

numediart

Research Program in Digital Art Technologies - Newsletter 5 Novo8

Period 5 - Jan-Mar'09

Call for Participation

numediart is a long-term research programme centered on Digital Media Arts, funded by Région Wallonne, Belgium (grant N°716631). Its main goal is to foster the development of new media technologies through digital performances and installations, in connection with local companies and artists.

It is organized around three major R&D themes (HyFORGE - hypermedia navigation, COMEDIA - body and media, COPI - digital luthery) and is performed as a series of short (3-months) projects, typically 3 or 4 of them in parallel, during which a 1-week "hands on" workshop is held.

numediart is the result of collaboration between Polytech.Mons (Information Technology R&D Pole) and UCL (TELE Lab), with a center of gravity in Mons, the cultural capital of Wallonia. It also benefits from the expertise of the Multitel research center on multimedia and telecommunications. As such, it is the R&D component of Mons 2015, a broader effort towards making Mons the cultural capital of Europe in 2015.

The numediart board now calls for participation to three short projects defined in the following pages, to be held (remotely) from Jan, 5th to Mar, 31st 2009. The project will involve a one-week workshop around mid-term in Louvain-la-Neuve. Results will be publicly presented at the end of the project. If you want to contribute, please send an email to contact@numediart.org mentioning which project you want to join, and what kind of expertise you could bring, before Dec, 19th. The number of participants is limited to 8 people per project. Participation to the workshop is mandatory. No funding is provided, but no fee is asked for either.



POLYTECH
MONS



UCL
Université
catholique
de Louvain



RÉGION WALLONNE

MATRIX: Natural Interaction Between Real & Virtual Worlds

Idea

The idea of this project is to test the best ways to handle naturalness in the interaction between virtual objects or avatars with real objects or people. This project is the direct continuation of project [Augmented Virtual Studio \(#4.3\)](#) achieved in the previous project session and also uses results from projects [Tracking-Dependent and Interactive Video Projection \(#3.1\)](#) and [Sensor-Based Mini-COMEDIA \(#1.2\)](#). Color and infra red video live capture and processing will be used.

Motivations

While interaction between real and virtual worlds is already possible and some implementations were already done in project [#4.3](#), it is still very challenging to get natural reactions and interaction between these different worlds. This project aims at investigating this natural interaction and providing tools for a better integration of virtual scenes and objects in the real world.

Natural interactions between real users and virtual avatars or agents

This part of the project is a continuation of project [#3.1](#) which led to a real-time attention algorithm implementation for space and time contexts. Virtual avatars that are able to pay attention to surprising behaviors have a real advantage in terms of interaction naturalness. Two enhancement ways will be followed:

- The enhancement of the short-term attention model by using the maximum of trajectory and body descriptors. In addition to the blob speed and the silhouette Quantity of Motion, many other descriptors can be added like blob direction or trajectory directness, etc... Those features can be used to get a realistic actor behavior description and to highlight the most interesting parts of it.
- The enhancement of the instantaneous attention model to avoid individual blob tracking. In that way complex scenarios like crowd motion and camera motion should also be handled.

Once those two approaches of motion attention are enhanced, a long-term attention approach which builds a motion model of a scene will be implemented. In that way, motion attention will handle instantaneous (no memory), short-term and long-term memory. An avatar which is able to pay attention should be able to have more human-like reactions.

Natural motion capture and toon-like motion simulation

In this part of the project we will study the use of our robust inertial motion capture suit IGS 190 for the production of 3D animations. Two different approaches will be considered:

- Using motion capture as a tool for the production of key frames for cartoon animations. As the data obtained with motion capture are too "human-like" to be used as such for the animation of most cartoon characters, we will investigate how we can develop algorithms for the automatic adaptation of motions to cartoon characters, by automatically exaggerating the amplitude of the motions for example.

Natural interactions between real users and virtual objects

- An interesting and difficult issue of augmented reality is the occlusion problem: how to manage the position (in deep or Z axis) of the objects in real video in order for them to more naturally interact with the 3D virtual objects? Several approaches (single and multiple camera-based ones) will be tested in order to select those which seem to be the most interesting depending on the applications.
- Another problem is in the feedback that the user can get from virtual objects. In the case where the virtual objects are linked to real objects that the user holds, this feedback is more natural. But in the case where virtual objects are manipulated without using real objects, it is very interesting to get a physical, haptic feedback (as opposed to a visual feedback only). One may for instance produce vibrations in a tangible object, whose amplitude could depend on the distance between a virtual object and the user's hand.



Figure 1: NeuroTV [6] avatar examples.



Figure 2: IGS 190 motion capture suit.

- Using motion capture data to directly animate in real-time a virtual avatar. The data from the motion capture suit will be sent through OSC on the motion capture side of the chain, and read and interpreted by the 3D rendering software for real time animation of the avatar at the animation rendering side. A comparison of this technique with the one already implemented in project #4.3 (IR markers-based motion capture) will be achieved. This comparison will provide criteria which will help in choosing when it is more interesting to choose the motion capture suit or the IR markers for motion capture.

The developments of these three scenarios will be mainly achieved by using the EyesWeb platform [8, 3] as already introduced in project #3.1. Other software like Processing [7] and Blender [2] will also be used as a continuation of project #4.3.

References

3D embedding in broadcast area is more and more present in modern television [6, 1] and extensions to more interaction especially a natural interaction between real actors and virtual objects are a novel challenge also in digital arts [5].

The EyesWeb software [8, 3], through the European IP Mega project, has brought very efficient algorithms for video-based tracking.

Processing [7] software can lead to interesting visualization effects by using a simplified Java language and numberless already existing visualization libraries. The results can be easily exported as Java applets which is very convenient for online demos.

Blender [2] is a very powerful 3D modeling tool which also enables scripting in order to automatically manipulate 3D virtual objects. The IGS 190 [4] is a wireless real-time motion capture suit which enables very accurate positioning of the whole human body.

Team

This project will be coordinated by Joëlle Tilmanne (FPMs/TCTS) and Matei Mancas (FPMs/TCTS).

Deliverables

- Report on the three tasks of the project.
- Software tools for building the applications.

Agenda

- 05/01/2009: Project start.
- Week 1-5: Study of the Z position extraction from real videos. Implementation of real time capture of coordinates in IGS 190. Enhancement of the short-term memory attention patch.
- Week 6-9: Study of occlusion handling by using the Z position of real objects and people. Real-time link between the IGS 190 coordinates and an avatar. Set-up of a long-term memory attention model.
- Week 10-11: Implementation of natural augmented reality set ups and study of physical feedback from virtual objects on the real actor. Key frames caption and toon-like motion synthesis by using the IGS 190 device. Enhancement of the instantaneous attention model.
- Week 12: Reporting and packaging deliverables.
- 31/03/2009: Project end and public presentation.

Scientific reference books

- [3] A. Camurri and G. Volpe, eds. *Gesture-based Communication in Human-Computer Interaction*. Vol. 2915. LNAI. Springer Verlag, 2004. P.: 3.

Software and technologies

- [2] *Blender software*. URL: <http://www.blender.org>. P.: 3.
- [4] *IGS 190 capture suit*. URL: <http://www.animazoo.com/IGS190.aspx>. P.: 3.
- [7] *Processing software*. URL: <http://www.processing.org>. P.: 3.
- [8] *The EyesWeb platform*. URL: <http://www.eyesweb.org>. P.: 3.

Artistic and scientific collaborators

- [1] *BlaBla*. URL: <http://www.rtbef.be>. P.: 3.
- [5] *La compagnie des hommes penchés*. URL: <http://www.leshommespenches.com>. P.: 3.
- [6] *NeuroTV*. URL: <http://www.neurotv.be>. Pp.: 2, 3.

Métabrâne, an individualized installation mirroring the subconscious mind

Idea

Métabrâne is a technical reconstruction of a symbolical process that we all know well but often repress: free association. A network of neural activities is being simulated and processes a certain amount of new events, past episodes and ancient memories that are presented to the “spect-actor” within an immersive environment. Its purpose: transposing self-disclosure and creating new meanings inside an interactive object.

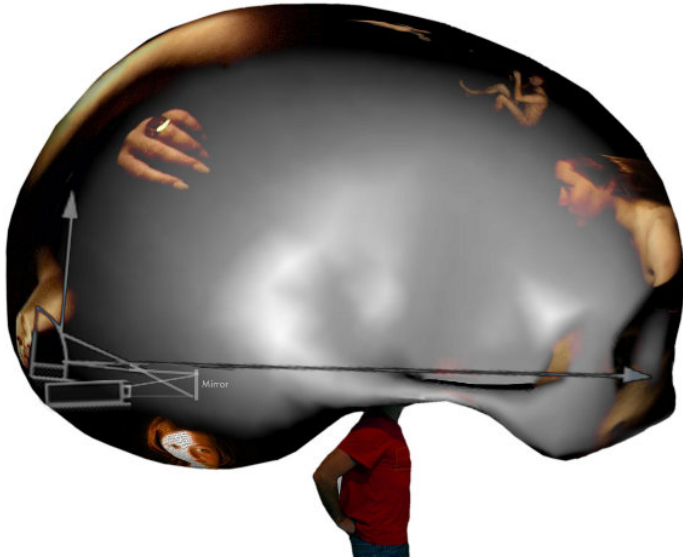


Figure 3: Side-view mockup of the installation being visited by a *spect-actor*.



Figure 4: A mashup of visuals extracted from Thomas Israël's installation *Caresse moi!* (2006).

When the *spect-actor* enters the *Métabrâne*, as depicted in figure 3, he gradually perceives the inherent and emergent behavior of the installation, while the flow of audiovisual fragments which are sequenced around him tends to take into consideration the color of his clothes, his body rhythm revealing his serenity or discomfort, and so on..., but without being granted “1:1” control. The initial content of the multimedia database presented to the *spect-actor* is composed of material extracted from Thomas Israël's previous productions, sequenced in real-time and processed with spatio-temporal effects, as in figure 4; and gets fed back from records of the *spect-actor*'s gestures and voice.

Motivation

In this project will be combined and merged several ongoing efforts undertaken within the HyFORGE and COMEDIA research axes by reusing results from various projects: adding video support to the audio loop database navigation, segmentation and manipulation techniques studied within [Audio Cycle \(# 4.1\)](#), [Audio Thumbnailing \(# 2.3\)](#) and [Audio Skimming \(# 1.1\)](#); selecting methods for bodily gestures acquisition from [Augmented Virtual Studio \(#4.3\)](#), [Stylistic Human Gait Modeling \(#3.2\)](#), [Tracking-Dependent and Interactive Video Projection \(#3.1\)](#); implementing realtime techniques for facial analysis and voice analysis/synthesis from [Multimodal Feedback from Robots and Agents in a Storytelling Experiment \(#3.4\)](#).

This is a great opportunity to learn new techniques, as many practical questions are left to be solved for this project:

- **3D projection:** How many projectors will be required? Will fisheye or spherical lenses better fit [2]? How to compute the realtime deformation of 2D images for a 3D surface projection? Using Eluminati's [OmniMap API](#) or Territoires Ouverts' [lightTWIST](#)?
- **Realtime spatio-temporal effects:** How about video seam carving [7]? [Audio Skimming](#)? What are the different layers of time and space involved? Visualization and sonification inspired by the localization of brain activity?
- **Affective multimodal interfaces:** What type of multimodal features, from gestures to speech, can be extracted from the *spect-actor* in real-time? What about modalities such as facial expression, gestural pose, eye gaze, finger pointing [4]? How can emotions be interpreted from these features?
- **Sound spatialization:** How many loudspeakers will suffice? How to generate a real-time multichannel signal? Using [audioTWIST](#)?

- **Audiovisual similarity computation:** What kind of multimedia features are relevant? Will semantic linking using tags prove itself to be useful? How can perceptual illusions be induced by such sequences?

Deliverables

Besides assisting Thomas Israël in building the hardware prototype of the installation, two major outcomes are expected, as reminded in figure 5:

1. adapting an existing tool for semi-automatic annotation of multimedia data (among others: the IRMA Browser, Advene, Anvil, Lignes de Temps, etc...) so as to prepare the initial content of the database;
2. extending a visual programming environment (such as the OpenInterface platform, EyesWeb, TroikaTronix Isadora, etc...) so as to aid the rapid prototyping of the multimodal installation, by adding the missing and required blocks built upon opensource libraries.

In order to disseminate the results and share our knowledge, we will also provide:

1. a report describing the methods studied and required to produce the aforementioned deliverables;
2. a video showcasing the first prototype of the installation, possibly featuring an example of a first-person immersion recreated using binaural spatialization.

Team

This project has been commissioned by Thomas Israël and will be coordinated by Christian Frisson (UCL-TELE).

We are looking forward to gathering students and researchers interested in the following research fields:

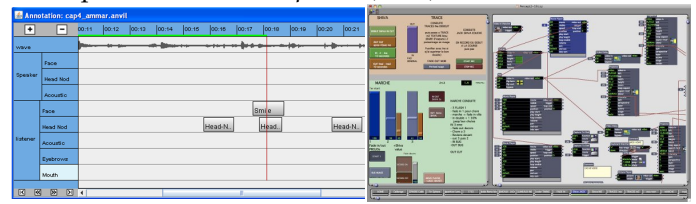
- machine learning for multimedia data mining and information retrieval using ontological databases [9],
- human-computer interaction [4] and affective multimodal interfaces [6],
- immersive multimedia installations [3, 10],
- computer graphics and computer vision [8, 5],
- sound processing and spatialization [1]...

We would enjoy collaborating with members of the IRMA (Interface de Recherche Multimodale dans le contenu Audiovisuel) and 3D-MEDIA projects.

Agenda

- 05/01/2009: Project start.
- Week 1-7: Building the prototype modules:
 - Hardware installation,
 - Semi-automatic annotation tool,
 - Enhanced visual programming environment.
- Week 8: Workshop week (Feb 23-27) - Assembly of the first standalone prototype linking the modules.
- Week 9-10: Preparation of the multimedia database.
- Week 11: Field tests with *spect-actors*.
- Week 12: Reporting and packaging deliverables.
- 31/03/2009: Project end and public presentation.

Figure 5: Screenshots of the Anvil annotator as used at eNTERFACE'08 in project #3.4 and of a TroikaTronix Isadora patch of Thomas Israël & Jacques Urbanska's interactive video performance *Percept 2.0* (2008).



References

Video demonstrations and visuals from Thomas Israël's previous works (including installations, performances, sculptures...) can be accessed on his website: <http://www.thomasisrael.be>.

Current techniques

- [2] Paul Bourke. "Digital Fulldome, Techniques and Technologies". In: *Course prepared for Graphite (ACM SIGGRAPH)*. 2007. P.: 4.
- [7] Michael Rubinstein, Ariel Shamir, and Shai Avidan. "Improved Seam Carving for Video Retargeting". In: *ACM Trans. on Graphics*. Vol. 27. 3. 2008. P.: 4.

Scientific reference books

- [1] Barry Blesser and Linda-Ruth Salter. *Spaces Speak, Are You Listening? - Experiencing Aural Architecture*. The MIT Press, 2007. ISBN: 978-0-262-02605-5. P.: 5.
- [3] Oliver Grau. *Virtual Art: From Illusion to Immersion*. Leonardo. The MIT Press, 2003. ISBN: 0-262-07241-6. P.: 5.
- [4] Philip Kortum, ed. *HCI Beyond the GUI: Design for Haptic, Speech, Olfactory, and Other Nontraditional Interfaces*. Morgan Kaufmann, 2008. ISBN: 978-0-12-374017-5. Pp.: 4, 5.
- [5] Gerard Medioni and Sing Bing Kang, eds. *Emerging Topics in Computer Vision*. IMSC Press Multimedia Series. Prentice Hall PTR, 2004. ISBN: 9780131013667. P.: 5.
- [6] Christian Peter and Russell Beale, eds. *Affect and Emotion in Human-Computer Interaction: From Theory to Applications*. 1st ed. Lecture Notes in Computer Science. Springer, 2008. ISBN: 9783540850984. P.: 5.
- [8] Linda G. Shapiro and George C. Stockman. *Computer Vision*. Prentice Hall, 2001. ISBN: 9780130307965. P.: 5.
- [9] Bhavani Thuraisingham. *Managing and Mining Multimedia Databases*. CRC Press, 2001. ISBN: 0-8493-0037-1. P.: 5.
- [10] Stephen Wilson. *Information Arts: Intersections of Art, Science, and Technology*. Leonardo. The MIT Press, 2002. ISBN: 0-262-23209-X. P.: 5.

Media Cycle: Browsing and Performing with Sound and Image Libraries

Idea

The aim of this project is to enhance in two ways the content-based navigation and performance tool for sound and music loop libraries that has been developed in the numediart [Audio Cycle](#) (# 4.1) project. First, it will be extended to image databases. Second, the databases of sounds/images will be organized in a mathematical graph, so that graph-theoretic approaches can be used during navigation. For example, the user will be able to connect sounds/images to a set of others, and use this information to assist in the creation of sequences of music or images with smooth transitions.

These navigation tools will have a wide range of potential applications, from browsing a photo collection to media art installation and live music or VJ (Video Jockey) performances. The graph framework can also lead to other visual representations of the library, as in the “Graph Theory” software [4], making use of libraries such as [igraph](#) [2].

In the case of sound and music, “content” was referring to rhythm, harmony, melody, timbre, . . . , whereas in the case of images it will refer to attributes such as color, shapes, texture, Algorithms for automatically extracting such attributes will be reused or developed. Possible extensions include navigation in video databases, where “content” can additionally be characterized by camera motion parameters (e.g., zoom, pan), object motion [3], and other dynamic attributes.

Finally, one well-known problem (called the “semantic/numeric gap”) resides in the possible discrepancy between the low-level attributes (color, shape, texture) and the meaning of the image (from the user’s point of view). To compensate this gap, machine learning techniques have to be incorporated to take into account the user’s preferences.

Motivations

Database search initially relied on metadata associated to each sound/image (such as captions or keywords) which may be difficult or tedious to generate, especially when the database reaches a critical size. Existing software for image [7, 8, 6] or sound [5] allow the user to retrieve media files related to the query. The efficiency of such queries has been shown to be enhanced by formulating a database search as a graph-clustering problem [1]. We propose to use graph-theoretic algorithms to provide new features to the query, such as the possibility of retrieving all media elements that lie between two others, which can be used to create music and visual sequences with smooth transitions. Algorithms such as the traveling salesman can also be used to find related elements of the database.

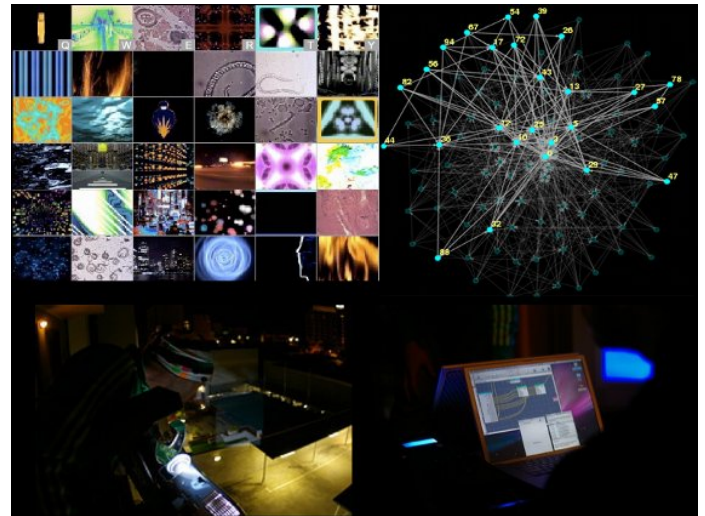


Figure 6: A visual mashup containing screenshots from [igraph](#) [2] and [LividInstruments Cell](#); and photos from [Create Digital Media's Flickr photostream](#).

Agenda

- 05/01/2009: Project start.
- Week 1: Kickoff and coordination.
- Week 2-3:
 - Explore existing techniques for image navigation [7, 8, 6] and VJ equipment/software such as [Resolume](#), [Livid Instruments Union VJ](#), and [Arkaos](#).
 - Investigate various descriptors of image similarities.
- Week 4-7:
 - Build a set of distances in an image database (e.g, the [Columbia University Image Library](#)).
 - Organize the sound/image databases into a graph and test simple graph algorithms.
- Week 8: Workshop week (Feb 23-27)
- Week 9-11: Overall improvements and optimizations, and final integration of the music/images browser prototype.
- Week 12: Reporting and packaging deliverables.
- 31/03/2009: Project end and public presentation.

Team

This project will be coordinated by [Stéphane Dupont](#) (FPMs/TCTS) and [Xavier Siebert](#) (FPMs/MathRO), in collaboration with [Philippe Fortemps](#) (FPMs/MathRO).

Deliverables

- Report on state-of-the-art software and equipment for image navigation and VJs.
- Algorithms for extracting image descriptors, such as shape, color, texture.
- Incorporation of a database (sound/images) into a graph.
- Music/Images browser prototype.

References

- [1] Selim Aksoy and Robert M. Haralick. “A graph-theoretic approach to image database retrieval”. In: *in Visual Information and Information Systems*. 1999. Pp. 341–348. P.: 6.
- [2] Gabor Csardi and Tamas Nepusz. “The igraph Software Package for Complex Network Research”. In: *InterJournal Complex Systems* (2006). P. 1695. URL: <http://igraph.sf.net>. P.: 6.
- [3] Myron Flickner et al. “Query by Image and Video Content: The QBIC System”. In: *Computer* 28.9 (1995). Pp. 23–32. ISSN: 0018-9162. DOI: <http://doi.ieeecomputersociety.org/10.1109/2.410146>. P.: 6.
- [4] Jason Freeman. “Graph theory: interfacing audiences into the compositional process”. In: *NIME '07: Proceedings of the 7th international conference on New interfaces for musical expression*. New York, New York: ACM, 2007. Pp. 260–263. DOI: <http://doi.acm.org/10.1145/1279740.1279794>. P.: 6.
- [5] S. Heise, M. Hlatky, and J. A. Loviscach. “Sound-Torch: Quick Browsing in Large Audio Collections”. In: *AES 125th Convention*. 2008. P. 7544. P.: 6.
- [6] Jorma Laaksonen et al. “PicSOM—content-based image retrieval with self-organizing maps”. In: *Pattern Recogn. Lett.* 21.13-14 (2000). Pp. 1199–1207. ISSN: 0167-8655. DOI: [http://dx.doi.org/10.1016/S0167-8655\(00\)00082-9](http://dx.doi.org/10.1016/S0167-8655(00)00082-9). P.: 6.
- [7] Stan Sclaroff, Leonid Taycher, and Marco La Cascia. “ImageRover: A content-based image browser for the world wide web”. In: *In Proc. IEEE Workshop on Content-based Access of Image and Video Libraries*. 1997. Pp. 2–9. P.: 6.
- [8] John R. Smith and Shih F. Chang. “VisualSEEK: a fully automated content-based image query system”. In: *Proceedings of the fourth ACM international conference on Multimedia*. ACM Press, 1996. Pp. 87–98. DOI: <http://doi.acm.org/10.1145/244130.244151>. P.: 6.

numediart Research Program

<http://www.numediart.org>

Editors

TCTS Lab

Faculté Polytechnique de Mons

<http://www.tcts.fpms.ac.be>

TELE Lab

Université Catholique de Louvain

<http://www.tele.ucl.ac.be>

Contact

contact@numediart.org

Credits

Cover photo © Loïc Reboursière

L^AT_EX design © Christian Frisson

Copyright © FPMs-UCL, Nov. 2008