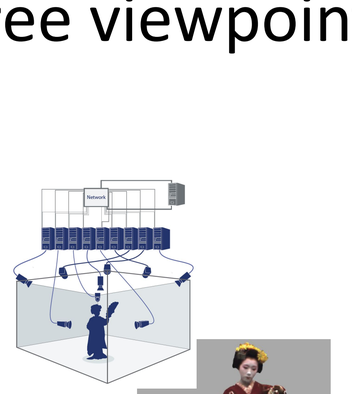
Action History Volume for Spatiotemporal Editing of 3D Video in Multi-party Interaction Scenes

QUN SHI SHOHEI NOBUHARA TAKASHI MATSUYAMA (Kyoto University)

Research Goal

Spatiotemporally synchronized multi-party interaction 3D Video synthesis for free viewpoint browsing, from separately captured data

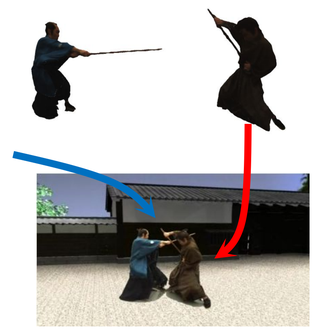


3D Video

Multi-view video capture

Mutual occlusion

Separate capture then synthesis scheme



Free viewpoint browsing

"Fake action" strategy unavailable

Frame wise 3D reconstruction

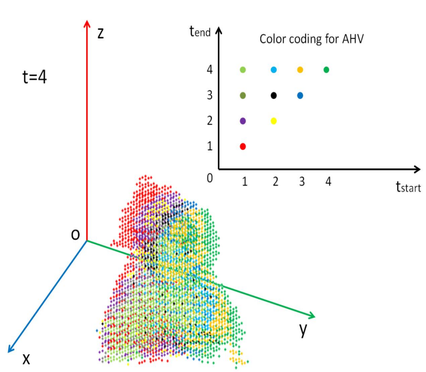
Unstructured mesh data

- *Difficulties:*
  - *1 Spatial temporal un-synchronization*
  - *2 View-independent consistency*
  - *3 Kinematic structure based conventional methods not applicable*

Approach

- Action History Volume for action representation

$$v_\tau(x,y,z,t) = \begin{cases} (t_{\text{start}}, t_{\text{end}}) & \text{if } \bigcup_{s=0}^{\tau-1} D(x,y,z,s), \\ \text{empty} & \text{otherwise,} \end{cases}$$



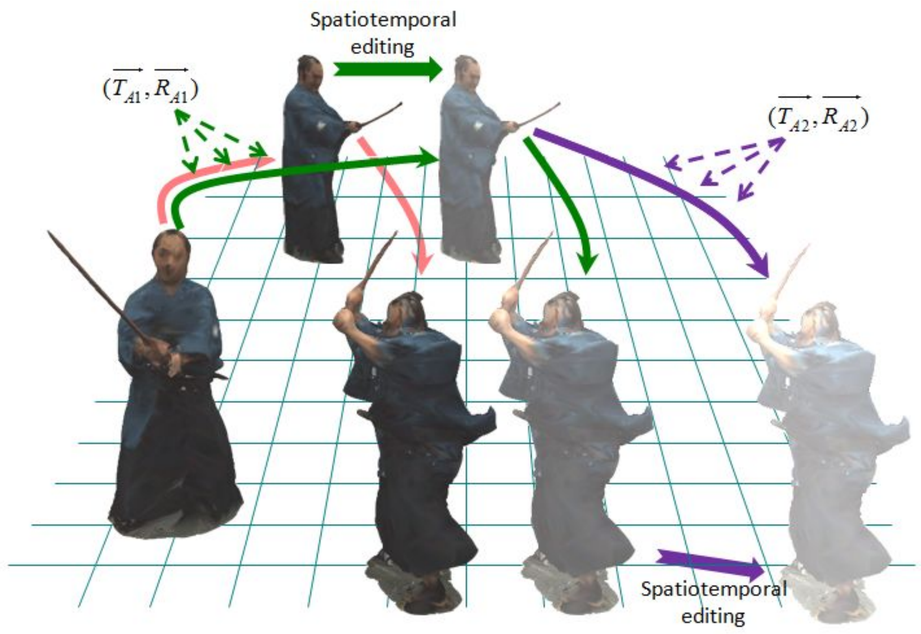
- Action History Volume for interaction representation

•Label
- start surface of motion
+ end surface of motion
0 lateral surface

AHV based Interaction Dictionary

Label Combination	+ & -	0 & 0	+ & +
Description	End surface contacts with start surface	Contact on lateral surfaces	Contact on end surfaces
Label Combination	**- & -**	**+ & 0**	**- & 0**
Description	Contact on start surfaces	End surface contacts with lateral surface	Start surface contacts with lateral surface

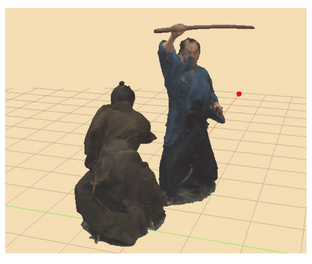
- AHV based multi-party interaction scene editing algorithm
  - 1 Data Segmentation
    Action sequence = Key segments + Transitional segments
  - 2 AHV computation in each key segment
  - 3 AHV based constraint definition
  - 4 Intra-key segment editing using constraint satisfaction
  - 5 Inter-key segment optimization for the entire sequence

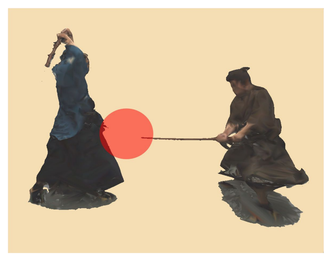


$$E = \sum_{i=1}^{N-1} (|\overrightarrow{T_{Ai}}|^2 + |\overrightarrow{T_{Bi}}|^2) + \sum_{i=1}^{m-1} \left\{ \gamma f_a(\overline{C_i}) + (1-\gamma) w_i f_v(\overline{C_i}) \right\}$$

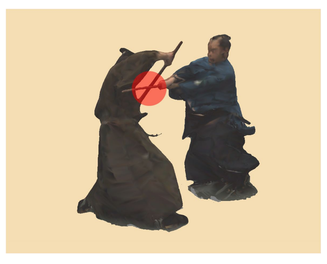
Experiment Results

Before Editing





After Editing



Conclusion

By applying the AHV based spatiotemporal motion editing we can successfully synthesize natural multi-party interaction 3D Video scenes from separately captured data, while protecting well the original motion dynamics for each object.

- Future Work
  - 1 Introduce skeletal based editing.
  - 2 Extend the proposed method onto three or more-party situations.