Multimedia practice session - proposal (basic details) (all-in max 3 500 words without section 8)

1a. Project title

Development of a Augmented Reality framework through a real-life application.

1b. Project acronym

INCAPS

1c. Applicants

Roberto Valenti , R. V. , rvalenti@science.uva.nl , 0493198
Jurjen Pieters, J. P. , mail@jurjenpieters.nl , 9877789

2a. Summary of research proposal (max 250 words)

In recent years, more applications are being developed to create new human-computer interfaces based on augmented reality(i.e.[8] [9] [15]). This is a model of the world that combines different sensors to enhance human-computer interaction or to display data in this world. Examples of this are vision aided hardware used in military helicopters, molecule manipulators in chemistry laboratories and camera-based devices that help surgeons in operating rooms ([34]). The number of this types of systems will grow enormously in the future both in numbers and in complexity. Most of the current systems are using the same types of hardware, but how these devices communicate with each other is fairly uncorrelated. We want to study the aspects of the communication and processing complexity of these systems and combine this into a general framework that will help the scientific community accelerate their projects and make it more easy to merge between different AR systems.

Our goal is to develop this framework with the use of a real-life application, which will consist of a automotive driving assistant, helping the driver recognise objects and situations during driving that communicates with the driver through different interfaces. In the developing process of the system we construct and explore the different aspects of designs and protocols which eventually will lead to the framework.

Because our work will touch all main aspects of computer vision, we hope that researchers, software architects and AR engineers can focus their effort on solving more pending and interesting problems.

2b. Abstract for laymen (max 250 words)

We will propose a framework that will allow computer vision scientists and other data fusion engineers to implement their algorithms easier and faster into AR systems. The framework will inhibit a model of the desired environment of study and all the communication to and from sensors in this world. The data from these sensors and can vary in type and range but the way this is retrieved and altered in the world model will be standardized.

The main idea behind this that it is easier and faster for different scientists working on the same world model to manipulate data streaming from a database in a known format than it is from reading signals from an (unknown) hardware device. This assumes that a former part of the computational complexity will be performed by the input hardware and this will eventually decrease the algorithmic complexity used by the different applications using the data.

When we incorporate the framework into the real life application, we will be able to adopt new technologies and systems very fast, because the only requirement is that we have to rewrite the communication and modeling layers between this systems and our model.

Our framework can be used in virtually every field that requires human-computer interaction through data fusion. The possibilities are unlimited.

3. Classification

Video/image analysis.

Research proposal

4. Description of the proposed research (max 2 000 words)

a. Research topic and envisaged results

The majority of current computer aided vision and data sensor fusion systems are correlated to each other only on the hardware aspect, and we believe that we can easily raise this correlation to a higher level. By this we mean that a lot of data that is retrieved from the environment will be lost in the model because it is not used. We want to use this fact in our framework, by giving the programmer tools to define in which area his algorithm must be applied, by defining rules and setting parameters depending on the situation at hand. In our driver assistance application for example, we want to recognize pedestrians and surrounding traffic. When using our framework it will be possible to restrict the search space to the street and to objects that are less than X meters away from the car. In this case we can reduce the computation time of the algorithm drastically. We believe that this can improve the most well known algorithms ([25]).

The testbed for our framework will be a driving assistance system for an automotive. This system will consists of a navigation system combined with collision, pedestrian and traffic signal detection supported by real time traffic information from an on-line database. For this we want to use an infrared camera for night vision, multiple cameras for instant close range traffic information, a GPS system for positioning, a short distance radar for collision detection and an wireless internet connection that communicates with different servers for weather and traffic information. We will have to analyze the input from all those devices, retrieving as much information as possible from the external environment, matching the information with internal and external databases and filtering all this information to match the desired model, which in out case will be a traffic oriented world. We want to be able to recognize all other users of the environment, such as cars and pedestrians, plus road signs, road features and directional features. With this model we then want to use this to supply the driver with this information, in the best possible way. For example we want to research the application of highlighting traffic signs and road features in the windshield. This can also be use full during the night and in low illuminated situations. But for example when a dangerous situation is ahead, one can think of sound signals to alert the driver of this imminent danger or lowering the volume of the radio to get the drivers attention.

If we wanted to implement the same application with common algorithms, we could face other problems such as noise and useless information. If, for example, there is an car trader on the side of the street, it is possible that it will be recognized as a traffic jam by all the current algorithms, and if there is a billboard on a curve displaying a picture of a person it can be easily misinterpreted as a close pedestrian. In our framework we will be able to filter out these effects because our model will predict that the information gathered from these parts from the sensor space will have a low probability of influence for the model. The model calculates how relevant information from the sensed input are, and can interact and steer its input from the sensors accordingly. When the assistant predicts an dangerous event ahead according to its current world model, it will provide resources for sensors (like computing cycles and data streams) needed to calculate the correctness of its prediction. This is an awareness model not yet used by current Augmented Reality Frameworks.

The framework major requirement will be that it is flexible and that it can grow, to maximize the change that manufacturers of sensors, components and vision systems will adopt the requirements and communication protocols needed for this framework. The more the framework will be adopted the more use full it will be for all parties, even when they are competing with each other.
But next to this framework, we believe that the knowledge gained and the effort invested in the driver assistance application will pay off the funding for this research.

b. Approach

The main framework idea is to simulate what humans are doing naturally. It has been proven ([35]) that the eyes are analyzing only the most interesting parts of human field of sight. We humans are only analyzing the interesting part of the scene, to catch the differences and react as soon as possible. The rest of the scene is still taken in consideration, but the focused area has the priority over the rest and is the center of attention/reaction.
In the case of our particular driver assistant application, the model keep track of the environment

and calculates its interests according to the current traffic situation. This we will do by introducing a "gazing" component into the framework that steers the computational resources according to its priority model. With this gazing component leading the data flow between the sensors and the database we ensure that the complexity of the models environment will be always limited by the maximum resources.
Furthermore, we need to have a dynamically world model of the environment. This model interacts with the environment via the protocols that are defined in the framework. When we connect a new device, we want that the model can updates itself automatically with the data from the new device, but only when the model is asking for this data. When an application connects to this framework, it thus only needs to ask the model for the types of available data of the world model, and uses its internal libraries to interact with this information.

The most complex part of the framework is definitely the information extraction. We will solve the first layer of this problem by using one of the existing algorithms of stereo vision that will help us to extract a simple 3d model of the world. The information extracted is going to be extended or validate through the data retrieved from all the other peripherals. Once the real world is sampled and rebuilt in a virtual representation we are going compute all the information and map them to 3d bounding boxes. The user will then select which boxes are relevant from the given list, and can apply further algorithms to the position of the boxes. The algorithms used to select those boxes may vary between the different applications, (for example we can use a color map to recognize the road color during the night, or we match objects through haar-like features and adaboost). Existing algorithms of based on color constancy and 3d object detection and will help us in this task as well. To speed up the basic retrieval, we will give to the user or programmer the chance to choose which of the algorithms wants to use, basing the decision on which kind of application he is going to build/use. To be consistent, all the algorithms that we are going to include in our framework to extract information from the internal world are required to be simple and to operate in real-time, and may vary from simple object matching to dynamic tracking and prediction in the 3d space (for example to avoid collisions). The combination of different algorithms to extract data in real-time may drop on the performance of few of those algorithms. Anyway, we believe that in the years that this framework will be commercialized, it will be possible to embed or integrate it in a single and portable system.

c. Scientific or economic relevance

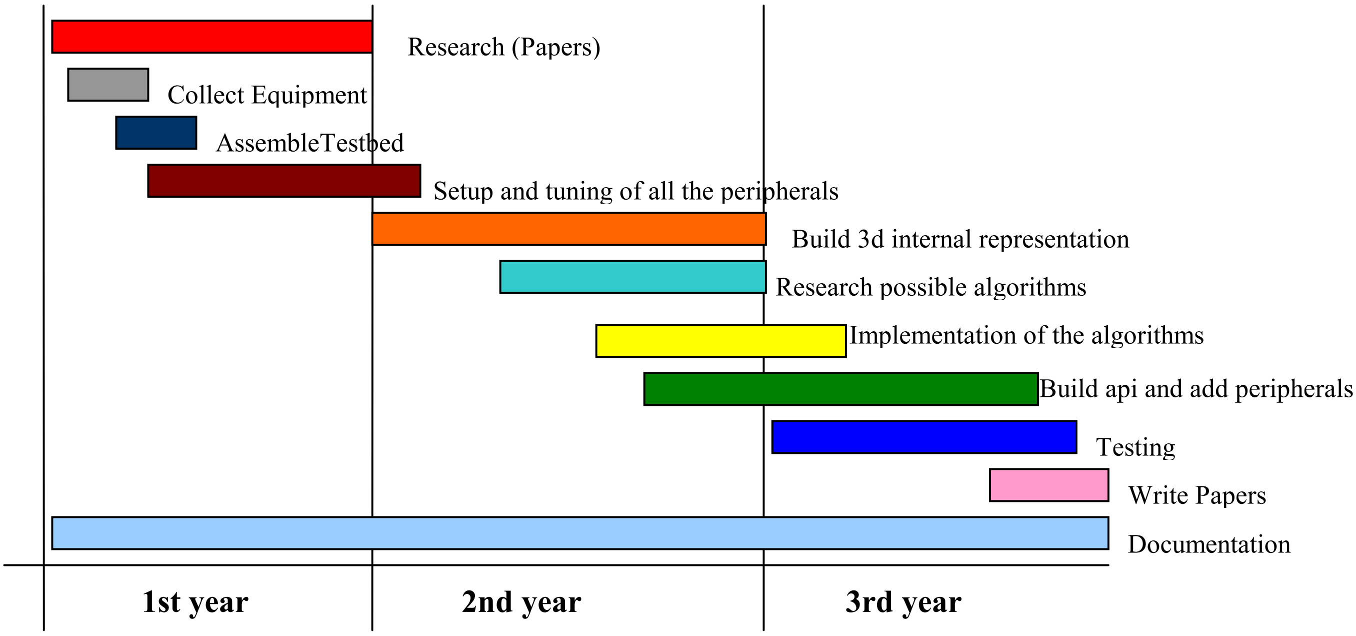
The proposed framework will give a background to almost every kind of video/image analysis subfields, allowing many computer vision scientists to apply their algorithms in an interactive way with the real world, and keeping them far from underlying problems such as data fusion.
By introducing the gazing component and allowing this to steer the availability of the resources and the size of the world model, we will be combining many aspects of computer science but most of them are all related to computer vision.
In the case that we will fail fully completing this framework, we will still be able to deliver of course an extraction of our research, which is the basis of our framework and of the world model. And next to that, we will have developed an innovative and hopefully highly efficient driving assistance navigation system that can be sold to many car manufacturers.

d. International developments

In the Augmented Reality field there has been some research in this direction ([36,38,39,41]). In the last years there seems to be growing activity related to AR. Frameworks have been a major concern, giving justification for the direction we want to take with this research. However, most of the current models lack the flexibility that we strive to. A lot of effort is put into the applications that deal with wearable computers ([39,43]). The model that matches our research the most is DWARF ([36]). This is a modular based software framework that enables users to add new components into their AR framework. But they lack the existence of a world model like we want to introduce, and that we need so much to reduce the amount of data absorbed by the system.

5. Work programme

Our work program can be represented with by the following graph :



During the course of the project, we will write project documentation describing what we did in each time step. This documentation will be useful when we are going to write the final documentation without loosing implementation details or missing any idea that we had during the first year. The documentation will consist of requirements documents, software design documents and software verification documents, enabling us to extract the framework from this basis and allowing third parties to test and comment on this framework, and therefore updating us with their demands. This will also allow us to respect the deliveries and to keep track of all the experiments and trouble that we will encounter during the development phase.

In the first year, our team will study all the most relevant papers on the field, focusing on the most relevant research that will help us for our purposes, and using the knowledge already acquired reading the background papers ([1..24]). In the mean time we will collect the equipment required to implement our night vision application testbed. As soon as we get enough equipment, we will assemble them on the experimental car and we will start to setup and fine tuning them in order to obtain homogeneous data.(i.e. Calibrating the instruments). This tuning and testing will conclude in the beginning of the second year, with a good and consistent "world sampling" from all of the used peripherals.

During the second year we will mainly focus on building an internal 3d world representation from all the sampled information, and we will research and develop algorithms that will be used to select or to extract from the internal world representation only the important information required for our specific application.

In the third year we expect our driving assistance application to be ready to use. We will test what we have done and we will create API's that will allow third parties to use and interface with our framework with a predefined subsets of augmented reality applications (we will choose them from the one studied the year before). The testing phase will include comparison with existing systems to evaluate the performances of our framework. We will work on the API's and test the overall framework all the time we implement a new feature, adding more and more applications and hypothetical peripherals until few months before the end of our third year. We will use the remaining time to extend and complete the documentation concerning the framework.

Equipment

6. Expected use of instrumentation

To implement the nightvision testbed for our frameworks, we are planning to buy a cheap test car that we will use to embed the hardware. We will assemble all the equipment on this car. The main processing unit should be a powerful computer, which we expect will be cheaper when the system will be complete. We want to equip the car with at least a stereo camera vision system, a close distance radar, an infrared camera (for the night vision) and we might need a database system and a wireless communication on board. Furthermore, we need few output peripheral that could vary between a classic LCD screen (HUD) or a beamer to VR glasses or a Translucent screen. The additional posts are covering trips to Augmented Reality conferences, which we will request for all the three years to be sure to have the latest advances in the field.

Cost estimates

7. Requested Budget:	**Amount (euro)**
For 3 years, your own gross salary + 70% overhead	306.000
Bench-fee	4 538
Specify any additional posts	10.000
Sub-total	**320538**
Equipment	30.000
Total requested funding	**350.538**

Total max. 400 000 Euros.

References

8. Literature

[1] Y. Baillot, D. Brown, and S. Julier. Authoring of Physical Models Using Mobile Computers. *Fifth International Symposium on Wearable Computers*, pages 39–46, October 7-10,2001.

[2] J.F. Bartlett. Rock 'n' Scroll is Here to Stay. *IEEE Computer Graphics and Applications*, 20(3):40–45, May/June 2000.

[3] U.S. Census Bureau. Topologically Integrated Geographic Encoding and Referencing System, http://www.census.gov/geo/www/tiger. 2002.

[4] A. Cheyer, L. Julia, and J. Martin. A Unified Framework for Constructing Multimodal Applications. *Conference on Cooperative Multimodal Communication (CMC98)*, pages 63–69,January 1998.

[5] W.J. Clinton. Statement by the President Regarding the United States' Decision to Stop Degrading Global Positioning System Accuracy. *Office the the Press Secretary, The White House*, May 1, 2000.

[6] P.R. Cohen, M. Johnston, D. McGee, S. Oviatt, J. Pittman,I. Smith, L. Chen, and J. Clow. Quickset: Multimodal Interaction for Distributed Applications. *ACM International Multimedia Conference*, pages 31–40, 1997.

[7] R.T. Collins, A.R. Hanson, and E.M. Riseman. Site Model Acquisition under the UMass RADIUS Project. *Proceedings of the ARPA Image Understanding Workshop*, pages 351–358,1994.

[8] D. Davis, T.Y. Jiang, W. Ribarsky, and N. Faust. Intent, Perception, and Out-of-Core Visualization Applied to Terrain. *IEEE Visualization*, pages 455–458, October 1998.

[9] D. Davis, W. Ribarsky, T.Y. Jiang, N. Faust, and Sean Ho. Real-Time Visualization of Scalably Large Collections of Heterogeneous Objects. *IEEE Visualization*, pages 437–440,October 1999.

[10] N. Faust, W. Ribarsky, T.Y. Jiang, and T. Wasilewski. Real-Time Global Data Model for the Digital Earth. *IEEE Visualization*,March 2000.

[11] T.L. Haithcoat, W. Song, and J.D. Hipple. Building Footprint Extraction and 3-D Reconstruction from LIDAR.*Data Remote Sensing and Data Fusion over Urban Areas,IEEE/ISPRS Joint Workshop*, pages 74–78, 2001.

[12] K. Hinckley, J.S. Pierce, M. Sinclair, and E. Horvitz. Sensing Techniques for Mobile Interaction. *ACM User*

Interface Software and Technology, pages 91–100, November 5-8, 2000.

[13] J.M. Kahn, R.H. Katz, and K.S.J. Pister. Mobile Networking for Smart Dust. *ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom 99)*, pages 116–122, August 17-19, 1999.

[14] D.M. Krum, R. Melby, W. Ribarsky, and L.F. Hodges.IsometricPointer Interfaces for Wearable 3D Visualization. *Submission to ACM CHI Conference on Human Factors in Computing System*, April 5-10, 2003.

[15] D.M. Krum, O. Omoteso, W. Ribarsky, T. Starner, and L.F.Hodges. Speech and Gesture Control of a Whole Earth 3D Visualization Environment. *Joint Eurographics - IEEE TCVG Symposium on Visualization*, May 27-29, 2002.

[16] D.M. Krum, O. Omoteso, W. Ribarsky, T. Starner, and L.F.Hodges. Evaluation of a Multimodal Interface for 3D Terrain Visualization. *IEEE Visualization*, October 27-November 1, 2002.

[17] D.M. Krum, W. Ribarsky, C.D. Shaw, L.F. Hodges, andN. Faust. Situational Visualization. *ACM Symposium on Virtual Reality Software and Technology*, pages 143–150, November 15-17, 2001.

[18] K. Lyons and T. Starner. Mobile Capture for Wearable Computer Usability Testing. *Fifth International Symposium on Wearable Computers*.

[19] S.L. Oviatt. Mutual Disambiguation of Recognition Errors in a Multimodal Architecture. *Proceedings of the Conference on Human Factors in Computing Systems (CHI'99)*, pages 576–583, May 15-20, 1999.

[20] W. Piekarski and B.H. Thomas. Tinmith-Metro: New Outdoor Techniques for Creating City Models with an Augmented Reality Wearable Computer. *Fifth International Symposium on Wearable Computers*, pages 31–38, October 7-10,2001.

[21] J.C. Spohrer. Information in Places. *IBM Systems Journal*, 38(4):602–628, 1999.

[22] J. Vandekerckhove, D. Frere, T. Moons, and L. Van Gool.Semi-Automatic Modelling of Urban Buildings from High Resolution Aerial Imagery. *Computer Graphics International Proceedings*, pages 588–596, 1998.

[23] T. Wasilewski, N. Faust, and W. Ribarsky. Semi-Automated and Interactive Construction of 3D Urban Terrains. *Proceedings of the SPIE Aerospace, Defense Sensing, Simulation and Controls Symposium*, 3694A, 1999.

[24] S. You, U. Neumann, and R. Azuma. Orientation Tracking for Outdoor Augmented Reality Registration. *IEEE Computer Graphics and Applications*, 19(6):36–42, Nov/Dec 1999.

[25] P. Viola and M. Jones, ÒFast and Robust Classification using Asymmetric Adaboost and a Detector CascadeÓ, *Advances in Neural Information Processing System* 14, MIT Press, Cambridge, MA, 2002.

[26] M. Ashdown and R. Sukthankar. Robust calibration of camera-projector system for multi-planar displays. Technical Report HPL-2003-24, HP Labs, 2003. January.

[27] A. Cockburn and B. McKenzie. Evaluating the effectiveness of spatial memory in 2D and 3D physical virtual environments. In *Proceedings of CHI*, 2002.

[28] C. Cruz-Neira, D. Sandlin, and T. DeFanti. Surround-screen projection-based virtual reality: The design and implementation of the CAVE. In *Proceedings of SIGGRAPH*, 1993.

[29] *M. Gross et al. Blue-C: A spatially immersive display and 3D video portal for telepresence. In Proceedings of SIGGRAPH, 2003.*

[30] C. Pinhanez. *The Everywhere display. In Proceedings of Ubiquitous Computing, 2001.*

[31] Fraser, Q.S., Robinson, P.: BrightBoard: *A Video-Augmented Environment. Proceedings of CHI'96. Vancouver (1996) 134-141*

[32] Leubner, C., Brockmann, C., Müller, H.: Computer-vision-based Human Computer Interaction with a Back Projection Wall Using Arm Gestures. 27th Euromicro Conference.(2001)

[33] Ashdown, Flagg ,Sukthankar , Rehg:A Flexible Projector-Camera System for Multi-Planar Displays

[34]Frank Rudolph : Cockpit for the Neurosurgeon http://www.zeiss.com/

[35] David Anthony Leopold :Brain Mechanisms of visual awareness: Using Perceptual Ambiguity to Investigate the Neural Basis of Image Segmentation and Grouping (1997)

[36] Martin Bauer, Bernd Bruegge, Gudrun Klinker, Asa MacWilliams,Thomas Reicher, Stefan Riß, Christian Sandor, Martin Wagner :*Design of a Component–Based Augmented Reality Framework* TU Munchen, Institut fur Informatik (2001)

[37] Selim Balcisoy, Marcelo Kallmann, Pascal Fua, Daniel Thalmann :*A framework for rapid evaluation of prototypes with Augmented Reality (2000)*

[38]Wayne Piekarski and Bruce H. Thomas: Developing Interactive Augmented Reality Modelling Applications

[39] Stuart Goose, Heiko Wanning, Georg Schneider: Mobile Reality: A PDA-Based Multimodal Framework Synchronizing a Hybrid Tracking Solution with 3D Graphics and Location-Sensitive Speech Interaction (2002)

[40]Yohan Baillot, Simon J. Julier, Dennis Brown, Mark A. Livingston: A Tracker Alignment Framework for Augmented Reality (2004)

[41] Andrej van der Zee :Multimedia Framework for Augmented Reality Applications in Ubiquitous Environments (2003)

[42] Blair MacIntyre, Maribeth Gandy, Steven Dow, and Jay David Bolter. "DART: A Toolkit for Rapid Design

Exploration of Augmented Reality Experiences." To appear at *conference on User Interface Software and Technology (UIST'04)*, October 24-27, 2004, Sante Fe, New Mexico

[43] Jie Yang, Weiyi Yang, Matthias Denecke, Alex Waibel :Smart Sight: A Tourist Assistant System