Flexible Low-Cost Digital Puppet System

Nanjie Rao(✉), Sharon Lynn Chu, and Ranger Chenore

University of Florida, Gainesville, FL 32611, USA
{raon,slchu,rangerchenore}@ufl.edu

Abstract. Puppet-basedsystems have been developed to help children engage in storytelling and pretend play in much prior literature.Many different approaches have been proposed to implement suchpuppet-based storytelling systems, and new storytelling systems are stillroutinely published, indicating the continued interest in the topic across domains like child-computer interaction, learning technologies, and the broader HCI community. This paper firstpresents a detailed review of the different approaches that have been usedfor puppet-based storytelling system implementations, and then proposesaflexible low-cost approach to puppet-based storytelling system implementation that uses a combination of vision- and sensor-based tracking. We contribute a framework that will help the community to make sense of the myriad of puppet-based storytelling system implementation approaches in the literature, and discuss results from a perceptionstudy that evaluated the performance of the system output using our proposed implementation approach.

Keywords: Puppet storytelling system · Digital puppetry · Perception study

1 Introduction

Storytelling plays an important role in education. As an approach of nurturing storytelling, digital puppetry has been used to help people engage in storytelling and pretend play. It is a great way to encourage creativity in children through expressive storytelling activities [23], such as the free-form pretend play (also referred to in literature as fantasy play [21], make-believe play [4] or story enactment [3]). This paper addresses the design, implementation and evaluation of a puppet-based storytelling system.

Different kinds of media serve as feedback to tangible interaction in all tangible narratives [6]. To enable tangible interaction with a puppet system, the tracking system is the stepping-stone. It defines the limitation of movements of the puppets, hence it forms the interaction style between puppeteer and puppet system [5]. It also is impacted by the physical affordances of the puppet as an object, as

Supported by National Science Foundation Grant #1736225 *To Enact, To Tell, To Write: A Bridge to Expressive Writing through Digital Enactment.*

J. Y. C. Chen and G. Fragomeni (Eds.): HCII 2021, LNCS 12770, pp. 676–694, 2021.
https://doi.org/10.1007/978-3-030-77599-5_46

the tracking system could rely on characteristics of the puppets [8,10,17,22] or constrain the form of the puppets [1,2,18].

Puppet systems presented in the literature were either settled with fixed combinations of technology and certain physical affordances [5] or set up in a way to address concepts outside of storytelling [6], which call for a flexible low-cost digital puppet storytelling system. The novel approach for our proposed storytelling system implementation is that the six components in our system are designed and employed in ways that they can each make it convenient for each other to work and communicate seamlessly. Mazalek et al.'s work [17] on the Xperimental Puppetry Theater stated that smooth communication and continuous mapping from the puppet data to the virtual artwork is the most challenging part of the implementation of puppet-based storytelling systems. Furthermore, in our approach, the combination of these six components are flexible yet robust in the way that other researchers could easily substitute our components to the ones they prefer, including the physical affordances as a puppet and object, the vision-based position tracking system, the sensor-based rotation tracking system, the display system, the story creation interface, and the data communication system.

In Harley et al.'s framework [6], 21 existing tangible narrative systems were reviewed, and the systems target both adults and children. In our consideration, children's behaviors are more unpredictable, and we intend to include children as potential users for our puppet system. Hence, much of our design rationale took into account children's preferences and developmental needs. We developed our puppet storytelling system using 3D printed puppet and objects, YOLO real-time object detection, the Aimxy and the BBC:microbit sensors, and Unity. We present a study that evaluates the quality of the implementation. However, as time goes by and new technology emerges, researchers could update each component as they wish. For example, the YOLO vision tracking and sensors could be replaced by ultra-wideband radio technology so that position and rotation tracking can be implemented on a single microchip.

In summary, our contributions are: (1) a review of the existing approaches that have been usedfor puppet-based storytelling system implementations; (2) a flexible low-cost approach to puppet-based storytelling system implementation; (3) an online perception study demonstrating the perceived quality of the output of our puppet storytelling system.

2 Background and Related Work

2.1 Digital Puppetry and Tangible Narratives

The manipulation of digitally animated 2D or 3D characters and objects in a virtual environment is considered to be digital puppetry. Ferguson defines digital puppeteering as a reflection of human motions onto digital animations with possible abstraction [5]. The productions of digital puppetry were widely used for movies, television series, live theaters, and Disneylands (and other interactive theme parks). However, it is not uncommon that people without advanced animation skills want to create digital stories. These digital stories can be for purposes that include arts and performance, learning and education, self-expression, product prototyping, etc. With many existing storytelling systems, storytelling requires specific hardware and software systems, training on specific techniques and skills, and even the coordination of multiple people.

To enable low-cost and easy to learn digital puppetry, laborious, and time-consuming key-frame based animation techniques are unsuitable, especially for non-expert animators [11]. Stories created by reflecting the performer's movement in real life may be more suitable. Pricey high-end motion capture systems like Vicon and OptiTrack are adopted among the professional filmmaking industry [11]. However, with recent developments in computer vision, inertial motion unit (IMU) sensors, and mass production of related consumer electronics, many different kinds of low-cost approaches to motion tracking have emerged. Acceptable accuracy of tracking can be achieved without a high budget. Despite the large number of motion-tracking-based storytelling systems that have been proposed thus far, we still have little understanding of what approaches work best. We highlight two prior efforts that have attempted to make sense of existing approaches to storytelling systems.

Ferguson conducted studies using commercially available production from three companies [5]. The systems were evaluated as they were out of the box, including digital characters, software, rigs, constraints, relations, general setup, joystick control, data glove control, 6DOF magnetic sensor control, and microphone control. As intended for single person multimodal live operation systems, the experience and results from these existing systems were disappointing. Issues were identified by experts including inconsistent data stream, cumbersome setup, device limitation, erratic concept mapping, mismatch in perceptual-motor coordination, movement constraint, and operation fatigue.

Different from Ferguson's 2015 paper which focused on commercial systems, Harley et al. presented a framework for tangible narratives in 2016 [6]. Their framework isolated the characteristics resulting from the storytelling systems that utilized physical objects as media to map into virtual environments. Seven categories were identified and employed across 21 systems. The categories were "primary user, media, the narrative function of the tangible objects, diegetic tangibles, narrative creation, choice, and position", which defined the structure of the narrative. Regardless of the different categories, among all systems, the stories were pushed forward by tangible interaction. Interestingly, although there were many different stated motivations behind creating these tangible narrative systems, the story itself was never the main purpose.

2.2 Existing Puppet Storytelling Systems

Based on a review of existing approaches for puppet systems, we specified a framework for decomposing puppet-based storytelling systems. Each aspect of the framework corresponds to a component of the puppet system, shaping how the puppet system is created and utilized:

1. The physical affordances as a puppet and object;
2. The vision-based tracking system to get 3D position data of the objects;

3. The sensor-based tracking system to get the raw, pitch, yaw rotation data of the physical objects;
4. The display system to animate the tracked data into virtual avatar animations and movement
5. The story creation interface to scaffold the storyline;
6. The data communication system to deliver real-time tracking data from 2 and 3 to 4.

We summarize 17 puppet systems created between 2008 and 2019, using the framework described above. We abstain from discussing specific concepts in the systems outside of storytelling, focusing on how the puppetry is captured and how the interaction goes. Here are the variations of approaches found for each component across all the 17 systems reviewed:

- **Puppet and object**: 2D paper puppets [1,2]; 3D printed puppets [19]; traditional puppets with realistic fabric [9,16,17,20,24]; human skeleton [12]; human hands [11,13–15]; robot-like puppets [8,10,22], and VR controllers [9,18,20].
- **Vision-based tracking**: Marker-based tracking [1]; color detection [2], Kinect [12,19]; VR base station [9,18,20]; and Leap Motion [11,13–15].
- **Sensor-based tracking**: Sensors in the joints of robot-like puppets [8,10,17,22]; IMU sensors (standalone module and inside VR controllers) [9,18–20].
- **Display system**: Unity engine [11–15,17,18]; non-Unity animator [1,2,8,10,19,22]; actual physical puppets [9,16,20,24].
- **Story creation interface**: Special cultural themes [16,17,24]; close loop control as in games [14,15]; no systematic instruction [1,2,8–13,18–20,22].
- **Data communication**: Database and data feeding software [17].

Table 1 shows system characteristics of 17 tangible systems published from 2008 to 2019 using the sections defined above. Only one paper [17] described their data communication system in detail, so we didn't list this component.

Below, we selected 5 representative puppet storytelling systems from prior literature that illustrated different combinations of the various approaches in detail above for the 6 components. Although the intended audience of the systems may include either adults or children, children's preference and development need were discussed specifically because their behavior was more unpredictable.

FingAR: Marker-Based Tracking+2D Paper Puppets. Bai et al. [1] developed the FingAR Puppet system shown in Fig. 1. It's based on multiple open-source software. Their system framework used Microsoft XNA Game Studio 4.0, AR registration, and rendering used GoblinXNA 4.1, marker tracking used ALVAR2.0, and image processing used Emgu CV2.4. FingAR is a good example of puppet systems that majorly used computer vision to achieve tracking, and physical referents (shaped card boards) as the puppet design.

Due to the nature of vision-based tracking, the system would lose track of the object when the camera had a blur which happened frequently when the object

Table 1. System components of 17 puppet systems published from 2008 to 2019 using our framework. NB: sjr = sensors in the joint of robot-like puppets, clc = close loop control, and nsi = no systematic instruction.

System	Puppet and object	Vision tracking	Sensor tracking	Display	Story creation interface
FingAR [1]	2D paper puppets	Marker-based	n/a	non-Unity animator	nsi
Video Puppetry [2]	2D paper puppets	color detection	n/a	non-Unity animator	nsi
Jacobson et al.'s [8]	robot-like puppets	n/a	sjr	non-Unity animator	nsi
Kawahara et al.'s [9]	VR controllers	VR	IMU	actual physical puppets	nsi
Lamberti et al.'s [10]	robot-like puppets	n/a	sjr	non-Unity animator	nsi
Anim-Actor [12]	human skeleton	Kinect	n/a	Unity	nsi
Mani-Pull-Action [13]	human hands	Leap Motion	n/a	Unity	nsi
Virtual Marionette [11]	human hands	Leap Motion	n/a	Unity	nsi
Liang et al.'s 1st [14]	human hands	Leap Motion	n/a	Unity	clc
Liang et al.'s 2nd [15]	human hands	Leap Motion	n/a	Unity	clc
Liu et al.'s [16]	traditional puppets	n/a	IMU	actual physical puppets	special cultural themes
Mazalek et al.'s [17]	traditional puppets	n/a	IMU	Unity	special cultural themes
Nitsche et al.'s [18]	VR controllers	VR	IMU	Unity	nsi
Figurines [19]	3D printed puppets	Kinect	IMU	non-Unity animator	nsi
Sakashita et al.'s [20]	VR controllers	VR	IMU	actual physical puppets	nsi
Yoshizaki et al.'s [22]	robot-like puppets	n/a	sjr	non-Unity animator	nsi
Zhao et al.'s [24]	traditional puppets	n/a	IMU	actual physical puppets	special cultural themes

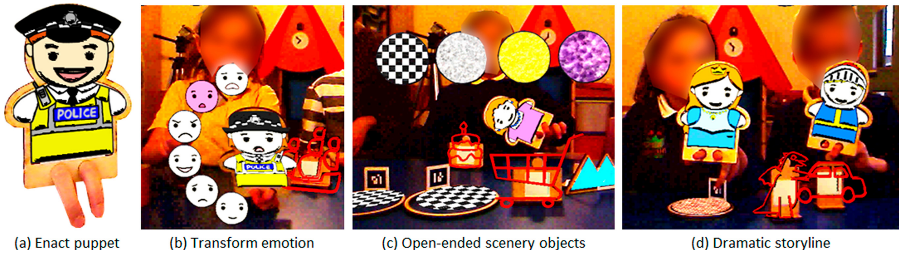

Fig. 1. FingAR puppet system overview in Bai et al.'s paper

was moving fast and the camera couldn't focus. It's relatively easy to implement because of the easy availability of open-source software and libraries including OpenCV and Aruco. It would be a "fast and dirty" solution for a puppet system that was meant to verify other concepts like emotions mentioned in Bai's work [1]. However, to achieve a puppet system that can support fluent and realistic storytelling where the focus is on the story creation, a higher tracking rate and a more robust system are required. And to bump the technical specs up, the cost would skyrocket. Also, there's no systematic instruction on storytelling creation in Bai et al.'s work.

Figurines: Fusion Tracking+3D Printed Puppets. Maxime Portaz et al. [19] proposed Figurines, adopted a hybrid system, using 2 RGB-D cameras and IMU sensors embedded figurines as vision-based and sensor-based tracking. They used 3D printed figurines and décor element as puppet design, and had 3d rendering offline after the recording of the narrative session. The designated area for puppet playing is the table shown in Fig. 2, sizing 70 cm × 70 cm. Obviously, taking advantage of RGB-D sensors and fusion tracking with IMU sensors made Figurines system more robust than the puppet systems that relied solely on vision-tracking. In fact, our framework was built on top of the fusion tracking framework since it combines the advantages of multiple techniques. It's incredible that some RGB-D sensors including Kinect and Intel RealSense are so commercially easily accessible at around 100 dollars. However, as we have tested, using RGB-D sensors had its own drawbacks. Firstly, it has a strict space limitation. This means that while holding the puppet, you cannot be too close or too far from the sensors. The IR sensor in this solution relies on the emission of the structured IR light pattern and the reflection it got back from the puppet. This also means that no mirror or other reflective material, including any smooth surface like a piece of paper, a side of a cabinet, or a cellphone screen should appear in the puppet playing space. This might be fine when the users are invited into a certain lab where the environments are strictly taken care of. But we demand more than a prototype in a lab environment. We want a truly robust and flexible puppet storytelling system so that we can bring it into a real elementary school where the children can freely enact stories on their own tables. Another important limitation of Maxime Portaz et al.'s puppet system is that the users cannot see their production on the digital display in real-time. And as it's been pointed out in the multiple works of Ferguson, Leite, and

Fig. 2. Figurines tangible storytelling system in Portaz et al.'s paper

Harley [5,7,11], it's important that the animation reflects the performer's movement in real life without post-processing. Also, there's no systematic instruction on storytelling creation in Portaz et al.'s work.

Liang et al.'s System: Leap Motion+Human Hand. Liang et al. [15] utilized the Leap Motion sensor to trigger pre-recorded animations of the crow puppet. For the purpose of letting young children easily control, they mapped different hand gestures to up, down, left and right, shown in Fig. 3.

This type of system has constrained degrees of freedom in mapping the real-world movement to the digital puppet. Also, in our experience of using the Leap Motion style sensor to capture the hand gesture, it felt more like using a game controller instead of the free form storytelling we want. The command list mapping might feel natural for a gamer, but it still differs from the original affordance of the physical puppet which all sorts of users could directly manipulate the physical puppet to do storytelling freely [2].

Nitsche et al.'s System: VR Base Station+VR Controllers. Nitsche et al. used 3 main sample mappings to exemplify different opportunities that open up through bottom-up inclusion of puppetry principles in VR controls. Rod mapping

Hand gesture	Movement	Target action
	Move right	Fly to the right
	Move left	Fly to the left
	Move down	Fly down
	Move up	Fly up
	Stretch	Hover
	Stretch to grip	Grasp pebble/ stick
	Grip to stretch	Drop pebbles/ stick

Fig. 3. Hand gesture-based interactive puppetry system in Liang et al.'s paper

Fig. 4. VR marionette in Nitsche et al.'s paper

is an example of variable control schemes and emphasizes the relationship of the puppet to the environment. The marionette mapping is equally variable through a changing control mechanism and it offers a possible solution for a 3rd person VR control scheme that might allow higher mobility through spatial tracking of the Vive controllers, shown in Fig. 4. The hand puppet mapping demonstrates varying granularity where controls shift between different levels of "distance" as outlined by Kaplin.

The approaches in VR were considered to be cool for children in elementary school, but many individuals have glasses or motion sickness that are not compatible with VR applications. Also, an important element of the puppet storytelling system is social engagement, which is still in an early exploring phase for VR. While wearing HMD, the user is visually isolated from the rest of the world. The isolation might be good for some other applications, but not for puppet storytelling where research had stated that children's cognitive development in skills and judgment, as well as the appropriation of augmented tangible objects, were not sufficiently nurtured if isolated. During the open-ended puppet storytelling process, children exercise their cognitive skill, imagination, and symbolic transformation which are essential for their excellence and competence in their adulthood [1]. So we want an open space for the children to engage in open storytelling and social interactions instead of an isolated HMD in a VR approach.

Jacobson et al.'s System: Sensors in the Joints+Robot-Like Puppets. Jacobson et al. [8] presented a tangible modular input device shown in Fig. 5. The sensors embedded in joints could infer the pose of the robot-like puppets. Since they adopted a modular design, the topology could be updated automatically with the alternation on the splitter parts. In terms of target acquisition and pose replication, robot-like puppets are preferred over mouse and keyboard.

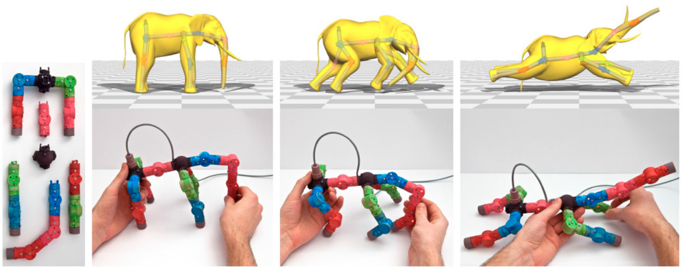

Fig. 5. Tangible and modular input device for character articulation in Jacobson et al.'s paper

Skeletal articulation made it accurate in terms of puppet limbs and action details, but it might still be a drag for a complete storyline where multiple scenes and characters would present. It's helpful for perfecting single poses, but the overall storyline was not addressed.

3 Goals for a New Flexible Low-Cost Digital Puppet Storytelling System

From our review of existing approaches, we can see that prior puppet-based storytelling systems have both strengths and limitations.

1. All the existing systems didn't focus on the story creation.
2. The systems usually required special setup and space, e.g., in the lab.
3. The operation for the systems required some expertise to use.
4. VR-based systems isolated the user from the real world.
5. The construction of the overall storyline was not scaffolded.
6. The systems would fail if any component didn't work.

To address the weaknesses identified, we propose six goals for a new flexible low-cost digital puppet storytelling system. The system should:

1. Support fluent storytelling where the focus is on the story creation.
2. Be robust enough to bring to an authentic classroom setting where children can freely enact stories on their own tables.
3. Be easy enough that persons of all ages and skill levels could readily engage.
4. Allow for social engagement.
5. Be able to support the overall storyline on top of enacting each scene.
6. Be modular enough such that each component is easily replaceable.

4 System Description

We developed a flexible low-cost puppet-based storytelling system that align with the six goals described above. Our proposed system can be described based on the six components of the framework specified in Sect. 2.2. Our system involves:

1. **Puppet and object**: 3D printed puppet and objects with the pattern for vision-based tracking and small slot left for sensor
2. **Vision-based tracking**: The YOLO vision tracking algorithm and training sets
3. **Sensor-based tracking**: The Aimxy and BBC:microbit
4. **Display system**: Self-created avatars and scripts inside Unity 3D
5. **Story creation interface**: A story creation interface inside Unity 3D
6. **Data communication**: UDP, Bluetooth, and shared memory mapping

The six components in our system are designed and employed so that they can each make it convenient for the other components to work and communicate seamlessly. For example, when we designed the 3D printed puppet and objects, we intentionally put in patterns on the base of puppet and object for better vision-based tracking and left small slots in the base for the sensor to be embedded, shown in Fig. 6. And the six components are so modular that other researchers could easily substitute our components for ones they prefer, including the physical puppet and object and supporting systems.

For better YOLO vision tracking results, we trained our own convolutional neural network using 148 picture samples marked by our researcher, shown in Fig. 7. The focus is on the base so that even if the 3D printed character and object change in other designs, the vision tracking would still work without training on the new set.

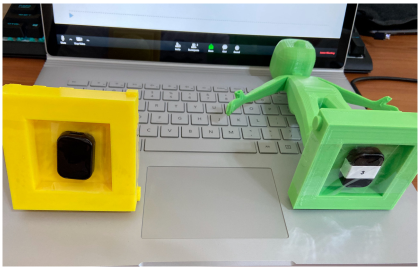

Fig. 6. Slot for sensor

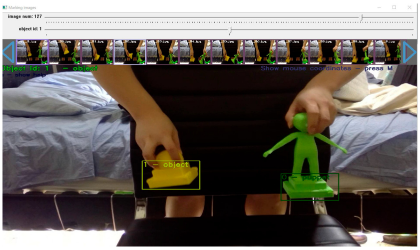

Fig. 7. YOLO pattern selection on the base

The story creation interface was based on the work of Zarei et al. [23] shown in Fig. 8. A collection of story scenes can be planned and ordered chronologically to construct the storyline before enactment. This supports the overall storyline on top of enacting each scene. This also brings the focus back to the story creation. Different backgrounds, characters, and objects can be chosen as the user wishes before acting out the scene. Some examples were shown in the Fig. 9.

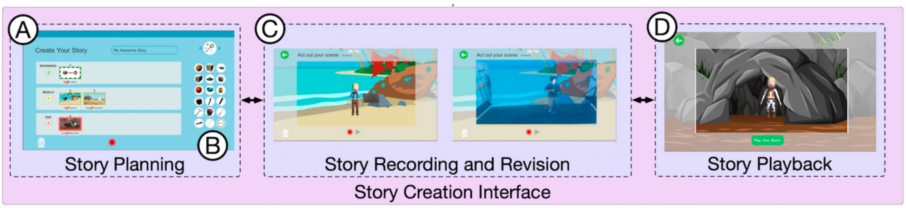

Fig. 8. Story creation interface

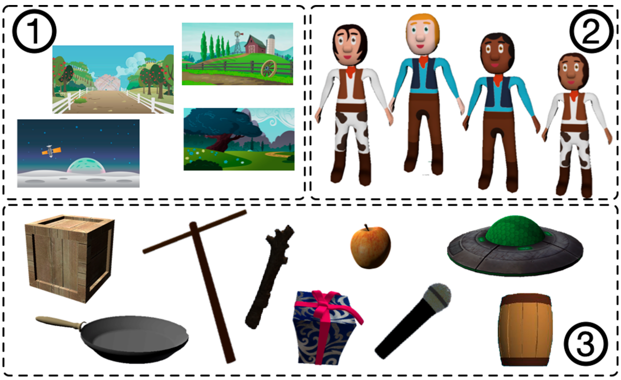

Fig. 9. Story backgrounds, characters and objects

Fig. 10. System setup

As mentioned in Mazalek et al.'s work [17], smooth communication, and continuous mapping from the puppet data to the virtual artwork was the most challenging part. The bridging software outside of Unity was also a crucial part of our puppet system. UDP (User Datagram Protocol) was used to broadcast the tracking data from the position tracking program to Unity, Bluetooth was used to send the rotation sensor data from the Aimxy and BBC:microbit to the receiving program on the computer, and shared memory mapping was used to send the received rotation data from the receiving program to Unity. As no fragile and expensive special devices were used, our system is both robust enough to bring to an authentic classroom setting and easy enough to use without expertise. Social engagement is also allowed because of no isolation of VR headset.

5 Evaluation of Our Puppet Storytelling System

Our research question for our evaluation study was as follows:

What is the perceived quality of the output of our proposed puppet storytelling system?

5.1 Evaluation Method

Due to the impact of COVID-19, no physical human subject study could be done. Instead, we conducted an online perceptionsurvey study to evaluate the performance of the system output using our proposed implementation approach. We recorded two stories with our system which were used in the evaluation. Both stories were enacted by the researcher. Both stories lasted less than three minutes. Story 1 was designed to be more complex as more "dramatic" movements were included by the researcher intentionally, while story 2 was simpler as the original storyline was created by a child in a previous full body enactment study. A sample physical set up of the system is shown in Fig. 10.

Our study had two independent variables: type of puppetry (IV1) and story design (IV2). Type of puppetry had 2 levels: Physical enactment; Virtual animation. Story design had 2 levels: Complex; Simple. For IV1, we were interested to see whether the virtual animation produced by the system faithfully represents the physical enactment of a story done by a user. Thus, we expected to see a replicated puppetry performance from physical puppetry to virtual animation (thus, no statistically significant difference between the 2 levels of the IV). An example of the physical puppetry and interface of virtual animation is shown in Fig. 11. From the raw footage (physical puppetry) and the system output (virtual animation), ratings from the participants were used as an evaluation of the puppet system. For IV2, complex stories were expected to be rated higher than simple stories, but we also expected an interaction between story designs (IV2) and the degree to which the virtual animation is perceived to faithfully represent the physical puppetry. Participants were expected to give their opinions in terms of the overall experience, appearance, clarity, degree of control, affective information, and importance of system components.

5.2 Study Participants and Study Protocol

We recruited workers on the Amazon Mechanical Turk crowdsourcing platform. Mturk workers were compensated via the Mturk platform as compensation. Out of 54 responses we collected, 23 of them (42.59%) were valid. Among these 23 Mturk workers, 13 (56.52%) of them are male, 10 (43.48%) of them are female.

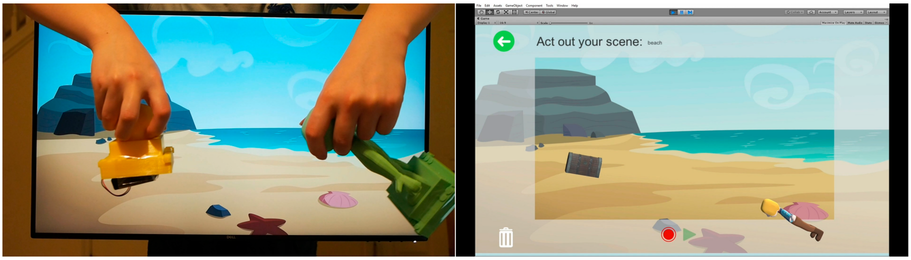

Fig. 11. Comparisons between type of puppetry

Each participant would go through one story design and see videos for both types of puppetry (physical enactment and its corresponding virtual animation). The average time needed for valid responses made by the Mturk workers was 35 min. In our complete survey study flow, the participant would accept our task on Mturk, click on the survey link on the MTurk task page which will guide them to our web page. On our first web page, the participant would read the instructions and consent to participate in the study, which would direct the participant to one of four study flows. Two study flows addressed one story, and the other two study flows addressed the other story. For each story, since we are counterbalancing the order in which participants engaged with IV1 (the type of puppetry), there were 2 study flows. The study instructions informed the participant that the study involved them watching story videos generated from a puppet-based storytelling system that the researchers developed.

After answering a few demographic questions, the participant would be directed to watch a story video (Physical puppetry or virtual animation recording using our puppet system), answer questions about this specific video based on his/her opinions about the video. Upon finishing the questionnaire in the previous web page, he/she will then be directed to another video of the same story (still recorded with our system, with same narration but with different setup - if the previous video was the physical puppetry setup then this one would be the virtual animation). The same set of questions would be answered with regard to the video of the different setups. After answering the questions for the two videos of physical enactment and virtual animation separately, the participant would proceed to a page where he/she would be asked to make a direct comparison between the physical enactment video and the virtual animation video. Both videos he/she had seen in the previous web pages would be shown on the same web page for this comparison, see Fig. 12.

5.3 Study Measures

Our survey contains sections: overall experience, appearance, clarity, degree of control, affective information, importance of system components, and some open-ended questions. For most of our questionnaire items, the 7-point Likert scale was used for the participants to rate their perception from 1 (not at all) to 7 (Very much) since almost all the questions were started with "How." Ratings were given by the participants for physical enactment and virtual animation separately about the angle of rotation of the movement, the carried narration, storyline, story details, and affective information, which used question template "How X did you feeltheYpuppetry was?".

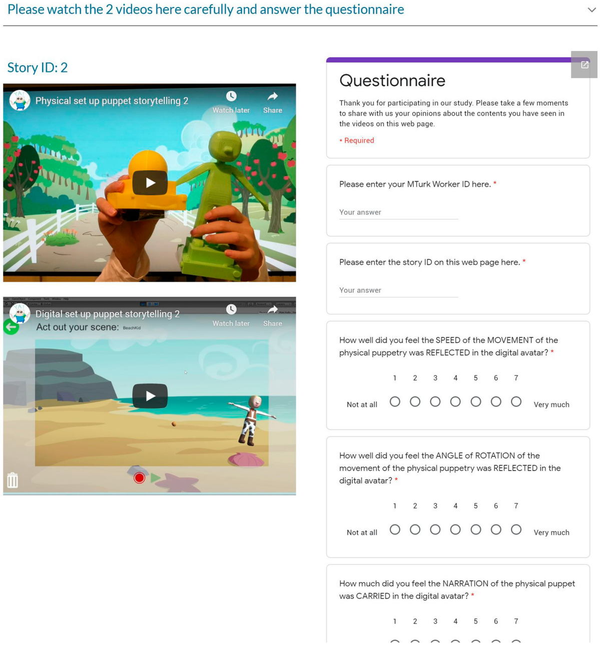

Fig. 12. The direct comparison web page in our evaluation

After separate ratings, participants were asked about direct comparison for speed of movement, angle of rotation of the movement, carried narration, storyline, story details, and affective information, which used question template "How well the virtual puppetry MATCH with the physical puppetry in terms of X?"

The participant would also describe their thought of the potential usage of the system, retell the story, and describe the details they observed between the physical and virtual animation. At the end of the questionnaires, the participant would be provided with a completion code that he/she need to enter in MTurk.

5.4 Response Validation

We ran a pilot test with the researcher observing the behavior of the pilot participants closely. Modifications were made to the parts where the participant felt unclear. We also ran some pilot tests with the Mturk batch, and it turned out that Mturk non-masters have a very high rate of not taking the survey carefully. 8 out of 9 (88.8%) pilot Mturk non-masters filled out invalid responses. A Mturk Master Worker is identified by the platform as they did good jobs consistently across different tasks assigned by multiple requesters. Thus, in the final iteration of our study, we decided to only recruit MTurk Master workers.

Some considerations were implemented in our web pages so that the study for the participants would involve less interruption and less invalid answers. On the website, the questionnaire items were arranged so that each web page only contained four questions, so that no scrolling is needed for the participants. Hence participants can refer back to the videos easily while answering the questions. Furthermore, we used this question at random locations in all three questionnaires for the purpose of an attention check: "While watching the television, how often have you ever had a fatal heart attack and died and were resurrected by a puppet and said:'where's my UFO?'?" And there are five options in the form of multiple-choice in the questionnaires, "Always, never, sometimes, rarely, very often." Only one of them was considered to pass the attention check, which was that "Never." The puppet and UFO were added into the question intentionally so unless the participants had read the contents of the questionnaire items, they would not have noticed this question was serving as attention check.

Participant responses were considered invalid if they did not finish the study, pass the attention check or did not provide quality entries to the open-ended questions in the questionnaire. All invalid responses were excluded from analysis.

5.5 Data Analysis and Evaluation Results

For questions asked separately on different web pages of physical enactment and virtual animation, two-way mixed ANOVA tests were run. There was no significant main effect of type of puppetry (IV1) on overall experience ($F(1, 21) = .42$, $p = .52$), appearance ($F(1, 21) = .12$, $p = .74$), clarity ($F(1, 21) = .26$, $p = .61$), degree of control ($F(1, 21) = .26$, $p = .61$), affective information ($F(1, 21) = .11$, $p = .75$), and importance of system components ($F(1, 21) = .37$, $p = .55$).

In addition, there were significant main effects of story design (IV2) on appearance ($F(1, 21) = 6.00$, $p < .05$), clarity ($F(1, 21) = 9.51$, $p < .05$), and affective information ($F(1, 21) = 7.02$, $p < .05$) between the two stories. Descriptive statistics showed that for appearance, story 1 (mean = 4.05, SD = .36) has higher ratings than story 2 (mean = 2.65, SD = .45); for clarity, story 1 (mean = 4.63, SD = .30) has higher ratings than story 2 (mean = 3.13, SD = .38); for affective information, story 1 (mean = 3.91, SD = .31) has higher ratings than story 2 (mean = 2.59, SD = .39). No significant main effect was found for overall experience ($F(1, 21) = 4.00$, $p = .06$), degree of control ($F(1, 21) = 0.00$, $p = .99$) and importance of components ($F(1, 21) = 2.19$, $p = .15$).

For the interaction between IV1 and IV2, there was no significant interaction on appearance ($F(1, 21) = 1.79$, $p = .20$), clarity ($F(1, 21) = 3.12$, $p = .09$), degree of control ($F(1, 21) = 0.09$, $p = .77$), affective information ($F(1, 21) = 2.72$, $p = .11$), and importance of system components ($F(1, 21) = 2.14$, $p = .16$). There was a significant interaction on overall experience ($F(1, 21)= 11.18$, $p <.05$). Descriptive statistics showed that while for story 1, virtual animation (mean = 4.86, SD = .42) was higher than physical enactment (mean = 3.79, SD = .39); story 2 showed the opposite pattern, virtual animation (mean = 2.78, SD = .52) was lower than physical enactment (mean = 3.50, SD = .49).

Aside from separate ratings, for questions that prompted participants to give direct comparison for the speed of movement, the angle of rotation of the movement, the carried narration, storyline, story details and affective information, we got an array of matching degree for reflection from physical enactment to virtual animation. One-sample t-tests were run to examine if there exist statistically significant differences between matching degree and baseline (greater or equal than 5 out of 7). No statistically significant difference was found.

An open coding process was done on responses from open-ended questions.

6 Discussion

Two-way ANOVA tests showed no main effect of type of puppetry (IV1) on all the dependent variables. So users perceived similarly which meant the virtual animation was perceived to faithfully represent the physical puppetry. Future puppeteers would feel consistency between the virtual animation they created and the physical puppetry they used in their real-world performance.

For IV2 - story design, there were significant main effects on appearance, clarity, and affective information where story 1 was rated higher. When using our puppet system, the complex story was perceived as more appealing, clearer, and more emotional than the simple story. This met our expected outcome.

And no significant main effect was found for overall experience, degree of control and importance of components. This meant our puppet system was consistent regarding technical perception no matter the story was complex or simple.

The results from direct comparison questions showed good matching degrees for reflection from physical enactment to virtual animation.

Some interesting findings from open-ended questions regarding usage scenarios and differences between physical enactment and virtual animation were that almost every participant thought this system was designed for kids, and very few participants discovered a limitation of the system, which is when the puppet and object collide, the animation is not ideal enough.

The evaluation done in this study helped us to begin to understand the benefits and challenges of such connections and combinations of the technologies that scaffolded the new puppet storytelling system.

7 Conclusion and Future Work

This paper firstcovered approaches usedfor puppet-based storytelling system, and then proposedaflexible low-cost approach to puppet-based storytelling system using a combination of vision- and sensor-based tracking. The results showed the consistency of the technical perception of our puppet system for different story design. The virtual animation was perceived to faithfully represent the physical puppetry. And the complex story was perceived as more appealing, clearer, and more emotional than the simple story. The novelty of our approach does not rely on any single part of the system but exists in the connections that glue each and every part together and takes advantage of the strength of every part. This flexibility warrants easy upgrades of new technology into the system. This helps to open up the design and implementation space of future interactive narrative authoring tools.

For future work, one promising avenue is to take advantage of the flexibility of this puppet system framework and expand the usage scenario. For example, with the advancing cellphone-based VR/AR technology, such as Apple ARKit and Google ARCore, the vision-based tracking and sensor-based tracking are already combined. And developers could train their custom tracking models faster with Lidar. Because it measures distances by illuminating the target with laser light and measuring the reflection with a sensor, it excludes most of the noise from data before feeding it to train the model. If designed and employed properly, mobile-based puppet system could emancipate the creativity of users by allowing greater freedom.

Another exciting avenue is to support more types of interaction. The technical aspects and storytelling-oriented design of this puppet system framework can be extended to scaffold more types of narrative creations, such as tabletop miniatures games (e.g., Warhammer) and building blocks.

We will also carry out the physical study with kids once the pandemic is over.

Acknowledgments. This research is supported by National Science Foundation Grant #1736225 *To Enact, To Tell, To Write: A Bridge to Expressive Writing through Digital Enactment*. Thanks to Lara Disuanco, and Grace Nemanic, who helped in technical development of this project.

References

1. Bai, Z., Blackwell, A.F., Coulouris, G.: Exploring expressive augmented reality: the fingar puppet system for social pretend play. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, pp. 1035–1044 (2015)
2. Barnes, C., et al.: Video puppetry: a performative interface for cutout animation. In: ACM SIGGRAPH Asia 2008 papers, pp. 1–9 (2008)
3. Chu, S.L., Quek, F., Tanenbaum, J.: Performative authoring: nurturing storytelling in children through imaginative enactment. In: Koenitz, H., et al. (eds.) ICIDS 2013. LNCS, vol. 8230, pp. 144–155. Springer, Cham (2013). https://doi.org/10.1007/978-3-319-02756-2_18

4. Dias, M., Harris, P.L.: The effect of make-believe play on deductive reasoning. Br. J. Dev. Psychol. **6**(3), 207–221 (1988)
5. Ferguson, J.: Lessons from digital puppetry: updating a design framework for a perceptual user interface. In: 2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing, pp. 1590–1595. IEEE (2015)
6. Harley, D., Chu, J.H., Kwan, J., Mazalek, A.: Towards a framework for tangible narratives. In: Proceedings of the TEI 2016: Tenth International Conference on Tangible, Embedded, and Embodied Interaction, pp. 62–69 (2016)
7. Harley, D., Tarun, A.P., Germinario, D., Mazalek, A.: Tangible vr: diegetic tangible objects for virtual reality narratives. In: Proceedings of the 2017 Conference on Designing Interactive Systems, pp. 1253–1263 (2017)
8. Jacobson, A., Panozzo, D., Glauser, O., Pradalier, C., Hilliges, O., Sorkine-Hornung, O.: Tangible and modular input device for character articulation. ACM Trans. Graph. (TOG) **33**(4), 1–12 (2014)
9. Kawahara, K., Sakashita, M., Koike, A., Suzuki, I., Suzuki, K., Ochiai, Y.: Transformed human presence for puppetry. In: Proceedings of the 13th International Conference on Advances in Computer Entertainment Technology, pp. 1–6 (2016)
10. Lamberti, F., Paravati, G., Gatteschi, V., Cannavo, A., Montuschi, P.: Virtual character animation based on affordable motion capture and reconfigurable tangible interfaces. IEEE Trans. Vis. Comput. Graph. **24**(5), 1742–1755 (2017)
11. Leite, L.: Virtual marionette. In: Proceedings of the 2012 ACM International Conference on Intelligent User Interfaces, pp. 363–366 (2012)
12. Leite, L., Orvalho, V.: Anim-actor: understanding interaction with digital puppetry using low-cost motion capture. In: Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology, pp. 1–2 (2011)
13. LEite, L., Orvalho, V.: Mani-pull-action: Hand-based digital puppetry. Proc. ACM Hum.-Comput. Interact. **1**(EICS), 1–16 (2017)
14. Liang, H., Chang, J., Deng, S., Chen, C., Tong, R., Zhang, J.J.: Exploitation of multiplayer interaction and development of virtual puppetry storytelling using gesture control and stereoscopic devices. Comput. Animation Vir. Worlds **28**(5), e1727 (2017)
15. Liang, H., Chang, J., Kazmi, I.K., Zhang, J.J., Jiao, P.: Hand gesture-based interactive puppetry system to assist storytelling for children. Vis. Comput. **33**(4), 517–531 (2016). https://doi.org/10.1007/s00371-016-1272-6
16. Liu, C.C., Liu, K.P., Wang, P.H., Chen, G.D., Su, M.C.: Applying tangible story avatars to enhance children's collaborative storytelling. Br. J. Educ. Technol. **43**(1), 39–51 (2012)
17. Mazalek, A., et al.: Pictures at an exhibition: design of a hybrid puppetry performance piece. In: Herrlich, M., Malaka, R., Masuch, M. (eds.) ICEC 2012. LNCS, vol. 7522, pp. 130–143. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33542-6_12
18. Nitsche, M., McBride, P.: A character in your hand: puppetry to inform game controls. DiGRA (2018)
19. Portaz, M., et al.: Figurines, a multimodal framework for tangible storytelling. In: WOCCI 2017-6th Workshop on Child Computer Interaction at ICMI 2017-19th ACM International Conference on Multi-modal Interaction, pp. 52–57 (2017)

20. Sakashita, M., Minagawa, T., Koike, A., Suzuki, I., Kawahara, K., Ochiai, Y.: You as a puppet: evaluation of telepresence user interface for puppetry. In: Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology, pp. 217–228 (2017)
21. Seja, A.L., Russ, S.W.: Children's fantasy play and emotional understanding. J. Clin. Child Psychol. **28**(2), 269–277 (1999)
22. Yoshizaki, W., et al.: An actuated physical puppet as an input device for controlling a digital manikin. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 637–646 (2011)
23. Zarei, N., Chu, S.L., Quek, F., Rao, N., Brown, S.A.: Investigating the effects of self-avatars and story-relevant avatars on children's creative storytelling. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–11 (2020)
24. Zhao, S., Kirk, D., Bowen, S., Chatting, D., Wright, P.: Supporting the cross-cultural appreciation of traditional Chinese puppetry through a digital gesture library. J. Comput. Cult. Heritage (JOCCH) **12**(4), 1–19 (2019)