File ID      228902
Filename     6: Conclusions

SOURCE (OR PART OF THE FOLLOWING SOURCE):
Type        Dissertation
Title       Intrinsic statistical techniques for robust pose estimation
Author      G. Dubbelman
Faculty     Faculty of Science
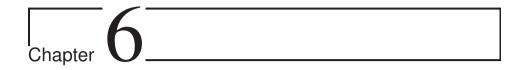Year        2011
Pages       157
ISBN        978-90-5335-427-8

FULL BIBLIOGRAPHIC DETAILS:
            http://dare.uva.nl/record/391318

# Chapter 6

# Conclusions

The focus of this Ph.D. thesis is on visual pose estimation. We identified the principles which limit the accuracy and the efficiency of state-of-art methods based on *random sample concensus* (RANSAC). By exploiting the manifold structure of pose spaces we developed alternative approaches which abolish RANSAC's deficiencies. The performance of our novel methods was evaluated extensively on challenging simulated data and on real image data. The latter includes image sequences recorded by a camera mounted on a vehicle which drove a total of 7.5 km. The obtained results showed that our novel methods are advantageous over current state-of-art. This scientific work was structured around 5 research questions stated in Chap. 1. Their answers are provided with detail in preceding chapters of this thesis. In this chapter we bring these answers together and formulate our conclusions. We first focus on our manifold related research of Chap. 2 and then on the specific algorithms of chapters 3, 4, and 5.

When the goal is to perform statistics in pose spaces, knowledge of the manifold structure of pose spaces is required. The pose of objects expressed in three dimensional Euclidean space is comprised of their position and orientation and generally has six degrees of freedom. Positions are often modeled by translation vectors, which themselves are elements of regular three dimensional Euclidean space. Certain circumstances, such as those of monocular vision, prevent the scale of translation to be estimated. In such cases one only has knowledge on the direction of translation, hence such scale-free translations can be normalized to unit length. The manifold structure of scale-free translations is therefore the unit sphere embedded in three dimensional Euclidean space. The manifold structure of rotations is most easily understood when reviewing their unit quaternion representation. Unit quaternions are four dimensional vectorial elements normalized to unit length. Their manifold structure is therefore similar to that of a unit sphere embedded in four dimensional Euclidean space. By combining translation with rotation one can express general Euclidean motions. In this thesis we showed that the combination of scale-free translation and rotation allows modeling the *epipolar geometry* related to monocular image pairs. This led us to the novel insight that the epipolar manifold is the combination of a sphere and a hypersphere. We furthermore showed that a previously proposed manifold structure for the epipolar manifold is not correctly founded by mathematical theory. As a

consequence this incorrect manifold structure does not facilitate maximum likelihood so-
lutions. To the contrary, our novel manifold structure does facilitate maximum likelihood
solutions. This was validated experimentally.

In order to develop statistical methods on pose manifolds we choose to follow the
methodology of intrinsic statistics. This approach performs statistical calculations in the
tangent spaces of a manifold. A tangent space is a local Euclidean approximation of the
manifold at one particular point and it preserves distances over the manifold with respect
to this particular point. The relation between the manifold and its tangent spaces is pro-
vided by its *exponential map* and its *logarithmic map* which both are differentiable. The
differentiability of these mappings allows expressing general statistical methods designed
for Euclidean spaces in manifolds. As all Euclidean spaces have similar structure, so do
all tangent spaces. Therefore, when one can extend a statistical algorithm in the tangent
space of one particular manifold, it can be extended to other manifolds as well. This pro-
cess only requires the availability of the distance preserving exponential and logarithmic
mappings belonging to the manifold. The benefit of intrinsic statistics is therefore that it
offers accuracy, stability as well as extensibility. In this thesis we extended several sta-
tistical methods to pose manifolds by using the methodology of intrinsic statistics. These
statistical methods are discussed next.

It was shown that pose hypotheses estimated on inlier image points can be modeled
by a Gaussian distribution in pose space. Furthermore, pose hypotheses estimated on
outlier image points can be modeled by a uniform distribution. A robust estimate for the
true camera pose can therefore be obtained by searching a Gaussian cluster in pose space.
To this purpose we developed a novel *intrinsic expectation maximization* algorithm. Our
approach has two benefits. Firstly, as it computes a mean on inlier hypotheses, it is more
accurate than returning a single hypothesis as RANSAC does. In fact, as the number of
inlier hypotheses increases, our method approaches the *maximum likelihood* lower bound
accuracy. Secondly, we can detect these inlier hypotheses directly in hypothesis space,
which is extremely efficient. Computationally intensive hypotheses verification as done
in RANSAC is therefore not required. The benefits of our approach were evaluated exper-
imentally within the context of *visual odometry*. Within this context it was shown to be
as accurate as the most accurate RANSAC approach and as efficient as the most efficient
RANSAC approach. On real data, comprising three urban trajectories of 600 m each, we
also showed that our EM approach is twice as accurate as a frequently applied RANSAC
approach.

In this thesis we also focussed on the fundamental issue of lower bound accuracy of
state-of-art RANSAC approaches. We showed that this lower bound is due to inefficient
hypotheses sampling and due to the relatively weak relation between the number of inliers
in the true accuracy of hypotheses. By exploiting the manifold structure of pose spaces
together with the number of inliers of hypotheses, we can approximately identify the loca-
tion of the ground truth pose. This location can then be targeted with artificially generated
hypotheses densely and efficiently. Whereas the number of inliers is not directly linked
to the true accuracy of hypotheses, a more accurate hypothesis has a higher probability
of receiving more inliers. Therefore, hypotheses near the location of the ground truth
will receive more inliers on average. Computing the mean of these hypotheses results in
a more accurate pose estimate than returning a single hypothesis. By using the method
of artificial sampling and the method of computing the mean on high ranking hypothe-
ses within an iterative algorithm, our approach improves on accuracy. Our experimental

evaluation showed that such a method is twice as accurate with respect to the maximum likelihood lower bound than state-of-art RANSAC approaches. We furthermore showed this gain in accuracy is approximately two times larger than the gain obtained by applying bundle adjustment to all inliers found by a state-of-art RANSAC method. This illustrates the significance of our conceptual improvements.

The similarity of RANSAC approaches and our novel methods is that they all provide relative pose estimates. The relative estimates are concatenated to track the absolute pose of the camera. As these relative pose estimates always contain some amount of error, be it structural (bias) or random, the absolute pose accumulates error and the estimated camera's trajectory will differ from its true trajectory. A more accurate relative pose estimator simply pushes forward the point at which the accumulated error becomes so significant that performance is no longer satisfactory. A structural approach to prevent error accumulation is to integrate absolute pose information. Absolute pose information can be derived from loop-detection or can be provided by absolute pose sensors. This absolute pose information will itself also exhibit uncertainty, it will however enforce an upper limit on the amount of accumulated error in the trajectory. In this thesis we presented a novel algorithm which can be used to incorporate absolute pose information into trajectories obtained from relative pose estimates. It minimally bends trajectories such that their final absolute pose is equal to the desired absolute pose. Our algorithm has linear complexity in the number of poses and operates in closed-form. It is therefore highly efficient. Extensive experiments show that its accuracy on realistic simulated data and on 800 m long real data is similar to an iterative maximum likelihood approach which is significantly less efficient. The proposed algorithm is used to perform *loop-closure* as well as *sensor-fusion*. The additional sensor used is an *absolute heading reference system*. This sensor provides accurate absolute orientation information without relying on off-board systems (like GPS does). By incorporating this data into the trajectory, the absolute orientation as well as the absolute position of every pose in the trajectory is improved. We are, in a sense, able to "close-the-loop" on orientational information only. This is demonstrated on a 5 km long urban trajectory which is one of the most challenging data sets used within state-of-art pose estimation research.

In this Ph.D. dissertation we extended the state-of-art in visual pose estimation with respect to efficiency and accuracy. We did so by exploiting the manifold structure of pose spaces. This also allowed us to provide solid theoretical foundations to our methods. The combination of our novel intrinsic pose estimator and our novel closed-form trajectory bending algorithm is shown to deliver accurate pose estimates under realistically challenging conditions. Given the efficiency of our methods, real-time implementations can be developed with relative ease. This research has therefore brought visual pose estimation closer to actual applications, especially those related to *intelligent mobile systems*.