Leveraging Feature Uncertainty in the PnP Problem

Luis Ferraz[1]
luis.ferraz@upf.edu
Xavier Binefa[1]
xavier.binefa@upf.edu
Francesc Moreno-Noguer[2]
fmoreno@iri.upc.edu

[1] Department of Information and Communication Technologies
Universitat Pompeu Fabra
08018, Barcelona, Spain

[2] Institut de Robòtica i Informàtica Industrial (CSIC-UPC)
08028, Barcelona, Spain

Introduction: The goal of the Perspective-n-Point (PnP) problem is to estimate the position and orientation of a calibrated camera from a set of n 3D-to-2D point matches. State-of-the-art PnP solutions assume that these correspondences may be corrupted by noise and show robustness against large amounts of it. Yet, none of these works considers that the particular structure of the uncertainty associated to each correspondence could indeed to be used to further improve the accuracy of the estimated pose. Specifically, existing solutions, as [3, 4], assume all 2D correspondences to be affected by the same model of noise, a zero mean Gaussian distribution, and consider all correspondences to equally contribute to the estimated pose, independently of the precision of their actual location.

Contributions: In this paper we propose a real-time and accurate PnP solution that exploits the fact that in practice the 2D position of not all 2D features is estimated with the same accuracy (see Fig.1(a,b)). Assuming a model of such feature uncertainties is known in advance, we reformulate the PnP problem as a Maximum Likelihood minimization approximated by an unconstrained Sampson error function, which naturally penalizes the most noisy correspondences. Pre-estimating feature uncertainty in real experiments is, though, not easy. In this paper we model it as 2D Gaussian distributions representing the sensitivity of the underlying 2D feature detectors to different camera viewpoints. When using these noise models with our PnP formulation we still obtain promising pose estimation results that outperform most recent approaches.

Method: Let $\mathbf{u}_i = [u_i, v_i]^\top$ be an observed 2D point obtained using a feature detector. This observed value can be regarded as the true 2D projection $\overline{\mathbf{u}}_i$ perturbed by a random variable $\Delta\mathbf{u}_i$,

$$\mathbf{u}_i = \overline{\mathbf{u}}_i + \Delta\mathbf{u}_i \tag{1}$$

We assume that $\Delta\mathbf{u}_i$ is small, independent and unbiased, and model it as a Gaussian distribution with expectation $E[\Delta\mathbf{u}_i] = \mathbf{0}$ and 2×2 covariance matrix $E[\Delta\mathbf{u}_i\Delta\mathbf{u}_i^\top] = \mathbf{C}_{\mathbf{u}_i}$, which is known in advance.

Taking into account these uncertainties the PnP problem can be solved as the following Maximum Likelihood for all n correspondences,

$$\arg\min_{\Delta\mathbf{u}_i,\mathbf{x}} \sum_{i=1}^{n} \|\Delta\mathbf{u}_i\|^2_{\mathbf{C}^{-1}_{\mathbf{u}_i}} \qquad \text{subject to} \quad \mathbf{M}_{\overline{\mathbf{u}}_i}\mathbf{x} = \mathbf{0} \tag{2}$$

where $\mathbf{M}_{\overline{\mathbf{u}}_i}\mathbf{x} = \mathbf{0}$ enforce the 3D-to-2D projective constraints of the noise-free correspondences and $\mathbf{x}$ represents a set of control points in camera coordinates. Since we assumed the uncertainty $\Delta\mathbf{u}_i = [\Delta u_i \;\; \Delta v_i]^\top$ to be small, the perspective constraint can be approximated using first order perturbation analysis

$$\mathbf{M}_{\overline{\mathbf{u}}_i}\mathbf{x} = \mathbf{M}_{\mathbf{u}_i}\mathbf{x} - \Delta u_i \nabla_u \mathbf{M}_{\mathbf{u}_i}\mathbf{x} - \Delta v_i \nabla_v \mathbf{M}_{\mathbf{u}_i}\mathbf{x} = \mathbf{0} \tag{3}$$

where $\nabla_u\mathbf{M}_{\mathbf{u}_i}$ and $\nabla_v\mathbf{M}_{\mathbf{u}_i}$ are the partial derivatives of $\mathbf{M}_{\mathbf{u}_i}$ with respect to u and v; and as in [2], $\mathbf{M}_{\mathbf{u}_i}$ encodes the perspective constraints.

Using Lagrange Multipliers Eq. 2 is rewritten as an unconstrained minimization of a Sampson Error function and solved using the Fundamental Numerical Scheme (FNS) approach [1].

Finally, once $\mathbf{x}$ is estimated, the PnP problem is solved following the Procrustes analysis proposed in [2].

Uncertainties estimation: Estimating 2D feature uncertainties $\mathbf{C}_{\mathbf{u}_i}$ in real images is still an open problem. Our approach starts by detecting features on a given reference view $\mathbf{V}_r$ of the object of interest. Then, we synthesize m novel views $\{\mathbf{I}_1, \ldots, \mathbf{I}_m\}$ of the object, which sample poses around $\mathbf{V}_r$. We then extract 2D features for each $\mathbf{I}_j$, and reproject them back to $\mathbf{V}_r$, creating feature point clouds (see Figure 1c).

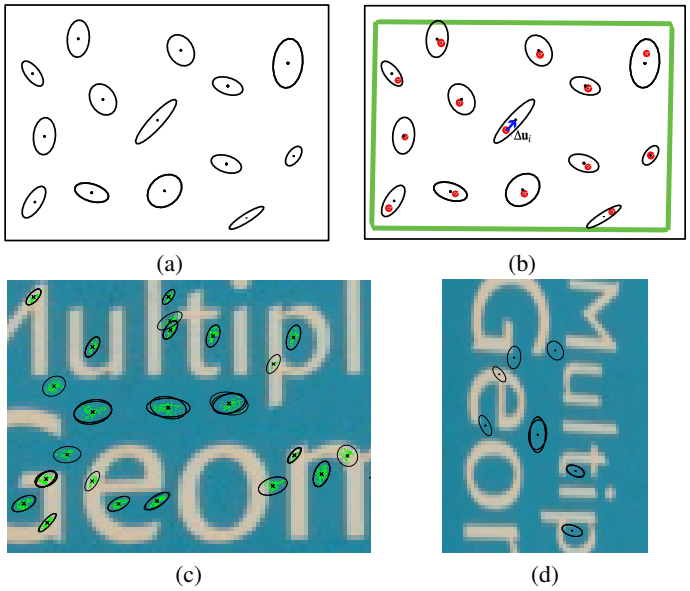

Figure 1: PnP problem with noisy correspondences. We assume 2D feature points are associated to particular noise models, as shown in (a). Our approach estimates a solution of the PnP problem that minimizes the Mahalanobis distances $\Delta\mathbf{u}_i$ shown in (b). The Green rectangle and red dots are the true projection of the 3D model and 3D points. Using our approach feature uncertainties $\mathbf{C}_{\mathbf{u}_i}$ (black ellipses) on real images are estimated for each reference view (c). In (d) uncertainties are aligned with a test image.

Once features are grouped we model each cluster i with a covariance matrix $\mathbf{C}_{\mathbf{u}_i}$. Note that this covariance tends to be anisotropic, thus it is not rotationally invariant. To achieve this invariance we use the main gradients as done by the SIFT detector. Fig.1(d) shows how each $\mathbf{C}_{\mathbf{u}_i}$ is rotated respect to the main gradients.

In practice, we found that $\mathbf{C}_{\mathbf{u}_i}$ accurately describes the uncertainties when the pose of $\mathbf{I}_j$ is close to the pose of the reference $\mathbf{V}_r$. This accuracy drops when camera moves away. In order to handle this, we defined a set of l reference images $\{\mathbf{V}_1, ..., \mathbf{V}_l\}$ under different poses and each one with its own uncertainty models. We experimentally found that a grid of reference images, taken all around the 3D object at every 20° in yaw and pitch angles, yielded precise uncertainty models.

Algorithm for real images is split into the following three main steps:

1. Estimate an initial camera pose without considering feature uncertainties using EPPnP. Let $[\mathbf{R}|\mathbf{t}]_{EPPnP}$ be this initial pose.
2. Pick the nearest reference view $\mathbf{V}_k$. Solving $\max_k \left(\frac{\mathbf{c}_k^\top}{\|\mathbf{c}_k\|} \cdot \frac{\mathbf{c}_{EPPnP}}{\|\mathbf{c}_{EPPnP}\|} \right)$, where $\mathbf{c}_*/\|\mathbf{c}_*\|$ are the normalized camera centers in world coordinates, being $\mathbf{c}_* = -\mathbf{R}_*^\top\mathbf{t}_*$.
3. Solve Eq. 2 using the covariances $\mathbf{C}_{\mathbf{u}_i}$ of the reference image $\mathbf{V}_k$, and $[\mathbf{R}|\mathbf{t}]_{EPPnP}$ for initializing the iterative process. The final pose $[\mathbf{R}|\mathbf{t}]_{CEPPnP}$ is obtained using Procrustes as in [2].

[1] W. Chojnacki, M. J. Brooks, A. Van Den Hengel, and D. Gawley. On the fitting of surfaces to data with covariances. *PAMI*, 2000.

[2] L. Ferraz, X. Binefa, and F. Moreno-Noguer. Very fast solution to the PnP problem with algebraic outlier rejection. In *CVPR*, 2014.

[3] C-P. Lu, G. D. Hager, and E. Mjolsness. Fast and globally convergent pose estimation from video images. *PAMI*, 22(6):610–622, 2000.

[4] Y. Zheng, Y. Kuang, S. Sugimoto, K. Aström, and M. Okutomi. Revisiting the PnP problem: A fast, general and optimal solution. In *ICCV*, 2013.