A Repository for the Evaluation of Image-based Orientation Tracking Solutions

Christopher Coffin*
University of California Santa Barbara

Jonathan Ventura†
University of California Santa Barbara

Tobias Höllerer‡
University of California Santa Barbara

ABSTRACT

We describe an online repository we have developed for evaluating image-based orientation tracking methods. We have collected many videos which contain rotation-only camera movement under a wide variety of conditions, such as changing illumination, position, and rotation speed and direction. The dataset is useful for testing the robustness of orientation tracking systems, as well as other systems which use panoramas as a data source. In this paper we discuss the design of the repository and give examples of various uses of the imagery and other data it contains.

Index Terms: B.8.0 [Computing Methodologies]: Image Processing and Computer Vision—Applications; I.6.3 [Computing Methodologies]: Simulation and Modeling—Applications

1 INTRODUCTION

Panoramas have become a useful tool for remote navigation of physical locations as demonstrated by systems such as Google Streetview [1]. Panoramas have also been used as tools to aid in the localization of cameras such as in [4]. Systems which allow for the robust creation of panoramas by end users can enhance the availability of panorama data and allow users to collect datasets for a specific environment or task. Such tools have been proposed, for example, by Wagner et al. [6] and Kim et al. [3]. However, the practicality of these solutions for the average non-expert user is somewhat limited and the development of more robust solutions is a subject of ongoing research. Our goal is therefore to provide tools and data for the evaluation of image-based orientation tracking solutions. To this end, we provide a repository of tracking data related to orientation tracking information, which can be found at the URL tracking.mat.ucsb.edu. Our tracking repository shares basic similarities to other repositories of video data such as the Archive of Many Outdoor Scenes project [2]. Our datasets provide useful and necessary tools for the evaluation of orientation tracking systems. The datasets we provide have been captured using a camera mounted onto a high-speed pan-tilt unit which enables very precise control and in turn, a very accurate ground truth. This configuration allows us to capture such effects as: realistic camera noise, natural movement of objects in the scene (e.g. leaves, etc), motion blur, and changing exposure settings for auto exposure cases.

2 DATASETS

The datasets we provide can be divided into several categories based on the nature of the evaluation they support.

2.1 Natural movement

One interesting dataset comes from orientation information collected from a user study in which a set of 23 users was tracked over a set of 9 tasks consisting of search and exploration tasks. In addition to the raw orientation information, we provide the camera information recorded during the study along with video captured using the previously mentioned pan-tilt unit and camera combination when playing back the orientation data in various locations. This provides a realistic dataset for the natural head motion (rotation) of users. Tracking accurately using these datasets is a very difficult task.

Figure 1: Examples of datasets from the repository. *Top:* Variable speed dataset exhibiting motion blur. *Middle:* Variable exposure settings. *Bottom:* Variable time of day – one of the samples collected over a period of six hours. The images represent only the first camera frame taken from the larger video dataset.

2.2 Variable speed

We provide video which can be used to evaluate the upper bounds of the speed of movement of an orientation tracking system. These videos were recorded from several locations, using different speeds from slow to fast. In each location, the position and orientation path of the camera is exactly the same; the only variable is the rate at which the camera moves. This dataset is particularly suited for stress-testing the robustness of real-world systems, as it provides natural noise and motion blur effects in the tracked video. For example, the orientation output of the tracker for each video can be compared to determine the effect of noise and motion blur on tracking accuracy.

2.3 Lighting changes

We also provide datasets for the evaluation of tracking systems with respect to changes in lighting. There are two aspects of this, which we consider separately. One difficulty is changes in lighting due to differences in ambient illumination or changes in the exposure setting of the camera. This is addressed by capturing panoramas both using a fixed exposure, and using the auto-exposure function of the camera. Ideally, an orientation tracking method should be invariant to exposure settings or ambient illumination changes, and should produce the same results under any lighting condition. A second issue is the complex change in illumination over the period of a day, which includes directional light, moving shadows, and changing color. To support evaluation of tracking robustness across time,

*e-mail: christopher.coffin@gmail.com
†e-mail:jventura@cs.ucsb.edu
‡e-mail:holl@cs.ucsb.edu

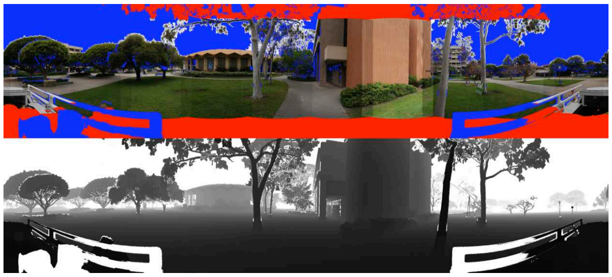

Figure 2: Example panorama with terrestrial LIDAR depth scan.

we collected video samples at various times of day, an example of which can be seen in Figure 1. At each location we recorded the exact same camera movement at half hour intervals, and captured a wide range of lighting conditions for each scene. Orientation tracking systems should be robust to such changes in illumination.

2.4 Position changes

Another set of panoramas in the repository provides orientation-tracked videos taken in a straight line along the ground, at evenly spaced intervals. These panoramas were created using a camera on a pan-tilt unit on a tripod. At each capture point we recorded a 360 degree view of the scene. The capture points were evenly spaced using a measuring tape on the ground. This dataset is useful for experiments which require a baseline between panoramas. For example, one could test the resilience of orientation-only tracking to slight translations. It is also possible to perform structure-from-motion using the panoramic sequences (see Section 3.2).

2.5 Orientation and Depth Datasets

Although many applications are possible using orientation-only tracking, it is also useful to have depth information about the scene. To support experiments which require 3D scene information, we have collected panoramas which are registered to 3D LIDAR scans (see Figure 2). The scanner produces a very dense point cloud as well as a high-resolution color panorama. We also have recordings of a single-point laser rangefinder which produces a sparse point cloud registered to the panorama. This type of data enables experiments in 3D camera localization, 3D scene annotations, and other applications which require a geometric scene representation.

3 Evaluation Examples

The repository presents a rich source of data for many different types of evaluations. Here we illustrate examples of some evaluations we have investigated using the panoramic datasets.

3.1 Comparing Patch Match Methods

We performed a comparison between two patch matching methods to determine which one is more suitable for tracking on images of outdoor urban scenes. Zero-mean summed squared differences (ZMSSD) is a technique which is fast to compute, and is invariant to bias (brightness). Normalized cross-correlation (NCC) is a bit more complex to compute, but is invariant to bias and scale (brightness and contrast). To compare the methods, we used images from the variable time of day dataset. This dataset offers images taken at precisely the same location and orientation, but at different times, thus exhibiting many changes in lighting conditions.

We detected SIFT features on a reference image, and then computed patch difference for each patch to all of the other images in the dataset taken from the same position and orientation. We also computed the patch difference to another image from the dataset which has a different orientation. Thus, all difference scores to the first image should be classified as correct matches (inliers), and all difference scores to the second image should be classified as incorrect matches (outliers).

Figure 3 shows box plots illustrating the distribution of scores by hour for the inlier image (top) and outlier image (bottom). The green area shows a possible region where matches can be considered inliers. We see that this makes for a clear delineation between inliers and outliers for the NCC metric, but not for the ZMSSD metric. Thus from our evaluation we conclude that NCC is a more useful metric than ZMSSD for patch-based trackers operating outdoors, and we suggest an inlier threshold of 0.7.

3.2 Structure from Motion

The panoramic datasets in our repository can be used not only for orientation tracking, but also for full 3D reconstruction. We experimented with point cloud reconstruction using the variable position datasets. An example reconstruction is shown in Figure 4. Our reconstruction pipeline uses the "upright constraint" which is the assumption that all cameras share a common vertical orientation [5]. This assumption is valid for our dataset, since the panoramas were carefully captured using a level tripod.

Because the panoramas are known to be evenly spaced, the drift in the reconstruction can be examined by computing the distance between reconstructed camera locations. This provides one way to rigorously test a structure-from-motion pipeline.

A second experiment which we are currently pursuing is to use the datasets to determine how many panoramas are needed to sufficiently model a scene. We have developed a 6DOF localization system which determines the pose of a mobile device by querying the panorama point cloud reconstruction [4]. A pertinent example research question would be how many panoramas are needed in a particular scene in order to robustly support localization queries. The dataset described here could be used to answer this question by testing the modeling and localization performance achieved using various numbers and spacings of panoramas.

4 Conclusions and Future Work

Our repository contains datasets which enable a multitude of experiments to evaluate various aspects of visual tracking and localization from panoramas. More generally, the rigorous nature of the capture conditions means that many experiments concerning illumination and geometric image analysis are possible.

In the future we would like to capture some videos which could be used to create extremely high resolution "gigapixel" panoramas, which cover the entire spherical field of view. We would also like to expand upon the variable time of day datasets to provide images across all times of day and seasons. This would provide a more comprehensive dataset like that of AMOS [2], but with higher resolution, full-view panoramas rather than small webcam images.

We have illustrated here some examples of evaluations and experiments that can be performed using the videos in our repository. We believe that the datasets in our repository can be widely used by researchers in computer vision and augmented reality.

References

[1] D. Anguelov, C. Dulong, D. Filip, C. Frueh, S. Lafon, R. Lyon, A. S. Ogale, L. Vincent, and J. Weaver. Google street view: Capturing the world at street level. *IEEE Computer*, 43(6):32–38, 2010.

[2] N. Jacobs, W. Burgin, R. Speyer, D. Ross, and R. Pless. Adventures in archiving and using three years of webcam images. pages 39–46, 2009.

[3] S. Kim, C. Coffin, and T. Höllerer. Relocalization using virtual keyframes for online environment map construction. In *VRST'09: Proc of the 16th ACM Symposium on Virtual Reality Software and Technology*, pages 127–134. ACM, 2009.

[4] J. Ventura and T. Höllerer. Outdoor mobile localization from panoramic imagery. In *Poster. Proc. IEEE/ACM ISMAR 2011 (10th Intl Symposium on Mixed and Augmented Reality)*, Basel, Switzerland, Oct 2011. IEEE Computer Society.

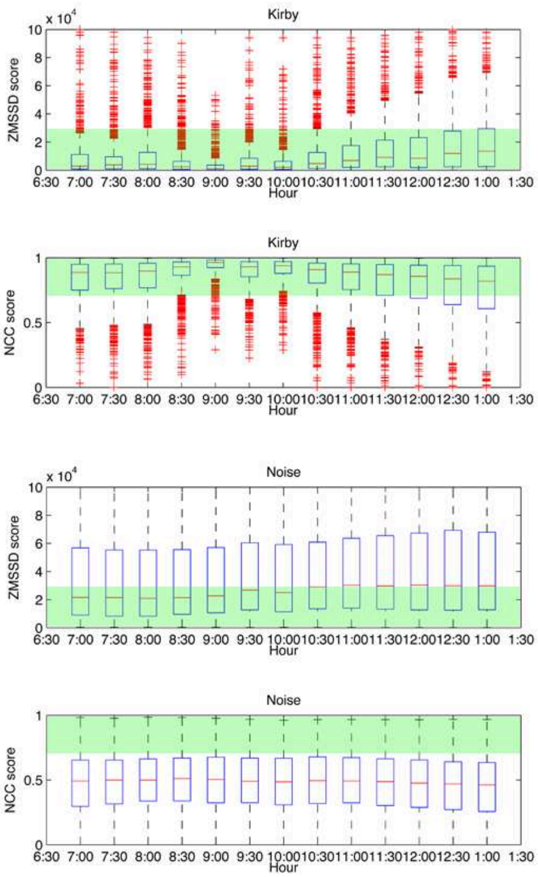

Figure 3: These box plots illustrate the distribution of patch matching scores for two difference methods (ZMSSD and NCC), given an identical image taken at a different time (top two plots) and an unrelated image (bottom two plots). These distributions help to determine a practical inlier threshold (shown in green) for the difference methods (shown in green).

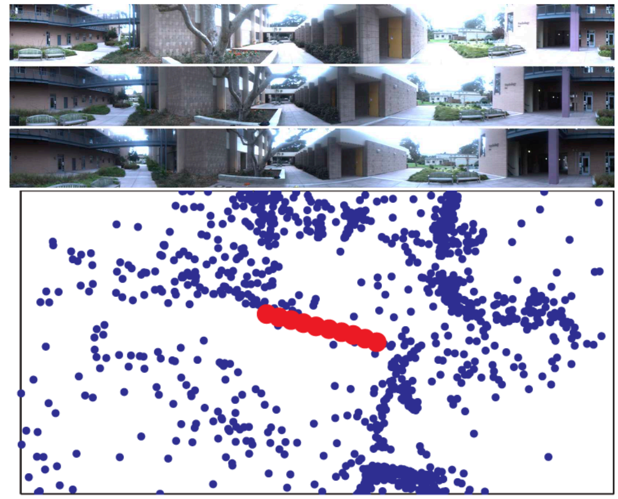

Figure 4: Top-down view of a point cloud reconstruction made using panoramas from our repository.

[5] J. Ventura and T. Höllerer. Structure and motion in urban environments using upright panoramas. *Virtual Reality (under review)*, 2011.

[6] D. Wagner, A. Mulloni, T. Langlotz, and D. Schmalstieg. Real-time panoramic mapping and tracking on mobile phones. In *Proc. IEEE VR*, pages 211–218, Washington, DC, USA, mar 2010. IEEE Computer Society.