Presence

The Integration of Classical Artistic Media in a Smart Space Prototype

Narvika Bovcon

Franc Solina

Borut Batagelj
Computer Vision Laboratory
Faculty of Computer and Information Science,
University of Ljubljana,
Ljubljana, Slovenia
narvika@bovcon.com
franc.solina@fri.uni-lj.si
borut.batagelj@fri.uni-lj.si

Aleš Vaupotič
ArtNetLab
Society for Connecting Art and Science
Ljubljana, Slovenia
ales@vaupotic.com

Damir Deželjin
Ljubljana, Slovenia
damir.dezeljin@gmail.com

Abstract—**In the mixed reality of the computer installation *Presence*, which functions as a prototype for a smart space, the visitor is placed in the position of a person in audience with the king. The interaction with a digital avatar is structured according to the rules of social behaviours and following the script of the Shakespeare's play. The paper explains different aspects of the conceptualisation of an interdisciplinary collaboration between artists and computer engineers.**

Keywords: computer vision, new media art, interactive installation, digital video, human-centered human-computer interaction design, digital animation, smart space, synthetic realism, avatar, multiple-screen video projection

I. INTRODUCTION

The project *Presence* has been initiated by new media artists Narvika Bovcon and Aleš Vaupotič, with the aim of integrating the experiences of art with those of science and new media technologies. The project has been developed first as an artistic concept and then tested as a prototype of a human-centered human-computer interaction system in the form of a gallery installation. Computer engineers Damir Deželjin, Jurij Porenta and Andraž Sraka, who were in 2007 students at the Faculty of Computer and Information Science, University of Ljubljana, have realized the technical part of the project. Thus, *Presence* is an interdisciplinary effort that combines expertise from various disciplines of knowledge, involving video artists and new media theorists, a graphic designer, a literary scholar and computer engineers.

With *Presence* we have tried to achieve a specific configuration of space, visitors, computer vision equipped sensors and two responsive video projections that would result in an interactive art installation and could be interpreted also as a prototype for a new media apparatus [1] of a so called "smart space". To structure the meaning of the interaction in this installation we have used a scene from Shakespeare's play *Richard III*. The re-enactment of the chosen scene poses boundaries on the behavioural patterns of the visitors in the gallery space, which gives also a framework for developing a digital avatar that interacts with the visitors.

We have used Maya modelling and animation techniques to construct the digital avatar of king Richard III. With the software VVVV [2] we have established the non-linear responsive multiple video projections of the avatar rendered as short video clips. For face detection and territory surveillance we have used OpenCV library [3].

The research into human-centred smart space technologies in the project *Presence* focuses on personal experiences that evolve from inter-human communication structured by subject positions. Social institutions, regulated by power-knowledge relations that define our perception of the world and ourselves [4], are the foundation for developing artificial intelligence as an integral part of our mixed reality.

II. ARTISTIC CONCEPT

It is significant at this point to present the development of the project *Presence*, since it didn't start as a schema for interaction based on computer vision but instead it developed gradually from artistic concepts. The artistic approach took into account the aspects of human perception and responses as researched in humanities, literature, theatre, painting, installation art, and finally, in media art. Artistic research is a relevant approach for building interactive systems in collaboration with sciences.

A. To Brecknock, while my fearful head is on.

The first step of the project *Presence* was made in 2006 with the digital video *To Brecknock, while my fearful head is on* represented with a frame in Fig. 1. This is a short (2 min) digital animation showing king Richard III that was rendered as digital video and cut into few seconds clips, which were then fed into the VVVV software for non-linear multiple projections display.

However, the artists Bovcon and Vaupotič explored the history play *Richard III* already in 2002 making a videotape and video installation on the basis of the literary text. The focus of the project was the deconstruction of the drama into non-linear elements of narration. Five dominant aspects have been extracted from the play that drove the narration: the battle, the peace, the oaths, the curses and Buckingham. Lord Buckingham is the princely cousin of Richard III; he helps him to the throne. They are friends and allies until a certain point, when Buckingham does not comply with the ruthless rule of Richard any more. There is a very subtle moment at the dramatic peak of the play, when the friendship breaks and from that point on everything goes downhill. The suspense at the moment of rejection is prolonged and symbolized with the strikes of the clock. We took this highly emotional and reflective moment to be re-enacted in the interactive installation *Presence* as an archetypical scene that involves the viewer.

Digital animation in *To Brecknock…* was done with Maya 7, ZBrush and Adobe After Effects 6. We chose a person as the model for the avatar, photographed him, recorded his speaking and used this reference data to make a digital portrait of him. We sculpted the head in ZBrush, textured it in Maya with a layered image texture composed of photographs and refined in Photoshop. We used mental ray sub-surface scattering shader (misss_fast_skin_phen) for skin and translucence of human flesh. We used Maya 7 hair and cloth simulations. The movement of the body of the avatar while being seated was captured with a mechanical motion capture suit and manually fine tuned in graph editor. The movements of the fingers of the hand were key-framed. We used an array of blend shapes for facial expressions and for the movement of lips while speaking. The animated king pronounces the sentence: "Well, let that rest." We paid special attention to optimising the Maya scene, omitting or simplifying all the elements that were not visible in the final rendered frames, and appropriating the amount of detail to the output screen resolution of 800 x 600 dpi.

Several short video clips contain different gestures of the avatar. One clip carries the image of a square with the text from the chosen point in the play so that the viewer can understand the context of the action. And a different video clip shows a travelling of the virtual camera around a digital model of the throne, in the varnished surface of which the very scene is reflected as acted out in a BBC production of the play, whereas the fly-through camera shows also the wild boar as the king's coat of arms. Thus, all the imagery in the video clips is chosen carefully and carries condensed meanings. The short video clips are played non-linearly according to the visitor's movements in the gallery space.

Figure 1. A frame from the digital animation *To Brecknock, while my fearful head is on*, showing king Richard III.

B. Theatre and Synthetic Realism

The challenge of the digital avatar to be as realistic as possible and finally confused for a real person was approached in *To Brecknock…* and resolved with the concept of targeting the overall painterly effect of the image. We combined the aspects of photo-realism as derived from the photographic apparatus and imposed over our perception, with the counter activity of realism and mannerism from the painting tradition as found in Caravaggio's *chiaro scuro* painterly technique. The dark and abrupt partitions of shade contrasted with gleaming whites allowed for less rendering detail, so that in the end we didn't need to use ambient occlusion layer at all, sparing us thus quite some rendering time. In 2006, when no adequate solution for a photo-realistic digital human has been achieved yet even in the industry, our approach presented a possible alternative solution to the problem by transferring the image closer to the representational traditions of painting. Today, in 2009, after the digital head of Benjamin Button was convincingly presented, our king cannot compete on the level of photographic similarity to the human any more, however it is still relevant as an example of "living painting", since it refers directly to the painterly elaboration of a certain art historic period.

On the other hand, the (not truly realistic) digital avatar can be regarded as a puppet or a marionette, even more so when found in a theatrical setting. The artificial theatrical situation determined by the script of the play positions human actors one-step away from real life situations and one step closer to puppet theatre. However, real actors bring to the stage along with the drama also their bodies, their human existence that lives on after the curtain comes down. The avant-garde theatrical traditions researched the possibilities of excluding the actor's bodily presence from the theatre - for instance with the notion of the superpuppet (Übermarionette) of Edward Gordon Craig [5] or with the experiments of replacing actors or more accurately covering their bodies with geometrical forms in the ballets of Oscar Schlemmer at Bauhaus [6]. Today, Craig's superpuppet could be interpreted in a new and challenging way with the photo-realistic procedures of digital animation.

The notion of synthetic realism as explained by Lev Manovich [7] could represent a bridge between the seemingly opposing sides: if modern theatre tried to exclude from the statement in the theatrical plurimedial language the serendipity of meanings that the actor's never fully controllable body carries around, the virtual computer worlds establish "realism in parts", synthetically, and thus on a completely different level. The image of a digital man contains no remnants of the physical presence of the human body and carries merely those meanings that are intentionally coded in it.

However, *Presence* addresses also a far more complex issue about the transposition of the theatrical medium into the mixed reality of new media technologies. It establishes a smart responsive space, where the action is monitored by computer vision software. In the gallery installation there is another meaningful excerpt from Shakespeare's drama *As You Like It* that is written on the wall of the smart space: "All the world is a stage." Smart spaces in general, learned to distinguish between standard human behaviour and deviations from behavioural patterns in the given social loci, resemble theatrical settings, where people behave according to the script [8].

III. CONCEPT OF INTERACTION

In the gallery, *Presence* functions as a double video projection, which is governed by two web cameras connected to computer vision software that recognize movements of visitors in the installation space. The two video projections show consecutive video clips and play certain successions of videos depending on the behaviour of visitors.

First, at setting up the interactive system, the web cams are calibrated to the empty space of the installation. The first camera records zone A in the middle of the gallery room, whereas the second camera records zone B right in front of the video projections and very close to them. Thus, when the interactive computer installation is launched, the first camera detects the person that steps into the zone A and moves around in it and the second camera detects the person that steps into zone B. Before the visitor enters any of the marked zones, one of the video projections shows short video clips of the animated digital king played in random succession, whereas the other plays the looped video clip of the throne, as in Fig. 2.

Figure 2.   Before entering zone A, the visitor encounters random video clips on two synchronized projections.

The first camera is connected to computer vision software and captures the face of the visitor, thus recognizing him/her as a person and not just any interference. If the camera captures the person's face, that means that the person is looking straight into the two video projections, above which the cameras are placed. The audience with the king poses specific protocol rules that reflect the power relations between the king as a sovereign and his subject, i.e. the visitor. The visitor has to regard these rules and perform as required. For example, it is forbidden to look straight at the king's face, because this action is regarded as disrespectful. If the visitor in the smart space of *Presence* does so, both projections become white squares and no image of the king is seen any more, so the communication with the avatar is blocked.

The second video projection shows a longer video clip with the throne, which is played repeatedly as the fundament of king's power (except when the king is looked straight in the face and both projections are interrupted by the white squares). If the visitor steps into zone B, the second camera records it and interprets this approach to be too close to the projection, which means too close to the king, and the video clip with a warning blinking flash light is triggered in place of the throne, see Fig. 3. Simultaneously, the excerpt from the drama that describes the situation in which the visitor is placed in *Presence* replaces the video clip of the king on the first projection. At this point, the visitor is very close to the projection, so he/she can read the text easily.

However, if the visitor approaches the king in the right manner, i.e. not too close and looking at him from the side, with a lowered head, then the computer vision software doesn't recognize his/her face; the camera only recognizes the entering into zone A. After playing short clips of the king in random order when there is nobody in zone A, now with the right behaviour of the visitor the video clips of the king are finally played in the right succession, thus presenting the whole action that ends with the king's speech and dismissal. So, if the visitor enters the relation with the avatar according to the rules, the king will finally speak to him/her. The speech is the only sound in the installation; therefore it is regarded as an important event in the communication. Fig. 4 shows the incorrect approach to the king, when the communication is obviously blocked.

Figure 3.   Stepping into zone B, the visitor triggers a warning flashlight; he/she came too close to the digital avatar.
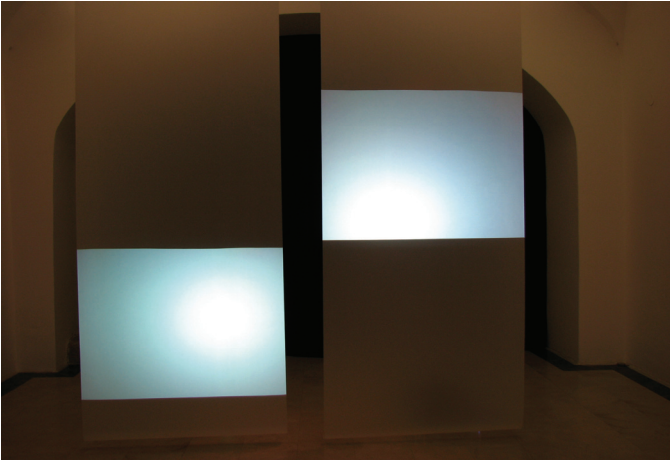
Figure 4.   In zone A, the visitor looks straight at the king.

From this description we understand that the interaction in *Presence* is planned carefully and according to the rules of communication, social institutions and inter-personal relations. Even more so for we have borrowed the scene re-enacted in *Presence* from a famous literary text, which guarantees a deeper humanist understanding of social behavioural patterns and a better dramatic elaboration for the constructed model of inter-personal communication. Artists relate to other great artist's works, e.g. Shakespeare's, with specific understanding and the ability to transpose the content into other media.

IV.   USER RESPONSE

The installation *Presence* functioned well in the gallery setting and had good reception from the visitors. The software worked as planned in the concept for interaction; however only after testing the interaction in the actual set-up we have recognized some aspects of the constructed model that we were not considering before.

First, there was the moment of entering the smart space, when the visitor had to be informed about the situation and the rules of interaction. As often encountered at new media art exhibitions, the visitors are unfamiliar with the interactive modules and have trouble experiencing the art works. We have realized that a simple description (delivered by the hostess in simple words) of how the installation is supposed to function and how the visitor must behave was enough to motivate the visitor to try it. The visitor usually played with all the possibilities and after some trial and error everyone succeeded to approach the king in the right manner. On the other hand, those visitors who were not informed about how the installation works moved freely in the space and observed the capricious behaviour of the king: one visitor described *Presence* as an abstract modernist installation. In a funny way, the visitor who doesn't know that he/she has to look at the king from the side after some time loses interest and turns away from the projection: at that point the king speaks to him/her and gets back the attention.

The installation functioned for one or many visitors as planned. Computer vision software found successively all the faces in the zone A that were looking straight at the king and

two white squares appeared on the projections for every person that did not obey the rules. So, after some time of trying to look at the king from the side, the visitors started to warn each other to behave appropriately, since they were sharing the experience. Thus social behaviour appeared not only between the visitor and the avatar but also among the visitors.

Finally, the background knowledge about the content of the drama *Richard III* that enabled a deeper understanding of the interaction was available at the gallery in the explanatory text printed in the catalogue that pointed the visitor towards reading Shakespeare's play. The hint to Shakespeare was also the writing on the wall: "All the world is a stage," which everybody recognized and interpreted in the new context of experienced mixed reality. Thus, the visitor gathered information about the interactive installation through a plethora of media and on different levels of meaning. The biggest challenge was to combine the experiences and knowledge in order to construct a meaningful interpretation of this new kind of visitor status: not any more a participant to a happening, not exactly a role in the theatre play nor a performer, but an inter-actor in a social situation comprised of real and virtual characters and suspiciously similar to the monitored public spaces in real life.

V.   TECHNICAL ASPECT

A.   VVVV Framework

VVVV is a toolkit for developing multimedia projections with special emphasis on real time video controlled by user interaction captured using user input modules like web-cams. VVVV contrarily to most other similar products has no separation between the multimedia project development and presentation layer. Basically the very same interface is used for development and presentation. This means the presentation is running during development, too.

Developing a VVVV project consists of graphically editing a VVVV model. This means adding nodes, which are 'active' elements, to a window and connecting them together. There are lots of different nodes a user can use for developing a multimedia project. These nodes are added to a window called a patch. Actually all VVVV windows are sub-patches as even the first window the user faces upon starting VVVV is a sub-patch of the root patch that is hidden in the background.

It is very important to understand the VVVV concept of in-schema values calculation. VVVV computes all values of a schema at once on a predefined frame rate (10 times per second by default). This means the user hasn't to worry about delays. However, this introduces problems when creating loops within the patch. To prevent unexpected results, VVVV prevents directly connecting an output of a node to any node that is on the input chain of the same node. This limitation is introduced with the concept of computing all values at once: it prevents computation of the input, which depends on the output, which is at its time dependant on the input again. This limitation can be workaround by using the Framedelay node that presents the value of the previous frame of the input signal to the output pin.

VVVV has a built-in clustering capability allowing distributing the logic computation as well output input / output processing (e.g. video output) and so making it possible to achieve certain presentation targets like smooth video projection. The mentioned clustering capability is called Boygrouping. Constructing a Boygroup involves the following steps:

- Adding host detection nodes to the schema (e.g. Boygroup (server) and Boygroup (client) nodes);

- Installing all required packages on all hosts the schema will run on. Part of this is setting the same directory structure on all nodes as well copying all materials beside the schema project (.p4v files) to hosts that will be used as clients;

- Starting VVVV in Server mode on one host (use the '/server' CLI switch). Example: 'vvvv.exe /server';

- Starting VVVV in Client mode on all other hosts (use the '/client <server_IP> <Project_dir>'). Example: 'vvvv.exe /client 192.168.100.100 C:\presence'.

- Opening the project schema (.p4v file) on the server host.

- After a VVVV schema is loaded on the server, the user has to select the nodes meant to be distributed followed by pressing <Ctrl>+B.
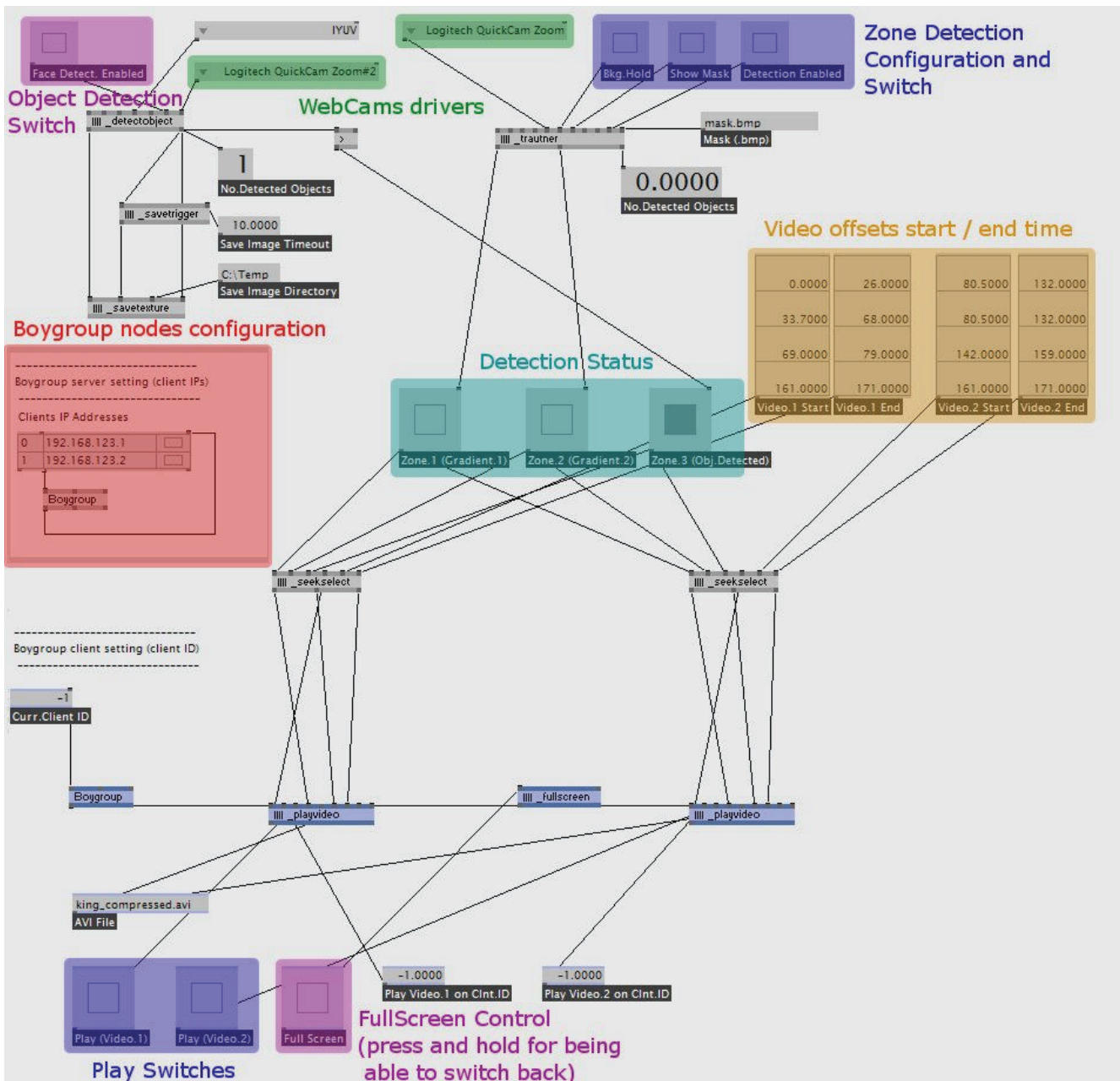

Figure 5.   VVVV project schema.

VVVV Boygroup clustering uses TCP and UDP ports 3333 for communication. For this reason it is important to configure the firewall to pass through the connection on mentioned ports. The project file is self-explanatory but for usage we must know, at least how to customize the main two features' settings for Boygrouping and Masking.

- Changing Boygrouping configuration; before you start using the Boygrouping functionality as well before any output is displayed to the screen, the Boygrouping clients should be configured. First the Boygrouping client IP-s should be updated in the project file. Afterwards the two Play 'Video.X on Client.ID' input boxes should be updated with client numbers where the individual video should be played – client numbers can be found in the Boygrouping table; the server ID is '-1', as in Fig. 5.

- Zones Bitmap Mask; the zones definition mask has to be an 8 bit greyscale file in BMP format containing just two colours that are used for defining both zones. Pay attention when using anti-aliasing, since creating or saving the mask file will end up defining a bunch of zones and so the zone detection will not work.

- Normal operation; it turned out VVVV has a bug (or a 'feature') preventing setting the Webcams input signals just once. The workaround is to select the driver every time the project is opened. However, sometimes this is not enough as a case when one or both input video windows don't appear. The workaround for the later one is changing the Video-In driver twice – once to a webcam driver that will not be used for the certain input and later back to the correct webcam driver. After correctly configuring the webcam drivers, all object detection input buttons should be switched ON and OFF before the user desired configuration is set.

B. Face Detection

The face detector is based on the AdaBoost [9] learning-based method because it is so far the most successful in terms of detection accuracy and speed. AdaBoost is used to solve the following three fundamental problems:

- learning effective features from a large feature set,

- constructing weak classifiers, each of which is based on one of the selected feature sets and

- boosting the weak classifiers to construct a strong classifier.

Weak classifiers are based on simple scalar Haar wavelet-like features, which are steerable filters. We use the integral image method for effective computation of a large number of such features under varying scale and location, which is important for real-time performance. Moreover, the simple-to-complex cascade of classifiers makes the computation even more efficient, which follows the principles of pattern rejection and coarse-to-fine search.

C. Movement Detection

Movement is calculated through subtraction of two images. If variable 'Hold Background' in VVVV framework is set to one, a static image is sampled and the current image is always subtracted from it. Otherwise always two consecutive images are subtracted.

VI. CONCLUSION

In the smart space of *Presence,* Buckingham's position is taken over by the viewer, who thus becomes a substitute for the character in the play as well as for the body of the actor on the stage. The apparatus of the theatre functions as the mediator when the model of a new media smart space is being established. As encountered increasingly often in the everyday use of smart devices, in this installation - as an example of a smart space - the visitor finds himself in a gallery space and in a narrative reality at the same time. The king's presence becomes even more realistic, for the computer vision is capable of recognising what is happening in front of the projection, which it then adjusts accordingly - thus adapting to the viewer's behaviour: one should not look the king in the eyes and the installation does "not work" if we look at it, for the computer vision includes face detection. At this the double use of the computer vision is important for the installation. The first change is caused as the viewer enters the shot in the active field of the installation as a smudge. At this level it is merely a sensor, however as soon as the computer decides that it is looking at a human and not a faceless object, its vision becomes similar to human vision and thus the computer can play a role in the ideological field in front of the installation. *Presence* thus opens the complex issues on artificial intelligence that - with a high level of analysing activities - surpasses the mere recording of surveillance cameras and poses questions as to the ideological connotation of the supposedly unbiased view of the machine.

REFERENCES

[1] http://www.let.uu.nl/~Frank.Kessler/personal/notes%20on%20dispositif.PDF . (Apr 5th 2009)

[2] http://vvvv.org/ . (Apr 25th 2008)

[3] Open Computer Vision Library, http://sourceforge.net/projects/opencvlibrary/ . (Apr 25th 2008)

[4] M. Foucault, L'Archeologie du Savoir. Paris, Gallimard 1997.

[5] E. G. Craig, "The actor and the über-marionette", The Mask, 1908.

[6] W. Gropius, A.S. Wensinger, The Theater of the Bauhaus. Wesleyan, 1961.

[7] L. Manovich, The Language of New Media. MIT Press, 2001.

[8] http://ralyx.inria.fr/2007/Raweb/prima/uid0.html . (Apr 5th 2009)

[9] P. Viola and M. Jones. "Robust real-time object detection", International Journal of Computer Vision, 57(2), 2004, pp. 137-154.