Robust Pose Estimation from a Planar Target

Gerald Schweighofer and Axel Pinz

Abstract

In theory, the pose of a calibrated camera can be uniquely determined from a minimum of four coplanar and not collinear points. In practice, there are many applications of camera pose tracking from planar targets, and there is also a number of recent pose estimation algorithms which perform this task in real-time, but all of these algorithms suffer from pose ambiguities. This paper investigates the pose ambiguity for planar targets viewed by a perspective camera. We proof that pose ambiguities - two distinct local minima of the according error function - exist even for cases with wide angle lenses and close range targets. We give a comprehensive interpretation of the two minima, and develop a new algorithm for a unique and robust solution to pose estimation from a planar target.

Index Terms

camera pose ambiguity, pose tracking

I. Introduction

THERE are many applications of pose estimation, where 6 degrees of freedom of a camera's pose have to be calculated from known correspondences with known scene structure. This can be done from a single image or from an image sequence. In photogrammetry this problem is known as space resectioning and it is often solved offline by bundle adjustment techniques, achieving very high precision and at the same time high robustness against outliers. In general, there are several ways to solve this problem, as long as many points can be used, and when the pose can be calculated offline (e.g. taking information from frames $n + k$ into account to calculate the pose for frame n). Online pose tracking requires to calculate the camera pose for each frame, in real-time. Often interest points are extracted from a target, and the camera pose is calculated relative to the target's pose in the scene. In theory, pose can be calculated from four or more coplanar and not collinear points, if the intrinsic parameters of the camera are known, while this remains a critical configuration for an uncalibrated camera, even if many coplanar points are available [1]. In the Computer Vision literature, several approaches to pose estimation are known. Most of them work for coplanar points [2], [3], some have been extended to use points and lines [4], [5], and some work also for arbitrary 3D target point configurations [6]. Recent success has also been reported for online structure *and* motion estimation [7], where many interest points are extracted, frame-to-frame correspondence is rather easy, and no a priori reference to a scene coordinate system is required (for early work in this direction see [8]).

Examples of vision-based tracking systems which are based on planar targets include ARToolkit ([9], widely used for many augmented reality (AR) applications), Malik et al. ([10], they claim to be more precise than ARToolkit due to an improved target design), and our own developments ([11], fusing inertial and vision sensors to increase robustness). Kawano et al. [12] discuss a number of further planar markers for AR, and present their own, coded planar target. Users of such systems observe that vision-based pose is not very precise, which results in significant jitter, and not very robust suffering from pose jumps and gross pose outliers.

We have performed a number of experiments comparing several pose algorithms and planar target configurations. Even if we use a well calibrated camera and avoid known critical configurations, we observe pose jumps, which should not occur.

G. Schweighofer is with the Inst. of El. Measurement and Measurement Signal Processing E-Mail: gschweig@emt.tugraz.at
A. Pinz is with the Inst. of El. Measurement and Measurement Signal Processing E-Mail: axel.pinz@tugraz.at

These pose ambiguities have also been discussed by Oberkampf et al. [13]. They give a straightforward interpretation for the case of orthographic projection and they develop their POSIT algorithm, which uses *scaled orthographic projection* at each iteration step. POSIT maintains two alternative solutions and finally decides for the better one based on a distance measure E, which is calculated from reprojection errors in the image space. However, this approach is only valid when the camera distance is very large compared to depth in the scene, so that perspective projection can be approximated by scaled orthographic projection.

This paper tackles the general case of perspective projection. We proof that pose ambiguities exist also for cases with wide angle lenses and close range targets. We give a comprehensive interpretation of these ambiguities, and develop a new algorithm for a unique and robust solution to pose estimation from a planar target.

II. Pose Ambiguity

We discuss here the camera pose estimation problem for a calibrated camera (with known interior parameters) as the problem of finding the six exterior parameters of the camera: orientation $R = f(\alpha, \beta, \gamma)$ and position $\mathbf{t} = [t_x, t_y, t_z]^T$ between camera and scene coordinate system. Figure 1 shows on the left side the center of projection C_C (camera center) and the image plane. On the right we see the model points P_i, P_j in the plane Π. C_C is also the origin of the camera coordinate system, and the origin of the model coordinate system is the center of the model C_M (model center)[1]. The distance between C_C and C_M is labeled with $\mathbf{t}$. By rotating the model around the y-axis for an angle α we obtain the model points $P_{i\alpha}$ and $P_{j\alpha}$ located in the plane Π_α. These points are projected to the image as $\mathbf{v}_i$ and $\mathbf{v}_j$.

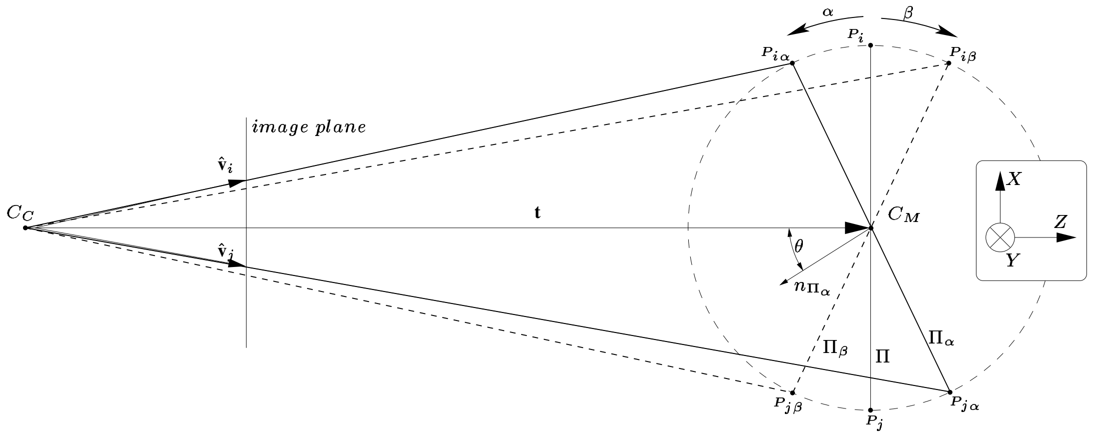

Fig. 1. Pose Ambiguities with perspective projection.

We assume n model points $\mathbf{p}_i$ in scene coordinates which are transformed to camera coordinates $\mathbf{v}_i$ by

$$\mathbf{v}_i \propto R\mathbf{p}_i + \mathbf{t}, \tag{1}$$

where $\propto$ denotes 'directly proportional', so $\mathbf{v}_i$ are measured up to an unknown scale factor. We further on refer to $\mathbf{v}_i$ as points in camera coordinates and to $\hat{\mathbf{v}}_i$ as their measurements in the image (also in camera coordinates, but imprecise due to noise). A *pose estimation* algorithm has to find values for $\hat{R}$ and $\hat{t}$ that minimize an error function. In this paper we use the *object-space* error function [2]

$$E_{os}(\hat{R}, \hat{\mathbf{t}}) = \sum_{i=1}^{n} \left\| \left(I - \frac{\hat{\mathbf{v}}_i \hat{\mathbf{v}}_i^t}{\hat{\mathbf{v}}_i^t \hat{\mathbf{v}}_i} \right) (\hat{R}\mathbf{p}_i + \hat{\mathbf{t}}) \right\|^2 . \tag{2}$$

For our example in Fig. 1 we want to point out that for every chosen α there may exist a second, different angle β which also leads to a local minimum of the error function E_{os}. The existence of such a minimum

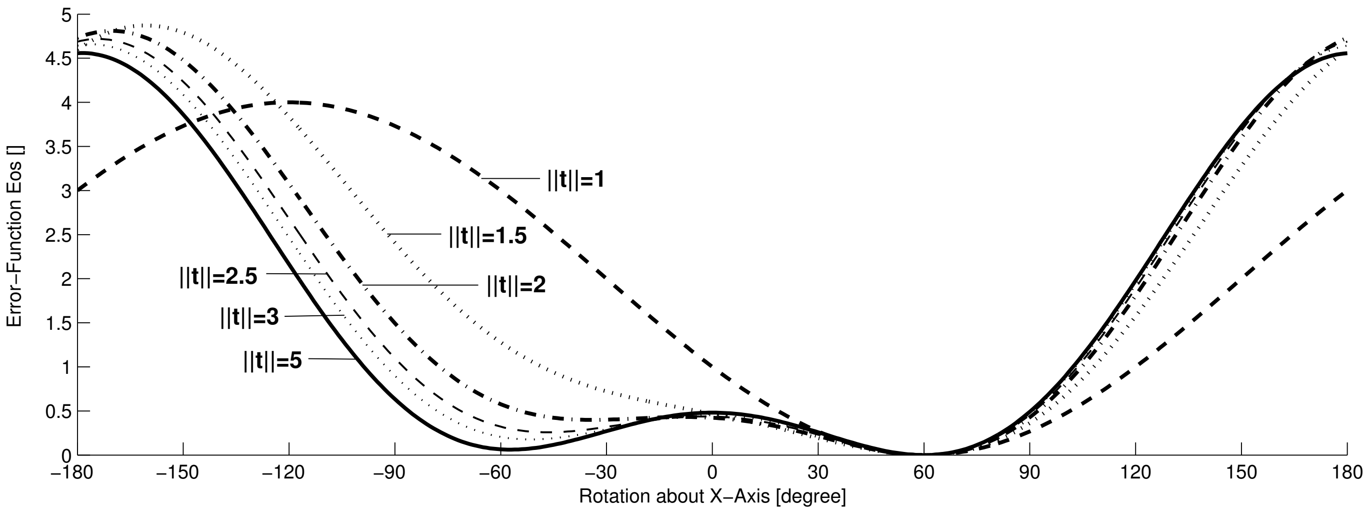

Fig. 2. Error value E_{os} for different distances $\|\mathbf{t}\|$.

depends on all parameters of (2). Figure 2 shows how the error function E_{os} changes depending on different distances $\mathbf{t}$ between model[2] center C_M and camera center C_C. In this case R is just a rotation about the y-axis, as shown in Fig. 1, with $\alpha = 60°$ and zero noise ($\hat{\mathbf{v}}_i = \mathbf{v}_i$). For all visualized curves there is one absolute minimum at $\alpha = 60°$ where the error function E_{os} is zero. For $\|\mathbf{t}\| = 2, 2.5, 3, 5$ we see local minima of increasing significance (decreasing E_{os}) at $\beta = -38.8°, -49.9°, -53.7°$ and $-58.0°$. With increasing distance $\|\mathbf{t}\|$ between model and camera the effect of perspective projection decreases. In the case $\|\mathbf{t}\| = \infty$ we would have *orthographic projection* with a second minimum of $E_{os} = 0$ at $\beta = -60°$. For $\|\mathbf{t}\| < 1.86$ there exists only one minimum of the error function ($E_{os} = 0$ at $\alpha = 60°$).

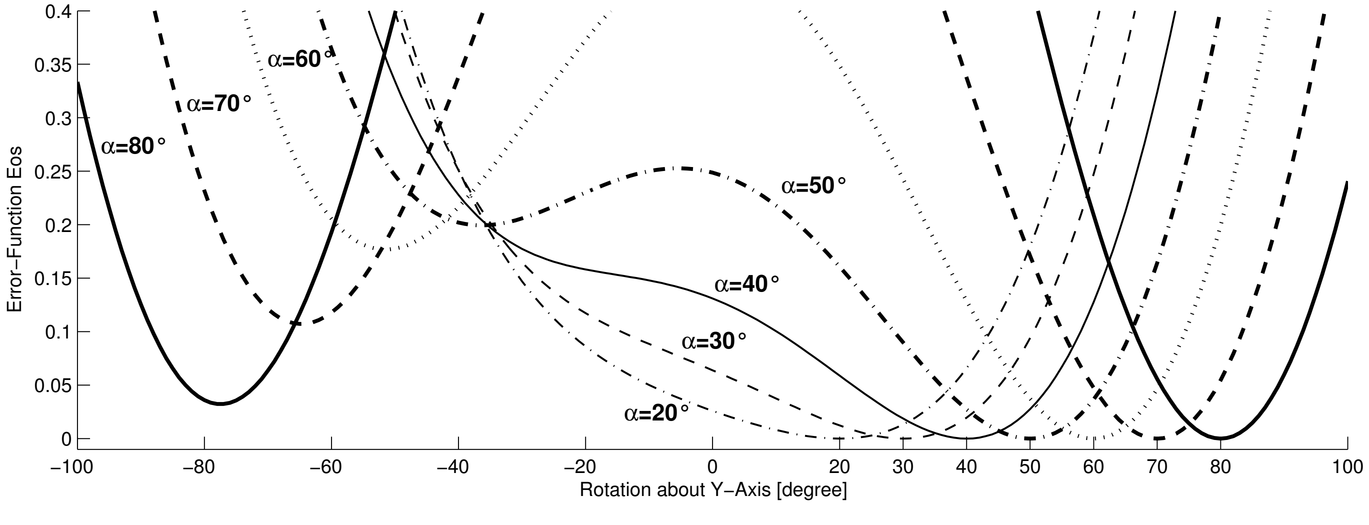

Fig. 3. Error value E_{os} for different rotation angles α.

Figure 3 depicts the dependency of E_{os} on varying rotation angles α. For this plot $\|\mathbf{t}\| = 3$. Again, we assume a noise-free case and obtain $E_{os} = 0$ for each α. For $\alpha = 50°, 60°, 70°$ and $80°$ we see a second local minimum of E_{os} at $\beta = -26.0°, -36.4°, -51.5°, -64.8°$ and $-77.5°$. For $\alpha < 43, 4°$ there exists no second local minimum of E_{os}.

[1] For the purpose of this paper, we can assume that scene and model coordinate system coincide without loss of generality

[2] The chosen model consists of four points: $p1 = [1\ 1\ 0]^T$, $p2 = [1\ -1\ 0]^T$, $p3 = [-1\ -1\ 0]^T$, $p4 = [-1\ 1\ 0]^T$

III. Robust Pose Estimation Algorithm

The previous section demonstrated, that there can exist two distinct local minima of E_{os}, even for the noise-free case. From now on we discuss the case of real images, where in general $\mathbf{v}_i \neq \hat{\mathbf{v}}_i$. One can expect more occurrences of two local minima than for the noise-free situation, but E_{os} will be above zero for both cases, the correct angle α, and the wrong one β. Our new algorithm is based on these major observations:

1) There are sometimes two local minima, depending on the actual configuration ($\hat{R}$, $\hat{\mathbf{t}}$, $\mathbf{p}_i$ and $\hat{\mathbf{v}}_i$).
2) The correct solution to the pose estimation problem should have the lower error value E_{os}.

To develop the algorithm we first assume a known pose ($\hat{R}_1$, $\hat{\mathbf{t}}_1$) which we get from any pose estimation algorithm. As we know from section II there could be sometimes two local minima. So it may happen that this pose estimation algorithm returns a wrong result. Experimental results in section IV show that this happens in nearly 50% of all cases for several common pose estimation algorithms.

We now show how to use a first guess of a pose $(\hat{R}_1, \hat{\mathbf{t}}_1)$ to estimate a second pose, which also minimizes the error function E_{os}.

This is done by successively changing the general transformation (1) in a way that finally E_{os} only depends on a rotation about the y-axis $R_y(\beta)$ and $\mathbf{t}$. Subsequently we switch between presentations of general equations ((3), (5) and (9)) and specific calculations (R_x and R_z) based on the initial first pose $(\hat{R}_1, \hat{\mathbf{t}}_1)$.

Assume model points $\mathbf{p}_i$ which are measured in the image as $\hat{\mathbf{v}}_i$ (see (1)) such that

$$\hat{\mathbf{v}}_i \approx \mathbf{v}_i \propto R\mathbf{p}_i + \mathbf{t}. \tag{3}$$

Without loss of generality[3] we can multiply both sides of (3) with R_t to get a transformed system

$$R_t\hat{\mathbf{v}}_i \approx R_t\mathbf{v}_i \propto R_tR\mathbf{p}_i + R_t\mathbf{t}, \tag{4}$$

such that $R_t\hat{\mathbf{t}}_1 = [0 \ 0 \ \|\hat{\mathbf{t}}_1\|]^T$. The outcome of this transformation of the coordinate system is, that the center of the model ($[0 \ 0 \ 0]^T$) will be projected into the image as $[0 \ 0 \ \|\hat{\mathbf{t}}_1\|]$. Let

$$\tilde{\mathbf{v}}_i = R_t\hat{\mathbf{v}}_i \quad \tilde{\mathbf{t}}_1 = R_t\hat{\mathbf{t}}_1 \quad \tilde{R}_1 = R_t\hat{R}_1. \tag{5}$$

The pose ($\tilde{R}_1$, $\tilde{\mathbf{t}}_1$) minimizes

$$E_{os}(\tilde{R}, \tilde{\mathbf{t}}) = \sum_{i=1}^{n} \left\| \left(I - \frac{\tilde{\mathbf{v}}_i\tilde{\mathbf{v}}_i^t}{\tilde{\mathbf{v}}_i^t\tilde{\mathbf{v}}_i} \right) (\tilde{R}\mathbf{p}_i + \tilde{\mathbf{t}}) \right\|^2 . \tag{6}$$

Without loss of generality we introduce a rotation matrix $\tilde{R}_z$, such that

$$E_{os}(\tilde{R}, \tilde{\mathbf{t}}) = \sum_{i=1}^{n} \left\| \left(I - \frac{\tilde{\mathbf{v}}_i\tilde{\mathbf{v}}_i^t}{\tilde{\mathbf{v}}_i^t\tilde{\mathbf{v}}_i} \right) (\tilde{R}\underbrace{\tilde{R}_z\tilde{R}_z^{-1}}_{I}\mathbf{p}_i + \tilde{\mathbf{t}}) \right\|^2 . \tag{7}$$

The rotation $\tilde{R}_z^{-1}$ rotates the planar model $\mathbf{p}_i$ only about the z-axis such that the rotated model $\tilde{\mathbf{p}}_i = \tilde{R}_z^{-1}\mathbf{p}_i$ is also planar with $z = 0$. The rotation matrix $\tilde{R}_1\tilde{R}_z$ can be decomposed into the product of three rotations

$$\tilde{R}_1\tilde{R}_z = R_z(\tilde{\gamma}_1)R_y(\tilde{\beta}_1)R_x(\tilde{\alpha}_1), \tag{8}$$

where $R_i(\phi)$ describes a rotation of ϕ degrees about axis i. By selecting $\tilde{R}_z$ such that $\tilde{\alpha}_1 = 0$ we obtain another transformed system

$$\tilde{\mathbf{v}}_i \approx R_z(\gamma)R_y(\beta)\tilde{\mathbf{p}}_i + \tilde{\mathbf{t}} \tag{9}$$

[3] Note that this transformation does not affect the shape of the error function. Thus, a minimum of the transformed system will also be a minimum of the original system.

with the corresponding error function

$$E_{os}(\gamma,\beta,\tilde{\mathbf{t}}) = \sum_{i=1}^{n}\left\|\left(I - \frac{\tilde{\mathbf{v}}_i\tilde{\mathbf{v}}_i^t}{\tilde{\mathbf{v}}_i^t\tilde{\mathbf{v}}_i}\right)\left(R_z(\gamma)R_y(\beta)\tilde{\mathbf{p}}_i + \tilde{\mathbf{t}}\right)\right\|^2 . \tag{10}$$

Equation (9) and (10) are equal to (4) and (6) in the case of using the known pose $(\tilde{R}_1, \tilde{\mathbf{t}}_1)$ or $\left(f(0,\tilde{\beta}_1,\tilde{\gamma}_1),\tilde{\mathbf{t}}_1\right)$.

From section II we know that there may be two minima of the error function E_{os} w.r.t. a rotation about one axis (y-axis) depending on the parameters. In our transformed system (9) we have a rotation about this axis (y-axis) and a rotation about the z-axis.

Let us discuss the effect of the Rotation $R_z(\gamma)$. We know from our normalization step (5) that $\tilde{\mathbf{t}}_1 = [0\ 0\ \|\hat{\mathbf{t}}_1\|]^T$. In this case we can rewrite (9) as

$$\tilde{\mathbf{v}}_i \approx R_z(\gamma)(R_y(\beta)\tilde{\mathbf{p}}_i + \tilde{\mathbf{t}}_1), \tag{11}$$

because $R_z(\gamma)\tilde{\mathbf{t}}_1 = \tilde{\mathbf{t}}_1$. Thus the rotation $R_z(\gamma)$ becomes a rotation just around the optical axis of the camera. This rotation leaves the geometric relation between image plane and model plane invariant and will just affect image coordinates. Thus, all our observations related to Fig. 1 are valid and we can just search for local minima of E_{os} with respect to β.

We thus propose the following "Robust Pose Estimation Algorithm for Planar Targets":

1) Estimate a first pose $\hat{P}_1 = (\hat{R}_1, \hat{\mathbf{t}}_1)$ by applying any existing iterative pose estimation algorithm. In our experiments we used the iterative algorithm proposed by [2]. $\hat{P}_1$ is one local minimum of E_{os}. Our goal is to analytically derive an estimate of the second local minimum, if such a minimum exists.
2) Transform the coordinate system according to (5) to get $\tilde{P}_1 = (\tilde{R}_1, \tilde{\mathbf{t}}_1)$.
3) Estimate $\tilde{R}_z$ as described in (7) to (9) to obtain the transformed system and the parameters of the first pose ($\tilde{\gamma}_1$ and $\tilde{\beta}_1$).
4) Fix $\gamma = \tilde{\gamma}_1$, and estimate all local minima of (10) for the parameters β and $\tilde{\mathbf{t}}$. For details see (III-A).
5) Undo the transformations of step 1 and 2 for all local minima to obtain poses $\hat{P}_i$.
6) Use all poses $\hat{P}_i$ as a start value for the iterative pose estimation algorithm [2] to get final poses $P_i^{\star}$.
7) Decide the final and correct pose, which has the lowest error E_{os}.

A. Estimation of the local minima

We discuss here how to estimate the local minima of (10) for given $\tilde{\mathbf{p}}_i$,$\tilde{\mathbf{v}}_i$ and $R_z(\gamma)$. To simplify the equations we set

$$\tilde{V}_i = \frac{\tilde{\mathbf{v}}_i\tilde{\mathbf{v}}_i^t}{\tilde{\mathbf{v}}_i^t\tilde{\mathbf{v}}_i}$$

We can solve in (10) for the optimal translation $\tilde{\mathbf{t}}_{opt}$ w.r.t. to the minimization of E_{os} by derivating (10) such that [2]

$$\begin{aligned}\frac{\partial E_{os}}{\partial \tilde{\mathbf{t}}} &= 0 \Rightarrow \qquad (12)\\ \tilde{\mathbf{t}}_{opt}(\beta) &= \frac{1}{n}\left(I - \frac{1}{n}\sum_j \tilde{V}_j\right)^{-1} \sum_j \left(\tilde{V}_j - I\right) R_z(\gamma)R_y(\beta)\tilde{\mathbf{p}}_j = G\sum_j \left(\tilde{V}_j - I\right) R_z(\gamma)R_y(\beta)\tilde{\mathbf{p}}_j,\end{aligned}$$

where G is a constant and depends only on measured image positions $\tilde{\mathbf{v}}_i$.

By plugging $\tilde{\mathbf{t}}_{opt}(R)$ into (10) we get an error function which only depends on β (the rotation about the y-axis):

$$E_{os}(\beta) = \sum_{i=1}^{n}\|(I - \tilde{V}_i)(R_z(\gamma)R_y(\beta)\tilde{\mathbf{p}}_i + G\sum_{j=1}^{n}(\tilde{V}_j - I)R_z(\gamma)R_y(\beta)\tilde{\mathbf{p}}_j)\|^2 \tag{13}$$

Simplifying $R_y(\beta)$, by substituting ($\beta_t = \tan\frac{1}{2\beta}$), we obtain

$$R_y(\beta) = \begin{bmatrix} \cos(\beta) & 0 & \sin(\beta) \\ 0 & 1 & 0 \\ -\sin(\beta) & 0 & \cos(\beta) \end{bmatrix}; R_y(\beta_t) = \frac{1}{1+{\beta_t}^2} \begin{bmatrix} 1-{\beta_t}^2 & 0 & 2\,\beta_t \\ 0 & 1+{\beta_t}^2 & 0 \\ -2\,\beta_t & 0 & 1-{\beta_t}^2 \end{bmatrix} \tag{14}$$

By plugging this into (13) we get a function $E_{os}(\beta_t)$ which now only depends on β_t. To get all extremal points of this function we need to solve

$$\frac{\partial E_{os}(\beta_t)}{\partial \beta_t} = 0,$$

which is a polynomial of degree four and can be easily solved. In general, we will obtain two maxima and *two minima*. The two minima are selected as those extrema with $\frac{\partial^2 E_{os}(\beta_t)}{\partial^2 \beta_t} > 0$.

IV. Experimental Results

To test the proposed algorithm we performed a huge amount of experiments. The setup was as follows:

1) For each test we generated a random model consisting of 10 points $p_i = [x_i\ y_i\ z_i]^t$ in the range $x_i = [-1\ 1]$, $y_i = [-1\ 1]$ and $z_i = 0$ to be planar.
2) For each test we generated a random rotation R of the pose.
3) The translation vector $\mathbf{t}$ of the pose was chosen such that the measured image points $\mathbf{v}_i$ were located in the image at a random position and with a maximum size of the Feret box of $200 Pixels^4$.
4) To each image point $\mathbf{v}_i$ Gaussian noise was added to get $\hat{\mathbf{v}}_i$.

We repeated this procedure for 13 different noise levels reaching from zero to 6 pixels. For each noise level we generated 1000 different models and poses. In Fig. 4 we see the result of the experiment.

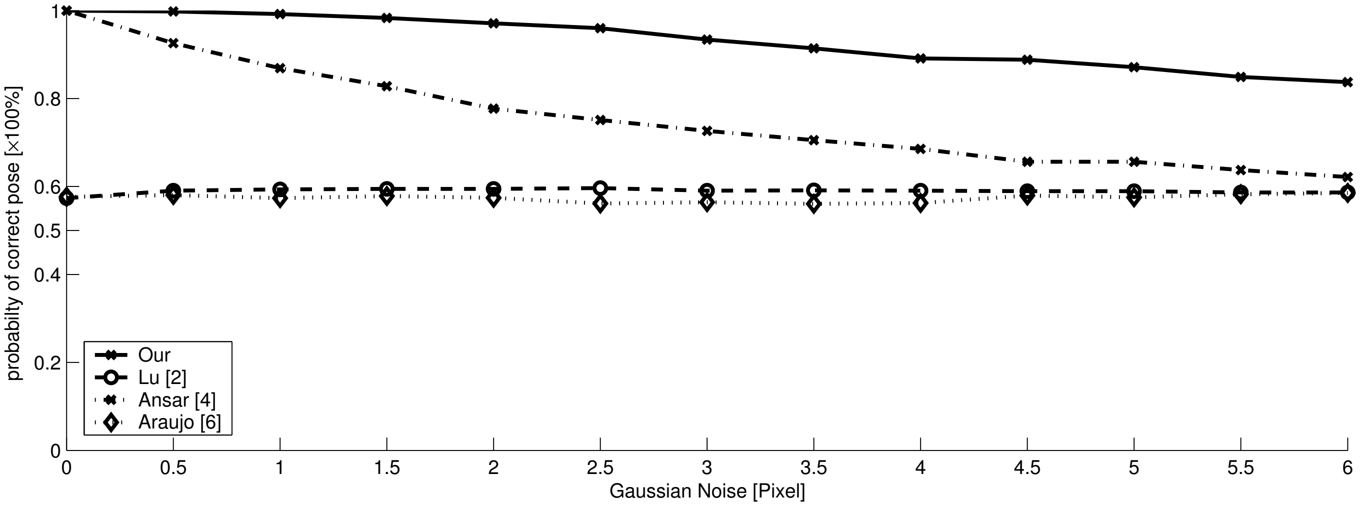

Fig. 4. Probability of choosing the correct Pose (Solid line: Our algorithm).

The solid line in Fig. 4 shows the result of our algorithm, which has in the noise-free case (Gaussian noise = 0 $Pixels$) a probability of 100 % to find the correct pose. With increasing noise the probability decreases down to 83.7 % for 6 $Pixels$ Gaussian noise. Results for existing common iterative pose estimation algorithms ([2], [6]) are also given. Because there are most of the time two local minima of the error function, the probability of finding the correct solution with these algorithms is just above 50 %. The algorithm of Ansar and Daniilidis [4] starts at 100 % for zero noise, because this algorithm calculates a direct solution, but its performance also decreases significantly with increasing noise level.

[4] Interior parameters: $f_x = f_y = 800$, $x_0 = 320$ and $y_0 = 240$ Image Size: 640 × 480

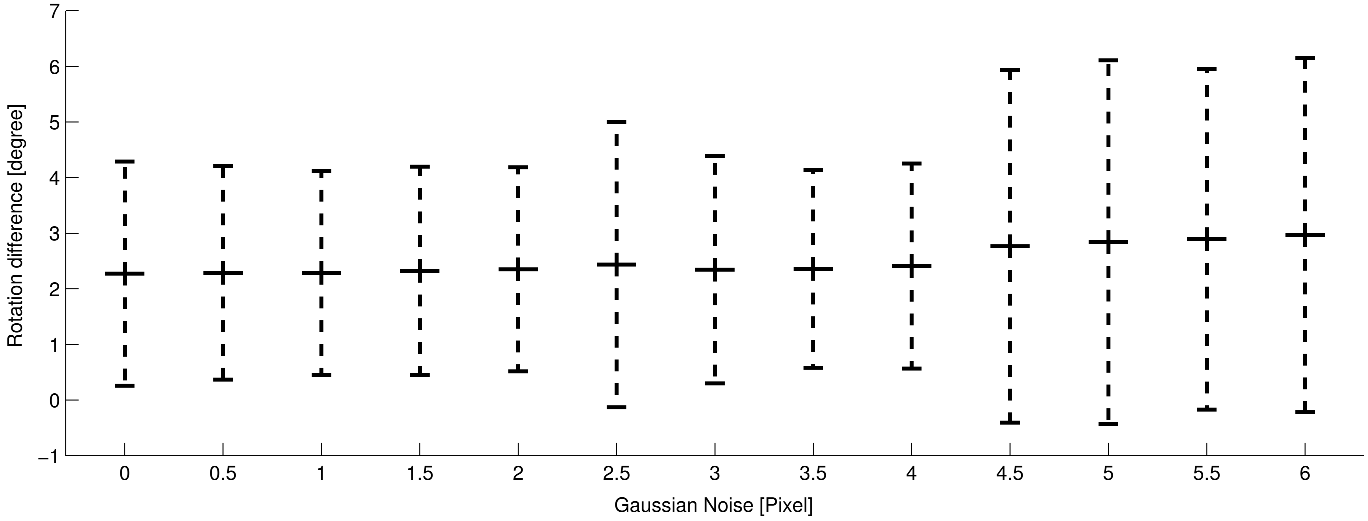

Fig. 5. Distance to the true minima of the second pose.

In step 6 of our proposed algorithm we used all poses $\hat{P}_i$ as start values for the iterative algorithm proposed by [2]. In Fig. 5 we show the distance between the rotations of the pose $\hat{P}_i$ and the refined pose $P_i^\star$. The plot shows for all different noise levels the mean and the standard deviation. We see that the mean error is always below three degrees and increasing with increasing noise level.

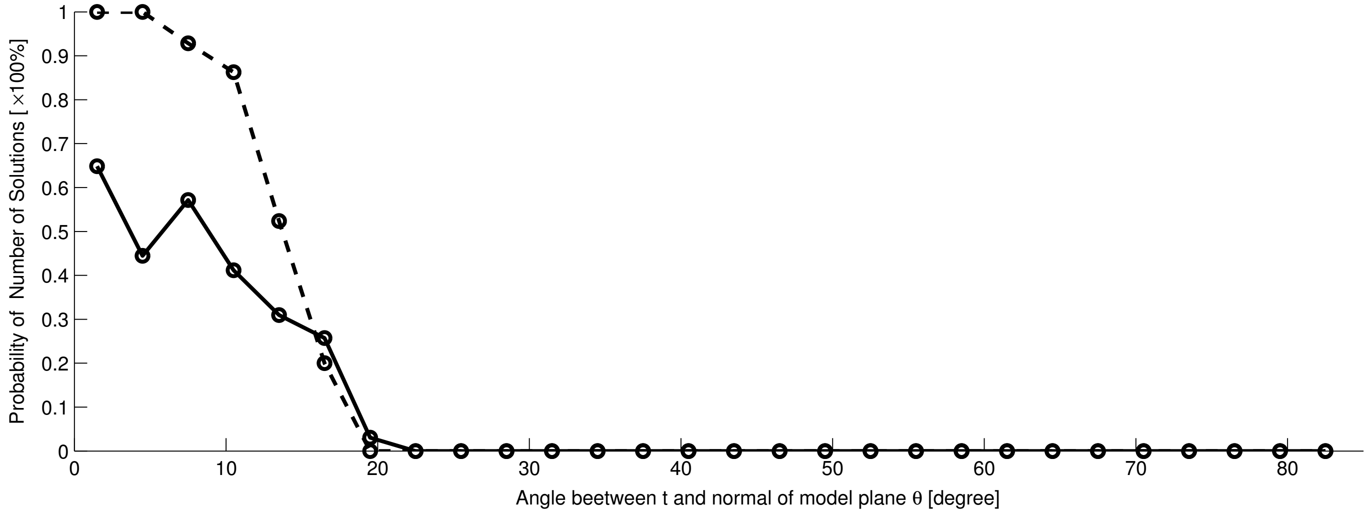

Fig. 6. Probability of having one solution vs. angle θ (Dashed line: noise-free case; Solid line: $noise = 2\ Pixels$)

From section II we know that the number of the minima of the error function depends on the rotation of the model plane. To visualize this, Fig. 6 plots the probability of having one minimum of the error function E_{os} against the angle between the normal of the model plane and **t** (for a better understanding this angle is visualized in Fig. 1 as θ). The solid line represents the probability of having only one minimum for the noise-free case. The dashed line shows the probability of having only one minimum for the case of 2 pixels Gaussian noise. If there is not one minimum there are two minima. We see, that if the image plane is nearly parallel to the object plane (angle lower than 15 degrees) there is only one solution for the pose. If the angle θ is above 20 degrees there are almost always two solutions.

Figure 7 compares the probabilities of choosing the correct solution for our algorithm (solid line) with other common pose estimation algorithms ([2], [4], [6]) plotted against the angle θ. Here we used only tests where the Gaussian noise was 2 $Pixels$. As we know already from Fig. 6 there is only one solution

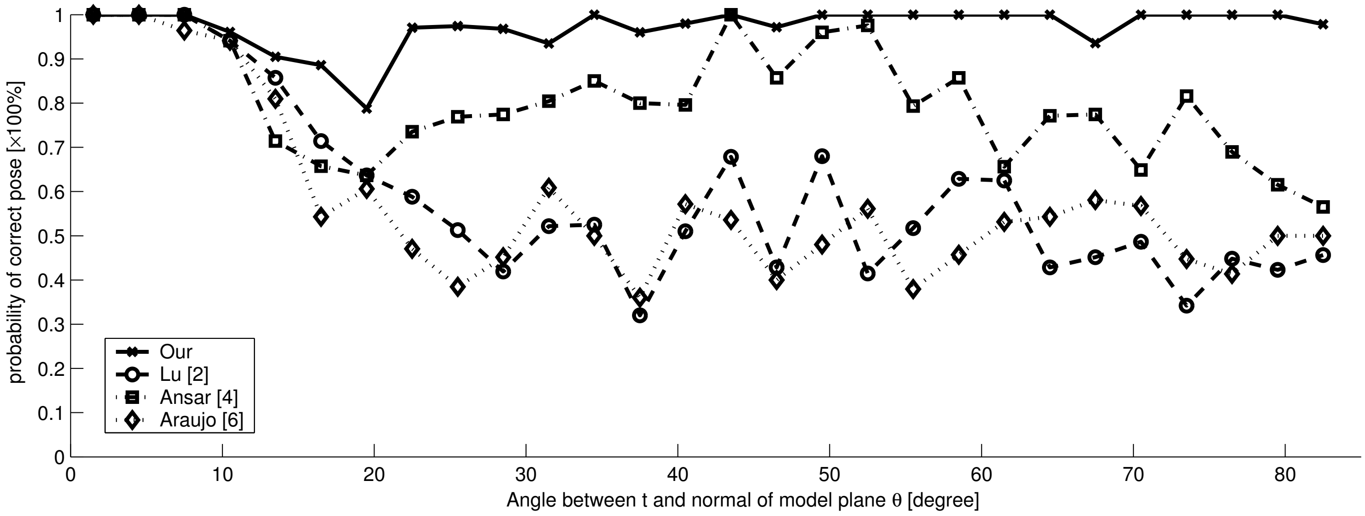

Fig. 7. Probability of correct Pose vs. angle θ for $noise = 2\ Pixels$ (Solid line: our algorithm)

in the case of small θ, so the probability of choosing the correct solution is one for all algorithms. In cases where two solutions occur, the performance of other algorithms drops down to probabilities as low as 50 % to find the correct solution. Our algorithm has a probability of above 95 % to estimate the correct solution for all angles θ except in the range from 10 to 20 degrees. This effect is caused by noise, that influences the results in this range of transition from one to two minima (see overlapping area of solid and dashed lines in Fig. 6).

V. Conclusion

Except for the work by Oberkampf et al. [13], ambiguities in pose estimation have been neglected in the literature. We have presented a thorough analysis and could proof, that in general there exist two local minima of the according error function. These two minima are the reason why pose jumps are observed in many pose tracking applications. Based on these findings, we presented a new “Robust Pose Estimation Algorithm for Planar Targets” which takes the two minima into account to give a robust pose. To our knowledge and experimental evidence this is the first algorithm which is robust against pose jumps in real applications using planar targets.

Acknowledgments

This work was supported by the Austrian Science Foundation (FWF, project S9103-N04), (FWF, project P15748) and VAMPIRE Visual Active Memory Processing and Interactive REtrieval (EU-IST Programme IST-2001-34401).

REFERENCES

[1] B. Wrobel, “Minimum solutions for orientation,” in *Calibration and Orientation of Cameras in Computer Vision*, Gruen and Huang, Eds. Springer-Verlag, 2001, ch. 2.

[2] C. Lu, G. Hager, and E. Mjolsness, “Fast and globally convergent pose estimation from video images,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 6, pp. 610–622, June 2000.

[3] M. Fischler and R. Bolles, “The random sample consensus set: a paradigm for model fitting with applications to image analysis and automated cartography,” *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[4] A. Ansar and K. Daniilidis, “Linear pose estimation from points or lines,” in *European Conference on Computer Vision*, A. H. et al., Ed., vol. 4. Copenhagen, Denmark: Springer, May 2002, pp. 282–296.

[5] P. Wunsch and G. Hirzinger, “Registration of cad-models to images by iterative inverse perspective matching,” *Pattern Recognition. Proceedings of the 13th International Conference on*, vol. 1, pp. 78–83, August 1996.

[6] H. Araújo, R. Carceroni, , and C. Brown, “A fully projective formulation to improve the accuracy of lowe’s pose-estimation algorithm,” *Computer Vision and Image Understanding*, vol. 71, no. 2, pp. 227–238, 1998.

[7] D. Nister, O. Naroditsky, and J. Bergen, “Visual odometry,” *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 652–659, 2004.

[8] C. Harris, “Geometry from visual motion,” in *Active Vision*, A. Blake and A. Yuille, Eds. MIT Press, 1992, ch. 16.

[9] H. Kato and M. Billinghurst, “Marker tracking and hmd calibration for a video-based augmented reality conferencing system,” *(IWAR 99) Proceedings. 2nd IEEE and ACM International Workshop on Augmented Reality*, pp. 85–94, 1999.

[10] S. Malik, G. Roth, and C. McDonald, “Robust 2d tracking for real-time augmented reality,” *Proceedings. Vision Interface 2002*, 2002.

[11] M. K. Chandraker, C. Stock, and A. Pinz, “Real-time camera pose in a room,” *3rd Int Conf. Comp. Vision Systems ICVS*, pp. 98–110, 2003.

[12] T. Kawano, Y. Ban, and K. Uehara, “A coded visual marker for video tracking system based on structured image analysis,” *(ISMAR 03) Proceedings. 2nd IEEE and ACM International Symposium on Mixed and Augmented Reality*, 2003.

[13] D. Oberkampf, D. F. DeMenthon, and L. S. Davis, “Iterative pose estimation using coplanar feature points,” *Computer Vision and Image Understanding*, vol. 63, no. 3, pp. 495–511, May 1996.