Evaluation of tracking robustness in real time panorama acquisition

Christopher Coffin* Sehwan Kim† Tobias Höllerer‡
University of California, Santa Barbara

ABSTRACT

We present an analysis of four orientation tracking systems used for construction of environment maps. We discuss the analysis necessary to determine the robustness of tracking systems in general. Due to the difficulty inherent in collecting user evaluation data, we then propose a metric which can be used to obtain a relative estimate of these values. The proposed metric will still require a set of input videos with an associated distance to ground truth, but not an additional user evaluation.

Keywords: Robustness metric, vision-based tracking, real-time panorama acquisition, expert evaluation, camera pose relocalization

Index Terms: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Virtual Reality; I.4.8 [Image Processing and Computer Vision]: Scene Analysis

1 INTRODUCTION

In this paper we present an extensive analysis of the performance of four existing orientation tracking systems. These methods are variations on an existing orientation tracking system Envisor [2]. We analyzed regular Evisor, Envisor with constant recovery, Envisor with selective recovery, and Envisor with constant recovery and prescanning. A detailed discussion of these methods can be found in [1]. In order to obtain an accurate understanding of the robustness of each system, we performed a quantitative analysis, an analysis of the performance output and a live evaluation. For the quantitative analysis, we collected distance to a known ground truth over a large set of input videos. However, ground truth error alone does not provide insight into the perceived robustness of the system. We obtain this through two qualititative analyses of the systems.

The first qualititative analysis was based on the final output of the systems, in our case, a set of environment maps. The second qualitative analysis focused on the results of a live expert evaluation of each system. While the analysis of the results allows for a larger breadth with respect to samples, expert analysis provides confirmation of the trends seen in the analysis of the results.

Based on the results of these experements, we propose guidlines for a metric which may eliminate the need to collect user evaluations for subsequent analyses. While these guidelines are meant to be generally applicable, the extension to other applications and tracking systems is left to future work.

2 DATA COLLECTION

For the ground truth and output analysis it was necessary to obtain a large set of meaningful motion data. As the primary use case for the analyzed systems is based on a head mounted camera, we collected a large set of head orientation information in a preliminary user study. Each of the 23 participants performed 9 different tasks. The tasks involved were a combination of search/exploration, counting, and casual observation. Some tasks were limited by time and others by completion. For the purposes of this study, we selected a random subset of those head motion sequences with a duration of a minute. The final set of data was then limited to 45 samples.

*e-mail: ccoffin@cs.ucsb.edu
†e-mail: skim@cs.ucsb.edu
‡e-mail: holl@cs.ucsb.edu

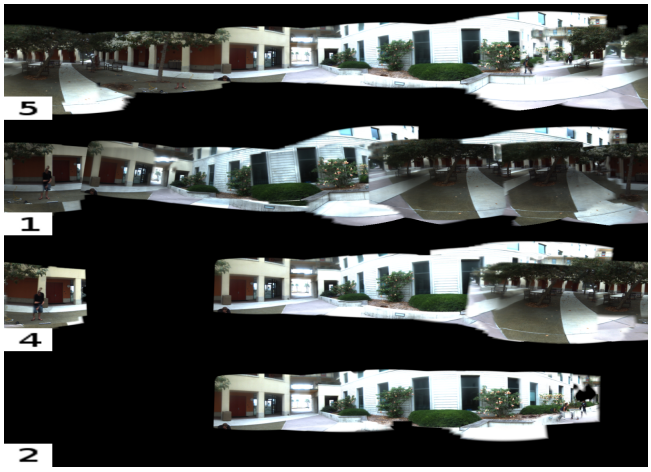

Figure 1: Users were asked to rank panoramas generated by each of the four methods used

3 EVALUATION

Tracking Error The ground truth values for each video were obtained by mouting a camera on a PTU-D46 pan tilt unit [3] from Directed Perceptions. This allowed us to precisely replay the orientation information collected previously, and to do so in multiple locations providing a sizeable representative set of tracking environments. A measure of absolute tracking error was then obtained by using each video as input to the tracking systems and comparing the resulting positional updates with the ground truth input to the PTU.

Result Evaluation For the analysis of the panoramas produced by each method, we designed a simple ranking program, with the user interface seen in Figure 1. For each data set, every expert was shown the panoramas generated by every method simultaneously. They then selected each panorama and rated them on a scale of 1 to 7. The assigned scores were then displayed on the left-hand side of the associated panorama. To assist in ranking, users were able to compare each panorama to a ground truth panorama.

The average ratings of the users for each set of environment maps was very consistent over each of the methods. Performing an analysis of variance single factor test with the independent variable being the method and the dependent variable being the ratings, resulted in a residual of 1:1940 with F = 572 and $p < 0.0001$. The results from a set of corresponding Tukey Post-hoc evaluations indicate that the methods performed with significant difference among all sequences. Every method was significantly pairwise different from each other with the exception of constant recovery compared with selective recovery in the indoor case. Note that this is important as it implies users were able to differentiate methods as more or less robust.

Live Evaluation For the live evaluation, we had 5 expert users evaluate each system in a live demonstration. In order to ensure a fair comparison, we had each user rank each method four times for a total of 16 randomly ordered runs. The evaluators were asked to rank each system from 1 to 7, and then we normalized and averaged the results for this evaluation similar to the panorama evaluation results.

4 Results

We present a relative comparison of each of the methods in Table 1. Note that for both of the qualitative analyses, the relative distance betwen the selective recovery method and the constant recovery method was very small. This distance is not only much greater in the ground truth error, but also reversed, indicating that there is not a direct linear mapping between eror and qualitative robustness.

Table 1: First row, relative distance to ground truth in degrees. Second row, ratings assigned to the panorama output data (scale 1 poor to 7 perfect). Third row, the robustness ratings from the live evaluation (scale 1 poor to 7 perfect). (CRS: Envisor with pre-scanning, SR: Envisor with selective recovery, CR: Envisor with constant recovery, NR: original version of Envisor (No Recovery))

	CRS	SR	CR	NR
Distance to ground truth	3.27	16.38	8.08	26.75
Panorama evaluation	5.41	3.54	3.12	2.03
Live evaluation	6.05	4.03	3.95	1.63

5 Toward User Free Analysis of Robustness

As mentioned in the introduction, both the result based and live evaluations contribute additional information to an understanding of the performance of the system. Along with our current findings, this implies there is not a simple linear mapping from the absolute error to the qualitative evaluations. Therefore, some types of error may have a greater effect on the qualitative result than others.

As a starting point for future discussion we propose guidelines for identifying three regions, in each of which, the effect of the tracking errors is of more or less importance.

The first region is determined by the application area of the tracking solution, in our case panorama construction. The format of the final results of the tracking determines the lower bound for noticable errors. There is some point at which a small amount of frame to frame noise or error is not noticable for the application. Therefore a system with a high percentage of errors in this region should not be penalized as heavily as a system with a high percentage of more obvious errors. In our case an empirical evaluation of artificially introduced errors showed that for a sphere map of 1536×512 pixels a shift of around 0.5 degrees at the equator is negligible. As an example of an alternative application area, an adaptation of this metric focusing on augmented reality displays would set this lower bound to a level at which augmentations have a drift or offset which becomes a distration to users.

We define the second and third regions by the threshold at which an error is non-recoverable. Such a tracking threshold is an upper-bound on the frame to frame error at which normal tracking breaks and some re-initialization method is needed. While the acceptable tracking threshold is based on the application area, this bound is dependent on the systems tested and must be determined for each newly introduced tracking system.

For our analysis the systems tested all share the same code base. Therefore, the recoverable tracking threshold for all systems was set to the point at which Envisor [2] using only frame to frame feature tracking is able to run without loosing frame to frame tracking. Our testing determined this to be 56° per second giving a maximum distance per frame of about 2.69°.

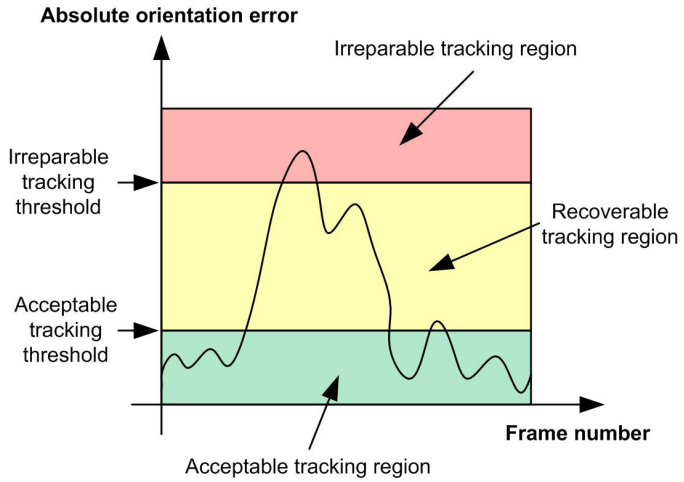

Figure 2: An illustration of acceptable, recoverable and irreparable tracking regions, and acceptable and irreparable tracking thresholds for an absolute orientation error graph

Given our thresholds, we organize the absolute tracking errors into three bins. We refer to these as the acceptable, recoverable, and irreparable tracking regions 2.

We propose the formula in Eq. 1 for determining a measurement of the robustness similar to the qualitative analysis. We base our metric around the percentage time spent in each error classification region. These percentages are obtained from N_T, N_A, N_R and N_I, where N_T is the number of total frames, and N_A, N_R and N_I denote the numbers of frames belonging to the acceptable, recoverable, and irreparable tracking regions, respectively.

$$Robustness = \alpha \cdot \frac{N_A}{N_T} + \beta \cdot \frac{N_R}{N_T} + \gamma \cdot \frac{N_I}{N_T} \quad (1)$$

Additionally, the equation requires weighting factors (α, β, γ). These weighting factors enable the mapping determining the qualitative robustness from the sets of ground truth data which are used as input. We derived the values for the weights using three of the four proposed methods and were able to confirm the accuracy of the metric by closely predicting the fourth.

Note that the metric itself does not contain references restricting its use to 3DoF tracking. All that is required is some measurement of distance to ground truth. It is unclear if the values of the mappings will remain consistent throughout multiple application areas and over multiple systems.

Acknowledgements

This work was supported in part by a research contract with KIST through the Tangible Space Initiative Project, and NSF CAREER grant #IIS-0747520. The authors also wish to thank Stephen DiVerdi for his development of the original Envisor system and his continued help.

References

[1] C. Coffin, S. Kim, and T. Höllerer. Evaluation of four methods for real time panorama aquisition. 2010.

[2] S. DiVerdi, J. Wither, and T. Höllerer. Envisor: Online environment map construction for mixed reality. In *IEEE VR*, pages 19–26, 2008.

[3] DPerception. http://www.dperception.com, June 2009.