

# PoCoMo: Projected Collaboration using Mobile devices

**Blind Review**  
affiliation  
address  
[email](#)

**Blind Review**  
affiliation  
address  
[email](#)

**Blind Review**  
affiliation  
address  
[email](#)

## ABSTRACT

We present a novel system and method that enables playful social interactions between multiple projected characters. The prototype consists of two mobile projector-camera systems, with lightly modified existing hardware, and computer vision algorithms to support a selection of applications and example scenarios. Our system allows participants to discover the characteristics and behaviors of other characters projected in the environment. The characters are guided by hand movements, and can respond to objects and other characters, to simulate a mixed reality of life-like entities.

## Author Keywords

mobile projection, collaborative play, social interaction, mixed-reality characters, animation, augmented reality

## ACM Classification Keywords

I.5.4 Applications: Computer vision, I.2.10 Vision and Scene Understanding: Video analysis, K.8 Personal Computing: Games, H.5.3 Group and Organization Interfaces: Synchronous Interaction, H.5.2 User Interfaces: Input Devices and Strategies.

## General Terms

Algorithms, Design, Experimentation, Human Factors

## INTRODUCTION

The use of mobile devices for social and playful interactions is a rapidly growing field. Mobile phones now provide the high-performance computing power that is needed to support high-resolution animation via embedded graphics processors, and an always-on Internet connectivity via 3G cellular networks. Further advances in mobile projection hardware make possible the exploration of more seamless social interactions between co-located users with mobile phones. PoCoMo is a system developed in our lab to begin prototyping social scenarios enabled by collaborative mobile projections in the same physical environment.

PoCoMo is a set of mobile, handheld projector-camera systems, embedded in fully functional mobile phones. The proposed system's goal is to enable users to interact with one another and their environment, in a natural, ad-hoc way. The phones display projected characters that align with the view of the camera on the phone, which is running

software that tracks the position of the characters. Two characters projected in the same area will quickly respond to one another in context of the features of the local environment.

PoCoMo is implemented on commercially available hardware and standardized software libraries to increase the likelihood of implementation by developers and to provide example scenarios, which advocate for projector-enabled mobile phones. Multiple phone manufacturers have experimental models with projection capabilities, but to our knowledge none of them have placed the camera in line of sight of the projection area. We expect models that support this capability for AR purposes to become available in the



**Figure 1. (a) Users hold micro projector-camera devices to project animated characters on the wall. (b) The characters recognize and interact with one another.**

near future.

PoCoMo makes use of standard methods in computer vision to sense the presence of other projected characters, and extract visual features from the environment as part of a game. The algorithm is lightweight, enabling operation in a limited resource environment, without a great decrease in usability.

Extending mobile phones to externalize display information in the local environment and respond to other players creates new possibilities for mixed reality interactions that are relatively unexplored. We focus in particular on the social scenarios of acknowledgment, relating, and exchange in games enabled between two co-located users. This paper describes the design of the system, implementation details, and our initial usage scenarios.

## RELATED WORK

The increasing portability of projection systems since 2003 has enabled researchers to explore geometrically aware collaborative systems based on orientation. ILamps and RFIG [10,11] by Raskar et al. examined the geometric issues necessary for adaptive images of multiple projectors into a single display. The Hotaru system [12] uses static projection to simulate multiple projections. Several researchers have introduced the flashlight metaphor [12,3,14] for interacting with the physical world, where a portion of the virtual world is revealed by the projection. Cao et al. [2] presented techniques for revealing contextual information on the fly also using the flashlight metaphor.

Other researchers have explored playful and game-like possibilities enabled by portable projection technologies. Twinkle [6] by Yoshida et al., describes a handheld projector-camera system projecting a character that interacts with drawings on a whiteboard. Their system specifically reacts to edges, either drawn or created by objects in the scene but does not respond other systems in the same area. VisiCon [7] and CoGAME [8], are collaborative robot-controlling games using handheld projectors. In this work the camera is mounted on the robot instead of being embedded in the mobile phone.

## SYSTEM OVERVIEW

PoCoMo is a compact mobile projector-camera system, built on a commercially available cellular smartphone that includes a pico-projector and video camera. In the current prototype, the projector and camera are aligned in the same field of view via a plastic encasing and a small mirror (Figure 1). The setup enables the program to track both the projected image and a portion of the surrounding surface. All algorithms are executed locally on the phones.

The characters are programmed to have component parts with separate articulation and trigger different sequences based on the proximity of the other characters in order to appear to be life-like in the environment. We have designed characters for a number of scenarios to demonstrate how such a system may be used to for playful collaborative experiences:

- **Character Acknowledgment** The projected characters can respond to the presence and orientation of one another and express their interest in interaction. Initially they acknowledge each other by turning to face one another and smiling.
- **Icebreaker Interactions** When characters have acknowledged each other, a slow approach by one character to the proximity to another can trigger a gesture of friendship such as waving (medium distance) or shaking hands (close proximity).
- **Mutual Exploration** When characters are moved through the environment they begin a walk cycle, which is correspondent the speed of change of position. If the characters are within view of each other, their cycles approximately match. The synchronized steps help create



Figure 2. Character rigs, and component parts for animation.

the feeling of a shared environment, and provide feedback about proximity.

- **Leave-a-Present** Characters can find and leave *presents* to one another at distinct features in the environment. We provide this as an example of collaborative interactions that can extend into multiple sessions and relate to specific physical locations.

## Detection and Tracking Methods

In our system, each device projects a character that should be detected and tracked, both by the device itself and other participating devices in the same projection area. For this purpose we implemented a simple marker-based tracking algorithm, inspired by the seminal work on Active Appearance Models [4] (AAM) and color-based tracking methods [13,6]. Each character has an identity that is represented by the color of its markers. For instance one character may have blue markers, while the other character may have red markers. The detection algorithm scans the image for the circular markers by first applying a threshold on the Hue, Saturation and Value, extracting the contour poly-lines. Following the work on AAM, we create a marker of a predefined shape, e.g. a circle, and match its contour to the blobs discovered in the masking stage. The two highest-ranking regions are considered to be the markers. In the following iterations, similar to [14], we consider the distance of the region to the previous detected markers as a weight to the rank equation:

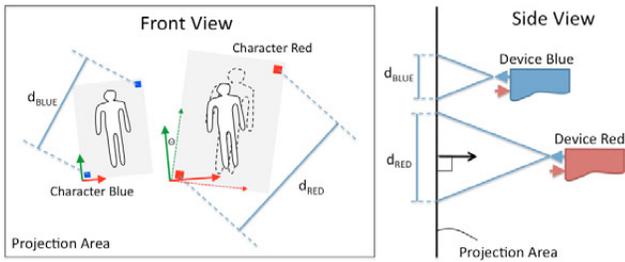
$$w(R) = R_{area} \cdot d(R_{center}, \hat{R}_{center}) e(R_{contour})$$

where  $R$  is the candidate region,  $d$  is a simple L2 distance function and  $e$  is the ellipse-matching measure.

The red character will look for blue markers, and vice versa, however both characters will also look for their own markers for calibration purposes.

The characters have markers on the opposite corners of the projection, i.e. top-right and bottom-left. Under the assumption of having both the camera plane and projector parallel to the projection plane this layout allows understanding the rotation angle and size of the projected character. These assumptions make the detection far quicker, in comparison to assuming arbitrary projective

transformations from the projection plane, while



**Figure 3. Rectification of the characters using the projected markers, under the assumption that the projectors and cameras are positioned perpendicular to the projection plane. The relation between  $d_{RED}$  and  $d_{BLUE}$  and the angle  $\theta$ , determine the rigid transformation applied to the red character.**

maintaining a lively experience. It also allows each system to determine the relation between the *self*-character and *friend* character, so the projection size and rotation can be adjusted to match, as Figure 3 shows. To maximize resolution, the larger projected character will try to match its size and rotation to the smaller one.

We are limited by the brightness (12 lumens) of our current projectors and therefore the scale and size of our current projected markers are exaggerated. Work by Costanza et al. [5] emphasizes the importance of incorporating the design of the markers with the visual content. In future work we will incorporate the markers into the design of the characters, and use them to provide feedback about the state of the character to the user.

## IMPLEMENTATION DETAILS AND CODE

### Character Interaction

**Following Head.** In the initial sensing stage, the projected character displays an *interested* look: The eyes follow the possible location of the other character. The presence of another character is determined once the two markers were observed in the scene for over 20 frames in the duration of the last 40 frames. Once the presence of another character in view is the *operated* character will face the *friend* character. If the characters linger for 80-100 frames they will wave in the direction of the other. If the characters are aligned they will attempt to do a *handshake*. Alignment is considered sufficient if the center point of the *friend* character projection is within a normalized range.

**Handshake.** The *handshake* is a key-frame animation of 10 frames, with 10 frame transitions from a standing pose. In order to synchronize the animation the *initiating* characters displays a third colored marker for a small period of time, to index the key frame timing. The *second* character reads the signal, and initializes their sequence as well. The handshake sequences of the characters are designed to have equal length animations for interoperability.

**Leave-a-Present.** The characters can also leave a *present* on a certain object for the other character to find. This is done

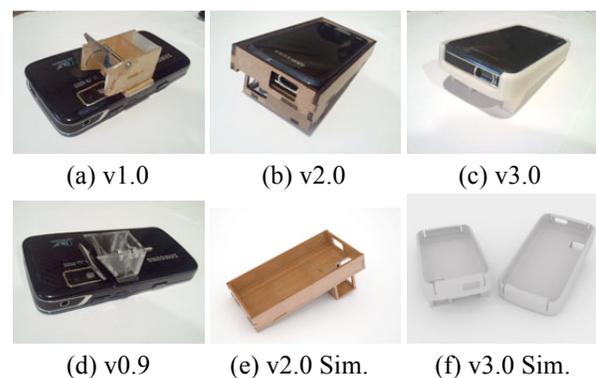
when the character is placed near an object, without movement, for over 3 seconds. An animation starts, showing the character leaning to leave a box on the ground, and the user is prompted for an object to leave: an image or any kind of file on the phone. The system extracts SURF descriptors of key-points on the object in the scene, and sends the data along with the batch of descriptors to a central server. The present location is recorded relative to strong features like corners and high contrast edges. When a character is in present-seeking mode, it tries to look for features in the scene that match those in the server. If a feature match is strongly correlated, the present will begin to fade in. We assume that users will leave presents in places they have previously interacted, to provide a context for searching for any virtual items in the scene.

### Gestures

Our system is responsive to the phone's motion sensing capabilities, an accelerometer and gyroscope. The characters can begin a walk sequence animation when the device is detected to move in a certain direction at a constant rate. The current accelerometer technology used in our prototype only gives out a derivative motion reading, i.e. on the beginning and end of the motion itself and not during. Therefore we use two cues: Start of Motion and End of Motion, which trigger the starting and halting of the walk sequence. The system also responds to shaking gestures, which breaks the current operation of the character.

### Hardware

We chose to use the Samsung Halo projector-phone as the platform for our system. This Android-OS system has a versatile development environment. The tracking and detection algorithms were implemented in C++ and compiled to a native library and the UI of the application was implemented in Java using Android API. We used the



**Figure 4. Development stages of the phone case prototype. (e,g) 3D rendering of the prototypes design.**

OpenCV library to implement the computer vision algorithms.

**The Prototype.** To align the camera with the projection, we designed a casing with a mirror that fits on the device. Figure 4 shows the development stages in the design of the prototype. First, our prototype was attached to the device,

and then we transitioned into full-device casing to support handling the phone. Finally, we designed a rounded case made from a more robust plastic to protect the mirror and allow it to be decoupled from the phone.

#### FUTURE WORK

In future systems we plan to make the detection more robust and integrate markers with the projected content. We also plan to develop animations that correlate to specific user profiles and migrate the application to devices with wider fields of views for the projector and camera. In addition, we will conduct a user study to assess the fluidity of the interaction and improve the scenarios.

#### CONCLUSION

We present PoCoMo, a system for collaborative mobile interactions with projected characters. PoCoMo is an integrated mobile platform independent of external devices and physical markers. The proposed methods do not require any setup and can be played in any location.

Our system allows users to interact together in the same physical environment with animated mixed-reality characters. The use of projectors for multiplayer gaming affords more natural and spontaneous shared experiences. Closely related systems support utilitarian tasks such as sharing [2] or inventory management [11]. We explore novel social interactions enabled by mobile projectors and hope developers will contribute other scenarios and games within the domain of collaborative mobile projection.

#### ACKNOWLEDGMENTS

Removed for blind review...

#### REFERENCES

- [1] G. Blasko, S. Feiner, and F. Coriand, "Exploring interaction with a simulated wrist-worn projection display," *9<sup>th</sup> Symposium on Wearable Computers*, 2005, pp. 2-9.
- [2] X. Cao, C. Forlines, and R. Balakrishnan, "Multi-user interaction using handheld projectors," *Proceedings of the 20th annual ACM symposium on User interface software and technology*, ACM, 2007, p. 43-52.
- [3] X. Cao, C. Forlines, and R. Balakrishnan, "Multi-user interaction using handheld projectors," ACM, 2007, p. 43-52.
- [4] T.F. Cootes, G.J. Edwards, and C.J. Taylor, "Active appearance models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, Jun. 2001, pp. 681-685.
- [5] Costanza E., Huang J., "Designable Visual Markers" In the Proceedings of ACM CHI2009, April 2009.
- [6] P. Fieguth and D. Terzopoulos, "Color-based tracking of heads and other mobile objects at video frame rates," *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 21-27.
- [7] K. Hosoi, V.N. Dao, A. Mori, and M. Sugimoto, "VisiCon : A Robot Control Interface for Visualizing Manipulation Using a Handheld Projector Cooperative Robot Navigation Game via Projection," *Human Factors*, 2007, pp. 99-106.
- [8] K. Hosoi, V.N. Dao, A. Mori, and M. Sugimoto, "CoGAME: manipulation using a handheld projector," *ACM SIGGRAPH 2007 emerging technologies*, ACM, 2007, p. 2.
- [9] S. Rapp, G. Michelitsch, M. Osen, J. Williams, M. Barbisch, R. Bohan, Z. Valsan, and M. Emele, "Spotlight Navigation: Interaction with a handheld projection device," *Advances in Pervasive Computing: At PERVASIVE*, 2004, pp. 397-400.
- [10] R. Raskar, J. van Baar, P. Beardsley, T. Willwacher, S. Rao, and C. Forlines, "iLamps: geometrically aware and self-configuring projectors," *ACM SIGGRAPH 2005*, p. 5.
- [11] R. Raskar, P. Beardsley, J. Van Baar, Y. Wang, P. Dietz, J. Lee, D. Leigh, and T. Willwacher, "RFIG lamps: interacting with a self-describing world via photosensing wireless tags and projectors," *ACM Transactions on Graphics (TOG)*, vol. 23, 2004, p. 406-415.
- [12] M. Sugimoto, K. Miyahara, H. Inoue, and Y. Tsunesada, "Hotaru: intuitive manipulation techniques for projected displays of mobile devices," *Human-Computer Interaction-INTERACT 2005*, 2005, pp. 57-68.
- [13] A. Yilmaz, O. Javed, and M. Shah, "Object tracking," *ACM Computing Surveys*, vol. 38, Dec. 2006, p. 13-es
- [14] T. Yoshida, Y. Hirobe, H. Nii, N. Kawakami, and S. Tachi, "Twinkle: Interacting with physical surfaces using handheld projector," *2010 IEEE Virtual Reality Conference (VR)*, Mar. 2010, pp. 87-90.