

Toward Long Distance Tabletop Hand-Document Telepresence



Chelhwon Kim¹, Patrick Chiu¹, Joseph de la Pena²,

Laurent Denoue¹, Jun Shingu², Yulius Tjahjadi¹

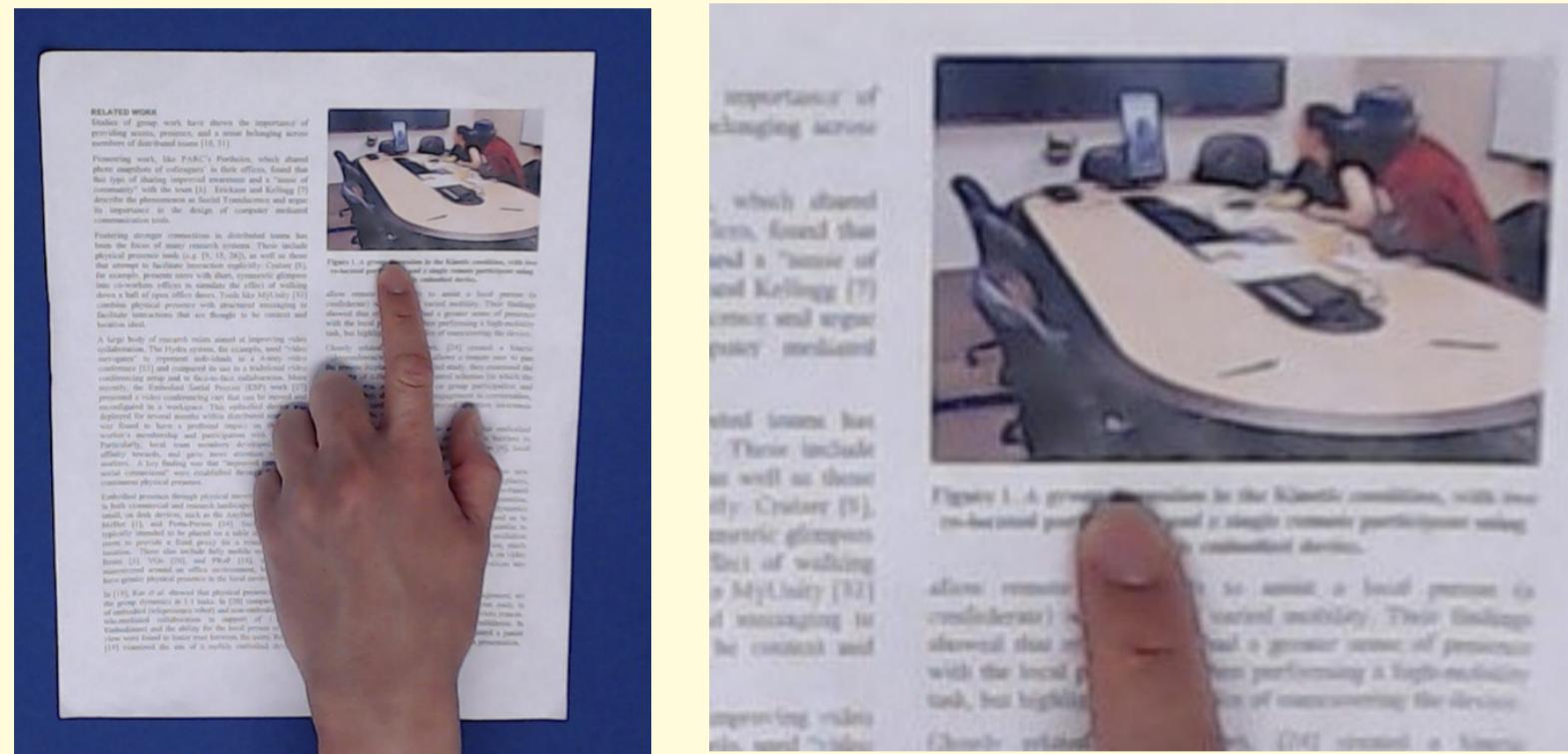
¹FX Palo Alto Laboratory, ²Fuji Xerox



Introduction

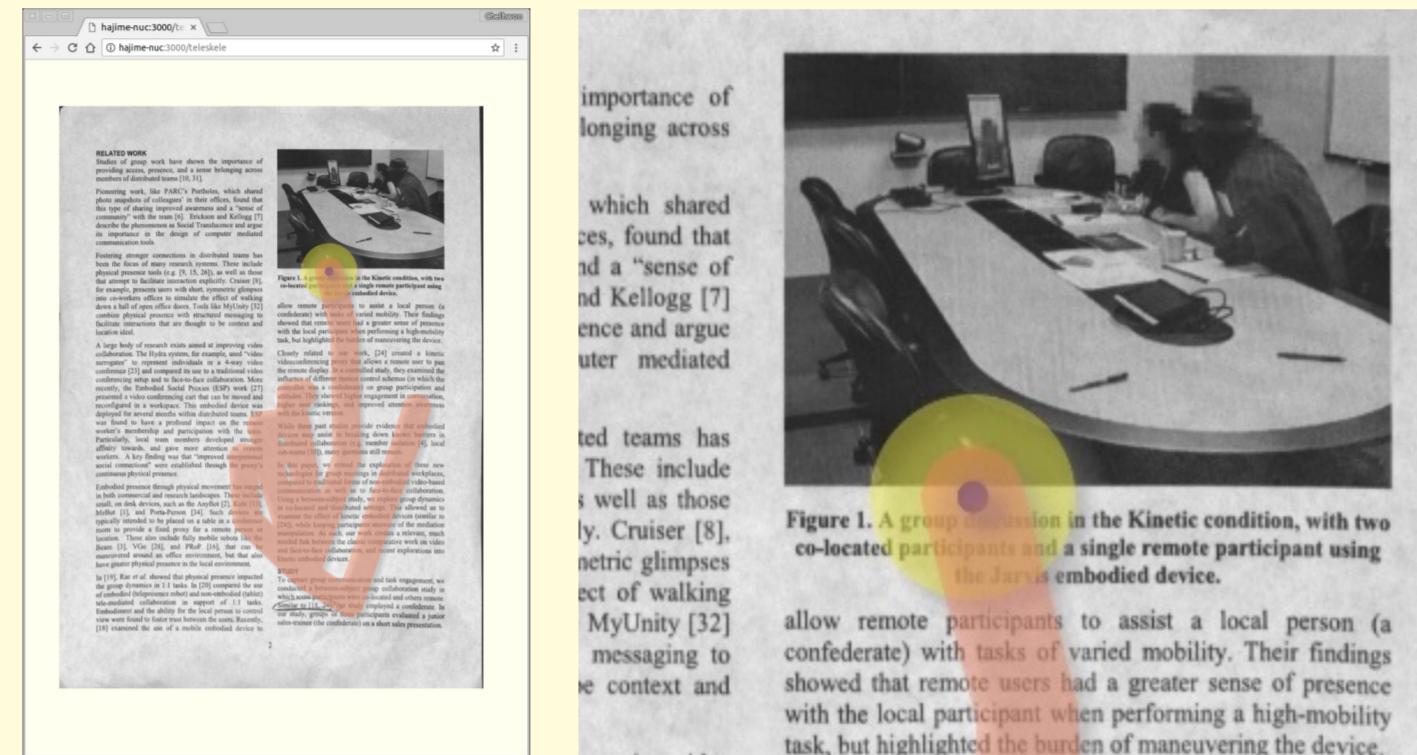
- Discussion over a document in a telepresence: showing the user's hand position on the document
- Problem:** How to capture and transmit the hand movements efficiently with high resolution document images

Standard video conferencing



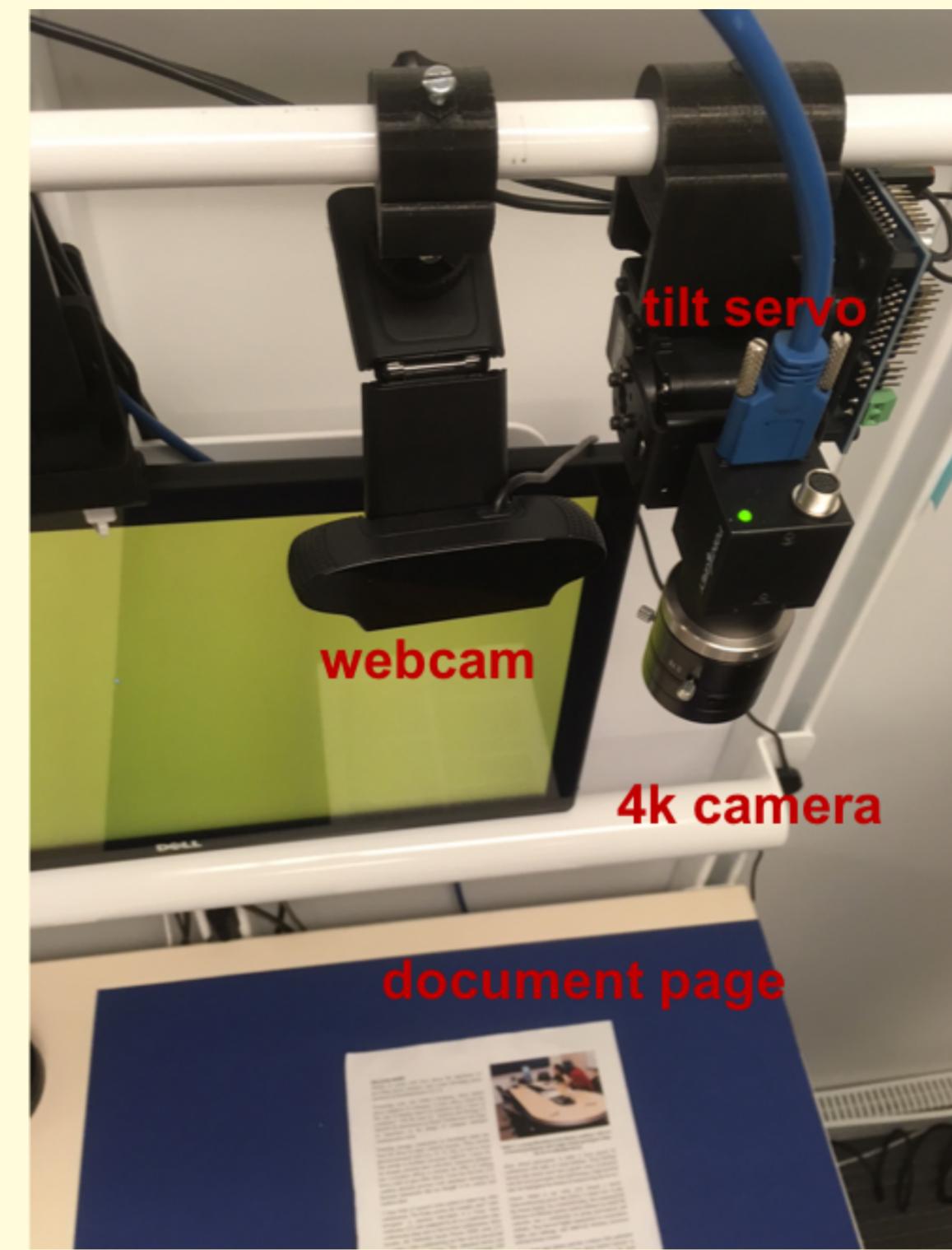
- Too low-res to read document
- Occlusion by hand

Our system



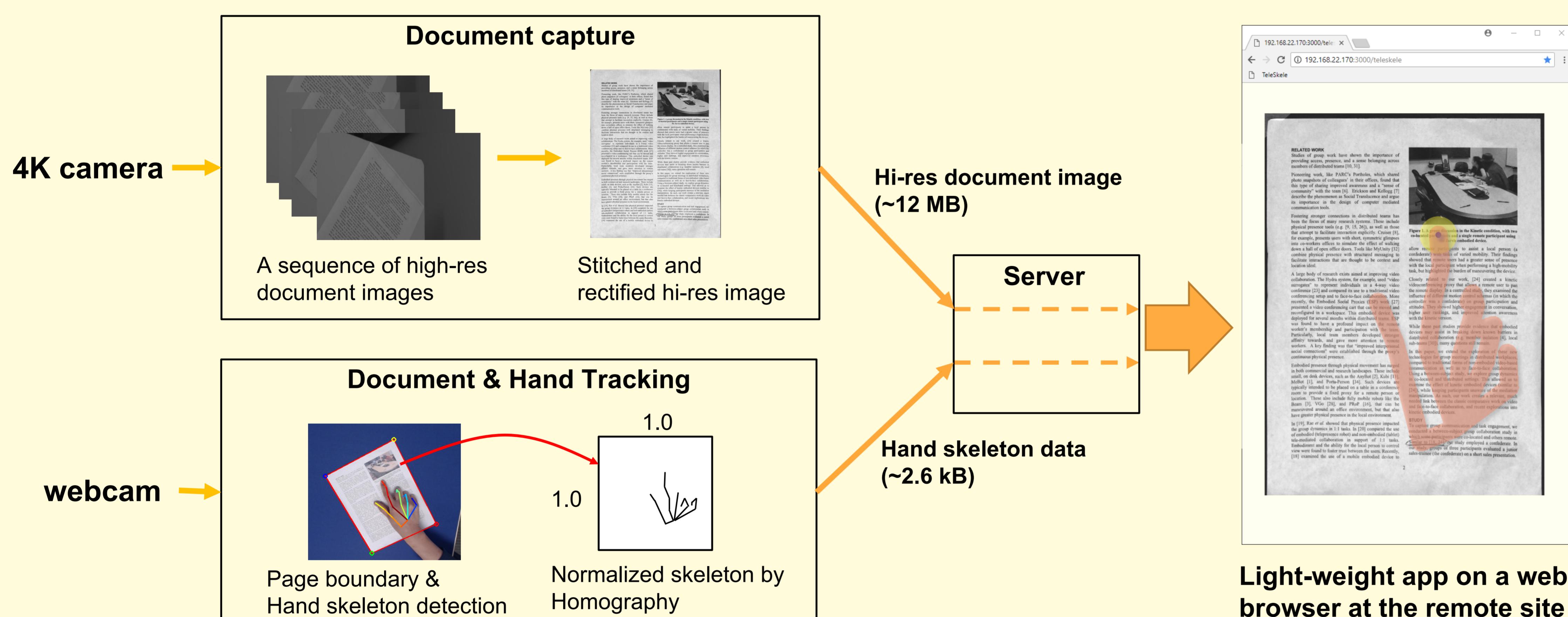
- Hi-res document capture
- Hand skeleton: less occlusion
- Two-channel data transmission at different rate

System Setup



- 4K camera for hi-res document capture [1]
- Webcam for hand tracking (Deep Learning based hand pose estimation [2]) & Document page tracking
- Document page on tabletop

Proposed Method

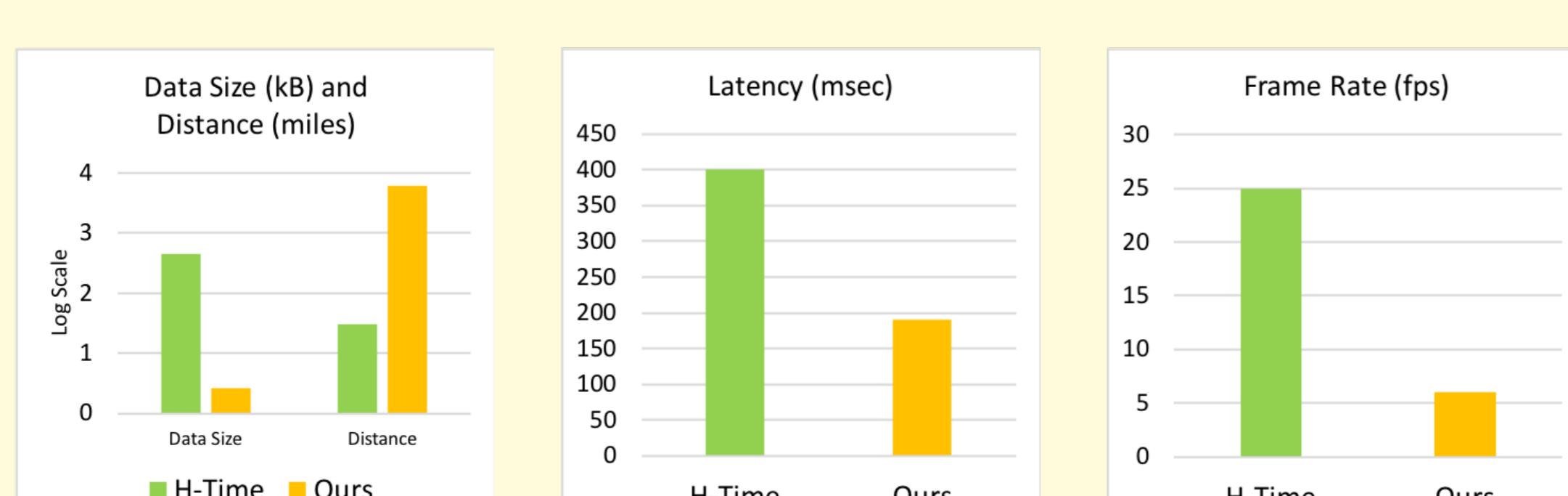


Data transmission with two channels

- Document page image (~12MB): only when the page has changed
- Hand skeleton data(~2.6kB): 5~6 fps

Evaluation

- Two distributed environment setups over long distances**
 - Palo Alto/US ~ Yokohama/Japan (~ 5100 miles)
 - Palo Alto/US ~ Verona/Italy (~ 6000 miles)
- Hand Skeleton Data (~ 2.6kB) Latency**
 - Palo Alto ~ Yokohama: ~ 180 msec
 - Palo Alto ~ Verona: ~ 191 msec
- High-res Document Data (~ 12MB) Latency**
 - Palo Alto ~ Yokohama: ~ 26.53 sec
 - Palo Alto ~ Verona: ~ 28.48 sec
- Our system can achieve relatively low latency over a long distance since we transmit the small hand skeleton data**
- Comparison with H-TIME [3] system that transmits large mesh data over shorter distance**



Conclusion & Future Work

- We presented a novel system for hand-document telepresence with high resolution document capture and hand skeleton tracking, and with two separate channels for transmitting these data.
- We evaluated our system over long distances, and compared our system to a tele-immersive system that was tested over much shorter distances.
- For future work, one direction is to evaluate our system on user level tasks over long distances, and to use our system in realistic situations with remote colleagues.

Related Work

- [1] 4K document capture: (Kim et al., *DocEng'15*)
- [2] Deep learning based hand tracking: (Zimmermann et al., *ICCV'17*)
- [3] H-Time: Haptic-enabled tele-immersive musculoskeletal examination: (Tian et al., *ACM Multimedia'17*)