

# A real-time tracking environment towards cross modal applications in public installations and education

*Tony Brooks*

Associate Professor

Aalborg University Esbjerg ,

Niels Bohrs Vej 8

DK-6700 Esbjerg

TLF +45 79127716

FAX +45 79127710

E-mail: [tonybrooks@cs.aue.auc.dk](mailto:tonybrooks@cs.aue.auc.dk)

*Volker Kruger*

Associate Professor

Aalborg University Esbjerg ,

Niels Bohrs Vej 8

DK-6700 Esbjerg

TLF +45 79127718

FAX +45 79127710

E-mail: [vok@cs.aue.auc.dk](mailto:vok@cs.aue.auc.dk)

*Michel Guglielmi*

Assistant Professor

Aalborg University Esbjerg ,

Niels Bohrs Vej 8

DK-6700 Esbjerg

TLF +45 79127717

FAX +45 79127710

E-mail: [mg@cs.aue.auc.dk](mailto:mg@cs.aue.auc.dk)

**ABSTRACT:** This paper presents aspects of an interactive installation proposal which has been researched and developed at Aalborg University Esbjerg (AUE) within the Medialogy education. The installation is a site specific human tracking system in a public space. It features a design that highlights the utilization of a computer system that enables assignable artistic content to relay non verbal indices of various forms. The design challenges the traditional interactive installation “rules” in respect of human ‘opt in’ - ‘opt out’ preference by presenting a situation where the system selects participants who are captured in a non-volatile space. The suggested resulting human reaction to such – provocation; meekness; exuberance (etc..) is of interest for the research team towards further refinement of the system concept and elements. The object of the project team is to create an installation so as to motivate people to acquaint and subsequently dialogue in a ‘human-to-human’ way which is a result of the system acting as a ‘virtual conduit for human social interaction’. The project - currently “work-in- progress” - has been accepted by the leaders of the initial site designated as specific to the design, i.e. the Metro in Copenhagen, Denmark.

**KEYWORDS:** ”Interactive psychological “drama”, ”non-verbal communication”, ”avatar mapping to tracking”, “expressive cues”, “interactive advertising”, ” Pattern recognition”, “Model-based background subtraction”, “Color-based segmentation”.

## INTRODUCTION

Many people can be observed not being conscious of their fellow commuters or indeed being aware of them to the point of seclusion (*how far away can I sit?*). A system

that is able to encourage personal communication between people gives opportunities for an improvement in life quality where the richness of human contact is realized. A system with such attributes has additional supplementary levels to explore at an empirical level within education and research; for example, in so far as content and human response to same is determinate. Such levels are hinted at in this contribution. Further, the questionable system characteristics of choice – e.g. random selection of participant or determinant selection - based on, for example fuzzy logic where various defining attributes based on system initial selection determines the corresponding subsequent choice (matched pairs) - or if no match (i.e. rules equated) defined and chosen within a predetermined time frame to deem a non match and select in an arbitrary fashion. In such a system the necessary rules enforced upon the selection process applied in dealing with the complexity involved in human social interactions may prove to be insufficient at this point without an implementation of “learning” algorithms such as in the case of a neural networked system of one exhibiting inherent artificial intelligence. Monitoring of the installation upon realization evaluates the rules and removes and refines where necessary (Elsea) (Cox).

## BACKGROUND

The abundance of online dating, chat rooms and other non-human-to-human (in person) communication forms point to the populous being aware of the need to communicate. In such online activities ‘identity’ is a potential myth, one can be anyone. The installation detailed in this paper deals with ‘anti-myth’ where the terminology of ‘identity’ is allied to a real person that has a potential of being selected by the system and thus an opportune phenomenology of space is a potent attribute therein. In other words it encourages one (if selected) to be themselves and to perceive others as themselves which in itself is a motivational encouragement to meet.

One can also relate to the “active programmable badges” that were popular in Japan in the nineties for encouraging

In CADE2004 Web Proceedings of Computers in Art and Design Education Conference, M. Agger Eriksen., L. Malmberg, J. Nielsen (eds.). Copenhagen Business School, Denmark and Malmö University, Sweden, 29 June – 1 July 2004. [asp.cbs.dk/cade2004/proceedings/](http://asp.cbs.dk/cade2004/proceedings/)

blind dating by their owners. These badges were programmed by the owner to be representational of their person; their desires and preferences, their background and job; interests etc. The badges would then be worn and when in the proximity of a secondary badge that exhibited similar characteristics a matching would occur and the persons were encouraged to dialogue. Such systems exhibit a novelty value that has an important element <fun>, as of course when the badge would sound or vibrate indicating a favourable potential match with another the individual's choice of whether to act upon it was inherent to the design. A similar simile can be attributed to the installation in this paper. However the initial (badge to badge seeking - before the actual personal communication) potential myth as described earlier in respect of online communication is also suggested as a temptation for the owners in these badges. So for example one may program oneself as incredibly healthy and wealthy, joyful and pleasant to be with, yet in reality they are <not>! In our installation one can assimilate the escalator descent as being parallel to the metaphorical badge seeking moments (- yet in our case only the select few are wearing a "badge"), followed by body language and potential flirtation (where appropriate). These latter moments of observation are deciding each <player/performer> on the next course of action.

It is in these moments, and the subsequent escalator exit to lower level, that make the installation interesting as relative to the earlier mentioned traditional interactive installation rule opposed - that of 'opt in' and 'opt out' - the system detailed here exhibits a more contemporary attitude such that each participant is the determinant (a factor that determines the outcome). In this way correspondence to the system and the other player/performer has been motivated and encouraged by the installation as still it enables the choice of whether to "acquaint" to another or not.

The concept could be overviewed as acting as a virtual social ice breaker where we are attempting to stimulate the consciousness such that the journey is - *largely a matter of perceptual experience, bodily feeling, and mood* (Tye 2003). Such a journey resulting in a social meeting effected from the installation.

We envisage that a selected participant will respond in a variety of ways which will indicate at their identity, their perception of the situation and their mood. We may even predict to stimulate an improved dress code and appropriate manner in the Metro as people become conscious of the opportunity for "contact" with a new friend or possible date.

### SCENARIO

Imagine, e.g., a scenario with two escalators descending, facing a big wall (see Fig. 1). This is a scenario which often appears, e.g., in metro stations. Cameras will be installed on the wall facing the escalators and the video stream will be projected onto this wall. In that sense, the

wall will become a big, artificial mirror of the scenario facing it. If an individual enters the scene on the top of the escalators and starts descending he/she sees ahead of him/her the whole environment reflected on a wall, including the escalator he is standing on. By automatically detecting the individual in the mirrored image, the individuals may be replaced by avatars or, e.g., speech bubbles may be inserted into the mirror-image. In that sense, a descending individual may notice that his reflection has an overlaid image. This image is following him/her as the escalator descends. A similar event may happen in the above scenario upon the other descending escalator. The vision system can detect this event and can enter into the mirror-image speech bobbles that may initiate some sort of communication between the two individuals.

### GOAL OF SUCH A SYSTEM

The expected result of the installation is a shared psychological "drama" - resulting from a defined number of people being instantaneously elected as support for a given message transmitted by the installation. The message can be in various forms and social as well as the obvious business angles are amongst those that have been explored.

One important goal was to consider safety and as it is not an interactive gesture tracking system but a system with a proposed core tracking algorithm with predefined trajectories and velocities (where movement thresholds determine the tracking fix) - individuals who try to influence cannot affect the avatar image feedback, this was an important point so as to ensure that there may be minimal disruption to the traffic on the escalators and to reduce disruptive "horseplay" which is a foreseen aspect of such installations. In fact due to the recent history of the Metro - (three people have died through misadventure) - such aspects are considered with the arranger towards utilizing the installation for even potentially improving safety. Thus the installation can be overviewed as having a goal as a tracker of human form with a designed limitation on its intrinsic interactive elements (passive/dynamic) to account for the public space and safety of same.

Another goal was the consideration of educational and research potentials from the installation and as such it is suggested that for example, the installation offers the potential for human behaviour analysis (considered profound) relative to the opportunities for exploration of content manipulation across subject disciplines

### EXPERIENCES

Often in interactive installations multiple sense stimulation of participants is intrinsically designed with subsequent intellectualised perspective analysis. This installation is conceptualized as favouring a non-intellectualist approach so as to balance the experience towards the emotional rather than favouring the scrutiny (yet as the previous above states analysis is a goal). Thus

an inquiry (either in the physical or social sciences) along naturalistic lines rather than positivism is suggested and a pre-empted supposition is hereby indicated.

It must be stressed, when considering the experience factor, that we are not intending the installation to approach a suggested “Unity of Perceptual Experience through time” where the five modality-specific experiences of the senses are stimulated (Tye 2003). Our concern is with image specific interplay solely, and a distinct choice was made to eliminate any sound elements in the feedback as the image intrusion is believed suffice for our purpose. However, considering the experience, when one does consider the reductive perspectives as to that which is influencing the multiplicity of stimulation one may presume that relatively: *At any given time, a subject has a multiplicity of experiences.... These experiences are distinct from each other.... But at the same time...they seem to be unified, by being aspects of a single encompassing state of consciousness.* (Bayne and Chalmers). The installation is an attempt to stimulate that state of consciousness to be motivated to act – “*And as we think, so do we act*” (Schwartz and Ogilvy, 1979).

Regarding “a unified experience” additional reference to Descartes and Kant regarding their writings on aspects of the unified consciousness is difficult to ignore due to the other opposing philosophy (Dennett, for example) that holds that consciousness is disunified. What makes such discussion interesting in respect of interactive installations is that as suggested where there is no one kind of consciousness and there is no one kind of unity (Tye 2003). Introspective awareness, as one kind of consciousness, is a potential also with the installation whereby ignorance of the interaction is possible. This is where the selected person is distanced through having a higher-order thought that he or she is undergoing a certain mental state, where this thought is not based on inference or observation (Rosenthal 1997) or alternatively a more inwardly directed state (Lycan 1997). This is termed as *I*-consciousness by Tye and he further details with examples another form that may be considered relative and what he has termed *D*-consciousness or “discriminatory” where cognitive aspects come into play so as to notice the events and it is in this state we would anticipate our audience. Of course there are other states which could prevail so as to influence the participant, such as for example, chosen ignorance following perception, and such choices (a form of ‘opt out’) can be implemented into a subsequent research.

Therefore relative to our expected audience we look to achieve communicational – communal experience - “play” aspects where the avatar is conveying non-intrusive third person messages to (and between) the main players as first person inferences. This is where the distances between the participants are ‘broken down’ by the avatar “bridge” interaction resulting in a physical contact (dialogue) or (in the very least) a body language

signal exchange. Or in the instance of *I*-conscious state – what could be observed as a form of blissful ignorance? The novelty value of this installation will be the use of a recognition device for the purpose of conveying real time interactive messages between people which through the inherent ‘playfulness’ of the system will be better assimilated by the player/audience. Messages that span the ground from those subliminal in character to more consciously perceived forms would be able to input to the system.

### **POSSIBLE APPLICATIONS**

Ideally the scenario is that two strangers arrive at the station and enter the escalator from either side without noticing each other and end up communicating with each other when exiting the escalator. This is observed and encouraged by others who may be present.

Another scenario is that where the system is used informing a larger amount of people in a way to insure best possible assimilation and acceptance of the message. By using an interactive medium such as the “electronic billboard” (the screened messages) the opportunity to give instructions (for example in respect of safety) is available. As stated earlier since the station opened in 2003, three people have died through misadventure - by running or trespassing and falling down the central atrium. Our system will of course not prevent such occurrences: Physical solutions are to be implemented such as higher escalator guard rails in an effort to curtail such “horseplay”. But safety messages like “slow down” or “don’t trespass” can be part of the content generated by the installation and it is suggested to have a potential to improve the situation.

As indicated earlier in the text one could of course argue that an opposite condition will prevail, i.e. the novelty and the interactive aspects, even when limited as outlined, may entice people and incite them to act dangerously while on the escalator. This hypothesis is taken seriously and is the object of discussion and will be closely scrutinized and tested.

As it is now, the client considers this installation as provisory and lasting over a period of 3-4 weeks. We will use this timeframe to initiate behavioural studies from the users.

### **ETHICAL QUESTIONS POSED.**

We are aware that we are projecting the image of people in the public space without giving them any option to avoid such situation. There are precedents to such cases (see reference section), but most important is that we consider the installation as mostly functioning like a mirror, a function which is usually accepted in public rooms. The exception in this case is of course the person chosen randomly to be replaced by an avatar. In this case, we made sure not to be offending by choosing some neutral figures (abstract silhouettes or iconic figures).



**Figure 1: Example from a Metro station access in Copenhagen of two escalators (outer left and right) descending with projection wall in place.**

**METHODOLOGY**

A major observation is that the imaging conditions are constant over time. This considerably simplifies the construction of a system that is expected to function independently of the day time. The key techniques in realizing the vision system are:-

- Model-based background subtraction
- Colour-based segmentation

**Model-based background subtraction with Eyesweb©**

Background subtraction (BGS) is being used to detect automatically the regions of interest in the video-stream. BGS results into a probabilistic image with pixel values between 0 and 1 stating the likelihood of a pixel belonging to an object novel in the scene. We employ a Model-based BGS (MBGS). MBGS uses a classical non-

parametric BGS approach.

During a training phase, which assumes a good image quality, exemplars are being learned. Exemplars are instances of objects directly extracted from the video. While the BGS deletes the background, exemplars appear as bright areas in the image. Different exemplars, e.g., different stances of a walking person, have different shapes. These shapes are learned as instances of likely objects appearing in the video.

During the learning process, the shapes are clustered for later access in a tree-like structure (see Fig. 2.). In our scenario, only frontal (as perceived from camera position) contours would appear.

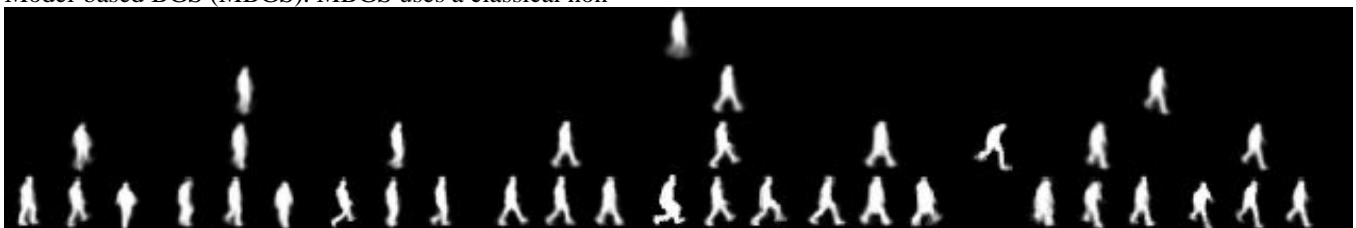
During the MBGS process, these contours are used to verify the output of the ordinary BGS. Given a probability image  $P_t$  as produced by the BGS process, the model knowledge is used to verify this data. A condensation-based approach is being used for this where we define a 4-D probabilistic variable  $X = (h, i)$ , where  $h$  denotes the geometric deformation and  $i$  is a number referring to a particular contour in the hierarchy.

Using the classical Bayesian propagation over time, we can compute the probabilistic density function (pdf)  $p$  over  $X_t$ , given the images  $P_1, \dots, P_t$ . Particle filtering is used to estimate the (pdf)  $p$  over the image  $P_t$ . Particle filtering has the advantage that it can estimate even multi-modal pdfs that are required for example due to several people being in the scene. The above method has the advantage that the estimation of  $X_t$  considers all images  $P$  that appeared so far, thus leading to a robust estimation of the random variable  $X$ .

In our present experimental setup, Eyesweb functions are used for image filtration and for computing BGS. The resulting image is then fed into an external MBGS engine.

**Color-based Segmentation**

The MBGS selects region of interests in the video input. These regions are further processed with statistical color segmentation. We describe an individual as a set of vertically aligned color blobs  $H = \{B_i\}$  that define each a region of homogeneous color. Thus, each blob can be described by a unimodal distribution  $h(c)$ . Following the spirit of BGS we define a likelihood measure for the Region of interest actually containing a person. This will

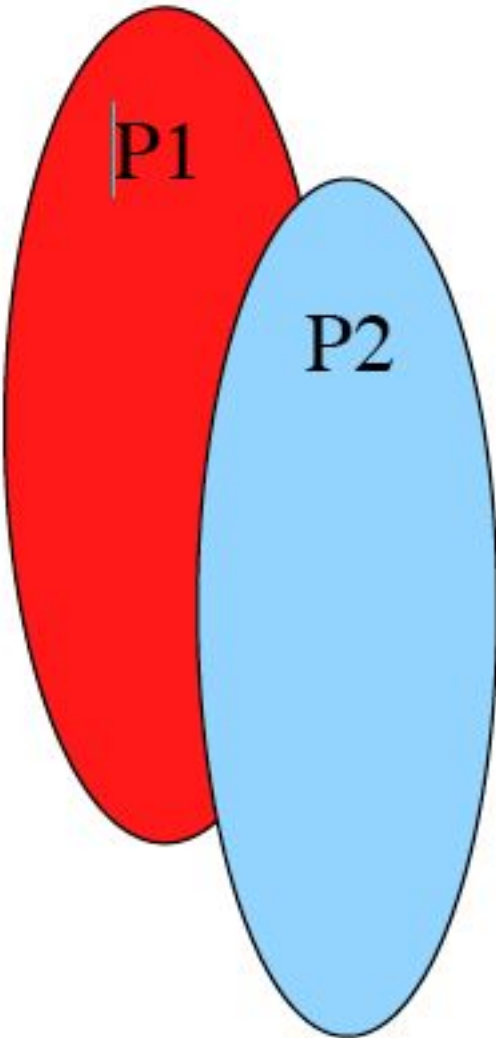


**Figure. 2. A typical hierarchical representation of walking humans.**

allow the system to handle occlusion which is likely to

happen on an escalator. We thus employ a technique

whereby the first uses a statistical model of the color distribution and diatribes each person as a unimodal distribution with a Gaussian kernel (see Fig. 3), the latter uses a ground plane constraint to reason about occlusion. We employ a voting scheme to find a robust depth-reasoning.



**Figure 3: An example of Occlusion.**

## CONCLUSION

The interactive installation as detailed in this paper has been developed for the specific context of the metro stations in Copenhagen. It is our belief that it could find further applications in a limited number of other related situations where escalators are located and as such it has much potential to be adaptable for various situations. It is an art piece that could result as a commercial vehicle for the business world. It is an installation piece that utilizes a common desired parameter from within conceptual art, which is the “control” aspect of the interactive audience.

It is a project with educational and research potentials. It is a vision system piece, which is able to cope with a real-world scenario, where imaging conditions are invariant but sub-optimal: The fact that the escalator scenario is a restrictive area and movement trajectory is linear at a pre-empted velocity offers a predetermined aspect. Variables are mostly time related and of course human related. It is anticipated that this project is realized in 2005.

## ACKNOWLEDGEMENTS

Eyesweb software © 1999-2003, InfoMus Lab (Laboratorio di Informatica Musicale) DIST, University of Genoa, Italy.

Students from the Medialogy education involved in the project: Vostrup, T., Galatius, S., Halse, K. A., Seest Christiansen, M., Ravnkilde, C., Pedersen, S.B.W.

## REFERENCES

- Bayne, T. and Chalmers, D. 2003 (pp. 23). “What Is the Unity of Consciousness?” In *The Unity of Consciousness: Binding, Integration Dissociation*, ed. A. Cleeremans, Oxford: Oxford University Press.
- Cox, E., Fuzzy logic for Business and Industry: Charles River Media 1995.
- Dennett, D. 1991. *Consciousness Explained*. Boston: Little, Brown.
- Descartes, R. 1641. *Meditations on first philosophy*. Paris: Michel Soly.
- Elgammal, A. and Davis, L. Probabilistic framework for segmenting people under occlusion. In *Proc. ICCV, ICCV01*, 2001.
- Elsea, P. Fuzzy Primer  
[ftp://arts.ucsc.edu/pub/ems/FUZZY/fuzzy\\_Logic](ftp://arts.ucsc.edu/pub/ems/FUZZY/fuzzy_Logic)
- Gavrila, D., The visual analysis of human movement: A survey. *Computer Vision and Image Understanding*, 73:82-98, 1999.
- Isard, M., and Blake, A. Condensation - conditional density propagation for visual tracking. *Int. J. of Computer Vision*, 1998.
- Kant, I. 1980. *Critique of Pure Reason*. Trans. Kemp Smith. London: Macmillan.
- Koller, D., Weber, J. and Malik, J. Robust multiple car tracking with occlusion reasoning. In *Proc. ECCV, ECCV94*, 1994.
- Krueger, V., and Zhou, S. Exemplar-based face recognition from video. In *Proc. European Conf. on Computer Vision*, Copenhagen, Denmark, June 27-31, 2002.

Li, B and Chellappa, R. Simultaneous tracking and verification via sequential posterior estimation. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Hilton Head Island, SC, June 13-15, 2000.

Lycan, W. 1997. "Consciousness as Internal Monitoring." In *The Nature of Consciousness*, ed. N. Block, O. Flanagan, and G. Guzeldere. Cambridge, Mass.: MIT Press/Bradford Books.

Rosenthal, D. 1997. "A Theory of Consciousness." In *The Nature of Consciousness*, ed. N. Block, O. Flanagan, and G. Guzeldere. Cambridge, Mass.: MIT Press/Bradford Books.

Schwartz, P., and Ogilvy, J., (1979) *The emergent paradigm: Changing patterns of thought and belief*. Analytical report 7, Values and Lifestyles program. Menlo Park, CA: SRI International.

Tye, M., *Consciousness and Persons*, Unity and Identity (pp 20) (pp. 1); MIT Press 0-262-20147-X

Notes to referenced related works below: these places become virtually connected. In both cases (2 & 3), people (User of the underground parking and people on the urban squares) are the object of media-content.

#### REFERENCES TO RELATED WORKS

1. **Co.in.cide**: An installation by Heimo Ranzenbacher & Ars Electronica Futurelab. Ars Electronica September 2003 (A project using video recording and form recognition real-time outlines of human bodies).
2. "Underground car park **ODE**": Zwarts & Jansma architects (A project in Oosterdokseiland, Amsterdam. where activities in an underground parking is filmed and projected on a giant screen in the public space above).
3. "**The Global village square project**": initiated by the "MCLUhan program in culture and technology" in Toronto. (A project where activities in well known urban squares are filmed and shown real-time on screens in corresponding places across the world).