Robust Frame Registration for Multiple Camera Setups in Dynamic Scenes

Xu Zhao, **Zhong Zhou**, Ye Duan, Wei Wu
State Key Laboratory of VR
Beihang University

Outline

- Background
- Related Work
- Overview
- Frame Registration Algorithm
- Experimental Results
- Conclusion

Background

- 3D vision is very active recent years in AI
  - Image-based modeling
    - Visual Hull & Binocular stereo
      - Can real-time reconstruct 3D models
    - Multi-View Stereo (MVS)
      - Can achieve nearly the same accurate reconstruction result compared with the 3D scanner
  - Camera calibration
    - The first step for image-based modeling and 3D vision
    - Depend the reconstruction accuracy

Background

- Frame registration in multi-camera system
  - Monocular structure from motion / SLAM
    - Accumulative errors across multiple cameras will be introduced
    - It is hard to unify the depth scales of all cameras when lacking reference geometry
  - How many cameras at least are needed for accurate reconstruction in the multi-camera system?
    - MVS can produce accurate reconstruction results but need a moderate of calibrated images

Related Work

- Multi-camera setups
  - CMU virtual reality project
  - ETHZ Blue-C, INRIA GrImage, Tsinghua Dome etc.

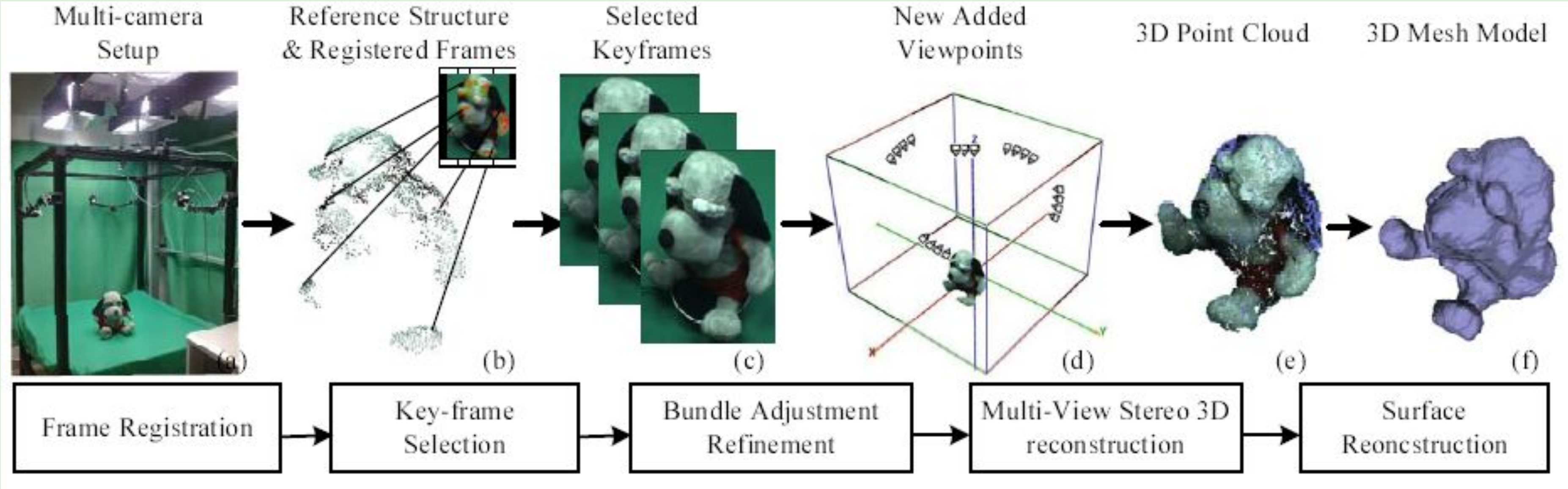
A COMPARISON OF DIFFERENT CAMERA SETUPS

	Main purpose	Advantage	Disadvantage
Single camera setup (hand-held)	Urban or interior scene modeling	Cheapest	It needs user interaction and can only reconstruct static scene
Stereo camera setup (2-3 cameras)	Stereoscopic 3D production	Simple but practical in industry	It can only reconstruct rough depth image
Multi-camera setup (5-8 cameras)	Motion capture	More common in real scenes and systems	It can only reconstruct rough 3D models
Multi-camera setup ($>$ 20 cameras)	Light field modeling	High quality 3D reconstruction	Expensive and for research purpose only

Related Work

- Frame registration algorithms
  - Select several key frames separately from multiple cameras' videos and then perform a global bundle adjustment for all images
    - matched 3D features are too sparse and noisy to estimate camera parameters
  - Perform a traditional structure from motion (SFM) or mono-SLAM algorithm for each camera, and then register all selected frames together
    - accumulative errors / different depth scales

Overview



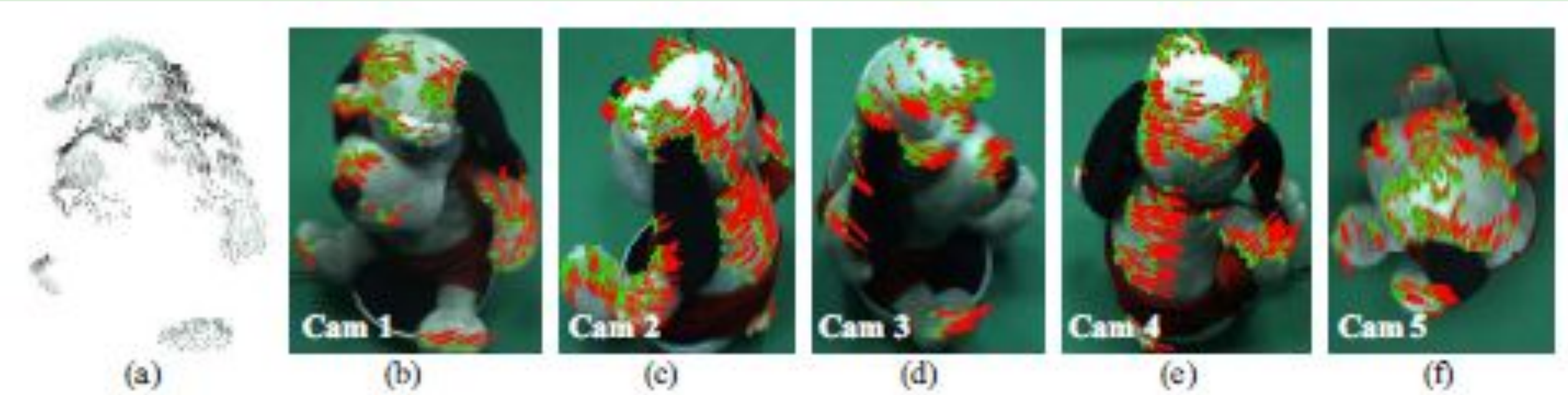
- 1. construct a global reference structure
- 2. use a RANSAC-based LM method to estimate frame poses
- 3. select key frames and combine other state-of-the-art techniques to reconstruct high quality 3D models

Overview

- Contributions:
  - A robust frame registration method for multiple camera setups
  - A high-quality 3D reconstruction system which can select a moderate number of calibrated key-frames
  - A more economic approach to recording motion with a limited number of cameras

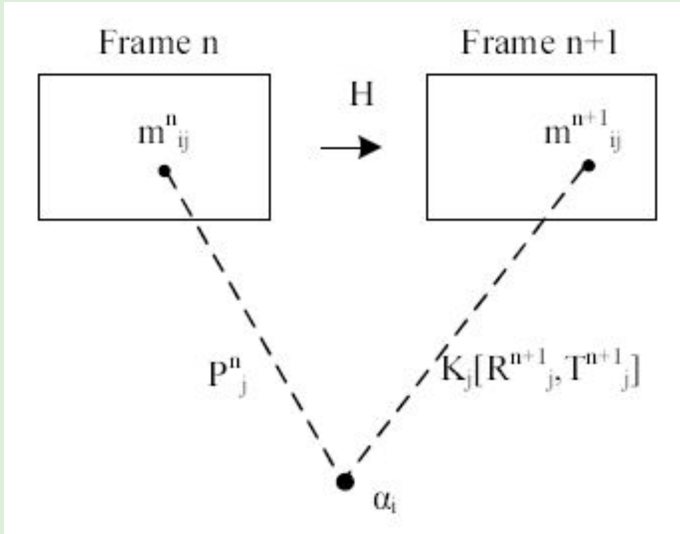
Frame Registration Algorithm

- Reference structure
  - Define it as a set of 3D patches
    - (position, visiblility, feature salience measure)
  - Use patch-based MVS method to construct
  - Project back to images and use Shi-Tomasi score to measure the salience of candidate features



Frame Registration Algorithm

- Frame Registration
  - Register during video tracking
    - Optical flow method
  - With the help of 3D patches and their track pairs, we derive a new frame pose optimizing equation
  - Use RANSAC-based LM method to solve it robustly

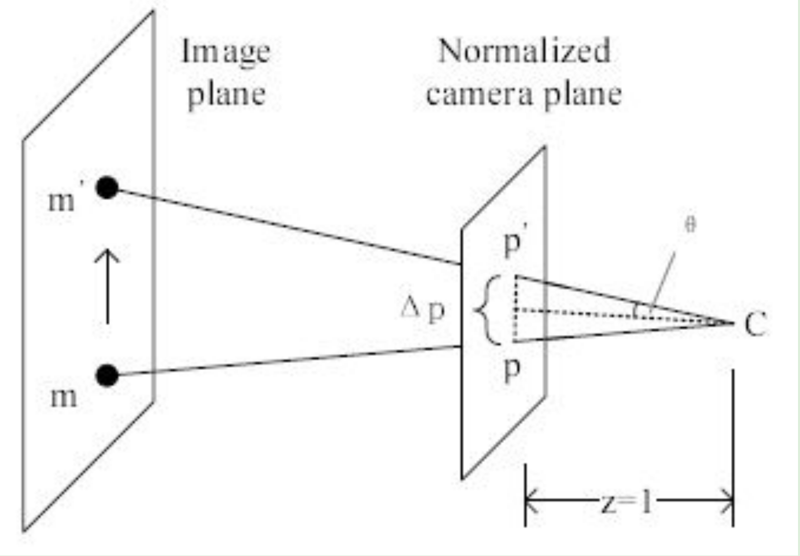


$$[R_j^n, T_j^n]^* = \arg \min_{[R_j^n, T_j^n]^*} \sum_{\alpha \in \Phi} (f(\alpha_i, K_j[R_j^n, T_j^n]), m_{ij}^n)^2$$

$$s.t. \det(R) = 1, R^T = R^{-1}$$

Frame Registration Algorithm

- Key-frame selection
  - A heuristic but effective scheme
    - Combine the narrow and wide baseline
  - Turn angular offset is used to measure the motion.
  - For each video sequence, we selected four to six key frames



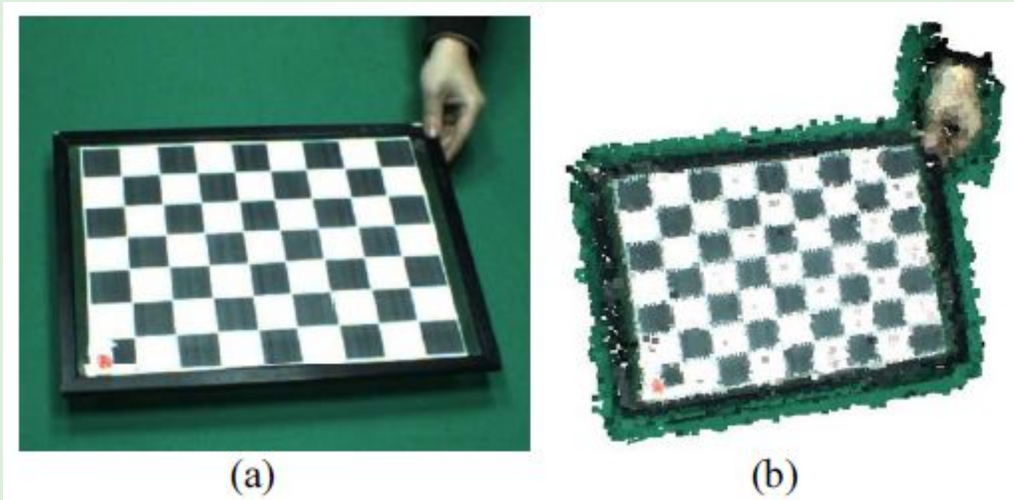
$$\theta \approx \arctan\left(\frac{\Delta p}{2z}\right) = \arctan\left(\frac{|p' - p|}{2}\right)$$

Frame Registration Algorithm

```
PROCEDURE FrameRegistrationForMultipleCameraSetups()
   recontructReferenceStructure()
   FOR each camera j
      captureFrame(j)
      FOR each frame i
            IF i=0 THEN
               determineTrackingFeatrues(i)
            ENDIF
               pyramidOpticalFlowTracking(i)
               estimateFramePoseRANSAC(i)
               //* used in the reconstruction system
            IF isKeyframe(i) = TRUE THEN
               saveFrameAndPose(i)
            ENDIF
      ENDFOR
   ENDFOR
   //* used in the reconstruction system
   bundleAdjustmentRefinement()
END
```

Experimental Results

- Frame registration with ground truth

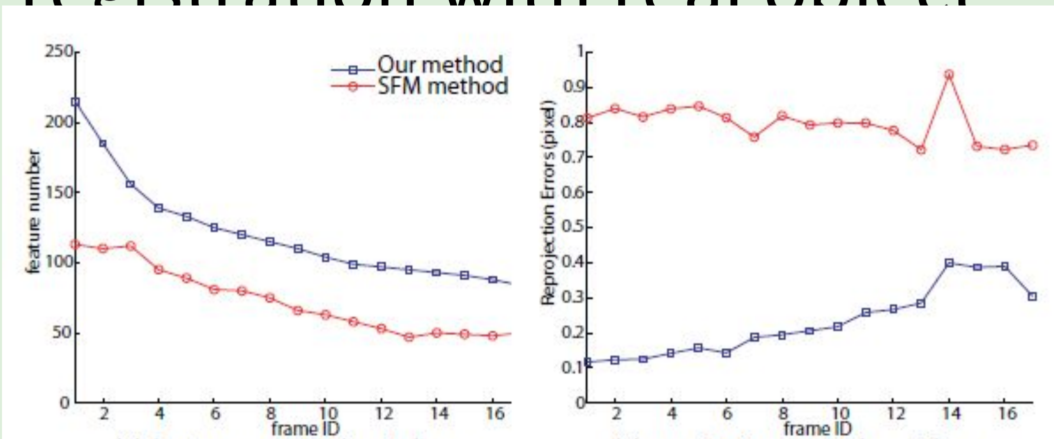


A COMPARISON WITH THE GROUND TRUTH

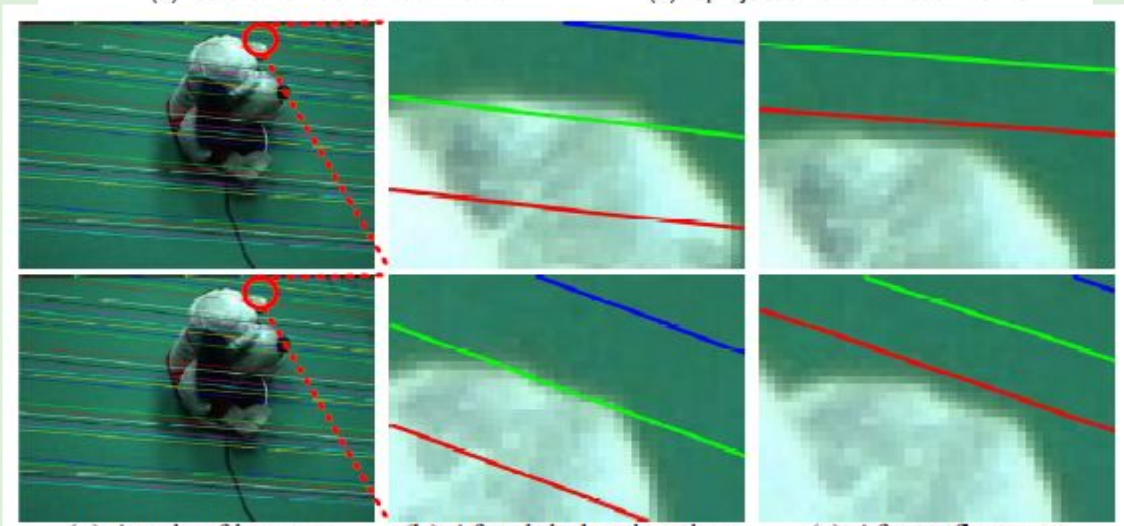
	Position diff.(mm)		Principal axis diff.(°)	
	Mean	Std. Dev.	Mean	Std. Dev.
Our method	1.3880	0.7698	0.0699	0.0380
SFM method	12.951	9.0739	0.6995	0.3764

Experimental Results

- Frame registration with real object



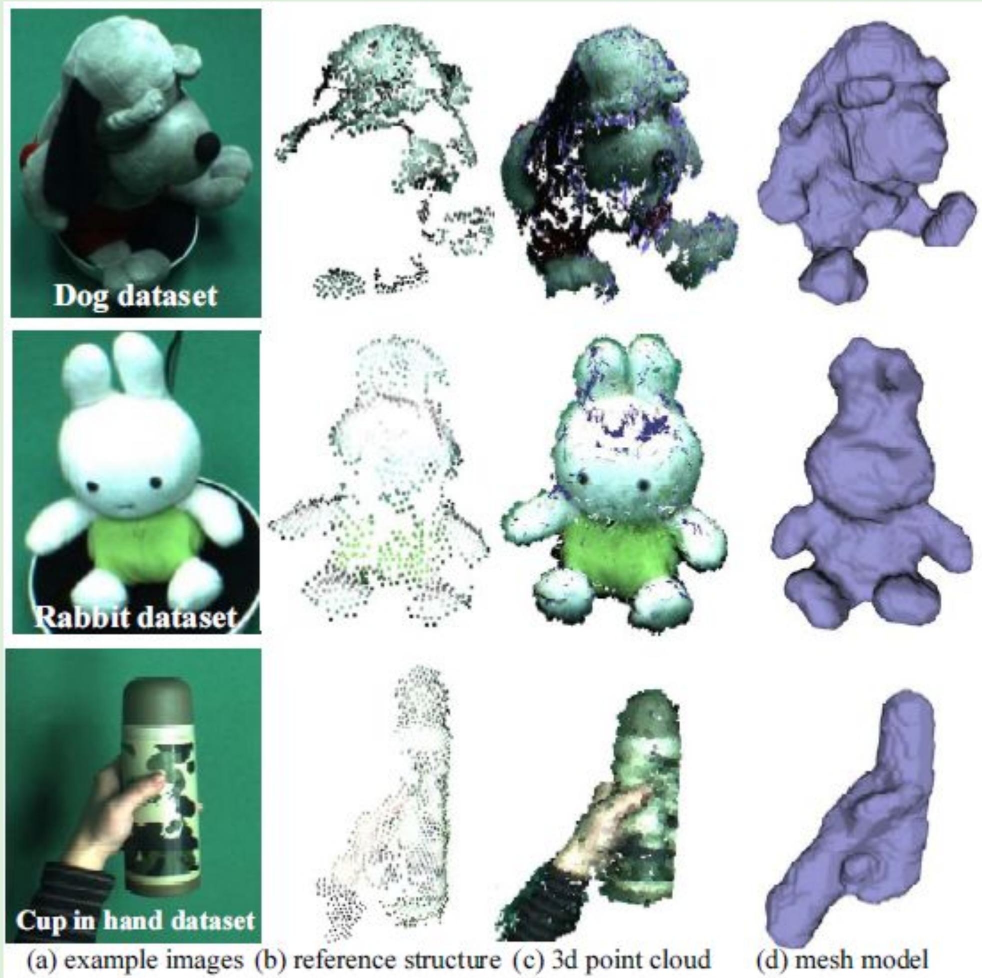
(a) feature number of each frame (b) reprojection error of each frame



(a) A pair of images (b) After inital estimation (c) After refinement

Experimental Results

- 3D reconstruction



Conclusion

- A robust frame registration algorithm
  - Frames are registered into one unified coordinate system
  - Avoiding accumulative errors compared with the separate SFM registration method
- A high quality 3D reconstruction system
  - Automatically select key-frames in dynamic scenes
  - Accurately reconstruction with a limited number of static cameras

Thanks for you listening!

Xu Zhao, **Zhong Zhou**, Ye Duan, Wei Wu

State Key Laboratory of VR

Beihang University