
TI-Photograph: a Tele-Immersive Photograph System for Distributed Parents and Children

Meiyu Huang

Beijing Key Laboratory of Mobile Computing and Pervasive Device, Institute of Computing Technology, Chinese Academy of Sciences, University of Chinese Academy of Sciences, Beijing, China
huangmeiyu@ict.ac.cn

Linglin Yin

Beijing Key Laboratory of Mobile Computing and Pervasive Device, Institute of Computing Technology, Chinese Academy of Sciences, University of Chinese Academy of Sciences, Beijing, China
yinlinglin@ict.ac.cn

Yiqiang Chen

Beijing Key Laboratory of Mobile Computing and Pervasive Device, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China
yqchen@ict.ac.cn

Wen Ji

Beijing Key Laboratory of Mobile Computing and Pervasive Device, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China
jiwen@ict.ac.cn

Abstract

With the social development, the demand of a natural remote communication platform for distributed families has greatly emerged. This work presents a Tele-Immersive Photograph system (TI-Photograph), which allows children and their remote parents to take pictures together in a virtual space with interactive behaviors. First, we propose a robust video object cutout method to segment the video of the child and the remote parent from their background surroundings. Second, we introduce a user behavioral intention driven video composition method to adaptively merge the segmented videos of the users into a same shared background customized by natural gesture interaction. Then we create an illusion that the child and the parent are immersed in the same environment. Experimental results demonstrate that the visual appearance in a shared interactive environment provides closer and efficient communication between remote family members.

Author Keywords

Parents and children, tele-immersive photograph system, video object cutout, video composition, gesture interaction.

ACM Classification Keywords

H.5.2 [User Interfaces]: Prototyping; I.4.9 [Image Processing and Computer Vision]: Applications

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s). Copyright is held by the author/owner(s).
UbiComp'13 Adjunct, September 8–12, 2013, Zurich, Switzerland.
ACM 978-1-4503-2215-7/13/09.

<http://dx.doi.org/10.1145/2494091.2494174>

Introduction

In recent years, the number of migrant workers increases rapidly with the globalization in economy, which becomes the main factor for the separation between children and their parents. Long-period and long-distant separation makes frequent face-to-face communication difficult. Recently, the research in [1] indicates that communication with children remains quite hard over current video conferencing systems due to only focusing on the conversation itself. The research in [2] has demonstrated that better communication with children may be achieved by structuring video conferencing interfaces with shared activities. The implicit under the study is the idea of playfulness as a means for engagement in distant communication. Motivated by these work, we present a novel implementation of a tele-immersive photograph system that enables remote participants to interact with each other in a shared virtual environment (Figure 1).



Figure 1: A father is “petting” his daughter’s head in TI-Photograph.

Tele-Immersive Photograph System

Aiming at providing tele-immersive photographing for distributed parents and children, TI-Photograph

incorporates video object cutout, video composition and gesture interaction techniques, as shown in Figure 2.

Robust video object cutout

Video object cutout is used to segment the video of the users (foreground) from their background scenes. We propose a robust video object cutout method based on Kinect. The method obtains an initial foreground mask through motion capture based on depth information, and takes a coarse-to-fine post processing procedure to eliminate mis-segmentation in depth occlusion area as well as on the boundary. The processing contains two major steps: 1) foreground hole detection, which identifies misclassified foreground holes due to depth loss based on pixel-level non-parametric background color models and piecewise-linear contrast models; 2) object boundary refining, which relabels each pixel among the boundary through making reliability fusion of decisions based on color, edge and temporal cues, eliminating mis-segmentation and flickering artifact caused by inaccurate and unstable depth estimation. Specifically, the decisions are derived from combined K-means and GMM(Gaussian Mixture Model) based local color models, four parameters S-function based local edge model and first-order HMM(Hidden Markov Model) based temporal model.

Immersive video composition

Video composition is to merge the video of the users (object videos) and the virtual background. A user behavioral intention driven video composition method is introduced to provide synchronous interaction between remote users. The approach automatically detects the intention of interactive users, and then adaptively composites videos of participants with intended interactive behaviors while keeping the geometry and perspective

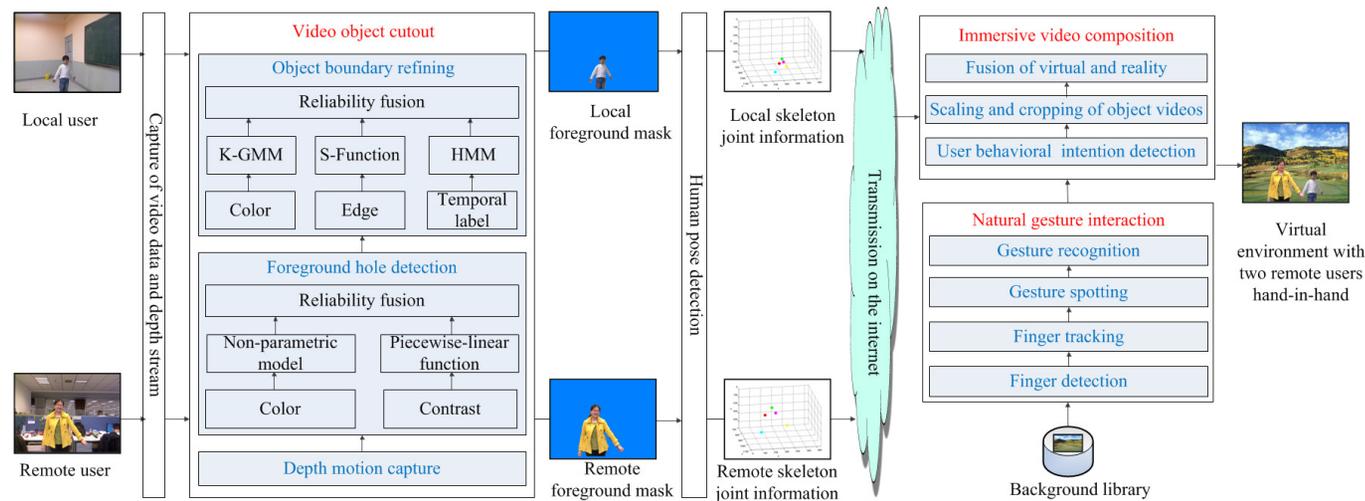


Figure 2: The implementation of TI-Photograph.

consistency. The method predicts users' behavioral intention based on both users' skeleton joint information. Then, according to the predicted behavioral intention, the method adaptively scales and crops two object videos. Finally, the method merges two coordinated objects videos into a shared background with reasonable position based on human being's visual habitats.

Natural gesture interaction

Gesture interaction is in charge of facilitating the customization of the virtual background and triggering of photographing. A user adaptive gesture interaction method is proposed to ease of the use for tele-immersive photograph experience. The method integrates accurate finger detection based on depth information, robust finger tracking using the improved Kalman filter, gesture spotting based on transition features, and gesture

recognition using the ELM (Extreme Learning Machine) classifier.

Evaluation

We implemented TI-Photograph using C++ on a pair of 2.7GHz computers. Experimental results show that TI-Photograph provides life-like composition video of VGA size at an up to 15 FPS speed with an average CPU load of 40%. To measure the usability of TI-Photograph, we test it with 20 parents and children pairs as well as a post-test questionnaire survey on six aspects (Figure 3). Those children are between 6 and 12 years old, and their parents are between 30 and 41 year old. 5 of the parents are high-educated engineers who leave their children working in different cities. The remain of the parents are farmer workers with little educations, who also work in

remote cities. All families in our experiments have used the conventional video conferencing systems.

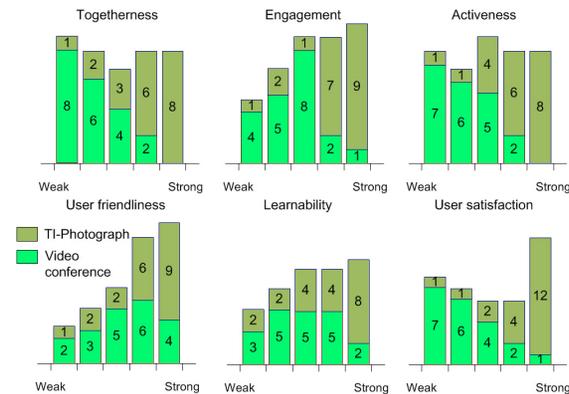


Figure 3: Results of post-test questionnaire asking participants to rate the usability of TI-Photograph and video conferencing with 5-point likert-scale questions.

The investigation results of the questionnaire show that TI-Photograph achieves a better performance than conventional video conferencing. TI-Photograph makes distributed parents and children feel closer to each other. Most Participants express a willing of togetherness through dialogs and actions such as “touching” each other. Also, TI-Photograph encourages engagement because most of the children are immersed in the shared activity of taking pictures with their remote parents as well as the magic functionality of gesture interaction. Moreover, TI-Photograph enhances communication between parents and children because it inspires sharing of experiences and desires through discussion of which background to pick. Well-educated parents tend to pick the images of interest places to share their travel experience. While children of farmer workers prefer to pick

the images with their desire, such as taking a plane, interacting with cartoons and playing with toys. In addition, benefited from the gesture interaction and heuristic user interface, TI-Photograph is quite user friendly and very easy to learn to use. Even participants with little education use the system flexibly. TI-Photograph also provides satisfying user experience. Most participants give positive remarks to TI-Photograph (e.g., “Interesting system”, “I enjoyed”, “Amazing”).

Conclusion

We design a tele-immersive photograph system to satisfy the remote interactive requirement of distributed parents and children. The highlight of the system is a robust video object cutout method with immersive video composition and natural gesture interaction techniques. Experiments show the proposed system achieves high-quality performance and high-interactive user experience. TI-Photograph promotes the feeling of togetherness and encourages experiences sharing and closer communication between children and their remote parents.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (61001194), and by the Beijing Natural Science Foundation (4112056, 4122078).

References

- [1] Ames, M. G., Go, J., Kaye, J., and Spasojevic, M. Making love in the network closet: the benefits and work of family videochat. In *Proc. of ACM CSCW* (2010), 145–154.
- [2] Follmer, S., Ballagas, R., Raffle, H., Spasojevic, M., and Ishii, H. People in books: using a flashcam to become part of an interactive book for connected reading. In *Proc. of ACM CSCW* (2012), 685–694.