Is Crowdsourcing feasible for optical flow Ground Truth generation?

Axel Donath, Daniel Kondermann – HCI Heidelberg

ICVS 2013, St.Petersburg
Overview

1. Introduction

2. Ground Truth via Mechanical Turk

3. Experiments & Results

4. Conclusion
1. Introduction
Motivation

Start

Sequence taken from [3]
Large scale dynamic outdoor scene

Frame 1
Large scale dynamic outdoor scene

Frame 2
Large scale dynamic outdoor scene

Frame 3
Large scale dynamic outdoor scene

Frame 4
Large scale dynamic outdoor scene

Frame 5
Large scale dynamic outdoor scene

End

Sequence taken from [3]
Flow field estimated by algorithm

Optical flow algorithm: **Classic+NL** [5]

Color **legend:**

Crowdsourcing for Ground Truth generation ICVS 2013
Principles to obtain Ground Truth

(1) Measurement with suitable setups
   E.g. Middlebury dataset [2]

(2) Simulate data with computer graphics

(3) Data can be annotated by humans
   „Human assisted motion annotation“ with Motion-Annotation-Tool, proposed by Liu et. al [1]
Manual labeling and tracking

Start

Sequence labeled with Motion Tool [1]
Manual labeling and tracking

Frame 1
Manual labeling and tracking
Manual labeling and tracking
Manual labeling and tracking
Manual labeling and tracking
Manual labeling and tracking

End

Sequence labeled with Motion Tool [1]
Idea

Outsource manual correction of outlines and finding of feature points to Mechanical Turk
2. Ground Truth via Mechanical Turk
General workflow

- Initial segmentation
- Tracking of outlines
- Correction of outlines
- Selection of feature points
- Selection of motion models
- Ground Truth

Trained user

Laymen via Mechanical Turk

Trained user
Mechanical Turk workflow

- Blur outlines
- 5 „HITs“ per outline
- Download & Review results
- Merge & import outlines

Correction of outlines
Selection of feature points

Laymen via Mechanical Turk

Webinterface DEMO
Mechanical Turk workflow

Webinterface DEMO

Correction of outlines
Selection of feature points

Laymen via Mechanical Turk

Divide image into patches
8 points per patch
Download & Review results
Import feature points
3. Experiments and Results
Outline correction of simple scenes

Outlines before...

...and after correction by the workers
Results on simple scenes I

Endpoint error of six runs on the „Rubber Whale“ sequence:

AEE = 0.37
AEE = 0.79
AEE = 0.51
AEE = 0.37
AEE = 0.47
AEE = 0.63

All images are normalized to max. endpoint error of 2 pix
Results on simple scenes II

Endpoint error with overlapping patches:

- Overlapping patches tend to result in better AEE!
- Largest deviation in region of background fabric due to non rigid motion
- Bias due to bad correspondences
Results on simple scenes III

Endpoint error with high resolution image:

AEE = 0.20

No significant improvement in endpoint error.
Results on complex scenes I

Endpoint error on „Dimetrodon“ and „Urban“ sequences:

- Larger AEE due to **non rigid motion** (Dimetrodon)
- Error due to **single layer building** in the foreground (Urban)
Outline correction of complex scenes

Outlines
before...

...and after correction by the workers
Results on complex scenes II

Endpoint error on Sintel [4] sequence:

- Larger deviations in complex regions (hair)
Flow field estimated by crowdsourcing

Estimated accuracy of 1 pixel
Crowdsourcing for Ground Truth generation ICVS 2013

Time effort

- Initial segmentation: 1h – 2h (simple scenes)
- Tracking of outlines: Trained user: 2h – 3h
- Correction of outlines: MTurk workers:
  - 1 – 2d in total
  - 2 - 4min. per HIT
- Selection of feature points: 1h - 2h
- Selection of motion models

Ground Truth
Costs

Simple scene (Rubber Whale)
- 3.5 $/frame
- 10 $/frame

Complex scene (Sintel)
- 17 $/frame
- 25 $/frame

Trained user (simple scene)
- 10 $/frame
- 20 $/frame

Outline correction
Feature points
4. Conclusion
Summary

- **Accuracy** is around 1 pixel
- **Reduced accuracy** when non rigid motion is present, due to improper motion models
- **Reduced precision** but similar accuracy compared to trained workers on simple scenes
- **Savings** up to 40% per frame

Suitable method, where otherwise no flow estimation at all would be available and pixel accuracy is sufficient
Future work

- Replace work of trained user:
  - Automatic estimation of flow field
  - Let MTurk workers do the initial segmentation
- Better and more suitable motion models

Thanks for your attention!
Future work

• Replace work of trained user:
  ➣ Automatic estimation of flow field
  ➣ Let MTurk workers do the initial segmentation

We can generate cheap ground truth for you! Ask Daniel!

Thanks for your attention!
References


Spare slides
Experiments

General procedure:

• Test method on datasets with known ground truth to evaluate **accuracy**
• Perform multiple runs to evaluate **precision**
• Test on **real** as well as **synthetic data**
• Test on **simple** as well as **complicated scenes** to find out limitations of human perception
• Accuracy is measured with **average endpoint error** (AEE) compared to GT
Segmentation webinterface

Adjust the outline of the object

Read these short instructions before you start, they will make completing the task easier. If you don’t follow the instructions, your work will be REJECTED.

We would be happy to receive suggestions for improvement, or any other feedback that you want to share in the form below.

You can leave your feedback here

Submit Results

Change contrast  Reset

Zoom 1.6
Feature points webinterface