

# Planar PØP: feature-less pose estimation with applications in UAV localization

Adrian Amor-Martinez, Angel Santamaria-Navarro, Fernando Herrero, Alberto Ruiz, and Alberto Sanfeliu

**Abstract**—We present a featureless pose estimation method that, in contrast to current Perspective-n-Point (PnP) approaches, it does not require  $n$  point correspondences to obtain the camera pose, allowing for pose estimation from natural shapes that do not necessarily have distinguished features like corners or intersecting edges. Instead of using  $n$  correspondences (e.g. extracted with a feature detector) we will use the raw polygonal representation of the observed shape and directly estimate the pose in the pose-space of the camera. This method compared with a general PnP method, does not require  $n$  point correspondences neither a priori knowledge of the object model (except the scale), which is registered with a picture taken from a known robot pose. Moreover, we achieve higher precision because all the information of the shape contour is used to minimize the area between the projected and the observed shape contours. To emphasize the non-use of  $n$  point correspondences between the projected template and observed contour shape, we call the method Planar PØP. The method is shown both in simulation and in a real application consisting on a UAV localization where comparisons with a precise ground-truth are provided.

## I. INTRODUCTION

The use of small Unmanned Aerial Vehicles (UAV), such as multirotors, in search and rescue missions (SAR) have gained popularity in the research community due to their low-cost and high manipulability. Examples are those cases where the vehicle localizes the victims [1], [2], [3] or even gives assistance in a first instance [4], [5]. Also promising applications are appearing using Unmanned Aerial Manipulators (UAM), i.e. UAVs equipped with robotic arms, to perform some manipulation operations [6], [7], [8].

The vehicle's pose estimation is a very challenging problem to be solved during SAR missions, allowing for robot navigation and control among many other important tasks. Immediately after a disaster it is impractical the use of external infrastructure to localize the vehicle in both indoor and outdoor scenarios (e.g. GPS). Moreover, not only a robust and fast pose estimation is strictly required but also the sensors to obtain it have to consider some constraints in terms of weight and power consumption. UAVs, and specifically multirotors have a limited payload and power

This work has been partially funded by the AEROARMS EU project H2020-ICT-2014-1-644271, the CICYT project DPI2013-42458-P, and by Spanish MINECO and EU FEDER grant TIN2015-66972-C5-3-R.

A. Amor-Martinez, A. Santamaria-Navarro, F. Herrero and A. Sanfeliu are with the Institut de Robòtica i Informàtica Industrial, CSIC-UPC, Llorens i Artigas 4-6, Barcelona 08028, Spain, aamor, asantamaria, fherrero, sanfeliu@iri.upc.edu.

A. Ruiz is with Universidad de Murcia, Spain, aruiz@um.es.

The paper has supplementary downloadable material available at <http://ieeexplore.ieee.org>, provided by the authors. The material includes a video of the pose estimation presented in the paper.

supply. Considering these restrictions, a good option is to obtain a pose estimation using visual information, thus using cameras.

In a classical approach to camera pose estimation, we first have to detect regions of interest in the image which must be repetitive and easily identifiable (e.g. points, edges, corners, convexities or ellipses), meaning that feature descriptors [9] are strictly required. There exist a great number of descriptors (e.g. [10], [11], [12]) and their use will be subject to computational costs, illumination problems, image texture or camera resolution. Once detected, these 2D features are associated with 3D model points, known a priori, allowing to compute  $n$  point correspondences between two images or one image and a 3D model. Most common methods for pose estimation [13], [14] rely on these explicit correspondences. Although these approaches have been widely studied and used, they require a 3D model known a priori and, specially during SAR operations, this represents a hard constraint because the scenario could have changed. Moreover, when using PnP methods, some significant problems can arise related to the repeatability of the correspondences due to, for example, illumination changes, low camera resolution or blurring. In those cases where image features cannot be extracted the addition of artificial markers is necessary [15], [16], however it may not be practical during SAR operations.

The use of contours for object pose estimation has been recently studied in [17], which presents a solution using a single image by learning an object model of view-dependent shape templates. This approach, although very useful for general pose estimation is not appropriate for UAV tasks considering the achieved precision and computational cost. In contrast, our approach reaches higher precision, it is simpler since it deals only with planar natural shapes as shown in Fig. 1, and it does not require an extensive learning process (only an image per template and its scale), hence it requires lower computational burden.

Other methods like [18], [19], [20] provide a homography from contours of a planar region without correspondences but they require complex optimization schemes and the homography does not consider camera restrictions. In the recent alignment approach for deformable contours [21] the camera constraints are again not taken into account in the key optimization step, preventing accurate pose estimation for high precision applications.

In this paper we present a novel Planar PØP method to extract the pose of a UAV using only contours of natural shapes instead of feature descriptors, and without the need for artificial markers. This method consists of 4 steps: 1)

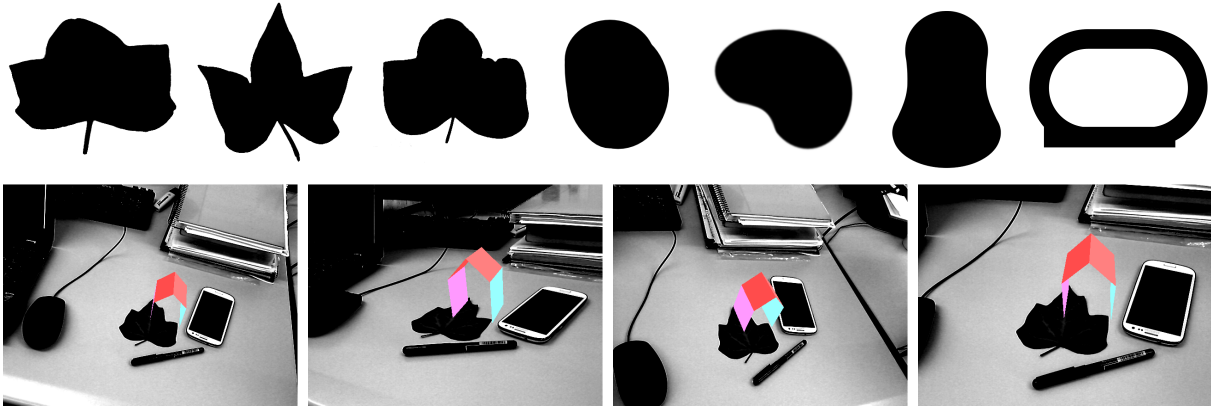


Fig. 1: Natural shapes (first row) and demonstration of an augmented reality application (second row). Because we are not using correspondences we can test our method with any natural (closed) shape that we might find. The method only needs initially a frontal view of the shapes.

initialization, to find a starting location to align the contours; 2) contour alignment, to look for the best homography alignment in the Homography-space using the method of the same authors explained in [21]; 3) pose extraction from the homography, to compute a first estimation of the pose using a novel method named Virtual Correspondences (VC); and 4) pose refinement in the pose-space. Finally, the Planar PØP method is validated using synthetic views and real-life experiments comparing the results using a precise Optitrack motion capture system as a ground-truth.

The remainder of this article is structured as follows. In the next section the Planar PØP method is presented. Validation and experiments are shown in Section III. Finally conclusions and remarks are given in Section IV.

## II. PØP: PERSPECTIVE FROM ZERO POINTS FOR PLANAR SHAPE CONTOURS

Let us consider a planar object, the *template contour*, that has been discretized as a polygon with  $n$  nodes  $\mathbf{T} = (T_1, \dots, T_n)$ , where  $T_i = (T_{i_x}, T_{i_y})$ , and an observed image, the *image contour*,  $\mathbf{O} = (O_1, \dots, O_m)$ , where  $O_i = (O_{i_x}, O_{i_y})$ . The image contour  $\mathbf{O}$  is obtained using standard image processing functions, contour extraction and polygon reduction.

We want to recover the pose of the *image contour* taken into account that this contour comes from the projection of the *template contour* in the image. The method that we will explain in this section does not require to find any corresponding points in the contours, instead we use the best global alignment between the image contour and a projection of the object template to find the pose. Because we do not use point correspondences we call the method PØP.

In order to find the best alignment, we define the alignment cost as the total area discrepancy between the projected *template contour* and the *image contour*. The error area between two regions can be efficiently obtained by an XOR (symmetric difference) clipping algorithm working on region boundaries [22], and their areas can be easily computed just from the contour nodes. For the sake of conciseness, we refer

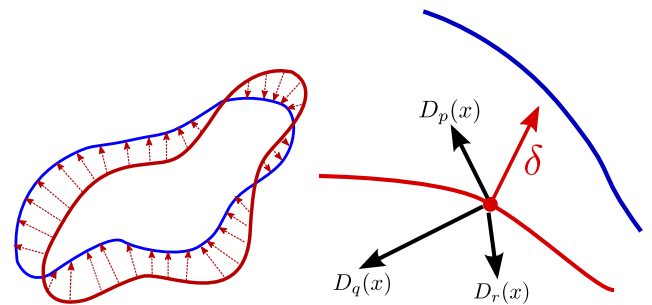


Fig. 2: The vector field  $\delta_p(x)$  shows the local deformation required to improve alignment and the effects of the transformation which must be combined to match  $\delta$ .

the reader to [22] for XOR operation details. In [21] this XOR cost was used in an efficient optimization algorithm to find the best homographic transformation between two contours. Briefly, this contour registration problem is formulated as follows. Given a transformation model  $\mathbf{x}' = \mathbf{W}_q(\mathbf{x})$  we find the parameters  $\mathbf{q}$  that minimize the error area for the image contour  $\mathbf{O}$  and template contour  $\mathbf{T}$ , which is expressed by

$$\mathbf{q}^* = \operatorname{argmin}_q \operatorname{XOR}(\mathbf{O}, \mathbf{W}_q(\mathbf{T})). \quad (1)$$

This is solved using Gauss-Newton's iterative optimization (GN): given a residual vector  $\mathbf{f}$  with Jacobian  $\mathbf{J} = \partial \mathbf{f} / \partial \mathbf{q}$ , the squared error  $C = 1/2 \mathbf{f}^T \mathbf{f}$  can be reduced by using the update rule

$$\Delta \mathbf{q} = -\mathbf{H}^{-1} \nabla C, \quad (2)$$

where  $\nabla C = \mathbf{J}^T \mathbf{f}$  and the Hessian is approximated by  $\mathbf{H} = \mathbf{J}^T \mathbf{J}$ .

Each node in the contour produces two residuals in  $\mathbf{f}$ , denoted by the vector  $\delta$ . Fig 2 shows the ideal  $\delta$  field in a hypothetical alignment example. Analogously, the corresponding two rows of the Jacobian will be denoted by  $D$ . Each component  $D_q$  quantifies up to first order the effect of parameter  $q$ . The optimization method combines the local

effects of all parameters of the transformation, trying to match the required correction  $\delta$ .

In this particular problem the required residuals and Jacobian can be efficiently computed in closed form just from the nodes of the XOR error polygons. The components of the gradient and the Hessian needed in (eq. 2), in terms of  $\delta$  and  $D_q$ , and for a polygonal contour with segments  $S_k$  joining nodes  $k$  and  $k + 1$ , can be expressed as:

$$\{\nabla C\}_r = \sum_{k=1}^n \int_{\mathbf{x} \in S_k} D_r(\mathbf{x}) \cdot \delta(\mathbf{x}) dx, \quad (3)$$

$$\mathbb{H}_{qr} = \sum_{k=1}^n \int_{\mathbf{x} \in S_k} D_q(\mathbf{x}) \cdot D_r(\mathbf{x}) dx. \quad (4)$$

The details of the computation of the above expressions can be found in [21]. The key idea is that the general homographic transformations form a group, and therefore we can use the *inverse compositional* approach [23]. The components  $D_q$  can be precomputed, yielding a very efficient alignment process. It only requires the signed XOR residual and low-order powers of the coordinates of the intersection nodes in each optimization step.

This alignment method presented in [21] is efficient and accurate for full projective models, but we are actually interested in the estimation of the best camera pose  $[R|t]$ , an element of the SE3 group of rigid motions. This constraint is essential to achieve high precision localization from noisy images.

There are several heuristic methods to estimate a pose from a more general homography H. One natural and simple method proposed in this work generates ‘‘Virtual Correspondences’’ (VC) using the estimated H and a standard PnP algorithm to get the pose that best fits this transformation. Another method called Infinitesimal Plane-based Pose Estimation (IPPE) has been proposed in [24] and works directly with the structure of H.

These methods obtain good estimates when the amount of noise is small. However, in more realistic scenarios with low quality and blurred images, the extraction of a consistent pose (6 DOF) just from an intermediate and unconstrained homographic model (8 DOF) is usually not accurate enough for high precision visual localization applications. Some distortions produced by image noise can be captured by homographic models that are inconsistent with rigid transformations.

The correct approach is to minimize the alignment error with respect to the true pose parameters. In this paper we present a reasonably efficient method to achieve this goal. The main complication is that the set of projections of rigid 3D transformations of a planar region is not a group, and hence the inverse compositional approach cannot be applied. Instead, a forward additive alternative must be used, which requires recomputation of the Jacobian in each optimization step. We will show that this problem is not as severe as it seems. While the contributions of the error segments can be no longer computed in closed form, an approximate numeric

integral can be still obtained for the final refinement steps, without a significant increase in the total computing time of the algorithm.

The PØP method has four steps:

- **Step 1: Initialization.** In this process we look for a good starting location of the first projection of the template contour  $T$  on the image contour  $O$ . We can use different techniques to obtain this first approximation. In our case we start from an affine transformation that can be computed in closed form from the ‘‘whitened’’ polygons, without the need of any correspondences. The output of this step is a coarse alignment approximation from where we will do the refinements. This initial step can be optionally improved with an inverse compositional step in the homography space using precomputed derivatives for some global contour properties.
- **Step 2: Refinement of the contour alignment in the Homography space.** We use the inverse compositional XOR method described above to look for the best homography alignment between the projection of  $T$  and  $O$ . The output of this step is the homography  $H_k$  that gets the minimum XOR error.
- **Step 3: Computation of the initial pose from the homography.** In contrast to [21], we here propose a Virtual Correspondences method to estimate a pose from the homography  $H_k$  obtained in Step 2, which consists on the following
  - Take  $m$  discretized points of the template  $T = (T_1, \dots, T_m)$ . These points of interest are obtained from standard image processing functions [25].
  - Project these  $m$  points using the homography  $H_k$ , obtaining the projected points  $T^I = (T_1^I, \dots, T_m^I)$ .
  - Use the correspondences between the pairs  $T_1$  with  $T_1^I$ ,  $T_2$  with  $T_2^I$ , ...,  $T_n$  with  $T_n^I$ .
  - Apply a classical PnP method with the  $m$  correspondences.

The output of this step is the pose  $p^{init}$ .

- **Step 4: Refinement of the pose in the pose-space.** In this novel step we use again the XOR method to find the best alignment, but now in the pose-space expressed as

$$\operatorname{argmin}_p \delta_{XOR}(O, W_p(T)). \quad (5)$$

The initialization starts with the pose  $p^{init}$ . The warping function is

$$W_p(T) = f(PT'), \quad (6)$$

with  $P = [R|t]$  the pose,  $T'$  the discretized points of the template contour in homogeneous coordinates, and  $f$  the transformation to inhomogeneous coordinates. The derivatives required by GN for the 6 parameters of

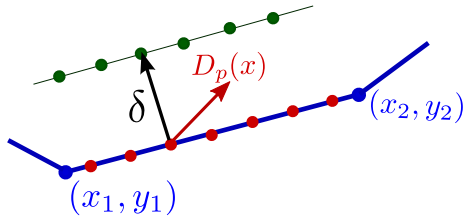


Fig. 3: Discretized contributions to  $\nabla C$  and  $H$ .

$P = [R|t]$  can be compactly expressed as

$$D_p = \begin{bmatrix} -y' & -x'^2 - 1 & -x'y' & z' & 0 & -x'z' \\ x' & -x'y' & -y'^2 - 1 & 0 & z' & -y'z' \end{bmatrix} \quad (7)$$

where  $[x \ y \ z]^T$  is the predicted 3D location of each contour node using the current pose, and  $x' = x/z$ ,  $y' = y/z$ , and  $z' = 1/z$ . The first three columns correspond to the orientation angles and the last ones correspond to the displacement. The above expression can be easily obtained from the incremental local parameterization of the pose at the current estimate. For example, for the first angle we have

$$\begin{bmatrix} x_h \\ y_h \\ z_h \end{bmatrix} = \begin{bmatrix} \cos(a) & 0 & \sin(a) \\ 0 & 1 & 0 \\ -\sin(a) & 0 & \cos(a) \end{bmatrix} [\mathbb{I}|0] \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}. \quad (8)$$

Unfortunately, the contributions to the integrals in eqs. (3) and (4) can no longer be expressed in terms of simple moments. Furthermore, the planar warping function arising from the 3D projection is not invertible. For this reason, we must resort to the more general forward additive optimization scheme. In each step the currently estimated pose is improved with an incremental correction that minimizes the XOR error area in image space. To accomplish this, the integrals in (3) and (4) are computed as a sum over a set of discrete points along each segment (Fig. 3).

The output of this step is the pose  $p^{final}$ .

This last discretization is quite different from the usage of Virtual Correspondences method (VC) in step 3. The auxiliary nodes are used just to compute approximate numeric integrals required by the minimization of the XOR regions. Correspondences are never used between points.

This final refinement requires very few iterations as it usually starts from an accurate initialization. An additional advantage of a final forward additive optimization step is that it minimizes the global geometric error in the image space, theoretically producing optimal estimates (the inverse compositional method is more efficient but it minimizes the error in the template space).

### III. VALIDATION AND EXPERIMENTS

To validate the proposed method, we created a database of different shapes (e.g. see Fig. 1). These shapes are designed in a way that either is very difficult to find image points corresponding to the template, or these points are

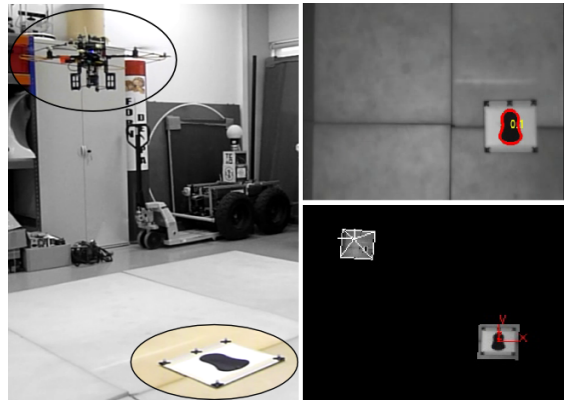


Fig. 4: Left: ASCETEC Pelican robot flight. Right: shape detection and camera pose provided by Planar PØP.

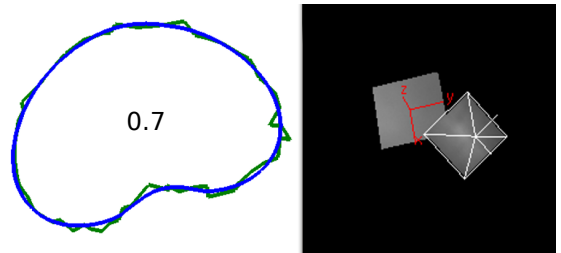


Fig. 5: Alignment and resulting pose of a noisy synthetic example. Green: Observed contour. Blue: Projected template with estimated pose. The number inside the alignment represents the normalized XOR area discrepancy between the projected template and the observed contour.

bad conditioned (e.g., they can be close to a straight line). Traditional methods for pose estimation are unable to extract pose in such shapes. In contrast, the proposed method only needs initially a frontal view picture for every shape (selected automatically or, as in our case, by a user selection done in the image plane), its scale and it is general enough to work with any closed contour.

We propose two different arrangements to evaluate the method. On the one hand, we generate synthetic views of the shapes contained in the database. Secondly, we present real pose estimation experiments with an ASCETEC PELICAN quadrotor (see Fig. 4) with a camera endowed below and pointing downwards. Finally we present some results in outdoor sequence of images.

#### A. Error evaluation with synthetic views

We generate synthetic shapes under plausible camera perspective views and add noise artificially as shown in Fig. 5. More specifically, we generate  $n$  arbitrary rotations and translations resulting in  $n$  camera poses  $P_i^{true}$ . Given  $\epsilon \sim N(0, \sigma)$ , the contours generated for the template  $T_j$  are:

$$O_{i,j}^{noisy} = W_{P_i^{true}}(T_j^{true}) + [\epsilon_1, \epsilon_2, \dots]^T$$

Note that this process must be done carefully to avoid self intersections in the resulting contour. For comparison purposes, once the observed contours are obtained we launch

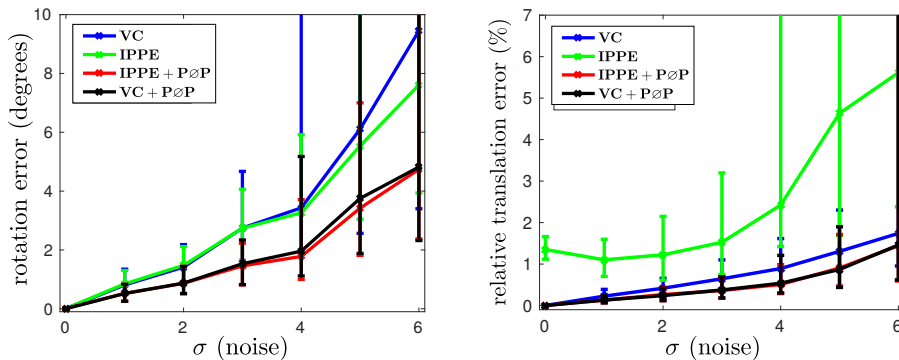


Fig. 6: Comparison of different methods using synthetic views: VC (Virtual Correspondences), IPPE [24] (Infinitesimal Plane-based Pose Estimation), and both of these methods refined with the proposed PØP method

different methods considering the frontal views  $T_j$  and the noisy contours  $O_{i,j}^{\text{noisy}}$  resulting in an estimation of the camera pose  $\hat{P}_i$ . The resulting errors are computed as the difference in rotation angles and normalized translation coordinates.

The following methods are compared:

- 1) Virtual Correspondences (VC) from homography ( $p^{\text{init}}$  from Step 3).
- 2) Infinitesimal Plane-based Pose (IPPE) [24] from homography.
- 3) PØP initialized with Virtual Correspondences (VC) method.
- 4) PØP initialized with Infinitesimal Plane-based Pose (IPPE) method.

The resulting comparison can be seen in Fig. 6 and we can draw several conclusions. On the one hand, it is interesting to note that the PØP method works well with both initializations and improves considerably the pose estimation. Also note that the IPPE method [24] needs to refine the initial translation estimation and for that purpose they propose a point correspondence based approach to minimize the reprojection error. In contrast, our PØP method is able to obtain a good pose without corresponding points even when the initialization is not good (note the difference between IPPE and IPPE+PØP in both graphs).

### B. Natural shape detections applied to UAVs localization

For the real robot experiments, we used an ASCTEC PELICAN quadrotor with a MATRIX mvBlueFOX camera (752x480 pixels of resolution) attached below. We calibrated the camera, thus we know the camera intrinsic parameters  $K$  and the radial distortion is corrected. An Optitrack motion capture system is used as a ground-truth by attaching several infrared markers to the quadrotor and the shapes to obtain an accurate pose measurement. These shapes are the same (resized) used in the synthetic validation section (see Fig. 4).

In each experiment the robot orbits above the shape autonomously thanks to the motion capture system. This is particularly important because human-driven trajectories are more abrupt than autonomous ones.

Table I shows the comparison between the different methods using Optitrack as a ground-truth. Notice how the align-

	Abs err (m)	Rel err (%)	normalized XOR
VC	<b>0.0167</b>	<b>1.2194</b>	0.1539
VC+PØP	0.0185	1.3448	<b>0.1332</b>
IPPE	0.0226	1.6213	0.2719
IPPE+PØP	<b>0.0217</b>	<b>1.5656</b>	<b>0.1393</b>

TABLE I: Comparison of different methods using Optitrack as a ground-truth: VC (Virtual Correspondences), IPPE [24], and both of these methods refined with the proposed PØP method

ment (XOR) error is always reduced. However, in the case of the pose error, the Virtual Correspondence method (VC) provides better estimate than the refinement, even though the XOR error was reduced.

The reason is that due to the use of low camera resolution in the robot, the edges in the observed contour are not in the correct place and the error is not gaussian (which is also bad for traditional PnP methods). In those cases, trying to improve the alignment of the observed contour (green) would reduce the XOR error, but the pose could be overfitted. In contrast, when using IPPE method as initialization the improvement is more evident, where both the XOR and the real pose errors are reduced.

### C. Camera pose in outdoor sequence of images

We have tested our method in an outdoor scenario. Because there is not the ground truth, the purpose has been to see that the error in object detection for a particular single shape was smaller than in the other objects in a sequence of outdoor images. Fig. 7 shows snapshots of two different experiments where the camera is moved around a simulated stones path. During these experiments a single shape is configured (top left corner of gray images) and, when the expected stone appears in the camera field of view, the method extracts the camera pose. Then, in case we set up all possible shapes, the method can allow a precise localization of the aerial vehicle while following the stones path.

## IV. CONCLUSIONS

This paper presents a camera pose estimation method, planar PØP, which does not require neither feature corre-

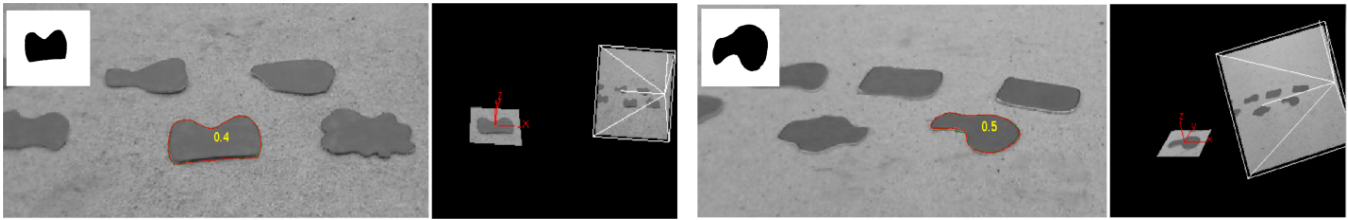


Fig. 7: Template detection and camera pose estimation using Planar PØP over a sequence of outdoor images consisting on a simulated stones path.

spondences nor corresponding points or artificial markers in the scene. Hence, we can precisely obtain the camera pose from natural shapes lacking of those attributes. The learning process of a new natural marker is done by capturing its shape from a vertical position and noting down the scale of the image shape. The method achieves higher precision on pose estimation than other methods (*e.g.* VC or IPPE), which is very important for UAV localization and specifically for aerial manipulation operations, although is sensible to contour extraction. The method uses an alignment cost defined as the total area discrepancy between the projected template and the image contours. The error area between them is obtained using an XOR (symmetric difference) clipping algorithm working on region boundaries. The minimization of the alignment cost is done in the XOR using a Gauss-Newton's iterative optimization process. Shape datasets have been used to compare and validate the proposed Planar PØP approach with different literature methods. First, a validation with synthetic camera views shows how the refinement of methods using Planar PØP increases the alignment and the pose accuracy. On the other hand, we provided comparisons of the pose estimation error between different methods and ground-truth measurements, obtained with a precise Optitrack motion capture system, in real robot experiments using a quadrotor with a camera endowed below.

## REFERENCES

- [1] M. A. Goodrich, J. L. Cooper, J. A. Adams, C. Humphrey, R. Zeeman, and B. G. Buss, "Using a mini-uav to support wilderness search and rescue: Practices for human-robot teaming," in *2007 IEEE International Workshop on Safety, Security and Rescue Robotics*, Sept 2007, pp. 1–6.
- [2] P. Rudol and P. Doherty, "Human body detection and geolocalization for uav search and rescue missions using color and thermal imagery," in *Aerospace Conference, 2008 IEEE*, March 2008, pp. 1–8.
- [3] T. Tomic, K. Schmid, P. Lutz, A. Domel, M. Kassecker, E. Mair, I. L. Grix, F. Ruess, M. Suppa, and D. Burschka, "Toward a fully autonomous uav: Research platform for indoor and outdoor urban search and rescue," *IEEE Robotics Automation Magazine*, vol. 19, no. 3, pp. 46–56, Sept 2012.
- [4] A. Momont, "Drones for good," Ph.D. dissertation, TU Delft, Delft University of Technology, 2014.
- [5] L. A. Haidari, S. T. Brown, M. Ferguson, E. Bancroft, M. Spiker, A. Wilcox, R. Ambikapathi, V. Sampath, D. L. Connor, and B. Y. Lee, "The economic and operational value of using drones to transport vaccines," *Vaccine*, 2016.
- [6] A. Santamaria-Navarro, V. Lippiello, and J. Andrade-Cetto, "Task priority control for aerial manipulation," in *Safety, Security, and Rescue Robotics (SSRR), 2014 IEEE International Symposium on*, Oct 2014, pp. 1–6.
- [7] V. Lippiello, J. Cacace, A. Santamaria-Navarro, J. Andrade-Cetto, M. Trujillo, Y. R. Esteves, and A. Viguria, "Hybrid visual servoing with hierarchical task composition for aerial manipulation," *Robotics and Automation Letters, IEEE*, pp. 259–266, to appear.
- [8] A. Jimenez-Cano, J. Martin, G. Heredia, A. Ollero, and R. Cano, "Control of an aerial robot with multi-link arm for assembly tasks," in *Robotics and Automation (ICRA), 2013 IEEE International Conference on*. IEEE, 2013, pp. 4916–4921.
- [9] T. Tuytelaars and K. Mikolajczyk, "Local invariant feature detectors: A survey," *Found. Trends. Comput. Graph. Vis.*, vol. 3, no. 3, pp. 177–280, July 2008. [Online]. Available: <http://dx.doi.org/10.1561/0600000017>
- [10] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *Computer vision—ECCV 2006*. Springer, 2006, pp. 404–417.
- [11] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [12] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "Orb: an efficient alternative to sift or surf," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 2564–2571.
- [13] V. Lepetit, F. Moreno-Noguer, and P. Fua, "Epnnp: An accurate o (n) solution to the pnp problem," *International Journal of Computer Vision*, vol. 81, no. 2, pp. 155–166, 2009.
- [14] C.-P. Lu, G. D. Hager, and E. Mjølness, "Fast and globally convergent pose estimation from video images," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 6, pp. 610–622, 2000.
- [15] H. Kato and M. Billinghurst, "Marker tracking and hmd calibration for a video-based augmented reality conferencing system," in *Augmented Reality, 1999.(IWAR'99) Proceedings. 2nd IEEE and ACM International Workshop on*. IEEE, 1999, pp. 85–94.
- [16] M. Fiala, "Designing highly reliable fiducial markers," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 7, pp. 1317–1324, 2010.
- [17] N. Payet and S. Todorovic, "From contours to 3d object detection and pose estimation," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 983–990.
- [18] P. K. Jain, "Homography estimation from planar contours," in *3D Data Processing, Visualization, and Transmission, Third International Symposium on*, 2006, pp. 877–884.
- [19] S. F. Zhai Fu JingMing Zhai and Z. Jing, "Homography estimation from planar contours in image sequence," *Opt. Eng.* 49, 2010.
- [20] A. Ruiz, P. de Teruel, and L. Fernández, "Robust homography estimation from planar contours based on convexity," in *ECCV'06, 2006*, pp. 107–120.
- [21] A. Amor-Martinez, A. Ruiz, F. Moreno-Noguer, and A. Sanfeliu, "On-board real-time pose estimation for uavs using deformable contour registration," in *Robotics and Automation (ICRA), 2014 IEEE International Conference on*. IEEE, 2014, pp. 2595–2601.
- [22] G. Greiner and K. Hormann, "Efficient clipping of arbitrary polygons," *ACM Transactions on Graphics (TOG)*, vol. 17, no. 2, pp. 71–83, 1998.
- [23] S. Baker and I. Matthews, "Lucas-kanade 20 years on: A unifying framework," *International Journal of Computer Vision*, vol. 56, no. 3, pp. 221–255, 2004.
- [24] T. Collins and A. Bartoli, "Infinitesimal plane-based pose estimation," *International Journal of Computer Vision*, vol. 109, no. 3, pp. 252–286, 2014.
- [25] D. H. Douglas and T. K. Peucker, "Algorithms for the reduction of the number of points required to represent a digitized line or its caricature," *Canadian Cartographer*, 1973.