Structure from motion for omnidirectional multi-camera system and its applications

Vision and Media Computing Lab.
Graduate School of Information Science
Nara Institute of Science and Technology (NAIST)

Tomokazu Sato

Overview of this talk

- SFM and PnP for multi-camera system
  - SFM (Structure from motion)
  - PnP (Perspective n point) problem
- GPS integration to SFM
- Applications of omnidirectional SFM

Applications of omnidirectional video using 3-D structure

(1) Omni-directional tele-presence system without invisible area

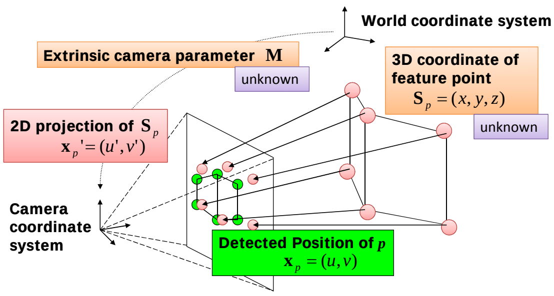
(2) Novel-view synthesis from omni-directional video using a deformable 3-D mesh model

(3) Feature landmark based marker-less augmented reality

Visual-SLAM (online SFM) and its application for video mosaic

* T. Sato, A. Iketani, S. Ikeda, M. Kanbara, N. Nakajima, and N. Yokoya: 'Video mosaicing for curved documents by structure from motion', ACM SIGGRAPH2006, Sketches, Aug. 2006.

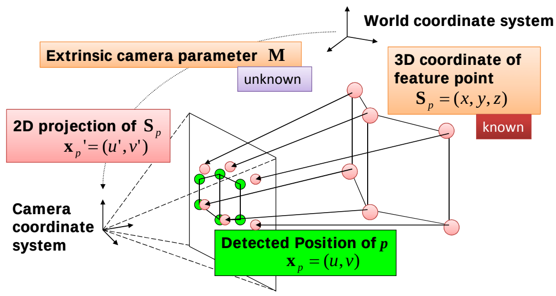
SFM and Reprojection error



Sum of re-projection errors: $E = \sum_p |x_p' - x_p|^2 \rightarrow \min$

Multiple images are necessary to solve this problem.

PnP problem and Reprojection error



Sum of re-projection errors: $E = \sum_p |x_p' - x_p|^2 \rightarrow \min$

Single images is sufficient to solve this problem.

Extrinsic camera parameter estimation (PnP and SFM)

- **PnP problem (Perspective n Point problem)**
  is for pairs of known 3-D positions and their 2-D observations.
  - Algebraic solution with minimum features (P3P, P4P, etc)
  - Linear solution for arbitral number of features (PnP)
  - Non-linear minimization of reprojection errors
    (For final refinement, it needs good initial guess)
- **SfM (Structure from Motion)**
  is for unknown 3-D positions and their 2-D observations.
  - Factorization-based method (For video, batch processing)
  - Visual-SLAM (For video, real-time)
  - Ego-motion estimation (For sufficiently long baseline image pairs)
  - Non-linear minimization of reprojection errors
    (For final refinement, it needs good initial guess)

Extrinsic camera parameter estimation (PnP and SFM)

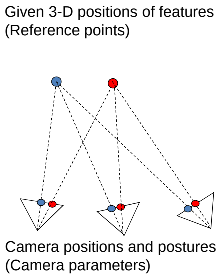
- **PnP problem (Perspective n Point problem)**
  is for pairs of known 3-D positions and its 2-D observation.
  - Algebraic solution with minimum features (P3P, P4P, etc)
  - Linear solution for arbitral number of features (PnP)
  - Non-linear minimization for reprojection errors

Solution for multi-camera system

- **SfM (Structure from Motion)**
  is for unknown 3-D positions and its 2-D observation.
  - Factorization-based method (For video, batch processing)
  - Visual-SLAM (For video, real-time)
  - Ego-motion estimation (For sufficiently long baseline image pairs)
  - Non-linear minimization of reprojection errors

Extension with GPS

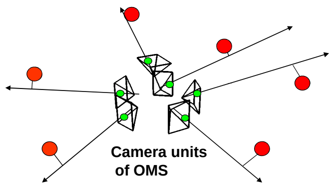
Basic strategy for visual-SLAM



Feature tracking
↓
Camera parameter estimation solving PnP
↓
3-D position update for image features
↓ (Iterate for each frame)
Bundle adjustment

9

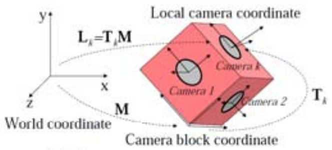
Linear solution for PnP problem in OMS (1/2)



We minimize distances between projecting lines and 3-D position of features instead of reprojection errors.

10

Linear solution for PnP problem in OMS (2/2)



$$\mathbf{A}\mathbf{m} = \mathbf{s},$$

$$\mathbf{A} = \begin{pmatrix} s1(k_1)\mathbf{S}_1 & s2(k_1)\mathbf{S}_1 & s3(k_1)\mathbf{S}_1 \\ \vdots & \vdots & \vdots \\ s1(k_j)\mathbf{S}_j & s2(k_j)\mathbf{S}_j & s3(k_j)\mathbf{S}_j \\ s5(k_1)\mathbf{S}_1 & s6(k_1)\mathbf{S}_1 & s7(k_1)\mathbf{S}_1 \\ \vdots & \vdots & \vdots \\ s5(k_j)\mathbf{S}_j & s6(k_j)\mathbf{S}_j & s7(k_j)\mathbf{S}_j \end{pmatrix}, \quad \mathbf{s} = \begin{pmatrix} -s4(k_1) \\ \vdots \\ -s4(k_j) \\ -s8(k_1) \\ \vdots \\ -s8(k_j) \end{pmatrix}.$$

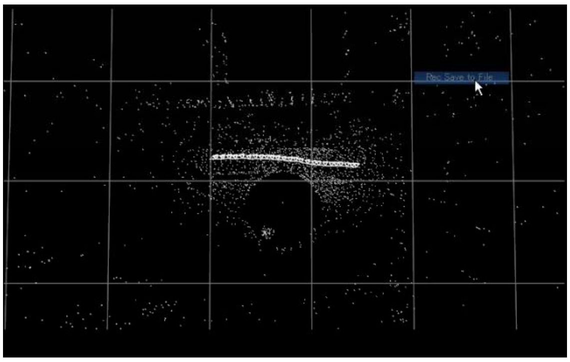
$$\mathbf{m} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{s}.$$

They can be computed from observations

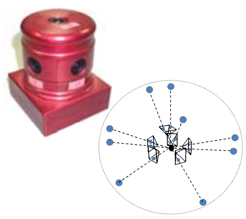
* Tomokazu Sato, S. Ikeda, and N. Yokoya 'Extrinsic camera parameter recovery from multiple image sequences captured by an omni-directional multi-camera system', *Proc. ECCV*, May. 2004.

Tracking of image features with reference points

Estimated camera path



In case **3D reference points** are **unavailable**



3D positions of features are initialized as they are on the sphere.

- Feature tracking
- Camera parameter estimation solving PnP
- 3-D position update for image features
- Local bundle adjustment

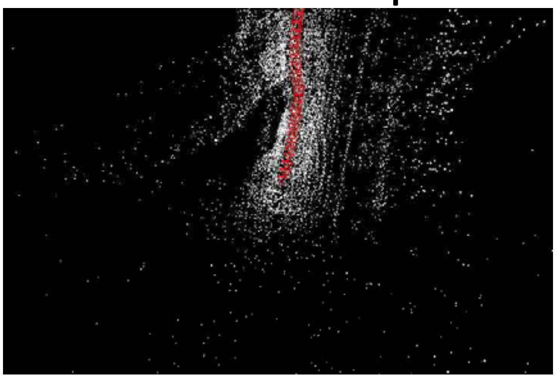
Iterate for each frame

- Global bundle adjustment

14

Tracking of image features without reference points

Estimated camera path without 3D reference points



Overview of this talk

- SFM and PnP for multi-camera system
  - SFM (Structure from motion)
  - PnP (Perspective n point) problem
- GPS integration to SFM
- Applications of omnidirectional SFM

Extension of SFM

Camera parameter estimation from video images and accuracy considered GPS

Hideyuki Kume,
Tomokazu Sato, Naokazu Yokoya

* 'Extrinsic camera parameter estimation using video images and GPS considering GPS positioning accuracy', *Proc. ICPR2010*, Aug. 2010 (accepted, to apper)

Extension of objective function

Penalty term for GPS positioning is added to objective function.

$$E = \sum_{i \in F} \Phi_i + \omega \sum_{i \in F_g} \Psi_i$$

Reprojection error

$$\Phi_i = \frac{1}{|P_i|} \sum_{j \in P_i} \mu_j (q_{ij} - \hat{q}_{ij})^2$$

Penalty term for GPS positioning

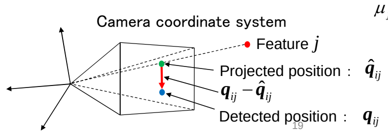
ω : weight

F : Group of input frames

F_g : Group of frames in which GPS data is acquired

P_i : Group of feature points tracked in the i-th frame

μ_j : Confidence of feature j



How should we design the penalty term for GPS data?

Positions are measured in fixed position using RTK-GPS (TOPCON GR-3) for 5 hours

Red axis : East, Green axis : South, Blue axis : Vertical up

Acquired positions in RTK-fix mode
Length of axe : 1 [m]

Acquired positions in RTK-float mode
Length of axe : 10 [m]

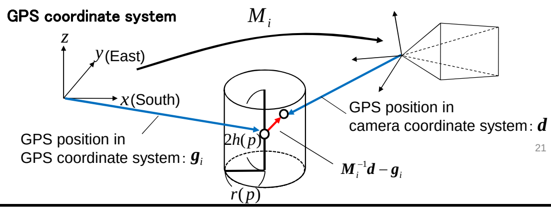
Assumption of normal distribution for GPS error in not effective for random walking errors for short time video streams.

Penalty term for GPS positioning

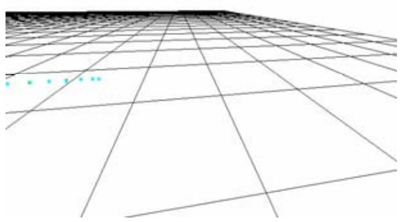
$$\Psi_i = \left(\frac{1}{r(p)}\sqrt{x_i^2 + y_i^2}\right)^{2n} + \left(\frac{1}{h(p)} z_i\right)^{2n}, \quad \begin{pmatrix} x_i \\ y_i \\ z_i \\ 1 \end{pmatrix} = M_i^{-1} d - g_i .$$

n : Large given number
p : Confidence of GPS

$$\Psi_i = \begin{cases} \text{Large number, if GPS is outside the cylinder} \\ \text{Almost 0, otherwise} \end{cases}$$

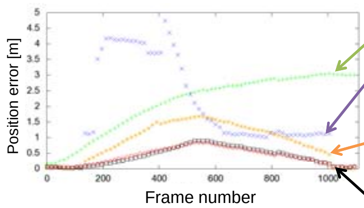


Estimated camera paths



: Proposed method
: Compared method [Anai et al., 09]
Pyramid : Camera posture
: Acquired GPS positions
: Ground truth
: RTK-fix
: RTK-float

Quantitative evaluation

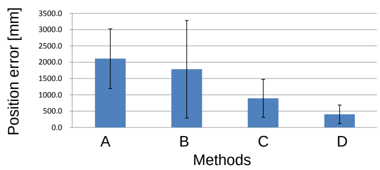


A: Vision only

B: Vision + GPS [Yokochi 06] w/o confidence info. with normal distribution

C: Vision + GPS [Anai 09] with confidence info. and normal distribution

D: Proposed method



Overview of this talk

- SFM and PnP for multi-camera system
  - SFM (Structure from motion)
  - PnP (Perspective n point) problem
- GPS integration to SFM
- Applications of omnidirectional SFM

Application 1/3

Omni-directional telepresence system without invisible area by image completion

Koutaro Machikita, Norihiko Kawai, Tomokazu Sato, Naokazu Yokoya

* 'Generation of an omnidirectional video without invisible areas using image inpainting', Proc. ACCV2009, Sep. 2009

Invisible area in OMS

Omnidirectional telepresence system using video captured by OMS.

Complete video capturing for all the direction by OMS is essentially difficult.

Invisible area in the video decreases a reality

Image from Google's street view

Image completion technique that relying only for geometric info. often generates unnatural images.

26

Exemplar based image completion (image inpainting) methods

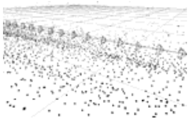
[Wexler et al., 2004]

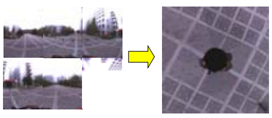
[Kawai et al., 2008]

Processes for image completion

1. Camera parameter estimation for omni-video

2. Estimation of ground surface and image projection to it





3. Determination of data region using geometric infomation

4. Image completion by using pattern similarity

Target　Reference frame　Target　Reference frame

29

Image completion based on pattern similarity

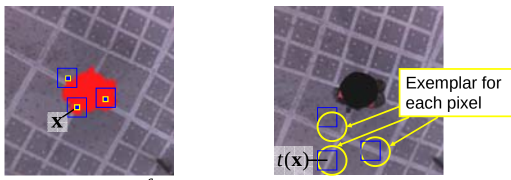
$$E = \sum_{\mathbf{x}_i \in \Omega'} w_{\mathbf{x}_i} \left[\sum_{\mathbf{p} \in W} \{ I_f(\mathbf{x}_i + \mathbf{p}) - I_k(t(\mathbf{x}_i) + \mathbf{p}) \}^2 \right]$$

$I(\mathbf{x})$: Pixel value for $\mathbf{x}$

$t(\mathbf{x})$: Corresponded pixel position with $\mathbf{x}$

$w_{\mathbf{x}}$: Weight for $\mathbf{x}$

Energy is defined as weighted sum of SSD.



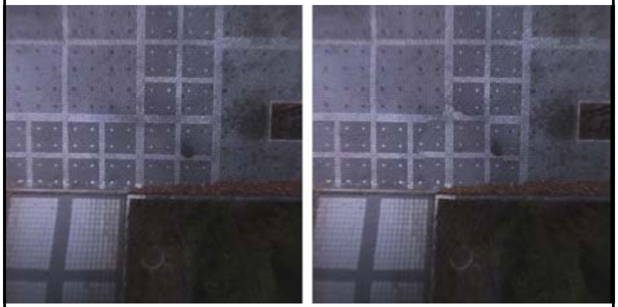
Target frame f　Reference frame k

Following two processes are repeated until energy convergence.

Searching for similar pattern ⟲ Update for pixel value

30

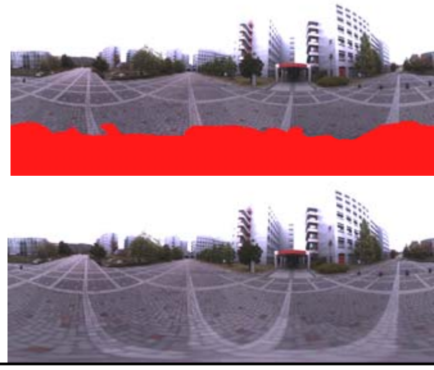
Completed result for projected images



Our method | The method that uses only geometric information

31

Before and after completion in panoramic image



32

Result in telepresence system

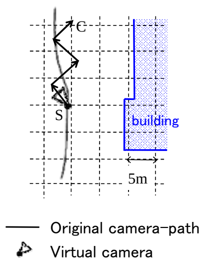
Application 2/3

Novel-view synthesis from omnidirectional video using a deformable 3-D mesh model

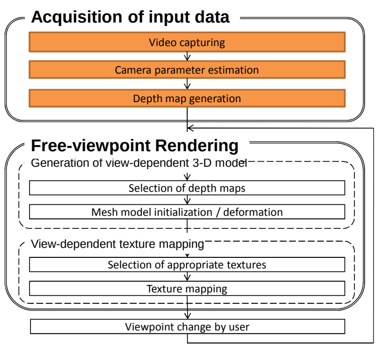
Hiroyuki Koshizawa, Takuya Ibuki, Tomokazu Sato, Naokazu Yokoya

* ‘Omnidirectional free-viewpoint rendering using a deformable 3-D mesh model’, Int. J. of Virtual Reality, Vol. 9, No. 1, pp. 37-44, March 2010.

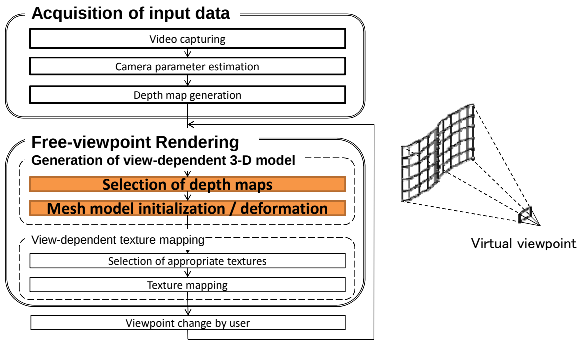
Generated video image for zigzag camera path



Flow diagram of proposed method

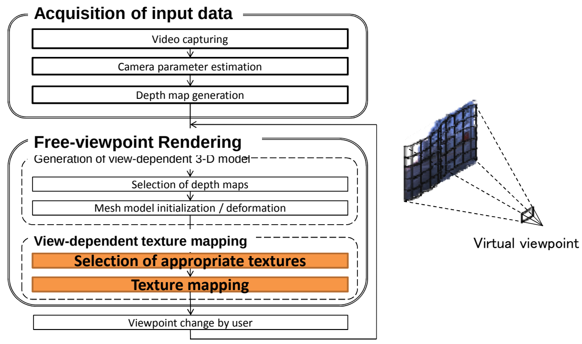


Flow diagram of proposed method



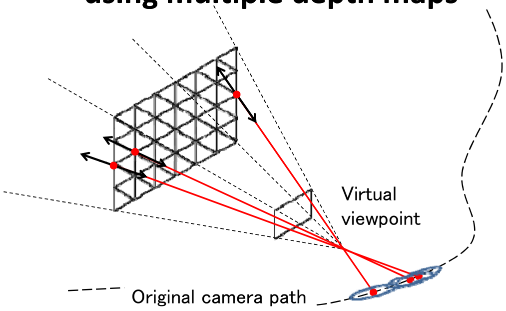
37

Flow diagram of proposed method



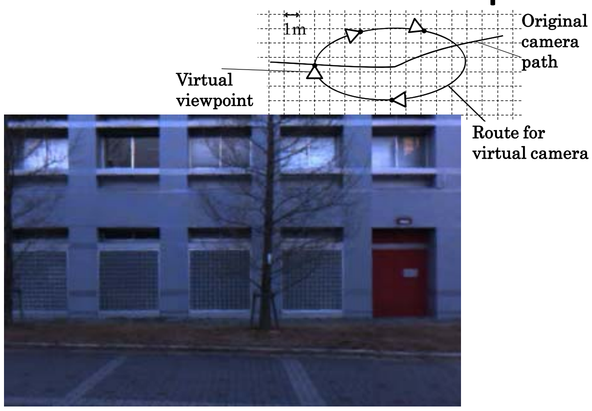
38

Mesh deformation using multiple depth maps



Each vertex on mesh is moved so as to minimize energy function that expresses consistency of the depth maps.

Generated video for circular path



Application 3/3

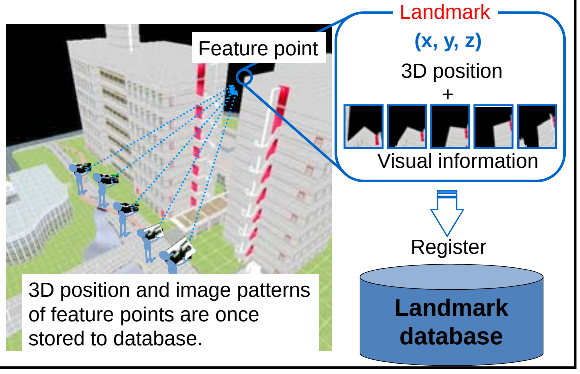
Feature landmark based augmented reality

Takafumi Taketomi,
Tomokazu Sato, Naokazu Yokoya

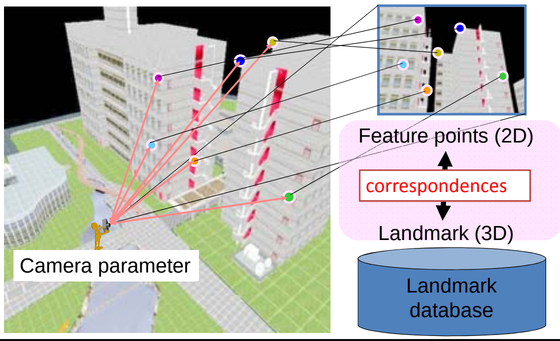
* 'Real-time camera position and posture estimation using a feature landmark database with priorities', Proc. ICPR2008, Dec. 2008.

Example of AR: Virtual sightseeing

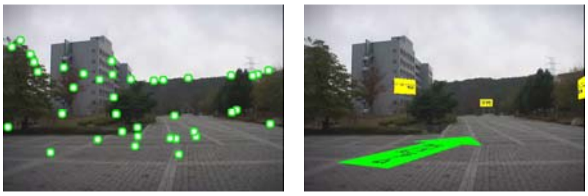
Construction of landmark database



Camera parameter estimation using landmark database



Human navigation using Augmented Reality

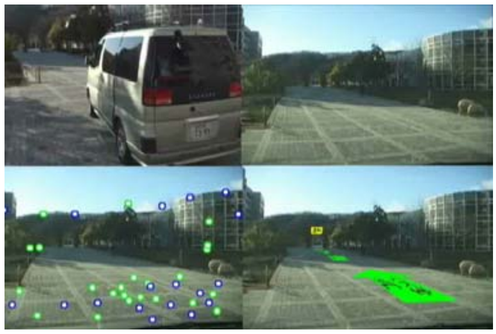


Detected Landmarks

Navigation information shown to the user

Intuitive navigation can be achieved in mobile phones.

Application for car AR navigation



Computational cost:59ms for 1 frame (17 fps)
(CPU: Core 2 Extreme 2.93GHz, Memory: 2GB)

Application for Pre-Visualization

Summary

- Introduction of omnidirectional SFM
  - PnP Solution for OMS
  - Visual SLAM for OMS
- GPS integration to SFM
  - Cylinder whose size is changed depending on measurement confidence is used for penalty term.
- Applications of omnidirectional SFM
  - Video generation without invisible area
  - Novel view synthesis
  - Landmark based Augmented Reality

E-mail: tomoka-s@is.naist.jp
URL: http://yokoya.naist.jp/~tomoka-s/index-e.html