

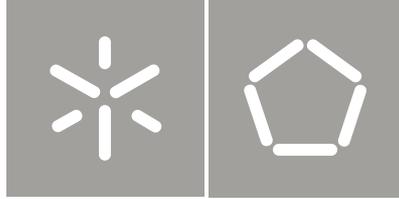
Universidade do Minho  
Escola de Engenharia

Gustavo dos Santos Mações

Augmented and Virtual Reality Activity  
Representation and Visualization

Tese de Mestrado  
Mestrado em Informática

Trabalho efectuado sob a orientação do  
Professor Doutor Luis Paulo Peixoto dos Santos



Universidade do Minho  
Escola de Engenharia

Gustavo dos Santos Mações

Augmented and Virtual Reality Activity  
Representation and Visualization

Tese de Mestrado  
Mestrado em Informática

Trabalho efectuado sob a orientação do  
Professor Doutor Luis Paulo Peixoto dos Santos

Outubro de 2010

## DECLARAÇÃO

Nome: Gustavo dos Santos Mações

Endereço electrónico: [gustavomacaes@hotmail.com](mailto:gustavomacaes@hotmail.com) Telefone: 252691194 / 963719175

Número do Bilhete de Identidade: 12940335

Título da Tese:

Augmented and Virtual Reality Activity Representation and Visualization

Orientador:

Professor Doutor Luís Paulo Peixoto dos Santos

Ano de conclusão: 2010

Designação do Mestrado:

Mestrado em Informática

É AUTORIZADA A REPRODUÇÃO PARCIAL DESTA TESE/TRABALHO APENAS PARA EFEITOS DE INVESTIGAÇÃO, MEDIANTE DECLARAÇÃO ESCRITA DO INTERESSADO, QUE A TAL SE COMPROMETE;

Universidade do Minho, 28/10/2010

Assinatura: \_\_\_\_\_

## Acknowledgment

I would like to leave a word of thanks to all who helped and supported me during the execution of this thesis.

# Augmented and Virtual Reality Activity Representation and Visualization

## Abstract

Nowadays many resources are spent creating technical documentation. To optimize this and to answer the needs of a new type of documentation aimed at assisting the development, production and maintenance of products, there is a technology that displays a fair amount of potential: Augmented Reality.

Augmented Reality mixes the real world with 3D objects and can be displayed within the user's field of vision with relatively simple eyewear displays. This is important in the industry, where mobility of the user must be assured.

The aim of this thesis is to contribute to the development of this new type of documentation. In order to achieve this, a structure called activity workflow was developed, capable of storing anything from simple tasks to tasks with a high degree of complexity. For that, each task will be decomposed into groups of atomic actions, in this process the entities involved in the activity and their movements in the task are stored. This structure has the advantage of allowing the reuse of movements; in other words, in the case of similar movements in the activity only one is stored and when needed the movement can be played in different starting points.

Then the activity workflow is read by a player that will generate the visualization in augmented or virtual reality. The visualization in virtual reality is based in the concept "*learn by demonstrating*" in which the operator learns how to perform a new task by watching a video. The augmented reality visualization is meant to be used by an inexperienced operator that is guided through the task step by step. This immersive approach is intended to keep the user focused in his task.

# Representação e Visualização de Actividades em Realidade Aumentada e Virtual

## Resumo

Nos dias de hoje muitos recursos são gastos na criação de documentação técnica. Para otimizar isto e responder às necessidades de um novo tipo de documentação capaz, de ajudar no desenvolvimento, produção e manutenção de produtos existe uma tecnologia que demonstra grande potencial: Realidade Aumentada.

A realidade aumentada mistura o mundo real com objectos 3D e pode ser visualizada no campo de visão do utilizador com recurso a uns óculos apropriados. Isto é importante na indústria onde a mobilidade do utilizador tem de ser assegurada.

O objectivo desta tese é contribuir para o desenvolvimento deste novo tipo de documentação. De modo a atingir isto, uma estrutura designada de *activity workflow* foi desenvolvida, que é capaz de conter desde simples tarefas a outras com elevado grau de complexidade. Para tal, cada actividade será decomposta em grupos de acções atómicas. Neste processo as entidades envolvidas e os seus movimentos são guardados. Esta estrutura tem a vantagem de permitir a reutilização de movimentos; no caso de estes serem similares na actividade apenas um é guardado e quando necessário o movimento pode ser reproduzido em diferentes pontos.

O *activity workflow* é depois lido por uma aplicação que vai gerar a visualização em realidade aumentada ou virtual. A visualização em realidade virtual baseia-se no conceito “aprender por demonstração” em que o utilizador aprende como executar uma nova tarefa através da observação de um vídeo. A visualização em realidade aumentada é para utilizadores inexperientes que são guiados pela tarefa passo a passo. Esta abordagem imersiva pretende manter o utilizador focado na sua tarefa.

## Index

1. INTRODUCTION.....	8
2. COGNITO .....	10
3. STATE OF ART .....	12
3.1. POWER SPACE.....	13
3.2. AMIRE .....	15
3.3. ARVIKA.....	17
3.4. MARS.....	19
3.5. TEMPLATE BASE AUTHORIZING .....	21
3.6. COMPARATIVE EFFECTIVENESS IN OBJECT ASSEMBLY .....	24
4. GRAPHIC PACKAGE.....	27
4.1. VIRTUAL AND AUGMENTED REALITY PLAYER.....	27
4.2. ACTIVITY WORKFLOW .....	29
4.3. CASE STUDY .....	32
4.4. USE-CASE DIAGRAM .....	34
4.5. CLASS DIAGRAM.....	35
4.6. RESULTS .....	37
5. CONCLUSION.....	39
REFERENCES.....	40

## Figures Index

Figure 1. COGNITO Hardware Representation .....	10
Figure 2. Overall architecture of COGNITO.....	11
Figure 3. PowerSpace in Development Environment .....	13
Figure 4. PowerSpace viewer in AR mode .....	13
Figure 5. Power Space Program Structure .....	14
Figure 6. AMIRE Guggenheim museum prototype.....	15
Figure 7. AMIRE Placing element in position animation.....	16
Figure 8. ARVIKA VW Golf and car crash test.....	17
Figure 9. ARVIKA Aircraft system prototype .....	18
Figure 10. MARS Authoring component with 3D viewer .....	19
Figure 11. MARS world with miniature map.....	19
Figure 12. MARS System architecture. ....	20
Figure 13. Representation of the task on the left and work step on the right. ....	22
Figure 14. View of the application in the user mode. ....	23
Figure 15. Test environments .....	25
Figure 16. AR system view mode.....	25
Figure 17. Graphic Result - Time spent in assembly task.....	25
Figure 18. Graphic Result - Average number of error in each treatment.....	26
Figure 19. Average score on NASA TLX rating .....	26
Figure 20. VR Player architecture .....	27
Figure 21. AR Player architecture .....	28
Figure 22. Activity Workflow Tree .....	29
Figure 23. Activity Workflow Representation .....	31
Figure 24. UML Case Diagram for augmented reality player scenario .....	34
Figure 25. Augmented Reality Player Class Diagram .....	35
Figure 26. COGNITO VR Visualization.....	37
Figure 27. COGNITO AR Visualization.....	38

## List of Abbreviations

AR – Augmented Reality

AW – Activity Workflow

BSN - On-Body Sensor Network

CAI - Computed Assisted Instruction

COGNITO - Cognitive Workflow Capturing and Rendering with On Body Sensor Networks

HMD – Head Mounted Display

OpenCV - Open Source Computer Vision Library

SLAM - Simultaneous Localization and Mapping

VR – Virtual Reality

XML - eXtensible Markup Language

# 1. INTRODUCTION

Cutting down the amount of time and effort spent in training and specialized technical support is very important in the industrial production lines, because they're directly attached to the costs of this process.

The pursuit of the automation of these processes has been the object of scientific research in recent years, with special focus on the gathering, analysis and treatment of human cognitive patterns, making use of on-body sensor networks and taking advantage of the development of automatic and interactive systems that support training and execution of critical, highly specialized tasks. Although the complexity in the stage of data collection and interpretation is already very high, other factors must be taken into account as just as important and equally conducting to other lines of research - data conversion to legible structures (called action models) and their subsequent visualization in an assisted and interactive way such as in a manual for an assembling task. At this point, some questions appear, such as the versatility and abstractions that are needed for the structure of the action models - which must support content that includes semantic, spatial or relational aspects of the generated visualization itself - that should guarantee a high usability level and a simple cognitive interpretation by the end user during task execution.

The purpose of this work is to create the representation of visualizations in Virtual or Augmented Reality systems that work as training manuals for specialized technicians, and as systems that support the execution of individual tasks.

This thesis is contained in the scope of the Work Package 7 Graphics Encoding, Editing and Viewing (WP7) from the project COGNITO: Cognitive Workflow Capturing and Rendering with On Body Sensor Networks, funded by the European Union Seventh Framework Programme (FP7). Emphasis is placed in tasks that involve the manual handling of objects and tools, given its relevance in industrial contexts.

WP7's goal is the development of the essential concepts, editors and players that allow:

- The (semi-)automatic composition of the augmented reality visualizations, based on the workflow descriptions and in data acquired in real-time through the sensors;
- The development of augmented reality systems that visualize the compositions mentioned before;

This thesis focuses on the creation of the visualization part and aims to establish foundations that will later potentiate the development of applications usable in production environments. This will imply:

- Identification and characterization of the abstract entities to be manipulated; examples of these entities include objects and tools, semantic and spatial relationships between them.
- Specification of the data structures for representation of the above mentioned entities, and their respective parameterization for concrete examples;
- Specification and development of the model that supports the description of the visualizations, based on: the manipulation of concrete entities (parameterized instances of the abstract entities); the description of the workflow associated to each specific task (action models); and the inputs from sensors that provide information about the user's actions;
- Augmented and virtual reality player responsible for creating the visualization of the action models;

The goal of COGNITO is to allow the capturing, processing, understanding and reproduction of cognitive patterns in the workflow. The intended final result is the conception of methodologies that allow the development of Augmented Reality (AR) systems that automatically recognize tasks performed by the users and accordingly provide them with support, either for training purposes, or to guarantee the successful conclusion of the task they're performing.

In section two is explained the overall architecture of the COGNITO system, in section three it is presented a state of art of the most relevant authoring tools in the past ten years. Section four will present the work developed in this thesis and finally in section five the conclusions.

## 2. COGNITO

The COGNITO system involves novel customized hardware with multiple sensors which will deliver a large amount of data at high speed. The software system will work with streams of heterogeneous sensor data in real-time besides rendering capabilities in an augmented reality display. The goal is to develop a mobile and wearable system, which requires hardware miniaturization, good ergonomic and optimization of the processing load.

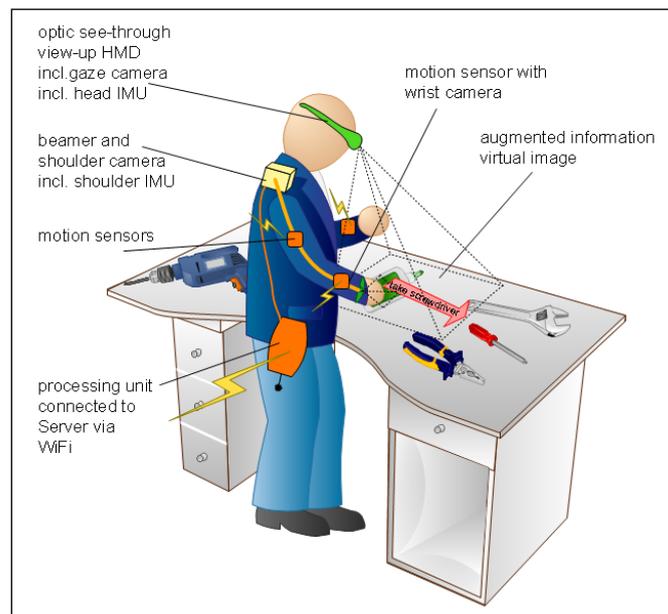


Figure 1. COGNITO Hardware Representation

Four main building blocks of the COGNITO system have been identified:

**On-Body Sensor Network (BSN) and Head-Mounted Display (HMD):** inertial measurement units (IMUs), cameras, eye-tracking and miniature projectors are combined in a sensor network. A monocular head-mounted display (HMD) provides the system feed-back and user assistance information.

**Low-Level Sensor processing:** processes the measurements from the BSN and estimates the positions of the operator, his hands, and relevant objects in the environment. The responsibility is shared between two modules Sensor Fusion and Computer Vision. The first is responsible for the integration of the inertial sensors, and high level vision information. The second is responsible for processing the raw image data from the cameras.

Workflow Recovery and Monitoring: receives a sequence of instantaneous configurations for the operator, her hands, objects and parts of objects contained within a work space from the Computer Vision. This building block processes it and provides the marker (start and stop timestamp) for ongoing/future atomic events in the current workflow sequence. It also estimates the position of objects in the same workspace and provides feedback to the Computer Vision.

Graphics Package: is used to produce the proper graphics for editing workflows as well as aiding the user during task execution using an augmented reality viewer. This thesis will focus on this block.

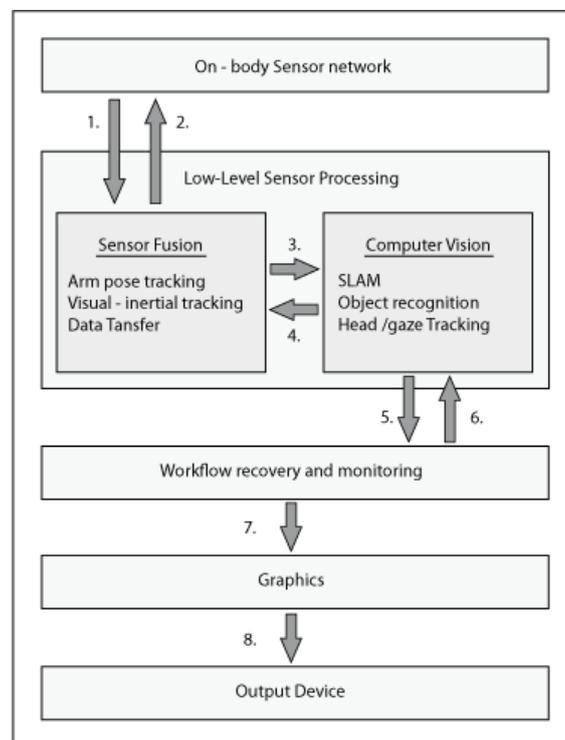


Figure 2. Overall architecture of COGNITO

### 3. STATE OF ART

Industrial environments need a solution to the high cost of technical documentation and to train their employees in the assembling and repairing of complex products. It is a difficult task to cover all versions and alternative models made nowadays.

The scenario of a worker switching between reading instructions in a manual and repairing the product itself must change. In some cases the worker has to accomplish an unusual task, when doing it for the first time or as young trainee who may make mistakes that could be avoided with an augmented reality system (AR).

AR enriches the user's environment by introducing three dimensional (3D) objects in the real world and mixing them. With AR it is possible to have animations, information and interaction in the task that the user has to do. There is a study (Tang A., 2003) that shows improvement in the assembly task using AR based systems and that computer assisted instructions are useful for complex processes. Although AR has evolved a lot (Azuma R., 2001) in the past years, there are challenges to this project such as the generation of high quality visualizations, precision in position and orientation of objects and the real time interaction.

Another problem faced is that the use of documentation typed by hand is not possible. Most of the manuals use common parts, but since they are not stored in a structured way (such as in a database), reuse is not possible and a cataloguing of the data is needed (Stock I. W. M., 2005). The revision of the documentation would be easier and less time would be spent in edition. In the industry that uses the computer in product creation this data must be used to improve the documentation in a new way (Stock I., 2006).

The solution for this is an AR authoring system where technical writers with no IT skills can create and rewrite manuals in an easy and fast way. With the information stored in this kind of systems it is possible to have a real-time response to the information received from the user. This approach takes manuals to a new level, transforming static information into a dynamic visualization of data.

In this section an overview of five projects in this context will be presented: PowerSpace a pragmatic approach to AR authoring; AMIRE an authoring mixed reality

component and framework based approach; ARVIKA Augmented Reality for development, production and service; MARS an Mobile AR system and Template based Authoring for AR bases service scenarios.

### 3.1.POWER SPACE

This project, PowerSpace (Haringer M., 2002)has an interesting approach as it uses Microsoft Power Point as the base of the program. The information that is shown could be arranged as shown in figure 3. Then it is exported in XML based format that is processed by the PowerSpace editor.



Figure 3. PowerSpace in Development Environment

After editing the presentation it can be played in virtual reality (VR) or AR mode by PowerSpace Viewer latestest one can be seen in figure 4. In this way the authors build an AR authoring system simple and easy to work with.

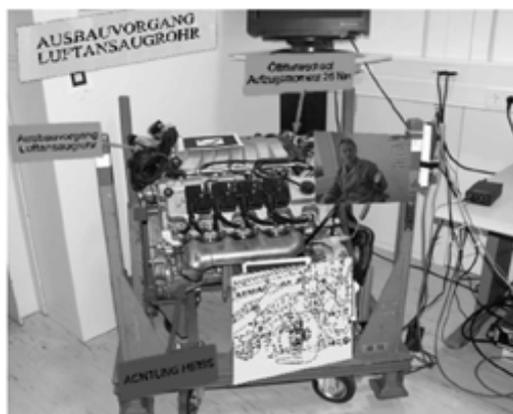


Figure 4. PowerSpace viewer in AR mode

It is possible to use images, videos, 3D objects and text in the system and due to the amount and diversity of the data that needs to be stored in a structured form they used XML. Figure 5 illustrates the structure of the program that was used.

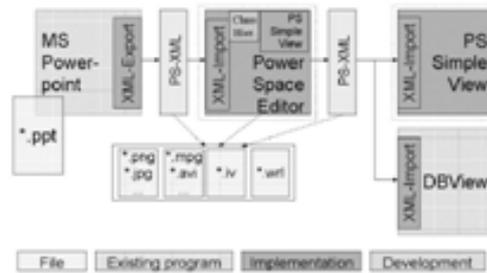


Figure 5. Power Space Program Structure

The core concept used in this project can be divided in four steps that need to be done to create a VR/AR guide:

- Create and arrange elements in Microsoft power point
- Arrange the 3D world in PowerSpace editor
- Define the order between slides and inside steps of each slide
- Export and test the application in PowerSpace viewer

This method allows building AR guides in a fast and easy way. Although there is some work to be done such, as problems with working with small parts and some issues when using the system in open spaces. It could be considered a good starting point to the creation of a generic AR authoring system.

### 3.2.AMIRE

AMIRE (Dorner R., 2002) is an authoring mixed reality component and framework. It has the objective of creating an application that allows users without programming skills to design and implement mixed reality applications, in a simple way. To allow this the AMIRE framework has a visual authoring tool. One of the aims of the project is to be a base for mixed reality standardization.

The framework allows the configuration of components by properties that represent the data. Communication between components is made by in and out slots and it is also possible to create new components that are easy to work and are handled like a native component. The framework also provides conventions for 2D and 3D components helping in placement and alignment, picking system and user interactions. The object detection in the framework is abstracted, it only needs to create a detection component.

Distinct samples were made to prove the flexibility of the application. The first is an application of an oil refinery with the objective of increasing the performance in assembly training systems. The second is an experience from Guggenheim museum in Spain that aimed to give visitors more information of pieces not available to the public. Figure 6 shows sample of that work. In this case the visitors could see the parts of a broken vase fit.



Figure 6. AMIRE Guggenheim museum prototype

Other example based on AMIRE is a mixed reality assembly instructor (Zauner J., 2003). Using a head mounted display the user receives information step-by-step on how to assembly a furniture piece. The engine of the application works like a state machine, checking conditions. The first step is to identify the base element.

Then when more parts are attached to this, it verifies if it is the last piece of the step in order to check the end of this step and move to the next one.

One interesting feature of the application is the animation when the element is interpolated from the origin to the end position showing the correct way of assembling as presented in figure 7. Another approach is the joint of the small elements that could not place markers to the big elements, solving one of the common problems in this kind of applications.

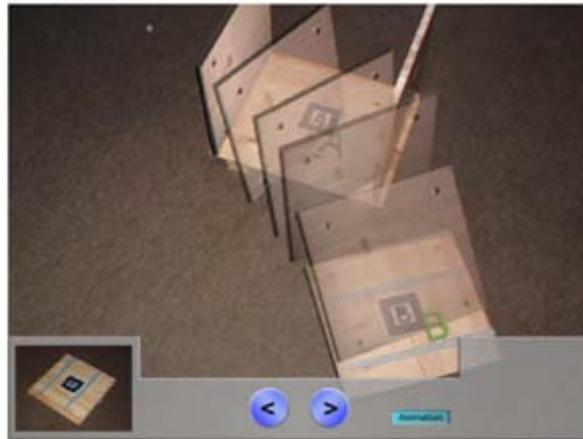


Figure 7. AMIRE Placing element in position animation

### 3.3. ARVIKA

ARVIKA (Friedrich W., 2002) is one of the first projects in this area. Its goal is to develop applications for the car and aerospace industries, power plants and service on production machines.

This project uses AR in the development of user oriented applications to support the implementation, production and servicing of complex products. The application is being tested in important automobile manufacturers such as VW, Audi, Ford, Daimler Chrysler and aircraft construction on Airbus. Small and medium companies can also use the application to help them reinforce their position in the market and to be more competitive.

The technical level is focused in having high quality visualization with pixel precision, a spatial volume of 28 m<sup>3</sup> and a hybrid rendering approach, allowing 3D real world depths, 2D video and 2D and 3D text information. The conditions given to the user via eyewear display replaces the instruction manual. It allows the presentation of information in the user field of vision, such as assembly instructions or warning messages.

ARVIKA is used in production lines to optimize assembly sequence; it gives the information to the user through an eyewear display using a networked wearable computer, helping them to investigate possible optimization and training workers in the learning phase of a new product. It is also used to verify the quality of the product in the finish area of car production.

It is also used in car crash tests as a comparative test by overlaying virtual and real components allowing fast results. In figure 8 is presented the virtual and the real car.



Figure 8. ARVIKA VW Golf and car crash test

It is used in complex systems in aircraft production such as electric cables and water pipes. The virtual scenario is built, based on digital data used in the product life cycle. The use of data such as plants, product information and others, when converted, allow the creation of work sequences. It also provides precise 3D objects for visual representation and synchronism with the real world, very important in AR systems. In figure 9 a prototype of the application used in this field is shown.



Figure 9. ARVIKA Aircraft system prototype

Another field of use of this system is the production of machines where all the history of the machine is available. It helps the user find the problem using fault trees, search engines or predefined routine tests. Troubleshooting is easier when compared with similar problems that have already occurred. One interesting approach is the use of a visual channel that gives the user the possibility to contact with the company support and they could look over the customer's shoulder, which means both see the same allowing a cutback in the time spent resolving the problem. For more interactive support by the company, besides audio, they can use circles or pointers to help the user find his next move.

In this project another important point is the earlier feedback from skilled users that worked with the prototypes. This approach helped the system to improve with the user's experience.

### 3.4.MARS

Mobile Augmented Reality System (MARS) (Guven S., 2003) is an authoring tool that was created to respond to the needs of journalism students in Columbia University, New York. The objective is to create documentaries without the help of programmers. The highlights of the project are the editor and the outdoor tracking system.

MARS is composed by an authoring component and a presentation component. Authoring component allows the construction of applications using audio, video, images, text and 3D objects for users without IT skills. One other feature of the application is the possibility to navigate in the 3D world allowing the content creator to position with precision the objects as shown in figure 10.

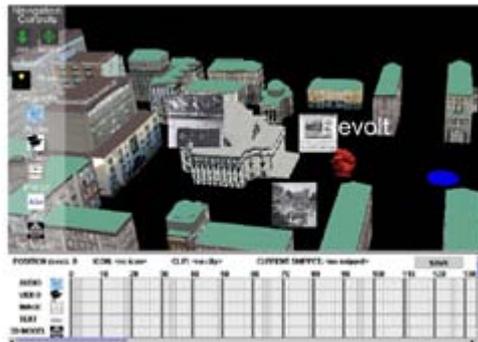


Figure 10. MARS Authoring component with 3D viewer

The presentation component gives the possibility to experiment viewing the documentaries in AR or VR mode. In this component there is a world in miniature that helps the viewer to locate places in the world and find the points of interest represented in figure 11.



Figure 11. MARS world with miniature map

At technical level they used VRML (ISO, 1997), Java and CMIL (Dietz, 2002). CMIL is a XML derived language that stores the relation between digital multimedia files and standard information. This language was extended for full support of project needs. This extension is called CMIL++. In figure 12 is presented a structure of the developed system.

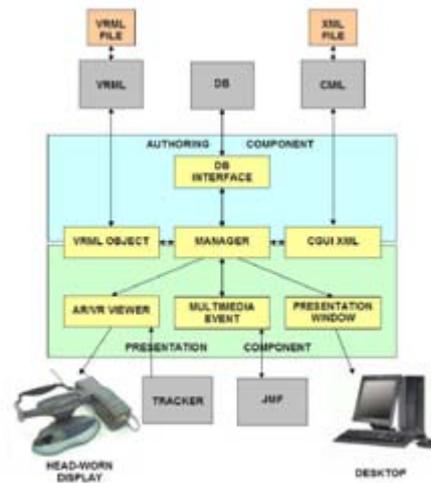


Figure 12. MARS System architecture.

This application was tested by the journalism students and the feedback received used to improve some functionalities of the program.

The results were good and the application gives students the possibility to work with great freedom without the need of a programmer to help them. They now can focus their work in the design and experimentation of new approaches and modify them as they want.

In conclusion the main objective was accomplished and there is an authoring tool to support the needs of the students. The authors also referred the possibility to expand and develop new functionalities to MARS to more advanced animations.

### 3.5.TEMPLATE BASE AUTHORIZING

This project is a response developed by BMW Group and some partners to the need of an AR based (Knoepfle, 2005)authoring tool to technical writers without IT skills.

The case study presented in this article is for car service. The user has an eyewear display and the information is displayed in his field of view combining the real world with animated tools and instructions.

In their analysis of the scenario the focus was to find a pattern in the operations. They realized that all the operations were a combination of atomic tasks that must be executed in a specific order and can be grouped in classes. A typical structure is:

- Kind of fastening
- Tool needed
  - Tool configuration
  - Initial position
  - Movement of the tool
  - Connection point between the tool and the part

Another information found is that most operations are similar. The difference resides only in the parameterization and that most common operations were release fastening, remove part and unlock plug.

The concept developed is that a template contains one operation. In order to arrange multiple operations an abstraction layer is needed to combine them and also allow the possibility to arrange them in a temporal order. A big advantage of this concept is that it fits very well with the way that professional writers work nowadays.

In the template, the operation is defined. It is here that animation is controlled and visibility of parts and tools defined. This approach of generic templates allows reusability. For instance the template remove screw is defined once, stored in a database and can be used by the editor when needed.

The abstraction layer of objects defines rules that allow templates to operate with objects in a simple way. Three kinds of objects could be defined in this scenario:

- Parts (e.g. engine, tire, etc)
- Standard Parts (e.g. screw, clamp, bolt, etc)
- Tools

Some of the rules are the definition that objects should be modeled in a standard way. Another important rule is the definition of the contact point between objects, very important to give precision to the animation.

Temporal order is where the order of the templates is defined. This concept is divided in two parts: work step and task.

- Work step consists in an animation composed by several operations and each of them has a start point in the time duration of the work step. The presentation is played in loop until the user says "next". This approach is used for simple tasks where more than one animation can run at the same time, without overloading the user's cognitive capacity.
- Task is the combination of all work steps that a repair procedure has. It is based in events and played in linear sequence. The system can adapt the work steps based on the feedback of the user.

The conclusions presented point to significant time reducing in creating AR based manuals when compared to standard modeling tools. But there is still some work to be done, for example, the need to develop a style guide for this kind of manual.

A use case of the methodology spoken above is presented below. The repair example is replacing the servomotor. It is a simple task and can be made in six steps.

In the user's point of view the first and second step consists in removing the acoustic cover and the ignition cover. The third step is removing the servomotor itself. This step can be called work step because it is the combination of three operations: unlock plug, release screws and remove part. In figure 13 is a graphic representation of the task and the work step.

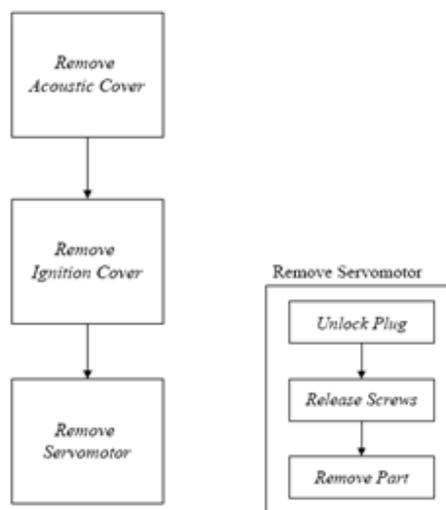


Figure 13. Representation of the task on the left and work step on the right.

From the author's point of view, to prepare this presentation he needs to arrange the steps in the correct order. In every step the author has to assign a component and specify the action that must be executed.

This approach of atomic operations allows the author to say that mounting is similar to dismounting in reverse order. In these cases the authoring tool can generate the mounting manual automatically. This technique of division can be used as many times as wanted allowing any process to be adapted to this hierarchical structure. The atomic operation is a manipulation of a component in a certain way. The motion sequence is contained in the metadata.

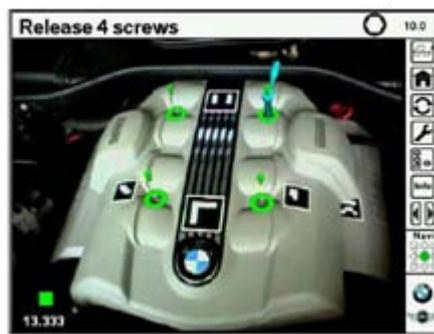


Figure 14. View of the application in the user mode.

### 3.6.COMPARATIVE EFFECTIVENESS IN OBJECT ASSEMBLY

This section presents the work of Tang et al. (Tang A., 2003), which has tested the effectiveness of AR systems in an assembly task. As shown in the previous section, AR is being used to increase productivity in assembly lines, product maintenance and to teach users to do their tasks. It is common to theorize that AR helps the users, allowing them to complete their task faster and reducing mistakes that they could have made. The goal of this study is to acquire knowledge about computer human interaction with AR environments. The theory presented says that AR reduces head and eye movement increasing in this way the eye-on-the-workspace ratio, leading to an improvement in performance. AR reduces the cost in attention switching and it superimposes information in the real world. AR systems can be a complement to human cognitive process.

AR systems reduce the overhead between the manual and the task and can be also used in increasing human attention with arrows, tags, object highlighting and animated tools. AR supports spatial cognition and mental transformation facilitating in task training. In the neuroscience field there are some theories suggesting a strong connection between spatial location and working memory. By connecting information to physical objects and location AR provides a strong leverage of spatial cognition and memory. Using an assembly process as case study three main points will be taken into account:

- Time to complete the task
- Accuracy and error reduction
- User cognitive load

The assembly task was chosen because it reveals most issues and claims made about the advantage of AR systems. It combines spatial registration of virtual and real objects, interaction of virtual and real objects and human coordination in procedural action.

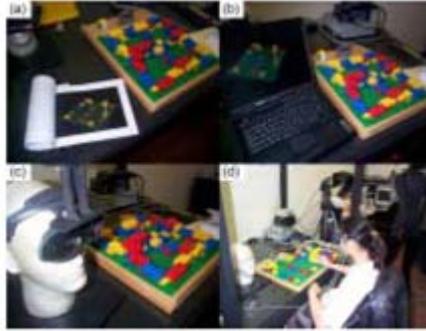


Figure 15. Test environments

The test has four levels. It compares the use of a printed manual, computed assisted instruction (CAI) using a liquid crystal display (LCD), CAI using a head mounted display (HMD) and the AR system. Figure 15 shows the four environments and figure 16 presents the assembly task environment in the AR system view mode.

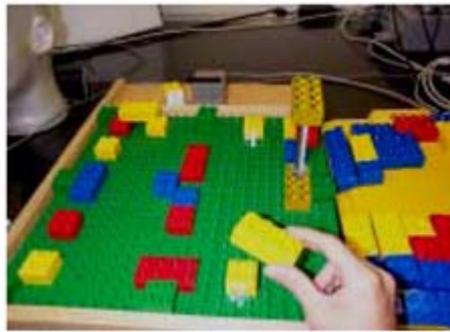


Figure 16. AR system view mode

This experiment was tested by 75 participants that have never experienced any AR environment. The results presented advantages of using a CAI or AR system when comparing time spent with a printed manual the gap was of almost four minutes as represented in figure 17.

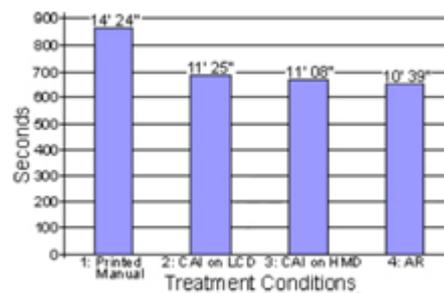


Figure 17. Graphic Result - Time spent in assembly task

When comparing the assembly's accuracy, the AR system is the best because it reduces error rate in 82 percent. In particular it reduces errors due to previous assembly mistakes. As it can be seen in figure 18 dependent errors (errors caused by a previous mistake) achieve the lowest values.

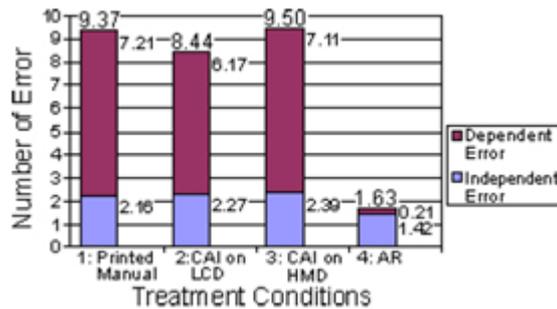


Figure 18. Graphic Result - Average number of error in each treatment

The study also demonstrates a reduction in mental work load when using the AR system with the NASA TLX - Task Load Index tool, as shown in figure 19. This tool assesses subjective workload, by deriving an overall workload score based on a weighted average of ratings on six subscales: Mental Demands, Physical Demands, Temporal Demands, Own Performance, Effort and Frustration.

Treatment Condition	NASA TLX Rating
Printed Manual	13.3/20
CAI on LCD	12.2/20
CAI on HMD	11.0/20
AR	10.0/20

Figure 19. Average score on NASA TLX rating

In conclusion this study provided evidence that AR can improve the performance and reduce mental work load on the user when performing an object assembly. The presentation of information in the user's scope of view creates a more effective tool. However there are still some concerns with hardware limitations in practical uses.

Taking into account all the data described above, which validate the objectives of this thesis, AR systems can help in the industry field allowing the operator to stay focused in his task taking less time to do his job and avoiding assembling errors.

## 4. GRAPHICS PACKAGE

This section presents the work made in the thesis which is divided in five sections, first the Graphics package is explained then a description of the activity workflow is presented, followed by a case study and finally the use case diagram and the class diagram of the system are presented.

### 4.1. VIRTUAL AND AUGMENTED REALITY PLAYER

The Graphics package is responsible for the design, development and implementation of the activity workflow (AW), a virtual reality (VR) player and the augmented reality (AR) player.

The VR player is based in the concept “*learn by demonstrating*”. This player works as a video player to an experienced operator showing how the task should be done. The interaction is limited - the user has the ability to play, pause and rewind the application.

Figure 20 shows how the system works, the player receives the activity workflow XML file, that contains the animations to be played, and then the information is processed by the parser that extracts the information to create the VR animation.

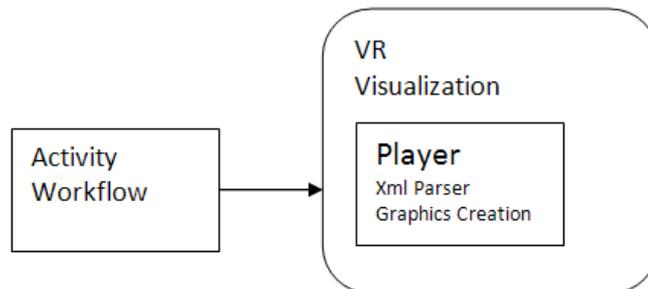


Figure 20. VR Player architecture

The AR player is meant to be used by an inexperienced operator, equipped with the COGNITO system who is guided through the task step by step. The system is adaptive, responding appropriately to uncertainty, difficulties and errors made by the operative, providing cognitive assistance in a timely and effective manner. In this approach the visualization will provide a true interactive scenario because the user actions will be checked out in real time by previous modules.

Figure 21 represents the way in which the AR player works, it is an improved version of the VR player and it uses the pre-recorded activity workflow as a database of animations. The animations are adjusted to the real time data received, which is the starting point where the animation should begin. It also includes the video camera input provided by the HMD to create the background video.

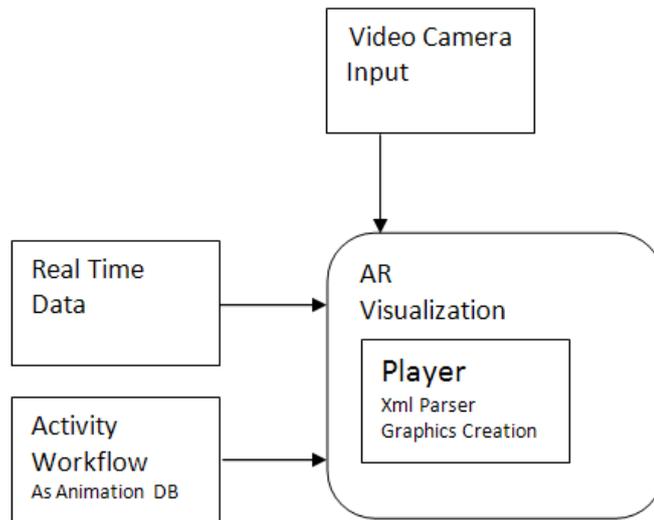


Figure 21. AR Player architecture

Finally, the concept of AW is assumed as being a formal description of compound events composed by primitive events which are atomic operations. The different combinations and sequences of these events describe a workflow which is explained with more detail in the activity workflow section below.

## 4.2. ACTIVITY WORKFLOW

The activity workflow that will be read and processed by the player and edited in the editor will obey a XML data schema. The file in XML format was chosen as a way of representing the activity workflow due to the need that the development teams, which belong to the COGNITO project, had to share data between them while the integration of the blocks described in section two does not occur. The XML format also has the advantage of being human readable, and is flexible enough to store the necessary data.

The AW is composed by two main blocks: Scene and Activity. Scene contains 3D objects and complementary information that will be used by the Activity. The activity contains the animations to be played and is composed of two types of elements: compound events and primitive events. Compound events are composed by other compound events and/or primitive events which are atomic actions. This structure can be visualized as a tree where the workflow is the root node. The Scene, Activity and Compound Events are nodes and the Primitive Event leaves. With this approach it is possible to expand the tree and adapt it to more complex activities.

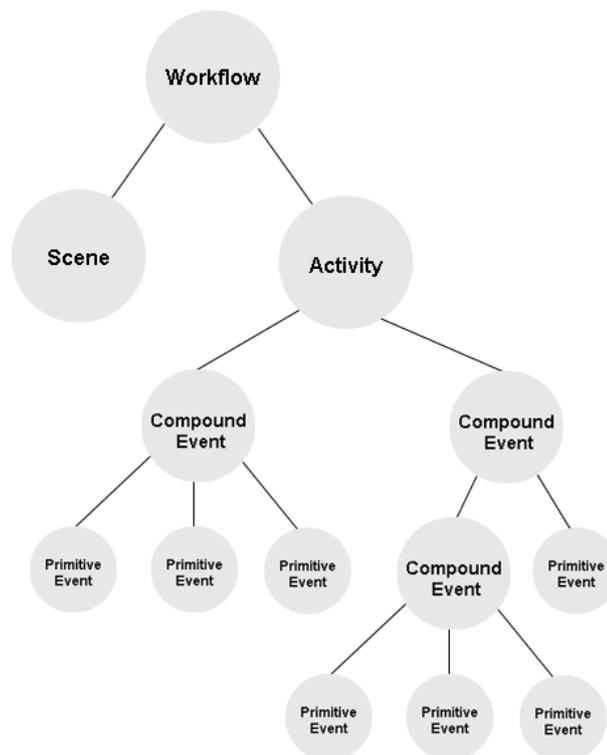


Figure 22. Activity Workflow Tree

This structure is also reusable. Once one animation is stored it can be reused as many times as needed. For instance, in the case study (see next section) the action hammering nail occurs three times, but there is only one animation stored that is played in three different places.

The overall structure of the XML file is described in detail above. It contains the following tags:

Workflow - It identifies the beginning and the end of a workflow. It must include one scene and one activity tag.

Scene - It identifies a group of element tags that are needed in order to represent a task visually. The execution of a task can usually involve people and tools and this tag can hold several elements.

Element – It has as parameters “name” and “id” that are unique strings used to better identify and describe what the element holds. The “type” can be one of these: tool, avatar, part or subpart. Inside the element tags a model tag and an info tag can be defined, according to each scenario requirements. These are used to describe in detail what the element is composed of.

e.g. `<element id="identifier" name="name" type="type">`

Model - Identification of the path and texture of the file that is going to be used to render the element.

Info - It is optional and holds all the complementary information that is going to be delivered to the user during the AR session. The following tags might be used: title, text, image, sound, video and link. Each of them might contain respectively the title, the image or video to be shown (in a part of the AR view), the hyperlink to be displayed or the texts to be written upon the AR visualization.

Activity - It can hold several Compound Event tags. The activity tag encompasses a group of elements that are needed to describe the motion that is needed to represent the scene. It contains the description of translations and rotations that are applied to scene objects during the animation.

Compound Event – It contains the value of “id” and “name” that are unique strings that help to identify the compound event itself. The value of “type” can be “error” (if the action is a corrective one) or “normal”. Each Compound Event contains several Primitive Events that describe an atomic action.

e.g. <CE id=“identifier” name=“name” time=“time” type=“type”>

Primitive Event - Contains the value of “identifier” and “name”, these are unique strings that help to identify the primitive event itself and the time that the event takes. For each primitive event exists a PEdata tag which contains the animation (3d position and orientation) of each element involved in the event.

e.g. <PE id=“identifier” name=“name” time=“time” >

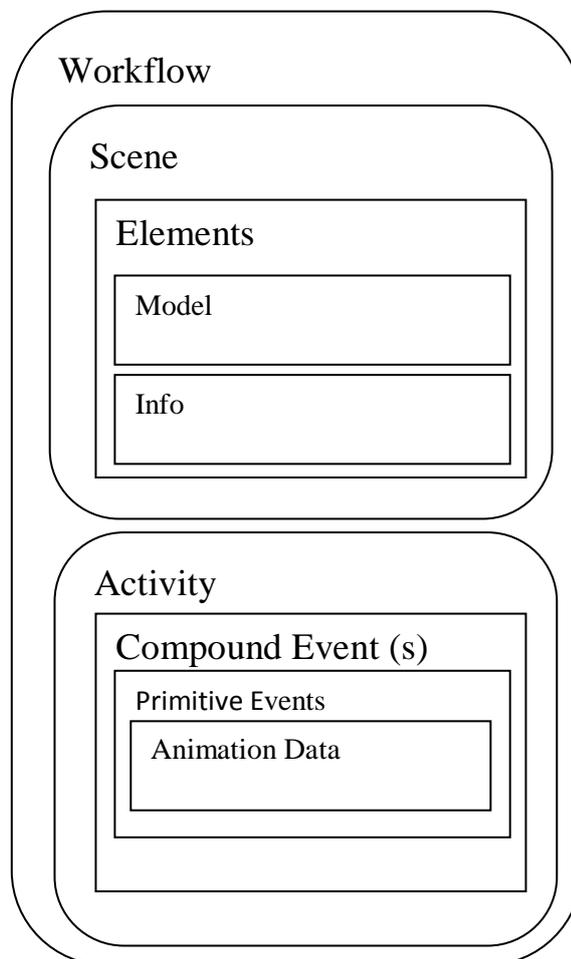


Figure 23. Activity Workflow Representation

This structure describes the characteristics of the classes of objects that are needed in order to create the visualization of the data received from the Workflow module.

### 4.3.CASE STUDY

This section presents the case study. Its aim is to serve as a proof of concept of the activity workflow and consequently the player.

In an initial assessment it may seem simple and too detailed, however these features have a purpose. It is simple to facilitate the first data acquisition, taking into account that the project is at an early stage and the body sensor network isn't ready. The high level of detail is explained taking into account the characteristics of the activity workflow in which the activity is divided into atomic actions, allowing that any activity can be processed and transformed into an activity workflow. In this first approach we have chosen to maintain a linear temporal order in which the atomic actions are executed sequentially.

Below is the description of the case study, the tools involved, the work to be done and the order in which it must be done. The tools involved are a hammer, an electric screwdriver, a box of nails (containing 3+ nails), a box of screws (containing 3+ screws), a piece of wood and two baton pieces with pre-drilled holes.

Using the tools above, the user is expected to hammer 3 nails, in the order specified by the nail markers ( $A > B > C$ ), onto a wooden piece. The user should then fasten 3 screws using an electric screwdriver, into another wooden piece, in the order specified by the provided screw markers ( $W > Y > Z$ ).

Activity	
<b>A1. Placing nails baton</b>	
A1.1	Hand lifts nails baton
A1.2	Places baton within the marked region on the piece of wood
A1.3	Hand releases baton
<b>Temporal constraints</b>	
Do A1.1 before A1.2 and A1.2 before A1.3	
<b>A2. Hammering nail(s)</b>	
A2.1	Hand approaches hammer and lifts it up [referred to as hand1]
A2.2	The "other" hand reaches out to the box of nails to retrieves one nail [referred to as hand2]
A2.3	Hand2 moves to nail marker and inserts the nail into the hole in the nails baton

A2.4	Hand1 moves the hammer to the top of the nail
A2.5	Hand1 moves vertically down to hit the nail then up again [1 or more times]
A2.6	Hand2 releases the nail during the hammering phase
A2.7	For subsequent nails, actions 1.2 - 1.6 are repeated in the same sequence
A2.8	Hand1 puts the hammer down
<b>Temporal constraints</b>	
Do A2.2 before A2.3, A2.1 before A2.4, A2.3 before A2.5, A2.4 before A2.5, A2.6 during A2.5 and A2.5 before A2.8	
<b>A3. Placing screws baton</b>	
A3.1	Hand lifts screws baton
A3.2	Places baton within the marked region on the piece of wood
A3.3	Hand releases baton
<b>Temporal constraints</b>	
Do A3.1 before A3.2 and A3.2 before A3.3	
<b>A4. Fastening screw(s)</b>	
A4.1	Hand approaches screwdriver and lifts it up [referred to as hand1]
A4.2	The “other” hand reaches out to the box of screws to retrieve one nail [referred to as hand2]
A4.3	Hand2 moves to nail marker and inserts the nail into the hole in the screws baton
A4.4	Hand2 releases the screw
A4.5	Hand1 moves the screwdriver to the top of the screw
A4.6	Hand1 switches the power on the screwdriver
A4.7	Hand1 pushes the screwdriver down slowly until the screw is fixed
A4.8	Hand1 powers off the screwdriver
A4.9	For subsequent screws, actions 2.2 - 2.8 are repeated
A4.10	Hand1 puts the screwdriver down
<b>Temporal constraints</b>	
Do A4.1 before A4.5, A4.2 before A4.3, A4.3 before A4.4, A4.4 before A4.6, A4.5 during A4.6, A4.6 before A4.7, A4.7 before A4.8 and A4.8 before A4.10	

#### 4.4. USE-CASE DIAGRAM

Figure 24 shows the augmented reality player scenario. The data needed to render the AR visualization is sent by the Computer Vision to the Workflow Module that analyzes it in real-time.

This data encompasses the virtual objects positions and orientations, and the AW that has to be played according to the action that the user performed: it can be correct or incorrect. The user starts his session choosing what training he wants to have, from a database. Once he loads his workflow and the Computer Vision and Workflow modules start running, the augmented reality module starts receiving the data mentioned above (several AWs are in sequence being read/received). The user is able to play, pause, stop or rewind the visualization (it means to stop, read or navigate through the actual AW being processed by the AR player). The AR visualization is created in real-time, and objects (multimedia or 3D models) are loaded dynamically from the database.

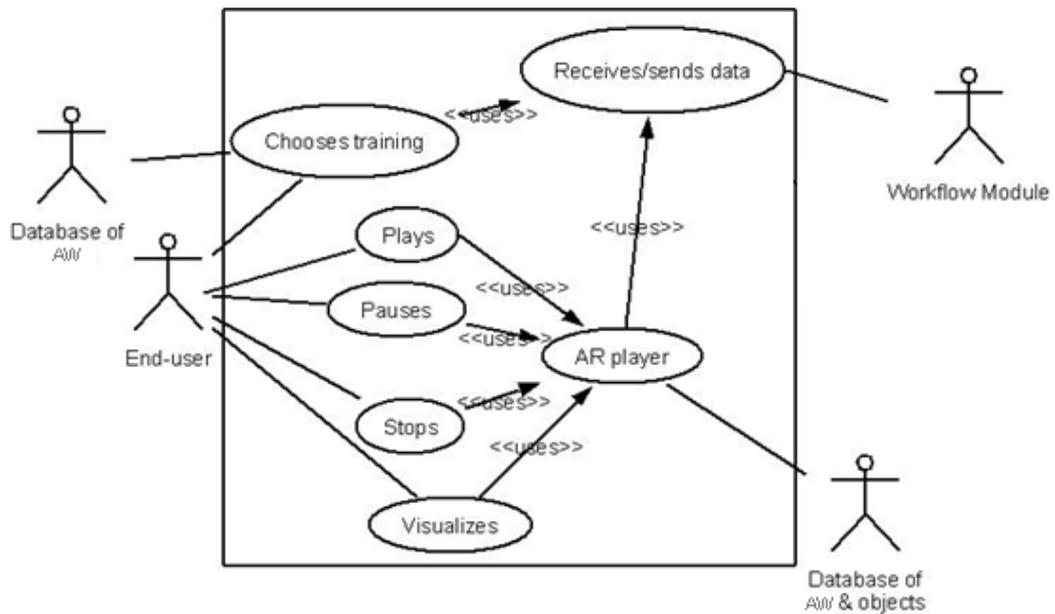


Figure 24. UML Case Diagram for augmented reality player scenario

## 4.5. CLASS DIAGRAM

This section presents the class diagram of the Augmented Reality Player. It is composed by three classes: ARPlayer, WFMModuleReceiverSender and Activity Workflow.

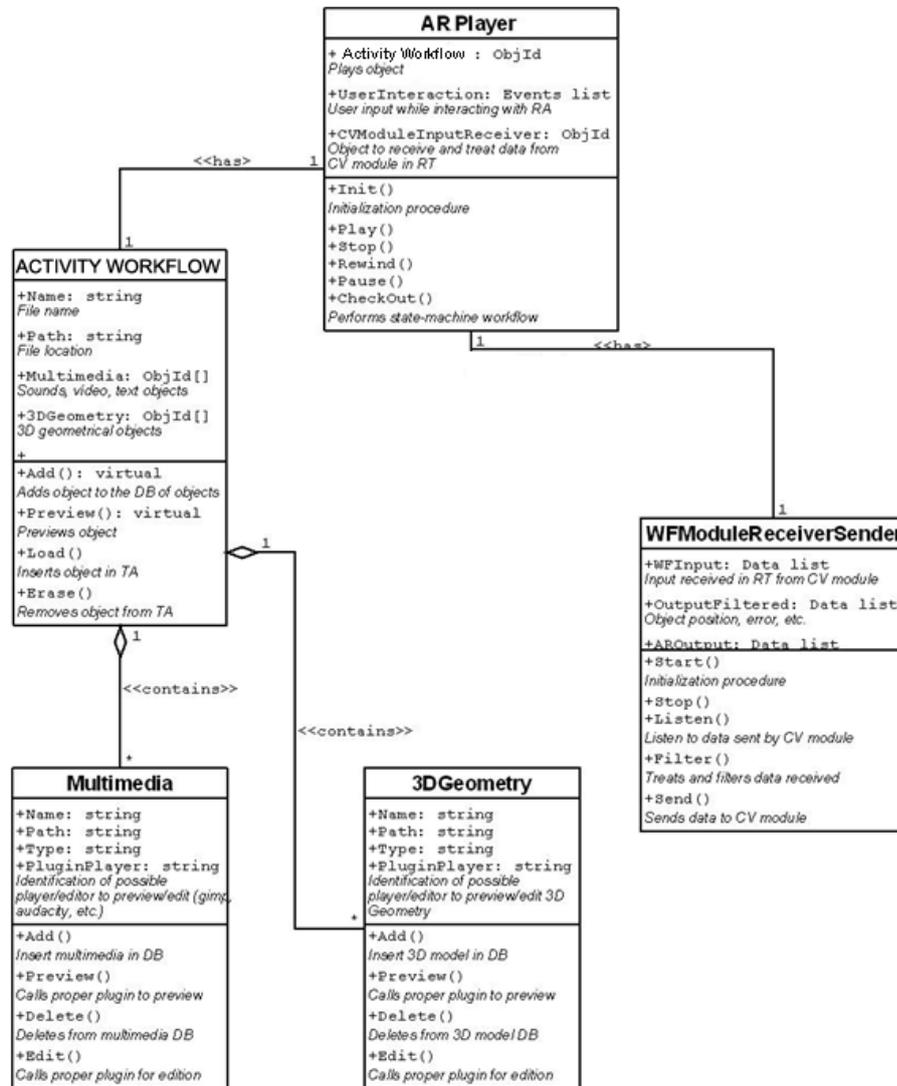


Figure 25. Augmented Reality Player Class Diagram

ARPlayer is the main class that is responsible for the end-user interface and the creation and management of two basic objects: Activity Workflow and WFMModuleReceiverSender.

It is associated to these other three classes in a proportion 1:1 (there can only be one instance of each object). It includes several methods that are used to control these objects after their instantiation. It also includes a method dedicated to check if

the pre-recorded AW is being properly executed. If not, the user should be warned and the piece of AR training repeated. This checking is performed through a state-machine workflow approach.

**WFModuleReceiverSender:** This class instantiates the object that manages the receiving and sending of real-time data between the AR player module and the Workflow module of COGNITO. The data received from the workflow module complements information to render objects properly on the video layer, indicating the next AW to be loaded. Data may contain errors and come in huge volume, which includes methods that constantly read filters and treats data before putting it in its attribute `OutputFiltered`. This attribute is used by the AR Player class. The data sent by the AR player module may be a command, an event or an error.

**ActivityWorkflow:** This class is responsible for the management of the activity workflow files. Because the AW is composed of multimedia and 3D model objects that are loaded or edited, this class has attribute objects that instantiate them (`Multimedia` and `3DGeometry`) and contains classes that are specialized in dealing with such objects. In the case of the multimedia objects, it should exist one class specialized for each type of multimedia - sounds, video, text, etc. These objects are instantiated according to the content of the loaded AW.

## 4.6. RESULTS

The players were developed in a Linux environment (Ubuntu 9.10). Irrlicht was used as the 3D engine and TinyXml to make the parser of the activity workflow. OpenCV was also used to create the background video to the AR player.

Figure 26 shows the current look of the application. It is possible to see in the top of the window the information of what is the compound event that is playing and also the atomic task to be executed in the second line. On the left side of the screen there is a menu of the elements that are used in the activity. When active the background color is white; when they aren't needed the background color is grey. In the center it is possible to see a representation in virtual reality mode of a recorded action in this case "Hammering Nail".

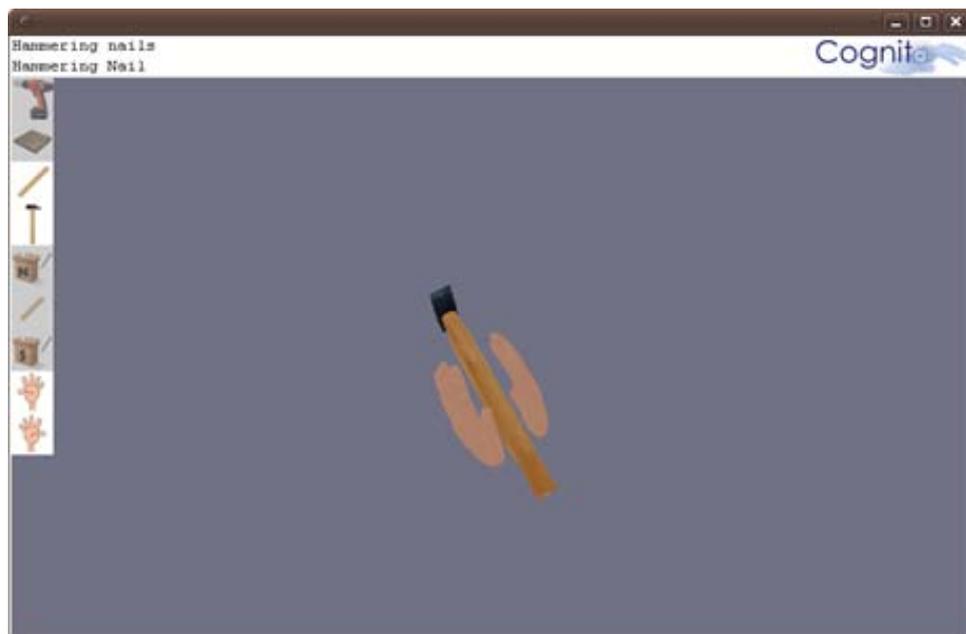


Figure 26. COGNITO VR Visualization

The application also has a camera system that allows viewing the scene in different perspectives; three options are available top, side and back view. This feature is only available in the VR player.

In what concerns the visualization mode there are two different types: movie and step by step. The movie mode is an option that makes the reproduction of the captured activity like a movie and the user has the possibility to play, pause and rewind. The step

by step mode plays the atomic actions one by one accompanying the user step by step and allowing him to set the pace of the activity.



Figure 27. COGNITO AR Visualization

The final purpose of these components is to be a part of an authoring tool that will be able to edit and generate visualizations in Virtual or Augmented Reality environments that work as training manuals for specialized technicians, and as systems that support the execution of individual tasks, providing the technician with information regarding the steps to take and correct them, if needed.

## 5. CONCLUSION

Authoring tools can be seen today as a viable solution to help industry to reduce costs in the creation of technical documentation. Create specific manuals by hand to each version of a particular product is becoming unfeasible. Also is necessary to have applications that allow the reuse of the data generated in the product creation process. With the technology available today there is an urgent need to be able to create a new approach where the user can maintain focus in his work without having to switch between the manual and his task.

This thesis aims to contribute to this new approach, the main goals were to create the action model and the player responsible for creating the visualization, which was achieved with the activity workflow where the data is stored according to the parameters defined, as well as creation of the virtual reality and augmented reality player. The developed components will be integrated in the first version of the COGNITO. However, they are not a final version as the project duration is three years and this is the work of the first ten months.

Future work will focus particularly in the system that supports the execution of individual tasks, providing the technician with information regarding the steps to take and correct them, if needed. A control input based on gesture recognition is also one option under study.

## REFERENCES

- Azuma R., Y. B. (2001). Recent Advances in Augmented Reality. *21* (34-37).
- Dietz, R. (2002). CMIL Specification 0.9.
- Dorner R., C. G. (2002). Authoring Mixed Reality - A Component and Framework-Based Approach. Japan: International Workshop on Entertainment Computing.
- Friedrich W. (2002). ARVIKA - Augmented Reality for Development, Production and Service. Darmstadt: The Second International Symposium on Mixed and Augmented Reality.
- Guven S., F. S. (2003). Authoring 3D Hypermedia for Wearable Augmented and Virtual Reality. New York, U.S.A.: International Symposium on Wearable Computers.
- Haringer M., R. T. (2002). A pragmatic approach to Augmented Reality Authoring. Darmstadt: Proceedings of the International Symposium on Mixed and Augmented Reality.
- ISO. (1997). The Virtual Reality Modeling Language, International Standard ISO/IEC 14772-1. ISO.
- Knoepfle, C. . (2005). Template based Authoring for AR based Service Scenarios. Germany: VR 2005.
- Stock I., W. M. (2006). Authoring Technical Documentation Using a Generic Document Model. *SIGDOC : ACM Special Interest Group on Systems Documentation* (pp. 172-179). USA: ACM.
- Stock I., W. M. (2005). Metadata based Authoring for technical documentation. United Kingdom: SIGDOC : ACM Special Interest Group on Systems Documentation.
- Tang A., C. O. (2003). Comparative effectiveness of augmented reality in object assembly. USA: SIGCHI conference on Human factors in computing systems.
- Zauner J., M. H. (2003). Authoring of a Mixed Reality Assembly Instructor for Hierarchical Structures. Tokyo: The Second International Symposium on Mixed and Augmented Reality.