Time-Offset Conversations on a Life-Sized Automultiscopic Projector Array

Andrew Jones, Koki Nagano, Jay Busch, Xueming Yu, Hsuan-Yueh Peng
Joseph Barreto, Oleg Alexander, Mark Bolas, Paul Debevec
USC Institute for Creative Technologies
12015 Waterfront Drive, Playa Vista, CA 90094
jones@ict.usc.edu

Jonas Unger
Linköping University
581 83 Linköping, Sweden

Abstract

We present a system for creating and displaying interactive life-sized 3D digital humans based on pre-recorded interviews. We use 30 cameras and an extensive list of questions to record a large set of video responses. Users access videos through a natural conversation interface that mimics face-to-face interaction. Recordings of answers, listening and idle behaviors are linked together to create a persistent visual image of the person throughout the interaction. The interview subjects are rendered using flowed light fields and shown life-size on a special rear-projection screen with an array of 216 video projectors. The display allows multiple users to see different 3D perspectives of the subject in proper relation to their viewpoints, without the need for stereo glasses. The display is effective for interactive conversations since it provides 3D cues such as eye gaze and spatial hand gestures.

1. Introduction

What would it be like if you could meet someone you admire, such as your favorite artist, scientist, or even world leader, and engage in an intimate one-on-one conversation? Face-to-face interaction remains one of the most compelling forms of communication. Unfortunately in many cases, a particular subject may not be available for live conversation. Speakers are both physically and logistically limited in how many people they can personally interact with. Yet the ability to have conversations with important historical figures could have a wide range of applications from entertainment to education.

Traditional video recording and playback allows for speakers to communicate with a broader audience but at the cost of interactivity. The conversation becomes a one-sided passive viewing experience when the narrative timeline is chosen early in the editing process. Particularly with first person narratives, it can be especially compelling to look the speaker in the eye, ask questions, and make a personal connection with the narrator and their story. Research has shown that people retain more information through active discussion over a passive lecture [18].

To solve this problem, we created a system that enables users to have interactive conversations with prerecorded 3D videos. In this paper, we simulate 3D conversations across time where one half of the conversation has already taken place. Our system presents each subject life-size on a dense automultiscopic projector array, combining both 3D immersion and interactivity. Automultiscopic 3D displays enable multiple users to view and interact with a speaker, and see the same 3D perspective as if he or she were actually present. For each subject, we record a large set of 3D video statements and users access these statements through natural conversation that mimics face-to-face interaction. Conversational reactions to user questions are retrieved through speech recognition and a statistical classifier that finds the best video response for a given question. Recordings of answers, listening and idle behaviors, are linked together to create a persistent visual image of the person throughout the interaction.

Our main contributions are:

1. A new dense projector array designed to show a life-sized human figure to multiple simultaneous users over a wide viewing area. The field of view can be easily customized with distributed rendering across multiple graphics cards and computers.

2. The 3D display is integrated with an interactive natural language interface that allows users to have simulated conversation with a prerecorded interview subject.

3. Display content is rendered using flowed light fields [4]. This technique allows for real-time resampling of sparse camera video directly to the projector array.

2. Interview Process

While it is impossible to anticipate and record all possible questions and answers, the system is based on the principle that any single prerecorded answer can serve as a viable response to a much larger set of potential questions. For some applications, it is possible to utilize scripted answers, carefully worded to be self-contained with a restricted topic of conversation. One of the first systems that allowed spoken interaction with a historical character was the August system [5]. This system used an animated ”talking head” fashioned after August Strindberg that could provide tourist information about Stockholm, as well as deliver quotes from and about Strindberg himself. The Virtual Human Toolkit [6] has been used to create multiple scripted characters such as Sgt. Blackwell [13] and the Eva and Grace virtual twins [25, 24], each serving as virtual educational guides that tell stories in response to user questions. All these systems utilize fictional characters modelled and animated by artists. In the late 1990s, Marinelli and Stevens came up with the idea of a ”Synthetic Interview”, where users can interact with a historical persona that was composed using video clips of an actor playing that historical character and answering questions from the user [16]. ”Ben Franklin’s Ghost” was a system built on this concept that was deployed in Philadelphia in 2005-2007 [22]. The system used speech recognition and keyword-spotting to select the responses.

It is not desirable or possible, however, to script all conversations with real people. Instead we utilize extensive interviews to gather a wide range of natural responses. The subjects interviewed for this project were experienced public speakers. By analyzing previous lectures and interviews, we gathered the most common audience questions. We also devised a set of prompts to further the interaction, including short factual biographical information, opinions, and stories. In cases where a question does not have a direct answer, a good story can often fill in the gaps. If no response is suitable, the subject will ask the user to restate the question or suggest a new topic. Additional details on the question development and analysis can be found in [2].

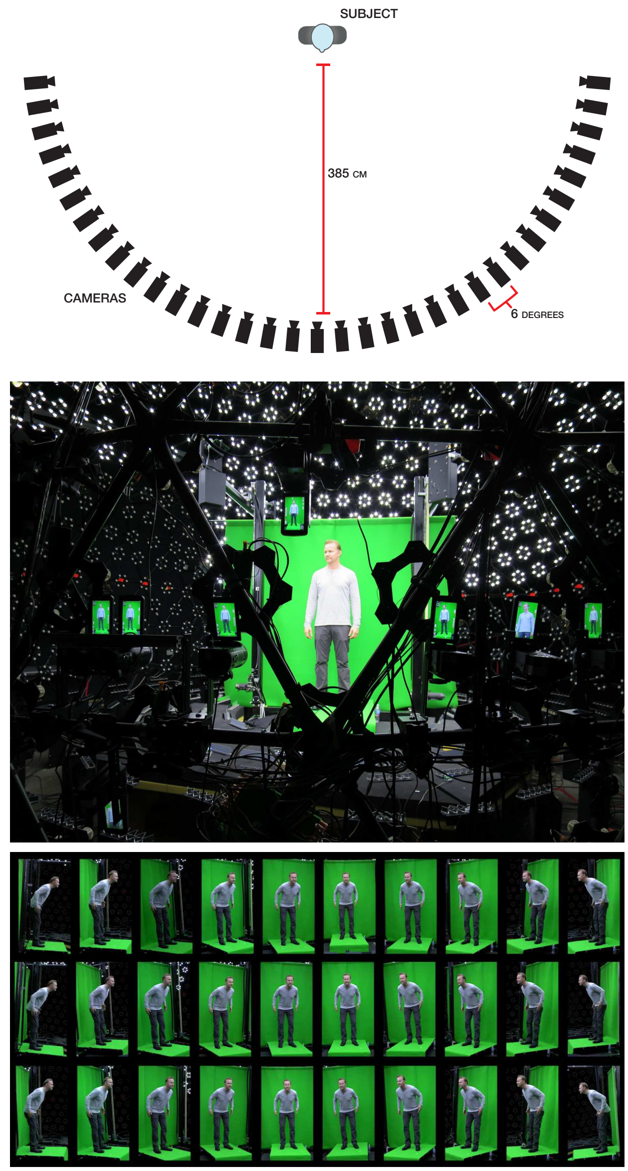

Figure 1. (top) Overhead diagram showing placement of cameras. (center) Seven of the Panasonic cameras mounted around the stage to record the performance. (bottom) Mosaic of all 30 camera views.

3. Data Capture

We record each subject with an array of 30 Panasonic X900MK cameras, spaced every 6 degrees over a 180 degree semi-circle and at a distance of 4 meters from the subject (see Figure 1). The cameras can record multiple hours of 60fps HD footage directly to SD cards with MPEG compression. As the Panasonic cameras were not genlocked, we synchronized our videos within 1/120 of a second by aligning their corresponding sound waveforms.

A major consideration during the interviews was maintaining video and audio continuity. This is important as the time-displaced interaction may jump back and forth between different takes and even different days of production. As much as possible, cameras were triggered remotely to avoid any unnecessary camera motion. We also prepared multiple identical outfits for the interview subject to wear on successive days. Between interview sessions we would try to match body posture and costume. A video overlay was used to rapidly compare footage between sessions. Even with all these efforts, maintaining complete continuity was not possible. In particular, we noticed changes in how clothing would fold and hang as well as changes in the subject’s mood over the course of days. Both types of changes may be noticeable when transitioning between disparate answers. Scene illumination was provided by a LED-dome with smooth white light over the upper-hemisphere (see Figure 1). This is a neutral lighting environment that also avoids hard shadows.

A key feature of natural conversation is eye-contact, as it helps communicate attention and subtle emotional cues. Ideally, future viewers will feel that the storyteller is addressing them directly. However, in early tests, when the interviewer was fully visible, the subject would tend to address the interviewer and not look at the camera. Alternatively, if the interviewer was completely hidden, the tone of the interview would feel subdued and less engaging. Our solution was to place the interviewer outside the stage and hidden behind a curtain, while the interviewer's face was visible as a reflection through a mirror box aligned with the central cameras.

After the interview we segmented the interview into stand-alone video responses. The initial rough edit points are marked during the interview transcription process. These in/out points are refined by automatically detecting the nearest start and end of the speech where the audio levels rose above a threshold. Occasionally, the detected audio start and end points will not exactly match the natural video edit points. For example, if the subject made silent hand-gestures prior to talking, these should be included in the video clip. In these cases we manually adjusted the audio and video edit points.

4. Display Hardware

Previous interactive "digital human" systems [22, 13] were displayed life-size but using conventional 2D technology such as large LCD displays, semi-transparent projection screens or Pepper's ghost displays [23]. While many different types of 3D displays exist, most are limited in size and/or field of view. Our system utilizes a large automultiscopic 3D projector array display capable of showing a full human body.

Early projector array systems [7, 17] utilized a multilayer vertical-oriented lenticular screen. The screen optics refracted multiple pixels behind each cylindrical lens to multiple view positions. Recent projector arrays [20, 11, 28, 10] utilize different screens based on vertically scattering anisoptropic materials. The vertical scattering component allows the image to be seen from multiple heights. The narrow horizontal scattering allows for greater angular resolution as it preserves the horizontal divergence of the original projector rays.

In order to reproduce full-body scenes, the projectors making up the array require higher pixel resolutions and brightness. Secondly as full bodies have more overall depth, we must increase the angular resolution to resolve objects further away from the projection screen. We use LED-powered Qumi v3 projectors in a portrait orientation, each with 1280 $\times$ 800 pixel image resolution (Figure 2). A total of 216 video projectors are mounted over 135° in a 3.4 meter radius semi-circle. At this distance, the projected pixels fill a 2 meter tall anisotropic screen with a life-sized human body (Figure 3). The narrow 0.625° spacing between pro-

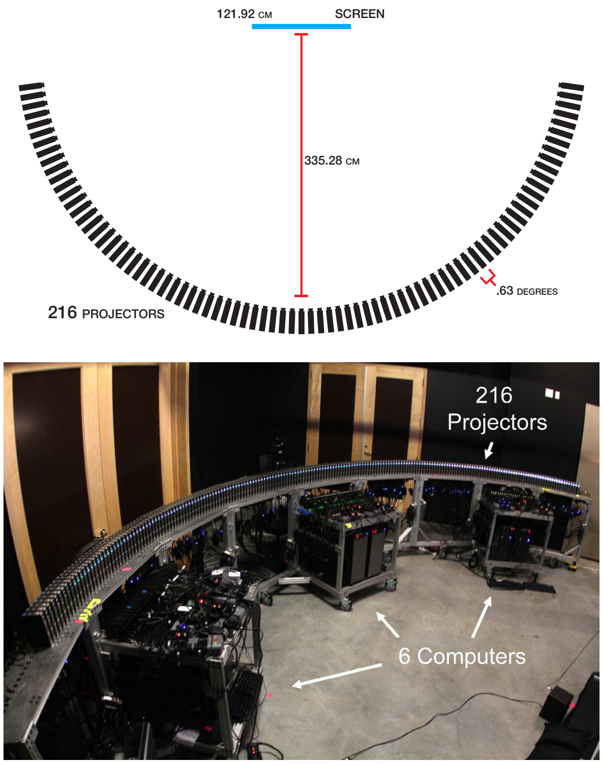

Figure 2. (top) Overhead diagram showing layout of projectors and screen. (bottom) Photograph showing the 6 computers, 72 video splitters, and 216 video projectors used to display the subject.

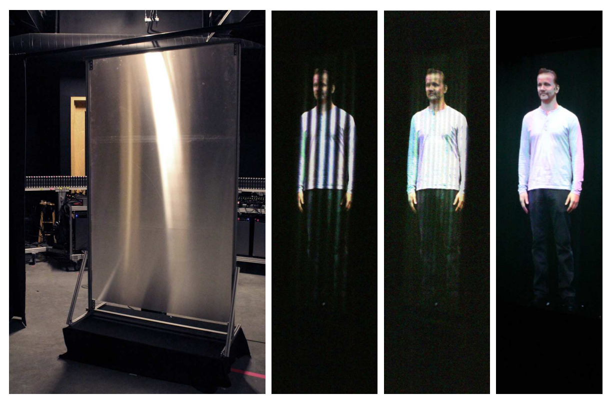
Anisotropic screen   2.5° 1.25° 0.625°

Figure 3. (left) The anisotropic screen scatters light from each projector into a vertical stripe. The individual stripes can be seen if we reduce the angular density of projectors. Each vertical stripe contains pixels from a different projector.

jectors provide a large display depth of field. Objects can be shown within about 0.5 meters of the screen with minimal aliasing. For convincing stereo and motion parallax, the angular spacing between views was also chosen to be small enough that several views are presented within the intraocular distance.

The screen material is a vertically-anisotropic light shaping diffuser manufactured by Luiminit Co. The material scatters light vertically (60 °) so that each pixel can be seen at multiple viewing heights and while maintaining a narrow horizontal blur (1 °). From a given viewer position, each projector contributes a narrow vertical slice taken from the corresponding projector frame. In Figure 3, we compare different projector spacings. If the angle between projectors is wider than the horizontal diffusion, the individual vertical slices can be observed directly. As the angular resolution increases, the gaps decrease in size. Ideally, the horizontal screen blur matches the angular separation between projectors thus smoothly filling in the gaps between the discrete projector positions and forming a seamless 3D image.

To maintain modularity and some degree of portability, the projector arc is divided into three separate carts each spanning 45 degrees of the field of view. We use six computers (two per cart) to render the projector images. Each computer contains two ATI Eyefinity 7870 graphics cards with 12 total video outputs. Each video signal is then divided three ways using a Matrox TripleHead-to-Go video DisplayPort splitter, so that each computer feeds 36 projectors. A single master server computer sends control and synchronization commands to all connected carts.

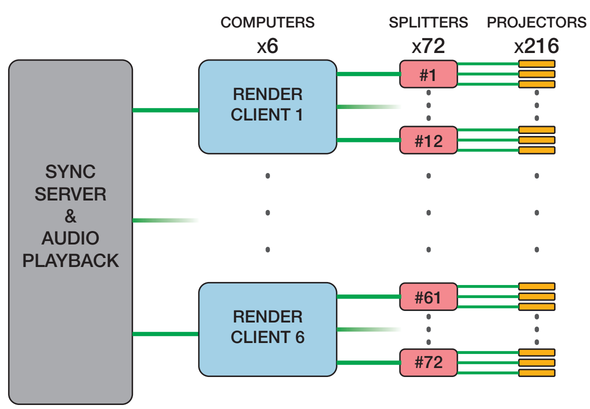

Figure 4. (top) Diagram showing connectivity between computers, splitters, and projectors. The render clients are synchronized and controlled by a single master server.

Ideally all projectors would receive identical HDMI timing signals based on the same internal clock. While adapters are available to synchronize graphics cards across multiple computers (such as Nvidia's G-Sync cards), the Matrox video splitters follow their own internal clocks and the final video signals no longer have subframe alignment. This effect is only noticeable due to the time-multiplexed color reproduction on single chip DLV projectors. Short video camera exposures will see different rainbow striping artifacts for each projector, however this effect is rarely visible to the human eye. Designing a more advanced video splitter that maintains the input video timing or accepts an external sync signal is a subject for future work.

We align the projectors with a per-projector 2D homography that maps projector pixels to positions on the anisotropic screen. We compute the homography based on checker patterns projected onto a diffuse screen placed in front of the anisotropic surface.

5. Light Field Rendering

The main problem in rendering images for the automultiscopic display is that the camera array used to capture the input video sequences is very sparse compared to the projector array. The cameras are placed 6 degrees apart while the angle between the projectors is only 0.625 degrees. It is therefore necessary to synthesize new views for projectors which are placed in-between the cameras. Furthermore rays emitted by each projector continue to diverge as they pass through the anisotropic screen. Rendering to such a display requires the generation of multiple center of projection (MCOP) imagery, as different slices of the projector frame must be rendered according to the varying viewpoints. Previous methods for rendering MCOP imagery on automultiscopic displays have required either high-density light fields [9] or existing geometry [8, 10].

Many techniques have been proposed to reconstruct 3D geometry from multiple cameras, however, this typically requires slower global optimization across all views [21]. Additional depth cameras [3] can accelerate quality or processing rates for playback on augmented reality headsets [1].

An intuitive way to view the recorded data is as a light field parameterized at the projection screen. Light fields are ideal for rendering complex non-convex shapes with a wide variety of materials for skin and clothing, and multiple occlusions from limbs. It also does not require global reconstruction of 3D geometry or perfectly synchronized data. A light field can be rendered by identifying the nearest cameras, and sampling pixel values that correspond to each projector ray [15]. This approach was used by Matusik et al. [17] to generate imagery on a 3D projector array based on a dense camera array. For sparse camera arrays such as the full-body camera array, linear sampling will cause noticeable aliasing (see Figure 6). Instead, we utilize flowed light fields [4] to predict intermediate camera positions. The core idea is to compute pair-wise optical flow correspondences between adjacent cameras as illustrated in Figure 5. All resampling is computed in real-time on the GPU [27], requiring only the original video streams and optical flow offsets.

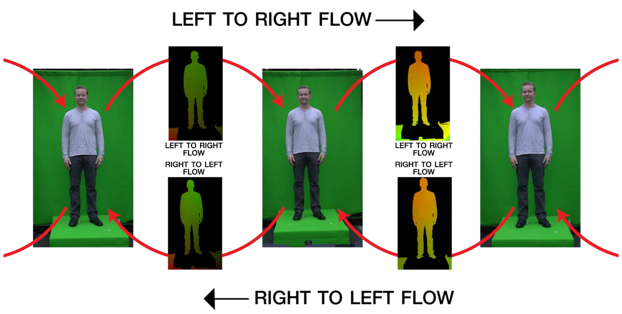

Figure 5. We compute bidirectional optical flow between adjacent cameras.

Figure 6. (left) View generated using bilinear interpolation exhibits aliasing. (center) View generated using optical flow interpolation has sharper edges and less aliasing. (right) Closeups of face.

Flow-based light field reconstruction assumes that optical flow offsets are locally smooth so we can use multiple nearby optical flow offsets to refine each sample camera coordinate. We use the screen surface as proxy geometry to find initial camera coordinates. For each target ray, we find the intersection point with the screen surface, and project it into each adjacent camera. As the projector positions are fixed, this mapping between projector pixels and camera coordinates is constant, and is precomputed as a lookup table.

The optical flow vectors correspond each camera coordinate to a second coordinate in the adjacent camera. In practice, each pair of coordinates references a slightly different surface point on either side of the ideal sample coordinate, since the screen surface does not match the true shape of the subject. We interpolate between the two coordinate pairs to get a single sample point for each camera, weighted by the angular distance between the target ray and each camera. Finally, we interpolate the pixel samples from the two cameras. An illustration of the two optical flow pairs and the interpolated sample positions is shown in Figure 7.

To compensate for rolling shutter effects, we also compute the temporal optical flow, i.e. between sequential frames in each individual video sequence. The temporal flow is then used to add an additional temporal offset to the final sample position weighted by the rolling shutter offset for each row on the sensor and distance from the global time.

We are able to compute light field sampling in real-time in our distributed rendering framework. The optical flow offsets and samples are combined in a pixel shader for each projector ray. For high resolution input, or if more projectors are connected to each host-PC, the algorithm may be limited by GPU upload speeds for the original video files and precomputed optical flow vectors. However, using modern motherboards and GPUs, this is less of a problem. For example, using PCIe 3.0 the maximum bandwidth is in the order of 900MB/s for each lane, and higher end motherboards usually have at least two 16x PCIe ports. As the optical flow is generally smooth, we downsample the flow to quarter resolution, and only upload videos associated with nearby camera positions. If there is insufficient bandwidth to upload both spatial and temporal flows, the input video files can be retimed as a separate preprocess.

6. Natural Language Interface

In a typical use scenario, the digital speaker presents a short introduction or overview to provide context and inspire a followup question and answer session. The audience watches the speaker on the 3D display, and interacts by speaking into a microphone and clicking a push to talk button to tell the system when to listen. We make use of Google API speech recognition to convert the initial audio to text, though the system is compatible with other general purpose recognizers [19]. The corresponding video response is chosen using the NPCEditor dialog manager [14]. The dialog manager is based on cross-language information retrieval, and calculates the relevance of words in a training data of user inputs to words in the set of subject recordings. A total ranking is provided of all possible responses, which is fairly robust to many speech recognition errors. At runtime, if the confidence for a selected response falls below the predetermined threshold, the the subject asks the user to rephrase the question or suggests an alternate topic. Message passing between speech recognizer, dialog manager, and video player is based on the publicly available virtual human toolkit [6].

7. Results

The first full application of this technology was to preserve the experience of in-person interactions with Holocaust survivors. Currently, many survivors visit museums and classrooms to educate, connect and inspire students. There is now an urgency to record these interactive narratives for the few remaining survivors before these personal encounters are no longer possible. Through 3D recording, display, and interacting, we seek to maintain a sense of intimacy and presence, and remain relevant to the future.

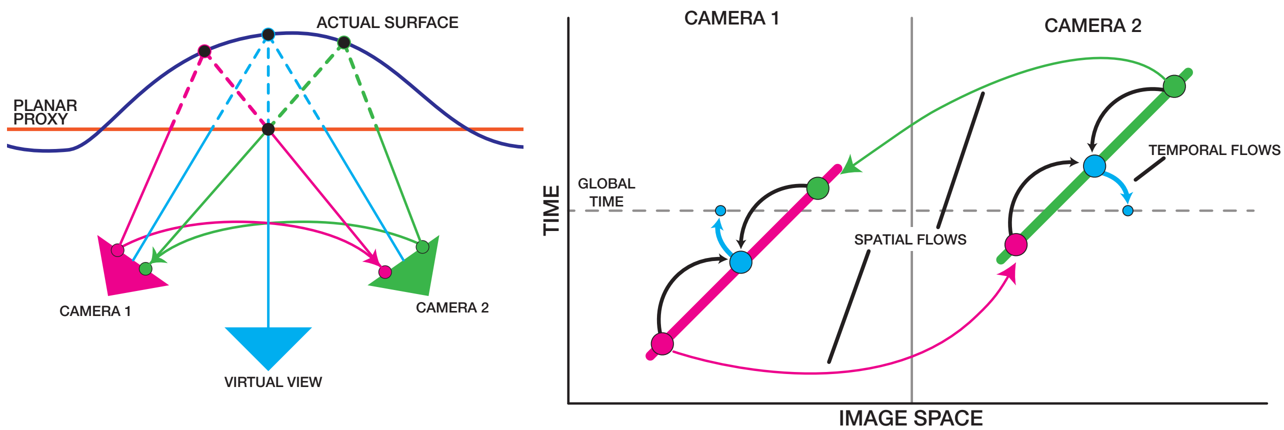

Figure 7. For each point on the screen, we sample the spatial optical flows fields between the two nearest cameras. Each optical flow pair represent a different point on the actual surface. We estimate an intermediate image coordinate using both spatial flow to interpolate between views and temporal flows to offset to a global timeframe. In the right diagram, each image plane is each image plane is represented as a slanted line due camera rolling shutter in each camera.

Figure 8. This is a sampling of four projector frames generated using flowed light field rendering. Each frame appears warped as it corrects for multiple centers of projection and foreshortening.

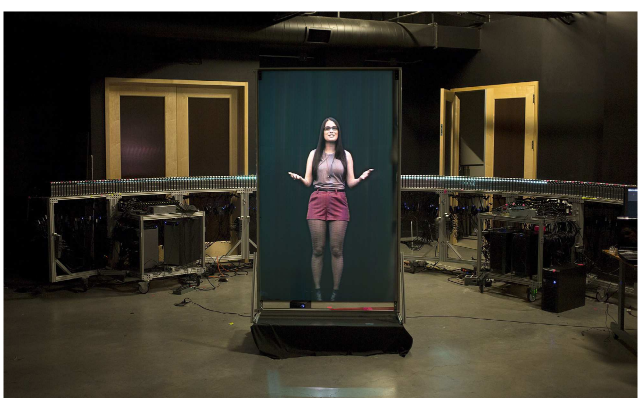

Figure 9. Photograph of subject shown on the automultiscopic projector array.

Our first subject was Pinchas Gutter. Mr Gutter was born in Poland in 1932, lived in a Warsaw ghetto and survived six concentration camps before being liberated by the Russians in 1945. The interview script was based on the top 500 questions typically asked of Holocaust survivors, along with stories catered to his particular life story. The full dataset includes 1897 questions totaling 18 hours of dialog. These questions are linked to 10492 training questions providing enough variation to simulate spontaneous and informative conversations. The interactive system was first demonstrated on an 80-inch 2D video screen at the Illinois Holocaust Museum and Education Center [26]. A user study based based on the 2D playback found that interactive video inspired students to help others, learn about genocide, and feel they could make a difference. Several students noted that that the experience felt like a video teleconference with a live person [12].

The 3D projector array system was tested in a public setting with several age groups. Viewers noted that the 3D display further increased their sense of presence with the survivor. Many older viewers responded on an emotional level. Anecdotally, many visitors act as if the survivor was present, apologizing for their suffering or if they interrupt. The most challenging cases were where other Holocaust survivors asked the system questions reflecting on their own personal experiences. A user study to quantitatively compare the 2D and 3D experiences is a subject for future work.

For this paper, we conducted two additional short interviews with standing subjects. Each interview was limited to 20-30 questions over 2 hours, but still allows for short moderated conversations. Figure 10 shows stereo photographs of all three subjects on the display. The accompanying video shows several 3D conversations with live natural language recognition and playback.

Figure 6 shows a comparison of view interpolation with and without optical flow correction. Optical-flow based interpolation dramatically reduces ghosting between adjacent camera positions. In a few cases, aliasing is still visible on the subject's hands where optical flow struggles to find accurate correspondences. The current optical flow settings sacrifice some quality in order to handle the large video dataset. Each individual optical flow takes 0.5 seconds on a nVidia GTX980. This adds up to 30 seconds per frame to precompute pair-wise optical flows for all camera views.

8. Future Work

Visual quality could be improved by specifically tracking critical regions such as the hands and face. Many body gestures are repeated throughout each interview. This redundancy could be exploited to further improve correspondence algorithms and compress the resulting dataset. Visual quality is also limited by camera and projector resolution. Image quality will improve as smaller, higher resolution projectors become available.

Another area for research, is seamless transitions between video clips. Our system currently uses hard cuts between clips, though dissolves or flowed transitions would be less noticalbe. It is interesting to note that changes in body pose between clips are more apparent in 3D than with traditional 2D video playback.

9. Conclusion

The problem of simulating natural human interaction is a long standing problem in computer science. Our system is able to imitate conversations with real human subjects by selecting from a large database of prerecorded 3D video statements. The interface is intuitive responding to regular spoken questions. We further increase the sense of presence by playing back each video clip in 3D on a dense projector array. We envisage that this system could be used to document a wide range of subjects such as scientists, politicians, or actors with applications in education and entertainment. We are working to generalize the interview framework to other domains, where less prior knowledge exists for each subject. 3D displays such as ours should become increasingly practical in the years to come as the core graphics and image projection components decrease in price and increase in capabilities. Our user interaction and rendering algorithms could also be adapted to other types of 3D displays. Our hope is that this technology will provide a new way for people communicate with each other and the past.

10. Acknowledgements

This work was sponsored by the University of Southern California Office of the Provost, U.S. Air Force DURIP, and U.S. Army Research, Development, and Engineering Command (RDECOM). The content of the information does not necessarily reflect the position or the policy of the US Government, and no official endorsement should be inferred.

We would like to recognize the hard work and inspiration of Steven Smith, Heather Maio, and the USC Shoah Foundation throughout concept development, interviewing, and production. Anton Leiuski, Ron Artstein, Kalliorroi Georgila, and David Traum provided valuable assistance with natural language integration. We would like to thank Kathleen Haase, Bill Swartout, Randall Hill, and Randolph Hall for their support and assistance;

This work was made possible by generous donations from private foundations and individuals. We are extremely grateful to the Pears Foundation, Alan Shay, Lucy Goldman, and the Wolfson Foundation for their support. Thanks to our interview subjects Morgan Spurlock and Cara Santa Maria. Special thanks to Pinchas Gutter for sharing his story, and for his tireless efforts to educate the world about the Holocaust.

References

[1] Microsoft research holoportation, 2016 (accessed: 2016-4-29). http://research.microsoft.com/en-us/projects/holoportation/. 4

[2] R. Artstein, A. Leuski, H. Maio, T. Mor-Barak, C. Gordon, and D. Traum. How Many Utterances Are Needed to Support Time-Offset Interaction? In *Proceedings of FLAIRS 28*, pages 144–149, Hollywood, FL, May 2015. AAAI Press. 2

[3] A. Collet, M. Chuang, P. Sweeney, D. Gillett, D. Evseev, D. Calabrese, H. Hoppe, A. Kirk, and S. Sullivan. High-quality streamable free-viewpoint video. *ACM Trans. Graph.*, 34(4):69:1–69:13, July 2015. 4

[4] P. Einarsson, C.-F. Chabert, A. Jones, W.-C. Ma, B. Lamond, T. Hawkins, M. Bolas, S. Sylwan, and P. Debevec. Relighting human locomotion with flowed reflectance fields. In *Rendering Techniques 2006: 17th Eurographics Symposium on Rendering*, pages 183–194, June 2006. 1, 4

[5] J. Gustafson, N. Lindberg, and M. Lundeberg. The august spoken dialogue system. In *In Proceedings of Eurospeech'99*, pages 1151–1154, 1999. 2

[6] A. Hartholt, D. Traum, S. C. Marsella, A. Shapiro, G. Stratou, A. Leuski, L.-P. Morency, and J. Gratch. All Together Now: Introducing the Virtual Human Toolkit. In *13th International Conference on Intelligent Virtual Agents*, Edinburgh, UK, Aug. 2013. 2, 5

[7] H. E. Ives. The projection of parallax panoramagrams. *Journal of the Optical Society of America*, 21(7):397–409, 1931. 3

[8] A. Jones, M. Lang, G. Fyffe, X. Yu, J. Busch, I. McDowall, M. Bolas, and P. Debevec. Achieving eye contact in a one-to-many 3d video teleconferencing system. In *ACM Transactions on Graphics (TOG)*, volume 28, page 64. ACM, 2009. 4

[9] A. Jones, I. McDowall, H. Yamada, M. Bolas, and P. Debevec. Rendering for an interactive 360 light field display. *ACM Transactions on Graphics (TOG)*, 26(3):40, 2007. 4

[10] A. Jones, K. Nagano, J. Liu, J. Busch, X. Yu, M. Bolas, and P. Debevec. Interpolating vertical parallax for an autostereoscopic 3d projector array. volume 9011, 2014. 3, 4

[11] M. Kawakita, S. Iwasawa, M. Sakai, Y. Haino, M. Sato, and N. Inoue. 3d image quality of 200-inch glasses-free 3d dis-

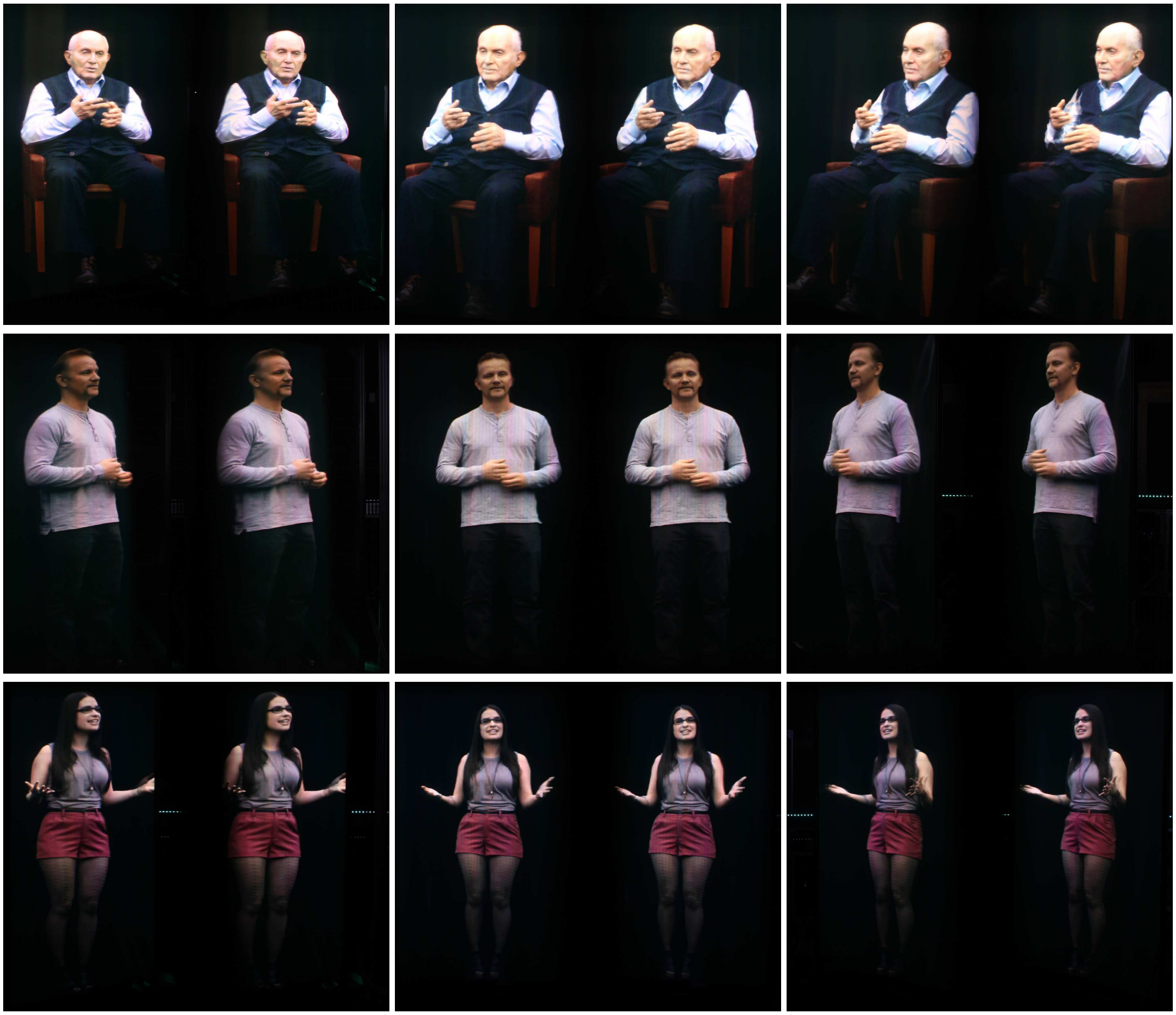

Figure 10. Stereo photograph of subjects on the display from three positions, left-right reversed for cross-fused stereo viewing.

play system. *SPIE Stereoscopic Displays and Applications XXIII*, 8288, 2012. 3

[12] K. Kim. New dimensions in testimony: Findings from student pilots. Technical report, USC Shoah Foundation, August 2015. 6

[13] A. Leuski, J. Pair, D. Traum, P. J. McNerney, P. G. Georgiou, and R. Patel. How to Talk to a Hologram. In E. Edmonds, D. Riecken, C. L. Paris, and C. L. Sidner, editors, *Proceedings of the 11th International Conference on Intelligent User Interfaces*, pages 360–362, Sydney, Australia, 2006. ACM Press New York, NY, USA. 2, 3

[14] A. Leuski and D. Traum. NPCEditor: Creating Virtual Human Dialogue Using Information Retrieval Techniques. *AI Magazine*, 32(2):42–56, July 2011. 5

[15] M. Levoy and P. M. Hanrahan. Light field rendering. In *Proceedings of ACM SIGGRAPH 96*, Computer Graphics Proceedings, Annual Conference Series, pages 31–42, 1996. 4

[16] D. Marinelli and S. Stevens. Synthetic interviews: The art of creating a ”dyad” between humans and machine-based characters. In *Proceedings of the Sixth ACM International Conference on Multimedia: Technologies for Interactive Movies*, MULTIMEDIA ’98, pages 11–16, New York, NY, USA, 1998. ACM. 2

[17] W. Matusik and H. Pfister. 3d tv: a scalable system for real-time acquisition, transmission, and autostereoscopic display of dynamic scenes. In *ACM Transactions on Graphics (TOG)*, volume 23, pages 814–824. ACM, 2004. 3, 4

[18] W. J. McKeachie et al. Teaching and learning in the college classroom. a review of the research literature (1986) and november 1987 supplement. 1987. 1

[19] F. Morbini, K. Audhkhasi, K. Sagae, R. Artstein, D. Can, P. G. Georgiou, S. Narayanan, A. Leuski, and D. Traum. Which ASR should I choose for my dialogue system? In *SIGDIAL*, Metz, France, Aug. 2013. 5

[20] T. Rodriguez, A. C. de Leon, B. Uzzan, N. Livet, E. Boyer, F. Geffray, T. Balogh, Z. Megyesi, and A. Barsi. Holographic and action capture techniques. In *ACM SIGGRAPH 2007 emerging technologies*, SIGGRAPH '07, New York, NY, USA, 2007. ACM. 3

[21] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 1, pages 519–528, June 2006. 4

[22] E. Sloss and A. Watzman. Carnegie Mellon's Entertainment Technology Center conjures up Benjamin Franklin's ghost. Press release, Carnegie Mellon Media Relations, June 28 2005. 2, 3

[23] J. Steinmeyer. *The Science behind the Ghost: A Brief History of Pepper's Ghost*. Hahne, 1999. 3

[24] W. Swartout, D. Traum, R. Artstein, D. Noren, P. Debevec, K. Bronnenkant, J. Williams, A. Leuski, S. Narayanan, D. Piepol, H. C. Lane, J. Morie, P. Aggarwal, M. Liewer, J.-Y. Chiang, J. Gerten, S. Chu, and K. White. Ada and Grace: Toward Realistic and Engaging Virtual Museum Guides. In *Proceedings of the 10th International Conference on Intelligent Virtual Agents (IVA 2010)*, Philadelphia, PA, Sept. 2010. 2

[25] D. Traum, P. Aggarwal, R. Artstein, S. Foutz, J. Gerten, A. Katsamanis, A. Leuski, D. Noren, and W. Swartout. Ada and Grace: Direct Interaction with Museum Visitors. In Y. Nakano, M. Neff, A. Paiva, and M. Walker, editors, *The 12th International Conference on Intelligent Virtual Agents (IVA)*, volume 7502 of *Lecture Notes in Artificial Intelligence*, pages 245–251, Santa Cruz, CA, Sept. 2012. 2

[26] D. Traum, A. Jones, K. Hays, H. Maio, O. Alexander, R. Artstein, P. Debevec, A. Gainer, K. Georgila, K. Haase, K. Jungblut, A. Leuski, S. Smith, and W. Swartout. *Interactive Storytelling: 8th International Conference on Interactive Digital Storytelling, ICIDS 2015, Copenhagen, Denmark, November 30 - December 4, 2015, Proceedings*, chapter New Dimensions in Testimony: Digitally Preserving a Holocaust Survivor's Interactive Storytelling, pages 269–281. Springer International Publishing, Cham, 2015. 6

[27] M. Werlberger, T. Pock, and H. Bischof. Motion estimation with non-local total variation regularization. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, San Francisco, CA, USA, June 2010. 4

[28] S. Yoshida, M. Kawakita, and H. Ando. Light-field generation by several screen types for glasses-free tabletop 3d display. In *3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON), 2011*, pages 1–4. IEEE, 2011. 3