P1A-08 Detecting Changes in 3D Structure of a Scene from Multi-view Images Captured by a Vehicle-mounted Camera

Ken Sakurada, Takayuki Okatani, Koichiro Deguchi (Tohoku Univ., Japan)

Goal: Estimate large-scale structural changes of a city from their two image sequences captured at different times
Background: Difficulty with dense depth estimation (i.e., multi-view stereo) from images captured by a ground vehicle
Idea: Estimate only the probability of a depth change at each pixel without explicitly estimating the depths

Motivation

- Visualize the damages and the recovery/reconstruction processes of the tsunami affected-areas
  - Since mid-April 2011
    - 2 or 3 months apart
  - 25 million images
    - Every 2m
    - 20 TB (as of Dec. 2012)


Recorded area (about 500km)

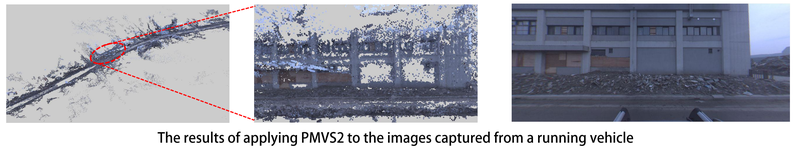
Examples of temporal changes


Panoramic images of tsunami-damaged area

Measurement vehicle

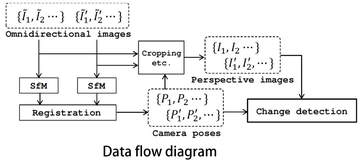
Dense 3D reconstruction from ground vehicle images

- A lot of missing parts tend to be missing
  - The differentiation of two reconstructions does not give good results

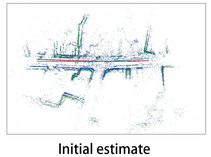

The results of applying PMVS2 to the images captured from a running vehicle

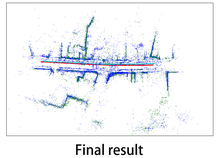
Estimation of relative camera poses

- Perform SfM independently for each of the two sequences
- Roughly align the two reconstructions based on GPS data
  - Reestablish the correspondences of feature points by incorporating a distance constraint
- Perform bundle adjustment over the two sequences



Data flow diagram

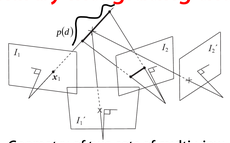

Initial estimate


Final result

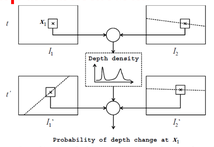
Detection of temporal changes of a scene

- Goal: Estimate the large-scale structural changes of a city.

- Estimate the scene depths not deterministically but probabilistically
- Estimate the probability of the structural changes independently at each pixel by integrating the estimated depth densities



Geometry of two sets of multi-view perspective images taken at different times



Outline of the proposed method

Estimating the probability of scene changes

- Inputs
  - The similarity of the local image patches among the multi-view images $\{s'_1, \ldots, s'_n\}$
  - (Camera poses that are already estimated)
- The probability that the scene changes its structure at the pixel x_1 of I_1.

$$p(c=1 \mid s'_1, \ldots, s'_n) = \frac{p(s'_1, \ldots, s'_n \mid c=1)\,p(c=1)}{p(s'_1, \ldots, s'_n)}$$

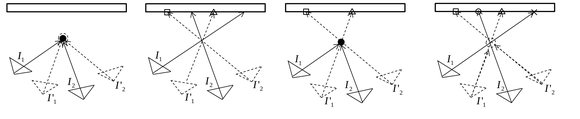
At the pixel x_1 of I_1,
the scene structure changed at time t' $c=1$, otherwise $c=0$
the scene depth at time t' is d $\delta'_d=1$, otherwise $\delta'_d=0$
the scene depth at time t is d $\delta_d=1$, otherwise $\delta_d=0$
the difference of the local patch for the depth d s'_d

Values of δ_d' for different pairs of c and δ_d

c \ δ_d	0	1
0	② 0	① 1
1	④ 0 or 1	③ 0

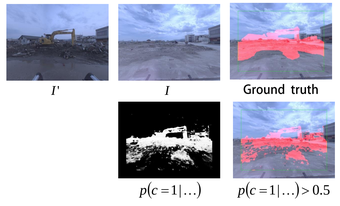
① The similarity between the patches of the same scene point $p(s'_d \mid \delta'_d=1) \propto \exp(-s'_d/\sigma')$
②-③ " the patches of the two different scene points $p(s'_d \mid \delta'_d=0) = \text{const}$
④ Both cases of above-mentioned $p(s'_d \mid \delta_d=0, c=1) \approx p(s_d \mid \delta'_d=0, \delta_d=0, c=1) = p(s'_d \mid \delta'_d=0)$
$(\because p(s_d \mid \delta'_d=1, \delta_d=0, c=1) \approx 0)$

① $p(s'_d \mid \delta_d=1, c=0) = p(s'_d \mid \delta'_d=1)$
② $p(s'_d \mid \delta_d=0, c=0) = p(s'_d \mid \delta'_d=0)$
③ $p(s'_d \mid \delta_d=1, c=1) = p(s'_d \mid \delta'_d=0)$
④ $p(s'_d \mid \delta_d=0, c=1) = p(s'_d \mid \delta'_d=0 \text{ or } 1)$



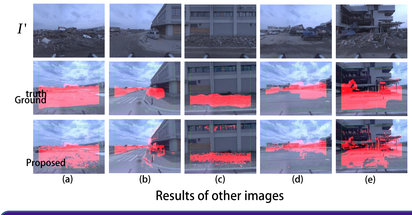
Experimental results

- The proposed method outperforms the MVS-based methods

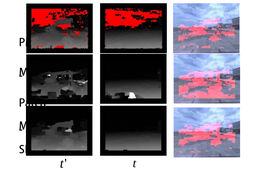


Proposed method
- Disparity space 128 blocks ($n=128$)
- No prior on the probability of scene changes $p(c=1)=0.5$

MVS-based methods
- The structures of a scene is reconstructed based on MVS
- Then, they are differentiated to detect scene changes



Results of other images



F_1 score of the detected changes shown in (a)-(e)

	(a)	(b)	(c)	(d)	(e)	⋯	Average
Proposed	0.88	0.67	0.77	0.85	0.82	⋯	0.83
PMVS2	0.49	0.30	0.65	0.66	0.56	⋯	0.56
Patch-MVS	0.66	0.28	0.69	0.60	0.70	⋯	0.62
SIFT-MVS	0.68	0.41	0.73	0.71	0.60	⋯	0.65

Conclusion

- The proposed method estimates the probability of the structural changes independently at each pixel by integrating the densities of estimated depths.
- Experimental results show that the proposed method outperforms the MVS-based methods.

Change detection dataset

- The dataset used in this study are available from our web site
  http://www.vision.is.tohoku.ac.jp/us/download/
  - Images of two different city streets
  - Data of each street consists of two image sequences captured at different times, the estimated camera poses, and several hand-labeled ground-truths