2014 IEEE International Conference on Robotics & Automation (ICRA)
Hong Kong Convention and Exhibition Center
May 31 - June 7, 2014. Hong Kong, China

New Marker for Real-Time Industrial Robot Programming by Motion Imitation*

Marcos Ferreira[1],Paulo Costa[1], Luís Rocha[1], A. Paulo Moreira[1], Norberto Pires[2]

***Abstract*— This paper presents a new marker for robot programming by demonstration through motion imitation. The device is based on high intensity LEDs (light emission diodes) which are captured by a pair of industrial cameras. Using stereoscopy, the marker supplies 6-DoF (degrees of freedom) human wrist tracking with both position and orientation data. We propose a robust technique for camera and stereo calibration which maps camera coordinates directly into the desired robot frame, using a single LED. The calibration and tracking procedures are thoroughly described. The tests show that the marker presents a new robust, accurate and intuitive method for industrial robot programming. The system is able to perform in real-time and requires only a single pair of industrial cameras though more can be used for improved effectiveness and accuracy.**

I. INTRODUCTION

Industrial manipulators are the ultimate automation tool. These machines deliver accuracy and repeatability aiding industrial processes at becoming increasingly efficient while reducing production costs. Yet, programming industrial manipulators is still extremely time consuming and usually require experienced and highly qualified workers. Overall this is not compatible with flexible setups neither with small companies budgets since both qualified programmers and reconfigurations downtime imply strong financial efforts.

Even though the former is a quite restrictive scenario, manipulators are still strongly desired at production lines due to a series of advantages over human work, e.g., the ability to work continuously while preserving output quality, immunity to fatigue, distractions and hazardous environments.

*This work was financed by the ERDF-European Regional Development Fund through the COMPETE Programme, and by National Funds through the FCT-Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) within project ńFCOMP-01-0124-FEDER-022701ż This work has also been supported by FCT through the project PTDC/EME-CRO/114595/2009 "High-Level programming for industrial robotic cells: capturing human body motion". This work is also part of the project "NORTE-07-0124-FEDER-000060" which is financed by the North Portugal Regional Operational Programme (ON.2 - O Novo Norte), under the National Strategic Reference Framework (NSRF), through the ERDF fund, and by national funds, through FCT.

[1]M. Ferreira et al. are with the Dept. of Electrical and Computer Eng., Faculty of Eng. – Univ. of Porto, and with INESC TEC – Institute for Systems and Computer Engineering of Porto, Portugal {marcos.ferreira,paco,lfr,amoreira}@fe.up.pt

[2]N. Pires is with the Dept. of Mechanical Eng., Faculty of Science and Technology, Univ. of Coimbra, Portugal jnp@robotics.dem.uc.pt

A. Proposed Solution and Aims

To overcome this reality, this work presents a methodology for fast industrial robot programming via human demonstration by gesture. The main goal is to achieve a new type of marker that enables the human to quickly show the robot how to do a concrete task with abstraction of the programming language and even completely avoiding the use of the teach pendant.

The focus of this paper is to describe a new 6-DoF marker that is captured by a pair of industrial cameras. The artificial vision system captures images can accurately tell us information of position and orientation. The resulting set of points that describe the human path are automatically transformed into a robot program. The human uses his natural abilities and skills to accomplish the demonstration process without needing further instruction on using new software packages, interfaces or tools. Unlike similar researches, the proposed marker is able to perform in real time and the required apparatus is reduced and cheap.

B. Related Work

The universe of human-robot interaction is vast and it is hard to cover all the current streams of research in the area. With a shorter scope, on motion and task demonstrations and industrial robot interface, there are still a number of contributions to consider: CAD based programming, such as in [1] and [2], where the user interacts with a simulated environment in order to draw the robot paths; gesture and voice recognition, [3] and [4] , where the robot is pre-programmed and the gestures are used to deploy actions. In more similar approaches to the work of this contributions there is the vast work of R. Dillman, using stereo vision and a data glove [5] for tracking, and adding tactile sensors for grasp recognition [6] and [7]. Another similar approach has been presented by B. Hein, [8] and [9], where a new tracking marker is also proposed, based on infra-red LEDs that can be coupled to any work-tool. This paper focus on the development of a new marker that uses visible light LEDs; with a new method for synchronized acquisition, the visible markers can perform faster than the infra-reds (either passive or active) because ours are easily distinguished using colour. The proposed solution also proves to be accurate and, unlike most of the previous solutions, cheap as it is employed with a minimum of just 2 standard industrial cameras.

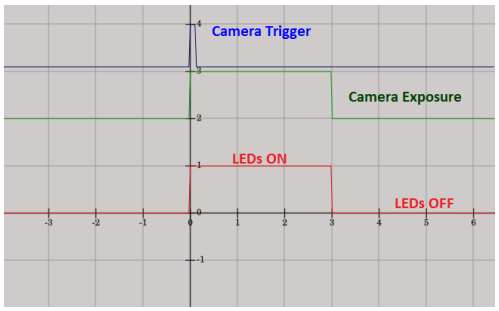


Fig. 1: **Sincrovision timings**: A timing diagram depicting the synchronous image acquisition. Both cameras are triggered at the same time. From that moment on they are acquiring the scene for a very limited time (3ms). The ultra bight LEDs are triggered at the same instant the cameras were triggered too and remain lit only during the camera exposure period.

II. Motion Imitation Framework

A. Stereoscopy using the sincrovision technique

The *sincrovision* concept was developed and patented [10] in the University of Porto - Faculty of Engineering. It implements a system of 3D acquisition based on stereoscopic vision synchronised with high intensity luminous markers. The key idea is to turn on the markers as soon as the cameras start acquiring image and turn them off after the camera exposure time has expired. Figure 1 shows a timing diagram of the system.

The high intensity lights will be very bright on the images whilst the background noisy data will have no time to be acquired by the camera. At the same time, blinking the markers for a short time makes it possible to stare at them; keeping them always on would cause eye damage. To ensure that unwanted light is not captured, the lenses aperture is reduced to minimal. This setup makes it possible to triangulate the markers positions in space in a robust way, independently of lighting conditions in the scene and ignoring most of the common noise sources in artificial vision applications. Figure 2 shows a typical image captured by the cameras using this synchronous feature: on the left, the scene is captured using standard camera aperture and exposition time; on the right, the *sincrovision* effect over the same scene. *Multiple View Geometry* [11] holds a complete guide for camera models, the stereoscopy principles and camera calibration, which are the necessary basis for retrieving 3D measures from a pair of images.

B. 6-DoF Marker

The major concerns for the development of this new tracking tool were to achieve a device that provides an accurate measure of the pose of the human hand/tool while maintaining costs low, reduced processing times and low impact on the process and on the human movements.

Using the *sincrovision* technique, this marker was built from a set of high intensity LEDs.

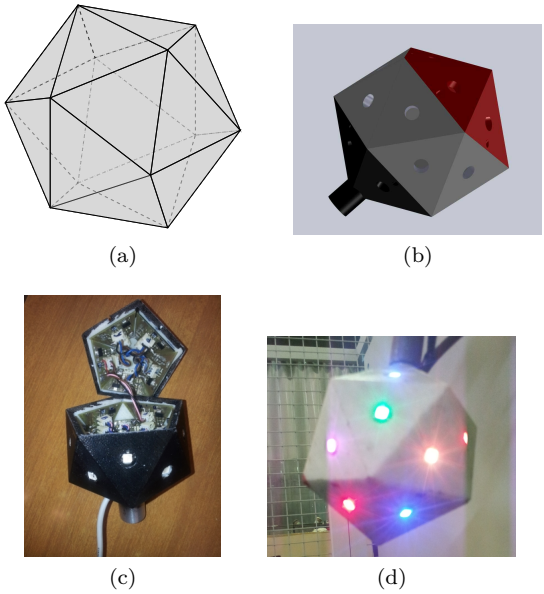
(a)

(b)

Fig. 2: (a) **Standard acquisition**: The scene is captured with no timings constraints. The exposure and lense aperture make the image too dirt to be processed. The background lighting obfuscates the marker. There are many noise sources, mainly the fluorescent lamp. (b) **Synchronous acquisition**: The lenses aperture are reduced to a minimal. Exposure is very low (around $3ms$). The background noise is eliminated and the markers shine robustly in the image. Under these conditions, image processing and clustering is straightforward.

While one LED is enough to track position (3D), in order to capture the other three degrees of freedom from orientation at least three non-collinear LEDs are needed. Nevertheless, such a scarce number of lights would fail to provide a complete freedom of movements to the end-user: all of those individual markers should have to visible at all times on both cameras otherwise pose estimation would fail due to occlusions. Increasing the number of cameras around the working area can fight back this problem but at a greater financial cost. On this line of reasoning, the proposed marker is based on 20 visible-light (RGB – red, green, blue) LEDs. These are distributed in a special manner, based on the shape of an icosahedron — see Fig. 3 — , as it showed to provide an interesting set of properties that aid in constructive and algorithmic aspects, described below.

1) Detection and Stereo Matching: Tracking the marker starts from a pair of synchronized images. The *sincrovision* technique (recall the section II-A and Fig. 2) provides very clean shots of the marker. The detection is accomplished by means of a global threshold which is very fast and still robust due to the acquisition effect. Each cluster (single LED) is then evaluated according to



(a) (b) (c) (d)

Fig. 3: (a) **Polyhedron 3D view**: A 3D representation of an icosahedron. (b) **CAD**: the CAD model of the designed marker. (c) **Construction Details**: details on the developed marker; cabling, electronics and the LEDs are visible. (d) **Lights ON**: real marker with LEDs in ON state.

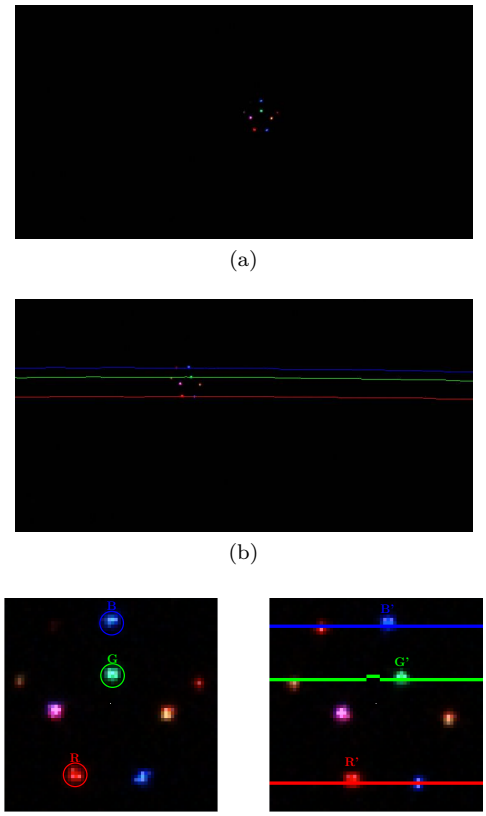
its colour.

Using the principles of stereoscopy, each LED turns into a 3D measure given that it is visible in both cameras. The stereo correspondence problem is easily solved — in real time — using the LED colour and the epipolar constraint as depicted in Fig. 4: once a cluster of a given colour is found in one image, the stereoscopy laws define a line (epipolar line) on the other image where the matching cluster must lie (provided the cameras have been properly calibrated). Searching along the epipolar allows finding a cluster of the same colour thus making possible to retrieve a 3D measure from the pair.

2) Position Estimation: One of the advantages of the icosahedron shaped marker relates to the position of the LEDs. Placing them of the centre of each face enables that there are always some visible LEDs on both cameras for whichever movement is made by the human holding the marker (unless he steps in-between the marker and the cameras). Also, when the LEDs are positioned on the faces, they lie on a sphere shell that touches every face centre. Due to this property, the 3-DoF related to the marker translation are computed taking advantage of this spherical positioning of the LEDs. Since the 3D coordinates of the matched clusters are already available, the world points are used to estimate the sphere shell in which they lie. From the sphere equation

$$(x_i - x_c)^2 + (y_i - y_c)^2 + (z_i - z_c)^2 = r^2 \qquad (1)$$

we use algebraic fitting [12] which solves the sphere fitting problem using least squares (the alternative method



(a) (b) (c)

Fig. 4: (a) **Camera 1 Image with clusters**: an image from camera 1 with a set of clusters. (b) **Camera 2 Image with Epipolar Lines**: the red, green and blue lines (not straight due to barrel distortion) are epipolar lines from some clusters in image (a). (c) **Zoom on Epipolar Lines and Cluster Correspondence**: the highlighted and labelled red, green and blue clusters — on the left, zoomed from (a) — need a matching pair in the other camera image so that $2D \leftrightarrow 3D$ is possible. Finding the correspondence is achieve by colour and the geometric constraint given by the epipolar lines — on the right, zoomed from (b).

is the geometric fitting, which requires an iterative minimization, [13]):

$$\kappa = {x_c}^2 + {y_c}^2 + {z_c}^2 - r^2 \qquad (2)$$

Using Eq. (2) in (1) holds:

$$\underbrace{\begin{bmatrix} 1 & 2x_1 & 2y_1 & 2z_1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 2x_n & 2y_n & 2z_n \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} -\kappa \\ x_c \\ y_c \\ z_c \end{bmatrix}}_{\theta} = \underbrace{\begin{bmatrix} x_1^2 + y_1^2 + z_1^2 \\ \vdots \\ x_n^2 + y_n^2 + z_n^2 \end{bmatrix}}_{\mathbf{b}} \qquad (3)$$

Fig. 5 shows a 3D scene with the luminous markers and the sphere shell in which they lie at.

The advantage of the fast least squares solution is that it can be used a number of times for the same image pairs

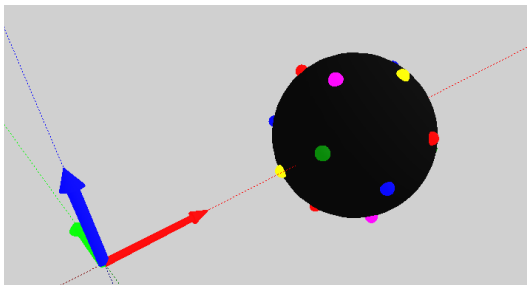


Fig. 5: **3D Scene with LEDs and a Sphere**: To estimate the whole marker position each individual LEDs contributes for a sphere fitting algorithm.

to eliminate stereo ambiguities without compromising the real time capability or minimization divergences: if some points are *phantom*, i.e., result from a bad stereo matching, using the algebraic fitting recursively while dropping some points ultimately comes up with the correct measure. Knowing when an estimate is good or not is done through the radius estimation: the marker radius is a well known constructive measure and the bad stereo matches produce a radius estimate different from the known value. When the radius approximates the real value, the estimate is considered a good measure and the marker position is known.

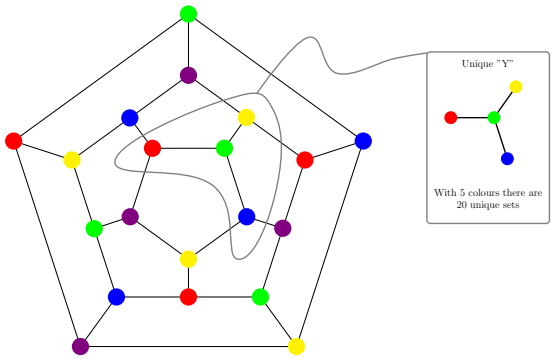
3) Orientation Estimation: After retrieving the translation vector, the missing 3-DoF with respect to angular displacement are computed from the known well-matched 3D points.

Using again the advantages of the icosahedron, it is possible to build a list of the LEDs' coordinates in the 3D world. When the icosahedron face centres are connected, we get the dual polyhedron, the dodecahedron. Each face centre of the icosahedron is a vertex of the dodecahedron. There is an orientation of the marker at which the LEDs lie at the given positions:

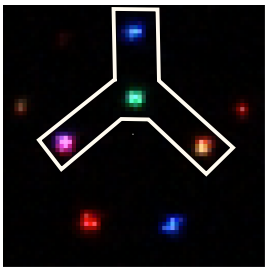
$$\begin{aligned} &s\,(\pm 1, \pm 1, \pm 1) \\ &s\,(0, \pm 1/\varphi, \pm\varphi) \\ &s\,(\pm 1/\varphi, \pm\varphi, 0) \\ &s\,(\pm\varphi, 0, \pm 1/\varphi) \end{aligned} \tag{4}$$

where s is a scale factor and $\varphi = \left(1+\sqrt{5}\right)/2$ is the so called golden ratio (if a and b are in golden ration, and $a > b$, then $\varphi = \frac{a+b}{a} = \frac{a}{b}$).

Moreover, the schlegel diagram of the dodecahedron (a planar representation of the 3D solid where the sides never cross) allows us to use a set of 5 different colours and distribute them around the marker in such a way that, given 4 visible LEDs, it is possible to instantaneously know the marker rotation. Fig. 6 (a) shows the coloured dodecahedron schlegel where a "Y" is highlighted; in (b), the actual detection of the "Y" in a captured image, enabling us to know which part of the marker is being seen. With 5 different colours there are 20 unique "Y"s, as many as there are in the dodecahedron.



(a)



(b)

Fig. 6: (a) **Coloured Dodecahedron Graph**: The schlegel diagram provides an intuitive way to view the distribution of LEDs around the dodecahedron. Using 5 colours allows a quick identification of the marker orientation as soon as a **Y** is completely visible in the images. (b) **Y Detection**: Detection of a complete **Y** in a real image capturing the luminous marker.

For the rotation estimation algorithm, the set of known 3D positions is designated by P, where P_i is a 3D vector $[x_i, y_i, z_i]^T$ with the coordinates of dodecahedron vertex i. P is considered the set of *stand-by* positions. From the detected "Y", it is possible to know which LED is which, for all visible LEDs in the images. After retrieving the 3D measures from the stereo analysis, these positions are stored in the matrix Q, the set of measured positions. The nth vector in Q stores the current world position of the nth stand-by LED in P — Q and P are said to be paired. To find the marker orientation, we find the rotation from points Q_i to points P_i.

The kabsch algorithm [14] provides a mean to solve this problem. This method finds the rotation matrix that optimally describes (in a root-mean-squared-error sense) the rotation from two paired 3D point lists:

1) P and Q must have origin-centred vectors so the first step is subtracting both sets their respective centroid.
2) Compute the covariance matrix A defined as: $A = P^T Q$
3) Compute the singular value decomposition of A:

$$A = USV^T$$

Fig. 7: (a) **Two Camera Arrangement for Stereoscopy**: The figure shows a cabinet that holds both cameras. These are separated by 700 mm and slightly rotated towards each other.

and the optimal rotation matrix W^* comes from :

$$d = sign(det(VU^T));$$

$$W^* = V \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & d \end{pmatrix} U^T \quad (5)$$

(The auxiliary parameter d is used to insure a right handed coordinated system.)

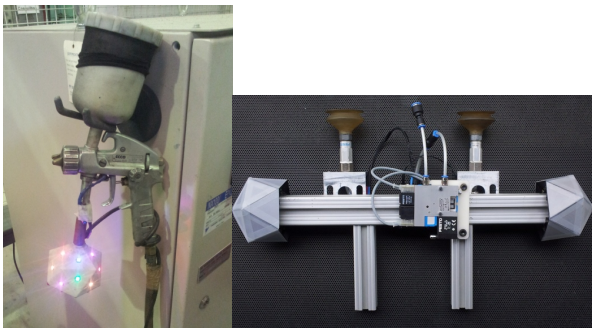
At this point we have the full (6-DoF) characterization of the movement of the marker, with both position (3-DoF) and orientation (3-DoF). Also note that using only 5 colours has an added advantage on the detection stage: ror instance, using HSV colour space analysis, the 5 colours can be chosen to have hue values of 0, 60, 120, 240 and 300 as it was used in the examples presented in this paper. This way colour misclassification is minimized.

III. TESTS & RESULTS

The system performance was studied using a single pair of industrial cameras (Fig. 7) and a marker with 50 mm radius.

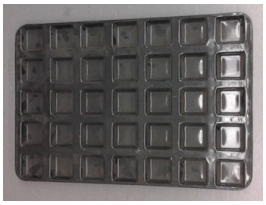
The cameras are both from *Imaging Source*, 1024×768 CCD sensor with USB 2.0 connectivity. The cameras and the marker are synchronously triggered using a microprocessor.

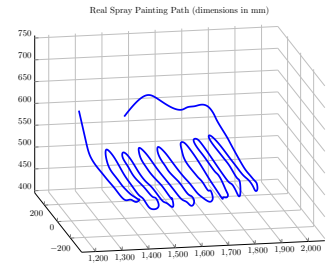
Table I shows how the pose estimation algorithms perform: the fail rate is below 0.2%. For this test the marker was attached to an industrial spray painting gun and also to an adapted tool with industrial suction cups for metal sheets bending operations. Fig. 8 shows both of these industrial tools: in (a) the marker is attacked to the spray painting gun so that it does not interfere with the painter movements; in (b), two markers are used in a single tool — this tool is used to manipulate metal sheets, and since large sheets can easily occlude one marker, two are used at each end of the tool to ensure a successful tracking during the whole demonstration. The workspace volume for testing was 1000×1000×400 mm.



(a) (b)

Fig. 8: Marker attached to an industrial spray painting gun and to industrial suction cups (one marker at each side).





(a) y (b)

Fig. 9: (a) Baking tray sample, used in our tests (b) spray coating tracked trajectory from the part shown in (a)

In order to find the marker precision in this workspace, the luminous device was attached to the industrial robot end-effector. Then, the robot was set to move in a 100 mm grid, i.e., moving to a point then stopping for measuring the marker pose, until the whole workspace was covered. At each stop, the coordinates of the position M_i and orientation (quaternion)q_{M_i} were read from the robot controller. The corresponding measured pose from the stereo is $\left(\tilde{M}, q_{\tilde{M}}\right)$. Table II summarizes the marker pose precision for the 2 camera arrangement. The maximum errors are found near the edges of the images (mostly due to barrel distortion); increasing the number of cameras and their resolution can effectively enhance these results.

Finally, the real time capability of this system is stated on Table III. The setup was used in an industrial environment, in spray coating processes. Fig. 9 (a&b) shows a sample tracked path using the proposed marker (b), captured during the spray coating of a baking tray, as shown in (a). The output was validated by the painters — check the motion mimic in the video attached to this contribution.

IV. CONCLUSIONS

A new marker was proposed, which enables an human operator to program an industrial manipulator by showing the required movements. Tests, in a real industrial scenario, show that the marker performs in real-time and

TABLE I: **Evaluation of the position and orientation estimation**: The success and failure rate for each algorithm is presented for a total of 50000 successfully paired image frames. The complete pose estimation performance is held at the last sub-table.

Stereo Frames: 50000 pairs

Position Estimation	
Success 50000 [100%]	Failure 0 [0%]
Orientation Estimation	
Success 49919 [$>$ 99.8%]	Failure 81 [$<$ 0.2%]
Complete Pose Estimation	
Success $>$ 99.8%	Failure $<$ 0.2%
average/per demonstration (1500 frames) $<$ 3 fails	

TABLE II: **Marker pose error**: Marker position error measured as the euclidean distance from the expected position M to the estimated $\tilde{M}$. For orientation, the absolute error is the shortest path (angle) between the two orientations which is computed from the angle between the two corresponding quaternions, q_M and $q_{\tilde{M}}$.

Marker Position Error: $E = \left\| M - \tilde{M} \right\|_2$, and Orientation Error: $\theta = 2\cos^{-1} q_M^{-1} q_{\tilde{M}}$ *
(all distances in mm, angles in degrees °)

	E	θ
Mean Error	3.8	1.7
Max Error	8.9	6.2
Std. Deviation	2.7	2.1

* possible implementation for the shortest angle between 2 quaternions

TABLE III: **Processing times for every stage of the motion demonstration**: The tests were run on a core i7 @2.8GHz , under a Fedora 16 installation.

	Routine	Time (ms)	
Online	Image Processing (Debayer and Clustering)	2.2	
	2-Camera Cluster Matching and 3D Retrieval	0.5	
	Pose Estimation	0.6	
Total			$<$ 6
Offline	Robot Program Generation	250	
	Communication and Upload	1000	
Total			$<$ 1300

has a low fail rate (below 0.2%). Furthermore, using only two cameras, a satisfactory precision was achieved for a spray painting application. The operators were able to program the robot with complete abstraction of the programming language and the robot was ready to perform as soon as the demonstration was over. The proposed system greatly reduces programming downtime, and it enables non-programmers to successfully develop robot programs without a single line of code.

ACKNOWLEDGMENT

Marcos Ferreira acknowledges FCT – Fundação para a Ciência e a Tecnologia for his PhD grant: SFRH/BD/60221/2009.

References

[1] H. C., W. S., N. Xi, M. Song, and Y. Chen, "Automated robot trajectory planning for spray painting of free-form surfaces in automotive manufacturing," in *Robotics and Automation, 2002. Proceedings. ICRA '02. IEEE International Conference on*, vol. 1, 2002, pp. 450–455 vol.1.

[2] P. Neto, N. Mendes, R. Arajo, J. N. Pires, and A. P. Moreira, "High-level robot programming based on cad: Dealing with unpredictable environments," *Industrial Robot*, vol. 39, no. 3, pp. 294–303, 2012.

[3] P. Kumar, J. Verma, and S. Prasad, "Hand data glove: A wearable real-time device for human-computer interaction," *Hand*, vol. 43, 2012.

[4] J. N. Pires, "Robot-by-voice: Experiments on commanding an industrial robot using the human voice," *Industrial Robot, an International Journal*, vol. 32, 2005.

[5] R. Dillmann, O. Rogalla, M. Ehrenmann, R. Zollner, and M. Bordegoni, "Learning robot behaviour and skills based on human demonstration and advice: the machine learning paradigm," in *Robotics Research-International Symposium*, vol. 9, 2000, pp. 229–238.

[6] R. Zollner, O. Rogalla, and R. Dillmann, "Integration of tactile sensors in a programming by demonstration system," in *Robotics and Automation, 2001. Proceedings 2001 ICRA. IEEE International Conference on*, vol. 3, 2001, pp. 2578–2583 vol.3.

[7] R. Zollner, O. Rogalla, R. Dillmann, and M. Zollner, "Understanding users intention: programming fine manipulation tasks by demonstration," in *Intelligent Robots and Systems, 2002. IEEE/RSJ International Conference on*, vol. 2, 2002, pp. 1114–1119 vol.2.

[8] B. Hein, M. Hensel, and H. Wo?rn, "Intuitive and model-based on-line programming of industrial robots: A modular on-line programming environment," in *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*, 2008, pp. 3952–3957.

[9] B. Hein and H. Worn, "Intuitive and model-based on-line programming of industrial robots: New input devices," in *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*, 2009, pp. 3064–3069.

[10] P. Malheiros, P. Costa, and A. P. Moreira, "Robust 3d motion capture and object positioning system using light emitting markers synchronized with stereoscopic camera system," *UPIN NPat.77/ Pat. 41, Int. Patent PCT/IB2009/007186.*

[11] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, ISBN: 0521540518, 2004.

[12] V. Pratt, "Direct least-squares fitting of algebraic surfaces," *SIGGRAPH Comput. Graph.*, vol. 21, no. 4, pp. 145–152, Aug. 1987.

[13] G. Lukács, A. D. Marshall, and R. R. Martin, "Geometric least-squares fitting of spheres, cylinders, cones and tori," Tech. Rep., 1997.

[14] W. Kabsch, "A solution for the best rotation to relate two sets of vectors," *Acta Crystallographica Section A*, vol. 32, no. 5, pp. 922–923, Sept. 1976.