Affordable content creation for free-viewpoint video and VR/AR application

Pagés et al.

Journal of Visual Communication and Image Representation

Yu-Ming Lai

2019/1/31

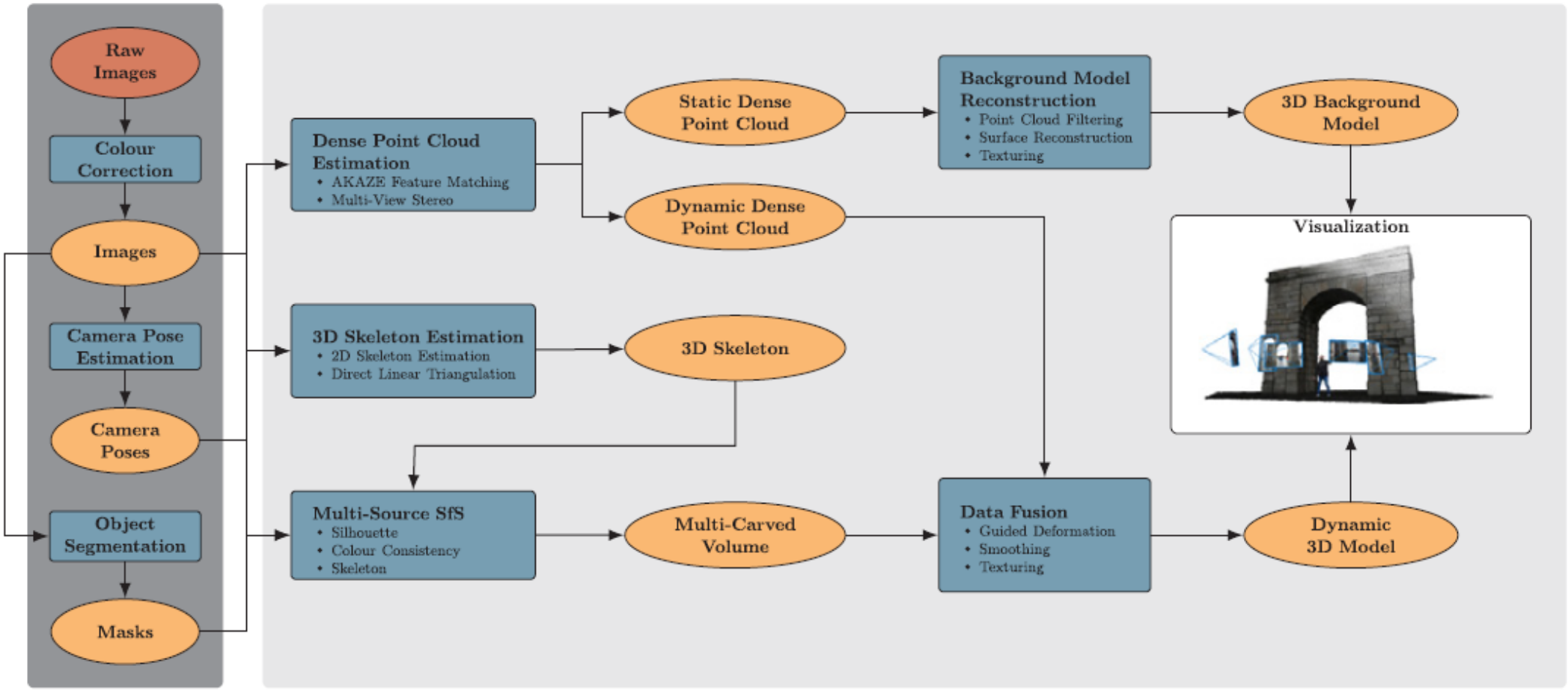
Introduction

- Empower the viewer to become the director can bring them closer to the action and give them a sense of immersion
  - 3D geometry information of the scene is necessary
  - Affordable 3D content creation tools will be necessary to satisfy the needs of these emerging consumer markets
- Free-Viewpoint Video (FVV) allows users to freely navigate within a recorded scene and select any viewpoint at any moment (6DoF) in time

Contributions

1. An end-to-end system to create and process FVV sequences that can be and visualized either in VR/AR or using a view-synthesis mode
2. A lightweight system that produces high quality content from a limited number of commodity cameras
3. Multi-source shape-from-silhouette (MS-SfS) and efficient fusion of different geometry data

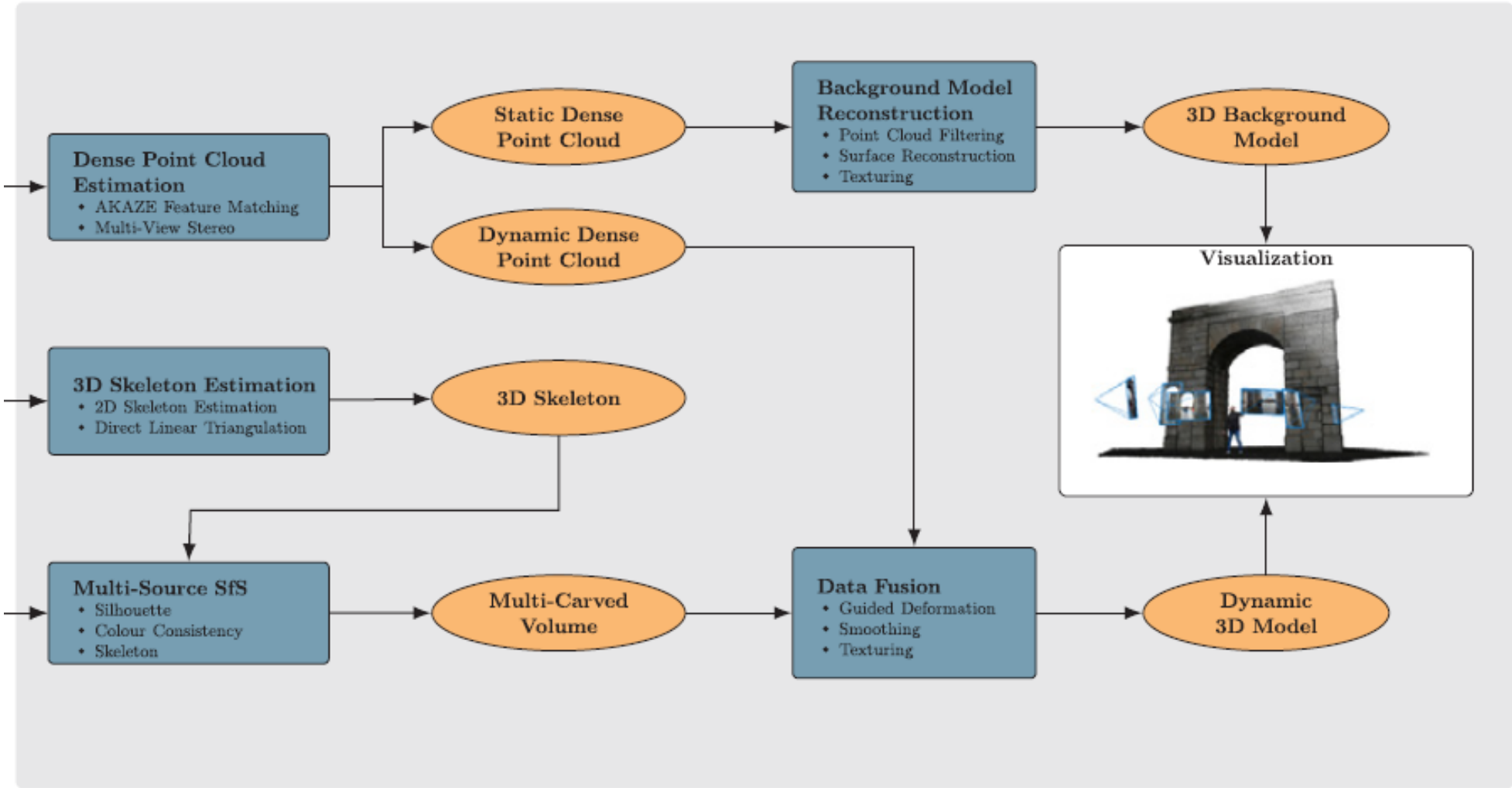
System Overview



Pre-processing

- Color correction
  - Transform the target image color distribution to match a palette image
  - Gaussian Mixture Models
- Camera pose estimation
  - Estimate calibration using SfM at certain time intervals and interpolate frames in between
  - Compute camera pose with EPnP algorithm
- Foreground segmentation
  - Use the up-to-date OSVOS approaches

Scene Reconstruction



Point Cloud Generation

- Use A-KAZE with multi-scale Retinex image enhancement to make feature detection easier
- Final point cloud is fed to the MVS system proposed by Schoenberger et al.

(c) SIFT+PMVS (d) Ours

Background Model

- Remove noisy points using a statistical outlier removal approach
- Compute final mesh using Poisson surface reconstruction (PSR)

3D Skeleton Estimation

- Estimate a 3D skeleton by triangulating a set of 2D skeletons detected in the input images
  - A CNN to detect and associate the 2D joints in the images
  - Use foreground mask to filter the unwanted skeletons
- Get final 3D joint coordinates by minimizing a set over determined linear triangulation problems

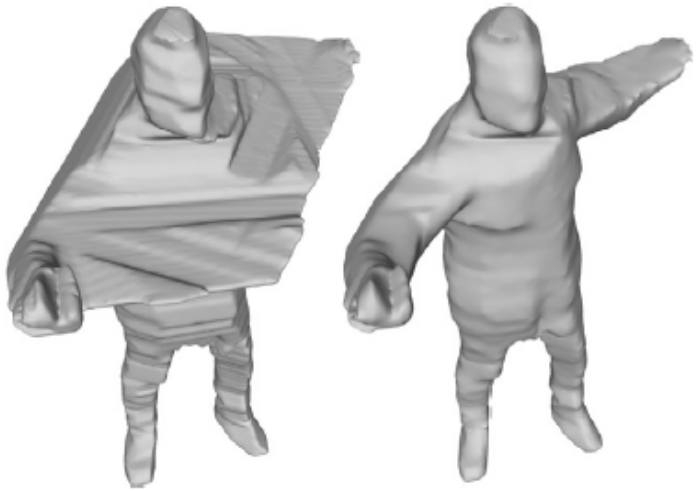
Multi-Source Shape-from-Silhouette (MS-SfS)

- Apply space carving techniques to handle the concavities and occlusions
  - Carving function:

$$\mathcal{F}_c(\mathbf{p}) = \phi_{sil}(\mathbf{p}, \mathcal{M}) \cdot \phi_{cc}(\mathbf{p}, \mathcal{I}) \cdot \phi_{skel}(\mathbf{p}, s)$$

A set of voxel — Foreground masks — Registered images — Skeletons

  - **Silhouette score** (ϕ_{sil}): binary factor
  - **Color consistency score** (ϕ_{cc}): the standard deviation σ of the hue value that $\mathbf{p}$ projected onto each image
  - **Skeleton score** (ϕ_{skel}): increase with the drop of the distance between $\mathbf{p}$ and its closest bone

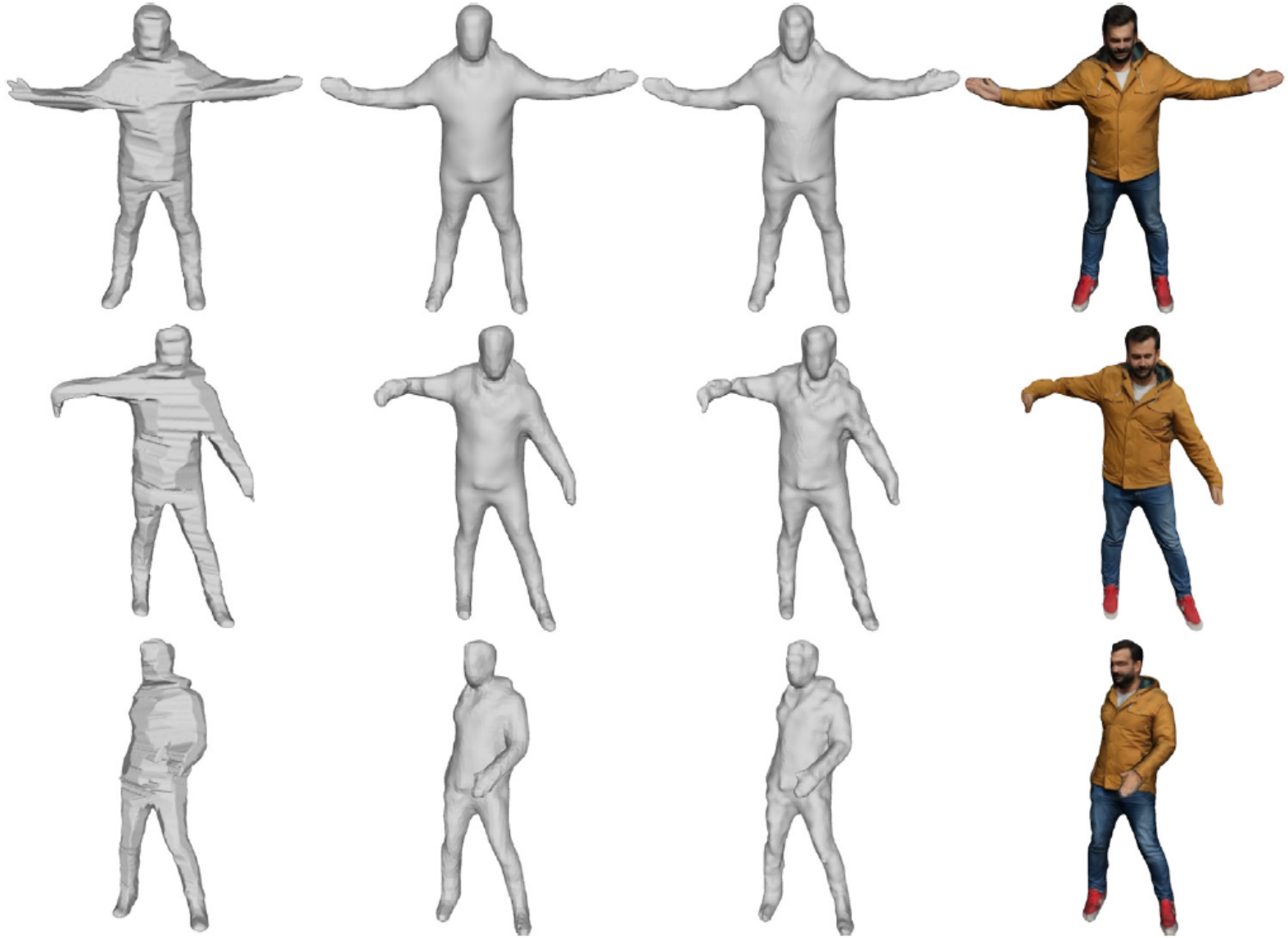


Data Fusion

- Combine the surface M_v, defined by MS-SfS and the surface M_f, defined by the foreground dense point cloud
  - Use M_f to guide a controlled deformation of M_v to have the features and the completeness
- Cast a ray from every vertex of M_f following its normal, searching for a intersection with M_v
  - If found, move the vertex to the intersection point
  - The displaced vertices define the handled points $\mathcal{H}$ and the deformation region $\mathcal{R}$
  - Displacement function:

$$d(\mathbf{v}_j^r) = \mathbf{v}_j^r + \mathbf{n}_j^r \cdot \mathbf{d}_i^h \frac{l - l_j}{l}, \qquad \mathbf{v}_j^r \in \mathcal{R}$$

Normal of $\mathbf{v}_j^r$ | Displacement vector of the closest $\mathbf{v}_j^h$ | Steps to $\mathbf{v}_j^h$



Original SfS model

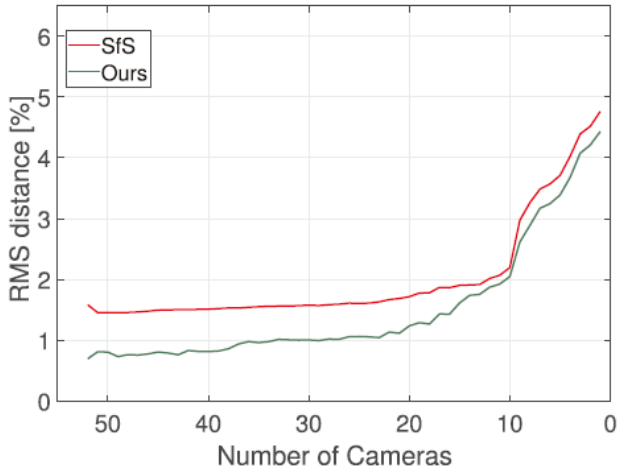
MS-SfS result

Data fusion result

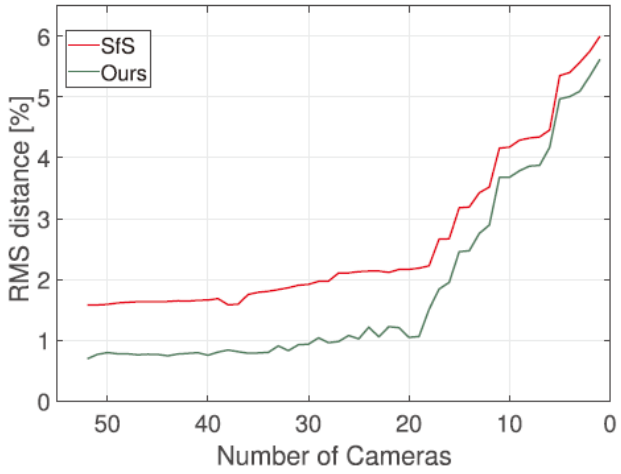
Final textured model

Reconstruction Evaluation

- Metrics: Hausdorff distance (overall distance of two sets)
- Baseline: Microsoft DancingDuo sequence, composed of a set of 53 RGB images
- Experiment: reduce camera number to 18
    1. prioritizing scene coverage
    2. prioritizing camera overlap



(a) Experiment 1



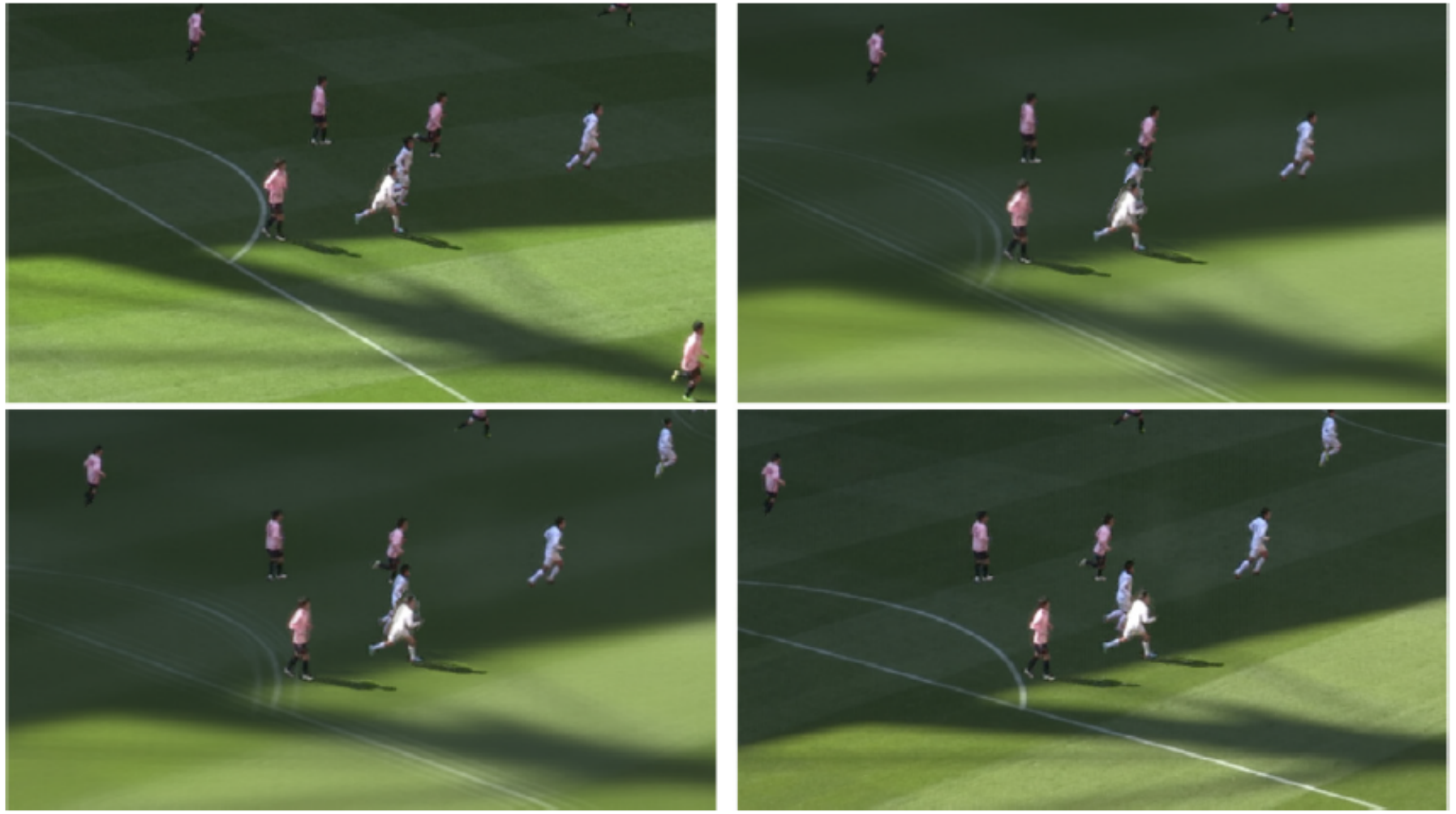
(b) Experiment 2

Reconstruction Evaluation

(a) 53 RGB cameras (b) Experiment 1: 18 cameras (c) Experiment 2: 18 cameras

VR/AR Application

View Synthesis



Conclusion

- Present a novel pipeline for affordable content creation for FVV, VR and AR
- Present an end-to-end system to process and visualize FVV sequences
  - Produces high quality 3D content with a small number of cameras
  - Effectively combine with fusion of different geometry data makes the result more reliable
- Evaluate the scene reconstruction using a quantitative analysis of several challenging examples

Thanks for listening

Any questions?