Four Metamorphosis States in a Distributed Virtual (TV) Studio: Human, Cyborg, Avatar, and Bot – Markerless Tracking and Feedback for Realtime Animation Control

Jens Herder[(✉)], Jeff Daemen, Peter Haufs-Brusberg, and Isis Abdel Aziz

Department of Media, FH Düsseldorf, University of Applied Sciences,
Josef-Gockeln-Str. 9, 40474 Düsseldorf, Germany
herder@fh-duesseldorf.de
http://vsvr.medien.fh-duesseldorf.de

Abstract. The major challenge in virtual studio technology is the interaction between actors and virtual objects. Virtual studios differ from other virtual environments because there always exist two concurrent views: The view of the TV consumer and the view of the talent in front of the camera. This paper illustrates the interaction and feedback in front of the camera and compares different markerless person tracking systems, which are used for realtime animations. Entertaining animations are required, but sensors usually provide only a limited number of parameters. Additional information based on the context allows the generation of appealing animations, which might be partly prefabricated. As main example, we use a distributed live production in a virtual studio with two locally separated markerless tracking systems. The production was based on a fully tracked actor, cyborg (half actor, half graphics), avatar, and a bot. All participants could interact and throw a virtual disc. This setup is compared and mapped to Milgram's continuum and technical challenges are described in detail.

Keywords: Markerless tracking · Virtual studio · Avatars · Virtual characters · Interaction feedback

1 Introduction

While today's TV and video productions got used to the benefits of virtual studio technology, the interaction between actors and virtual objects inside a virtual world remains challenging. This contribution offers deep insights to actor tracking in virtual studios and their advantages for live productions. Different solutions as well as approaches with their strengths and weaknesses are compared

Fig. 1. The cyborg was created by partly overlaying the real, tracked actor with computer generated graphics within the green box.

to each other. Major functionalities were evaluated in exemplary productions at the Fachhochschule Düsseldorf, University of Applied Sciences.

The interaction control as well as multiple interfaces of a virtual set environment have been discussed and classified in [1]. An overview about the origin of virtual set environments and an introduction to virtual studios can be found in [2]. Virtual studio's developments were accompanied by experiments with distributed live productions using virtual studios and avatars all along [3]. The EU-funded Origami project addressed the interaction challenge within virtual sets by capturing the actor's volume and projecting feedback on a retro-reflective background [4].

Virtual studios allow the realtime combination of camera images and virtual elements, which brings advantages in flexibility and efficiency by using virtual scenery. Through keying techniques (e.g. chroma keying [5]) a separation of actors or objects and background can be accomplished. For this purpose a possible approach is the green- or blue-screen compositing technique, which requires a chroma keyer and a concolorous background. The separated background is replaced by a virtual environment, which must be rendered in realtime. This is done with the help of a high-performance computer system and special rendering software. Apart from the virtual background image, the render engine transmits a matte-out signal, the so-called external key signal, to the keyer. It allows displaying virtual

objects in front of the camera image. In order to provide realistic virtual camera perspectives and orientations in a virtual scenery, the camera's position and setup has to be tracked. The determination of that information, as well as the rendering of the virtual environment requires a small amount of time, by which the camera image has to be delayed. To ensure a smooth production flow, this offset must not exceed the length of 8 frames. Nowadays, most commercial virtual studios use a similar configuration as just described. An operator in the control room manages the virtual set. Newer approaches give the actor the possibility to control the set by using tablets, smartphones or computer-displays in the speaker's desk. More intuitive ways of control are gestures or a scene, which reacts on the actors' movements. This can be accomplished by tracking the actor's position or even his full skeleton. Because this should be done unobtrusive, only markerless tracking is suitable. In Sect. 2 different methods of markerless actor tracking are described. The determination of the actor's position has to be taken into account when setting the video delay. If the delay surpasses the value needed to accommodate the camera tracking determination, the camera tracking data has to be delayed additionally. This applies in reverse as well. Figure 2 describes the signal and data-flow in a virtual studio, using actor tracking. The render engine processes the tracking data, renders the background and external key signal, and transmits them to the chroma keyer. The delayed video signal and the virtual background get combined regarding the external key signal to display parts of the virtual image in front of the camera image.

2   Markerless Talent Tracking

Nowadays actor tracking is successfully used for movie and game productions, as well as for medical applications. The applied systems often require markers, which can have a disturbing influence in some situations. In virtual studio productions for example, the audience shall not notice, that the actor is tracked, so only markerless actor tracking is suitable. In medical applications, a reasonable compromise between precision and interference for the patients has to be found. Markerless tracking would be absolutely unpersuasive but less precise. Because of the significant accuracy improvements of markerless motion capturing systems, they are suitable for the study of the biomechanics of human movement like e.g. gait analysis [6]. There are different approaches to determine the position and orientation of people or even of their whole skeletons. Some systems use depth cameras, to some extent in combination with colour or monochrome image streams, to identify body parts and determine their spatial position. With the aid of a 2D Laser-scanner, the position of people's feet can be located. Other systems analyse the images of multiple cameras, which capture the tracking area from different perspectives. In the following Section, several tracking methods are described, followed by a comparison of some commercially available systems.

In virtual studios, actor tracking can be used for an automatic occlusion handling[1], interactions with virtual objects are possible and the motions of the

---

[1] Display virtual elements in front or behind the actor, dependent on her/his tracked position.
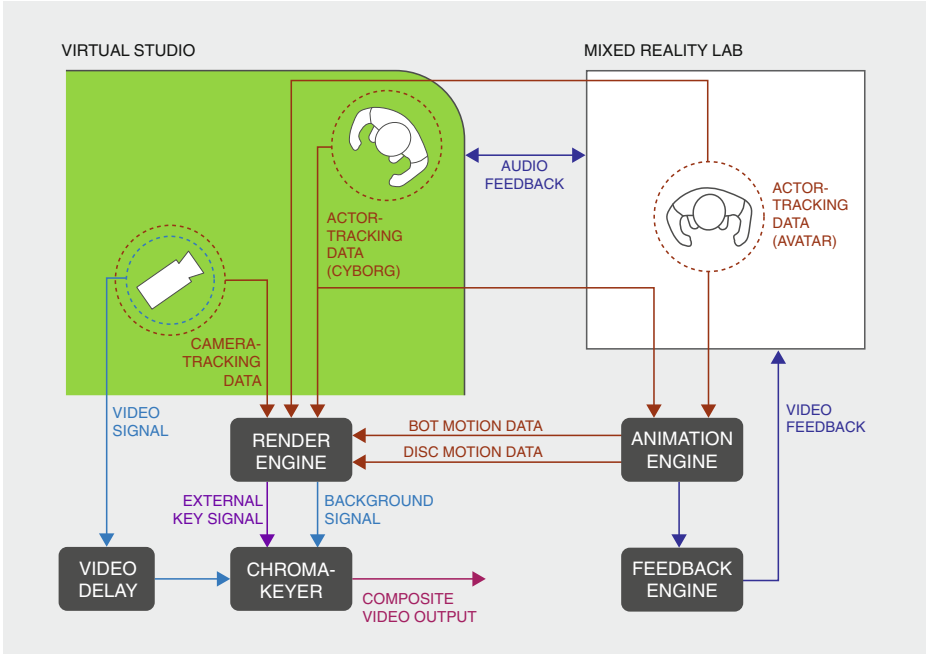
Fig. 2. System layout for the distributed live production with actor, avatar (controlled from separate locations), cyborg and bot.

tracked actors can control avatars. Mammoth Graphics and Kenziko developed an interactive control system for virtual studio sets for the broadcasting of the Olympic Games in London by *BBC* using the *Microsoft Kinect* [7,8]. The anchor was able to display a menu of virtual objects, navigate through them and make a choice just by gestures. Price et al. presented the Prometheus Project, where an MPEG-4 stream of a virtual 3D-production was transmitted [9]. An auxiliary camera, attached at the edge of the studio's ceiling identified the silhouette of the actor, which was used to determine the position of the actor's feet on the floor. In an example, the actor was mapped as a texture of a plane in the virtual set. The position of that plane was adapted to the position of the actor's feet. As a result, the occlusion was handled automatically.

Gibbs et al. described the idea of virtual actors in virtual studios and gave the example of Hugo, a German game show where a hobgoblin-like character was controlled by a human actor outside the set using an improvised cyber suit [10]. In the experimental production described in Sect. 3, markerless actor tracking was used to control different types of virtual characters.

Kim et al. [11] proposed a 3D system enabling natural-looking interactions of actors with synthetic environments and objects. A stereo camera was used to capture a 3D environment. The image and the spatial position of an actor were captured by a multiview camera. A realtime registration and rendering software processed and combined all information. Several examples of applications of the proposed approach were illustrated, e.g. direct interaction, enabled through

collision detection or automatic occlusion handling. The problems of missing visual feedback, as well as possible solutions like background-coloured props, vibrotactile display devices and monitors in the sight field of the actor were discussed. The main weakness of the system was that it did not allow any camera movement.

The amount of tracking information is decisive for the level of the interaction's complexity. By knowing the position of a whole person, simple interactions can be accomplished. To enable more precise interactions, at least the positions of the person's hands are required. At best, the tracking system should be able to determine every joint's position of the whole skeleton of a person.

2.1   2D Tracking

A simple way of 2D person tracking can be achieved by using a laser scanner, e.g. the *radarTOUCH* system [12]. An infrared laser beam in combination with a precisely timed rotating mirror creates an invisible plane. The system detects intersecting objects and determines their areal positions. This system was used by Marinos et al. to acquire the position of a person's feet. By means of that information, a virtual interactive set could be controlled. A person was able to open a virtual door by getting closer to it and a virtual display could be faded in [13]. *Orad Hi-Tec Systems* had announced a markerless actor tracking system named *X-PLORO*. This system has been developed by *Xync*, a GMD start-up company, which was bought by *Orad*. Two overhead cameras on the studio's ceiling or walls capture the actor in the scene. By analysing both image streams, the actor can be identified and located in the 2D area, which allows an automatic occlusion handling.

2.2   3D Tracking

Because 2D position information only allow very simple interactions, 3D tracking is much more attractive for the application in virtual studios. Different systems using depth cameras, as well as multi-camera systems, which capture the scene from several perspectives, are distinguished.

Depth-Based. The different depth cameras vary in the techniques to determine the depth map of their field of view. Some systems consist of infrared or visible light sources combined with sensors, whereas others require a specific illumination of the environment. The depth values, often in combination with the colour- or brightness information of additional sensors, have to be analysed to identify the person or its body parts, where then a depth value can be assigned to. Common methods[2] to acquire depth maps are the passive stereo technique (e.g., *Point*

---

[2] Comparison of 3D imaging technologies issued by Texas Instruments can be found on http://www.ti.com/ww/en/analog/3dtof/.

Grey Bumblebee), structured light coding (e.g., *Microsoft Kinect* 1) and the time of flight method (e.g., *Microsoft Kinect* 2). A middleware like *SoftKinetic iisu* allows tracking the full body of up to four people using the depth map. Systems like the Microsoft Kinect 1 or 2 allow skeleton tracking within their own SDK. Time-of-flight cameras were already used to gather the position of a whole person or even of some of its body parts in a virtual studio. Using the monochromatic image, which the time-of-flight camera provides additionally to the depth map, the distance between camera and actor could be determined by finding the 2D-position of the head in the monochromatic image and combining that information with the depth value. Automatic occlusion handling was possible with this technique [14]. Flasko et al. used an auxiliary HD-studio camera to track the position of the head and the hands of a person, to allow interaction in a virtual studio [15]. The skeleton information, acquired by a *Microsoft Kinect*, was used by Hough et al. [16] to handle the occlusion of virtual objects in a virtual studio. Because one position information for a whole person is not sufficient for advanced occlusion handling, especially the hand's positions were added.

Multiple Camera Images. Carranza et al. [17] describe an approach to reconstruct a 3D geometry of a person by means of camera images, captured with far distance between each other. In addition, the applied method allows determining the movements of multiple people's skeletons. This requires the camera's imaging properties, as well as their relative position to each other, to be known. This information can be established by a system calibration. By means of the silhouettes of the people in the tracking area, their individual visual hull can be determined, in which a skeleton model is fitted in. The *OpenStage* system by Organic Motion is the first commercially available tracking system, which uses multiple camera images to track people's motions. Table 1 shows a comparison of the properties of the *OpenStage* 2, *Microsoft Kinect* 1 and 2. The *Kinect* 2 delivers the largest number of additional parameters including biometric identity, activities, leaning, appearances, expressions, engagement, and heart rate.

Comparison of Systems. The different approaches shown in the previous Sections offer very different possibilities and features. All systems possess very different strengths and weaknesses depending on the lighting conditions. This means, after evaluation of the environment, an appropriate system has to be chosen. The lighting conditions do not only depend on the illumination of a virtual studio, for some systems, e.g. infrared based, the wavelength of light has to be considered as well. Sunlight for example might influence the quality of tracking.

Systems which capture the scene from only one perspective (e.g. *Microsoft Kinect* 1&2) do not allow a full 360° tracking. Motions, which cannot be seen by the sensor, e.g. body parts which occlude others, can be estimated to a certain extent, but mostly lead to tracking errors. Combining multiple sensors can solve

Table 1. Comparison of commercially available full body tracking systems.

	Organic Motion OpenStage 2	Microsoft Kinect 1	Microsoft Kinect 2
Method	Multiple camera images	Structured Light Coding	Time-of-Flight
Tracking volume	Up to approx. $100\,m^3$ $6\,\mathrm{m} \times 6\,\mathrm{m} \times 3\,\mathrm{m}$ for 18 camera system	Up to approx. $16\,m^3$ full body tracking: $1.8\,\mathrm{m}$–$4\,\mathrm{m}$ distance & $57\,^\circ \times 43\,^\circ$ field of view	Up to approx. $45\,m^3$ full body tracking: $1.8\,\mathrm{m}$–$4.5\,\mathrm{m}$ distance & $70\,^\circ \times 60\,^\circ$ field of view
Calibration effort	A few minutes	Pre-calibrated	Pre-calibrated
Sampling rate	60–120 fps	30 fps	30 fps
Latency	25 ms–50 ms	150 ms–250 ms	ca. 60 ms
Max. number of tracked people	5	2	6
Skeleton model	21 joints	20 joints	25 joints
Parameters:			
Head orientation	No	No	Yes
Foot orientation	Yes	No	Yes
Hand orientation	Limited	No	Yes
Hand state	No	No	Open/Closed/Lasso (Pointing with two fingers)
Additional parameters	No	Voice Recognition	Voice Recognition Person-related Parameters
References	[18, 19]	[18, 20]	[21]

this issue, but some sensors do not work perfectly when using them simultaneously. The *Microsoft Kinect* 1, which projects infrared patterns on the environment to determine the depth map, produces errors, when the pattern of a second *Kinect* interferes with the other one. On the contrary, there are no problems occurring when combining multiple *Microsoft Kinect* 2 sensors.

As shown in Table 1, the *OpenStage* system does not allow head and hand tracking. To enhance the *OpenStage* tracking data, it can be combined with the information of the head's orientation and the state of the hands, collected by the *Kinect* 2. The realisation of such an approach is discussed in Sect. 6.

2.3   Benefits of Markerless Actor Tracking

Markerless actor tracking offers a wide variety of new possibilities for virtual studios, medical applications, and virtual simulations. The key benefits of markerless actor tracking are the fast and easy operation. Compared to marker based tracking systems, no markers have to be attached to the actors. The actor is allowed to enter the tracking area and being recognised without any special preparations. In the context of virtual studios, marker based tracking systems

influence the actor's behaviour as well as the audience's suspension of disbelief. Markers might interrupt the illusion of plausibility, produced in the virtual studio. By controlling the virtual set by the motions of the actors, for example the applications mentioned in Sect. 3, interactions can appear more natural and realistic. Moreover, markerless actor tracking enables new kinds of interaction, which could lead to new TV formats.

2.4   Limitations

Besides many advantages of markerless tracking in comparison to marker-based approaches, there are still some limitations. Depending on the system and lighting conditions, the precision, as well as the reliability can be insufficient. During live broadcasting, accurate tracking has to be provided permanently. However, the stability of today's systems is not good enough to guarantee this. Most systems are limited to track humans, which is sufficient for most applications. Sometimes, the tracking of objects or animals could be advantageous. For experimental use, systems like the *OpenStage* 2 allow the definition of new kinds of skeleton models, like animals or simple objects (e.g. Sticks). Another crucial problem is the missing feedback in a virtual environment, which will be discussed in Sect. 5.

3   Use Cases – Experimental Production

The simple story is based on three players, cyborg, avatar, and bot, who throw a disc to each other. All players are located within a virtual arena. A judge oversees the game and is played by a human. Every character represents one of the four metamorphosis states: Human, Avatar, Bot, and Cyborg as seen in Fig. 3. The cyborg is shown in Fig. 1. A tracked actor is partly overlaid with graphics, assembling an armour. The generated mask from the render engine was slightly larger than the armour graphics to compensate alignment errors and tracking noise. The armour was mapped to the exact position and orientation of every joint provided by the markerless tracking system. A human skeleton divided in 21 different joints can precisely be covered with virtual elements. The avatar's motions are controlled via a remote *OpenStage* markerless motion capture system. The bot's motions are controlled by an animation engine (*Unity3D*).

   In a previous work [22], necessary software (plugin) was developed to receive and process *OpenStage*'s motion data within a commercial virtual studio renderer from Vizrt. The production took place at the FH Düsseldorf's virtual studio with an *OpenStage* system, using 18 cameras. The avatar was controlled by a second *OpenStage* system with 10 cameras in the separate Mixed Reality lab. The system layout and data flow is shown in Fig. 2. Beside the renderer, an animation engine (*Unity3D*) was used for controlling the bot and the disc's flight. The disc's target was chosen based on the situation e.g. the cyborg's or avatar's tracked hand or the wall/floor. This means that the disc is automatically catched by the avatar or cyborg, which ensures perfect story flow without special training for the talents
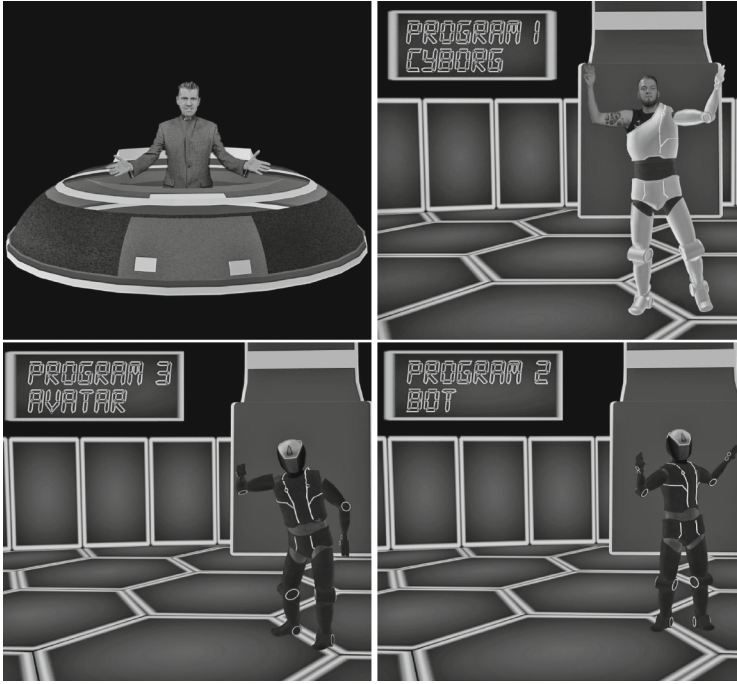
Fig. 3. Four metamorphosis states in a distributed virtual (TV) studio: Human, Cyborg, Avatar (controlled from a separate location), and Bot.

in the virtual environment. But still their movements had to match in time. The bot's arm was controlled with inverse kinematics to catch and throw the disc. Other behaviour of the bot was based on pre-recorded motions. For skeleton data Organic Motion's SDK was used. For other communications, messages using Open Sound Control (OSC) were applied.

The final video and some pictures from the production can be found online[3]. The tracking accuracy was good enough for wide shots.

4   Classification: Yet Another Continuum

The classification from Milgram et al., showing a continuum from reality to virtuality [23], requires another dimension for expressing the amount of control a user in our case actor has over avatars and virtual characters (see Fig. 4). Mixed control is necessary because input devices do not capture all parameters, required to animate an avatar. For example the current version of the *OpenStage* does not capture head or hand rotations. For a good animation those parameters have to be generated or captured by other means. In case of synthetic generated parameters, the instant control is reduced. If a reduced number of parameters is

---

[3] http://vsvr.medien.fh-duesseldorf.de/productions/vron2013/
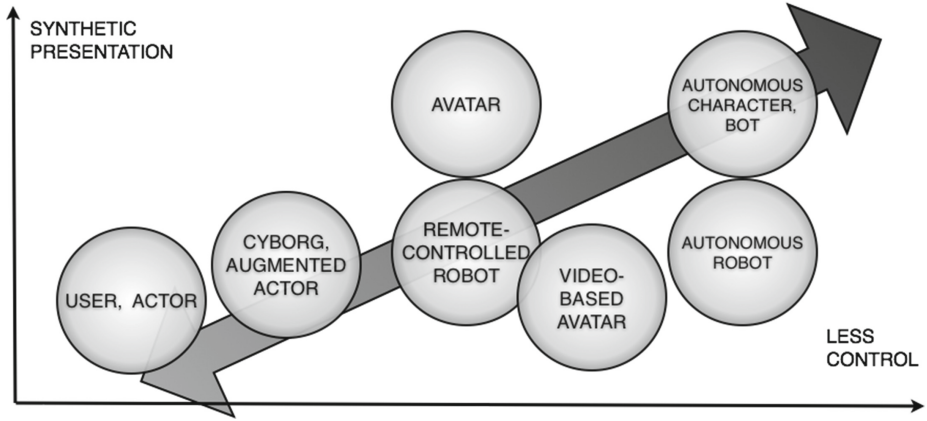
Fig. 4. User – Avatar – Virtual Character Continuum with different levels of control.

in use, the classification is deferred to the area of bots. For social interaction with systems, Holz et al. organised different incarnations (robots and social agents) in a continuum, which is also inspired by Milgram [24]. In principle, the bot in this article corresponds to the social agents. Also the listed mixed reality agents share the mixing of real and virtual images with the cyborg, but are not driven by a human.

4.1   Reduction and Expansion: Limited Data, but Extensive Animations

Organic objects, for instance humans, animals or plants are usually in constant motion. Complex characteristics – reduced to simple features – are hard to transfer to 3D animations in a natural and fluent way. Continuity is the key factor in this matter. Only then it is possible for the viewer to get the impression of realistic dynamics. Movements of animated beings are very complex processes. To be easily understood from computer programs they have to be reduced to a very basic type of data. This reduction can take different forms. In virtual studios for example recognising or tracking the body of a person can be interpreted by the render engine, as described earlier in Sect. 2, into a more or less rudimental skeleton which then can be used to control different kinds of animation in realtime. To give a rough example: 18 cameras record the moving actor, with an output of millions of pixels, which then form a huge volume with voxels, which are then converted into only 21 skeleton joints (as described in Sect. 2.2). These limited values may suggest that the animation itself has to be broken down to a very simple state, but this is not true. Even if the involved software receives only a very finite number of data, it can still be instructed to display a continuous and logical visualisation of the animation, so to speak expanding again after being reduced.

As mentioned in Sect. 2.2 the *Kinect* 2 is one of the commercially available full body tracking systems, but besides making the skeleton of a person available,
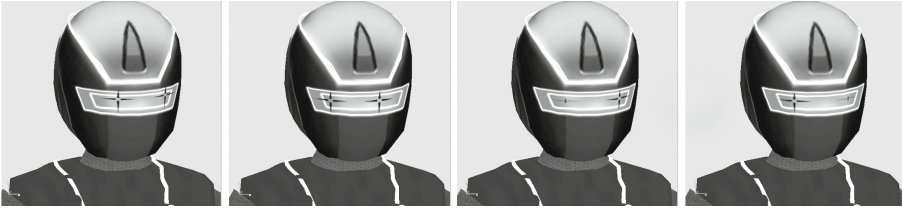
Fig. 5. Eye Animation: closing animation of left and right eye; idle eye animation, using time and context as parameters

other more distinguished parameters were added to the newer version of the sensor (see also "additional parameters" in Table 1). This includes recognition of eyelid movements, emotional expressions (e.g. neutral, happy), and appearances (e.g. wearing glasses). Activities are left eye closed, right eye closed, mouth open, mouth moved and looking away. To be more concrete, the camera is going to identify two conditions of the eyes. These simple and reduced commands can be integrated in the virtual environment in a more complex way than it may seem at first sight. Instead of just animating the shutting of the left and right eye (see Fig. 5) it is possible to extend this motion beyond the actual restricted data. This could mean, for example, letting the 3D character have a random look around throughout the area while in an idle state (see Fig. 5). So even if the sensor does not send out usable data, the avatar is able to behave autonomous.

One other new and very useful feature of the *Kinect* 2 is the implemented face tracking, which delivers an emotional state. This offers the chance to recognise human expressions, which can then be translated to a synthetic avatar with a smooth transition between different moods or can be used to trigger custom animations according to the mood of the person in front of the camera.

In certain cases a completely precise tracking is not always possible. To avoid visible inconsistencies in the animation it can be helpful to interpret variables providing information about the current tracking quality. If the sensors in the camera cannot operate correctly, the respective state needs to be estimated. In this case the value is not necessarily Boolean, but could be referred to as a maybe, an interference, or even a percentage figure for the accuracy. This can be used to prioritise different data sources inside the render engine and prefer better working sensors (see explanation for sensor fusion, Fig. 6) or using a default animation. Another important aspect regarding realistic interactions between different scene objects or the actor and the 3D animations is a context-based reaction. This means that the respective object responds in accordance with another, for example a real-life actor during a realtime recording. The control of the virtual character was put into practice through a remote rendering scenario, where actors were in charge of the movement and voice response of the 3D animation [10]. In the case of the animated bot (Fig. 5) a context related interaction could mean letting it follow another entering 3D object or person with both eyes depending on the object's movement.

If one wants to focus the user's attention on a specific object, e.g. an animated character, automatic light and camera control will become crucial. As described
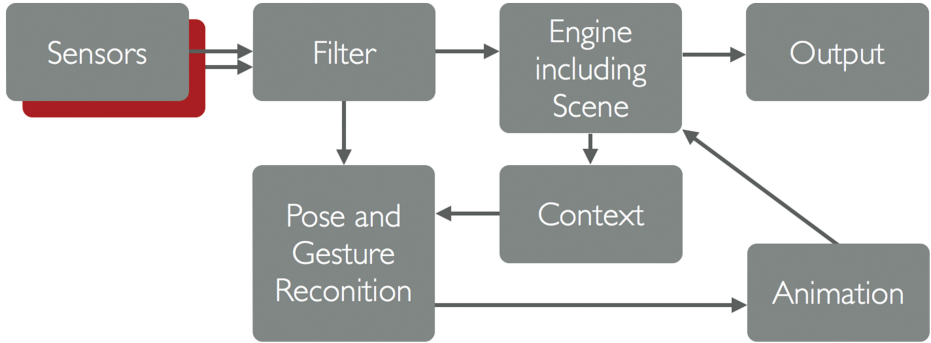
Fig. 6. Filtered sensor data and gesture/pose detection drive the animations, influenced by application context, which is mostly derived from the scene (graph).

by Herder et al. [25] the context, meaning the scene, can be used to trigger certain animations, which focus the user's attention on specific parts in the scene. In this scenario a server was integrated that captured relevant data from the scene and user and then decided, which object needed most of the viewers attention. In that case a lighting animation got this input from the server and highlighted the according objects.

Figure 6 gives an overview of the correlation between the actual animation, the scene context and the influence it has on the pose and gesture recognition using sensors. The illustration deliberately indicates that the virtual scene can be dependent on more than one sensor. The *Kinect*, for example, has not just a depth camera to capture the person's position, but can also acquire additional tracking data through the integrated microphone array. For sensor fusion, multiple sensors are combined for gaining better tracking results.

5   Talent Feedback

One major problem is the orientation inside the virtual studio. Virtual objects visible for the audience are invisible for the actor and do not allow reliable orientation. This problem is especially affecting precise actor tracking to determine interaction between the actor and virtual objects. This includes touch interactions with small virtual objects inside the virtual scenery. Very common techniques to provide some kind of feedback are displays and identification marks on the floor. These techniques provide a reliable orientation for the actor. Depending on the kind of production and virtual environment, these techniques for actor feedback have a serious disadvantage in influencing the actor's behaviour and bias. In the succeeding Sections, various kinds of feedback are described in detail and a connection to the experimental production is shown.

5.1   Visual Feedback

Any kind of visual feedback is only recommended, if the actor is not influenced in a noticeable manner or the production does not need the actor's real appearance.
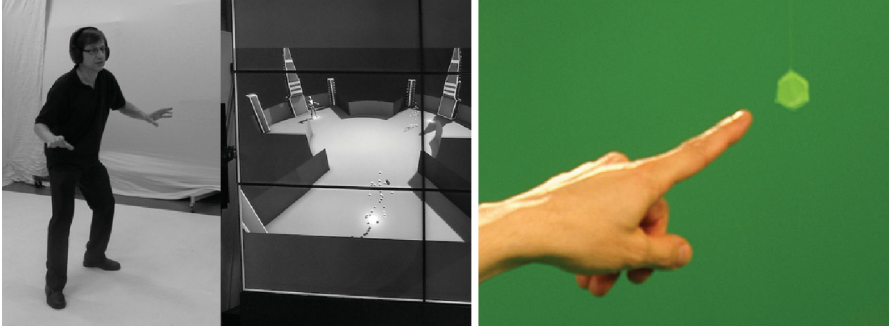
Fig. 7. Mixed reality lab with powerwall and headphone as feedback for the animator of the avatar (left & middle); Invisible props as visual feedback (right).

This for example means that actor tracking is only used to determine motions for virtual characters without any real world elements. A very reasonable approach is the utilisation of non-visible markers on the floor and mounted on transparent fishing lines, e.g. a green coin as shown in Fig. 7 (right). These invisible props are very efficient and less cost-intensive. In the experimental production mentioned in Sect. 3 visual feedback was an integral part and required to make a distributed live production possible. As a result of the different states (Human, Avatar, Bot, and Cyborg) very different kinds of visual feedback were in use. For example, the avatar was fully animated and only controlled by a tracked person. This offered the full attention of the person to a feedback-engine providing information about the actors positions and a robust scheduling for motions like throwing a virtual disc. The Feedback-Engine was based on a *Unity3D* scene providing detailed real time information about all tracked actors and their position in a virtual scene. This information was mirrored to a powerwall (see also Fig. 7, left).

Monitors and Projections in Green. Another quite common approach to provide some kind of feedback and orientation for the actor are displays placed in the studio. Those displays are not visible to the audience and show the mirrored camera output. The camera output can also be projected into the green or blue areas of the virtual studio. The advantage is that the audience does not recognise any special behaviour of the actor because of the projection is placed within the actor's natural field of view. In a production captured at the virtual studio of the FH Düsseldorf in 2012, a touchscreen was coated with a green semi-transparent tissue (see Fig. 8). The actor was able to identify the content on the touchscreen and could interact with it. Because of the keying process in the virtual studio, the screen was not visible in the final video output.

5.2   Vibrotactile Feedback

A waist-belt with vibrotactile elements can help to determine the orientation inside a virtual studio. Depending on the distance between the actor and specific virtual objects, signals can be sent to the vibrotactile elements to signalise the
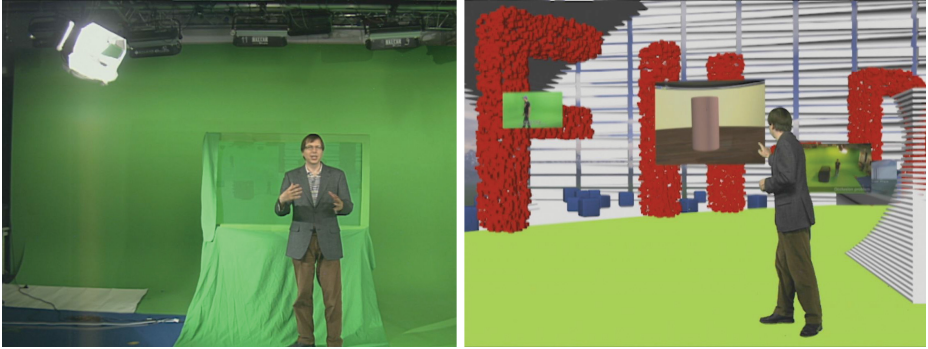
Fig. 8. Green monitor with infrared touch frame for interaction and feedback.

distance to the object. This technique can be used in very different contexts and provides solid information without any noticeable influence on the actor. This means that safe and exact interactions and motions can be performed. Different forms of vibrotactile feedback/patterns were evaluated by Vierjahn et al. [26,27].

5.3   Acoustic Feedback

In contrast to visual feedback, acoustic feedback is not in sight of the audience. Depending on the kind of production acoustic feedback can provide precise and inconspicuous help for the actor. Acoustic feedback can either be audible to all actors and participants in the virtual studio or it can be realised with not visible in-ear headphones exclusive for the actor. Acoustic feedback can provide simple feedback if the actor enters a certain area of the studio or touches a virtual object. For a more precise feedback, head-related transfer functions can be used to give directional and distance cues of virtual objects [28]. In the experimental production mentioned in Sect. 3, acoustic feedback was used to coordinate the actors placed in different locations. To harmonise closely spaced movements like throwing a disk between Human, Avatar, Bot, and Cyborg, acoustic feedback is of great help.

6   Conclusions and Future Development

The clear progress and acceptance of virtual studios in nowadays TV productions offer enormous potential for further development and new approaches. For a lot of TV productions an enhancement in interactivity enabled by precise markerless actor tracking could lead to new TV formats and shows. Although there are a lot of exemplary productions and approaches, there is still a lot of research and improvement needed to maximise its deployment. We showed new ways of interaction using markerless talent tracking in a virtual studio and combined it with a game engine for physics and inverse kinematics. Major issues like robust tracking and instant feedback still remain.

For an evaluation of markerless actor tracking for virtual (TV) studio applications [22], several experts experienced in the field of virtual studios were asked about their opinion. Most of the experts reported concerns connected to newscasts and recommended the far more possibilities of interactivity connected to game shows or alike. They mentioned subtle effects like raising dust or virtual footprints as reasonable applications of markerless actor tracking. The feedback considering the new type of actor composed of real and virtual elements was also quite positive.

The story in the production illustrates the user avatar virtual characters continuum diagram. Having all metamorphosis states in one distributed real-time system is challenging. The problems and solutions were addressed. Future development needs to focus on the actor's feedback in virtual environments. While head mounted displays might be used for actors controlling an avatar, this is not an option for actors in a green box. The use of virtual acoustics might be a solution [28].

As already mentioned in Sect. 4.1, the combination of different tracking systems can enhance the stability and quality of the tracking data, as well as the amount of available information. In order to be able to combine multiple data flows, which often follow different logics, a mutual way of transfer, as well as unified messages have to be defined. By respecting a predefined dictionary for the joints' names and the naming of additional information, the data of different systems can be processed within one framework (e.g. OscCalibrator[4] by Marinos).

References

1. Herder, J.: Interactive content creation with virtual set environments. J. 3D-Forum Soc. **15**(4), 53–56 (2001). Japan
2. Gibbs, S., Arapis, C., Breiteneder, C., Lalioti, V., Mostafawy, S., Speier, J.: Virtual studios: an overview. IEEE Multimedia **5**(1), 18–35 (1998)
3. Bunsen, O.: Verteilte Virtuelle TV-Produktion im Gigabit-Testbed West. Abschließender Bericht, Laboratory for Mixed Realities, Institut an der Kunsthochschule für Medien Köln, GMD Forschungszentrum Informationstechnik GmbH Institut für Medienkommunikation, February 2000

---

[4] https://github.com/fubyo/osccalibrator.

4. Grau, O., Pullen, T., Thomas, G.: A combined studio production system for 3-d capturing of live action and immersive actor feedback. IEEE Trans. Circuits Syst. Video Technol. **14**(3), 370–380 (2004)
5. Ray, A., Blinn, J.F.: Blue screen matting. In: SIGGRAPH'98, Conference Proceeding, pp. 259–268 (1996)
6. Corazza, S., Mndermann, L., Chaudhari, A., Demattio, T., Cobelli, C., Andriacchi, T.: A markerless motion capture system to study musculoskeletal biomechanics: visual hull and simulated annealing approachD. Ann. Biomed. Eng. **34**(6), 1019–1029 (2006)
7. Vizrt: Kenziko and mammoth graphics at IBC 2012: Kinetrak 2012. http://www.vizrt.com/news/newsgrid/35347/Kenziko_and_Mammoth_Graphics_at_IBC_2012
8. Mammoth Graphics and Kenziko Ltd.: Kinetrak (2014). http://www.kinetrak.tv
9. Price, M., Thomas, G.A.: 3d virtual production and delivery using mpeg-4. In: International Broadcasting Convention (IBC). IEEE (2000)
10. Gibbs, S., Baudisch, P.: Interaction in the virtual studio. In: SIGGRAPH Computer Graphics, vol. 30, pp. 29–32. ACM Press, New York, November 1996. ISSN:0097–8930
11. Kim, N., Woo, W., Kim, G., Park, C.M.: 3-d virtual studio for natural inter-"acting". IEEE Trans. Syst. Man Cybern. Part A: Syst. Hum. **36**(4), 758–773 (2006)
12. Wöldecke, B., Marinos, D., Pogscheba, P., Geiger, C., Herder, J., Schwirten, T.: radarTHEREMIN - creating musicalexpressionsin a virtual studio environment. In: Proceeding of ISVRI 2011 (International Symposium on VR Innovation), Singapore, pp. 345–346 (2009)
13. Marinos, D., Geiger, C., Herder, J.: Large-area moderator tracking and demonstrational configuration of position based interactions for virtual studios. In: 10th European Interactive TV Conference, Berlin (2012)
14. Herder, J., Wilke, M., Heimbach, J., Göbel, S., Marinos, D.: Simple actor tracking for virtual tv studios using a photonic mixing device. In: 12th International Conference on Human and Computer, Hamamatsu / Aizu-Wakamatsu / Düsseldorf, University of Aizu (2009)
15. Flasko, M., Pogscheba, P., Herder, J., Vonolfen, W.: Heterogeneous binocular camera-tracking in a virtual studio. In: 8. Workshop Virtuelle und Erweiterte RealitŁt der GI-Fachgruppe VR/AR, Wedel (2011)
16. Hough, G., Athwal, C., Williams, I.: Advanced occlusion handling for virtual studios. In: Lee, G., Howard, D., Kang, J.J., Skezak, D. (eds.) ICHIT 2012. LNCS, vol. 7425, pp. 287–294. Springer, Heidelberg (2012)
17. Carranza, J., Theobalt, C., Magnor, M.A., Seidel, H.P.: Free-viewpoint video of human actors. ACM Trans. Graph. **22**(3), 569–577 (2003)
18. Brooks, A., Czarowicz, A.: Markerless motion tracking: Ms kinect & organic motion openstage. In: 9th International Conference on Disability, Virtual Reality and Associated Technologies, vol. 9, ICDVRAT and The University of Reading, pp. 435–437 (2012)
19. Organic Motion Inc.: Openstage 2.0 technical overview, June 2014. http://www.organicmotion.com/openstage-2-0-technical-overview/
20. Livingston, M., Sebastian, J., Ai, Z., Decker, J.: Performance measurements for the microsoft kinect skeleton. In: Virtual Reality Short Papers and Posters (VRW), pp. 119–120. IEEE (2012)
21. Microsoft: Kinect for windows sdk documentation, July 2014

22. Daemen, J., Haufs-Brusberg, P., Herder, J.: Markerless actor tracking for virtual (tv) studio applications. In: International Joint Conference on Awareness Science and Technology & Ubi-Media Computing, iCAST 2013 & UMEDIA 2013. IEEE (2012)
23. Milgram, P., Takemura, H., Utsumi, A., Kishino, F.: Augmented reality: a class of displays on the reality-virtuality continuum. Proc. SPIE **2351**, 282–292 (1995)
24. Holz, T., Dragone, M., O'Hare, G.: Where robots and virtual agents meet. Int. J.Soc. Robot. **1**(1), 83–93 (2009)
25. Herder, J., Cohen, M.: Enhancing perspicuity of objects in virtual reality environments. In: CT'97 – Second International Cognitive Technology Conference on IEEE, pp. 228–237. IEEE Press, August 1997. ISBN 0-8186-8084-9
26. Vierjahn, T., Wöldecke, B., Geiger, C., Herder, J.: Improved direction signalization technique employing vibrotactile feedback. In: 11th Virtual Reality International Conference, VRIC'2009 (2009)
27. Wöldecke, B., Vierjahn, T., Flasko, M., Herder, J., Geiger, C.: Steering actors through a virtual set employing vibro-tactile feedback. In: TEI 2009: Proceedings of the 3rd International Conference on Tangible and Embedded Interaction, pp. 169–174. ACM, New York (2009)
28. Ludwig, P., Büchel, J., Herder, J., Vonolfen, W.: InEarGuide - a navigation and interaction feedback system using in ear headphones for virtual tv studio productions. In: 9. Workshop Virtuelle und Erweiterte Realität der GI-Fachgruppe VR/AR, Düsseldorf (2012)