Mathematisch-Naturwissenschaftliche Fakultät

Computergrafik

Bachelorarbeit

Realistic MVS dataset

Eberhard Karls Universität Tübingen
Mathematisch-Naturwissenschaftliche Fakultät
Wilhelm-Schickard-Institut für Informatik
Computergrafik
Peter Trost, peter.trost@student.uni-tuebingen.de, 2019

Bearbeitungszeitraum:	von-bis
Betreuer/Gutachter:	Prof. Dr. Hendrik Lensch, Universität Tübingen
Zweitgutachter:	Prof. Dr. Max Mustermann, Universität Tübingen

Selbstständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Bachelorarbeit selbständig und nur mit den angegebenen Hilfsmitteln angefertigt habe und dass alle Stellen, die dem Wortlaut oder dem Sinne nach anderen Werken entnommen sind, durch Angaben von Quellen als Entlehnung kenntlich gemacht worden sind. Diese Bachelorarbeit wurde in gleicher oder ähnlicher Form in keinem anderen Studiengang als Prüfungsleistung vorgelegt.

Peter Trost (Matrikelnummer 4039682), May 22, 2019

Abstract

Template

Acknowledgments

If you have someone to Acknowledge ;)

Contents

1. Introduction

What is this all about?

Cite like this: [AFS+11]

2. Related Work

2.1. Synthetically rendered datasets

2.1.1. A naturalistic open source movie for optical flow evaluation

[BWSB12]

Figure 2.1.: Example images (ground truth flow in uneven rows and RGB in even ones) from the Sintel dataset

Overview

In this paper the authors provide a dataset for optical flow estimation derived from the open source 3D animated short film Sintel **TODO: cite Sintel: https://durian.blender.org/**. The dataset contains long sequences, large motions, specular reflections, motion blur, defocus blur, atmospheric effects and more. Its scenes are rendered in varying

complexity through the source graphics data provided by the authors of the film. Because of this aforementioned variety the dataset can be used to improve optical flow methods.

Render passes

As mentioned above the dataset contains image sequences rendered in the following variying complexity:

- Albedo Pass: Flat and unshaded. Surfaces exhibit constant albedo over time
- Clean Pass: Illumination including smooth shading and specular reflections adds realism
- Final Pass: Full rendering with all effects including blur due to camera depth of field and motion, and atmospheric effects.

Main aspects

The main aspects of the Sintel dataset are the following:
It contains varying and more challenging (for existing methods) scenes than older datasets. Sequences are 50 frames long and are provided with 49 ground truth flow fields which are a measure of changes in position for objects in the scene from frame to frame. Some frames include motions of well over 100 pixels. There are 1628 frames with 564 for testing and 1064 for training. The Sintel dataset contains sequences having real-world challenges like lighting variations, shadows, complex materials, reflections and more.

Meta

The authors modified Blender's internal motion blur pipeline to give accurate motion vectors at each pixel which provide ground truth optical flow maps. Although the clips are selected so that optical flow is realistic, one still has to be cautios when training and evaluating algorithms that strongly rely on real-world laws of physics. The images are saved as 8-bit PNG files and the clips have a framerate of 24 fps.

2.1.2. Playing for data: Ground truth from computer games

[RVRK16]

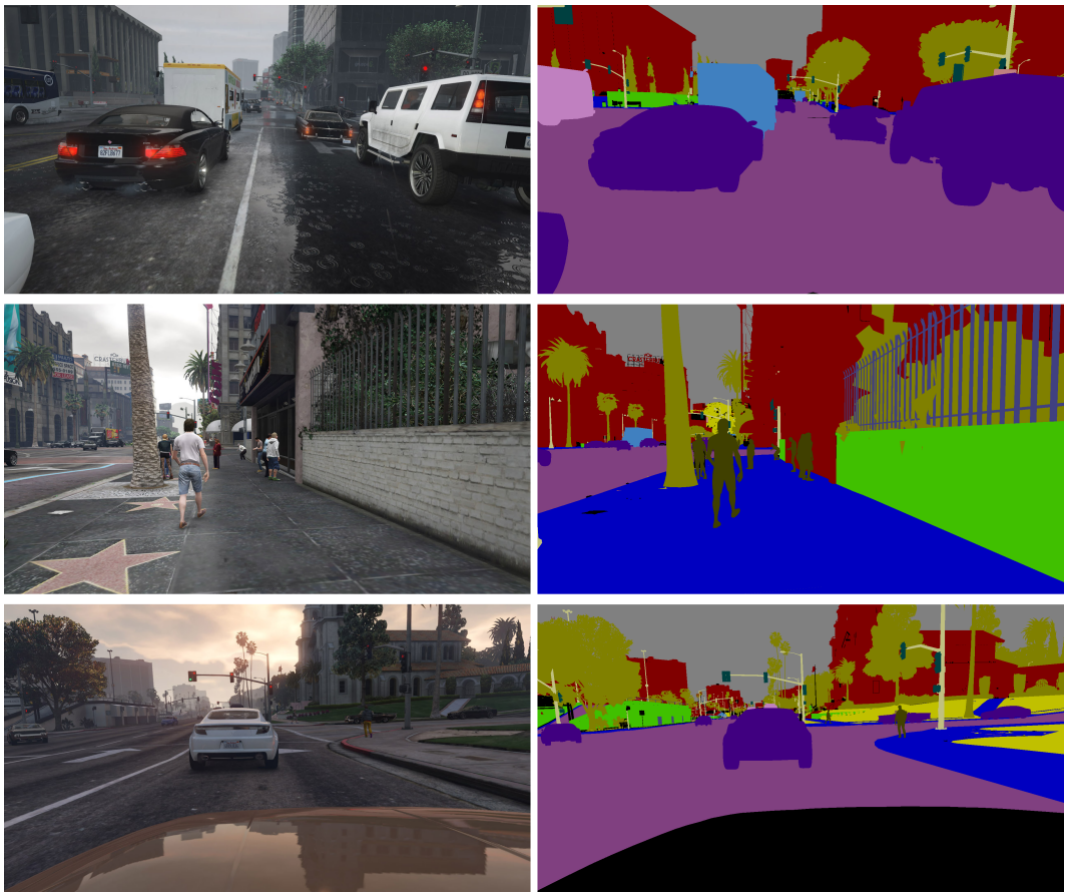

Figure 2.2.: Example images (RGB on the left and semantic segmentation right) from the ”Playing for data“-dataset

Overview

The main aspect of this paper is to get pixel-accurate ground truth of synthetic data and therefore be able to label objects in the images accurately and efficiently. The authors use detouring (i.e. injecting a wrapper between the game and the operating system) to record, modify and reproduce rendering commands from the game Grand Theft Auto 5 (GTA5). They retrieve the distinct rendering resources (geometry, textures, shaders) which they hash in order to create object signatures. These signatures are then used to label the objects pixel-accurately. The signatures then enable them to propagate these labels across time and instances that share distinctive resources. The dataset of the paper contains 25,000 images from GTA5

with pixel-level semantic segmentation ground-truth. Labeling the data took 49 hours which is 3 orders of magnitude faster than other semantic segmentation datasets with similar annotation density). This is achieved through the object signatures: When an object is labeled in a given image this label is propagated to every image that contains this object using the object signatures.

Extracting information from the rendering pipeline

Games communicate with the hardware through APIs (Application Programming Interface) like OpenGL, Direct3D, Vulkan and more. The authors implemented a wrapper for the DirectX 9 API and used the program RenderDoc **TODO: link renderdoc.org** for wrapping Direct3D 11. This enables them to monitor creation, modification and deletion of resources used to specify the scene and synthesize an image. With this approach every 40th frame during a Gamingsession was recorded. These recorded frames are processed in batch after a gameplay session. To make the data suitable for annotation the authors modified RenderDoc.

2.1.3. The SYNTHIA dataset: A large collection of synthetic images for semantic segmentation of urban scenes

[RSM+16]

Overview

This dataset consists of synthetic images of urban scenes and was generated to aid semantic segmentation in the context of autonomous driving. It provides photo-realistic frames from multiple view points together with associated depth maps and pixel-level semantic annotations for 13 classes. The dataset was generated by rendering a virtual city created with Unity development platform **TODO: cite unity website** and includes four different seasons (see 2.3) with drastic change of appearance due to simulated weather conditions, a variety of illumination conditions (day- and nighttime) and more.

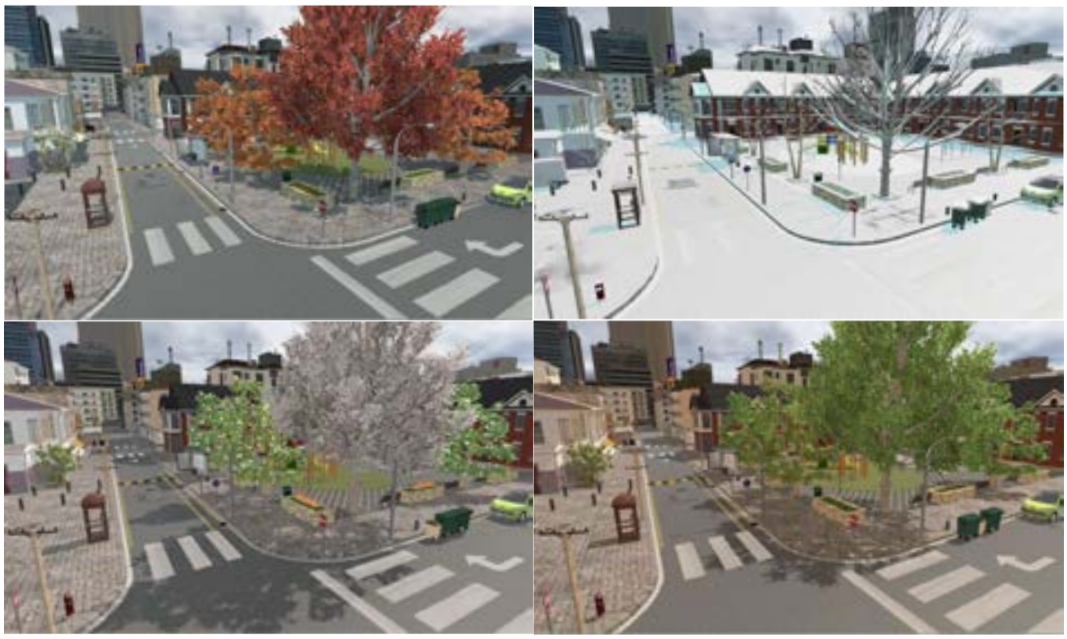

Figure 2.3.: Example images from the SYNTHIA dataset of each of the four different seasons (from top left to bottom right: autumn, winter, spring, summer)

2.1.4. SYNTHA-Rand and SYNTHIA-Seq

The dataset contains two categories: SYNTHIA-Rand and SYNTHIA-Seq. SYNTHIA-Rand is a collection of images gathered by moving the camera around the city randomly. It contains 13,400 frames of the city. SYNTHIA-Seq contains four video sequences with approximately 50,000 frames each. There is one sequence per season

provided that simulates a car moving through the city. Including interactions with objects, speeding up and slowing down and omnidirectional view (cameras in all 4 directions (see 2.4)).

Figure 2.4.: Example images from the omnidirectional view provided in the SYNTHIA-Seq dataset including corresponding depth maps

2.1.5. SyB3R: A Realistic Synthetic Benchmark for 3D Reconstruction from Images

[LHH16]

Overview

The authors propose a framework to evaluate 3D reconstruction algorithms using realistically rendered images. "Realistic" is defined as not only photo-realistic (images look real) but also in a physical sense. For this they use path tracing instead of ray tracing to be able to simulate more complex light-surface interactions. Real world effects like motion blur and noise are simulated during image rendering or post-processing. For this framework all camera parameters and the 3D structure of the scene are known. The image formation process is split into rendering the 2D image and then post-processing that image for additional effects. The authors also provide a dataset with realistically rendered images and their corresponding ground truth depths (see 2.5).

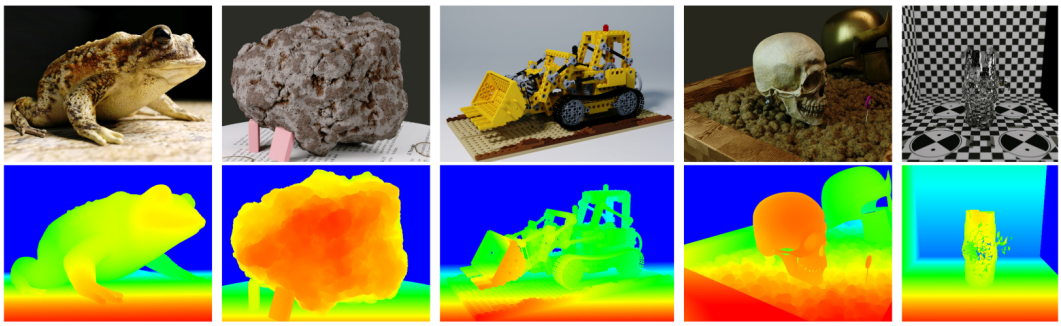

Figure 2.5.: realistically rendered Images from the SyB3R dataset (top) with corresponding ground truth depth (bottom)

Image Rendering

To render the images "Cycles" is used in Blender. "Cycles" contains a Monte-Carlo path tracer for accurate propagation of light through the scene **TODO: link cycles-renderer.org**. It handles scene properties like lighting, surface texture and more, object motion, large camera motion and camera properties including focal length, principal point, resolution, depth of field (DoF) and field of view). The rendered images are stored in HDR format to retain full floating-point precision of all intensity values.

Post-processing

For post-processing the authors provide a modular pipeline seen in 2.6 where the modules are interchengable.

Figure 2.6.: SyB3R post-processing pipeline

2.2. Problem Statement

TODO: what you have to do here :)

3. Conclusion

To conclude...

A. Blub

Bibliography

[AFS+11] Sameer Agarwal, Yasutaka Furukawa, Noah Snavely, Ian Simon, Brian Curless, Steven M. Seitz, and Richard Szeliski. Building rome in a day. *Commun. ACM*, 54(10):105–112, October 2011.

[BWSB12] D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black. A naturalistic open source movie for optical flow evaluation. In *European Conf. on Computer Vision (ECCV)*, Part IV, LNCS 7577, pages 611–625. Springer-Verlag, October 2012.

[LHH16] Andreas Ley, Ronny Hänsch, and Olaf Hellwich. *SyB3R: A Realistic Synthetic Benchmark for 3D Reconstruction from Images*, pages 236–251. Springer International Publishing, 2016.

[RSM+16] German Ros, Laura Sellart, Joanna Materzynska, David Vazquez, and Antonio Lopez. The SYNTHIA Dataset: A large collection of synthetic images for semantic segmentation of urban scenes. 2016.

[RVRK16] Stephan R. Richter, Vibhav Vineet, Stefan Roth, and Vladlen Koltun. Playing for data: Ground truth from computer games. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, editors, *European Conference on Computer Vision (ECCV)*, volume 9906 of *LNCS*, pages 102–118. Springer International Publishing, 2016.