KinectFusion: Real-Time Dense Surface Mapping and Tracking

Gabriele Bleser

Thanks to Richard Newcombe for providing the ISMAR slides

Overview

- General: scientific papers (structure, category)
- KinectFusion: deep understanding of core parts
  - Paper: very detailed information
  - Slides: general concepts/illustrations
  - This lecture: discussion based on both, focus on key aspects
- Some interesting links
- Announcements

General info for me

- How familiar are you with reading scientific papers?
- How many papers have you read before?
- Did you write a paper yourself?

- Did you read KinectFusion? Did you read the slides?
- How much time did you take to work through the paper?
- What was easier to understand? Paper or slides?
- Did you understand the approach: not at all, partly, completely?

Typical scientific paper categories

- Technology paper
  - new approach, improved method, new combination of methods; more precise, more efficient, new capabilities…
- System paper
  - Put existing methods together, engineering, make it work, validate variations
- Survey
  - Summarize what is there
- Taxonomy
  - (Hierarchically) classify/group existing approaches
- Application paper
  - Apply existing technology to a specific application, evaluate, end users

Structure of scientific papers

- Typical length: 6-10 pages
- **Abstract:** short summary of whole paper, sell highlights, get reader interested
- **Introduction:** general topic and problem you try to solve
- **Background/related work:** how did others do it (position yourself with respect to state-of-the-art)
- **Method overview:** outline; possibly introduce notation
- **Core of method (structure appropriately)**
- **Experiments/evaluation:** compare to state-of-the-art, compare variations; criteria e.g.: robustness, accuracy, computational efficiency, …
- **Conclusion:** general summary, highlights, key features, limitations (!); open up: indicate future work
- **Bibliography**

KinectFusion

Real-time dense surface mapping and tracking

http://research.microsoft.com/en-us/projects/surfacerecon/

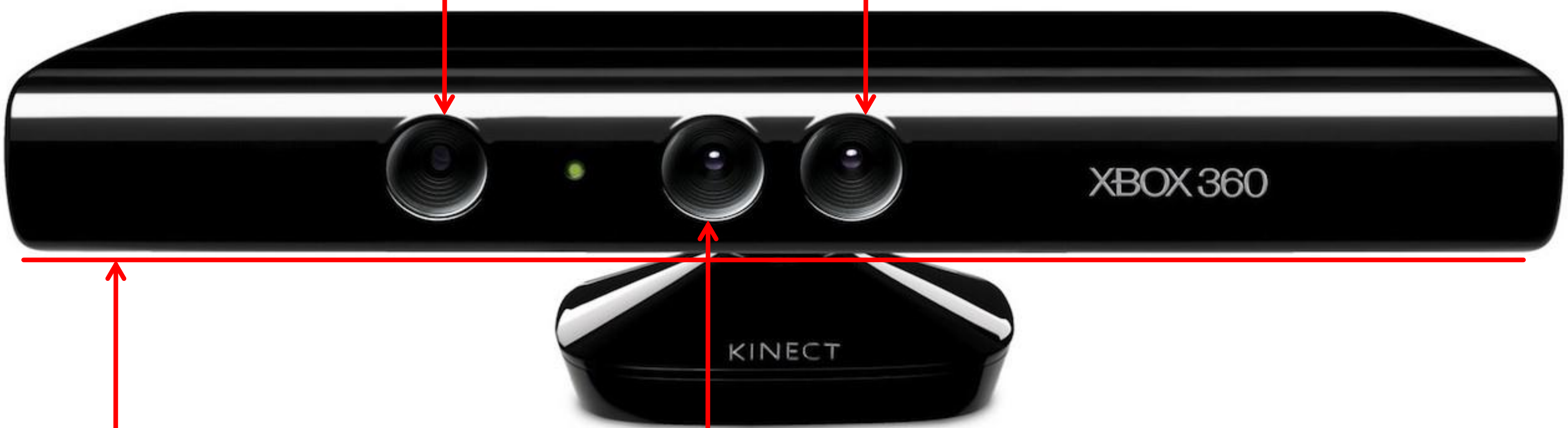
General questions

- Explain the general KinectFusion approach.
- Position the approach with respect to the methods taught in the lecture.
  - Which knowledge from the lecture could you apply and which methods did you recover?
  - What is different?
- What are the core parts of the KinectFusion method? Where are these in the paper (which equations)?

Kinect hardware

Infrared video projector

Infrared camera



Color camera

Microphone array for sound source localization

Kinect: encoded SL grid

KinectFusion paper/slides structure

- Introduction and state-of-the-art (1-2) – slides 1-14
- Method overview (3 intro, 3.1) – slides 15-27
- Surface measurement (3.2) – no slides
- Surface reconstruction (3.3) – slides 28-37, 42-45
- Surface prediction (3.4) – slides 38-41
- Sensor pose estimation (3.5) – slides 46-57
- Experiments (4) – slides 58-63
- Conclusion (5) – slide 65

Comment

- The following slides list the questions from the assignment and the slides of the KinectFusion presentation that we looked at.
- Questions in black are the ones that we discussed.
- Questions in grey are the ones that we did not explicitly discuss.
- The illustrations from the lecture are also included. Note, these are just sketches for the purpose of helping the understanding.
- The paper has a lot of aspects that we did not explicitly discuss. However, the focus of today‘s session was to understand the key parts of the mapping and tracking method. You are not expected to know all the other details of the paper.
- However, if you want help concerning specific questions, write an email to gabriele.bleser@dfki.de (could potentially be included in the question lecture).

Introduction and state-of-the-art (1-2)

Slides 1-14

1. What is meant with: *active and passive computer vision*, *infrastructure-free tracking*, *frame-to-frame tracking*?

2. In how far differs KinectFusion from frame-to-frame tracking?

3. What is the difference between dense reconstruction and sparse reconstruction and what are advantages and disadvantages of the one over the other?

4. Can you summarize shortly, how PTAM works and what are the major components.

5. What is meant with *loop closure*, *large-scale metrically consistent maps*?

Introduction and state-of-the-art (1-2) cont‘d

6. What does the first paragraph of Section 2.4 want to tell us?
7. What is a truncated signed distance function?
8. What is the definition of the L1,2 norm? Why is the L1 norm more robust against outliers during least squares estimation?
9. What means *total variation regularisation*?
10. What are *marging cubes* and *ray casting* used for, what is the major difference?

Mathematical notation

- p-norm: $\|x\|_p := \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}}$
- L1-norm (p=1): sum of absolute values
- L2-norm (p=2): Euclidean norm (length of vector)

Method overview (3 intro, 3.1)

Explain the general method, e.g. using Figure 3 in the paper

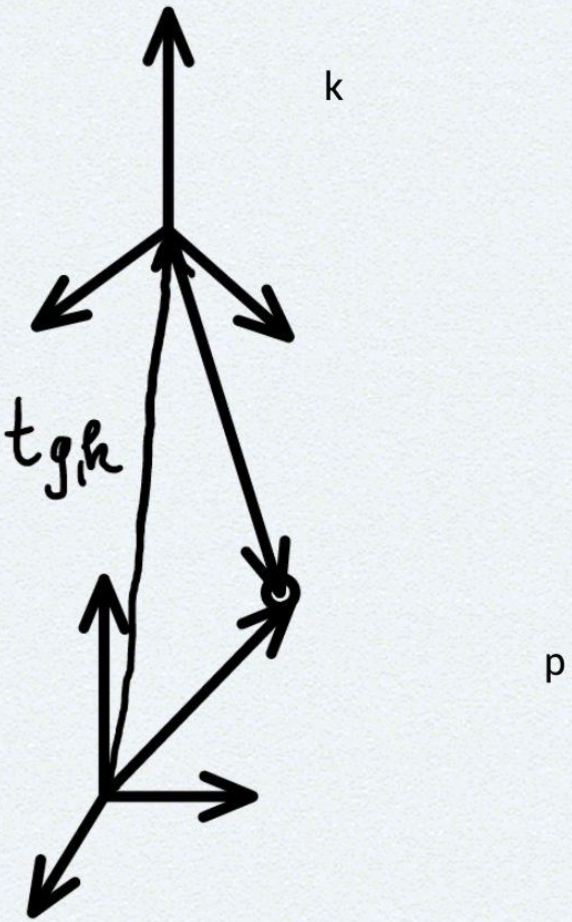
Method overview (3 intro, 3.1)

Slides 15-27

1. Using Equation (1), write the transformation of point p from the Kinect system into the world system in Euclidean form.

2. Mathematical notation: While reading, mark new symbols and their meaning, so that you can look them up easily later.

3. Illustrate the transformation and the vectors $t_{k,g}$ and $t_{g,k}$ with a simple drawing.



$$p_g = R_{g,k}\, p_k + t_{g,k}$$

Surface measurement (3.2) – no slides

The Kinect provides a depth image. How do we obtain a surface measurement (vertices and normals)?

4. Equation $p_k = R_k(u)K^{-1}\dot{u}$: What do the different symbols mean? Can you illustrate this with a simple drawing?

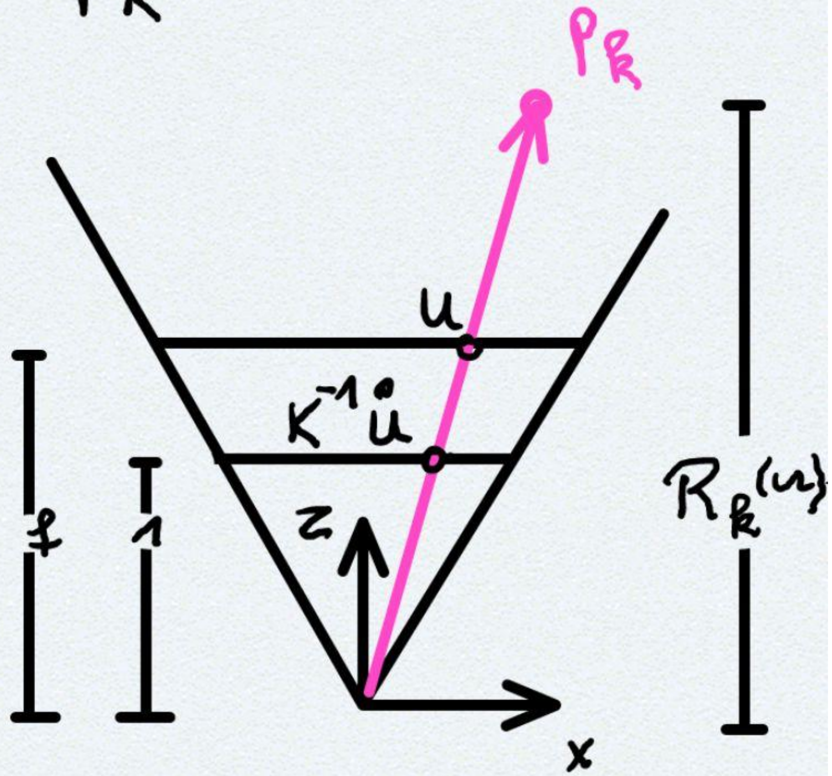
5. Explain the different terms of the bilateral filter in Equation (2). Why is the filtered result *discontinuity-preserved* and what does it mean?

6. Why do you think is the depth map pyramid computed? Do you know this principle from the lecture?

7. At the end of paragraph 3.2, why is the transformation of vertices different from the transformation of normals?

Pixel + depth → vertex

$$p_k = R_k(u)\, K^{-1} \dot{u}$$



Surface reconstruction (3.3)

Slides 28-37, 42-45

8. What is the TSDF? What are advantages of the TSDF over *e.g.* an explicit surface representation?

9. How would you initialize the TSDF at system start-up?

11. What happens in Equations (6) to (9)? Can you illustrate this with vectors and coordinate systems?

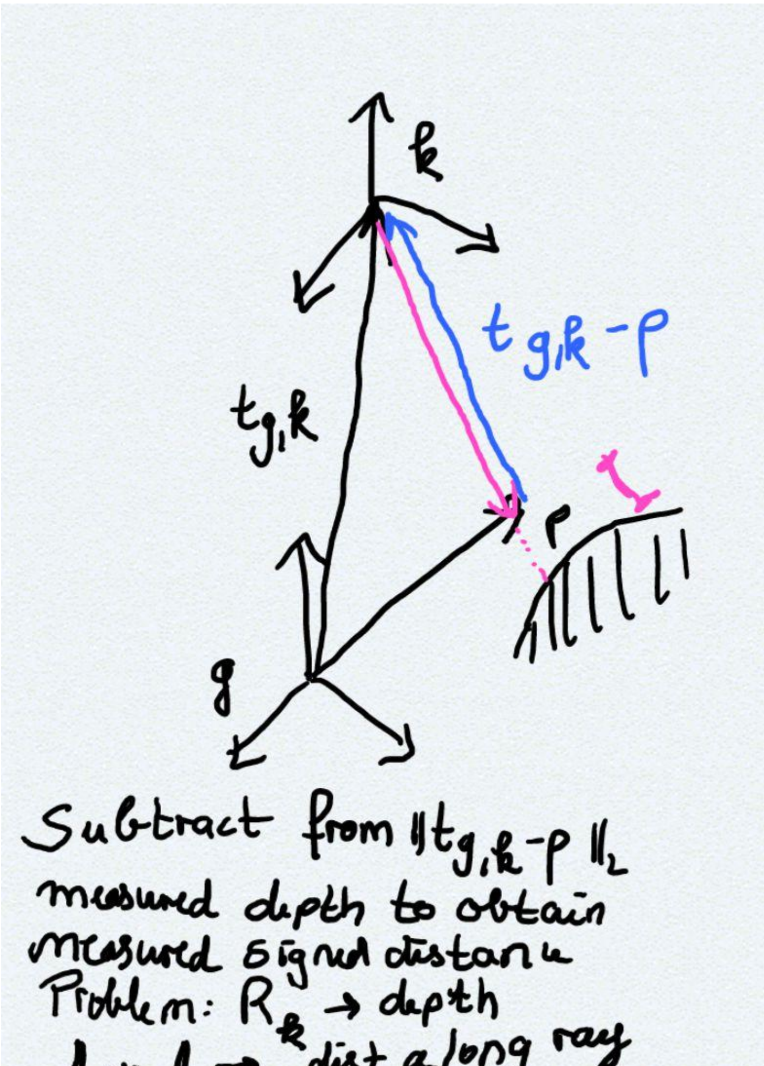
10. What is meant with a *projective TSDF* and why is it used?

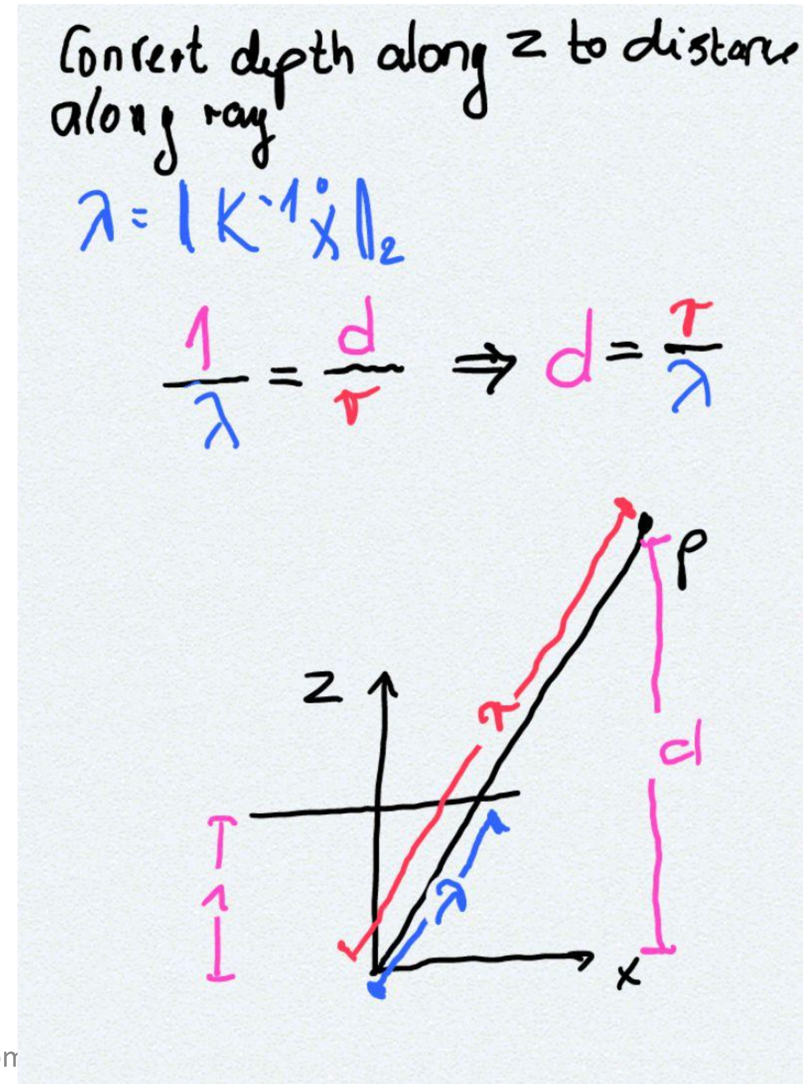
12. Why is the weight associated to a TSDF value computed as $\cos(\theta)/R_k(x)$? When does it get bigger, when smaller?

13. In (12), how do the weights W_k evolve over time and what is the effect of this?

Projective signed distance computation

Signed distance along viewing rays, not to actually closest surface!





Surface prediction (3.4)

Slides 38-41

10. What are *marging cubes* and *ray casting* used for, what is the major difference?

14. In (14), why is the gradient of the TSDF orthogonal to the zero level set, *i.e.* the surface? Can you illustrate/explain this in 2D?

15. Can you illustrate Equation (15) with a simple drawing and derive it? What is the mathematical principle behind?

Sensor pose estimation (3.5)

Slides 46-57

17. What is the difference between *frame-to-frame* and *frame-to-model tracking* and why is the second method preferred?

18. What are the principles of the tracking approach used in this paper?

16. What are the requirements for ICP to work and why is this the case?

19. Explain Equation (16). How are the vertices associated? Why is the point-plane metric used? How is the iteration obtained? Are there configurations that cause problems (*e.g.* so that the pose can't be estimated)?

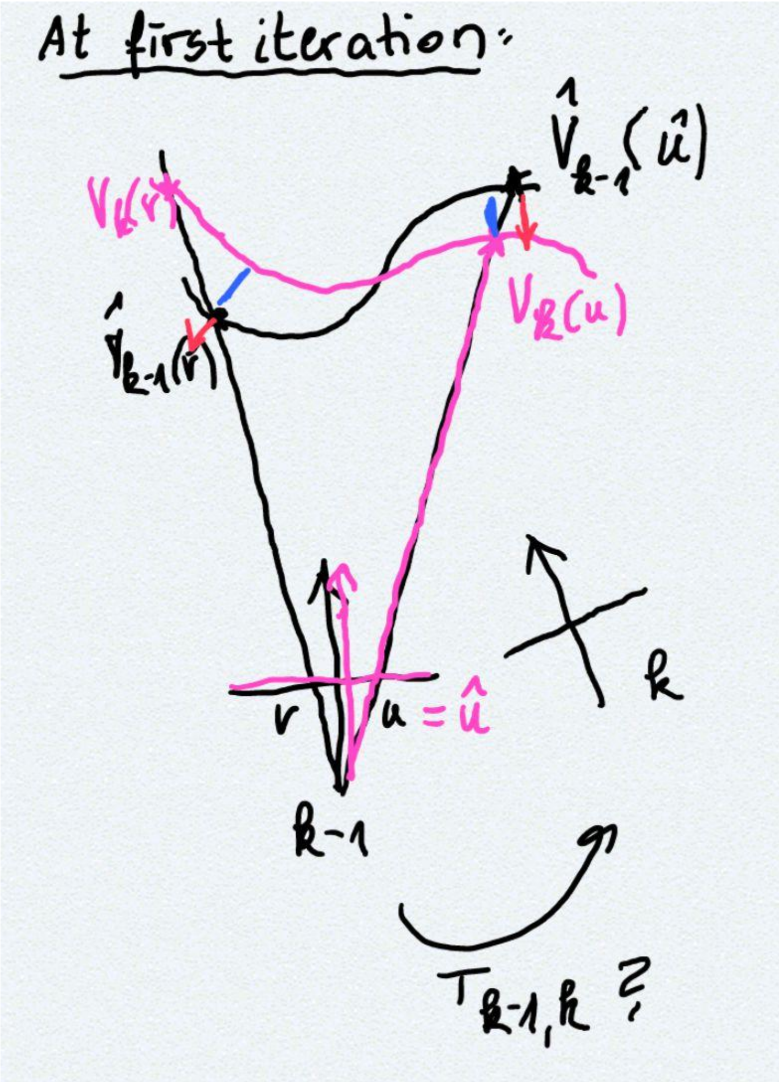
20. In Equation (17), what is tested here?

21. In Equation (18), what is the small angle approximation and how is it derived?

Sensor pose estimation (3.5) cont'd

22. What is the purpose of the reformulations in Equation (18) to (23)? Why do we compute the derivative in Equation (23)?
23. Why is a *coarse-to-fine* approach used for pose estimation and what does it mean?
24. What are the effects of large frame-to-frame motion?
25. When minimizing the cost function in Equation (23) (or the original version in (16)), is the pose estimate always fully defined or can you find an example, where it is not fully defined?

ICP with projective data association



- Point plane metric compensates for error through projective data association
- Doesn‘t work, if you look at just one plane!
- The reformulations in (18)-(23) are just done for efficiently solving the nonlinear minimisation problem in (16)!
  - Linearisation
  - Set derivative to zero → provides optimum
  - Solve linear equation system (see e.g. parameter estimation lecture)

Experiments (4)

[Slides 58-63]

1. What do the terms *metrically consistent model*, *local loop closure*, *explicit joint estimation* and *convergence properties* mean?
2. What are the major differences in the experimental setups 1-5 and what are the outcomes?
3. Why does dropping key frames in Experiment 5 reduce drift?
4. What are the evaluation criteria?
5. In Figure 13, which part of the processing depends most on the resolution of the voxel grid and why?

Conclusion (5)

Slide 65

1. Which limitations does the system have and what are the suggested improvements?
2. Do you have ideas for improvements?

Open source software – try it out

- Point Cloud Library (PCL)
- Large-scale KinectFusion implementation: http://www.pointclouds.org/documentation/tutorials/using_kinfu_large_scale.php#using-kinfu-large-scale

Announcements

- SS 2012
  - Computer Vision: Object and People Tracking (V2+1)
  - 3D Computer Vision & Augmented Reality (2S and 4P)
  - http://av.dfki.de/seminars_and_projects/
- Anytime:
  - Bachelor and master theses
  - http://av.dfki.de/thesis/ (or write an email)

Oral exams/question lecture

- Qualification: A minimum average score of 60% in the exercises
- Period: 11.03.2013 - 15.03.2013
- DFKI, room 1.27
- Registration: email to Leivy Michelly Kaul (kaul@dfki.uni-kl.de)

- No lecture next week
- Question lecture:
  - last week before exams (4.3.-8.3.), date will be published on the website