# Optimization of Target Objects for Natural Feature Tracking

Lukas Gruber[1], Stefanie Zollman[2], Daniel Wagner[3] and Dieter Schmalstieg[4]
ICG
Graz University of Technology, TUG
Graz, Austria
lgruber[1],zollmann[2],wagner[3],schmalstieg[4]@icg.tugraz.at

Tobias Höllerer[5]
Four Eyes Lab
University of California, UCSB
Santa Barbara, California US
holl[5]@cs.usb.edu

*Abstract*— **This paper investigates possible physical alterations of tracking targets to obtain improved 6DoF pose detection for a camera observing the known targets. We explore the influence of several texture characteristics on the pose detection, by simulating a large number of different target objects and camera poses. Based on statistical observations, we rank the importance of characteristics such as texturedness and feature distribution for a specific implementation of a 6DoF tracking technique. These findings allow informed modification strategies for improving the tracking target objects themselves, in the common case of man-made targets, as for example used in advertising. This fundamentally differs from and complements the traditional approach of leaving the targets unchanged while trying to optimize the tracking algorithms and parameters.**

***Keywords-Natural feature tracking; tracking target optimization; simulation***

## I. INTRODUCTION AND RELATED WORK

Much work has gone into improving algorithms for detecting feature-points and estimating a 6DoF camera pose from a set of point correspondences. In the following such an algorithm is simply called 'tracker'. In comparison, the analysis and specifically the improvement of the natural objects to be tracked themselves received little attention from the research community so far. However, in industrial scenarios such as Augmented Reality (AR) applications, it is often the case that a certain tracker is given and can only be minimally tuned, whereas the texture of the object or environment to be tracked may be subtly or even drastically changed to improve tracking.

Many common 3D tracking techniques rely on feature points on textured surfaces, which are often part of man-made objects, for example, a billboard used in advertising or product packaging, on top of which shoppers may want to experience AR annotations. In this case, the application developer would like to understand the trackability of the target object, i.e., which parts of the target can be tracked accurately and robustly. The trackability assessment can aid the AR designer in making modifications to the target object to improve the quality of the tracking, as long as certain aesthetic goals are not sacrificed.

Previous work focuses on the analysis of feature sets, feature set improvement, and the evaluation of descriptors and tracking algorithms. Shi and Tomasi [4] used feature set analysis for the selection and monitoring of features during the tracking process to optimize the Kanade-Lucas-Tomasi (KLT) tracker. For feature set improvements, Knapek et al. [1] explore methods for selecting promising features from an image. The methods were tested empirically by simulating the possible changes of a feature window. All this work focuses on the analysis and improvement of the feature set selected from a given set of fixed objects, without considering modifications to the tracked objects themselves.

An evaluation system for local descriptors was presented by Mikolajczyk and Schmid [3], which has become a standard framework in the computer vision community. Similarly, Moreels and Perona [2] used a computer-controlled turntable and real objects to match 3D object features under different viewpoint and lighting conditions, in order to evaluate feature detectors and descriptors under more realistic conditions. Both methods compared algorithms, not objects/environments as we do here.

In this paper, we present a tool chain for the simulation of tracking target objects using a given feature-point based 3D tracker. In a pilot experiment, the tool chain was used to evaluate the trackability of a large number of planar, textured objects, to gather statistical relevant data on the quantitative influence of object parameters such as the number, texturedness, spatial distribution and similarity of features as well as environmental conditions such as lighting and camera pose. The accuracy of the simulation was verified with a robotic arm setup. We then show case studies how the tools can be applied to systematically improve the trackability of a given tracking target. Based on the findings about the quantitative influence of tracking target characteristics to the trackability of an object, we have chosen different image manipulation techniques such as contrast enhancement or content and structure adding to improve the trackability of a tracking target.

1051-4651/10 $26.00 © 2010 IEEE
DOI 10.1109/ICPR.2010.880

## II. TRACKING TARGET SIMULATION

We define a tracking target object as any physical object or environment with a surface texture that can be used to estimate the 6 DoF pose of a camera directed at it, using a vision based feature detection algorithm. This includes single objects as well as complete environments.

The success of the tracking process is dependent on the geometry and appearance of the tracking target, the tracking algorithms (e.g. feature detection, feature descriptors, and frame-to-frame pose estimation) and their parameters, as well as the environmental conditions (e.g., lighting, camera choice and parameters). In this paper, we address the situation that the designer of a real-time tracking experience wants to optimize tracking performance for a given tracker and fixed range of environmental conditions, by optimizing the tracking target.

Our work focuses on visual detection and tracking methods. Hence, a tracking target T can be represented by a set of natural feature points. As indicated above, the feature point set is dependent on the employed tracking algorithms (e.g. feature detector) and their parameters, the environmental conditions, and of course the surface texture and appearance of the target object itself. In most cases, the surface representation of T consists of simple 2D geometry with a certain topology (often piece-wise planar). We are interested in how to optimize these surfaces to improve tracking, and thus we explore the characteristics of different textures and their influence on tracking performance under various environmental conditions.

## III. EXPLORING DOMINANT TRACKING TARGET CHARACTERISTICS BY SIMULATION

To this end, we developed a simulation framework by synthesizing a vast variety of views onto sample tracking targets under varying conditions, thus gathering a sizeable amount of data for statistical evaluation.

In this work we only consider planar tracking targets, which can be printed on a sheet of cardboard in real live, or rendered as a single textured polygon for simulation purposes. Consequently, the pose can be computed as a homography. The natural feature tracker described in [7] was used as the tracking technique to collect the data. However, all functionality of the tracker such as feature point detection and pose estimation are invoked via a public interface, and consequently the tool chain is open to accommodate other tracking techniques.

## IV. SIMULATION OF TRACKING TARGET OBJECTS

As input for our simulation tool chain we used a database of 1188 images with a wide variety of subjects (peoples, cars, bikes, objects, animals) as tracking target sources. For each of the 1188 tracking targets, we created views for 200 camera poses at varying distances and rotations each under 6 lighting conditions, resulting in 1425600 synthesized views rendered at a resolution of 320x240 pixels, which is very common for real-time computer vision applications as for example in the AR domain. The views were analyzed with three feature sets per targets (500, 1000, 1500 key-points, total 20490 datasets). To compensate for variations introduced by random-seed RANSAC outlier removal in the tracker, we repeated each combination of view and feature set 10 times. The overall computation time for the 42.7 million tracker runs was 2.5 days on 4 desktop computers.

## V. SIMULATION VS. REAL WORLD SETUP

Synthesizing images for evaluation raises the question how reliably the results characterize real world situations. We therefore validated the results of our simulation framework using a Mitsubishi RV-1A robotic arm with six degrees of freedom. We attached printed tracking targets to the tip of the arm and recorded them with a Logitech Quickcam 9000 Pro mounted at a fixed position.

We calibrated the intrinsic and extrinsic camera parameters using the method of Tsai et al. [5]. After calibration, position and orientation of the robotic arm relatively to the camera coordinate system are known with an accuracy of 0.4 mm. The calibrated setup therefore enables the determination of exact ground truth data.

We evaluated 155 poses with 3 lighting conditions each resulting in 465 captured images per target. In total we attached seven different tracking targets to the robotic arm (see Fig. 2f). We then compared the ground truth from the robotic arm with corresponding synthesized images from our simulation pipeline. The result of our simulation vs. real-world comparison (see Fig. 1) is the median of the consistency rate C: 0.78. The average of C is 0.74 and the standard deviation is 0.23. We define the consistency rate C as follows in equation (1) where DSv and DSv are the detection rates of synthesized and real views, and views is the number of views per targets.

$$C = 1 - \sum_{v}^{views} \frac{|DS_v - DR_v|}{views} \quad (1)$$

The detection rate itself is the percentage of the views where a pose is detected successfully. The criteria for a positive pose detection is a minimum of eight detected key-points (after RANSAC optimization). The detection rate is used to compare the tracking performance or trackability of the simulation and the real-world test data.

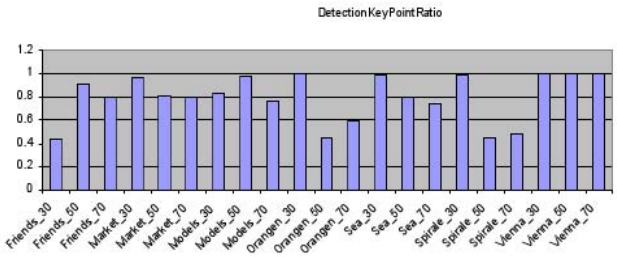


Figure 1. Validation setup results: Consistency C (y-axis) of detection ratio of simulated and real data.

## VI. EVALUATING TRACKING TARGET CHARACTERISTICS

For evaluating the relationship of inherent tracking target characteristics on the detection rate R we computed the correlation between R and the target properties as detailed

below using the Spearman correlation test. In the following, we present characteristics which show highly significant correlations ($p<.01$, $n = 20490$). Our expectations that texturedness and distribution have high impact on R were confirmed. Other characteristics such as feature similarity do not have such high importance for the employed natural feature tracker.

### A. *Feature texturedness (FT)*

The texturedness of a feature provides information on how strongly the image intensity in the feature window varies. Whereas Shi and Tomasi [4] used the Eigenvalues of the second moment matrix to determine the texturedness of features, we use the standard deviation of 8 bit intensities in the (in our case 8x8) feature window as a general and simple approach. The result from our evaluation for texturedness is a positive Spearman correlation of r=0.201. The greater the texturedness, the more poses are successfully detected. Investigating the influence of texturedness under varying lighting conditions shows that the correlation increases for underexposed and overexposed lighting. An ambient lighting model was employed. The global intensity of each light level corresponds to the brightness of the final simulated image. The correlation between texturedness and pose ratio gets stronger for extremer light levels: Level 2 (36% brightness, highly underexposed) r=0.289; Level 4 (52% brightness, moderate) r=0.187; and Level 7 (76% brightness, highly overexposed) r=0.326.

### B. *Spatial feature distribution (FD)*

We use a dispersion index d [8] to describe the distribution pattern on feature points in 2D over the surface of a tracking target. The dispersion index depends on the standard deviation s and the median m of the number of features per cell: $d = s^2/m$. As expected, the dispersion index has a negative correlation with the pose detection: 0.405. The correlation is getting weaker for extremer light levels: Level 2 (36% brightness, highly underexposed) r=0.460; Level 4 (52% brightness, moderate) r=0.610; and Level 7 (76% brightness, highly overexposed) r=0.443

### C. *Feature similarity (FS)*

Feature similarity of a keypoint describes how many similar keypoints exist in the same feature set, which are likely to be confused with this keypoint. Naturally this metric strongly depends on the way the used tracker builds descriptors (in our case the length of the descriptor is 36), such as on the size of the support region used to describe keypoints. A keypoint's similarity value is given by the number of keypoints with feature descriptors that are similar to the original keypoint descriptor within a summed Euclidean distance below a certain threshold (in our case, 19200). The Spearman test shows that the similarity has a very weak negative correlation with the pose detection (r=0.053).

### D. *Number of features (FN)*

The total number of features of a tracking target was intuitively expected to have an influence on the pose detection quality: Targets, which are poor on features, are generally hard to detect. The correlation test confirms a positive correlation (r=0.196).

## VII. TRACKING TARGET OPTIMIZATION

In the following section, we discuss possibilities how to improve tracking target objects with common image manipulation tools. We applied sigmoidal non-linear contrast modification without saturating highlights or shadows (see [9]) using the image magic toolbox [10] to 40 target images and then submitted the manipulated images to the simulation tool chain, calculated R for 200 camera poses and 6 light levels. The "α" value indicates how much the contrast is increased and the "β" value defines where the mid-tones fall in the output image (0: white, 50%: middle-gray, 100%: black).

The results were compared to the results from the original images. The analysis of R showed that the contrast enhancement has a positive effect to the tracking rate. For example, a contrast enhancement of α=14 and β=50% increases the tracking rate from 0.14 to 0.35, which can be explained by the corresponding increase in average texturedness from 21 to 33. Especially for difficult light situations, a large improvement can be achieved: for dark scenes (level 2) from 0.16 to 0.32 and for overexposed scenes (level 7) from 0.05 to 0.19. In Table I the results of the two examples - "castle" (C) and "motor bike" (B) - are reported (entities defined in Section III).

TABLE I. SIMULATION RESULTS AND TRACKING TARGET CHARACTERISTICS FOR 2 EXAMPLES.

| | R | FN | FT | FS | FD |
|---|---|---|---|---|---|
| **C Original** | 0.15 | 1520 | 24 | 2.6 | 1.6 |
| **C α=14, β=50%** | 0.44 | 1394 | 40 | 3.1 | 1.3 |
| **B Original** | 0.55 | 1431 | 33 | 1.9 | 1.9 |
| **B α=14, β=50%** | 0.67 | 1349 | 52 | 3.1 | 1.5 |

To show the potential for improving detection results by optimizing the tracking target object for concrete examples, we present two randomly selected tracking target objects, which have been modified with different techniques. Fig. 2 shows the target image "castle" and the modified images of the tracking target object.

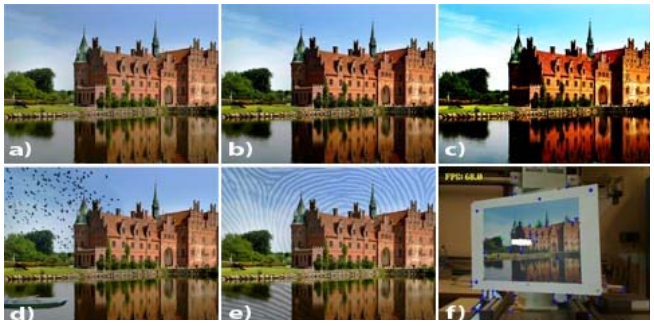


Figure 2. Fig. 1. a) Original image, b) contrast α=5, β=50% c) contrast α=14, β=50%, d) added content, e) added structure, f) captured target object.

Our feature distribution metric for the image indicated that features were predominantly clustered in specific subareas, which we alleviated by adding new content which fits into the image context (in this case we added a swarm of birds) to largely homogeneous texture regions (Fig. 2d). Increased contrast (Fig. 2b/c) and background structure which is unobtrusively embedded in the image (Fig. 2e) increase both, texturedness and feature distribution.We explored the effect of tracking target optimization with a real setup with a webcam and the printed targets. To ensure comparability of the tests, we used the robotic arm to capture images from predefined positions. In Table II, the detection rates for 4 illumination levels are presented for various modifications of the target image "castle" (Fig. 2). The controllable light source in this setup was a projector. An illumination level states a certain percentage of the maximum intensity of the projector: 100% equals white with max light emission and 0% equals to no light emission. The used illumination levels were: 30%: weak, 40%: moderate, 50%: overexposed and 60%: highly overexposed. The results show that each of the optimization techniques achieves an improvement. Only the strong contrast enhancement for highly illuminated images shows a decline; which is likely caused by over-modulation.

TABLE II. POSE RATIO R FOR TARGET "CASTLE".

| Illumination Intensity | 30% | 40% | 50% | 60% |
|---|---|---|---|---|
| **Original** | 0,58 | 0,86 | 0,81 | 0,53 |
| **Contrast α=5, β=50%** | 0,94 | 0,96 | 0,90 | 0,73 |
| **Contrast α=14,β=50%** | 0,94 | 0,98 | 0,79 | 0,40 |
| **Added content** | 0,91 | 0,92 | 0,90 | 0,63 |
| **Added structure** | 0,72 | 0,95 | 0,83 | 0,69 |

These results suggest various possibilities of tracking target improvement. Based on the flexibility a designer has with the motives on the tracking target, a slight contrast enhancement or subtly added content, or a combination may be preferable.

The optimization of the second target image "motor bike" (see Fig. 3) was also successful. For this target, we only used the contrast enhancement, since the feature distribution was already randomized. For instance the spatial feature distribution at the most used scale levels has an average of 0.977. This means that the features are distributed randomly over the entire target. Therefore no extra feature points have been added.

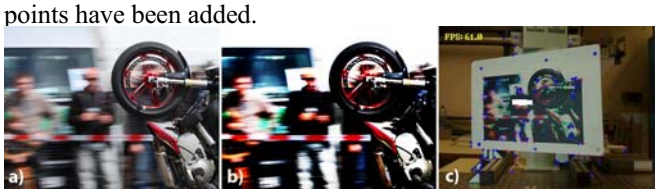


Figure 3. a) Original image, b) contrast α=14, β=50% c) captured target object.

An optimization for the lower light levels was not possible, since detection rate was already at 1.0, but the contrast enhancement allowed an improvement in light level 50 and 60 (see Table III).

TABLE III. POSE RATIO R FOR TARGET "BIKE"

| Illumination Intensity | 30% | 40% | 50% | 60% |
|---|---|---|---|---|
| **Original** | 1 | 1 | 0,97 | 0.82 |
| **Contrast α=14,β=50%** | 1 | 1 | 1 | 0,97 |

## VIII. CONCLUSION

This paper investigated how to characterize and optimize the appearance of tracking target objects and how the quality of pose detection can be predicted through simulation. Based on these results we demonstrated how to improve the pose estimation by slightly altering the tracking target object itself, which is a relatively new and unexplored approach.

In the future, we plan to extend our work by considering additional tracking algorithms, which should lead to even more generalizable guidelines for tracking target design. The possibility of analyzing 3D targets instead of only 2D images, which can be generally supported by our system, offers a wide space for further investigations.

## ACKNOWLEDGMENT

The authors want to thank Matthias Ruether and Martin Lenz (ICG - Graz) for their assistance with the Mitsubishi RV-1A robotic arm. This work has been funded by the Christian Doppler Laboratory (CDL), the Austrian Science Fund FWF under contracts Y193, W1209-N15 and the FIT-IT 820922.

## REFERENCES

[1] Knapek, M., Oropeza, R., Kriegman, D., Selecting promising landmarks, *In Proceedings of International Conference on Robotics and Automation*, pp. 3771-3777, 2000

[2] Moreels, P., Perona, P., Evaluation of features detectors and descriptors based on 3D objects, *International Conference on Computer Vision (ICCV)*, pp. 800-807, 2005

[3] Mikolajczyk, K. Schmid, C., A Performance Evaluation of Local Descriptors, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, pp. 1615-1630, 2005

[4] Shi, J. Tomasi, C., Good Features to Track, In Proceedings of Computer Vision and Pattern Recognition (CVPR'94), pp. 593-600, 1994

[5] Tsai, R.Y., Lenz, R.K., A new technique for fully autonomous and efficient 3D robotics hand-eye calibration, *IEEE Trans. on Robotics and Automation 5(3)*: 345–358, 1989.

[6] L. Gruber, S. Zollmann, D. Wagner, and D. Schmalstieg, Evaluating the trackability of natural feature-point sets, *Proceedings of the 2009 8th IEEE ISMAR*, 2009, pp. 189-190.

[7] Wagner, D., Reitmayr, G., Mulloni, A., Drummond, T., Schmalstieg, D., Pose Tracking from Natural Features on Mobile Phones, *Proceedings of the 7th IEEE/ACM ISMAR 2008*, pp. 125-134, 2008

[8] Cox, D. R., Lewis, P. A., Statistical Analysis of Series of Events, Chapman & Hall, 1966

[9] Farid H., Fundamentals of Image Processing, p. 44, http://www.cs.dartmouth.edu/farid/tutorials/fip.pdf, (last visit 25.01.2010)

[10] ImageMagick, http://www.imagemagick.org/script/index.php, (last visit 25.01.2010)