# Model-based load localisation for an autonomous Hot Metal Carrier

Jesus Nuevo
Department of Electronics
University of Alcala, Spain
jnuevo@depeca.uah.es

Cédric Pradalier
Autonomous Systems Laboratory
CSIRO ICT Centre
PO Box 883 Kenmore, QLD 4069, Australia
cedric.pradalier@csiro.au

Luis M. Bergasa
Department of Electronics
University of Alcala, Spain
bergasa@depeca.uah.es

***Abstract*— Hot Metal Carriers (HMCs) are large forklift-type vehicles used to move molten metal in aluminium smelters. The molten metal is contained in bucket-like crucibles, that the HMC picks up. In this paper we explore the feasibility of using active appearance models to recognise and localise the handle of crucible, from a camera on board an autonomous HMC. A two-dimensional model is built that mimics the apparent perspective deformations of the three-dimensional handle. The model is fitted to the handle using efficient algorithms, that include M-estimators for improved robustness. The fitting algorithm also provides and estimate of the actual distance to the crucible. We evaluate the accuracy and robustness of the approach in different lighting conditons.**

## I. Introduction

Hot Metal Carriers (HMC) are massive forklift-like vehicle operating in aluminium smelters. Their purposes and main task is to transport molten metal, in bucket-like crucibles, from the pots where alumina is smelted to the cast-house where the hot metal is cast into ingots. The industry is considering HMC automation both for safety reasons and to improve efficiency through reliabiliy and repeatibility.

In this context, our team is working on the complete automation of HMCs [1]. See figure 1. This paper will focus on one specific aspect of this process: the autonomous pickup of the crucibles. As will be developped later, various techniques have been proposed to tackle detection and handling of a load by a forklift vehicle. Most of them use laser or indoor vision. In this work, we aim at developing a robust vision system, able to work outdoor in various lighting conditions, based only a model of the crucible and no additional markers.

This objective raises several challenges: first we need to be able to identify the crucible from significantly different view points; then we must be able to deal with strong variation in lighing such as the difference between a bright tropical sunshine and the interior of an industrial shed. Also, to implement successful vision-based pickups (more than 99.9% success), we require the localisation and orientation of the crucible with respect to the HMC to be estimated with more than $5cm$ or $5°$ accuracy.

In this article we will investigate the suitability of Active Appearance Models approach to solve the problem at hand. This will include details on how to improve the robustness of the approach and quantitative results.

Fig. 1. Autonomous Hot Metal Carrier and crucible

The rest of the paper is structured as follows. In section II we present the work related to industrial vehicle automation and previous works on active appearance models. Section III describes the model fitting algorithm and how the distance to the crucible is estimated. We then present our test setup and the results. Finally, section V concludes with a brief discussion of the results and outlines our future work.

## II. Background

Extensive research efforts have been made in the automation of vehicles for cargo transport in industrial environments. In [2], Mora *et al.* present a complete system for controlling forklifts in a warehouse. The vehicles were automated, and commanded from a centralised controller that managed all factory process. Garibotto *et al.* [3], [4] have presented ROBOLIFT, an autonomous forklift that had computer vision as its main sensor. In [5], a humanoid robot was developed to drive a forklift. In [6], [7], Roberts *et al.* presented an autonomous underground mining vehicle that navigated using scanning lasers.

Along with navigation, load detection and handling is the other main task for these systems. It usually relies on the use of fiducials [3], [8]. Nygards *et al.* [9] presented a system that used the image of a visible laser to locate and dock to a pallet.

In our autonomous carrier, crucible handling relied in detecting and ranging of fiducials attached to the handle of the crucible [1]. This system has proved to be very reliable, with a success rate of 100% in multi-hour tests. However, using fiducials may not always be possible. In a environment such an aluminium smelter, fiducials may suffer from fast material degradation, dirt accumulation and other circustances that degrade the performance of the detection algorithm. We try to address these possible problems by exploring alterate methods that can work without any artificial mark on the crucible.

Methods following the paradigm of "analysis through synthesis" have recieved considerable attention over the past few years. These methods try to parameterize the contents of an image by generating a synthetic image as close as possible to the one given, or a region of it. Active Appearance Models (AAM) [10] are among the most widely used methods. They have been successfully applied to face modelling and tracking, and in medical imaging [11]. AAMs have proved to be robust enough to be used in driver monitoring [12], [13].

Active Appearance Models are able to model changes in both appearance (or texture) and shape in an object, and thus can model non-rigid objects. While the handle of the crucible is rigid, its projection on the camera sensor depends on the point of view and the camera parameters, and makes it appear as a non-rigid object. A similar problem was explored in [14]. In that work, face models include some deformations in 3D, using orthographic projection. Our problem is in that sense a little simpler, but the robustness we need to achieve is much higher (99.9%).

## III. Recognition and localisation

In this section, a brief description of the model features and fitting algorithm are presented. We refer to the literature for further details [10], [15]. We then introduce some extensions that allow us to estimate the position of the crucible with respect to the HMC.

### A. 2D texture and shape models

An active appearance model (AAM) contains two vector bases that generate the spaces of valid shapes and appearances. The shape is defined by the coordinates of its $n$ points

$$\mathbf{s} = (x_0, y_0, x_1, y_1, x_2, y_2, \cdots, x_{n-1}, y_{n-1})^t \qquad (1)$$

Any valid shape is a linear combination of the *shape vectors* and the *base shape*. Without loss of generality, the vectors are assumed to be orthonormal.

$$\mathbf{s} = \mathbf{s_0} + \sum_{i=1}^{n} p_i \cdot \mathbf{s_i} \qquad (2)$$

The shape points are commonly triangulated in a meshed surface, using an algorithm such as Delaunay's [16]. See figure 2 for an example. The values of the pixels that fall in the mesh triangles represent the appearance of the AAM. Let $\mathbf{x} = (x, y)^t$ also be the pixels inside the triangles, an apperance is generated as a linear combination of the *appearance vectors*, $A_i(\mathbf{x})$, and the *base appearance*

$$A(\mathbf{x}) = A_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i \cdot A_i(\mathbf{x}) \qquad (3)$$

As with the shapes, the *appearance vectors* are supposed to be orthonormal.

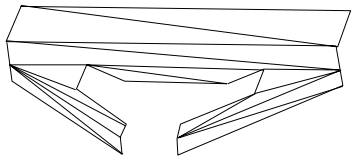


Fig. 2. A triangulated shape

Although there may be some correlation between the appearance and shape bases, they can be safely considered independent. Some AAM implementations combine both shape and appearance parameters in a single set [10].

### B. Model building

Model building is difficult, as it istherefore usually done by hand. The process involves manual anotation of the shape's points in a number of images that can be very high, resulting in a time-consuming task.

Once the training set is ready, marked shapes are normalized and aligned using the Procrustes algorithm [17]. The mean of the aligned shapes is chosen to be the *base shape*, $\mathbf{s}_0$. Principal Component Analysis (PCA) [18] is performed on the aligned shapes. The appearance base is build by warping the textures inside the marked shapes to a conveniently scaled version of $\mathbf{s}_0$, and then performing PCA over them.

In our case, model building is simpler as we deal with a rigid object. To generate the shape training set, a CAD

model of the handle was constructed, and 3D to 2D projections of it were generated for a extensive range of camera positions and orientations. The appearance training set was built from a few manually marked images, as the texture of the handle does not change significantly with the point of view.

### C. 2D model fitting

Model fitting is the process of obtaining the parameters that minimize the distance between the model and a new image $I$, i.e. the error image

$$\min \sum_{\mathbf{x}\in\mathbf{s}_0} g(Err(x)^2) \tag{4}$$

with

$$Err(x) = A_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i A_i(\mathbf{x}) - I(\mathbf{W}(\mathbf{x};\mathbf{p})) \tag{5}$$

where $g$ is a positive function, usually the $L_2$ norm, and $W(\mathbf{x};\mathbf{p})$ is the warp defined between the triangles of $\mathbf{s}$ and those of $\mathbf{s}_0$.

We use the approach of Baker and Matthews [19] to minimization, and their *inverse compositional algorithm* (IC) for its computational efficiency, that allows the algorithm to run in real-time. If, for simplicity, we only consider the shape vectors, the algorithm iteratively minimizes

$$g(Err(x)^2) = g((A_0(\mathbf{W}(\mathbf{x};\Delta\mathbf{p})) - I(\mathbf{W}(\mathbf{x};\mathbf{p})))^2) \tag{6}$$

In the $L_2$ case, using Gauss-Newton, the update $\Delta\mathbf{p}$ is obtained as

$$\Delta\mathbf{p} = \mathbf{H}^{-1} \sum_{\mathbf{x}\in\mathbf{s}_0} \mathbf{SD} Err(\mathbf{x}) \tag{7}$$

with

$$\mathbf{SD}(\mathbf{x}) = \left[\nabla A_0(\mathbf{x}) \frac{\partial \mathbf{W}}{\partial \mathbf{p}}\right], \quad \mathbf{H} = \sum_{\mathbf{x}\in\mathbf{s}_0} \mathbf{SD}^T\mathbf{SD} \tag{8}$$

The hessian $\mathbf{H}$ is constant as it only depends on the image $A_0$, greatly reducing the computational cost of each iteration. The new set of parameters $\mathbf{p}$ is obtained by composing the deformations

$$\mathbf{W}(\mathbf{x};\mathbf{p}) \leftarrow \mathbf{W}(\mathbf{x};\mathbf{p}) \circ \mathbf{W}(\mathbf{x};\Delta\mathbf{p})^{-1} \tag{9}$$

In [19], the authors show that, to a first order approximation,

$$\mathbf{W}(\mathbf{x};\Delta\mathbf{p})^{-1} \equiv \mathbf{W}(\mathbf{x};-\Delta\mathbf{p}) \tag{10}$$

This approximation is important, because the set of piecewise affine warps that AAM normally use doesn't form a semi-group, and thus $W(x,p)^{-1}$ may not exist. See also [20] for details.

### D. 2D model fitting using robust statistics

Using the $L_2$ norm as function $g$ in equation 4 allows many well-known methods to be used. But, in turn, the algorithm sensitivity to noise and outliers (spurious points or areas that result in high error values) becomes very high. As robustness is critical for our application, we use M-estimators to reduce outlier influence [21], [22]. In this we follow the modified fitting algorithm proposed by Gross in [23].

We tested several estimators, choosing Huber function as it produced the best results [24]. Huber function is defined as

$$\rho(r) = \begin{cases} r^2/2 & \text{if} \quad |r| \le \sigma \\ \sigma(|r| - \sigma/2) & \text{if} \quad |r| > \sigma \end{cases} \tag{11}$$

The derivate of the Huber function (its *influence function*) $\psi$ and its weight function $w$ are

$$\psi(r) = \frac{\partial\rho(r)}{\partial r} = \begin{cases} r & \text{if} \quad |r| \le \sigma \\ \sigma(sign(r)) & \text{if} \quad |r| > \sigma \end{cases} \tag{12}$$

and

$$w(r) = \frac{\psi(r)}{r} = \begin{cases} 1 & \text{if} \quad |r| \le \sigma \\ \sigma(|r|) & \text{if} \quad |r| > \sigma \end{cases} \tag{13}$$

We estimate the value of the scale parameter $\sigma$ as a function of the median of the values of $r$, that in our case is the square error image in equation 6.

Unlike in the $L_2$ case above, when using an M-estimator the hessian depends on the error image, and has to be recomputed at each iteration. The hessian now is

$$\mathbf{H} = \sum_{\mathbf{x}\in\mathbf{s}_0} w(Err(x)^2)\mathbf{SD}^T\mathbf{SD} \tag{14}$$

The equation shows that $w$ is a ponderation of the values of the hessian. It is a safe assumption that outliers appear with some degree of locality. Thus, we can consider $w(Err(x)^2)$ constant in an area. As it is the only element that changes and the hessian can be computed localy, this property can be used to efficiently calculate an approximation to the hessian. A simple choice is to take the triangles of the mesh as subdivisions of $\mathbf{H}$. If

$$\mathbf{H}_i = \sum_{\mathbf{x}\in\mathbf{T}_i} \mathbf{SD}^T\mathbf{SD} \tag{15}$$

is the hessian of pixels on the $i$-th triangle, the hessian can be approximated as a weighted sum

$$\mathbf{H} \equiv \sum_{i=1}^{k} w_i \mathbf{H}_i \tag{16}$$

where $w_i$ is an estimate (usually the mean or median) of the value of $w(Err(x)^2)$ in the $i$-th triangle.

The model we use in our system has both appearance and shape vector bases. In this case, we calculate the

update to both parameter sets alternatively. In each iteration, we obtain the appearance parameters update, $\Delta\lambda$ in a similar fashion of that described above, with the exception that the update is added to the previous value, instead of composed.

$$\lambda \leftarrow \lambda + \Delta\lambda \tag{17}$$

*E. Model fitting with 3D restrictions*

The projections of the 3D CAD model of the crucible produce some *apparent* deformations in 2D that the shape vectors mimic. These deformations have some constraints, imposed by the orthogonality of the projection matrix, that are lost because the fitting process does not impose any restrictions on the values of the parameter set.

A solution to this is to use priors in the fitting algorithm [14], to force the combination of the 2D shape vectors to be an (approximately) valid projection of the 3D model. The function to minimize is now

$$\min \sum_{\mathbf{x}\in\mathbf{s}_0} g(Err(x)^2) + K\|\mathbf{P}\mathbf{s}^{3D} - (\mathbf{s}_0 + \sum_{i=1}^{m} p_i\mathbf{s}_i)\|^2 \tag{18}$$

where $K$ is a constant, $\mathbf{s}^{3D}$ is the 3D model and $\mathbf{P}$ is the projection matrix, that has to be estimated in the minimization process.

The use of this prior improves the fitting by constraining the values of the parameter set $\mathbf{p}$. It also provides a mean for estimating the projection matrix $\mathbf{P}$. This is an interesting point, as obtaining the distante between the HMC and the crucible is useful for the navigation system of the vehicle. We briefly outline the algorithm here, and refer to [25], [26] for details.

Let $\mathbf{s}$ be a model shape in 2D, and $\mathbf{x}_i = (x_i, y_i)^t$ one of its points, and $\mathbf{X}_i = (X_i, Y_i, Z_i)^t$ its corresponding point in the 3D model. We can define two orthogonal planes passing through the origin $o$ and $p_i$ by their normals

$$N_i^1 = \frac{o\vec{p}_i \wedge \vec{v}}{\|o\vec{p}_i \wedge \vec{v}\|} \qquad N_i^2 = \frac{o\vec{p}_i \wedge \vec{v}}{\|o\vec{p}_i \wedge \vec{v}\|} \tag{19}$$

were $u$ and $v$ are the reference system horizontal and vertical unit vectors. Let $\mathbf{X}'_i$ be the transformation of $\mathbf{X}_i$ for the current set of parameters $\mathbf{A} = (\alpha, \beta, \gamma, t_x, t_y, t_z)^t$

$$\mathbf{X}'_i = \mathbf{R}_{\alpha\beta\gamma}\mathbf{X}_i + \mathbf{T}_{xyz} \tag{20}$$

The aim of the estimation process is to obtain the parameters $\mathbf{A}$ that align $o$, $p_i$ and $\mathbf{X}'_i$. This is equivalent as to null the projection of $\mathbf{X}'_i$ on the normals defined above, i.e.

$$\begin{aligned} N_i^1\mathbf{X}'_i &= N_i^1[\mathbf{R}_{\alpha\beta\gamma}\mathbf{X}_i + \mathbf{T}_{xyz}] = 0 \\ N_i^2\mathbf{X}'_i &= N_i^2[\mathbf{R}_{\alpha\beta\gamma}\mathbf{X}_i + \mathbf{T}_{xyz}] = 0 \end{aligned} \tag{21}$$

To solve the system, standard methods such as Gauss-Newton or Levenberg-Marquandt can be used. Although the jacobian and hessian have to be recomputed in each iteration, the matrices involved are small, so the computational load is not remarkable. Each iteration of the algorithm is performed in parallel to the 2D model fitting.

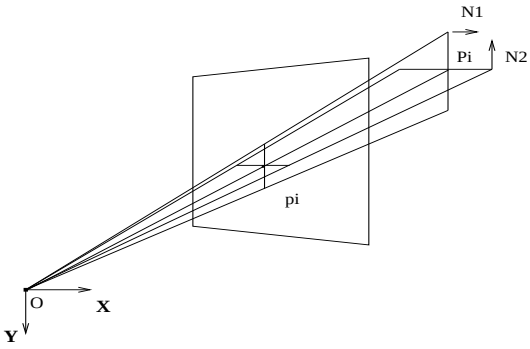



Fig. 3. Estimation of projection matrix

## IV. Results

In this section, we present the results of the different parts of our system. The image sequences used were recorded with the HMC moving automously, driven by the fiducial-based recognition system. Several sequences were recorded in bright and cloudy days.

One common problem to any non linear minimization algorithm is that it requires a starting point close enough to the minimun so it does not diverge.

To initialize the 2D fitting algorithm, we use the MeanShift method [27] to obtain an estimate of the position of the handle of the crucible. See figure 4. The square with thick outline represents the best estimate.

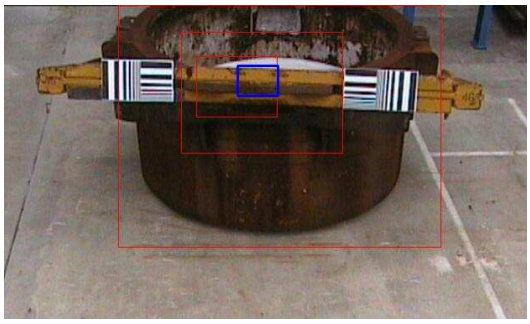
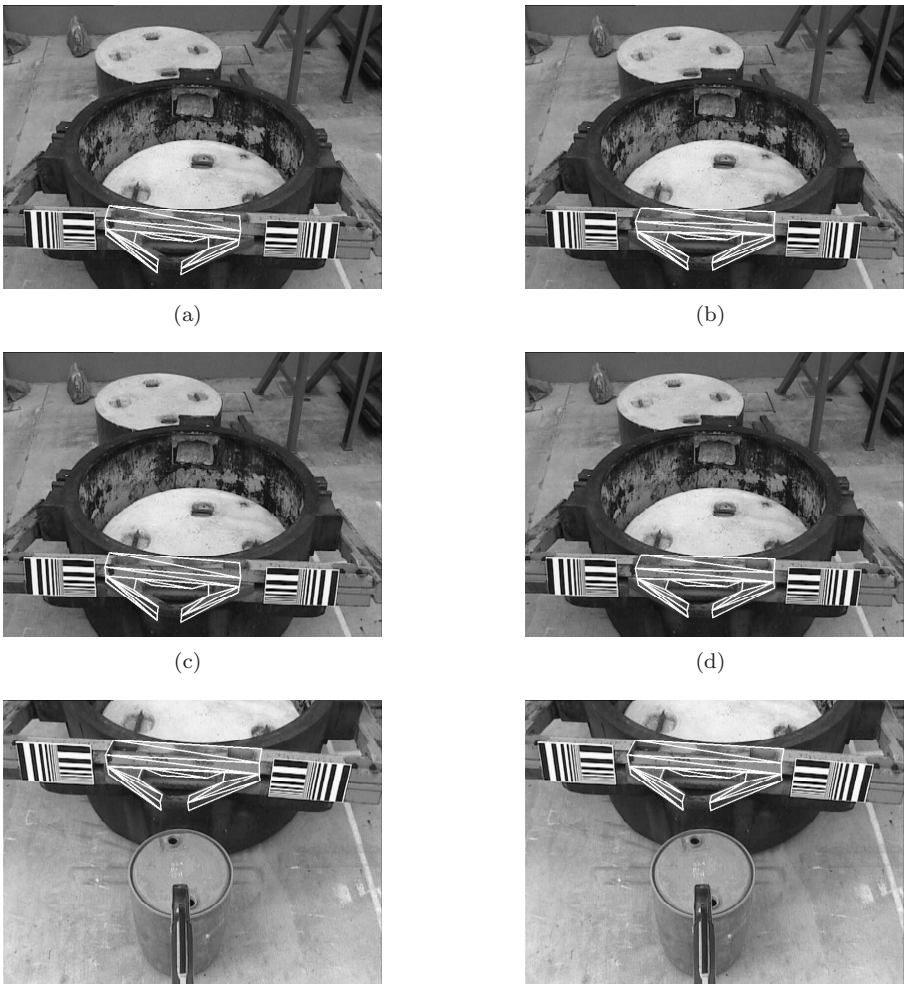


Fig. 4. Initialization of the AAM parameters

To improve the chances of having a successful fitting, several models with different positions and sizes are iterated around the estimate given by MeanShift. Most of them are discarded after very few iterations, so the computational cost of the process is small. We observed that, for a model of 150x50 pixels of size the range of convergence is around 20 pixels of distance. An initial value for the rotation angle within $\pm 10°$ of the actual angle

The initial values of the projection matrix $\mathbf{P}$ are calculated with Dementhon's method [28].



(a) (b) (c) (d) (e) (f)

Fig. 5. Models fitted to different views of the crucible

Figure 5 shows different views of the crucible where the model has been fitted. As can be seen, we use two different models, one when the HMC is far away from the crucible and another one for close-ups. The second model only covers the hookeye and the area of the handle around it, and its vectors deform to conform to the stronger perspective. The transition is simple as both models share many common points.

While the test whose results are shown in figure 5 were run imposing the 3D restrictions, the improvements over the results of the pure 2D model were very small. Also, the estimate of the matrix $\mathbf{P}$, although good, was not as accurate as expected.We are actively revising our implementation to improve it, and solve any weaknesses that may be present.

## V. Conclusion

We have presented a system for recognition and localisation of a crucible based on 2D appearance models, eliminating the need for fiducials to be placed on the crucible. The system models the handle and hookeye of the crucible, and uses M-estimators to improve the robustness of the fitting algorithm. The system also imposes restrictions on the model instantiation, forcing it to be a valid projection of a 3D model.

The next steps on this project will include improving how the 3D restrictions are imposed, and the accuracy of the estimation of the projection matrix. Also, different representation of the appearance will be tested, moving from an RGB space to others such as HS or HS+edge map. The model for the hookeye section will be extended to include finer details. We also plan to develop measurements that would help to evaluate the correctness of the alignment between the HMC and the crucible.

## Acknowledgment

This work was funded in part by CSIRO's Light Metals Flagship project and by the CSIRO ICT Centre's ROVER and Dependable Field Robotics projects. The authors would like to acknowledge the contribution of the Autonomous Systems Lab's team, and in particular to Jonathan Roberts. J. Nuevo is also working under a researcher training grant from the Education Department

of the Comunidad de Madrid and the European Social Fund.

## References


[1] A. Tews, C. Pradalier, and J. Roberts, "Autonomous hot metal carrier," in *Proceedings of IEEE Int. Conf. on Robotics and Automation*, 2007.

[2] M. Mora, V. Suesta, L. Armesto, and J. Tornero, "Factory management and transport automation factory management and transport automation," in *Emerging Technologies and Factory Automation*, September 2003, pp. 508–515.

[3] G. Garibotto, S. Masciangelo, M. Ilic, and P. Basino, "Robolift: a vision guided autonomous fork-lift for pallet handling," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, November 1996, pp. 656–663.

[4] 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 *International Conference on Robotics and Automation*, vol. 2, May 1998, pp. 1459–1464.

[5] H. Hasunuma, K. Nakashima, M. Kobayashi, F. Mifune, Y. Yanagihara, T. Ueno, K. Ohya, and K. Yokoi, "A teleoperated humanoid robot drives a backhoe," in *International Conference on Robotics and Automation*, Taipei, Taiwan, September 2003, pp. 2998–3004.

[6] J. Roberts, E. Duff, P. Corke, P. Sikka, G. Winstanley, and J. Cunningham, "Autonomous control of underground mining vehicles using reactive navigation," in *Proceedings of IEEE Int. Conf. on Robotics and Automation*, San Francisco, USA, 2000, pp. 3790–3795.

[7] J. Roberts, E. Duff, and P. Corke, "Reactive navigation and opportunistic localization for autonomous underground mining vehicles," *The International Journal of Information Sciences*, vol. 145, pp. 127–146, 2002.

[8] M. Seelinger and J.-D. Yoder, "Automatic pallet engagment by a vision guided forklift," in *International Conference on Robotics and Automation*, 2005.

[9] J. Nygards, T. Hogstrom, and A. Wernersson, "Docking to pallets with feedback from a sheet-of-light range camera," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 3, October 2000, pp. 1853–1859.

[10] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 23, pp. 681–685, Jan. 2001.

[11] M. Stegmann and D. Pedersen, "Bi-temporal 3 D active appearance models with applications to unsupervised ejection fraction estimation," *Proc. SPIE*, vol. 5747, pp. 336–350, 2005.

[12] S. Baker, I. Matthews, J. Xiao, R. Gross, T. Kanade, and T. Ishikawa, "Real-time non-rigid driver head tracking for driver mental state estimation," in *11th World Congress on Intelligent Transportation Systems*, October 2004.

[13] J. Nuevo, L. Bergasa, M. Sotelo, and M. Ocana, "Real-time robust face tracking for driver monitoring," in *Intelligent Transportation Systems Conference, 2006. ITSC '06. IEEE*, 2006, pp. 1346–1351.

[14] J. Xiao, S. Baker, I. Matthews, and T. Kanade, "Real-time combined 2D+ 3D active appearance models," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 535–542, 2004.

[15] I. Matthews and S. Baker, "Active appearance models revisited," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 135–164, November 2004.

[16] B. Delaunay, "Sur la sphere vide," *Izv. Akad. Nauk SSSR, Otdelenie Matematicheskii i Estestvennyka Nauk*, vol. 7, pp. 793–800, 1934.

[17] J. Gower, "Generalized procrustes analysis," *Psychometrika*, vol. 40, no. 1, pp. 33–51, 1975.

[18] I. Jolliffe, *Principal Component Analysis*. Springer, 2002.

[19] 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, March 2004.

[20] ——, "Equivalence and efficiency of image alignment algorithms," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 1090–1097, 2001.

[21] P. J. Huber, *Robust Statistics*, ser. Wiley Series in Probability and Mathematical Statistics. Wiley-Interscience, 1981.

[22] V. Barnett, *Outliers in Statistical Data*, ser. Wiley series in probability and mathematical statistics. John Wiley & Sons, 1978.

[23] R. Gross, I. Matthews, and S. Baker, "Constructing and fitting active appearance models with occlusion," in *Proceedings of the IEEE Workshop on Face Processing in Video*, June 2004.

[24] Z. Zhang, "Parameter estimation techniques: A tutorial with application to conic fitting," *Image and Vision Computing Journal*, vol. 15, no. 1, pp. 59–76, 1997.

[25] D. Lowe, *Perceptual Organization and Visual Recognition*. Kluwer Academic Publishers Norwell, MA, USA, 1985.

[26] M. Dhome, M. Richetin, J. Lapreste, and G. Rives, "Determination of the attitude of 3D objects from a single perspective view," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 12, pp. 1265–1278, 1989.

[27] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 5, pp. 603–619, 2002.

[28] D. Dementhon and L. Davis, "Model-based object pose in 25 lines of code," *International Journal of Computer Vision*, vol. 15, no. 1, pp. 123–141, 1995.