An Interactive Visual Canon Platform

Mathias Funk and Christoph Bartneck

Abstract The canon is composition pattern with a long history and many forms. The concept of the canon has also been applied to experimental film making and Japanese television. We describe our work in progress on an Interactive Visual Canon Platform (IVCP) that empowers creators of visual canons to design their movements through rapid cycles of performance and evaluation. The IVCP system provides real time support for the actors; they can see the resulting canon from their movement while they still perform. We describe principle solutions and their implementation. The hardware has reached a stable status, while we are still optimizing the visual processing of the system. A first user test is planned to give us guidance for the improvement of the system.

1 Introduction

In music, a canon is a composition that uses with one or more imitations of that melody played after a certain duration [3]. The initial melody is called the leader, while the imitative melody is called the follower. The follower must be based on the leader by being either an exact replication or a transformation of the leader. Several types of canons are available, including simple canon, interval canon, crab canon, table canon, mensuration canon. “Row, Row, Row Your Boat” and “Frère Jacques” are two well-known examples of a simple canon.

Experimental film maker Norman McLaren introduced the concept of the canon to visual arts an received the Canadian Film Award in 1965 for his movie “Canon”

Mathias Funk
Department of Electrical Engineering, Eindhoven University of Technology, Netherlands, e-mail: m.funk@tue.nl

Christoph Bartneck
Department of Industrial Design, Eindhoven University of Technology, Netherlands, e-mail: c.bartneck@tue.nl

[2]. This movie can be considered a visual canon. The actor enters the screen to perform a certain movement in his role as the leader. After the completion of this "voice" he walks forward. A copy of himself, the "follower", enters the screen and repeats this movement while the leader introduces another movement. This process continues until four copies of the same actor, "voices", are present on the screen (see Figure 1). McLaren continues with variations of this canon by introducing transformations, such as mirroring. Instead of walking on the stage, the "mirror" voice now walks on the ceiling. Moreover, he introduced causal relationships between the voices. One voice, for example, might kick a second voice. This is particularly interesting if the two voices are walking backwards. Then, the relationships are introduced also in the reversed time sequence. First the leader performs the reception of the kick before he moves on to perform the actual kick. The voices can also have a spatial relationship. While one voice bends down, a second may swing its arms.

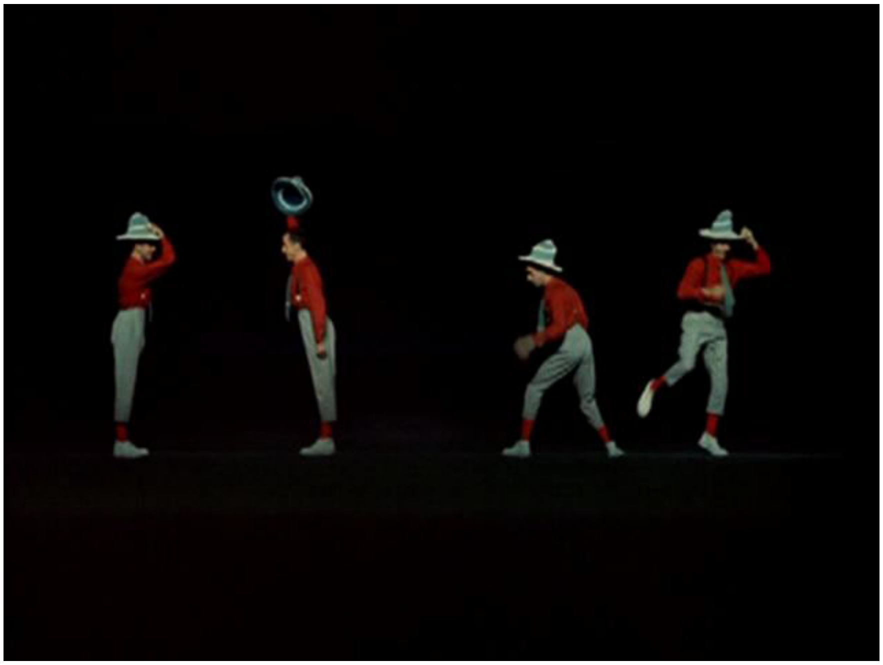

Fig. 1 Canon by Norman McLaren.

More recently, the comedian duo Itsumo Kokokara, consisting of Kazunari Yamada and Hidenori Kikuchi, perform a visual canon in as part of the daily show "Pythagoras Switch" on the National Japanese television channel NHK. Similar to McLarens movie, followers imitate the movements of the leader. The Algorithm March is constraint to the format of a simple canon in which no variations or transformations are being introduced by the followers. The duo named their visual canon "Algorithm March" and invite different groups of followers into their dance performance. Prominent followers include the NHK news team, the Japanese Polar Research team, Sonys Qrio robots, and many more.

The creating of new visual canon is very difficult, since a whole team needs to study the specific new movements. Only when all actors perform the canon correctly, the design of it as a whole can be evaluated. The long duration of each iterative cycle is a major obstacle in the design process of visual canons. In this paper we will describe our work in progress on an interactive visual canon platform (IVCP) that supports the creation of visual canons and that can be used to extend them in real time.

2 Requirements

The IVCP requires a large screen on which the design can be displayed. Naturally, the actors should be shown in life size. The screen can also be used for the capturing of the leading actor. The Algorithm March uses eight voices and given an average step distance of around 75 centimeters, a total of 6 meter is necessary to complete the complete cycle. The screen should therefore be at least 2 meters high and 6 meters wide.

Besides enabling actors to design and train new visual canons, the IVCP should also enable a broader audience to experience a visual canon. Art festivals and exhibitions are a ideal events for this and hence it is desirable to have a portable system. It follows, that the IVCPs structure must support itself. It should not have to rely on the availability of walls to which it could be mounted. It should also be possible to quickly assemble and disassemble the IVCP and the individual components should not exceed a size that would make its transportation difficult. Shipping a six-meter tube, for example, is very unpractical.

Needless to say that the IVCP must operate in real time. Simply recording an actor and afterwards playing multiple copies of the recording with a time delay would be insufficient. The actor needs to be able to interact with the followers immediately. Only a direct interaction would allow the actor to quickly experiment with movements and gestures. If, for example, the actor likes to take a swing at the follower, then he/she must know where exactly the follower is (see figure 2).

It is also desirable if the IVCP can be controlled without the any additional input devices. Sonys Eye Toy game, for example, allowed the players to control the game with gestures. This gesture control can be the starting point for the IVCPs gesture control.

3 Solution

The main characteristic of a canon is the structural consistency: only the leader has freedom to perform; the follower(s) must conform to the leader movements with the

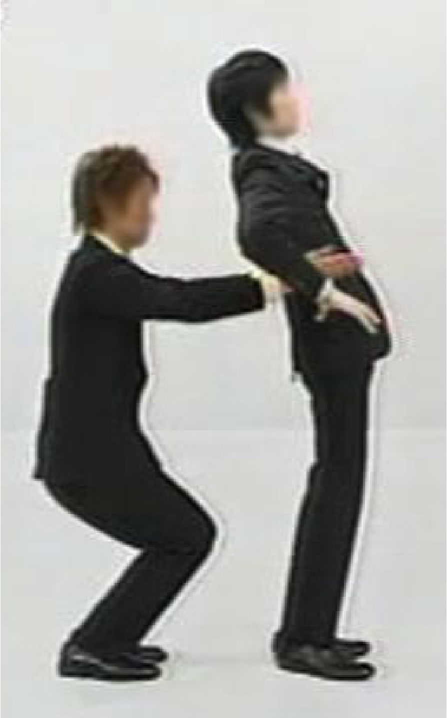

Fig. 2 Interdependency between leader and follower that can only be achieved through a real-time system.

utmost precision and discipline. Also, when looking at a canon in its performance, one will realize that each movement stays basically at the same horizontal position. It is only performed by all participants (voices) of the canon one after the other. The algorithm in the solution software takes advantage of the strict algorithmic nature of the canon. The only degree of freedom in the actor's performance lies in the movements of the leader which are constraint to the forward direction. One could even say that this motion is strictly monotonous in the mathematical sense: once a direction is chosen, the leader must move forward in that direction or otherwise risk that the followers overlap with him/her.

The solution algorithm includes several delay units that record a short video of the leader's movements from the camera, wait a certain amount of time, and project the video back onto the wall. During the waiting period the leader has moved forward, so he/she does not interfere with the playback. Next, the first playback unit is fed into a second delay unit, which records, waits and plays the video of the first delay unit. This results in two followers appearing on the screen. This chain of feedback units can be extended for as many followers as needed. However, this simple approach works only if the playback is projected back onto the wall only after the leader has left the scene. Otherwise the playback interferes with the recording of the video. Not only the leader, but also the followers would also be captured and spawn their voices. These voices of voices populate the screen. They may overlap and cause even more mutilated voices. Soon the output is an abstract beautiful color cloud, a chaos which contradicts the strict minimalism of the canon.

It becomes clear, that the major challenge is to control the feedback. The followers shall only be based on the leader. Several approaches exist to find the original (human) leader by masking the rest of the recorded picture, including the all projected followers. This way, the infinite feedback loops can be broken and the system works as intended.

The first practical solution concept we tried was to subtract the projected video from the captured video. The resulting difference picture would not contain the newly added followers, but only the elements which were initially not projected: the leader. Since, we already know what to project (the followers), it should have been possible to use it ask a mask to isolate the leaders. At first sight, this solution appears elegant and simple. However, aligning the recorded video with the projected video turned out to be very difficult. First, the known video that is to be projected is not the same as what is in the end shown on the wall. The projector itself has got optical properties and so does the camera that records this video. The camera could never be placed in the exact same position as the projector which would always result in slight optical discrepancies. Second, the two videos need not only be aligned in space, but also in time. The projection and recording process takes a certain amount of time. The video that arrives back into the computer is slightly delayed from its original. The geometric and temporal alignment of the projected with the recorded video proofed to be difficult to control, which resulted in unreliable results. Thus, a more robust solution was needed as the system should be usable in all kinds of different demonstration environments.

Another approach would be to place an additional camera behind the projection screen. This second camera would only record the shadow of the leader, because he/she is the only body that can possibly cast a shadow on the screen. An alternative to this idea would have been to place an infra-red camera next to the projector. The infra-red camera could isolate the leader by its heat signature. Both approaches would require additional cameras that need to be calibrated in space and time with the original setup. Adding these components would increase the complexity of the system and introduce additional sources of errors that would have a negative impact on the system's reliability.

A solution that works with only one camera and projectors is preferable. One option would be to use visual cues. The bounding boxes of each follower, for example, could be described with visual markers, such as cropping crosses, that are easy to recognize by the image processing algorithm. The followers could then easily be masked out from the recorded video. Since the visual cues would be added in the previous rendering iteration, they are part of the projected picture and thus perfectly aligned to the followers. However, visual cues may distract the observer and would have a negative impact on the overall aesthetics. They would not only clutter the screen, but they would also take the attention of the observer away from the dancing performance. Instead of enjoying the performance, the observer might focus on the artifacts of the technical implementation.

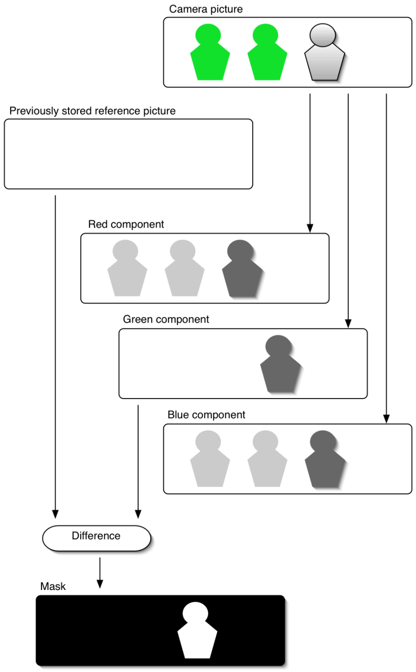

Fig. 3 Color separation for masking of the leader.

For our final solution, we took the concept of visual cues to the next level. We integrated the visual cues into the underlying concept of the performance: the first follower projected by the system is filled with a light green color that is easily recognizable by the image processing system. It become unnecessary to try hiding the markers since they become visually appealing and part of the performance. This principle reduces the followers - in case should the leader looks back - to moving shadows. One by one the projected green shapes will fade to back to recorded picture of the leader providing a natural and smooth transition from bare "shadows" to real followers.

What comes to mind first is the chroma key technique [1] commonly used in the movie industry. Light-green items that are filmed, including the background, can be replaced with other film material in the post-processing phase. The movie "300", for example, extensively used this method. All the actors were filmed against a light green background, which was replaced with a computer rendering of the ancient Greece in the post production. Light-green is one of the colors that least resemble the human skin tone and therefore can be substitued without causing an alienating look of the person on screen. Hence, using this "green-screen" technique, the system could mask the area behind the leader and avoid the notorious feedback cycle.

Due to possible physical limitations of the demonstration room, additional lighting or properties of the projector, the chroma key technique would not be robust enough. Instead we exploit basically the color separation of the RGB video. The background of the demonstration is pure white and the green color of the followers is so light that it is very close to white if seen only in the green component of the RGB video. Figure 3 shows this principle: the original captured picture contains the leader and two light-green followers. The color separation shows small traces of the followers in the red and blue component, because the first projected and then recorded green might be not as pure as intended due to the factors explained above. This does not matter as long as the green is light enough to appear as white in the green color component. In this case the followers merge with the white background of the screen and only the shape of the leader remains. To build a mask from this picture it is necessary to substract a white reference picture from the green channel. This reference can be captured once before the demonstration. The complete processing chain consists of first a masking component that uses visual cues the mask all the screen except for the picture of the leader. The resulting image which is most black is fed into several delay units that "clone" the followers. All cloned pictures are then combined and visual cues are set accordingly. Finally the picture is projected on the screen.

4 Realization

Given the mobility requirement for the system and the solution algorithm described above, the IVCP can be realized with a single camera and a single projector. This eliminates the necessity of aligning and synchronizing multiple screens and cameras. Otherwise, this difficult and lengthy procedure would have to be repeated every time the system is being moved. Full HD cameras and projectors have recently entered the market at a reasonable price. They have a resolution of 1920 x 1080 pixel, which means that they have a 16:9 aspect ration [4]. The projection screen should be optimized for this proportion. Given this proportion, the screen would have to be 3.37 meters high to achieve the required 6 meter widths. However, even tall human rarely exceed 2 meters of heights. We therefore decided to exclude one third of the vertical dimension, which resulted in a final dimension for the screen of 592 cm x 222 cm. It follows that the projection will use 1920 x 720 pixel. The Optoma HD80 projector we used had to be placed at a distance of 14 from the screen to achieve this projection size.

The image projection and capturing hardware is connected to a computer running Mac OS X, and Cycling74's graphical development environment for music and multimedia, Max/MSP, together with the video processing sub system Jitter. This provided a convenient basis to rapid prototype the IVCP algorithm and also makes the system easily extendable with additional components for gestures or audio control.

Figure 5 shows an essential part of the system architecture. In order to give a proper overview and to hide over-complex parts, several details of the algorithm have been omitted. The camera provides a picture of the screen which is fed into three different blocks: first into a reference picture block that simply stores an image of the empty screen before the demonstration begins. Second, the image is routed

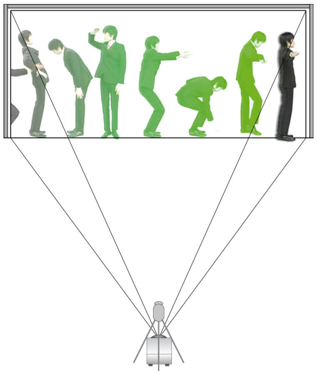

Fig. 4 Model of the IVCP setup.

into a component which extracts the green channel from the RGB video stream. As described in the solution section above a difference picture is created and after some further smoothing blended with the live image, the third block, using the alpha channel. This subsystem groups all masking functionality.

The isolated image of the leader must be cloned to obtain followers. This is the purpose of the next processing stage: the video is fed into a delay component which simply stores the data and outputs after a certain predefined amount of time. As explained in the solution section this process of storing and delayed playback is repeated several times. The output of each delay unit is a layer containing one follower's picture. In order to project the followers together on the screen, those layers must be composed.

5 Conclusions

The hardware of the system has been build and tested. Though, the biggest challenge was to find a large enough room to setup the system. It turned out that a close proximity of projector and camera is not only important for image alignment, but also to minimize the amount of shadow captured: as a most white image is projected any 3-dimensional object in front of the white screen casts a strong shadow. As long as projector and camera are horizontally close this does not matter much, but more distance will result in a bigger shadow captured by the camera which in turn is hard to remove by image processing.

Next steps will cover first improvements of performance, visual appearance and general robustness of the system. Capturing and processing a Full HD video in real time pushes currently available personal computer power to its limits. Especially for user testing the system should be capable of stable activity over time. A user test will confront participants who already know about the canon as well as participants who will just experience the system as it is. Additionally, is crucial to keep the

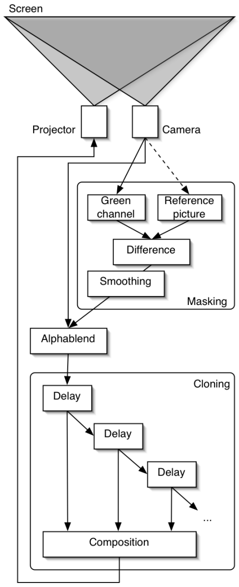

Fig. 5 System architecture.

image processing core modular and reusable, something which also concerns the robustness of the system architecture.

Future work will deal with integration of audio control, music, and a comprehensive gesture interface. Most importantly the outcome of the user test will be further requirements for system improvements. Later on we might think of more sophisticated visual effects or special guidance for first-time users. Moreover, game-like attributes can be added to the system, e.g. different levels of canon complexity, obstacles, moving object that must be used in dances and the like.

Generally the approach taken looks promising and could probably lead to a new kind of entertainment experience which not only encourages full body interaction, but also supports the development of mental skills as well as body control - and especially the connection thereof.

References

1. Jack, K. (2004). Video demystified: a handbook for the digital engineer (4th ed.). Oxford Burlington, MA: Newnes.
2. McLaren, N. (Writer) (2006). Norman McLaren: The Masters Edition: Homevision.
3. Norden, H. (1982). The Technique of Canon: Branden Books.
4. Weise, M. and Weynand, D. (2004). How Video Works: Focal Press Boston.