

Vision-based handling tasks for an autonomous outdoor forklift

Cédric Pradalier

► **To cite this version:**

Cédric Pradalier. Vision-based handling tasks for an autonomous outdoor forklift. 6th International Conference on Field and Service Robotics - FSR 2007, Jul 2007, Chamonix, France. Springer, 42, 2007, Springer Tracts in Advanced Robotics. <inria-00194757>

HAL Id: inria-00194757

<https://hal.inria.fr/inria-00194757>

Submitted on 7 Dec 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Vision-based handling tasks for an autonomous outdoor forklift

Cédric Pradalier

CSIRO ICT Centre, Autonomous Systems Laboratory, Brisbane, Australia

Summary. In the aluminium industry, Hot Metal Carriers (HMCs) perform the task of transporting molten aluminium from the smelter (where the aluminium is made) to the casting shed where it is turned into block products. The vehicles weigh approximately 20 tonnes unloaded and resemble forklifts except they have a dedicated hook for manipulating the load, called the crucible. In the perspective of a full automation of this kind of vehicle, this paper main contribution is the implementation of robust outdoor crucible handling capacities, relying on the autonomous tracking of especially designed fiducials attached to the crucible handle.

1 Introduction



Fig. 1. Hot Metal Carrier picking up a crucible

Hot Metal Carriers (HMCs, see fig. 1) perform the task of transporting molten aluminium from the smelter (where the aluminium is made) to the casting house where it is turned into block products. The vehicles weigh approximately 20 tonnes unloaded and resemble forklifts except they have a dedicated hook for manipulating the load rather than fork tines. The molten aluminium is carried in large metal crucibles. The crucibles weigh approximately 2 tonnes and they can carry 8 tonnes of molten aluminium at 700 degrees Celsius. Therefore HMC operations are considered heavy, hot, and hazardous, with safety of operation a significant issue.

Our long term objective is the full automation of a HMC in order to fulfil tasks coherent with its industrial usage. The current state of our HMC automation has been described in [6]. This paper will address especially the implementation of robust outdoor crucible handling capacities based on computer vision. In an

industrial context, it may be argued that laser-based implementation would provide better overall robustness. Nevertheless, we chose computer vision for its versatility and wider field of view.

The vehicle is equipped with a mast that supports the pickup hook. The mast is controllable in tilt and the hook is also controllable in height. Mechanically, the hook is linked to the top of the mast and raising the top of the mast with an hydraulic cylinder raises the hook. These mechanical aspects of the vehicle will be important for the remaining of the paper since we equipped the vehicle with two pan/tilt/zoom cameras attached to the top of the mast. This is required in order to have a good vantage point toward the pickup scene, but as we will see later, this introduces complexity in the evaluation of the exact position of the camera in the vehicle frame of reference.

1.1 Related work

As expected for such a widely used industrial task, forklift handling has attracted a lot of attraction in the robotics research community, as well as in the industry, and commercial products are already available. These products are, to our knowledge, mainly using laser range finders for most of their activities and limited to indoor environments.

In the research community, some work can be found on vision-based cargo handling operations. The list below is not exhaustive but rather an overview of what we have found to be typical approaches. Garibotto *et al.* [3] worked on an indoor forklift detecting and localising pallets using vision. The localisation is achieved by detecting a H-shaped landmark painted on the ground and the pallets are identified by the dark appearance of the tine engagement holes. This method is not generic enough to be adapted to our cargo and we estimated that then relying on shadows for outdoor vision based operations is unlikely to provide the required robustness.

Also, Seelinger *et al.* [5] presented a fiducial-based pallet localisation system. The fiducial they describe are circular discs used in the motion capture industry. They provide very good response to corner detectors such as Harris' detector, and may be good enough for an indoor controlled environment, but they are not distinctive enough for reliable outdoor detection.

Most of the control algorithms discussed in the above-mentioned research into load handling could be applied to our crucible handling task. However, the pallet sensing methods are not suitable due to their low saliency and reliability in an outdoor, industrial setting. Nevertheless, for an industrial application we decided that the most reliable approach was to use fiducials attached to the payload, and to design them so as to maximise the distinctiveness in the environment. The next sections will elaborate on this approach.

2 Crucible Operations

The key functionality of a Hot Metal Carrier is its ability to handle the crucible. Two main operational phases can be distinguished: crucible pickup and crucible drop off.

2.1 Crucible drop off

Drop off is an easy manoeuvre from an automation point-of-view. No sensing is required and a simple ballistic manoeuvre is sufficient (Figure 2).

2.2 Crucible pickup

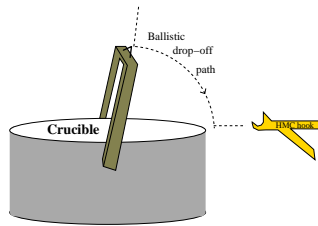


Fig. 2. Schematic of the hook movement during a drop off manoeuvre.

The pickup manoeuvre is much harder than the drop off. It can be divided into two steps: first, an approach step where the hook is visually guided toward the pickup point in the middle of the crucible handle (figure 7), then the actual pickup. The latter is an easy manoeuvre, again a ballistic movement, similar to drop off (Figure 2).

The approach part is more complex. It is principally based on the on-board camera detecting the crucible from about 5 m. Using sensor-based servoing and not a memory of the position where the crucible was last dropped off is justified by two main argument: first, in a real site, several agents handle the crucible (cranes,

lifts, and other HMCs). If some are not computer controlled or instrumented, there will be inaccuracies and uncertainties on the crucible drop off positions. Secondly, even during computer controlled drop off from our autonomous HMC, the crucible can swing around the hook and produce a different drop off position than previously. Most significantly, the crucible's orientation can vary in an unpredictable manner.

During the pickup manoeuvre the critical part is the crucible localisation in the vehicle frame. The description of our implementation of the crucible localisation forms the major part of this paper. Using the localisation, an approach trajectory can be implemented as described in section 5.

3 Landmark based localisation

3.1 Landmark choice

Our aim is to implement a robust vision-based handling task for an outdoor forklift. To maximise the resilience of our system to external perturbation, we decided to use artificial fiducials that can be attached to our payload.

As mentioned before, some existing vision-based industrial applications use motion capture fiducials, i.e. small disks with black and white quadrants. These fiducials are not reliable enough in an outdoor uncontrolled environment, so we chose to use self-similar landmarks, as described by Scharstein and Briggs[1].

Self-similarity

A p -similar function for $0 < p < 1$ is a function $f : \mathbb{R}^+ \rightarrow \mathbb{R}$ such that $\forall x > 0, f(x) = f(p \cdot x)$. It is essential to note that the p -similarity is invariant to change of scale ($\forall x > 0, \forall k > 0, f(k \cdot x) = f(p \cdot k \cdot x)$). In the context of a computer vision application, this property is interesting since, if a method to detect self-similarity is available, it will be possible to do so from whatever distance¹ since, with a pin-hole camera, a change of range results in a change of scale in the observed image.

When designing a self-similarity detector, the fact that all constant functions are p -similar for all p is problematic since any uniform region in an image will give a strong response on the self-similarity detector. To solve this, [1] introduces the notion of anti-similarity: a function f is p -antisimilar if $\forall x > 0, |f(x) - f(p \cdot x)| = 1$.

The p -similarity and \sqrt{p} -antisimilarity are then used by Scharstein and Briggs to define the “Self-similar square wave” function which is both p -similar and \sqrt{p} -antisimilar (with $\lfloor \cdot \rfloor$ representing the floor function):

$$s_p(x) = \lfloor 2(\log_p x - \lfloor \log_p x \rfloor) \rfloor \quad (1)$$

This function maps \mathbb{R}^+ to 0, 1. It is used to print black and white landmarks, as shown in fig. 3.



Fig. 3. Self-similar landmark as described in [1], with $p = \frac{2}{3}$

In order to recognise the self-similar landmarks, [1] defines a matching function that will respond strongly to a sequence of pixels looking like a p -similar and \sqrt{p} -antisimilar function. For a pixel located at (x, y) in image I , this is achieved by integrating the p -similarity and \sqrt{p} -antisimilarity conditions on a window of length w starting at (x, y) :

¹ as long as the resolution provides enough information on the observed object

$$\begin{aligned}
m_{+x}^w(x, y) &= \frac{1}{w} \int_0^w \left| I(x + \xi, y) - I(x + \sqrt{p} \cdot \xi, y) \right| d\xi \\
&\quad - \frac{1}{w} \int_0^w \left| I(x + \xi, y) - I(x + p \cdot \xi, y) \right| d\xi
\end{aligned} \tag{2}$$

In this equation, the window is made of the w pixels after (x, y) , on an image row. We identify this function by the $+x$ subscript. The function can be obviously modified to use the w pixels before, above, or below (x, y) . We name these functions m_{-x}^w , m_{-y}^w and m_{+y}^w respectively. In a practical implementation, w would take values in the $[20, 100]$ pixel range. We use 40, as in [1] as a good balance between robustness and computation cost.

Landmark improvement

The single landmark used in [1] and shown in fig. 3 is very reliably detected, especially in an indoor setting. In a more variable setting such as outdoors, we found it suffers from two main defects. First, even if the vertical line of peaks in $m_{+x}^w(x, y)$ is always very salient and very well located horizontally, its vertical position is more uncertain: strong contrasts, JPEG encoding, sensor saturation or image slant can make the vertical position of the line very unreliable. Then, if we consider an application in an industrial environment, corrugated iron used to create shed walls can create a nearly self-similar pattern when observed with enough perspective. In this case, this would generate a substantial number of false positive.

To increase the robustness of our landmark, we have tested two alternatives: an orthogonal landmark made of two self-similar pattern patched together with a 90° rotation, and a circular landmark where the self-similarity is radial instead of being axial. The circular landmark is less sensitive to rotation and motion blur, but more sensitive to occlusions than the orthogonal one. Consequently, the circular landmark was used for an underwater robot (see [4]) and the orthogonal one was used in this article. Both alternatives are depicted in fig. 4. Orthogonal landmarks can be made in two versions easily distinguishable: a *left* version, where the horizontal landmark goes from left to right, on the left of an up-to-down vertical landmark, and a *right* version, resulting of a flip of the other one over a vertical axis. The former is the one that appears in fig. 4.

To detect an orthogonal landmark, we use the self-similar landmark detector form [1]. For a given search orientation, this gives us a segment identified as laying on the right hand border of the landmark (see fig. 3). After the extraction of all the segments, for left-right, right-left and top-down orientation, we select only points where one segment is close enough to an other one, and roughly orthogonal. The image position of the landmark will then be the intersection of its two orthogonal segments.

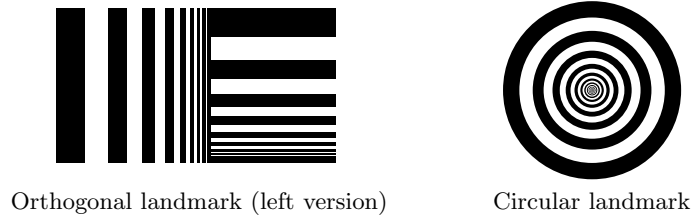


Fig. 4. Improved self-similar landmark

3.2 Localisation

In order to localise the vehicle load, we need to detect and identify the fiducials attached to it, and from their location in the image, estimate their location with respect to the vehicle end-effector. In the context of our forklift handling task, our camera is mounted on a pan-tilt-zoom (PTZ) head, attached to the forklift mast. To map the image observations to world coordinates, we have two options: either calibrate the full kinematic chain linking the camera frame to the forklift reference frame, or calibrate directly the relation between pixel coordinates and world coordinates for a fixed mast pose. The remaining of this section will compare both approaches.

Calibrating the kinematic chain

From a scientific point of view, this is the most attractive solution: it is more generic and closer to the physical system. On the other hand, the full kinematic chain, as described in fig. 5, is quite complex. On a large vehicle, most of these lengths or angles cannot be accurately measured, so they would have to be calibrated from indirect measurements. Even if this can be done, the accumulation of uncertainties along the kinematic chain will inevitably result in inaccuracies greater than what is acceptable for our application.

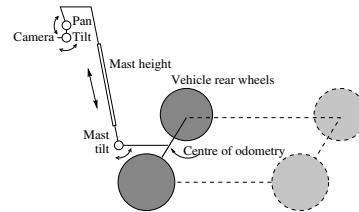


Fig. 5. Kinematic chain linking the camera lens to the vehicle centre of odometry

Calibration of the pixel-to-world relation

As an alternative to the full calibration, we can assume that the payload localisation will only be required while the forklift mast is in a known constant position. To constrain the problem even more, we will assume that the vehicle moves in a planar environment, and that the payload fiducials are located at

a known constant height above the ground. This simplifies the problem since the fiducials need only to be localised in a plane parallel to the ground plane. From now on, this plane will be named *target plane*.

Our objective now is to calibrate the relationship between pixel coordinates and target position in the target plane. With a fixed camera, this would only require to estimate an homography, which is easily estimated using the mapping of at least four points.

When considering a PTZ camera, an observation is not only the pixel coordinates (i, j) but also the pan (p), tilt (t) and zoom (z) of the camera. Nevertheless, a similar technique can be used. Let us first define two frames: the camera frame is aligned with the image plane and the optical axis, and the base frame is the frame of the PTZ head. The relationship between these frames is a rotation defined by the pan and tilt angles.

We can now define our mapping in two phases. First, for a given pan, tilt and zoom, an image pixel can be mapped to a unit vector going out of the origin, in the base frame. This mapping is bijective. Then, let us consider a plane \mathcal{P} located between the camera and the target plane², at unit distance of the camera for instance. The mapping associating each unit vector v to the intersection of the half-line supported by v with \mathcal{P} , is invertible as long as the intersection exists. This mapping can be computed from purely geometric considerations and does not require calibration.

Finally, the image in P of at least four points from the target plane is enough to estimate an homography mapping P to the target plane.

It is important to notice that the resulting mapping, from (i, j, p, t, z) to coordinates in the target plane, is the composition of two bijective mappings, and thus bijective too.

Note: It may be argued that this mapping model is less valuable than an estimation of the fixed $6DOF$ pose of the base frame in the vehicle reference frame. When developing this application, we effectively tried this method first, as well as, modelling the mapping using a 3^{rd} -degree multivariate polynomial function. In our setting, the two-phase homography resulted in more accurate results.

4 Payload localisation and tracking

Our objective here is to describe how the vision system described in the previous sections can be used to localise and track our payload. Let us first recall the setup: our vehicle is equipped with a single hook that must be passed into the eye of the payload handle in order to realise a successful pickup. The pickup phase is the only time when the payload localisation is required.

² Theoretically only a plane parallel to the target plane would verify this, but in practice we only want a plane separating the camera and a region of interest in the target plane.

The approach to the handle must be done on a trajectory perpendicular to the handle, consequently, we not only need to estimate the pose of the handle's eye, but also the orientation of the handle. Nevertheless, to simplify the problem, the payload handle is a massive rigid object and we can safely assume that it will always lay inside (or close enough to) the target plane.

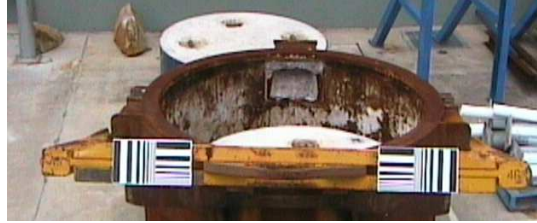


Fig. 6. Fiducial and their placement on the crucible

In order to reliably and accurately obtain the orientation of the handle, we set up two orthogonal self-similar landmarks (one *left* and one *right*). From the camera located on the forklift mast, a typical tracking image is illustrated in fig. 6.

Using the landmark extractor described in section 3 and the mapping described in section 3.2, it will be nearly straight forward to localise the handle. Three aspects have to be taken into considerations:

Conservation of field of view: this is required to account for the vehicle movement while approaching, and to ensure that the landmark is big enough in the image frames. This is achieved by controlling the PTZ head to put the current position estimate in the centre of the image, with enough resolution.

Long processing time: The pickup of our handle requires a position accuracy better than 0.05 m. Consequently, even at very small speeds, the time spent processing the images must be accounted for when integrating each observation.

Handle identification: In our application, with only one payload available, we have not developed any specific identification software so far. Our landmarks are distinctive enough to avoid any false positives and to identify each side of the handle without ambiguity.

In order to be robust to multiple possible perturbations, the handle localisation is integrated into a particle filter[2] and fused with movement information coming from the vehicle odometry.

5 Crucible approach

This section will describe the final piece of our application: the control of the approach. The payload localisation provides us with a continuous measurement of the payload position and orientation in the vehicle frame. Considering the non-holonomic properties of the vehicle, the best we can do is try to servo to the line orthogonal to the payload passing through the pickup point (see fig. 7).

The approach trajectory is done at constant speed v while the steering of the vehicle is controlled using a standard control law:

$$\dot{\theta} = K_{\theta}\Delta\theta + K_y v \frac{\sin \Delta\theta}{\Delta\theta} \Delta y \quad (3)$$

where K_{θ} and K_y are tunable gains, and other variables are illustrated on fig. 7. According to the initial error,

this controller may bring the vehicle to the pickup line. In this case the pickup can continue successfully. In other cases, we need to reliably detect this and generate a mitigation strategy. We currently detect this kind of failure by checking that the final position of the vehicle presents a value of $|\Delta y|$ smaller than 0.15 m and an orientation error smaller than 20 degrees. These values are arbitrary and adapted empirically to our system. Fig. 8 gives an example of an approach trajectory captured during 5 hours of continuous operations[6]. During this experiment, 100% of the pickups were successful.

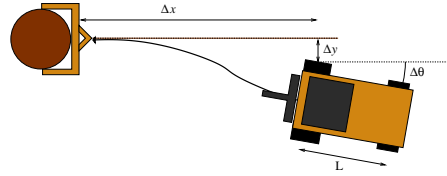


Fig. 7. Crucible approach trajectories and variables appearing in the control law



Fig. 8. Example of a real approach trajectory

The reliability and predictability of the HMC autonomous manoeuvre is critical to its acceptance by the light metal industry. As an illustration of the repeatability of our current implementation, figure 9 shows the superimposition of half of the 58 pickups performed during our 5 hour run. It can be clearly seen that the paths are well contained in an envelope whose width is correlated with the variation of crucible position (dots and ellipse around crucible handle).

6 Conclusions

The automation of Hot Metal Carriers, or general forklift, requires the implementation of robust handling capabilities. In this paper, we presented how



Fig. 9. Pickup approach trajectories, superimposition of 29 pickups. Dots and ellipse around the crucible handle show the variability of the crucible location across this set of manoeuvres. The left figure shows an overview of the stage, right figure shows a close-up on the paths themselves. For reference, the handle width is about 2 m and the pickup hole approximately 20 cm across.

we used especially designed artificial fiducials to implement a robust payload detection and localisation using pan-tilt-zoom cameras. In an ideal world a fiducial-less system will be more desirable, but within the current state of the art, this is unlikely to provide a robust enough solution for a real industrial deployment. Nevertheless, we are trying to develop such a system to evaluate how robust it can be, in comparison with our current one. Also, in this article, we did not address the long range detection of the payload, when a single view is not enough to detect the fiducials and a scanning strategy is required. We are working on this aspect of the problem and hope to present results soon.

References

1. A.J. Briggs, D. Scharstein, D. Braziunas, C. Dima, and P. Wall. Mobile robot navigation using self-similar landmarks. In *proc. of IEEE Conf. on Robotics and Automation*, 2000.
2. Arnaud Doucet, Nando De Freitas, and Neil Gordon, editors. *Sequential Monte Carlo methods in practice*. 2001.
3. G. Garibotto, S. Masciangelo, P. Bassino, C. Coelho, A. Pavan, M. Marson, and G. Elsag Bailey. Industrial exploitation of computer vision in logistic automation: autonomous control of an intelligent forklift truck. In *proc. of IEEE Conf. on Robotics and Automation*, 1998.
4. A. Negre, C. Pradalier, and M. Dunbabin. Robust vision-based underwater target identification & docking using self-similar landmarks. In *submitted to Conf. on Field and Service Robotics*, 2007.
5. M. Seelinger and J.-D. Yoder. Automatic pallet engagement by a vision guided forklift. In *proc. of IEEE Conf. on Robotics and Automation*, 2005.
6. A Tews, C. Pradalier, and J. Roberts. Autonomous hot metal carrier. In *proc. of IEEE Conf. on Robotics and Automation*, 2007.