like the video door entry systems installed in the access points of houses and apartment buildings. The results of the analysis of these data will be used to enable entering or exiting the building.

The access operation involves the opening of the entrance door, disabling of security services and enabling of the welcome service (e.g. switching on lights, activation of sound system, etc.). The exit operation entails the activation of the security services and the operations usually carried out in this situation (e.g. switching off lights, configuring the thermoregulation system, lowering of the shutters, etc.).

The management system that controls the entrances will be characterized by a double check to increase the level of security, therefore both the biometric features of the voice and the iris are verified. The exit operation is characterized by a lower security level; therefore, the system could verify only the voice features. For a more natural human-machine interface, the system will be endowed with a speech synthesizer able to produce artificial human voice used to generate welcome and goodbye messages and to request iris verification and voice verification.

This first system will be the starting point to develop more advanced services characterized by voice analysis and non-vocal signals, which could be used to identify intruders in the house or malfunctions of household electrical appliances, and to allow the contextualization of the vocal information with the events that have occurred inside the house. Furthermore, the system could allow the analysis of the emotional states of people and this could consent to identify a potential fraud in the identification phase and allow a more natural interaction between the home automation system and the inhabitants of the house.

Implementation steps

The SHL scenario will rely on the implementation of four functional modules: the speaker identification module, the iris recognition module, the biometric score fusion module and the speech synthesis module. The four modules are going to be implemented on the Xilinx Zynq SoC FPGA, which has been selected in this project for the AXIOM board (Milestone MS1 and Deliverables D6.1, D7.1).

The implementation of these four modules requires the analysis of specific algorithms, the testing of open source libraries, the development of the software code and the acceleration through the AXIOM tools. The acceleration task is composed of the intra-node FPGA acceleration (OmpSs@FPGA) and inter-node Cluster acceleration (OmpSs@Cluster). The acceleration task will involve testing and dealing with the scalability issues that typically arise on such distributed architectures. This specific task will be carried on with the close cooperation and iteration with the AXIOM partners.

Modular programming will be used to develop the four steps in independent modules.

The sequential implementation steps will be:

Step 1: Speaker identification module

Develop an advanced access control system capable of extracting the biometric characteristics of the voice of the people and select if they are authorized or not.

1.1. Retrieve the audio stream
1.2. Pre-processing and Voice Activity Detection
1.3. Speech parameterization
1.4. Statistical modeling
1.5. Scoring normalization
1.6. Decision

The audio stream is received and stored in the memory by the microprocessor and its peripherals. The pre-processing and extraction of the parameters which describe the voice biometric characteristics (Speech parameterization) is characterized by a pipeline of filters that processes the audio stream; this is an operation that can be better performed by architecture that allows concurrent processing rather than a sequential architecture. This task is a candidate to be processed on FPGA.

The processing of the statistical model could require iterative algorithms and they could be easily processed and managed by sequential architectures such as microprocessors. This task will be probably performed by a microprocessor.

The steps of normalization and decision could be carried out in both types of architecture and they will be decided according to the final algorithm.

Step 2: Iris recognition module

Introduction of the verification of the biometric features of the iris to extend the reliability of the access control system. The system checks the iris biometric features deciding if the person is authorized to the access.

2.1. Frame retrieval
2.2. Frame decompression
2.3 Eye region detection
2.4. Segmentation
2.5. Normalization
2.6. Encoding
2.7. Matching
2.8. Decision

The video stream is received and stored in the memory by the microprocessor and its peripherals. Since the input video stream is compressed, it could be decompressed by optimized libraries for ARM architecture. The first part of the iris recognition process, like eye region detection, segmentation and normalization requires trigonometric operations, floating point operations as well as scanning the pixels disorderly inside the frame. It is preferable to implement these operations with the microprocessor because it is equipped with dedicated instruction set for floating point arithmetic and it does not require to scan data in a continuous way.

The operations of encoding and matching require to perform filtering in cascade of a stream of data and comparisons in parallel. This operation could be speeded up with peripheral devices in FPGA.

Step 3: Biometric score fusion module

Define the acceptance policy to authorize the person on the basis of the results of the voice timber and iris recognition.

Step 4: Speech synthesis module

Introduction of a speech synthesis module to allow vocal feedback to the system

4.1. Speech synthesis
4.2. Define a set of texts and events

Computational challenges

Audio processing

- Noise filtering
  - Pre-Filtering
- Voice Activity Detection (VAD)
- Speaker recognition
  - Speech parameterization with filterbank-based cepstral parameters
  - Statistical modeling with Gaussian Mixture Models (GMM) for Speaker recognition Text-independent or Hidden Markov Models (HMMs) for Speaker recognition Text-dependent
- Normalization

Video processing

- Decompression
  - Frames decompression (MJPEG or H.264)
- Color space conversion
  - YUV to Y
- Iris recognition
  - Eye region Detection
  - Segmentation
  - Normalization
  - Encoding
  - Matching

Hardware requirements

Input/output connections

The input data of the system is an audio-video stream. The data are received from IP network. The output data of the system are actuation commands sent over an IP network. The Ethernet connection on board AXIOM is therefore required. Since the board is designed to process multiple audio and video streams in parallel, it is required that the connection is Gigabit Ethernet. The Gigabit allow to broadcast a minimum of 2 audio streams and 2 video streams FullHD compressed in H264/MJPEG.

Processor and programmable logic

The data stream is received through the Ethernet connection and stored in the DDR memory. The dataflow is received, extracted from the TCP/UDP package and stored in DDR by the microprocessor and its peripherals. The memorized data is a candidate to be later processed by tasks running on the

FPGA or to be executed on the microprocessor, according to the logic defined in the application software, in order the achieve maximum performance.

Shared memory between the microprocessor and FPGA and the internal connections between FPGA and microprocessor allow for the possibility to accelerate and run more efficiently the tasks of our algorithm.

Memory

The microprocessor and the FPGA must co-process streams of video and audio data. The system needs a memory area in which the data are stored and shared.

DDR memory is required to store the data received from the network and to share and store the data used by the applications. In order to perform these operations on multiple streams the memory required is then 1 GB.

2.5 Output

The design and implementation of the selected scenarios is planned to be executed using an iterative model in which a working prototype is built starting from the evaluation of a predefined set of benchmarks. These benchmarks isolate and implement fundamental core algorithms structured in data-parallel kernels that are going to be benefitted from the AXIOM architecture. The aim of such design/space exploration is to obtain an early picture of all the involved operations and parallel constructs that will be required for dynamically offloading the algorithms to FPGAs by means of the OmpSs programming model. Thanks to the preliminary unoptimized kernels, BSC, UNISI and EVI partners will have a minimal starting application that closely represents the final workloads, and thus can start working on low-level compiling, scheduling and run-time optimizations since the beginning of year 2.

2.5.1 Definition of the early benchmark set

On the SVS use case scenario, the benchmark set is expected to be based on the kernels defined on the table presented below.

Table 3. *Proposed benchmarks representing the SVS scenario workload.*

Kernel Name	Description
`H264_video_decoding`	H.264 codec decoder working at slice level
`LBP_cascade_evaluation`	Face detection based on LBP patterns
`CNN_inference`	Convolutional neural network inference engine
`YUV_to_RGB`	Color space conversion for displaying frames

Starting from the input video stream broadcasted from a remote H.264 camera, the AXIOM board will unpack and demux the transport protocol (typically RTSP) and the container format up to the slice level of the H.264 video codec. At this point, the disclosed `H264_video_decoding` kernel will decode the video slices by means of either the FPGA logic or the ARM cores. Initially, this kernel will

rely on the LibAv software video decoding engine. In the case the OmpSs compiler is not able to automatically parallelize annotations of the most time-consuming loops of such implementation, H.264 decoding will be offloaded to an IP block synthesized on the FPGA. Since the design of the IP block is out of the scope of this project, it has been decided that it will be purchased from a third party company or research institution in the case it is required.

The output of the abovementioned kernel will serve as the input of the upcoming kernel `LBP_cascade_evaluation`, which will evaluate the boosted cascade of classifiers based on local binary patterns by means of a sliding-window approach. The output of this latter kernel is expected to be a vector with the coordinates and dimensions of all faces appearing on the processed input video frame or recently-decoded H.264 slice.

For each region of the image identified as a face, the `CNN_inference` kernel will evaluate several convolutional neural network (CNN) models with multiple levels of hidden layers to automatically determine the age, gender and ethnicity of each face region. Therefore, this kernel it is expected to implement the CNN inference engine using the standard forward propagation technique widely used to evaluate neural networks. At the lowest, level this kernel will rely on standardized BLAS-SGEMM matrix multiply operations for performing the inference. The final output of the CNN will be a vector of integers that will encode the demographic categorization of the face sub image used as the input of the kernel.

Finally, it will be necessary to somehow show the results of processed faces, demographic features and the decoded video to end users using the HDMI output included on the AXIOM board. All this information is expected to be drawn on a given video frame, and the converted to RGB for displaying it on a computer or TV screen. This color space conversion is going to be performed by the `YUV_to_RGB` kernel.

On the SHL use case scenario, the benchmark set is based on audio and video processing involved in biometric analysis. The audio and video processing will be analyzed in sequence.

Table 4. *Proposed benchmarks representing the SHL scenario workload.*

Kernel Name	Description
`Voice_Activity_Detector`	Algorithm to label speech frames
`Speaker_Recognition`	Pattern matching for speaker identification
`Iris_Recognition`	Pattern matching for iris recognition

The audio stream used in speaker recognition will be received from the IP network. Firstly, the `Voice_Activity_Detector` will pre-process the audio stream in order to extract only the parameters used to describe the voice biometric features and label the audio samples to identify the voice frames into the stream. This kernel constitutes a pipeline of filters able to process the raw audio input and obtain the parameters that defines the biometric voice characteristics. The output of this pipeline will be used by the `Speaker_Recognition` core to define whether the speaker is a known enabled speaker or an unknown speaker.

The `Voice_Activity_Detector` and the `Speaker_Recognition` will be profiled with BSC and UNISI to define the best solution to reduce the processing time using the AXIOM infrastructure.

The video data used for the iris recognition will be received from the IP network. The frames received will be decompressed to allow the extraction of information from inside the frames. To perform iris recognition, the video data, provided in color space YUV, will be converted in grayscale Y to simplify the algorithms used in the next steps. After these two passages, the images obtained will be used to extract the biometric features which will be compared to the iris features already stored. The `Iris_Recognition` requires four kernels in cascade, as follows: segmentation core, normalization core, encoding core and matching core. This pipeline of kernels will define the iris contours, transform the iris area into a size-invariant strip, and finally it will extract the iris texture and compare iris codes using Hamming distance. As for the audio benchmarks set, the `Iris_Recognition` will be profiled with BSC and UNISI to define the best solution to reduce the processing time using the AXIOM infrastructure.

All the kernels summarized above will constitute the initial benchmark set of the SVS and SHL scenarios. As the project progresses, proposed optimizations on the Mercurium/OmpSs compiler and runtime will target them and try to both reduce the execution time and increase throughput by optimizing the efficiency of memory access patterns. Since the kernels are designed to be executed on a pipeline fashion, partner HERTA and VIMAR will provide a set of expected inputs and outputs for each kernel. With these input/output test datasets, each kernel can be productively refined, simulated, and/or debugged separately.

2.5.2 Proposal of a modular architecture based on two units

The definition of the two scenarios and the results of the survey pointed out the opportunity to consider two basic units for the AXIOM architecture: one entry level, and one advanced level. Both will have high speed connections, and at the present we suggested the same hardware specification. Nevertheless, according to the evolution of the two case studies the specific features of the two basic units could be differentiated.

2.5.3 Definition of testing based on modularity and acceleration

The two scenarios are structured differently and data is processed in two ways. The SVS is characterized by algorithms able to process the input data in cascade, while the SHL is characterized of algorithms which process the input data concurrently. This enable the possibility to analyze the AXIOM architecture in two separate ways.

Due to the decision of structuring fundamental operations in kernels, it is possible to easily test them separately as black boxes with minimal scripting or coding. The testing process will involve, first, ensuring the correctness of the code of a given kernel parallelized by OmpSs, just by comparing the differences between the actual output and the output provided in the benchmark data set. Once this is done, automated fuzzy testing with synthetic inputs will take place just by determining whether the outputs of both the serial and parallel versions of each kernels precisely match.

The next crucial step will involve the testing of the monolithic SVS/SHL prototype applications on a production-ready single AXIOM board. These applications will internally rely on the parallelized versions of the kernels previously described, and it is expected to involve human interaction during the testing process. If parallelization bugs are detected on this step, the same tests will be conducted again

with debugging and logging messages enabled to reproduce them, and thus exactly determine the root cause of behind errors. This information will then be submitted to partners BSC and EVI with the purpose of fixing bugs on the Mercurium compiler and/or OmpSs runtime software.

Finally, if testing succeeds on a single board, the last step will involve the evaluation of the proposed AXIOM interconnect by leveraging multiple boards. At this stage, tasks and loop iterations inside kernels will be (tentatively) automatically offloaded across multiple AXIOM boards through the high-speed interconnect (AXIOM-link). The testing process will thus involve providing challenging inputs to the SVS/SHL use case applications that we expect are going to trigger high computational requirements. As it is impossible to compute challenging inputs (e.g. dozens of simultaneous faces on a given video frame, or multiple audio streams) on a reasonable time on a single AXIOM board, OmpSs will split kernel iterations in blocks, and then automatically schedule the computations on multiple AXIOM boards. This exhaustive testing procedure will involve all AXIOM partners.

3 Confirmation of DoA objectives

PLANNED	DELIVERED
DELIVERABLE: SCENARIOS DEFINITION	
• State of the art application review	The most trending applications and services related to SHL/SVS have been reviewed and presented in section 2.1
• Scenarios envisioning	We envisioned 16 scenarios, see section 2.2
• FPGA board survey	Even if unplanned, we benchmarked the existing and planned most interesting FPGA boards.
• Analysis on data flow, algorithms and computation power needed	An approximate analysis is stated in section 2.4
DELIVERABLE: SCENARIOS REFINEMENT	
• Scenarios refinement using the AXIOM CPS platform	The refined scenarios are presented in section 2.4
• Benchmark set definition	An early set of possible benchmark to assess the hardware and software stack is presented in section 2.5

4 Conclusion

For the task T3.1, we took in consideration SHL and SVS services/applications and reviewed some of the most promising ones. This review highlighted that most leading trends are tied on the user experience customization, based both on the user and the context of use, capitalizing on the past experience. The expression recognition, demographics estimation and activity context recognition are major challenges often addressed through the complementary use of audio and video analysis.

All the partners in WP3 were invited to submit their own vision of forthcoming services and products without specific constraints. The partners produced several envisioned scenarios in the domain of both SVS and SHL.

In order to support the decision process about the AXIOM architecture, we prepared a survey with twenty recent proposals (successful or not) of FPGA boards. We invited all of the AXIOM partners to express their opinion, feeling, intuition about the projects. A year after the presentation of the project, all the elements of the objectives proposed by AXIOM are still targets that are not in any of the existing technologies. In addition, the survey shows how the features related to AXIOM's objectives are important objectives for some boards, although none of them combines all the features of AXIOM.

For the task T3.2, starting from the envisioned scenarios, we selected and refined two scenarios (one per domain) which will represent a common ground on which both simulations and porting will be made. For each scenario, the computational challenges have been articulated in growing steps of complexity, offering bases for the valuation and comparison of the software and hardware definition belonging to the other work packages. In this way, these scenarios contributed to define a first precise hardware specification too.

The next steps, still tied to T3.2, will concern the porting of the Smart Living/Home Application and Smart Video-Surveillance to the OmpSs Programming Model available from WP4.

Appendix A

Applications and services related to SHL/SVS

ERSATZ Labs

Field: Machine learning

Value Proposition: Ersatz [5] is a web-based machine learning platform designed to assist with data wrangling, machine learning, and model deployment.

Description: Deep learning advances the state of the art by being able to process and analyze video better than ever before.

For instance, a retailer might be interested in reducing theft. Cameras are cheaper than ever, but people to watch them are not. Deep learning enables the development of algorithms that can watch these feeds automatically and obtain a high degree of accuracy and a very short development cycle.

One of the key tools in making deep neural networks practical is the use of specialized hardware, specifically Graphics Processing Units (GPUs). The use of GPUs allows an approximately 40x speedup on training time. So now, a model that might have taken 40 days to train will only take one day. That's a tremendous boost.

Erstaz claims to speed up neural networks and have been running them on GPUs

Press Coverage: Ersatz Labs Launches First GPU Platform for Deep Learning with Cloud- and Appliance-Based Technology [6]

iFLYTEK

Field: Smart speech-recognition technology

Value Proposition: iFLYTEK [7] is focused on R&D of intelligent speech technology, software, chip products, and services.

Description: Voiceprint Recognition compares the speaker's voice with the voiceprints registered in the database, checks and verifies the speaker's identity, and determines whether the voice matches the voiceprint of the speaker or another person in a group. Voiceprint recognition delivers high security performance comparable to other biological recognition technologies (such as fingerprint, palm print, and iris identification).

Raised so far: $6.16M in 5 Rounds from 3 Investors [8]

Press Coverage: Chinese Manufacturers, Carriers and Developers Form an Alliance to Block Apple's Siri in China [9]

Percepto

Field: Machine Learning, Image Processing

Value Proposition: Percepto [10] aims to develop an add-on board for drones in order to add computer vision capabilities.

Description: Percepto is a drone add-on. Meaning it's a device (about the size of a deck of cards) and a small camera that you connect to your existing drone. It enables your drone to use the power of computer vision in real time. With computer vision drones can "see" and understand the environment better than with today's sensors. Using your smartphone, mark any object and have the drone follow it, circle it or create any filming maneuver you desire with a click of a button. A customizable heads up display, driven by our computer vision algorithms which can produce smart alerts such as obstacle advisory, object searching or even locations of additional Percepto powered drones flying nearby.

Raised so far: 1M$

Press Coverage: Mark Cuban and Richard Parsons invest in drone computer vision startup Percepto [11]

Kairos

Field: Face Recognition-as-a-Service

Value Proposition: Kairos AR [12] is a tech platform that provides advanced face recognition API for its users' apps and services.

Description: Kairos AR is a tech platform that provides advanced face recognition API for its users' apps and services.

It delivers fraud management and work automation solutions with features such as 3D facial recognition, one-to-many identification, anti-spoofing, mood detection, gender analysis, security and compliance, mobile authentication, and more.

Raised so far: $3.3M in 6 Rounds from 9 Investors [13]

Press Coverage: Kairos Lands $500K to Bring Facial Recognition Tech to Employee Time Clocks [14]

Clarify

Field: Data mining of visual and audio content

Value Proposition: Clarify [15] makes searching audio/video files simple for developers to integrate into their applications.

Description: Clarify aims at giving people access to rich data extracted from an unstructured database of video and audio content. Signal processing, language analysis, and complex math lie behind an easy-to-use API with useful libraries and quick start guides to help users get started in minutes. The startup focus around extrapolating information such as environment, subject matter, emotions, and identity in order to provide users with relevant information about the context in order to understand the speaker's intention.

Raised so far: $1.32M in 2 Rounds from 3 Investors [16]

Press Coverage: PE HUB [17]

Ecovent

Field: Smart thermostats and heat management

Value Proposition: Ecovent [18] provides one single smart thermoregulation system powered with machine learning.

Description: Most homes have only one thermostat, so they operate like a house with only one light switch. Everything is either all on or off. That leaves some rooms boiling hot while others are freezing cold. It also leads to heating and cooling empty rooms. It's uncomfortable, and it's inefficient.

Ecovent fixes that for the over 200 million Americans with forced air heating and cooling systems. We've built a system of wireless vents and sensors which enables homeowners to: (i) Get room-by-room temperature control, (b) save energy by automatically directing airflow to the rooms that need it most, (c) choose custom temperature levels for different floors, rooms, or times of day, (d) control home heating and cooling anywhere from a mobile phone, tablet or web app and (e) install the whole system hours without special tools or invasive construction.

Raised so far: $9.71M in 3 Rounds from 6 Investors [19]

Press Coverage: Smart-home heating startup Ecovent lands $6.9m investment, touts $1m in pre-orders [20]

Dato

Field: Machine learning as a B2B SAAS.

Value Proposition: Dato [21]: a large-scale image-oriented machine learning platform optimized for production environment and big data analytics on top of which third-party players could eventually develop apps powered by such intelligence.

Description: GraphLab Create™ is a machine learning platform that enables data scientists and app developers to easily create intelligent apps at scale. Building an intelligent, predictive application involves iterating over multiple steps: cleaning the data, developing features, training a model, and creating and maintaining a predictive service. GraphLab Create™ does all of this in one platform. It claims to be easy to use, fast, and powerful.

Raised so far: $25.25M in 2 Rounds from 6 Investors [22]

Press Coverage: TechCrunch [23]

Emotient

Field: Facial expression recognition and analysis

Value Proposition: Emotient [24]: emotion detection, identification and analysis, also with Google Glasses.

Description: Emotion measurement integrated into an online survey. Demographic insights – quickly and at scale.

Emotient Analytics delivers >95% accuracy in the detection of primary emotions, expressed in as little as a single video frame, in both real-world conditions as well as controlled environments. These brief flashes of emotion, or “micro-expressions”, were shown by Dr. Ekman’s work to be important to the understanding of people’s gut reactions to stimuli. Practically speaking, they reveal quite a bit about a person’s beliefs and their propensity to act or buy. These are the tiny “tells” that occur in the short period between unfiltered reaction and conscious masking of emotions.

Raised so far: $6M in 1 Round from 2 Investors [25]

Press Coverage: TechCrunch (Emotient and Google Glasses) [26]

Enlitic

Field: Machine Learning, Health Diagnostics

Value Proposition: Creating data driven medicine using deep learning, mostly focused on deep learning of visual content.

Description: Enlitic [27] uses recent advances in machine learning to make medical diagnostics faster, more accurate, and more accessible. The company's mission is to provide the tools that allow phy-

sicians to fully utilize the vast stores of medical data collected today, regardless of what form they are in - such as medical images, doctors' notes, and structured lab tests. To realize this vision, the company is building on state-of-the-art deep learning algorithms and partnering with top research hospitals and medical device manufacturers.

Data could include X-rays, MRIs, CT scans, 3D medical images. Enclitic's software could plug into systems medical institutions already used to share or view medical images. Researchers could annotate images and click to find similar elements in other data.

Raised so far: $5M in 2 Rounds [28]

Press Coverage: TechCrunch [29]

Eyeris

Field: Sensors, Software, Artificial Intelligence, Face Recognition, Computer Vision, Machine Learning, Developer APIs, Video Processing, Enterprise Software, Image Recognition, Predictive Analytics, Human Computer Interaction

Value Proposition: Eyeris [30]: deep Learning-based emotion recognition software that reads facial micro-expressions.

Description: Emotions are hardwired into our brains at birth and manifest as facial expressions. Research has long proven their universality, regardless of age, gender or race. Facial expressions are classified as joy, surprise, sadness, disgust, fear and anger. We created novel techniques that use Deep Learning architectures, which employ Convolutional Neural Networks (CNN) to train our algorithm and deploy it to the real world for continuous learning. Our holistic facial expression recognition methodology imitates human vision and allows our algorithm to learn prototypic expressions directly from the face instead of relying on decomposed action units.

Press Coverage: Sentiment Analysis Innovation Source [31]

Orbeus

Field: Facial/Object/Scene Recognition, Image/Text Processing

Value Proposition: Orbeus [32], an innovator in automated facial, object and scene recognition, image-to-text analysis and video indexing.

Description: Orbeus taps the power of deep learning to provide scalable image and face recognition solutions for businesses and consumers. With expertise in state-of-the-art machine learning and ImageNet (award winners in 2013 and 2014), Orbeus understands both the tremendous value, and limitations, of a purely academic approach for everyday uses. Orbeus bridges this gap to meet identification requirements has developed what could be the best application yet for letting everyday consumers benefit from advances in deep learning, called PhotoTime, which got featured by Apple store and mass medias.

Raised so far: $1.47M in 2 Rounds [33]

Press Coverage: We Asked Some of the Smartest Computers to Identify This Picture [34]

Facety

Field: Machine Learning, Image Processing

Value Proposition: The startup aims to develop a beauty-recognition system to help fashion agency in hiring models.

Description: At Facety [35] we are passionate about connecting beauty and technology. Our mission is to teach computers to understand human beauty. Through that we want to benefit industries that are connected to human aesthetics. Right now the company is building a software for modeling agencies that will help them dramatically increase speed of scouting people with model potential on Instagram. This solution alone will change the way scouting is done and increase model agencies revenue and minimize scouting expenses.

Clarifai

Field: advertising and content curation.

Value Proposition: Clarifai [36] provides advanced image recognition systems for customers to detect near-duplicates and visual searches.

Description: Clarifai's image recognition systems recognize various categories and tags in images, as well as find similar images. The company's image recognition systems allow its users to find similar images in large uncategorized repositories using a combination of semantic and visual similarities.

Raised so far: $10M Series A on April 28, 2015 from 6 investors [37]

Press Coverage: TechCrunch [38], YourStory [39]

Special notes: 1) Clarifai uses convolutional neural networks; 2) the value proposition is not so unique: "the world's best image recognition technology on the market", this was true just for the ImageNet 2013 challenge [40]

Camio

Field: Home video surveillance through machine and deep learning

Value Proposition: An easy way to monitor life at home.

Description: Camio [41] is a free web and mobile app that turns any tablet, smartphone, or computer into a home video monitoring camera. With Camio, you can see what's happening with your family or

pets at home when you're away. Camio leverages machine learning and computer vision technology to learn what you find important, so you can see the video that matters to you.

Press Coverage: Using Deep Learning to Make Video Surveillance Smarter [42]

Appendix B

Envisioned SHL/SVS scenarios

1. Smart marketing

Molly is looking for some new dresses and accessories for the upcoming season. She enters a new shopping mall and notice an advertising digital signage station equipped with a camera. As soon as she stays in front of it, the station displays targeted advertisement according to her demographic group carried out over captured facial snapshots. The analysis of the time she spent looking at a particular product or ad gives an estimation of its impact to the demographic segment she belongs. She decides to visit two women shops with special prices and then to eat at the restaurant downstairs because today it offers a vegan gluten-free menu.

All the shops are equipped with cameras that are distributed along certain passages. Molly is tracked from one camera to another, so as to discover the main paths she takes through the mall and how long she stays at different locations.

All the statistical information collected about Molly and other customers, such as the demographic profiles, or how do they distribute into gender and age segments, are used by shopping mall owners which can define marketing strategies and better determine the rent of all their shops based on the observed traffic of people, and decide the best location for a store on a customer base.

Technologies: Video analysis (anonymous face identification, demographics estimation)
HERTA possible exploitation relevance: 10/10
VIMAR possible exploitation relevance: 2/10

2. Special moments

Mr. Scarlet and his new wife have planned to stay at the Smart Resort for two weeks during their honeymoon. The resort already offers a large number of Photo Spots at the main places of interest, in which customers just have to place themselves for a few seconds to have an instant photograph automatically taken. It’s amazing for them because they do not have to worry about bringing their own camera or device to take a picture of them.

The Photo Spots are in all the main places of interest including the swimming pool, the tennis court, the restaurants, viewpoints and Spa. The Photo Spot system tags each photograph with the identity of people appearing on it, so that later, The Whites can immediately check out all their photos.

Technologies: Video analysis (intent recognition), face identification
HERTA possible exploitation relevance: 6/10
VIMAR possible exploitation relevance: 1/10

3. Smart home comfort "autopilot"

Igor’s house is equipped with smart thermostats, video door entry system and other smart sensors that are distributed in each room and house ambient.

The devices perform their primary "normal" function, but also collect different kind of information, ranging from presence detection, temperature, humidity, windows and doors opening, air quality or audio.

All devices are set to collect statistical and behavioral information about Igor habits. Real time data analysis is carried out in order to create and maintain a map of the room and house usage during the day and the week, for example the rate of room usage, kind of movement occurring, people moving or people staying still, helping to identify recurring patterns and forecast possible activities and energy needed to guarantee the desired level of comfort.

Identifications of pattern of usage has the purpose of better managing house resources and increasing the efficiency of power consumption management, making use of traditional and renewable power sources. Comfort perception and necessities can be different for young and healthy people, moving and running around, from elder or ill people, standing still for most of the time. The smart home is required to identify and manage the different situations, and to react at the people indications in an easy and smooth way. The objective of the smart home comfort autopilot is to minimize power consumption and to guarantee people's comfort and well-being, without giving the impression of reducing people freedom and capacity of control. The comfort autopilot of the single smart home or apartment can share part of the data collected and analyzed with a broader environment, being it the local area community or the building, in order to achieve a more global level of optimization and power saving.

Technologies: Presence detection, data analysis, pattern mining, activity prediction
HERTA possible exploitation relevance: 0/10
VIMAR possible exploitation relevance: 7/10

4. Smart home and office “Security Guard”

George works in a small but smart office that is equipped with all kind of sensors and cameras and smart thermostats, with on board cameras, and capable of analyzing real time data. His smart office could detect peculiar and unexpected situations that can be recognized as a dangerous situation for people or goods inside the premise.

Data analysis can be used for facial recognition, expression recognition, audio stress situations and unpredictable movement recognition; from the deriving knowledge the Smart Home Security Guard autopilot can set off different kind of actions, such as silent alarm issuing in case of robbery.

Technologies: Video analysis (intent recognition, face identification, expression recognition), audio analysis, data analysis
HERTA possible exploitation relevance: 6/10
VIMAR possible exploitation relevance: 6/10

5. Home Broken Home

The Reds are ready to rent a furnished home for a period of six months and want to check the condition of the infrastructures of the house and its appliances (refrigerator, washing machine, etc.). They are thinking to call some experts to do the check but the owner say “No problem, this house has the AXIOM maintenance supervision system”. “What’s this?” inquire Mrs. Reds. “It’s simple – replay

the owner - Every appliance has a performance log with consume, life span track, and remote maintenance preventive control. The Heating Ventilation and Air Conditioning (HVAC) system has his own monitoring system made by a distributed sensor system inside the components plus a set of camera furnished with infrared sensor for detecting possible leaks, produced internally of coming from outside, or dangerous overheating of electrical components or other potential risky source. Finally, there is a brand new smell system not only for smokes but also for air quality control, including germs!". "Terrific!!!" - say Mrs. Red – "We want to stay longer!!!" "Well, as a matter of fact, there is more: The AXIOM maintenance supervision system allows progressive inspection of the infrastructures in the remote case of a fault, in this way the home infrastructure have the property of graceful degradation, that is it very rare that you lose the service completely, you should be able to downplay the service while getting also an idea of where the failure occurred!". "Amazing! I want to know more" – was the replay of Mr. Reds.

Technologies: Data collection, data mining
HERTA possible exploitation relevance: 0/10
VIMAR possible exploitation relevance: 7/10

6. Expression Recognition: The game

A new game is spreading on the web; Giulia wants to give it a try. Giulia connects with the site where you can play the game and choose a random companion. A picture of a human face is presented to both players and they have to reproduce the facial expression while reporting the key feature that according to their knowledge describe the expression (metadata). The PC cameras record the expression and a picture is taken. After both players submit their data (picture and text) the results are exhibited to both players: Three pictures (the probe, plus the two pictures of the players), plus the metadata. If the metadata match and the system recognizes the expression, they get the best score; otherwise the score is proportional to their degree of agreement (only facial expression, only metadata, neither). The idea behind the game is to use the computational power of humans to improve the face recognition process and at the same time to get data processing in real time on how people try to mirror facial expressions. The Game can be extended so to include short movies for trend of expressions.

Technologies: --
HERTA possible exploitation relevance: 0/10
VIMAR possible exploitation relevance: 1/10

7. Helping Firefighters in Danger

Stephan is a firefighter. During the fast emergency he and his colleagues have saved a group of people from a restaurant on fire.

The building was equipped with cameras and sensors that have been useful to find the best way to drive the firefighters to find people locked inside and get them out of the building as soon as possible.

Stephan learned that the building was equipped with the AXIOM platform. This system can quickly recognize/process all the main information to guide the firefighters depending on their objective: extinguish the fire or get out of the zone. The same cameras may be used to detect/recognize any health problem in any of the firefighters to request for external help.

Technologies: Indoor localization, presence detection, video analysis (intent recognition)
HERTA possible exploitation relevance: 7/10
VIMAR possible exploitation relevance: 4/10

8. Guiding an Ambulance on Rush Hour

An ambulance has to arrive as fast as possible to the “closest” hospital when there are some traffic jams on the roads. All traffic cameras of all possible paths from the current position of the ambulance to the "closest" hospitals send the information to the AXIOM platform. The AXIOM platform can quickly recognize traffic jams and figure out the best path from the current position of the Ambulance to the closest hospital thanks to the accelerated process of analyzing images. A variant can be the case of a fire truck that should find the fastest path, on rush hour, to a smart AXIOM building that has automatically detected and communicated to emergencies that there is a possible fire in one of their flats/offices.

Technologies: video analysis
HERTA possible exploitation relevance: 8/10
VIMAR possible exploitation relevance: 0/10

9. Smart Building Communication

In a big office building a large company has several employees that work in several different places. However, when anyone is needed the caller only has to dial the number of the desired person and the AXIOM system will redirect the communication to the corresponding room where this person is currently working.

Technologies: indoor localization
HERTA possible exploitation relevance: 0/10
VIMAR possible exploitation relevance: 2/10

10. A hard day's night

Sara’s day begins and end at home but all day long she is outside for work or other activities.

She decides to install AXIOM intelligence inside her home to help her keeping track of all the activities she runs outside home: to monitor health diary, to read her agenda dates, to collect data from her personal wearable devices and communicate with her. This smart home not only manage her home but suggest actions to do when she is out.

E.g. Sara’s home could: -suggest what to buy for dinner: it knows the food that is ended in the fridge and the calories that you have burned today; -recognize when you are outside home in a bad weather for too much time and set the home temperature etc. accordingly; -alert you to hurry up because all your family is already inside the house.

Technologies: Data collection, data analysis, pattern mining, activity prediction
HERTA possible exploitation relevance: 0/10
VIMAR possible exploitation relevance: 3/10

11. My Home Just Cares for Me

A health station inside your home: a camera with special sensors and a dedicated app that helps you when you are ill or you need a first general checkup from doctors. Without going outside, you can communicate with doctors and other caregivers to let them temporary access data analysis about you and your home: temperature, qualitative metrics, etc. The video recording recognizes the different members of your family, lets you have a medical consultation from your bed and collects data from the cloud storage.

Technologies: Video analysis (face identification), data collection
HERTA possible exploitation relevance: 4/10
VIMAR possible exploitation relevance: 2/10

12. If, Then, Video streaming

A special set of cameras with nocturnal and temperature sensors and microphone that can be freely and easily installed inside or outside the house. The cameras interact with an app and a series of IF and THEN statements can be easily set to control their interaction and the interaction with the user device. When the conditions set are true the cameras immediately connect with your device and stream the content live on your mobile phone allowing you to perform different actions. E.g.: - While Don is out for work a postman rings at his home door. A package arrived for him and he is not at home. The postman rings and the camera at the entrance detects movements, it activates, checks the cameras inside and if no movements are detected it streams the content on the user's phone. The user can see the postman and can speak through the camera to decide what to do with the package. - Security Mode: all the cameras inside are set to detect any kind of movement or particularly loud noise. If something happens they stream the content to your phone and start recording saving directly to the cloud.

Technologies: Presence detection, data analysis
HERTA possible exploitation relevance: 5/10
VIMAR possible exploitation relevance: 5/10

13. We all move

A passenger transport coach company installs an AXIOM-based system on their vehicles and main stop points. The modules and the sensors collect data in order to monitor the fleet and to provide a better experience for the passengers. The passengers can see on the screen in the stop points or in their mobile device when the bus they are waiting for will arrive since each vehicle transmits its GPS position. The location data can be also used to monitor the fleet and raise an alarm when a vehicle is stopped for an unexpected time or goes out from a geofence. A passenger can then buy a ticket in the bus with a mobile device using RFID/NFC communication and have a sit. The coaches are equipped with cameras to detect the damage/theft of the equipment's and to detect the number of passengers, making the coach company able to reduce unused journeys or increment the crowd ones. The sensors are also used to monitor the mechanical components with greater risk of breakage for automatic signaling to the mechanics. In order to save fuel, exhaust gases are tracked too.

Technologies: Geolocalization, data analysis, video analysis (intent recognition)
HERTA possible exploitation relevance: 5/10
VIMAR possible exploitation relevance: 0/10

14. Autonomous drones for infrastructure control

Autonomous drones used to analyze structures and buildings that are not easily accessible as skyscrapers, dams, etc. The drones are moving independently and analyze the surface of the crystals of a skyscraper through thermal cameras to see if there are any breaks that may be caused by birds and wind. In the same way they can check the status of a dam. The analysis of the video stream is done directly on the drone and in case of problems the base is notified of the problem by code. Drones have their charging base that can be fixed on the roof of the skyscraper. The drones can also have multiple sensors: thermo room, laser, camera for recognizing objects and, of course, GPS. AXIOM The hardware could be very useful because of low consumption and high computing capacity in real time.

Technologies: Video analysis
HERTA possible exploitation relevance: 5/10
VIMAR possible exploitation relevance: 0/10

15. Natural light control

This scenario describes an evolution of the standard automatic light control system based on PIR sensors. The AXIOM light system in the house or in the office adapts itself accordingly to the activities carried out in various environments or to the external luminosity, and learns the habits of people. The lightning gets modeled according to the movements of the inhabitants through the rooms of the building. AXIOM recognizes the skeleton of people and automatically turns on the light system with a dimming light effect, calibrated on the luminosity of the entrance, when someone gets into a room. AXIOM identifies how many people are in the room, the objects inside the room (like computers, desks, tables, chairs, TV...) and for example, whether people are sitting on a desk or lying down on a sofa watching television. According to different activities, AXIOM optimizes the surrounding lightning, providing a perfect one on the desk without reflecting it on the computer screen. In the same way, after someone spends a few minutes in the room, AXIOM optimizes the light efficiency, progressively reducing the luminosity so that the person inside doesn't notice the change. The AXIOM system identifies the transitional spaces like corridors, and can manage lightning profiles which react to the movement of the person inside the building, turning on and off the lights as he or she passes through the rooms: the lightning system follows the movement of the inhabitant. Other profiles are dedicated to other environments, like the bathroom, where the lightning system is organized differently, according to what the person inside the bathroom is doing, like getting closer to the mirror or having a shower. Also RGB profiles can be managed, to be configured depending on particular situations, considering also the power of colors. During the night the system turns on the light very softly, following the person who is going to the bathroom and bewaring if someone is sleeping. Specific profiles may be selected for the morning wake up, profiles which pair with the alarm clock on the mobile device and/or pair with web services which inform you through lightning about messages receiving. The AXIOM lightning system may be also controlled and configured by smartphone and tablet, so that users to manage lights at home and organize the system as they prefer. Wireless power switches may be installed, allowing a more direct control of the lightning system. Is it possible also to activate

the voice control system through selected words. Sensors: LED light, Motion tracking camera, depth sensor, microphone arrays Pros: energy-consumption efficiency, optimized light design according to schedules and disposition of people in space, natural presence of artificial lightning, learning light.

Technologies: Presence detection, data analysis, video analysis (object recognition, intent recognition), audio analysis (voice recognition)
HERTA possible exploitation relevance: 4/10
VIMAR possible exploitation relevance: 8/10

16. Luxury Resort and Spa

A luxury Resort and Spa give to his customers exclusive ways to enjoy a stay with personalized and smart services offer.

Mario chooses Tuscany Smart Resort because of the architecture seems to be the highest expression of the new smart concept of luxury. Here sensors, cameras, smart objects and spaces blend together in the unique scenery of a prestigious resort.

At the arrival the customer receives a smart wristlet that will service him like a personal smart concierge. The smart device collects health and wellbeing data about the customer and guides him in a personalized stay. Mario enters his room and the ambient temperature changes in line with the personal temperature and heartbeats of his body monitored by the wristlet. Entering the restaurant, he could read personalized menu based on the GI level and cholesterol pathology he reached today so to reach the best benefit from his diet. All the calories burned during sporting activities and Spa are monitored and the diet customized accordingly.

In rooms and other area, the music and lights change according to the personal mood of Mario in order to let him relax and feel as good possible during the stay.

All the data about the ambient and the customers are monitored and modified in order to optimize energy consumptions.

Technologies: Data collection, data analysis
HERTA possible exploitation relevance: 0/10
VIMAR possible exploitation relevance: 4/10

Appendix C

FPGA boards survey

LOGI FPGA

Logi Fpga [43] PROs

- Conceived as shield for Raspy/Beaglebone
- Wishbone bus
- Edu edition useful to teach VHDL to newbies
- Easy bitstream building and programming (via GUI)
- SATA connector (for LVDS signals)
- Wiki with quickstart and manual
- Github: schematics, HDL code, projects, apps, …
- 256 MB SDRAM
- See also LOGI Pi 2

Logi Fpga CONs

- Not standalone, requires Raspy/Beaglebone
- Spartan 6 LX9 (9152 logic cells, and not anymore supported by Xilinx tools)

MiniSpartan6+

MiniSpartan6+ [44] PROs

- Lot of I/O ports for sensors and actuators
- Features a DAC and an audio connector
- Provides an IDE (Scarab) for increasing the programmer's productivity.
- Low cost 75/105$
- FPGA programmer and 32MB of RAM (to implement a microcontroller)
- Standalone board

MiniSpartan6+ CONs

- Scarab IDE does not generate bitstream, Xilinx tools are still needed. Not suited to write code, but only for connections.
- Pinout not Arduino compatible
- Lacks a proper framework / API
- No fast interconnections (seems not suitable/easy to connect to a main processor)
- Low performance FPGA, Spartan6 LX9 or LX25
- Looks discontinued (could not find where to buy it)

Papilio DUO

Papilio DUO [45] PROs

- Papilio DesignLab (~ArdIDE) allows to edit the FPGA circuit (via Xilinx ISE) using Papilio Circuit Library; useful for educational purposes
- Uses Wishbone and ZPUino
- Papilio wings: extension boards with common pins.
- Arduino mega hard-core and synthesizable soft-cores (AVR8, Zpuino)
- Low cost, 89/125€
- Schematics and more on GitHub

Papilio DUO CONs

- ZPUino may be slow
- Up to 2MB of on-board SRAM
- No fast interconnections
- Spartan6 LX9

MOJO

MOJO [46] PROs

- ATmega microcontroler with 84 digital I/O pins and 8 analog inputs.
- Tutorials available
- Low cost, 55$
- Mojo IDE (~ArdIDE) for easing the development.
- Lots of LEDs, push buttons, LED display
- An LED indicates successfully bitstream load
- Arduino bootloader to upload the sketch (embedded into bitstream)

MOJO CONs

- This board is only useful for managing remote devices and sensors.
- No on-board SDRAM
- No hard-core CPU
- Low performance
- No fast interconnections

SMARTZynq

SMARTZynq [47] PROs

- Zynq 7010 or 7020
- Hight performance
- Fast interconnections
- 8Gb main memory

SMARTZynq CONs

- Lack of general I/O ports

- No software/development tools
- Expensive: 412€ + 202€ carrier board
- Not standalone, needs a carrier board

Parallella

Parallella [48] PROs

- Powerful platform: 16-core Epiphany + Zynq7010/20
- 1GB RAM
- The SDK and APIs for Adapteva's chip seems to be solid, with a target backend for GCC
- The company provides a customized port of GDB
- You can download a ready-to-deploy Ubuntu distribution designed for this board
- Very power efficient
- 4 high speed connectors
- Good price (119/149$)

Parallella CONs

- The Epiphany processor has been a commercial failure.
- There are other approaches like NVIDIA's Tegra platform which are also massive parallel, cost effective, and backed by volume/economies of scale.
- Not easy to scale with high bandwidth
- FPGA available only for communication (not computing)
- Not easily extensible with sensors/actuators

Aijuboard

Aijuboard [49] PROs

- Xilinx Zynq 7015
- Cortex-A9 dual core 666MHz
- It's designed for running Plan 9, an operating system developed by Bell Labs
- High performance
- 1GB main memory
- SATA, displayport, SD slot, gigabit Ethernet

Aijuboard CONs

- The board lacks an open source community or company behind the project
- Price (500$)
- No longer produced

Red Pitaya

Red Pitaya [50] PROs

- Designed for replacing oscilloscopes, signal generators and spectrum analyzers.
- Based on the Xilinx Zynq 7010 FPGA so it has Linux support and production ready compilers.
- Cortex-A9 dual-core
- On-board SDRAM 512MB
- Open source software in GitHub
- Board includes Xilinx licences

Red Pitaya CONs

- Price is ~200 euro, not too bad for the hardware
- Expensive, fast analog inputs not of interest for AXIOM project
- Antenna connectors too much specific, removing them it would be possible to lower the price.
- Hard to find any other realistic uses cases beyond signal processing.

OHO

OHO [51] PROs

- Small FPGA-based board designed for replacing microcontrollers.
- Features a considerable amount of I/O pins.
- Small price (~20 euros)

OHO CONs

- A modern high-end microcontroller such as the ATMega would be a much better fit for the use case scenarios promoted by the board manufacturer.
- Spartan 3 is rather old FPGA thought to replace 8/16-bit CPUs (Z80, 502, 8080...)
- No hard-core CPU
- No RAM
- Just for learn FPGA programming, low performance, no interconnections

RetroCade Synth

RetroCade Synth [52] general comments

- Born for musicians and for making retro sound
- Very vertical market (musician hackers)
- MIDI capabilities so it could be potentially used by amateur musicians
- Suitable for audio applications
- SSG (Software-Controlled Sound Generator)
- No other uses
- Spartan 3E 500K or Spartan 6 LX9
- Not stand alone

Papilio

Papilio general comments

- Arduino FPGA shield
- Wishbone bus
- Synthesizable soft-cores (AVR8, Zpuino)
- 4/8 MB of RAM
- Old project (2011), Spartan 3
- Kickstarter campaign failed. Users reported the lack of a FTDI chip for easy programming and asked for more RAM
- Later funded Papilio Duo

TRIFDEV

TRIFDEV [53] general comments

- Add-on board for Arduino based on a Lattice FPGA
- Arduino support via I2C
- Increasing the number of IO of Arduino
- Not made for playing FPGA but targets Arduino users that needs mor I/Os
- No on-board SDRAM
- Low performance, no interconnections

owlBoard

owlBoard [54] general comments

- Big prototyping area
- Lot of LEDs, push buttons
- Aims to be an open source project
- Based on the Spartan6 FPGA
- No hard-core CPU
- Basic prototyping board for FPGA developing, low performance, no Arduino compatible, no fast interconnections
- Board design does not convince, too much DIY
- The Kickstarter campaign failed
- Use MACHXO2-1200 FPGA, only 1280 LUTs
- Kickstarter campaign failed; No production

ALAN

ALAN [55] general comments

- Arduino microcontroller + FPGA
- Easy to program through Arduino IDE
- FPGA can be up to XC6SLX45 (43K LUTS)
- No on-board SDRAM
- Low performance, no fast interconnections

- No CPU
- Arduino seems to be used for device configuration only
- Unsuccessful funding

4CH signal generator

4CH signal generator [56] general comments

- This is a 4-channel fully differential USB powered signal generator based on FPGA + 4x 100 MS/s DAC
- 2 available FPGA devices (Lattice and Xilinx)
- Low performance, no fast interconnections, the functions are just limited to signal generator
- No CPU
- Not useful for the average user
- Kickstarter campaign failed

Logitraxx

Logitraxx [57] general comments

- A board designed for amateur wheeled robots
- Provides onboard connectors for plugging in small motors
- On-board SDRAM (8/64MB)
- Compatible with many Arduino shields
- Price is 165$
- No software/SDK provided
- No fast interconnection
- Only builds a simple robot
- No CPU
- Kickstarter campaign failed 4 times. Users reported VHDL is hard and asked for GUIs

KromaLights

KromaLights [58] general comments

- Aims to manage LEDs
- FPGA + ARM hard-core CPU
- On-board 256MB SDRAM
- Arduino Due (SAM3X) compatible, SDK available
- Easy to use led panel
- The uses cases are limited. Designed for controlling LEDs.
- Low performance, no fast interconnections
- CPU/FPGA not integrated
- FPGA only used to control the LEDs
- The funding was cancelled.

Crystal Board

Crystal Board [59] general comments

- Based on the Chinese Rockchip RK3188 4-core ARM SoC attached to a Spartan-6 FPGA and an AVR microcontroller
- Features similar to ours
- Powerful processing, Arduino compatible, high number of peripherals available, different programming languages
- No fast interconnections, slow interconnect between the small FPGA and SoC
- Difficult FPGA programmability
- Kickstarter campaign failed

PSHDL Board

PSHDL Board [60] general comments

- Based on an Actel FPGA that provides connectivity and scalability across multiple boards
- Good connectivity with lots of IOs
- Uses PSHDL for easy programming, a simplified VHDL
- The connectivity between several boards seems designed for just sending low-bandwidth I/O signals
- No purpose except learn FPGA programming
- No CPU
- Can be difficult to program
- Funding cancelled

Helix-4

Helix-4 [61] general comments

- Just an FPGA module for controlling I/O signals
- Seems to be designed for education purposes.
- Altera Cyclone 4 FPGA, 22k logic elements
- Good price
- Arduino compatible
- No hard-core CPU
- 4MB of RAM
- Just for learn FPGA programming, low performance, no fast interconnections
- Only Arduino UNO and USB connectivity (no network)
- Fundraising failed

NOTE: The survey will go on until the end of the projects, the next board to be included are:

- Numato Opsis [62]
- XLR8 [63]
- Snickerdoodle [64]

References

[1] https://scholar.google.it/scholar?q=%22smart+video+surveillance%22&btnG=&hl=it&as_sdt=0%2C5&as_ylo=1970&as_yhi=2014
[2] https://scholar.google.it/scholar?q=%22smart+video+surveillance%22+learning&btnG=&hl=it&as_sdt=0%2C5&as_ylo=1970&as_yhi=2014
[3] https://scholar.google.it/scholar?q=%22smart+home%22&btnG=&hl=it&as_sdt=0%2C5&as_ylo=1970&as_yhi=2014
[4] https://scholar.google.it/scholar?q=%22smart+home%22+learning&btnG=&hl=it&as_sdt=0%2C5&as_ylo=1970&as_yhi=2014
[5] http://ersatzlabs.com
[6] http://finance.yahoo.com/news/ersatz-labs-launches-first-gpu-110000774.html
[7] http://www.iflytek.com/
[8] https://www.crunchbase.com/organization/anhui-ustc-iflytek-science-and-technology-co-ltd/funding-rounds
[9] http://techcrunch.com/2012/08/09/chinese-manufacturers-carriers-and-developers-form-an-alliance-to-block-apples-siri-in-china/
[10] http://www.percepto.co
[11] http://venturebeat.com/2015/08/11/percepto-funding/
[12] https://www.kairos.com/
[13] https://www.crunchbase.com/organization/kairos-ar/funding-rounds
[14] http://techcrunch.com/2012/09/17/kairos-lands-500k-seed-funding-timeclock-facial-recognition/
[15] http://clarify.io
[16] https://www.crunchbase.com/organization/clarify-inc/funding-rounds
[17] https://www.pehub.com/2014/10/projector-ventures-silverton-partners-join-1-mln-seed-funding-in-clarify/
[18] https://www.ecoventsystems.com/
[19] https://www.crunchbase.com/organization/ecovent/funding-rounds
[20] http://www.betaboston.com/news/2015/07/27/smart-home-heating-startup-ecovent-lands-6-9m-investment-touts-1m-in-pre-orders/
[21] https://dato.com/
[22] https://www.crunchbase.com/organization/graphlab/funding-rounds
[23] http://techcrunch.com/2015/01/08/machine-learning-startup-graphlab-gets-a-new-name-and-an-18-5m-check/
[24] http://emotient.com
[25] https://www.crunchbase.com/organization/emotient/funding-rounds
[26] http://techcrunch.com/2014/03/06/emotient-raises-6m-for-its-facial-expression-recognition-tech-debuts-sentiment-analysis-app-for-google-glass/
[27] http://www.enlitic.com/
[28] https://www.crunchbase.com/organization/enlitic/funding-rounds
[29] http://techcrunch.com/2015/05/02/rise-of-the-quants/
[30] http://www.eyeristech.com/
[31] http://breakthroughanalysis.com/2014/02/21/sentiment-analysis-innovation/
[32] http://orbeus.com/
[33] https://www.crunchbase.com/organization/orbeus/funding-rounds
[34] http://www.bloomberg.com/news/articles/2015-02-18/even-the-smartest-ai-can-mistake-zuckerberg-for-a-cardigan
[35] http://www.facety.me
[36] http://www.clarifai.com/
[37] https://www.crunchbase.com/organization/clarifai/funding-rounds
[38] http://techcrunch.com/2015/06/18/deep-learning-and-the-future-of-search-engine-optimization/#.jn8js7:Uwgw
[39] http://yourstory.com/2015/08/hyperverge-funding/
[40] http://www.clarifai.com/about
[41] https://www.camio.com/
[42] http://www.technologyreview.com/news/540396/using-deep-learning-to-make-video-surveillance-smarter/
[43] https://www.kickstarter.com/projects/1575992013/logi-fpga-development-board-for-raspberry-pi-beagl
[44] https://www.kickstarter.com/projects/1812459948/minispartan6-a-powerful-fpga-board-and-easy-to-use
[45] https://www.kickstarter.com/projects/13588168/papilio-duo-drag-and-drop-fpga-circuit-lab-for-mak
[46] https://www.kickstarter.com/projects/1106670630/mojo-digital-design-for-the-hobbyist
[47] http://www.trenz-electronic.de/products/fpga-boards/soc-e.html
[48] https://www.parallella.org
[49] https://www.indiegogo.com/projects/aijuboard#/story
[50] http://www.trenz-electronic.de/products/fpga-boards/red-pitaya.html
[51] http://www.trenz-electronic.de/products/fpga-boards/oho-elektronik.html
[52] https://www.kickstarter.com/projects/13588168/retrocade-synth-one-chiptune-board-to-rule-them-al
[53] https://www.kickstarter.com/projects/2050495186/trifdev-a-re-programmable-fpga-i-o-expander-for-ar
[54] https://www.kickstarter.com/projects/1903665837/owlboard-fpga-devkit-open-hardware-lab-at-oregon-t
[55] https://www.kickstarter.com/projects/1034459656/alan-an-arduino-compatible-mircoprocessor-and-xili

[56] https://www.kickstarter.com/projects/333918172/usb-powered-4-channel-signal-generator

[57] https://www.kickstarter.com/projects/989837359/logitraxx-tracked-robot-kit

[58] https://www.kickstarter.com/projects/105536416/kromalights-led-controller-and-led-panels

[59] https://www.kickstarter.com/projects/redcrytal/crystal-board-fpga-arduino-arm-soc-in-a-single-dev

[60] https://www.indiegogo.com/projects/pshdl-board#/story

[61] https://www.indiegogo.com/projects/helix-4-the-fpga-module-for-everyone#/story

[62] https://forums.xilinx.com/t5/Xcell-Daily-Blog/Cool-349-Open-Source-FPGA-based-Video-Dev-Platform-supports-HDMI/bap/653642

[63] https://www.kickstarter.com/projects/814866572/xlr8-arduino-compatible-fpga-based-application-acc

[64] https://libav.org/

[65] https://www.crowdsupply.com/krtkl/snickerdoodle

[66] S. Wong, A. Brandon, F. Anjam, R. Seedorf, R. Giorgi, Z. Yu, N. Puzovic, S. Mckee, Magnus Sjaelander and Georgios Keramidas, "Early Results from ERA – Embedded Reconfigurable Architectures", 9th IEEE Int.l Conf. on Industrial Informatics (INDIN), Lisbon, Portugal, Jul 2011, pp. 816-822.

[67] ERA Benchmark suite: http://www.dii.unisi.it/~giorgi/ebs/

[68] S. Wong, L. Carro, S. Kavvadias, G. Keramidas, F. Papariello, C. Scordino, R. Giorgi, S. Kaxiras, "Embedded reconfigurable architectures", ACM Proc. 2012 international conference on Compilers, architectures and synthesis for embedded systems (CASES), New York, NY, USA, 2012, pp. 2.