An Interactive Visual Canon Platform

Mathias Funk and Christoph Bartneck

Abstract The canon is a composition pattern with a long history and many forms. The concept of the canon has also been applied to experimental film making and on Japanese television. We describe our work-in-progress on an Interactive Visual Canon Platform (IVCP) that enables 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 canon resulting from their movements while they are still performing. We describe some possible approaches to a solution, and reasons for choosing the approach that we have implemented. The hardware has reached a stable state, but we are still optimizing the visual processing of the system. A first user test is planned to provide us with information for improving the system.

1 Introduction

In music, a canon is a composition employing one or more repetitions of a melody, played with a certain delay [3]. The initial melody is called the leader, and 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 canon are possible, including the simple canon, the interval canon, the crab canon, the table canon, and the mensuration canon. Two well-known examples of a simple canon are “Row, Row, Row Your Boat” and “Frère Jacques”.

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

Fig. 1 Canon by Norman McLaren.

Experimental film maker Norman McLaren introduced the concept of the canon to the visual arts and, in 1965, received the Canadian Film Award for his movie "Canon" [2]. This movie can be considered a visual canon. The actor enters the picture 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 picture 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 introduces 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 also introduced in the reversed time sequence. First the leader performs the receiving 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.

More recently, the comedy duo Itsumo Kokokara, consisting of Kazunari Yamada and Hidenori Kikuchi, has included the performance of a visual canon as part of the daily show "Pythagoras Switch" on the Japanese National television channel NHK. As in McLarens movie, followers imitated the movements of the leader. The duo named their visual canon "Algorithm March". It is constrained to the for-

mat of a simple canon in which no variations or transformations are introduced by the followers. They invited different groups of followers into their dance performance. Prominent followers have included the NHK news team, the Japanese Polar Research team, Sonys Qrio robots, and many more.

The creation of a new visual canon is very difficult, since a whole team needs to study the specific sequence of new movements. Only when all actors perform the canon correctly, can the design of it as a whole be evaluated. The long duration of each iterative cycle is a major obstacle in the design process of visual canons. In this paper we shall 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.

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

2 Requirements

The IVCP requires a large screen on which the design can be displayed. Naturally, the followers have to be shown life-size (to match the leader - the leader can also be shown on the screen). The Algorithm March uses eight voices, and with an average

step distance of around 75 centimeters, a total of 6 meters is needed to display the whole cycle. The screen should therefore be at least 2 meters high and 6 meters wide.

Besides enabling actors to design, learn, and practice new visual canons, the IVCP should also enable a broader audience to experience a visual canon. Arts festivals and exhibitions are ideal events for this, so it is desirable to have a portable system. It follows that the IVCP must be a free-standing structure. It should not have to rely on the availability of walls on which it could be mounted. It should also be possible to assemble and disassemble the IVCP quickly, and the individual components should not be of a size that would make transportation difficult. For example, shipping a six-meter tube would be very impractical.

Needless to say, the IVCP must operate in real time. Simply recording an actor and 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 experiment quickly with movements and gestures. If, for example, the actor wants to take a swing at the follower, then he/she must know exactly where the follower is (see Figure 2). It is also desirable that the IVCP can be controlled without any additional input devices. For example, in Sonys Eye Toy game, the players can control the game with gestures. This gesture control can be the starting point for the IVCPs gesture control.

3 Solution

The key characteristic of a canon is its structural consistency; only the leader has the freedom to perform, the follower(s) must conform to the leader movements with the utmost precision and discipline. Also, when watching the performance of a canon, one realizes that, in principle, each movement is performed at the same position by all the actors (voices) of the canon, one after the other. The algorithm in the solution software takes advantage of the strict algorithmic nature of the canon. In practice, the only degree of freedom in the actors performance lies in the movements of the leader. These are restricted to the forward direction to avoid risk of overlap with the followers.

The solution algorithm includes several delay units that record a short video of the leaders movements, wait a certain amount of time, and project the video back onto the screen. 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 not projected back onto the screen before 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 be captured and create their own followers (voices). These voices of voices would populate the screen. They might

overlap and cause even more distorted voices. Soon the output is a beautiful, abstract cloud of color, a chaos that contradicts the strict minimalism of the canon. It becomes clear that the major challenge is to control the feedback. The followers must be based only on the leader. There are several possible approaches to singling out the original (human) leader by masking out the rest of the recorded picture, including all the projected followers, thus breaking the infinite feedback loops and enabling the system to work as intended.

The first approach 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 leader. Since, we already know what to project (the followers), it should have been possible to use this picture as a mask to isolate the leader. At first sight, this solution appears simple and elegant. However, it turned out to be very difficult to align the recorded video with the projected video. In the first place, the known video that is to be projected is not the same as what is ultimately shown on the screen because of the optical properties of the camera that records this video and the projector itself. The camera can never be placed in the exact same position as the projector, and this will always result in slight optical discrepancies. Secondly, the two videos must be aligned not only in space but also in time. The projection and recording process takes a certain amount of time. The video that is fed back into the computer is slightly delayed relative to its original. The geometric and temporal alignment of the projected image with the recorded video proved to be difficult to control, which led to unreliable results. Therefore a more robust solution was needed, since 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 record only 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 could have been to place an infra-red camera next to the projector. The infra-red camera would isolate the leader using his/her heat signature. Both approaches would require additional cameras that would need to be calibrated in space and time against 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 systems reliability.

A solution that works with only one camera and projector is preferable. One option would be to use visual cues. For example, the position of each follower could be indicated with visual markers, such as cropping crosses, that can be recognized easily by the image processing algorithm. It would then be easy to mask out the followers from the recorded video. Since the visual cues would be added in the rendering of the previous iteration, they are part of the projected picture and therefore perfectly aligned with 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 distract the attention of the observer from the performance. Instead of enjoying the performance, the observer might focus on the artifacts of the technical implementation.

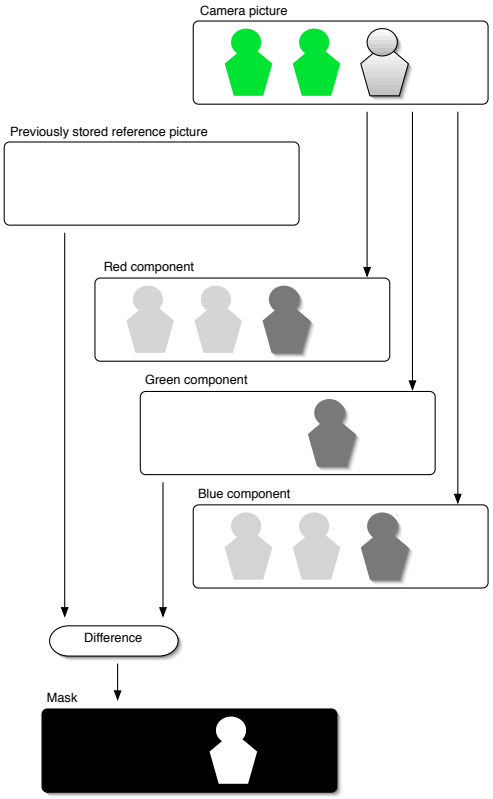

Fig. 3 Color separation for masking the leader.

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 displayed in a light green color that is easily recognizable by the image processing system. It then becomes unnecessary to hide the markers since they are visually appealing and become part of the performance. This principle means that if the leader looks back he/she will see only moving "shadows". One by one the projected green shapes revert to the recorded picture of the leader, providing a natural and smooth transition from bare "shadows" to real followers.

What first comes to mind is the chroma key technique [1], commonly used in the movie industry. Light green items, including the background, can be substituted by other film material in the post-processing phase. For example, the movie "300" used this method extensively. All the actors were filmed against a light green background that was replaced, in the post production, with a computer rendering of ancient Greece. Light green is one of the colors that least resembles human skin tone, and can therefore be substituted without making the person on screen look like an alien. Using this "green-screen" technique, the system could mask the area behind the leader and avoid the unacceptable feedback cycle mentioned earlier. The chroma key technique would not be sufficiently robust to withstand changes in the

physical environment such as the size of the demonstration room, the lighting, or the properties of the projector. Instead we basically exploit the color separation of the RGB video. The leader performs in front of a pure white background, 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 illustrates this principle, showing a moment at which the camera sees the leader and two light-green followers. The green component of the color separation shows only the leader, while the red and blue components show small traces of the followers, because the green first projected and then recorded might not be 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, so that 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 subtract a white reference picture from the green channel. This reference picture can be captured before the performance. The complete processing chain is as follows. First a masking component uses visual cues to mask the whole screen except for the picture of the leader. The resulting image, which is mostly black, is then 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 onto the screen.

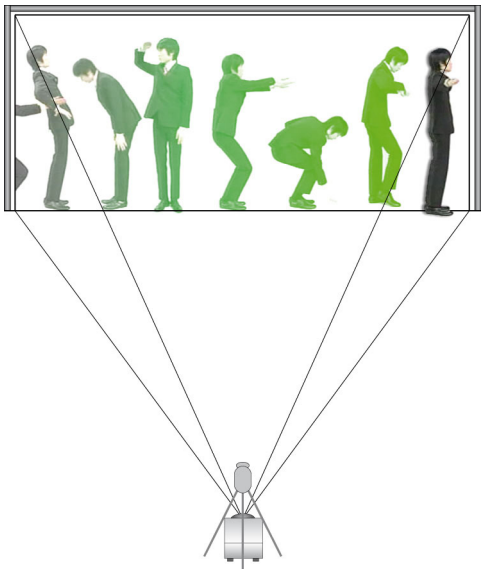

Fig. 4 Model of the IVCP setup.

4 Realization

Given the portability requirement for the system, and the solution algorithm described above, the IVCP can be implemented with a single camera and a single projector (see Figure 4). This eliminates the need to align and synchronize multiple screens and cameras. Otherwise, this difficult and lengthy procedure would have to be repeated every time the system was moved. Full HD cameras and projectors have recently entered the market at a reasonable price. They have a resolution of 1920 x 1080 pixels, which means that they have a 16:9 aspect ratio [4]. The projection screen should be optimized for this aspect ratio, meaning a height of 3.37 meters to achieve the required 6 meter width. However, even tall people rarely exceed 2 meters in height. We therefore decided to exclude one third of the vertical dimension, which resulted in final dimensions for the screen of 592 cm x 222 cm. It follows that the projection will use 1920 x 720 pixels. The Optoma HD80 projector that we used had to be placed at a distance of 14 meters from the screen to achieve this projection size.

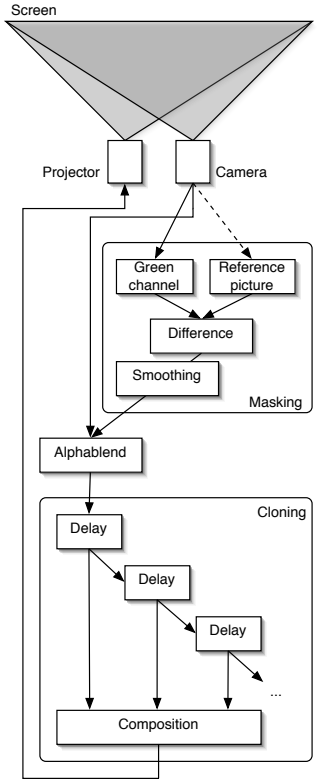

Fig. 5 System architecture.

The hardware for image projection and capture is connected to a computer running Mac OS X and Max/MSP, and Cycling74s graphical development environment for music and multimedia, together with the video processing sub system Jitter. This provides a convenient basis for rapid prototyping of the IVCP algorithm, and also results in a system that is easy to extend with additional components for gesture 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. The first block simply stores a reference picture of the empty screen taken before the start of the performance. The second block extracts the green channel from the RGB video stream.

As described in section 3, Solution, a difference picture is created and, after some further smoothing, is blended with the live image in the third block, using the alpha channel. This subsystem incorporates 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 that simply stores the data and outputs it 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 picture. In order to project the followers together on the screen, these layers must be merged.

5 Conclusions

The hardware of the system has been built and tested, but the biggest challenge was to find a room large enough to set up the system. It turned out that positioning the projector and camera at the same height is not only important for image alignment, but also for minimizing the amount of shadow captured. (Since a mostly white image is projected, any three?dimensional object in front of the white screen casts a strong shadow.) As long as projector and camera are at the same height this does not matter much, but increased vertical separation will result in a bigger shadow being captured by the camera, which is hard to remove using image processing.

Our next steps will be to provide improvements in the performance, visual appearance, and general robustness of the system. The capture and processing of a Full HD video in real time pushes the processing power of currently available personal computers to its limits, but, especially for user testing, the system should be capable of sustained stable activity. Participants facing a user test will include those who already know about the canon as well as those who are encountering the concept for the first time. Additionally, it is crucial to keep the image processing core modular and reusable, something which also affects the robustness of the system architecture. Future work will deal with integrating audio control, music, and a comprehensive gesture interface. Most importantly the user test will produce further requirements for system improvements. Later on we might think of more sophisticated visual ef-

fects 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, and moving objects that must be used in the performance.

Generally, the approach taken looks promising and could probably lead to a new kind of entertainment experience that would not only encourage full body interaction, but also support the development of mental skills as well as body control and especially the connection between the two.

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.