

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

ABB INC.,
Petitioner,

v.

ROBOTICVISIONTECH, INC.,
Patent Owner

Case: IPR2023-1426
U.S. Patent No. 8,095,237

PETITION FOR *INTER PARTES* REVIEW OF U.S. PATENT NO. 8,095,237

TABLE OF CONTENTS

I.	MANDATORY NOTICES UNDER 37 C.F.R. § 42.8.....	1
A.	Real Party-in-Interest.....	1
B.	Related Matters	1
C.	Lead and Back-up Counsel and Service Information.....	1
II.	GROUND FOR STANDING.....	2
III.	REQUESTED RELIEF	2
IV.	REASONS FOR THE REQUESTED RELIEF	2
A.	Summary of the '237 Patent	2
B.	Prosecution History	4
V.	STATUTORY GROUNDS FOR CHALLENGES	5
VI.	THE CHALLENGES ARE BASED ON PRIOR ART PATENTS AND PRINTED PUBLICATIONS	6
VII.	LEVEL OF ORDINARY SKILL IN THE ART	7
VIII.	CLAIM CONSTRUCTION	8
IX.	STATE OF THE PRIOR ART	9
A.	Corke.....	9
X.	GROUND 1: CLAIMS 1-4, 6-10, 17-20, AND 24-28 ARE UNPATENTABLE AS OBVIOUS OVER CORKE IN VIEW OF THE KNOWLEDGE OF A POSITA.....	11
A.	Claim 1.....	11
B.	Claim 2.....	22
C.	Claims 3 and 4	26
D.	Claim 6.....	30
E.	Claim 7.....	31
F.	Claim 8.....	31
G.	Claim 9.....	34
H.	Claim 10.....	36
I.	Claims 17, 24, and 28	37
J.	Claim 18.....	39

K. Claim 19.....	41
L. Claims 20 and 25	43
M. Claims 26 and 27	55
XI. GROUND 2: CLAIMS 5, 12-16, AND 21-24 ARE UNPATENTABLE AS OBVIOUS OVER CORKE IN VIEW OF WEI-I.....	56
A. Wei-I	56
B. Motivation to Combine Corke and Wei-I.....	57
C. Claims 5 and 12	59
D. Claim 13.....	64
E. Claim 14.....	65
F. Claim 15.....	67
G. Claim 16.....	68
H. Claim 21.....	73
I. Claims 22 and 23	76
XII. DISCRETIONARY DENIAL UNDER § 314 OR § 325(D) IS NOT WARRANTED.....	79
A. § 314	79
B. § 325(d).....	81

TABLE OF AUTHORITIES

Cases	Page(s)
<i>Apple Inc. v. Fintiv, Inc.</i> , IPR2020-00019, Paper 11 (PTAB March 20, 2020)	79
<i>Apple Inc. v. Fintiv, Inc.</i> , IPR2020-00019, Paper 15 (PTAB May 13, 2020)	79
<i>Becton, Dickinson & Co. v. B. Braun Melsungen AG</i> , IPR2017-01586, Paper 8 (P.T.A.B. Dec. 15, 2017)	81
<i>Bristol-Myers Squibb Co. v. Ben Venue Lab'ys, Inc.</i> , 246 F.3d 1368 (Fed. Cir. 2001)	11
<i>Harris Corp. v. Ericsson Inc.</i> , 417 F.3d 1241 (Fed. Cir. 2005)	8
<i>Koninklijke Philips N.V. v. Google LLC</i> , 948 F.3d 1330 (Fed. Cir. 2020)	14, 15
<i>KSR Int'l Co. v. Teleflex, Inc.</i> , 550 U.S. 398 (2007)	14, 60, 61
<i>Sand Revolution II, LLC v. Continental Intermodal Group - Trucking LLC</i> , IPR2019-01393, Paper 24 (PTAB June 16, 2020)	80
<i>TecSec, Inc. v. Int'l Bus. Machs. Corp.</i> , 731 F.3d 1336 (Fed. Cir. 2013)	8
Statutes	
35 U.S.C. §102	6
35 U.S.C. § 103	5, 7, 14
35 U.S.C. §112(6)	9
35 U.S.C. § 314	79
35 U.S.C. §315(e)(2)	80

IPR2023-01426

U.S. Patent No. 8,095,237

35 U.S.C. §316(a)(11).....79

§ 325(d) 79, 81

PETITIONER’S EXHIBIT LIST

Exhibit	Description
1001	U.S. Patent No. 8,095,237 (the “’237 Patent”)
1002	File History of U.S. Patent No. 8,095,237
1003	Declaration of Seth Hutchinson, Ph.D. Regarding Invalidity of U.S. Patent No. 8,095,237
1004	Peter I. Corke, <i>Visual Control of Robots: High-Performance Visual Servoing</i> (Aug. 1996) (“Corke”)
1005	Guo-Qing Wei, et al., “ <i>Active Self-calibration of Robotic Eyes and Hand-eye Relationships with Model Identification</i> ,” TRANSACTIONS ON ROBOTICS AND ANIMATION, vol. 14, No. 1, Feb 1998 (“Wei-I”)
1006	“Multisensory Visual Servoing by a Neural Network,” Guo-Qing Wei et al., IEEE Transactions on Systems, Man and Cybernetics, April 1999 (“Wei-II”).
1007	<i>RoboticVISIONTech, Inc. v. ABB Inc.</i> , No. 1:22-cv-01257, Summons in a Civil Action (served Sept. 23, 2023).
1008	U.S. Patent No. 4,146,924 to Birk et al.
1009	U.S. Patent No. 5,959,425 to Bieman et al.
1010	Declaration of Sylvia Hall-Ellis, Ph.D.
1011	Docket Sheet, <i>RoboticVISIONTech, Inc. v. ABB Inc.</i> , Case No. 1:22-cv-01257-GBW (D. Del.)

IPR2023-01426

U.S. Patent No. 8,095,237

1012	Curriculum Vitae of Seth Hutchinson, Ph.D.
1013	Scheduling Order, <i>RoboticVISIONTech, Inc. v. ABB Inc.</i> , Case No. 1:22-cv-01257-GBW (D. Del.)

I. MANDATORY NOTICES UNDER 37 C.F.R. § 42.8**A. Real Party-in-Interest**

ABB Inc. (“ABB”) is the Petitioner and real party-in-interest.

B. Related Matters

U.S. Patent No. 8,095,237 (the “237 Patent”) is asserted in *RoboticVISIONTech, Inc. v. ABB Inc.*, Case No. 1:22-cv-01257-GBW (D. Del.) (“Related Litigation”).

C. Lead and Back-up Counsel and Service Information

Lead Counsel	Backup Counsel
Andrew R. Sommer USPTO Reg. No. 53,932 Greenberg Traurig, LLP 1750 Tysons Boulevard Suite 1000 McLean, VA 22102 Phone: 703-749-1370 Fax: 703-749-1301 sommera@gtlaw.com	Benjamin Schladweiler (<i>pro hac vice</i> forthcoming) Greenberg Traurig, LLP 222 Delaware Avenue Suite 1600 Wilmington, DE 19801 Phone: 302-661-7352 Fax: 302-661-7360 schladweilb@gtlaw.com
Backup Counsel	Backup Counsel
Trenton A. Ward USPTO Reg. No. 59,157 Greenberg Traurig, LLP 333 Piedmont Road NE Suite 250 Atlanta, GA 3035 Phone: 678-553-2100 Fax: 678-553-02212 Trenton.ward@gtlaw.com	Kathryn E. Albanese USPTO Reg. No. 78,153 Greenberg Traurig, LLP One Vanderbilt Avenue New York, NY 10017 Phone: 212-801-6533 Fax: 212-801-6200 Katie.albanese@gtlaw.com

Please address all correspondence to lead and back-up counsel at ABB-RVT-LITIGATION@gtlaw.com. Petitioner consents to electronic service.

II. GROUNDS FOR STANDING

Petitioner certifies that the '237 Patent is available for *inter partes* review and that Petitioner is not barred or estopped from requesting *inter partes* review challenging the claims on the grounds identified in this Petition.

III. REQUESTED RELIEF

Petitioner requests institution of *inter partes* review of Claims 1-10 and 12-28 (“Challenged Claims”) of the '237 Patent, and cancelation of those claims as unpatentable.

IV. REASONS FOR THE REQUESTED RELIEF

For the reasons discussed in this Petition, the Board should institute trial and find the Challenged Claims unpatentable.

A. Summary of the '237 Patent

The '237 Patent describes the use of machine vision for 3D pose estimation. The methods of the '237 Patent include three steps: “a) calibration of the camera; b) teaching the features on the object; and c) finding the three-dimensional pose of the object.” EX1001, 2:60-67.

Figure 1 of the '237 Patent depicts a “vision-guided robot” 10 with a manipulating arm 12 on which camera 16 and tool 14—designed to manipulate a target object—are mounted. EX1001, 2:29, 2:53-59.

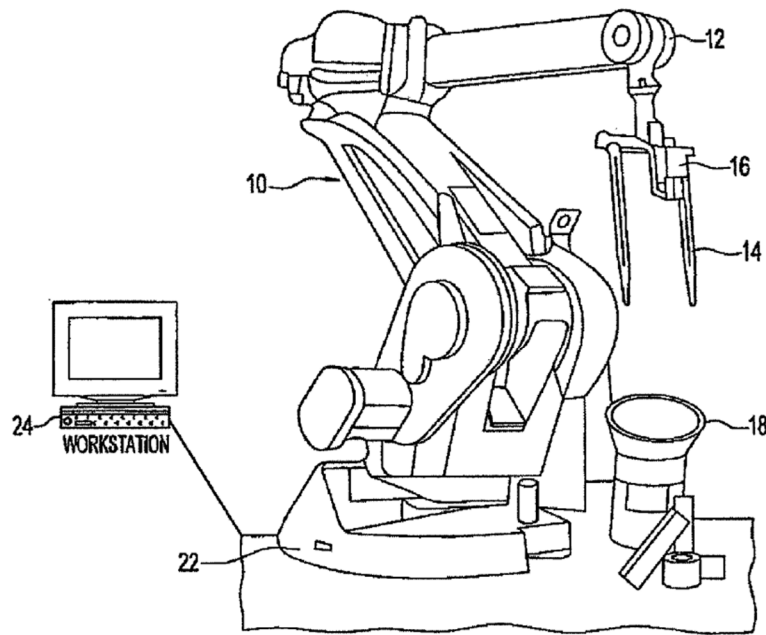


FIG. 1

The first step is to calibrate the camera. The '237 Patent describes three well-known types of calibration: (1) intrinsic calibration, which involves finding the “camera intrinsic parameters” describing “how the camera forms an image,” including the focal length of the camera, a radial distortion coefficient, coordinates of the center of radial lens distortion, and a scale factor, (2) extrinsic calibration, which involves finding the camera’s position and orientation (*i.e.*, “pose”) in the world coordinate frame and (3) hand-eye calibration, which involves finding the position and orientation of the camera “relative to the tool of the robot.” EX1001, 3:36-38, 3:56-67, 4:1-10, 5:51-65, 7:19-25, 7:48-49, 8:11-14, 8:30-39, 9:25-32, 9:44-65; EX1003, ¶¶24-41.

B. Prosecution History

The '237 Patent was filed as Application No. 10/634,874 (“the '874 Application”) on August 6, 2003 as a continuation-in-part of Application No. 10/153,680, which claims the benefit of a Canadian patent application filed on January 31, 2002. EX1001. The earliest possible effective filing date of the '237 Patent claims is January 31, 2002.

In eight years of prosecution, the Office rejected the claims in six Office Actions. During prosecution, Patentee canceled the original claims and presented new claims 33-61. EX1002, pp.64-65. These claims were rejected as anticipated by Wei, et al., “*Multisensory Visual Servoing by a Neural Network*, IEEE (1999) (“Wei-II”).¹ Patentee sought to distinguish Wei-II: “Importantly...Wei[-II]...is directed to the use of a stereo *pair of cameras*” but the claimed invention “employ[s] *single camera* three-dimensional (3-D) vision for robotic guidance.” EX1002, p.129. “It is the problems associated with *stereo vision* based systems and other *multi-camera* systems which are addressed by the various single camera embodiments described in Applicants’ specification and claims.” EX1002, p.129. Patentee argued that Wei-II’s techniques “are not directly applicable” to the claimed

¹ Wei-I and Wei-II were coauthored by Wei fourteen months apart, but disclose completely different methods. Whereas Wei-I discloses a method for “Active Calibration of Robotic Eyes,” Wei-II discloses a visual servoing method that “avoids all such calibrations.” EX1005, p.1; EX1006, p.276.

“*single camera* vision based methods and apparatus.” EX1002, p.130. Wei-II also sought to “*avoid* calibration,” and the ’237 Patent’s independent claims recited calibration steps. EX1002, p.130. On this basis, the Examiner allowed the claims. EX1002, p.193.

Several rejections followed, including obviousness rejections based on U.S. Patent No. 4,942,539 to McGee in view of Wei-II, and McGee in view of U.S. Patent Application Publication No. 2002/0159628 to Matusik. EX1002, pp.858-860. In response to the last of those rejections, Patentee argued that for application claim 50, the prior art failed to show “determining an object space-to-camera space transformation for the target object based at least in part on a position of at least some of the located features using only the single captured image.” EX1002, p.945.

V. STATUTORY GROUNDS FOR CHALLENGES

Petitioner presents the following challenges:

No.	Claims	Ground	Statute
1	1-4, 6-10, 17-20, 24-28	Obviousness over Peter I. Corke, <i>Visual Control of Robots: High-Performance Visual Servoing</i> (Aug. 1996) (“Corke”) in view of the knowledge of a POSITA	§103(a)
2	5, 12-16, 21-24	Obviousness over Corke in view of Guo-Qing Wei, et al., “ <i>Active Self-calibration of Robotic Eyes and Hand-eye Relationships with Model Identification</i> ,” TRANSACTIONS ON ROBOTICS AND ANIMATION, vol. 14, No. 1, Feb 1998 (“Wei-I”) and the knowledge of a POSITA	§103(a)

VI. THE CHALLENGES ARE BASED ON PRIOR ART PATENTS AND PRINTED PUBLICATIONS

Based on the references provided herein, Petitioner assumes, without admitting, that the claims of the '237 Patent are entitled to a filing date of January 31, 2002. Corke² was published in 1996 and was available to those working in the field no later than July 9, 1997 because, at least by this date, Corke was cataloged by subject matter in various databases including the OCLC bibliographic database and accessible to the public in the British Library and the Morgan Library at Colorado State University. EX1010, ¶¶40-45. Corke was distributed and accessible to the public based on (1) the print copy in the Morgan Library and Colorado State at least of July 9, 1997, (2) that it was cataloged by subject matter on or around that date, (3) and was widely held by libraries and was cataloged by subject matter on the OCLC bibliographic database. EX1010, ¶¶40-45. Regardless of which of these dates is the date of first publication, Corke is a prior art printed publication under §102(b). Corke is analogous to the '237 Patent because it is in the same field as the '237 Patent—vision guided robotics. EX1001, 1:13-15; EX1004, p.3 (discussing background of robotic vision).

² The version of Corke submitted with the Petition was obtained from the Morgan Library at Colorado State University. EX1010, ¶39.

Wei-I³ was included in the February 1998 edition of *Transactions on Robotics and Animation*. EX1005; EX1010, ¶46. Wei-I is expressly referenced in the '237 Patent specification and is Applicant Admitted Prior Art. EX1001, 8:30-39, 6:6-7:2. Specifically, the '237 Patent states that its “algorithm is developed using the approach described in” Wei-I. EX1001, 6:6-7:2. Wei-I was distributed and accessible to the public based on (1) the date stamp on the cover page showing that it was received in the Linda Hall Library on or about February 18, 1998, (2) that it was cataloged by subject matter on or around that date, (3) and was widely held by libraries and was cataloged by subject matter on the OCLC bibliographic database. EX1010, ¶¶47-52. Wei-I was thus publicly accessible on or shortly after February 18, 1998, making it prior art under at least §102(b). Wei-I is analogous to the '237 Patent because it is in the same field as the '237 Patent—vision guided robotics. EX1001, 1:13-15; EX1005, p.1 (referring to “camera self-calibration” achieved in “robotics and computer vision.”).

VII. LEVEL OF ORDINARY SKILL IN THE ART

As of January 31, 2002, a POSITA would have had a Bachelor's degree in robotics, mechanical engineering, computer science, electrical engineering, or an equivalent, and at least three years of professional experience working in the field

³ The version of Wei-I submitted with the Petition was obtained from the Linda Hall Library of Science, Engineering & Technology. EX1010, ¶46.

of computer vision or three years of graduate education, including a focus on computer vision applications. The POSITA would have knowledge about machine vision and camera calibration techniques. EX1003, ¶¶77-78.

VIII. CLAIM CONSTRUCTION

For this Petition only, Petitioner relies upon the definitions provided in the '237 Patent, including for “object space” as “a reference frame defined with respect to, and therefore rigid to, the object” and “training space” as “a reference frame defined with respect to a point on the calibration template, and aligned to its main axis.” EX1001, 3:3-34.

Claims 20-28 recite “means for calibrating the camera” and “means for estimating a pose of a target object.” EX1001, 13:41-14:65. The presumption that these are means-plus-function claims can be overcome if the claims recite structural, rather than functional, language. *TecSec, Inc. v. Int’l Bus. Machs. Corp.*, 731 F.3d 1336, 1347 (Fed. Cir. 2013). “[I]t is sufficient if the claim term is used in common parlance or by persons of skill in the pertinent art to designate structure.” *Id.* The Federal Circuit has established that the corresponding structure for a “computer-implemented means-plus-function” is “the algorithm.” *Harris Corp. v. Ericsson Inc.*, 417 F.3d 1241, 1253 (Fed. Cir. 2005).

In the '237 Patent, the functional language provided in Claims 20 and 28 for the “means for calibrating the camera” is the calibration algorithm itself, as the claim

itself recites “determining a set of intrinsic parameters” and “determining a set of extrinsic parameters” “from at least one of the images of the calibration object,” which results in performing the calibration algorithm recited in the specification. EX1001, Fig. 5; EX1003, ¶88. The same is true with the recited “means for estimating the pose of a target object” because completing the claimed steps of “capturing a two-dimensional image of a volume containing a target object,” “locating” a number of “features in the captured image,” and “determining an object space-to-camera space transformation,” results in performing the pose estimation algorithm recited in the specification. EX1003, ¶89; EX1001, Fig. 7. Therefore, Claims 20-28 are not means-plus-function claims and no construction is necessary. A POSITA reading the specification of the '237 Patent would have understood that Claims 20-28 provide sufficient structure to perform the algorithm cited therein; thus, these terms do not invoke 35 U.S.C. §112(6). EX1003, ¶90.

IX. STATE OF THE PRIOR ART

A. Corke

Corke is a textbook published in 1996 about the “application of high-speed machine vision for closed-loop position control, or visual servoing, of a robot manipulator.” EX1004, p.ix; EX1003, ¶¶79-81. Corke explains that “visual servoing involves the use of one or more cameras and a computer vision system to control the position of the robot’s end-effector relative to the workpiece as required by the task.”

EX1004, p.1. In Figure 5.4, Corke discloses a robot having a single camera to implement a method for 3D pose estimation:

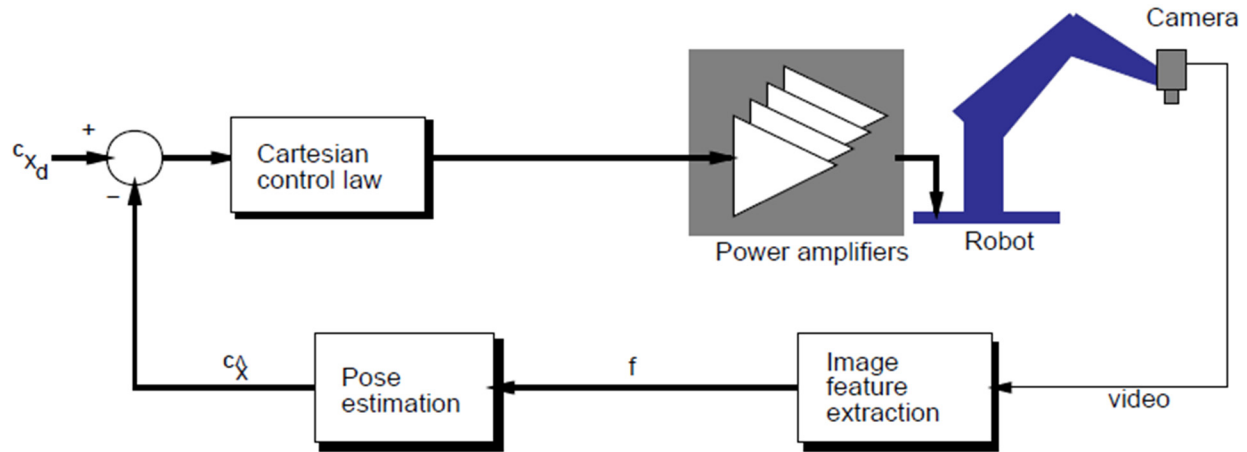


Figure 5.4: Position-based visual servo (PBVS) structure as per Weiss.

EX1004, Figure 5.4.

Corke explains that a key component of visual servoing systems is the use of vision sensing and a closed “feedback” loop through which the system (for each video frame) takes an image of the target object, extracts features from the object, determines the “relative pose between the camera and the target” using that single image and “knowledge of the geometric relationship between several feature points on the target,” and sends the pose estimation to the robot, causing the robot to adjust the position of the robot’s end-effector relative to the object. EX1004, pp.3, 152-55.

X. GROUND 1: CLAIMS 1-4, 6-10, 17-20, AND 24-28 ARE UNPATENTABLE AS OBVIOUS OVER CORKE IN VIEW OF THE KNOWLEDGE OF A POSITA

A. Claim 1

[1.0] A useful in three-dimensional pose estimation for use with a single camera mounted to a movable portion of a robot, the method comprising:

To the extent the Board construes the preamble as limiting, Corke discloses the preamble. EX1003, ¶¶93-96. A preamble is limiting if it gives “life, meaning, and vitality” to a claim. *Bristol-Myers Squibb Co. v. Ben Venue Labs, Inc.*, 246 F.3d 1368, 1373 (Fed. Cir. 2001). Here, the recitation in the preamble of “a single camera mounted to a moveable portion of a robot” is limiting in view of Patentee’s reliance upon this limitation for allowance. As detailed above, Patentee distinguished the claims over the prior art as “directed to methods and apparatus that employ *single camera* three-dimensional (3-D) vision for robotic guidance.” EX1002, p.129.

As to the method recited in the preamble of Claim 1, Corke discloses methods for 3D pose estimation for which “[t]he camera may be fixed, or *mounted on the robot’s end-effector* in which case there exists a constant relationship, ${}^{6}x_c$, between the pose of the camera and the pose of the end-effector.” EX1004, pp.152, 207 (disclosing a “single camera” mounted on robot wrist). Corke’s techniques designed to estimate the “3D position and orientation”—*i.e.*, “pose”—of an object are useful

in three-dimensional pose estimation.⁴ EX1004, p.3 n.2; EX1003, ¶95. Placing the camera on the robot's arm—corresponding to the claimed moveable portion of a robot—provides certain advantages, including “the ability to avoid occlusion, resolve ambiguity and increase accuracy, by directing its attention.” EX1004, pp.3, 166. Figures 4.11, 5.1, 6.10 and 6.11 of Corke illustrate “a single camera mounted to a moveable part of the robot”:

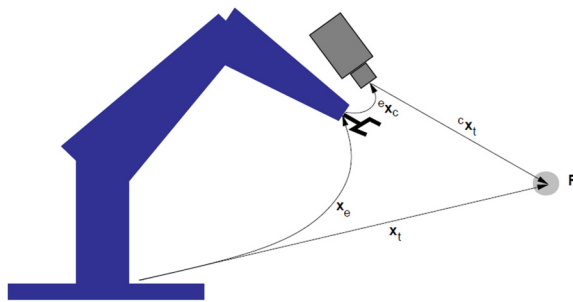


Figure 5.1: Relevant coordinate frames: world, end-effector, camera and target.

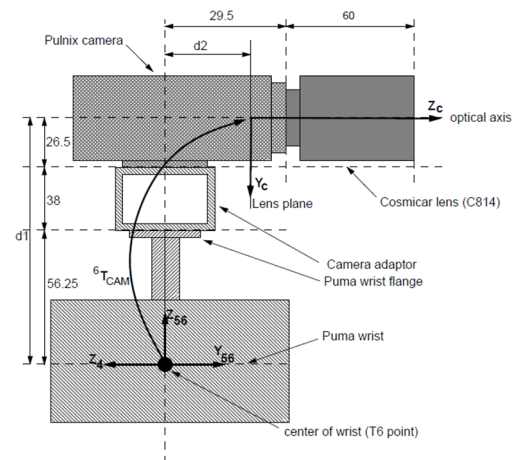


Figure 4.11: Details of camera mounting (not to scale, dimensions in mm). Robot wrist is in the zero angle pose.

⁴ Underlined text denotes language from the claims, unless otherwise noted.

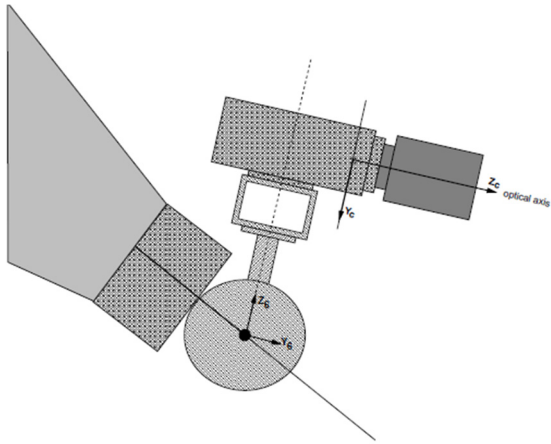


Figure 6.10: Camera mount used in this work. In this pose $\theta_0 \approx -\pi/2$.

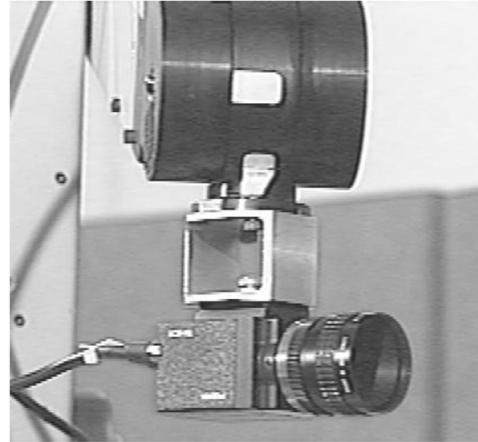


Figure 6.11: Photograph of camera mounting arrangement.

EX1004, pp.148, 152, 186-87; EX1003, ¶95. Corke in view of the knowledge of a POSITA discloses or renders obvious Preamble [1.0]. EX1003, ¶¶93-96.

[1.1] capturing a two-dimensional image of a volume containing a target object;

Corke teaches a method that captur[es] a two-dimensional image of a volume containing a target object. EX1003, ¶¶97-101. Corke discloses that “[t]he camera contains a lens which *forms a 2D projection of the scene on the image plane* where the sensor is located.” EX1004, pp. 152, 159 (“A broad definition of position-based servoing...includes methods based on *analysis of 2D features*”), 159 (disclosing method to “determine the 3D relative pose of an object...from 2D image plane coordinates”); EX1003, ¶98.

As shown in Figure 5.4, Corke discloses a system that uses a video camera to capture and process multiple two-dimensional images of a target object, one image at a time at the “camera frame rate,” which is “effectively the sample rate in a visual

servoing system.” EX1004, p.103, 73 (“the camera’s shutter acts as the sampler in a visual control loop”):⁵

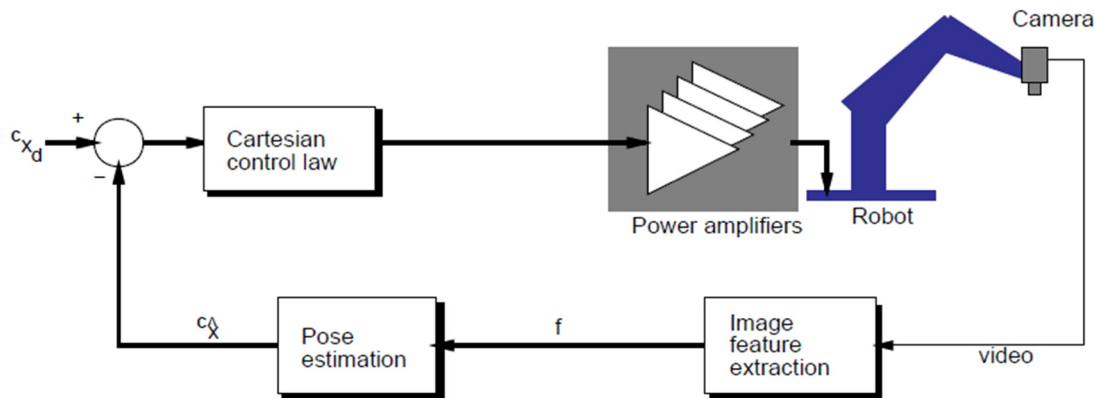


Figure 5.4: Position-based visual servo (PBVS) structure as per Weiss.

EX1004, p.155; EX1003, ¶99.

The claim requires that the target object reside in a volume, which Corke teaches is the space from which the image of the target object is captured. EX1004, pp.152-153 (referring to “an image of the target” as “the object of interest, that is, the object that will be tracked,” so that “features” can be “extracted from the image”); EX1003, ¶100. In *KSR Int’l Co. v. Teleflex, Inc.*, the Supreme Court held the obviousness analysis should consider the “background knowledge possessed by a person having ordinary skill in the art.” 550 U.S. 398 (2007).⁶ A POSITA would

⁵ The portions of Corke cited herein are from a single embodiment referenced in both Chapter 3 (“Fundamentals of image capture”) and Chapter 4 (“Machine vision” concepts), and used as a baseline for the position-based visual servoing methods discussed in Chapter 5. EX1004.

⁶ In *Koninklijke Philips N.V. v. Google LLC*, the Federal Circuit determined that “under 35 U.S.C. § 103, the obviousness inquiry turns not only on the prior art, but

have understood that all 3D objects occupy a volume. EX1003, ¶100. Corke discloses that “[t]he camera contains a lens which forms a 2D projection of the scene on the image plane,” and “each point on the image plane corresponds to a ray *in 3D space*.” EX1004, pp.152. Accordingly, Corke’s teaching of capturing a two-dimensional image of a target object by a single camera would have included capturing a two-dimensional image of a volume containing a target object. EX1003, ¶100. Corke in view of the knowledge of a POSITA renders obvious Element [1.1]. EX1003, ¶¶97-101.

[1.2] locating a number of features in the captured image of the target object;

Corke discloses locating a number of features in the captured image of the target object. EX1004, pp.136, 168; EX1003, ¶¶102-107. For example, Corke discloses a method of “*locating features*” in each captured image where “[d]etermining the initial *location of features* requires *the entire image* to be searched.” EX1004, p.136, 168; EX1003, ¶103. Corke discloses that “[a] good *feature* is one that can be *located* unambiguously” such as “a hole in a gasket [] or a contrived pattern.” EX1004, pp.153; EX1003, ¶103.

whether the differences between the claimed invention and the prior art are such that the claimed invention as a whole would have been obvious . . . to a person having ordinary skill in the art to which the claimed invention pertains.” 948 F.3d 1330, 1337 (Fed. Cir. 2020) (affirming PTAB’s determination of unpatentability based upon cited prior reference in view of knowledge of POSITA).

Corke also discloses that in “*position based control, features are extracted from the image*” of a target object to “determine the pose of the target with respect to the camera.” EX1004, p.152-53 (“The task in visual servoing is to control the *pose* of the robot’s end-effector, ${}^xT_6$, using visual information, *features*, extracted from *the image*”). EX1003, ¶104. Figure 5.4 shows the “Image feature extraction” step disclosed in Corke:

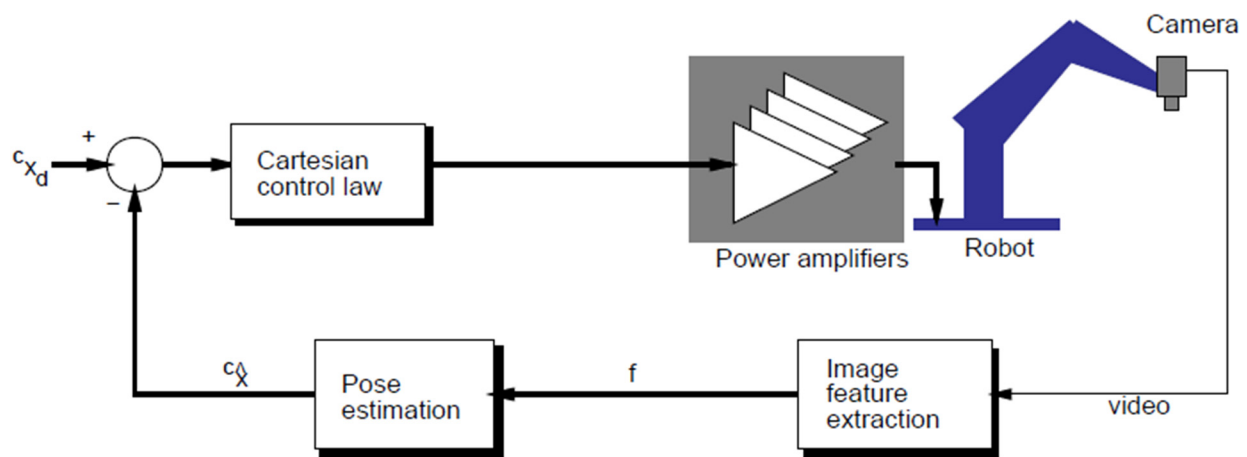


Figure 5.4: Position-based visual servo (PBVS) structure as per Weiss.

EX1004, p.155. A POSITA would have understood that Corke teaches that features are located and processed one image at a time in view of Corke’s express disclosure that “features are extracted from *the image*.” EX1004, p.153, 152 (“*features*” “extracted from *the image*”); Figure 5.4 (showing “*Image* feature extraction”). EX1003, ¶105. As an example of single-image processing, Corke teaches use of the Datacube DIGIMAX for acquiring an individual image from the camera and processing this individual image to locate features, stating “[t]he incoming analog

video signal is digitized to form a *512x512 pixel digital image*,” after which various features are detected, such as “moments up to second order, perimeter and bounding box.” EX1004, p.180; EX1003, ¶105.

Corke also discloses that when machine vision is used for “part inspection and quality control,” the process involves “extraction” of “features from the image[.]” EX1004, p.123. By disclosing that “features are extracted,” Corke teaches the POSITA that a number of features are located in the captured image of the target object. EX1003, ¶106.

Corke in view of the knowledge of a POSITA renders obvious Element [1.2]. EX1003, ¶¶102-107.

[1.3] determining by a processor an object space-to-camera space transformation for the target object based at least in part on a position of at least some of the located features using only the single captured image and an algorithm that employs a known or determinable physical relationship between at least some of the located features.

Corke discloses determining by a processor an object space-to-camera space transformation for the target object utilizing the position of the features from a single image. EX1003, ¶¶108-116. The '237 Patent provides that “object space” is a 3D coordinate frame “defined with respect to a point on, and therefore rigid to, the object,” “camera space” is a coordinate frame rigid to the camera, and “transformation” is the “three-dimensional rotation & translation between two spaces” or coordinate frames. EX1001, 3:1-34.

Corke discloses that in “*position based* control, features are extracted from *the image*” of the target object “and used in conjunction with a geometric model of the target to determine the pose of the target with respect to the camera.” EX1004 p.153. A POSITA would have understood that determining the target’s pose with respect to the camera is equivalent to the claimed step of determining the object space-to-camera space transformation. EX1003, ¶110. This object space-to-camera space transformation is depicted in Figure 5.1 of Corke, where **P** designates the target object and ${}^c\mathbf{x}_t$ denotes “the relative pose between the camera and the target,” *i.e.*, the transformation between the camera and the target object—the “object of interest”:

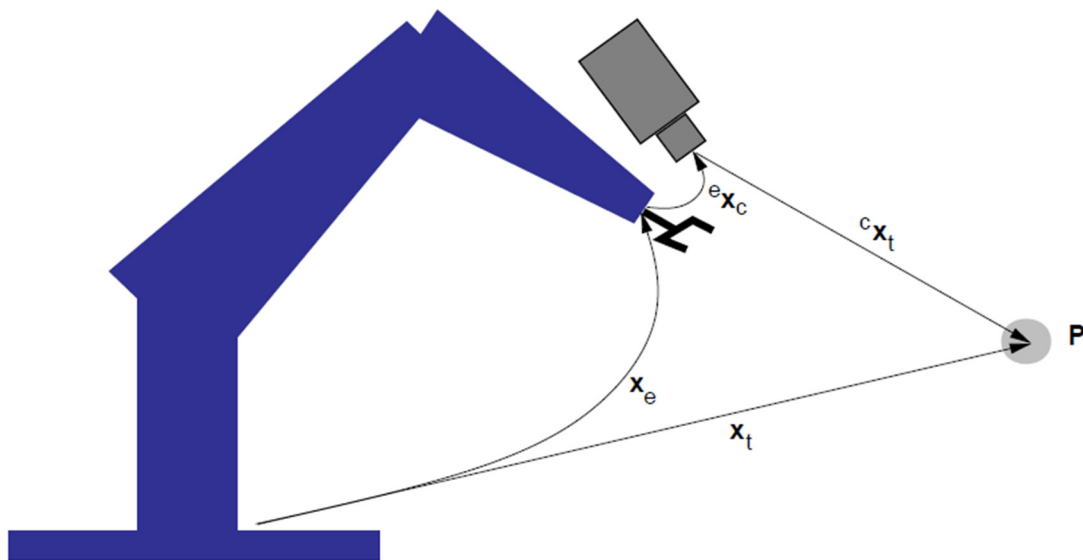


Figure 5.1: Relevant coordinate frames; world, end-effector, camera and target.

EX1004 pp.152, 138; EX1003, ¶110.

A POSITA would have understood that Corke's references to "*the image*" teaches that only features from a single captured image are used to determine the object space-to-camera space transformation. EX1004, pp.152 ("The task in visual servoing is to control the *pose* of the robot's end-effector, ${}^xT_6$, using visual information, *features*, extracted from *the image*."); 153 ("*The image plane coordinates* of three or more visual features can be used to determine the pose...of the target relative to the camera, given knowledge of the geometric relationship between the feature points."); EX1003, ¶111. As an example of single-image processing, Corke teaches use of the Datacube DIGIMAX for acquiring an individual image from the camera and processing this individual image to locate features, stating "[t]he incoming analog video signal is digitized to form a **512x512 pixel digital image**," after which various features are detected, such as "moments up to second order, perimeter and bounding box," EX1004, p.180. EX1003, ¶111.

Corke teaches an algorithm that uses the position of features from a single image to determine the object space-to-camera space transformation for the target object based at least in part on a position of at least some of the located features. EX1004, p.138; EX1003, ¶112. Corke teaches in equation 4.60 that camera calibration matrix, **C**, "represents the relationship between 3-D world coordinates and their corresponding 2-D image coordinates as seen by the computer," with i_x , i_y

and i_z encoding the position of a feature in the image plane, and x , y , and z encoding the position of the same feature in 3D “world coordinates”:

$$\begin{bmatrix} i_x \\ i_y \\ i_z \end{bmatrix} = \mathbf{C} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad (4.60)$$

and with the matrix $\mathbf{C}$ defined in equation 4.61:

$$\mathbf{C} = \begin{bmatrix} \frac{\alpha_x n_x f - X_0 n_z}{f - c_z} & \frac{\alpha_x \sigma_x f - X_0 \sigma_z}{f - c_z} & \frac{\alpha_x \alpha_x f - X_0 a_z}{f - c_z} & \frac{\alpha_x c_x f - X_0 c_z + X_0 f}{f - c_z} \\ \frac{\alpha_y n_y f - Y_0 n_z}{f - c_z} & \frac{\alpha_y \sigma_y f - Y_0 \sigma_z}{f - c_z} & \frac{\alpha_y \alpha_y f - Y_0 a_z}{f - c_z} & \frac{\alpha_y c_y f - Y_0 c_z + Y_0 f}{f - c_z} \\ -\frac{n_z}{f - c_z} & -\frac{\sigma_z}{f - c_z} & -\frac{a_z}{f - c_z} & 1 \end{bmatrix} \quad (4.61)$$

EX1004, p.138. Corke teaches methods for determining matrix $\mathbf{C}$ using the position of multiple features in a single image. EX1004, pp.139-142; EX1003, ¶112. Corke teaches “[t]he calibration matrix, $\mathbf{C}$, encodes the 6 extrinsic parameters,” noting these extrinsic parameters are the “3D position vector and 3 angles for camera pose.” EX1004, p.144. Corke teaches an algorithm to recover these parameters from the matrix $\mathbf{C}$ by decomposing the camera calibration matrix. EX1004, pp.143-144; EX1003, ¶112. A POSITA would have understood that if the 3D coordinates are specified with respect to the object space, then the camera pose obtained by Corke’s algorithm is defined with respect to the object space, and therefore determines the object space-to-camera space transformation. EX1003, ¶112.

Using Corke’s equation 4.60, a POSITA would have understood that by multiplying the camera calibration $\mathbf{C}$ by the 3D position of any feature in 3D space,

the corresponding position for that feature in the 2D image frame (${}^i\mathbf{x}$, ${}^i\mathbf{y}$, ${}^i\mathbf{z}$) is obtained, and those coordinates can be converted to two-dimensional camera frame pixel coordinates— ${}^iX$, ${}^iY$ —by Corke’s equations 4.56 and 4.57:

$${}^iX = \frac{{}^ix}{{}^iz} \quad (4.56)$$

$${}^iY = \frac{{}^iy}{{}^iz} \quad (4.57)$$

EX1004, pp.138, xi; EX1003, ¶113. Corke discloses that this algorithm works with features extracted from a single captured image. EX1004, p.138; EX1003, ¶113.

The algorithm disclosed in Corke employs a known physical relationship between some of the located features using stored knowledge of model feature points in 3D space. EX1004, p.159; EX1003, ¶114. Corke explains that to “determine the 3D relative pose of an object...from 2D image coordinates,” “additional information is needed.” EX1004, p.159. This includes “***knowledge of the relationship between the observed feature points*** (perhaps from a CAD model) and also the camera’s intrinsic parameters.” EX1004, p.159; 153 (“three or more visual features can be used to determine the pose...***given knowledge of the geometric relationship between the feature points***”). Corke also explains that the object space-to-camera space transformation is determined by an algorithm using “features extracted from the image” “***in conjunction with a geometric model of the target***,” and “each feature point yields two equations and six equations are needed to solve the six elements of

the pose vector.” EX1004, pp.153, 159-60; EX1003, ¶114. A POSITA would have understood that Corke teaches an algorithm using the position of “features extracted from the image” (only the single captured image) and “a geometric model of the target” (known physical relationship) processed by a “computer” (processor) to determine the “3D relative pose of an object” (object space-to-camera space transformation for the target object). EX1004, pp. 153, 159-160 (“features are extracted from the image and used in conjunction with a geometric model of the target to determine the pose of the target with respect to the camera”); EX1003, ¶114.

Corke also discloses that the object space-to-camera space algorithm can be performed “by a computer,” which a POSITA would have understood includes the claimed processor. EX1004, p.138; EX1003, ¶115. Corke in view of the knowledge of a POSITA renders obvious Element [1.3]. EX1003, ¶¶108-116.

B. Claim 2

[2.0] The method of claim 1, further comprising: [2.1] capturing a number of images of a calibration object by the camera; and

Corke discloses a camera capturing a number of images of a calibration object. EX1003, ¶¶117-119. In Corke’s experiments, “[f]our images of the SHAPE calibration target, shown in Figure 4.8 below, were taken from different viewpoints.” EX1004, p.145; EX1003, ¶118.

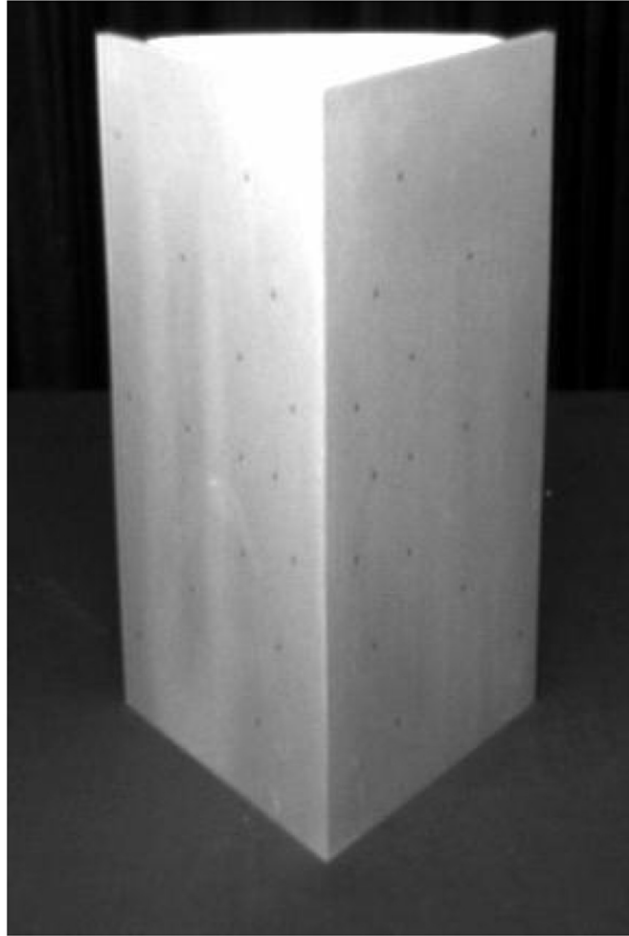


Figure 4.8: The SHAPE calibration target used for intrinsic parameter determination. The surface is brushed aluminium with black marker dots at accurately known locations on the two surfaces.

EX1004, p.140.

Corke in view of the knowledge of a POSITA renders obvious Elements [2.0] and [2.1]. EX1003, ¶¶117-119.

[2.2] determining a set of intrinsic parameters of the camera from at least one of the number of images of the calibration object captured by the camera;

Corke discloses determining a set of intrinsic parameters of the camera from images of the calibration object by the camera. EX1004, p.143-44; EX1003, ¶¶120-123. Corke discloses that four images taken by the camera of the SHAPE

calibration target were used to compute the “calibration matrix”—a mathematical object that determines “intrinsic parameters of the camera.” EX1004, p.143-144 (the “calibration matrix” encodes the “5 intrinsic parameters: f , α_x , α_y , X_0 and Y_0 ”). EX1003, ¶121. Table 4.2 of Corke discloses using features extracted from those four images to calculate the calibration matrix, including 5 intrinsic parameters:

Trial	Residual (pixel)	δ (deg)	X_0 (pixel)	Y_0 (pixel)	$\alpha_x f$ (pixel)	$\alpha_y f$ (pixel)	α_y/α_x
1	0.552	-0.0277	270.1	213.7	625.7	948.9	1.52
2	0.898	0.0672	271.4	223.1	614.0	933.0	1.52
3	0.605	0.0685	278.0	205.6	617.3	938.2	1.52
4	0.639	-0.1198	275.8	196.8	616.8	939.2	1.52
mean			273.8	209.8	615.5	939.8	
σ			3.2	9.7	5.0	6.6	

Table 4.2: Summary of calibration experiment results.

EX1004, pp.146, 138-139. Corke also defines camera calibration generally as “the process of determining the internal camera geometric and optical characteristics (intrinsic parameters).” EX1004, p.139; EX1003, ¶122. Corke in view of the knowledge of a POSITA renders obvious Element [2.2]. EX1003, ¶¶120-123.

[2.3] determining a set of extrinsic parameters of the camera from at least one of the number of images of the calibration object captured by the camera, the set of extrinsic parameters comprising a camera space-to-training space transformation defining a transformation between a camera space reference frame and a training space reference frame.

Corke discloses determining a set of extrinsic parameters of the camera from at least one of the images of the calibration object captured by the camera. EX1004,

p.146; EX1003, ¶¶124-127. Corke explains that four images of the SHAPE calibration target were used to compute the “calibration matrix”—a mathematical object that “contains information about the position of the camera with respect to the world coordinate frame”—the extrinsic parameters. EX1004, p.143-144 (the “calibration matrix” “encodes the 6 extrinsic parameters”—*i.e.*, the “3D position vector and 3 angles for camera pose”). Corke also defines camera calibration generally as “the process of determining the...orientation of the camera frame relative to the world coordinate system (extrinsic parameters).” EX1004, p.139 (“the camera calibration matrix...relates a world coordinate system to the image plane coordinates for a given camera position and orientation”); EX1003, ¶125.

Corke further teaches that the extrinsic parameters include a transformation between a camera space reference frame and a training space reference frame. EX1003, ¶126. Specifically, Corke teaches that “*extrinsic* parameters” relate to “the position and orientation of the camera.” EX1004, p.138. One of those “extrinsic” parameters is ${}^C T_P$ —“the position of the [calibration] object with respect to the camera.” EX1004, p.127. A POSITA would have understood this was the “transformation”— T —between the coordinate frame defined by the camera— C —and the coordinate frame defined by the calibration object— P . EX1004, p.xi; EX1003, ¶126. Because “training space” is a “reference frame defined with respect to a point on the calibration” object, a POSITA would have understood

that Corke discloses determining a set of extrinsic parameters of the camera from at least one of the number of images of the calibration object captured by the camera, the set of extrinsic parameters comprising a camera space-to-training space transformation. EX1001, 3:22-24; EX1003, ¶126; *see also* EX1004, p.3 (transformations described in Corke provide “positioning information” “directly in Cartesian or task space”—*i.e.*, training space).

Corke in view of the knowledge of a POSITA renders obvious Element [2.3]. EX1003, ¶¶124-127.

C. Claims 3 and 4

Corke discloses positioning the camera with respect to the calibration object. EX1003, ¶¶128-133. In Corke’s experiments, a camera mounted to a moveable part of the robot took “[f]our images of the SHAPE calibration target...from different viewpoints.” EX1004, p.145.

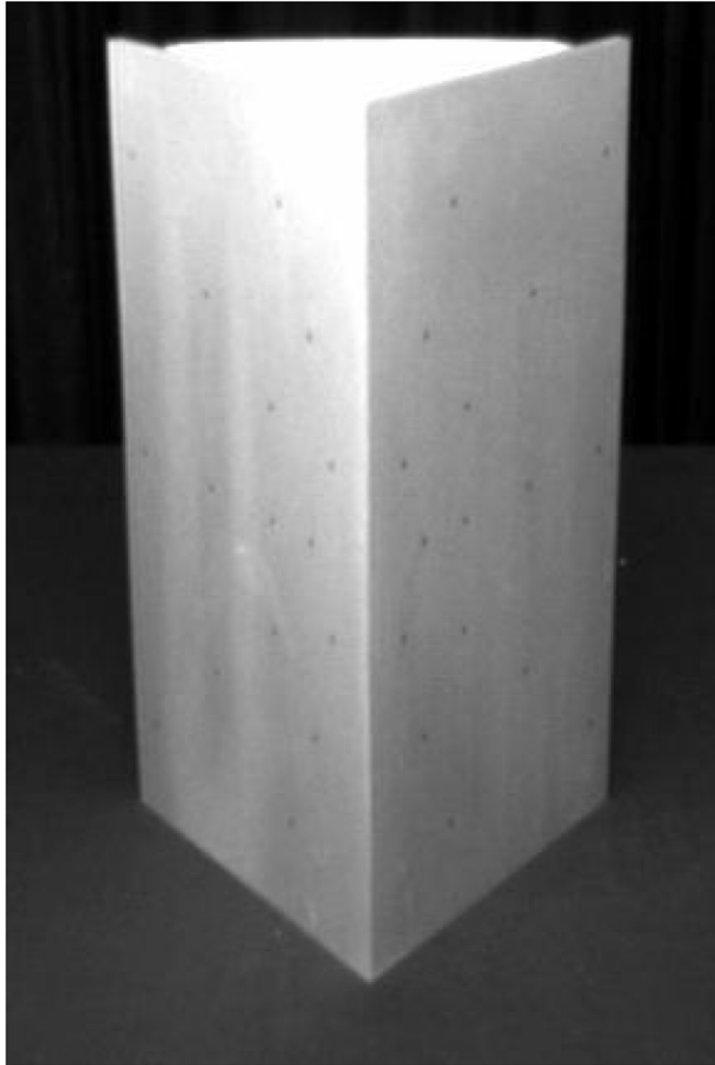


Figure 4.8: The SHAPE calibration target used for intrinsic parameter determination. The surface is brushed aluminium with black marker dots at accurately known locations on the two surfaces.

EX1004, p.140.

Corke further discloses positioning the camera orthogonal to a ruled template with a known physical relationship between the features. EX1003, ¶130. For example, Corke discloses acceptable camera calibration techniques, including the two-plane camera model. EX1004, pp.141-43. This method, illustrated in Figure 4.9, shows the camera's "image plane" is parallel to the two "calibration planes,"

meaning that images were captured by a camera placed orthogonal to the calibration templates:

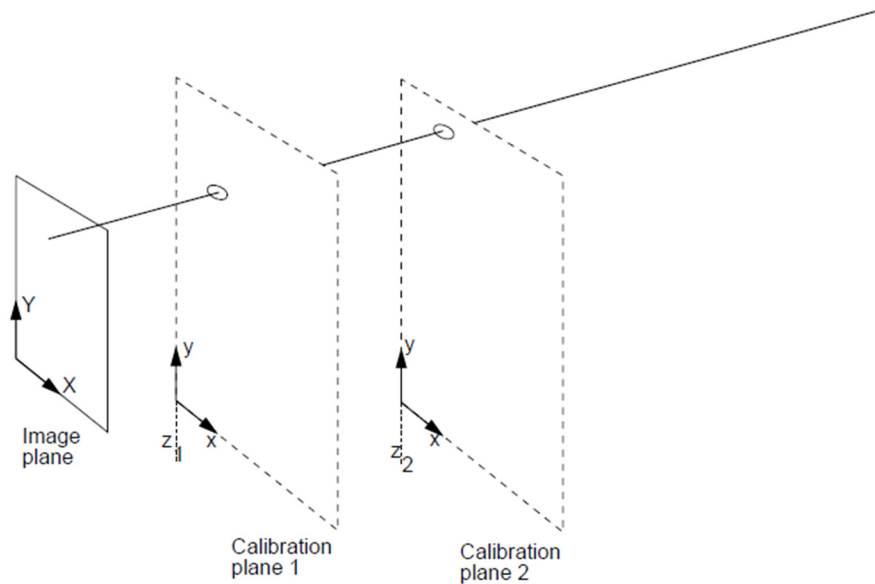


Figure 4.9: The two-plane camera model.

EX1004, p.142; EX1003, ¶130. The calibration templates used in Corke's two-plane calibration method have a ruled template or "test pattern containing an array of dots at known distances from the camera." EX1004, p.143; EX1003, ¶131.

Corke discloses that visual servoing can generally be performed to position a camera at an arbitrary orientation with respect to a target. EX1004, pp.161-162; EX1003, ¶132. A POSITA would have understood that visual servoing could be used to position the camera to be orthogonal to the calibration target, this simply being a special case of visual servoing. EX1003, ¶132. Figure 5.6 shows positioning a camera orthogonal to the object, and illustrates moving a camera from the initial position (left) to a final position orthogonal to the object in question (right):

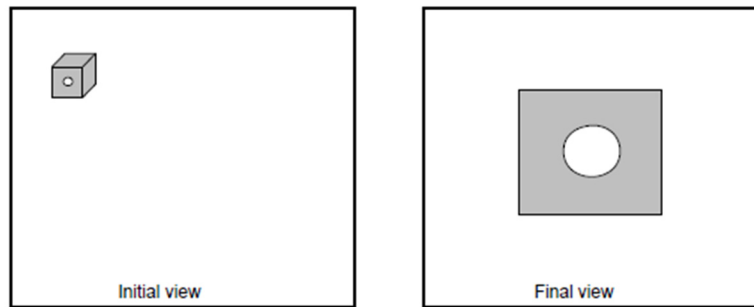


Figure 5.6: Example of initial and desired view of a cube

EX1004, p.162; EX1003, ¶132. Corke describes that “[t]he feature vector may comprise coordinates of vertices, or areas of the faces. Implicit in f_d is that the robot is *normal* to, and centered over the face of the cube, at the desired distance” (EX1004 p.161), where a POSITA would have understood “normal to” as equivalent to placing the camera “orthogonal to” the target. EX1003, ¶132. Section 5.4 of Corke says that the ROBOT is “normal” to the object. EX1004, p.161; EX1003, ¶132. From this paragraph and Figure 5.6, a POSITA would have understood that the robot *and* the end-effector mounted camera would be normal to the object. EX1004, p.161; EX1003, ¶132. Thus, Corke in view of the knowledge of a POSITA teaches positioning the camera orthogonally with respect to a ruled template with a number of features, where a known or determinable physical relationship exists between at least some of the features. Corke in view of the knowledge of a POSITA renders obvious Claims 3 and 4. EX1003, ¶¶128-133.

D. Claim 6

Corke teaches capturing a number of images of a calibration object at different positions. EX1003, ¶¶134-136. In Figure 4.9, Corke discloses a “two-plane calibration” method where at least two different images are taken of the calibration object. EX1004, p.142; EX1003, ¶135. Because the camera’s “image plane” is parallel to the two “calibration planes,” Corke teaches that the images were captured by a camera placed orthogonal to the calibration templates:

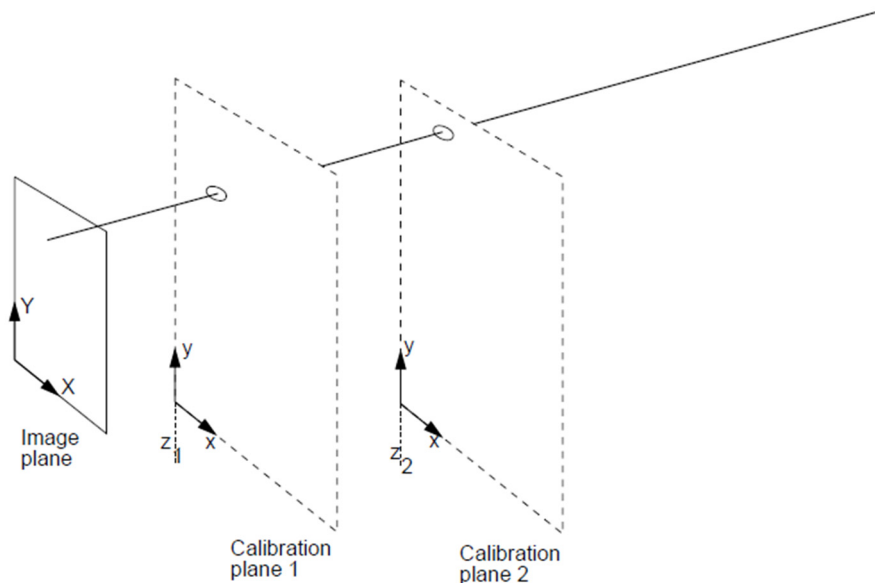


Figure 4.9: The two-plane camera model.

EX1004, p.142.

Specifically, a POSITA would have understood “spaced perpendicularly from the target” to teach that the camera locations are spaced along a line perpendicular to the target. EX1003, ¶135. For example, Corke discloses in Figure 4.9 that the image plane (perpendicular to the camera’s optical axis) is parallel to the two

calibration planes. EX1003, ¶135. Thus, a POSITA would have understood that motion of the camera along the optical axis (a line perpendicular to the two calibration planes) would enable capturing a number of images of a calibration object by the camera, including capturing at least one image at each of a plurality of positions spaced perpendicularly from the calibration object. EX1003, ¶135. Corke in view of the knowledge of a POSITA renders obvious Claim 6. EX1003, ¶¶134-136.

E. Claim 7

Corke discloses capturing several images of a calibration object at different orientations with respect to the calibration object. EX1003, ¶¶137-139. Corke discloses experiments in which “four images of the SHAPE calibration target...were taken from different viewpoints.” EX1004, p.145.

Corke in view of the knowledge of a POSITA renders obvious Claim 7. EX1003, ¶¶137-139.

F. Claim 8

Corke teaches determining a set of intrinsic parameters of the camera from the at least one of the number of images of the calibration object captured by the camera, including at least one of a [1] focal length, [2] a first order radial lens distortion coefficient, and [3] coordinates of a center of a radial distortion (called the principal point). EX1003, ¶¶140-147. Corke explains that “[i]ntrinsic parameters are innate

characteristics of the camera,” and “include the [1] *focal length*, [2] *lens distortion*, and the [3] coordinates of the *principal point*”—*i.e.*, “the intersection of the optical axis and the [] sensor plane.” EX1004, pp.138-39. A POSITA would have understood that determining “coordinates of the principal point” is the same as [3] determining a set of coordinates of a center of a radial distortion. EX1003, ¶141.

As for [1], Corke discloses calibration of the focal length, given the known pixel size in the sensing array. EX1003, ¶142. In Corke, pixel size is encoded in parameters α_x and α_y . EX1004, p.137. Corke notes “it is not possible to uniquely determine all intrinsic parameters, only the products of $f\alpha_x$ and $f\alpha_y$,” where f denotes focal length. A POSITA would have understood that camera manufacturers typically supply the specification of pixel size α_x and α_y such that Corke discloses calibration of focal length f . EX1003, ¶142. For example, Table 3.9 of Corke discloses α_x and α_y for the Pulnix TM-6 camera and DIGIMIAX digitizer, computed using the manufacturer’s specifications from Table 3.8:

Parameter	Value
Sensor width	6.5 mm
Sensor height	4.8 mm
Horizontal active pixels	752 pixels
Vertical active lines	582 pixels
Pixel width (p_x)	8.6 μm
Pixel height (p_y)	8.3 μm

Table 3.8: Manufacturer’s specifications for the Pulnix TM-6 camera.

Parameter	Value
α_x	79.2 pixel/mm
α_y	120.5 pixel/mm (frame mode)
α_y	60.2 pixel/mm (field mode)

Table 3.9: Derived pixel scale factors for the Pulnix TM-6 camera and DIGIMAX digitizer.

EX1004, p.111; EX1003, ¶142.

As for [2], Corke explains that radial distortion “causes image points to be translated along radial lines from the principal point.” EX1004, p.86. According to Corke, *radial distortion* may be approximated by a polynomial, where r is the distance between the image point and the principal point, k_i are the radial *coefficients*, and k_1 is the first order radial distortion coefficient:

$$\Delta r = k_1 r^3 + k_2 r^5 + k_3 r^7 + \dots \quad (3.26)$$

EX1004, p.86. Corke adds that the *radial distortion coefficients* in the above equation “would be determined by a calibration procedure.” EX1004, p.86. A POSITA would have understood that the intrinsic calibration method discussed in Corke would determine a focal length, a first order radial lens distortion coefficient, and a set of coordinates of a center of radial distortion (*i.e.*, “the principal point”). EX1003, ¶¶143-144. As for [3], a POSITA would have understood that the “principal point” is the center of radial distortion because Corke discloses that “[r]adial distortion causes image points to be translated along radial lines from the

principal point,” noting that the principal point is “[t]he point where the camera’s optical axis intersects the image plane.” EX1004, p.86; EX1003, ¶144.

Corke discloses the experimental results of its calibration, which solves for scaling factor, focal length, and the coordinates for the principal point:

α_x X-axis scaling factor in pixels/mm (intrinsic)
 α_y Y-axis scaling factor in pixels/mm (intrinsic)
 X_0 image plane offset in pixels (intrinsic)
 Y_0 image plane offset in pixels (intrinsic)
 f focal length (intrinsic)

EX1004, p.137. These results are shown in Table 4.2:

Trial	Residual (pixel)	δ (deg)	X_0 (pixel)	Y_0 (pixel)	$\alpha_x f$ (pixel)	$\alpha_y f$ (pixel)	α_y/α_x
1	0.552	-0.0277	270.1	213.7	625.7	948.9	1.52
2	0.898	0.0672	271.4	223.1	614.0	933.0	1.52
3	0.605	0.0685	278.0	205.6	617.3	938.2	1.52
4	0.639	-0.1198	275.8	196.8	616.8	939.2	1.52
mean			273.8	209.8	615.5	939.8	
σ			3.2	9.7	5.0	6.6	

Table 4.2: Summary of calibration experiment results.

EX1004, p.146; EX1003, ¶¶145-146.

Corke in view of the knowledge of a POSITA renders obvious Claim 8.

EX1003, ¶¶140-147.

G. Claim 9

As discussed above for Claim 2, Corke discloses determining a set of extrinsic parameters of the camera from at least one of the number of images of the calibration

object captured by the camera, the set of extrinsic parameters comprising a camera space-to-training space transformation defining a transformation between a camera space reference frame and a training space reference frame. EX1003, ¶¶148-153.

Corke also discloses an extrinsic calibration that determines a respective translation component along three orthogonal axes, and a respective rotation component about the three orthogonal axes. EX1003, ¶150. Corke explains that “[r]obots typically have 6 degrees of freedom (DOF), allowing the end-effector to achieve, within limits, any pose in 3D space.” EX1004, p.153. The 6 degrees include **“3 in translation and 3 in rotation.”** EX1004, pp.10. Corke discloses a “calibration matrix” that determines “6 extrinsic parameters”—including the “3D vector” (x, y, z coordinates) and “3 angles for the camera pose.” EX1004, p.144.

Corke defines “3-D pose” as “comprising translation along, and rotation about the X, Y, and Z axes.” EX1004, p.xi. Figure 4.7 shows the X, Y, and Z axes that make up the three translational and three rotational components of the 3D pose estimation are orthogonal to each other:

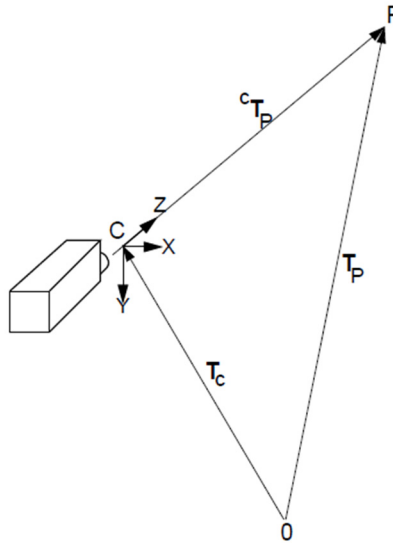


Figure 4.7: Relevant coordinate frames.

EX1004, p.139; EX1003, ¶151. A POSITA would have understood that a 3D orientation can be decomposed into three rotations about orthogonal axes, such as is obtained using an Euler angle representation or a roll-pitch-yaw angle representation. EX1003, ¶152. That a translation in 3D can be decomposed into translation along three orthogonal axes is well known Euclidean geometry, which would have been understood by a POSITA. EX1003, ¶152.

Corke in view of the knowledge of a POSITA renders obvious Claim 9. EX1003, ¶¶148-153.

H. Claim 10

Corke discloses determining a camera space-to-tool space transformation based on at least two images of the calibration object captured by the camera of the calibration object. EX1003, ¶¶154-156. Corke discloses that “[o]nce the target pose

relative to the camera is determined, it is then necessary to determine the target pose relative to the robot's end-effector. This requires robot eye-hand calibration." EX1004, p.160. Corke explains that "[r]obot eye to hand calibration is the process of determining the fixed transformation between the gripper and the camera coordinate system"—the camera space-to-tool space transformation—and involves "the robot making a number of moves and measuring the change in image plane coordinates of a fixed target." EX1004, p.147. This requires at least two images be captured of the "calibration target." EX1004, p.147; EX1003, ¶155. Corke in view of the knowledge of a POSITA renders obvious Claim 10. EX1003, ¶¶154-156.

I. Claims 17, 24, and 28

Corke discloses, or at least renders obvious, adjusting the position of the robot and capturing additional two-dimensional image of the volume containing the target object if the number of features located in the original captured image of the target object is determined to be insufficient. EX1003, ¶¶157-161. Corke explains that image features are chosen and extracted "on the basis of a confidence measure computed from the feature window." EX1004, p.168. Corke also states that selecting the appropriate threshold is crucial, and discloses algorithms that select which "measurable features give [the] best control." EX1004, pp.125, 165. It would have been obvious to a POSITA based on Corke that certain features should be ignored or given less weight than others during pose estimation, and that when the number

of viable features drops below the required minimum, it is necessary to obtain additional feature points. EX1003, ¶158.

Corke further discloses that while “at least three feature points” are needed to solve for pose, “[s]ix or more points always yield unique solutions, as well as the intrinsic calibration parameters.” EX1004, pp.159-60. Thus, Corke teaches a minimum number of features should be used for successful pose estimation. EX1003, ¶159.

Corke does not expressly disclose adjusting the position of the camera to take additional images if an insufficient number of features are found and capturing an additional two-dimensional image of the volume containing a target object, but, as described above, Corke acknowledges that a certain minimum number of features are required to estimate pose and that certain features should be ignored or given less weight in the analysis. EX1004, pp.159–160, 125, 165. Given the disclosures in Corke, it would have been obvious to a POSITA to move the camera to a different position if the number of features is inadequate capture another image and to determine pose. EX1003, ¶160.

Corke in view of the knowledge of a POSITA renders obvious claims 17, 24, 28. EX1003, ¶¶157-161.

J. Claim 18

As discussed above, Corke discloses determining an object space-to-camera space transformation of a target object. EX1003, ¶¶162-165. Section 5.4 of Corke says the ROBOT is “normal” to the object. EX1004, p.161. A POSITA would have understood, from this paragraph and Figure 5.6, that the robot AND the end-effector mounted camera would be normal to the object:

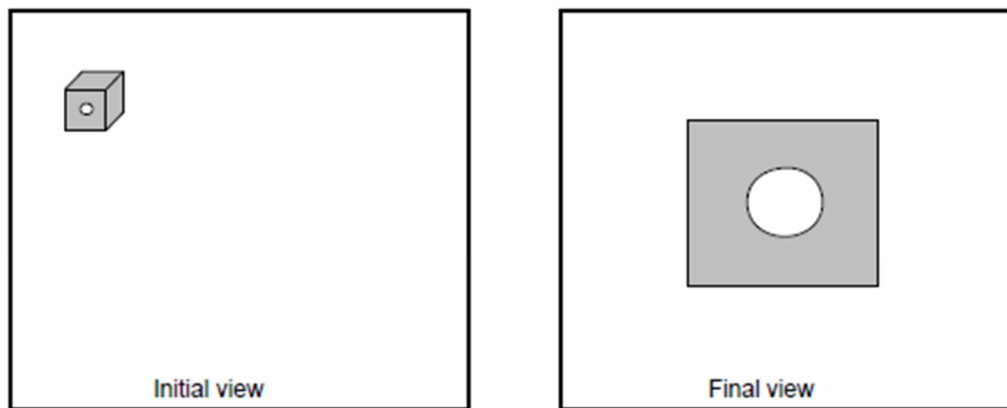


Figure 5.6: Example of initial and desired view of a cube

EX1004, p.161; EX1003, ¶162.

Corke describes methods to determine “the relative pose between the camera and the target, ${}^c\mathbf{x}_t$.” EX1004, p.152; EX1003, ¶163. The object space-to-camera space transformation is depicted in Figure 5.1 of Corke, where $\mathbf{P}$ designates the target object and ${}^c\mathbf{x}_t$ is the transformation between the camera and the target object—the “object of interest”:

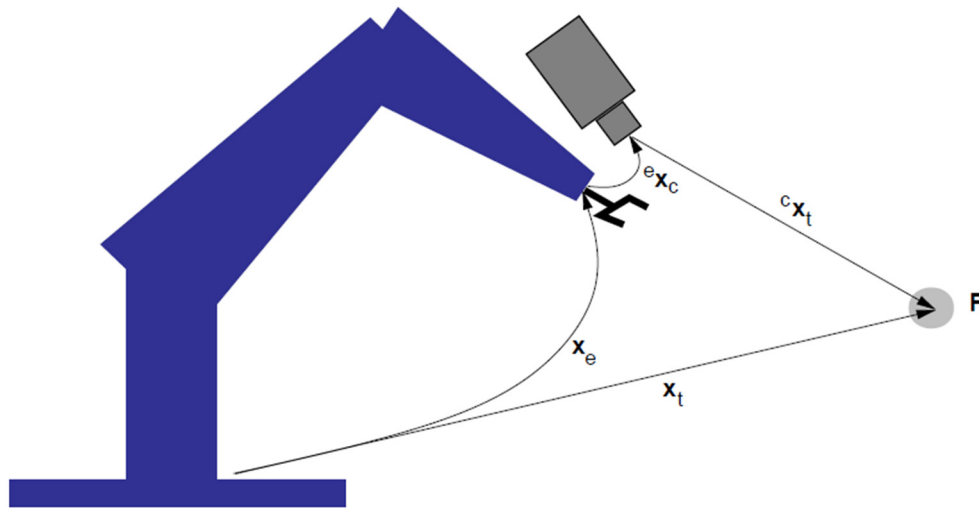


Figure 5.1: Relevant coordinate frames; world, end-effector, camera and target.

EX1004 pp.152, 138.

Corke also discloses determining at least one movement of the robot based on the object space-to-camera space transformation that orients the camera orthogonally with respect to the target object. EX1003, ¶164. According to Corke, “sensory data”—including a transformation determined by the disclosed algorithms—can modify a robot’s movements. EX1004, p.179. Corke discloses “[m]oving [the robot] to a position defined in terms of [a] **transformation**” such that the robot’s motion is “controlled by the vision system.” EX1004, p.179. From Corke, a POSITA would have understood that a camera could be moved orthogonally to the target object based on the object pose determined from the object space-to-camera space transformation. EX1003, ¶164. As described above, this could be

accomplished, for example, using visual servo control as illustrated in Figure 5.6.

EX1004, p.161; EX1003, ¶164.

Corke in view of the knowledge of a POSITA renders obvious Claim 18.

EX1003, ¶¶162-165.

K. Claim 19

Corke discloses determining a position of the object frame in the tool space reference frame. EX1003, ¶¶166-171. For instance, for Corke's visual servoing methods it is "necessary to determine the target pose relative to the robot's end-effector"—*i.e.*, the position of the object frame in the tool space reference frame. EX1004, p.160; EX1003, ¶167. Figure 5.1 of Corke depicts the location and orientation of an object (P) in relation to a tool space reference frame (e):

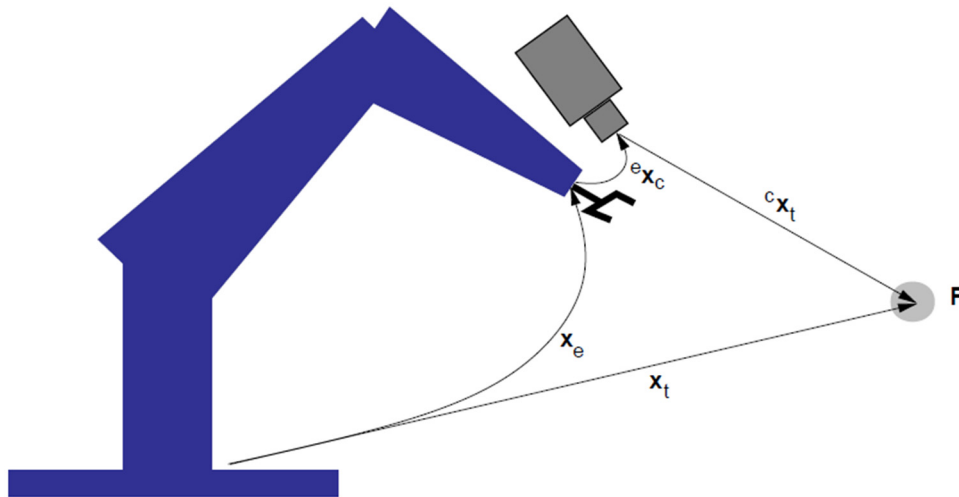


Figure 5.1: Relevant coordinate frames; world, end-effector, camera and target.

EX1004, p.152; EX1003, ¶167. A POSITA would have understood that once the object space-to-camera space and the camera space-to-tool space transformations are determined, the position of the object frame in the tool space reference frame is necessarily determined. EX1003, ¶168.

Corke also discloses providing the object frame to the robot. EX1004, p.179. According to Corke, “sensory data”—including an object frame determined by the disclosed transformations—can modify a robot’s movements. EX1004, p.179; EX1003, ¶169. Corke discloses “[m]oving [the robot] to a position defined in terms of [a] transformation” such that the robot’s motion is “controlled by the vision system.” EX1004, p.179.

Figure 5.4 shows that in Corke’s position-based visual servoing method, the target object’s “pose estimation”—the object frame—is sent to the robot:

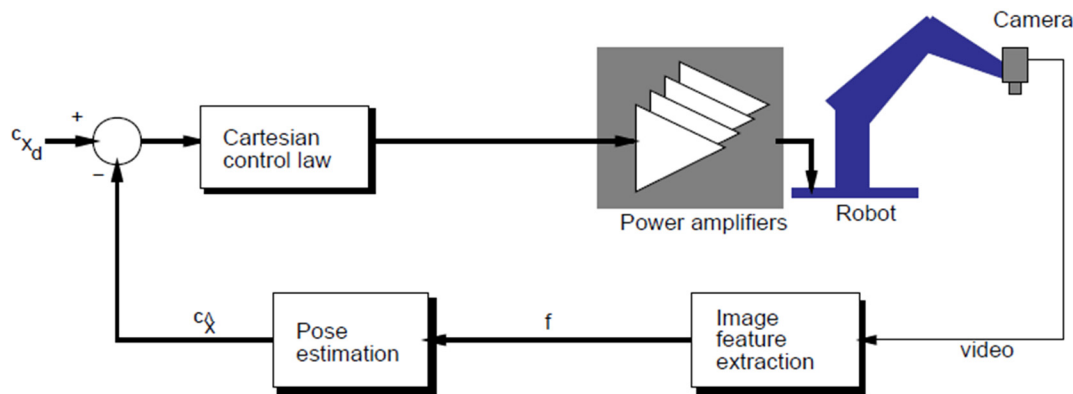


Figure 5.4: Position-based visual servo (PBVS) structure as per Weiss.

EX1004, p.155; EX1003, ¶170.

Corke in view of the knowledge of a POSITA renders obvious Claim 19.

EX1003, ¶¶166-171.

L. Claims 20 and 25

[20.0] An apparatus useful in robotics, the apparatus comprising:

[20.1] a single camera operable to capture a number of images of a calibration object;

[25.0] [similar to [20.0]]

[25.1] [similar to 20.1]

Regardless of whether the preamble is limiting, Corke discloses an apparatus useful in robotics. EX1003, ¶¶172-176. Corke describes a “computer vision system” using “one or more cameras” to “control the position of the robot’s end-effector relative to the workplace as required by the task,” “allow[ing]” robots to perform “tasks that were not strictly repetitive.” EX1004, pp.1-2; EX1003, ¶173.

Corke also discloses a single camera operable to capture a number of images of a calibration object. EX1003, ¶174. The single camera used in Corke’s experiments is shown in Figures 4.11, 6.10, and 6.11:

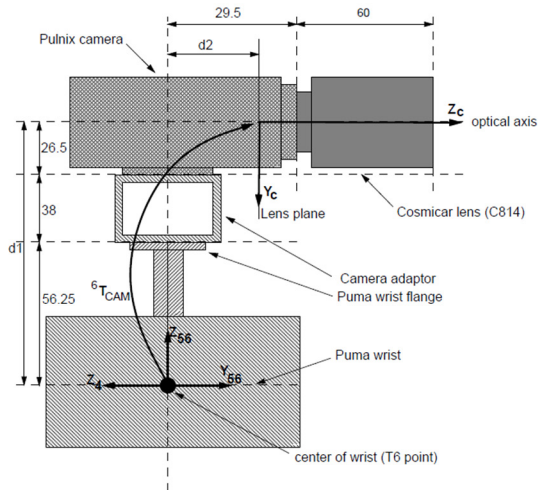


Figure 4.11: Details of camera mounting (not to scale, dimensions in mm). Robot wrist is in the zero angle pose.

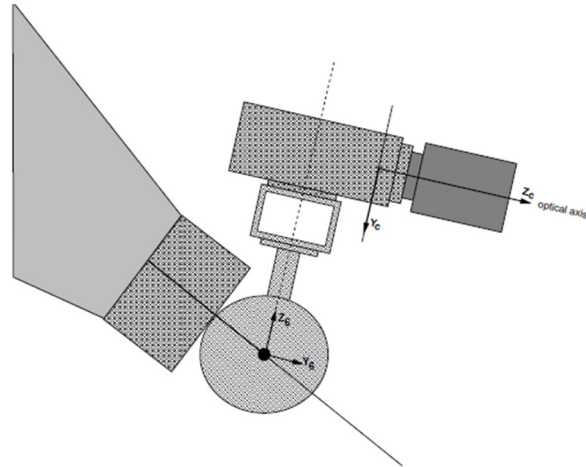


Figure 6.10: Camera mount used in this work. In this pose $\theta_6 \approx -\pi/2$.

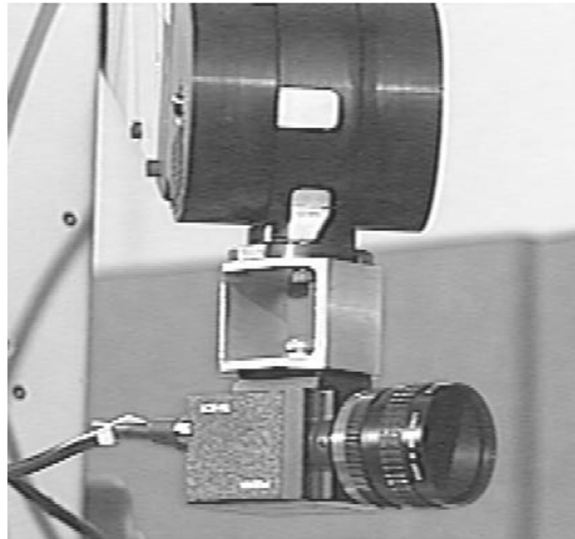


Figure 6.11: Photograph of camera mounting arrangement.

EX1004, p.148, 186-87. Using this single camera, Corke states that “[f]our images of the SHAPE calibration target...were taken from different viewpoints.” EX1004, p.145. Figure 4.8 shows one such image:

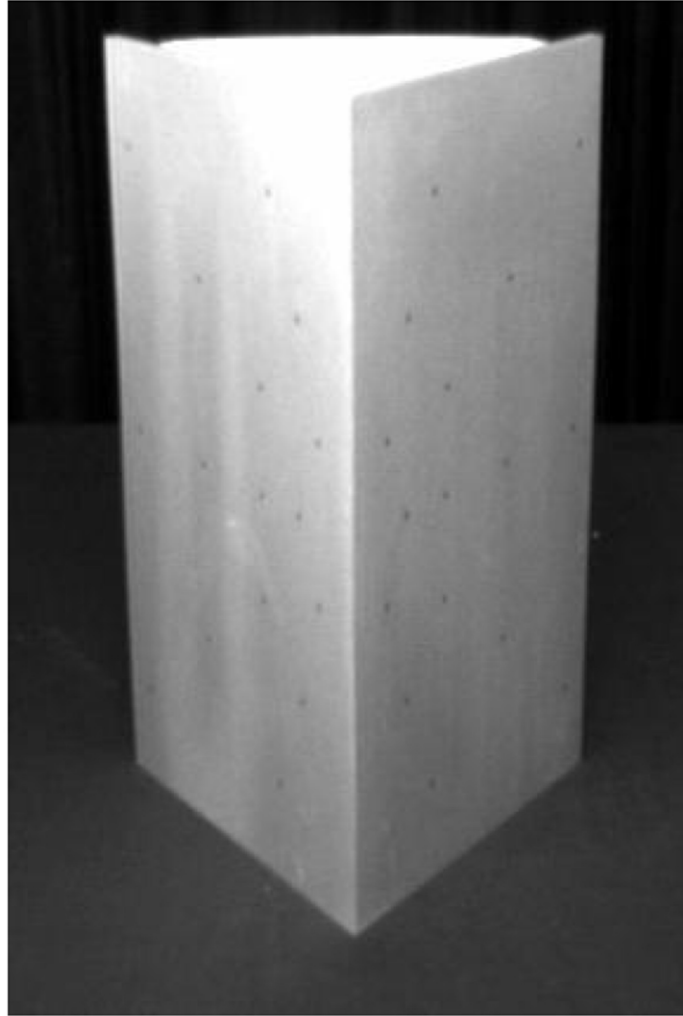


Figure 4.8: The SHAPE calibration target used for intrinsic parameter determination. The surface is brushed aluminium with black marker dots at accurately known locations on the two surfaces.

EX1004, p.147.

Corke in view of the knowledge of a POSITA renders obvious Elements

[20.0], [20.1], [25.0], and [25.1]. EX1003, ¶¶172-176.

[20.2] means for calibrating the camera, by: determining a set of intrinsic parameters of the camera from at least one of the number of images of the calibration object captured by the camera; and determining a set of extrinsic parameters of the camera from at least one of the number of images of the calibration object captured by the camera,

[25.2] [similar to 20.2]

Corke discloses determining a set of intrinsic parameters and extrinsic parameters from at least one of the number of images of the calibration target captured by the camera. EX1003, ¶¶177-180. In Corke's experiments, "[f]our images of the SHAPE calibration target, shown in Figure 4.8, were taken from different viewpoints." EX1004, p.145. Corke explains that the four images were used to compute the "calibration matrix"—a mathematical object that "contains information about the position of the camera with respect to the world coordinate frame"—the extrinsic parameters—"as well as intrinsic parameters of the camera." EX1004, pp.143-144 (the "calibration matrix" "encodes the 6 extrinsic parameters" ("3D position vector and 3 angles for camera pose") and "5 intrinsic parameters: f , α_x , α_y , X_0 and Y_0 "). By comparing the location of the calibration object's features in the images to their "known 3D marker location," Corke determines the extrinsic and intrinsic parameters. EX1004, pp.139-142; EX1003, ¶178. Additionally, Corke defines camera calibration generally as "the process of determining the internal camera geometric and optical characteristics (intrinsic parameters) and the 3-D position and orientation of the camera frame relative to the world coordinate system (extrinsic parameters)." EX1004, p.139 ("the camera calibration matrix which relates a world coordinate system to the image plane coordinates for a given camera position and orientation."). EX1003, ¶179.

Corke in view of the knowledge of a POSITA renders obvious Elements [20.2] and [25.2]. EX1003, ¶¶177-180.

[20.3] the set of extrinsic parameters comprising a camera space-to-training space transformation defining a transformation between a camera space reference frame and a training space reference frame; and

[25.3] [similar to 20.3]

Corke discloses that extrinsic parameters include a transformation between a camera space reference frame and a training space reference frame. EX1003, ¶¶181-183. Specifically, Corke teaches that “*extrinsic* parameters” relate to “the position and orientation of the camera.” EX1004, p.138. One “extrinsic” parameter is ${}^C T_P$ —“the position of the [calibration] object with respect to the camera.” EX1004, p.127. A POSITA would have understood this was the “transformation”— T —between the coordinate frame defined by the camera— C —and the coordinate frame defined by the calibration object— P . EX1004, p.xi; EX1003, ¶182. Because a “training space” is a “reference frame defined with respect to a point on the calibration” object, a POSITA would have understood that Corke discloses extrinsic parameters, including a camera space-to-training space transformation. EX1001, 3:22-24; EX1003, ¶182; EX1004, p.3 (the transformations described in Corke provide “positioning information” “directly in Cartesian or task space”—*i.e.*, training space).

Corke in view of the knowledge of a POSITA renders obvious Elements [20.3] and [25.3]. EX1003, ¶¶181-183.

[20.4] means for estimating a pose of a target object, by: capturing a two-dimensional image of a volume containing a target object;

[25.4] [similar to 20.4]

Corke discloses means for estimating a pose of a target object. EX1003, ¶¶184-191. Corke discloses that in “*position based* control, features are extracted from the image” of a target object to “***determine the pose of the target*** with respect to the camera.” EX1004, pp.152-53; EX1003, ¶185.

As shown in Figure 5.4, Corke discloses using a video camera to capture and process multiple two-dimensional images of a target object one image at a time at the “camera frame rate,” which is “effectively the sample rate in a visual servoing system.” EX1004, p.103, 73 (“the camera’s shutter acts as the sampler in a visual control loop”):⁷

⁷ The portions of Corke cited herein are from a single embodiment as Chapter 3’s “Fundamentals of image capture” and Chapter 4’s “Machine vision” concepts are used as a baseline for Chapter 5’s position-based visual servoing methods. EX1004.

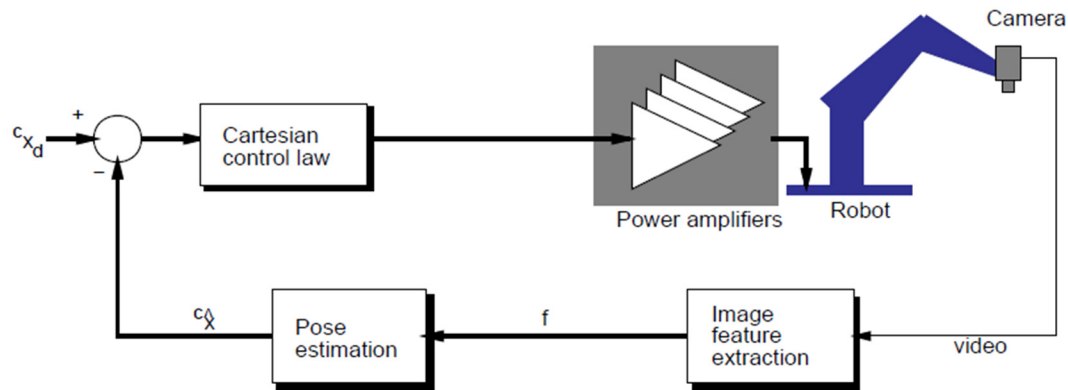


Figure 5.4: Position-based visual servo (PBVS) structure as per Weiss.

EX1004, p.155; EX1003, ¶186.

The claim requires that the target object reside in a volume, which Corke teaches is the space from which the image of the target object is captured. EX1004, pp.152-153 (referring to “an image of the target” as “the object of interest, that is, the object that will be tracked,” so that “features” can be “extracted from the image”); EX1003, ¶187. A POSITA would have understood that any 3D object occupies a volume. EX1003, ¶187. Corke discloses that “[t]he camera contains a lens which forms a 2D projection of the scene on the image plane.” EX1004, p.152. Corke’s teaching that capturing a 2D image of a target object by a single camera would have included capturing a two-dimensional image of a volume containing a target object. EX1003, ¶187.

Corke in view of the knowledge of a POSITA renders obvious Elements [20.4] and [25.4]. EX1003, ¶¶184-191.

[20.5] locating at least six features in the captured image of the target object;

[25.5] [similar to 20.5]

Corke discloses locating at least 5 or 6 features on the target object. EX1003, ¶¶189-191. Indeed, Corke states that locating at least 6 features is preferred since that will yield “unique solutions” to pose estimation:

It can be shown that at least three feature points are needed to solve for pose. Intuitively, the coordinate of each feature point yields two equations and six equations are needed to solve for the six elements of the pose vector. In fact three feature points will yield multiple solutions, but four coplanar points yield a unique solution...***Six or more points always yield unique solutions***, as well as intrinsic camera calibration parameters.

EX1004, pp.159-60. Corke in view of the knowledge of a POSITA renders obvious Elements [20.5] and [25.5]. EX1003, ¶¶189-191.

[20.6] and determining an object space-to-camera space transformation based at least in part on a position of at least some of the located features in solely the captured image using an algorithm that employs a known or determinable physical relationship between at least some of the located features.

[25.6] [similar to 20.6]

Corke discloses determining an object space-to-camera space transformation for the target object utilizing features from a single image. EX1003, ¶¶192-200. Corke discloses that in “*position based* control, features are extracted from ***the image***” of the target object “and used in conjunction with a geometric model of the

target to determine the pose of the target with respect to the camera.” EX1004 p.153.

A POSITA would have understood that determining the target’s pose (*i.e.*, position and orientation) with respect to the camera is the same as determining the object space-to-camera space transformation. EX1003, ¶193. This transformation is depicted in Figure 5.1, where **P** designates the target object and ${}^c\mathbf{x}_t$ is the transformation between the camera and the target object—the “object of interest”:

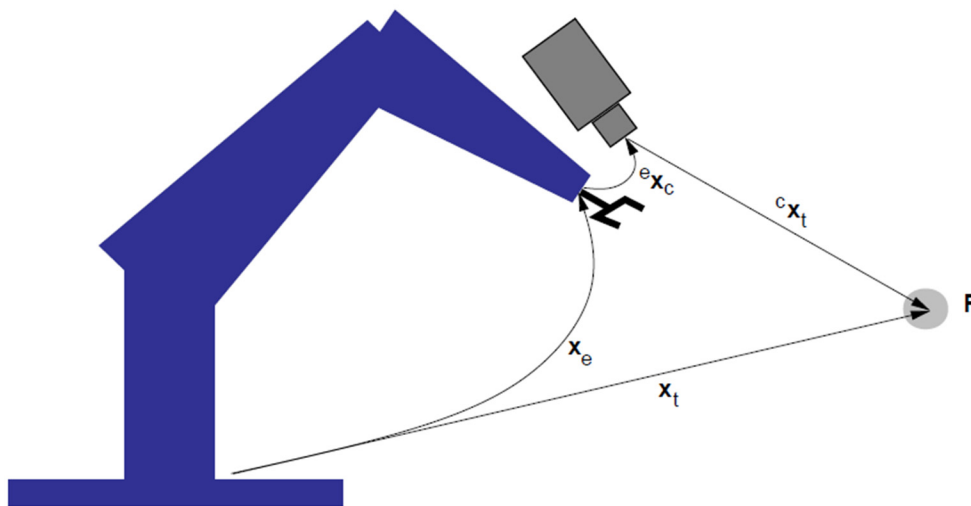


Figure 5.1: Relevant coordinate frames; world, end-effector, camera and target.

EX1004 pp.152, 138; EX1003, ¶193.

A POSITA would also have understood that Corke’s references to “*the image*” teaches that only features from a single captured image are used to determine the object space-to-camera space transformation. EX1004, pp.152 (“The task in visual servoing is to control the *pose* of the robot’s end-effector, ${}^x\mathbf{T}_6$, using visual information, *features*, extracted from *the image*.”); 153 (“*The image plane*

coordinates of three or more visual features can be used to determine the pose...of the target relative to the camera, given knowledge of the geometric relationship between the feature points.”); EX1003, ¶194. As an example of single-image processing, Corke teaches use of the Datacube DIGIMAX for acquiring individual images from the camera and processing these individual images to locate features, stating “[t]he incoming analog video signal is digitized to form a *512x512 pixel digital image*,” after which various features are detected, such as “moments up to second order, perimeter and bounding box,” EX1004, p.180. EX1003, ¶194.

Corke teaches an algorithm that uses the position of features from a single image to determine the object space-to-camera space transformation. EX1004, p.138; EX1003, ¶195. Corke teaches in equation 4.60 that camera calibration matrix **C** “represents the relationship between 3-D world coordinates and their corresponding 2-D image coordinates as seen by the computer,” with *i***x**, *i***y** and *i***z** encoding the position of a feature in the image plane, and **x**, **y**, and **z** encoding the position of the same feature in 3D “world coordinates”:

$$\begin{bmatrix} i_x \\ i_y \\ i_z \end{bmatrix} = \mathbf{C} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad (4.60)$$

and with the matrix **C** defined in equation 4.61:

$$\mathbf{C} = \begin{bmatrix} \frac{\alpha_x n_x f - X_0 n_z}{f - c_z} & \frac{\alpha_x \sigma_x f - X_0 \sigma_z}{f - c_z} & \frac{\alpha_x \alpha_x f - X_0 a_z}{f - c_z} & \frac{\alpha_x c_x f - X_0 c_z + X_0 f}{f - c_z} \\ \frac{\alpha_y n_y f - Y_0 n_z}{f - c_z} & \frac{\alpha_y \sigma_y f - Y_0 \sigma_z}{f - c_z} & \frac{\alpha_y a_y f - Y_0 a_z}{f - c_z} & \frac{\alpha_y c_y f - Y_0 c_z + Y_0 f}{f - c_z} \\ -\frac{n_z}{f - c_z} & -\frac{\sigma_z}{f - c_z} & -\frac{a_z}{f - c_z} & 1 \end{bmatrix} \quad (4.61)$$

EX1004, pp.138. Corke teaches methods for determining matrix $\mathbf{C}$ using the position of multiple features in a single image. EX1004, pp.139-142; EX1003, ¶196. Corke teaches that “calibration matrix, $\mathbf{C}$, encodes the 6 extrinsic parameters,” noting that these extrinsic parameters are the “3D position vector and 3 angles for camera pose.” EX1004, p.144. Corke teaches an algorithm to recover these parameters from the matrix $\mathbf{C}$ by decomposing the camera calibration matrix. EX1004, pp.143-144; EX1003, ¶196. A POSITA would have understood that if the 3D coordinates are specified with respect to the object space, then the camera pose obtained by Corke’s algorithm is defined with respect to the object space, and therefore determines the object space-to-camera space transformation. EX1003, ¶196.

Using Corke’s equation 4.60, a POSITA would have understood that by multiplying the camera calibration $\mathbf{C}$ by the 3D coordinates of any feature in 3D space, the corresponding position for that feature in the 2D image frame— ${}^i\mathbf{x}$, ${}^i\mathbf{y}$, ${}^i\mathbf{z}$ —are obtained, and those coordinates can be converted to two-dimensional camera frame pixel coordinates— ${}^iX$, ${}^iY$ —using Corke’s equations 4.56 and 4.57:

$${}^iX = \frac{{}^ix}{{}^iz} \quad (4.56)$$

$${}^iY = \frac{{}^iy}{{}^iz} \quad (4.57)$$

EX1004, p.138; EX1003, ¶197. Furthermore, Corke discloses that this algorithm works with features extracted using only the single captured image and without additional captured images. EX1004, p.138; EX1003, ¶197.

The algorithm disclosed in Corke also employs a known physical relationship between some located features using stored knowledge of model feature points in 3D space. EX1004, p.159; EX1003, ¶198. Corke explains that to “determine the 3D relative pose of an object...from 2D image coordinates,” “additional information is needed.” EX1004, p.159. This includes “*knowledge of the relationship between the observed feature points* (perhaps from a CAD model) and also the camera’s intrinsic parameters.” EX1004, p.159, 153 (“three or more visual features can be used to determine the pose...*given knowledge of the geometric relationship between the feature points*”). Corke also explains that the object space-to-camera space transformation is determined by an algorithm using the position of “features extracted from the image” “*in conjunction with a geometric model of the target,*” and “each feature point yields two equations and six equations are needed to solve the six elements of the pose vector.” EX1004, pp.153, 159-60; EX1003, ¶198. A POSITA would have understood that Corke teaches an algorithm that employs “features extracted from the image” (only the single captured image) and “a geometric model of the target” (known physical relationship) to determine the “3D relative pose of an object” (object space-to-camera space transformation for the

target object). EX1004, pp. 153, 159-60 (“features are extracted from the image and used in conjunction with a geometric model of the target to determine the pose of the target with respect to the camera”); EX1003, ¶198. Corke also discloses that the object space-to-camera space algorithm can be performed “by a computer,” which a POSITA would have understood includes the claimed processor. EX1004, p.138; EX1003, ¶199.

Corke in view of the knowledge of a POSITA renders obvious Elements [20.6] and [25.6]. EX1003, ¶¶192-200.

M. Claims 26 and 27

Corke discloses using at least one programmed computer and/or computer-readable medium for calibration and pose estimation. EX1003, ¶¶201-205. For example, Corke discloses that visual servoing is a method used for “computer vision”—the “application of a computer system for receiving and processing information,” including image interpretation. EX1004, p.123. Corke also discloses a calibration matrix run on a computer. EX1003, ¶202. This matrix is a “3x4 homogenous transform which performs scaling, translation, and perspective” and “represents the relationship between 3-D world coordinates and their corresponding 2-D image coordinates as seen by the computer.” EX1004, pp.138, 167 (“image processing” is handled “by a conventional computer”).

One of the computer systems used by Corke is shown in Figure 6.2:

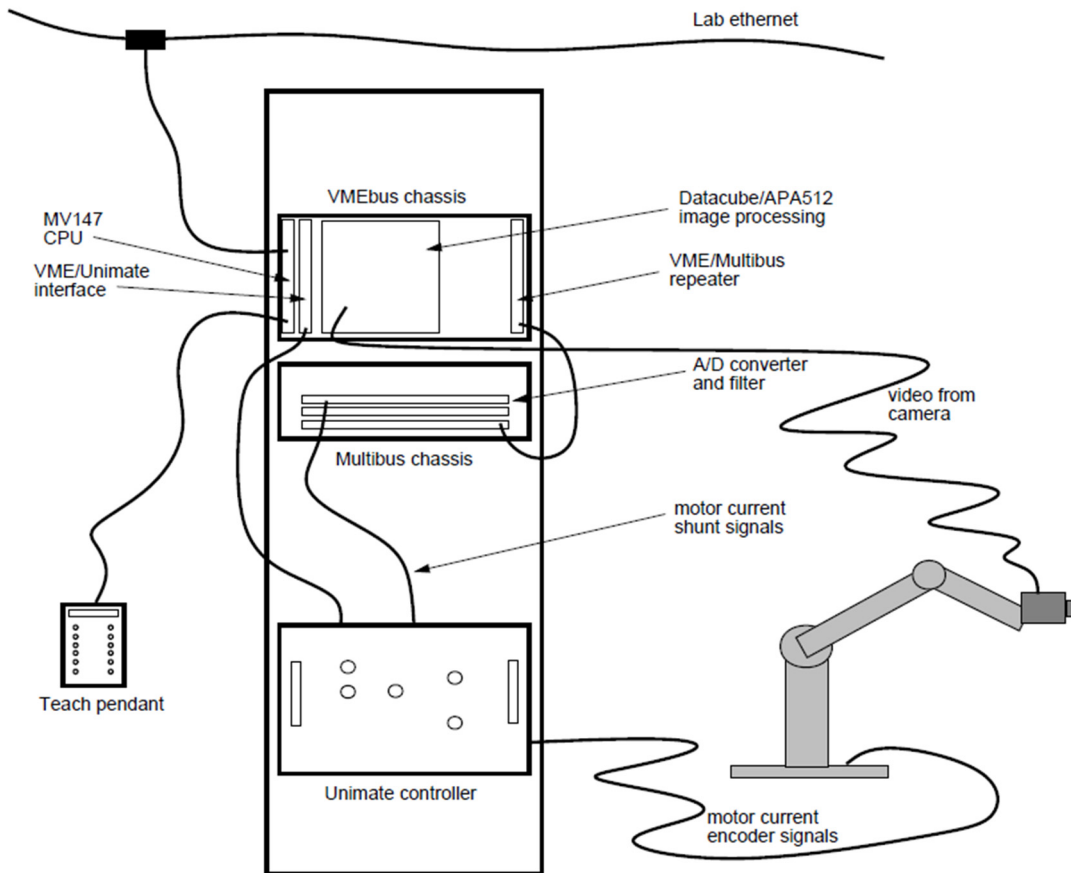


Figure 6.2: Overall view of the system used for visual servo experiments.

EX1004, p.176. The computer of Figure 6.2 processes and uses information received from the robot to calculate calibration parameters and determine pose. EX1003, ¶203. Corke in view of the knowledge of a POSITA renders obvious Claims 26 and 27. EX1003, ¶¶201-205.

XI. GROUND 2: CLAIMS 5, 12-16, AND 21-24 ARE UNPATENTABLE AS OBVIOUS OVER CORKE IN VIEW OF WEI-I

A. Wei-I

Wei-I discloses “a method for self-calibration of robotic hand cameras by means of active motion.” EX1005, p.1; EX1003, ¶¶82-83. The method involves

“tracking a set of world points” on an object “during robot motion” to determine “the camera internal and external parameters” and “hand-eye calibration.” EX1005, pp.1-2.

Figure 4.7 shows the object used to perform self-calibration in Wei-I, the “Orbital Replacement Unit (ORU),” which was an “octagonal shape[d]” object used in a “1993 German Spacelab mission D2” during which it was “grasped in a predefined position by a space robot using telesensor programming.” EX1005, p.279:

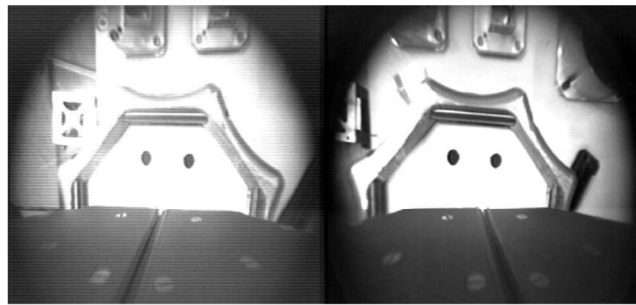


Figure 7: The images of the ORU from the hand cameras. The two black blobs are used as calibration points.

EX1005, p.15.

B. Motivation to Combine Corke and Wei-I

As noted above, Corke’s pose estimation methods require intrinsic, extrinsic, and hand-eye calibration. EX1004, pp.136-147. Similarly, Wei-I states that “[i]n order to use cameras for estimating robot motion for object manipulation, it is usually necessary to do the following three calibrations: camera calibration, hand-eye calibration, and robot calibration.” EX1005, p.2. Wei-I explains that

“camera calibration” “determines the camera internal and external parameters.”

EX1005, p.2; EX1003, ¶¶207-208.

Wei-I “propose[s] a complete autonomous approach for self-calibration of hand cameras and hand-eye relationships” that has several advantages over other calibration methods, including (1) no control points (*e.g.*, calibration plates), (2) the initial values of all unknowns are found automatically, (3) it accounts for lens distortion, and (4) it does not rely on any prior calibration. EX1005, p.3; EX1003, ¶¶209-210.

A POSITA would have been motivated to combine the teachings of Corke (which disclosed methods and systems that require calibration) with those of Wei-I (which provides an advantageous self-calibration method). EX1003, ¶¶207-213. This is because Wei-I discloses a way for a system to self-calibrate (using a teaching object) for each of the intrinsic, extrinsic, and hand-eye calibration requirements of Corke. EX1004, pp.136-147; EX1005, pp.1-3; EX1003, ¶211. The combination of Corke and Wei-I would have used Wei-I’s known calibration techniques to improve Corke’s similar device and methods in the same way as it improved Wei-I’s system. EX1003, ¶211.

A POSITA would have understood that Wei-I discloses a streamlined way to perform the calibration required by Corke. EX1003, ¶212. Whereas Corke requires use of a “target object” for calibration, Wei-I performs self-calibration using a

teaching object of the type a robot will manipulate during calibration. EX1005, p.1.

However, a POSITA would have understood that a “teaching object” is just a subset of a “target object” that is used specifically for self-calibration. EX1003, ¶212. A “target object” is simply the object that is being used to calibrate the robot. EX1003, ¶212. In a method of “self” calibration, like in Wei-I, that “target” object is known as a “teaching” object because the features of the object are “determinable”—*i.e.*, not already known, but can be “taught.” EX1003, ¶212. The combination of Corke with Wei-I would eliminate the need for a calibration plate and would streamline, but not otherwise change, Corke’s calibration process. EX1003, ¶¶213.

C. Claims 5 and 12

[5] The method of claim 3 wherein positioning the camera with respect to the calibration object comprises positioning the camera with respect to a sample of a type of object the robot will manipulate, the sample having a number of features, where a known or determinable physical relationship exists between at least some of the features.

[12.1] The method of claim 1, further comprising: capturing an image of a teaching object of a type of object that will be manipulated by the robot;

[12.2] selecting a number of features from the captured image of the teaching object

[12.3] determining a set of object space coordinates for each of the selected features from the captured image of the teaching object.

As explained above, Corke discloses or renders obvious all limitations of Claims 1 and 3, from which Claims 5 and 12 depend. Corke does not specifically disclose using a teaching object of the type a robot will manipulate during calibration

as recited in Claims 5 and 12. Yet Corke does disclose the use of a target object that has features where a known or determinable relationship exists between the features, and teaching objects and target object are the same—they are both objects “that will be manipulated by the robot.” EX1004, p.152; EX1003, ¶215. As described above, a POSITA would have found it obvious to improve Corke’s calibration techniques (using a calibration object) with Wei-I’s calibration techniques (using a “teaching object of the type a robot will manipulate”). EX1003, ¶¶214-224. This combination would eliminate the need for a calibration plate and would streamline, but not otherwise change, Corke’s calibration process discussed above. EX1003, ¶216. A POSITA would have found it obvious to improve Corke’s calibration techniques with Wei-I’s teachings regarding the use of a teaching object of the type a robot will manipulate in place of a calibration object. EX1003, ¶216.

The combination of Corke and Wei-I would have used Wei-I’s known calibration techniques, including a teaching object of a type of object that will be manipulated by the robot, to improve Corke’s similar device and methods by eliminating the need for a specially constructed calibration target, thus improving Corke in the same way that it improved Wei-I’s techniques. EX1003, ¶216.

“[I]f a technique has been used to improve one device, and a person of ordinary skill in the art would recognize that it would improve similar devices in the same way, using the technique is obvious.” *KSR*, 550 U.S. at 417. In other words,

this combination would have involved “the mere application of a known technique”—Wei-I’s teachings regarding the use of a teaching object—“to a piece of prior art”—Corke’s calibration techniques—“ready for the improvement.” *Id.*; EX1003, ¶217. Furthermore, Corke states that “[c]alibration using non-coplanar points requires a mechanically complex calibration frame,” which provides additional motivation for the use of a teaching object over a specially constructed calibration target. EX1004, p.141; EX1003, ¶216.

Wei-I discloses positioning the camera with respect to a sample of a type of object the robot will manipulate during the calibration step and capturing an image of a teaching object of a type of object the that will be manipulated by the robot. EX1003, ¶218. Wei-I discloses a method “for self-calibration of robotic hand cameras” using a teaching object where a determinable—rather than known—physical relationship exists between at least some of the features, eliminating the need for a calibration plate. EX1005, p.1; EX1003, ¶218. As described above, a POSITA would have understood that a “teaching object” is just a subset of a “target object” that is used specifically for self-calibration. EX1003, ¶218. In a method of “self” calibration, like in Wei-I, that “target” object is known as a “teaching” object because the features of the object are “determinable”—*i.e.*, **they are not already known, but can be “taught.”** EX1003, ¶218.

Wei-I discloses using an ORU “for camera calibration,” which is an “octagonal shape” that was “to be grasped in a predefined position by a space robot using telesensor programming.” EX1006, pp. 13, 279. The ORU was a sample of the type of object the robot will manipulate. EX1003, ¶219.

The ORU has a number of features where a discernable physical relationship exists between some of the features. EX1003, ¶220. Wei-I discloses selecting a number of features from the captured image of the teaching object. EX1003, ¶220. Wei-I discloses “two black blobs artificially marked on the object” used as calibration points. EX1005, p.13. Wei-I also discloses capturing an image of a teaching object. EX1003, ¶220. As described above, Wei-I’s “teaching object” is simply one type of “target object” like the one disclosed in Corke. EX1003, ¶220. Figure 7 of Wei-I shows images captured of the ORU teaching object with the two black blobs:

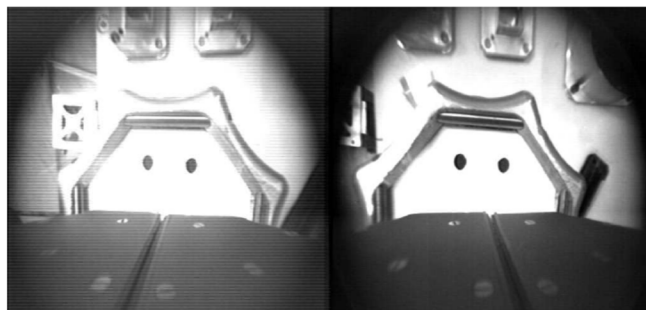


Figure 7: The images of the ORU from the hand cameras. The two black blobs are used as calibration points.

EX1005, p.15. Wei-I explains that by tracking the black blobs “at various positions and orientations of the robot,” “the camera’s internal parameters (including distortion coefficients) and the mounting parameters can be calibrated.” EX1005, pp.15-16. EX1003, ¶221. Given the advantages of Wei-I’s teaching object techniques, it would have been obvious to a POSITA to improve Corke with Wei-I’s disclosures regarding capturing an image of a teaching object. EX1003, ¶221.

Wei-I also discloses determining a set of object space coordinates for each of the selected features from the captured image of the teaching object. EX1003, ¶222. Wei-I explains that by tracking the black blob features on the ORU teaching object “at various positions and orientations of the robot,” the “*world points*” (“3D points”) of the blobs were calculated. EX1005, pp.15-16. Since these 3D feature coordinates are defined with respect to the teaching object (the target object), these coordinates are in “object space.” EX1001, 3:25-26 (“object space” is “a reference frame defined with respect to a point on, and therefore rigid to, the object”); EX1003, ¶222. A POSITA would have understood that use of a teaching object is preferable to the use of a special purpose calibration target, and the methods disclosed in Wei-I enable the algorithms of Corke to function without need for a calibration target. EX1003, ¶222. In view of Wei-I’s disclosure, it would have been obvious to a POSITA to improve Corke with Wei-I’s disclosures regarding determining a set of object space coordinates for each of the selected features from the captured image of the teaching

object. EX1003, ¶¶222-223. Corke in view of Wei-I discloses Claims 5 and 12. EX1003, ¶¶214-224.

D. Claim 13

Corke in view of Wei-I discloses, or at least renders obvious, Claim 13. EX1003, ¶¶225-230. Corke discloses that position-based visual servoing selects features extracted from the image of a target object to determine its pose. EX1004, pp.152; 153 (“In position based control, features are extracted from the image and used in conjunction with a geometric model of the target to determine the pose of the target with respect to the camera.”); EX1003, ¶226.

Corke also discloses that while “at least three feature points” are needed to solve for pose, “[s]ix or more points always yield unique solutions, as well as the intrinsic calibration parameters.” EX1004, pp.159-60. Thus, according to Corke, use of six features is preferable. EX1003, ¶227.

Although Corke does not specifically disclose extracting features from an image of a *teaching* object, it discloses capturing images of a target object, extracting features, and determining the feature coordinates—and both teaching and target objects are types of objects a robot would manipulate. EX1004, pp.145, 152; EX1003, ¶228. For the reasons above, it would have been obvious to a POSITA to improve Corke’s calibration procedure, which preferably selects six or more

features, with Wei-I's disclosure of a using a teaching object during calibration described above. EX1003, ¶¶228-229.

Corke in view of Wei-I discloses Claim 13. EX1003, ¶¶225-230.

E. Claim 14

[14] The method of claim 12 further comprising: determining an object space-to-camera space transformation defining a transformation between an object space reference frame and the camera space reference frame.

Corke discloses determining an object space-to-camera space transformation of a target object, which, like a teaching object, is the type of object a robot would manipulate. *See, e.g.,* EX1004, p.152; EX1003, ¶¶231-236. The object space-to-camera space transformation is depicted in Figure 5.1 of Corke, where **P** is the object and ${}^c x_t$ is the transformation between the camera and the target—*i.e.*, the “object of interest”:

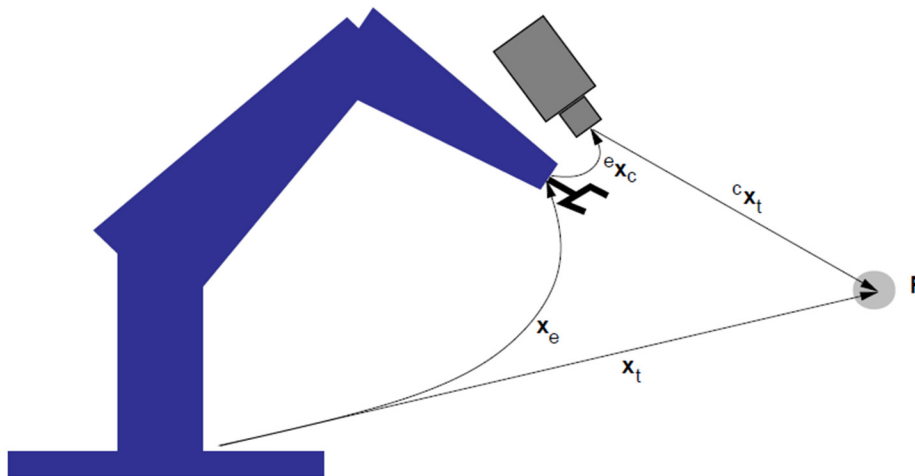


Figure 5.1: Relevant coordinate frames; world, end-effector, camera and target.

EX1004, p.152; EX1003, ¶232.

Corke does not expressly disclose determining an object space-to-camera space transformation using a teaching object, but as described above, it discloses determining an object space-to-camera space transformation using a target object. EX1004, p.153; EX1003, ¶233. For the same reasons, it would have been obvious to a POSITA to improve Corke with Wei-I's disclosure of using a teaching object for calibration. EX1003, ¶233.

Wei-I discloses a self-calibration method using a teaching object—the ORU—to determine a camera's extrinsic parameters, including the “transformation from the sensor space to the world space.” EX1005, p.15; EX1006, p.279. Using such a method, the “3D points” for the features on the ORU teaching object in a world space can be determined. EX1005, pp.7, 13-15. Since these 3D feature coordinates are defined with respect to the teaching object, they are in “object space.” EX1001, 3:25-26 (“object space” is “a reference frame defined with respect to a point on, and therefore rigid to, the object”); EX1003, ¶234. Thus, Wei-I teaches a camera space-to-object space transformation for a teaching object. EX1003, ¶234. Given the advantages of Wei-I's teaching object techniques, it would have been obvious to a POSITA to improve Corke with Wei-I's disclosures regarding a camera space-to-object space transformation for a teaching object. EX1003, ¶¶234-235.

Corke in view of Wei-I discloses Claim 14. EX1003, ¶¶231-236.

F. Claim 15

[15] The method of claim 14 further comprising: determining a position and an orientation of an object frame in the tool frame reference frame based at least in part on the object frame-to-camera space and camera space-to-tool space transformations.

Corke discloses determining object space-to-camera space and camera space-to-tool space transformations. EX1003, ¶¶237-242. Corke discloses that in “*position based* control, features are extracted from the image” of the target object “and used in conjunction with a geometric model of the target to determine the pose of the target with respect to the camera” (object space-to-camera space transformation) and the “pose of the camera” with respect to “the pose of the end-effector” (the camera space-to-tool space transformation). EX1004, p.152. Corke explains that “[o]nce the target pose relative to the camera is determined, it is then necessary to determine the target pose”—the position and orientation of the object frame—“relative to the robot’s end-effector”—*i.e.*, the tool reference frame—based at least in part on the object space-to-camera-space and camera-space-to-tool-space calibrations. EX1004, p.160. “This requires robot hand-eye calibration.” EX1004, p.160; EX1003, ¶238.

Corke does not expressly disclose determining an object space-to-camera space or camera space-to-tool space transformation using a teaching object, but as described above, it discloses those transformations using a target object. EX1004, p.152; EX1003, ¶239. For the same reasons stated above, it would have been obvious

to a POSITA to improve Corke with Wei-I's disclosure of using a teaching object to perform these transformations. EX1003, ¶239.

Wei-I discloses a self-calibration method using a teaching object to determine an object space-to-camera space transformation (extrinsic calibration) and a camera-space-to-tool-space transformation (hand-eye calibration). EX1005, pp.1-3. Corke states that calculating the hand-eye calibration necessarily involves a determination of "the target pose relative to the robot's end-effector." EX1004, p.160. Thus, Wei-I also discloses using a teaching object to determine the position and orientation of the of the object frame in the tool frame reference frame. Given the advantages disclosed with Wei-I's teaching object techniques, it would have been obvious to a POSITA to improve Corke with Wei-I's disclosures regarding determining an object space-to-camera space transformation and a camera space-to-tool space transformation. EX1003, ¶¶240-241.

Corke in view of Wei-I discloses Claim 15. EX1003, ¶¶237-242.

G. Claim 16

The method of claim 15, further comprising: providing the position and orientation of the object frame to the robot; and training an intended operation path inside the object frame.

Corke discloses providing the position and orientation of the object frame to the robot. EX1003, ¶¶243-252.

The claimed “training an intended operation path” merely specifies the locations of “way points” (*i.e.*, points that should explicitly be visited by the robot) with respect to the object frame, or specifying a tool trajectory with respect to the object frame. EX1003, ¶244. In both cases, these points can then be converted to any of the other coordinate frames (*e.g.*, tool space, camera space, robot space, etc.) using the calibration information. EX1003, ¶244. Transformations of this kind are given by Corke in, for example, Equations (4.58) and (4.59):

$${}^0\mathbf{T}_P = {}^0\mathbf{T}_c {}^c\mathbf{T}_P \quad (4.58)$$

$${}^c\mathbf{T}_P = ({}^0\mathbf{T}_c)^{-1} {}^0\mathbf{T}_P \quad (4.59)$$

EX1004, p.138; EX1003, ¶245. Transformations of this kind are also illustrated in Figures 4.7 and 5.1 of Corke:

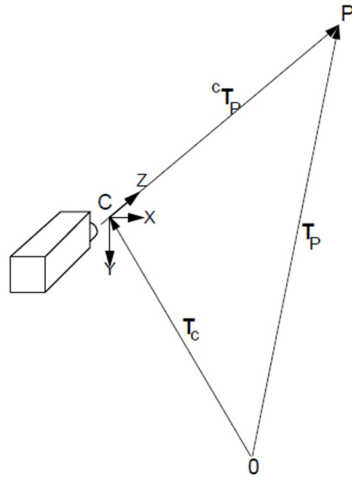


Figure 4.7: Relevant coordinate frames.

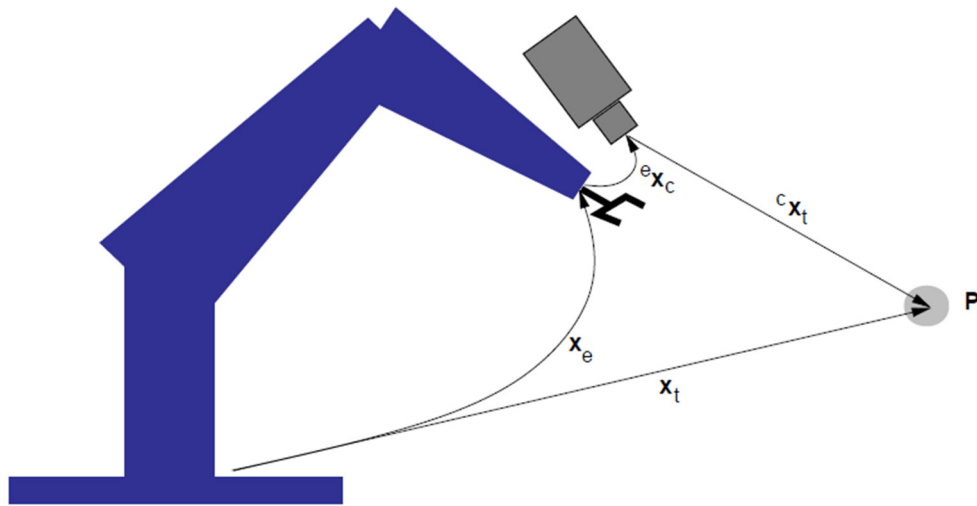


Figure 5.1: Relevant coordinate frames; world, end-effector, camera and target.

EX1004, pp.139, 152; EX1003, ¶246.

Figure 5.4 of Corke shows that the position-based visual servoing method disclosed in Corke sends the target object's “pose estimation”—*i.e.*, the object frame—to the robot:

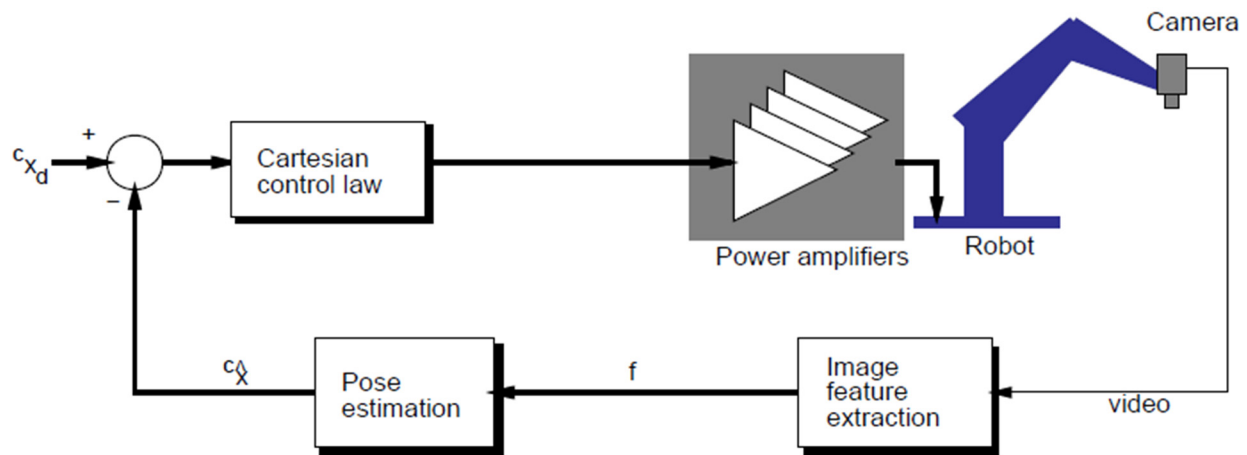


Figure 5.4: Position-based visual servo (PBVS) structure as per Weiss.

EX1004, p.155; EX1003, ¶247.

Corke also describes training an intended operation path inside the object frame. EX1003, ¶248. According to Corke, “sensory data”—including an object frame determined by Corke’s transformations—can modify a robot’s movements. EX1004, p.179. Corke discloses “[m]oving [the robot] to a position defined in terms of [a] transformation” such that the robot’s motion is “controlled by the vision system.” EX1004, p.179. Figure 6.4 of Corke illustrates a method by which the trajectory of the robot is altered based on information received about the object’s pose by a “trajectory generator”:

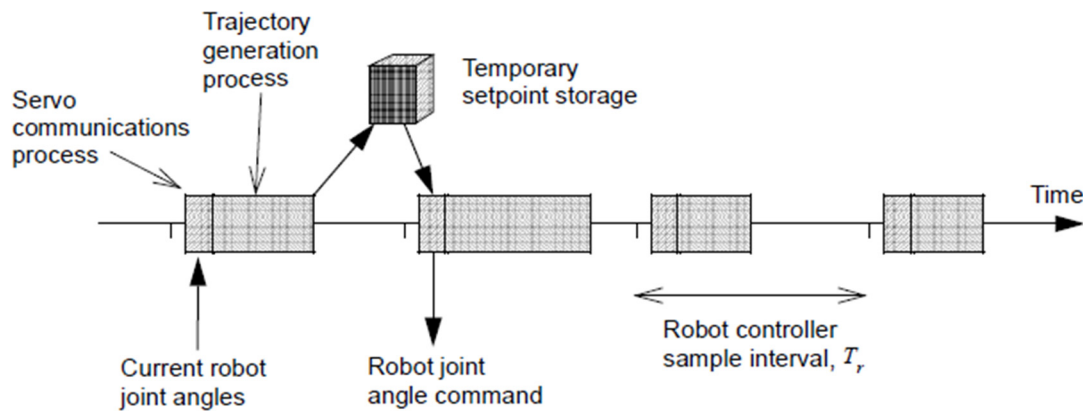


Figure 6.4: Timing of ARCL setpoint generator and servo communication processes.

EX1004, p.179. Corke's "trajectory generator" "move[s] the robot to the general vicinity of the target." EX1004, p.265. A POSITA would have understood that Corke's trajectory generator can train an intended operation path inside the object frame based on pose information (*i.e.*, an "object frame). EX1003, ¶249.

Corke does not expressly disclose a teaching object, but as described in great detail above, it discloses providing the position and orientation of an object frame to the robot and training the robot's operation path. EX1003, ¶250. For the same reasons stated above, it would have been obvious to a POSITA to improve Corke's teaching of providing the object frame to the robot with Wei-I's disclosure of using a teaching object to generate the object frame. EX1003, ¶¶250-251.

Accordingly, Corke in view of Wei-I discloses Claim 16. EX1003, ¶¶243-252.

H. Claim 21*[21.1]*⁸

Corke discloses, or at least renders obvious, Element [21.1]. EX1003, ¶¶253-258. For example, as explained above, Corke discloses all limitations in Claim 20 from which Claim 21 depends. Corke does not expressly disclose training of a teaching object, but as described above, it discloses capturing images of a target object, extracting features, determining the feature coordinates, and determining a target object space-to-camera space transformation. For the same reasons stated above, it would have been obvious to a POSITA to improve Corke's calibration techniques with Wei-I's disclosure of using a teaching object to perform the calibration. EX1003, ¶254.

Wei-I discloses using a teaching object of a type that will be manipulated by the robot. EX1003, ¶255. Wei-I discloses a self-calibration method that does not use calibration plates, but rather any object that has features where a determinable relationship exists between those features. EX1005, p.2. Wei-I discloses using an ORU "for camera calibration." EX1005, p.13. The ORU is an "octagonal shape" that was "to be grasped in a predefined position by a space robot using telesensor programming." EX1006, p.279. During prosecution when application claim 54

⁸ The cited portions of EX1003 provide the claim language for the corresponding portions of the claim.

(issued Claim 21) was rejected over Wei-II (which disclosed a method of machine vision that “avoids” all calibration), the Patent Owner did not argue that the object used in those experiments was not “the type of object the robot will manipulate.” EX1002, p.642. An image of the ORU is shown in Figure 7 of Wei-I:

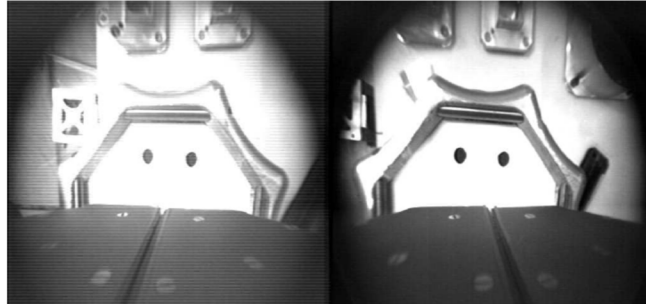


Figure 7: The images of the ORU from the hand cameras. The two black blobs are used as calibration points.

EX1005, p.15. The ORU used in Wei-I’s calibration method was a sample of the type of object the robot will manipulate. EX1003, ¶256. Given the advantages disclosed with Wei-I’s teaching object techniques, it would have been obvious to a POSITA to improve Corke with Wei-I’s disclosures regarding capturing an image of a teaching object of a type that will be manipulated by the robot. EX1003, ¶¶256-257. Corke in view of Wei-I discloses Element [21.1]. EX1003, ¶¶253-258.

[21.2]

Wei-I discloses selecting a number of features from the captured image of the teaching object. EX1003, ¶¶259-263. Wei-I states that “two black blobs” were “artificially marked on the object” and used as calibration features. EX1005, p.13; EX1003, ¶260.

Wei-I also discloses determining a set of object space coordinates for each of the selected features from the captured image of the teaching object. EX1003, ¶261. Wei-I explains that by capturing images of the black blobs “at various positions and orientations of the robot,” the camera’s extrinsic parameters, including the “transformation from the sensor space to the world space,” and the “world points”—“3D points”—of the features, are calculated. EX1005, pp.15-16, 1. Since these 3Dfeature coordinates are defined with respect to the teaching object, they are in “object space.” EX1004, pp.137-38; EX1001, 3:25-26 (defining “object space” as “a reference frame defined with respect to a point on, and therefore rigid to, the object”). For the same reasons stated above, it would have been obvious to a POSITA to improve Corke’s calibration techniques with Wei-I’s disclosure of using a teaching object to perform the calibration. EX1003, ¶¶261-262. Corke in view of Wei-I discloses Element [21.2]. EX1003, ¶¶259-263.

[21.3]

Wei-II discloses determining an object space-to-camera space transformation defining a transformation between an object space reference frame and the camera space reference frame. EX1003, ¶¶264-268. The algorithms of Wei-I “can compute the pose of an object from an image containing several noncoplanar feature points of the object.” EX1005, p.24. Put another way, Wei-I determines the “transformation from the sensor space”—camera space—“to the world space.” EX1005, pp.15-16;

EX1003, ¶265. Since these world coordinates are defined with respect to the teaching object, they are in “object space.” EX1004, pp.137-38; EX1001, 3:25-26 (defining “object space” as “a reference frame defined with respect to a point on, and therefore rigid to, the object”).

Moreover, the '237 Patent states that “the calibration and teaching steps” “can be combined by using a self calibration method of robotic eye and hand-eye relationship with model identification as described in” Wei-I. EX1001, 8:30-39. The '237 Patent states that the self-calibration method in Wei-I can be used “for computing the 3D position of the selected features automatically, without any prior knowledge of the part.” EX1001, 6:6-7:2. For the same reasons stated above, it would have been obvious to a POSITA to improve Corke’s calibration techniques with Wei-I’s disclosure of using a teaching object to perform the calibration. EX1003, ¶266–267. Accordingly, Corke in view of Wei-I discloses Element [21.3]. EX1003, ¶¶264-268.

I. Claims 22 and 23

Corke discloses using at least one programmed computer and/or computer-readable medium for calibration and pose estimation. EX1003, ¶¶269-275. For example, Corke discloses that visual servoing is a method used for “computer vision”—*i.e.*, the “application of a computer system for receiving and processing information,” including image interpretation. EX1004, p.123. Corke also

discloses a calibration matrix run on a computer. EX1003, ¶270. This matrix is a “3x4 homogenous transform which performs scaling, translation, and perspective” and “represents the relationship between 3-D world coordinates and their corresponding 2-D image coordinates as seen by the computer.” EX1004, p.138; 167 (stating that “image processing” is handled “by a conventional computer”).

One of the computer systems used by Corke is shown in Figure 6.2:

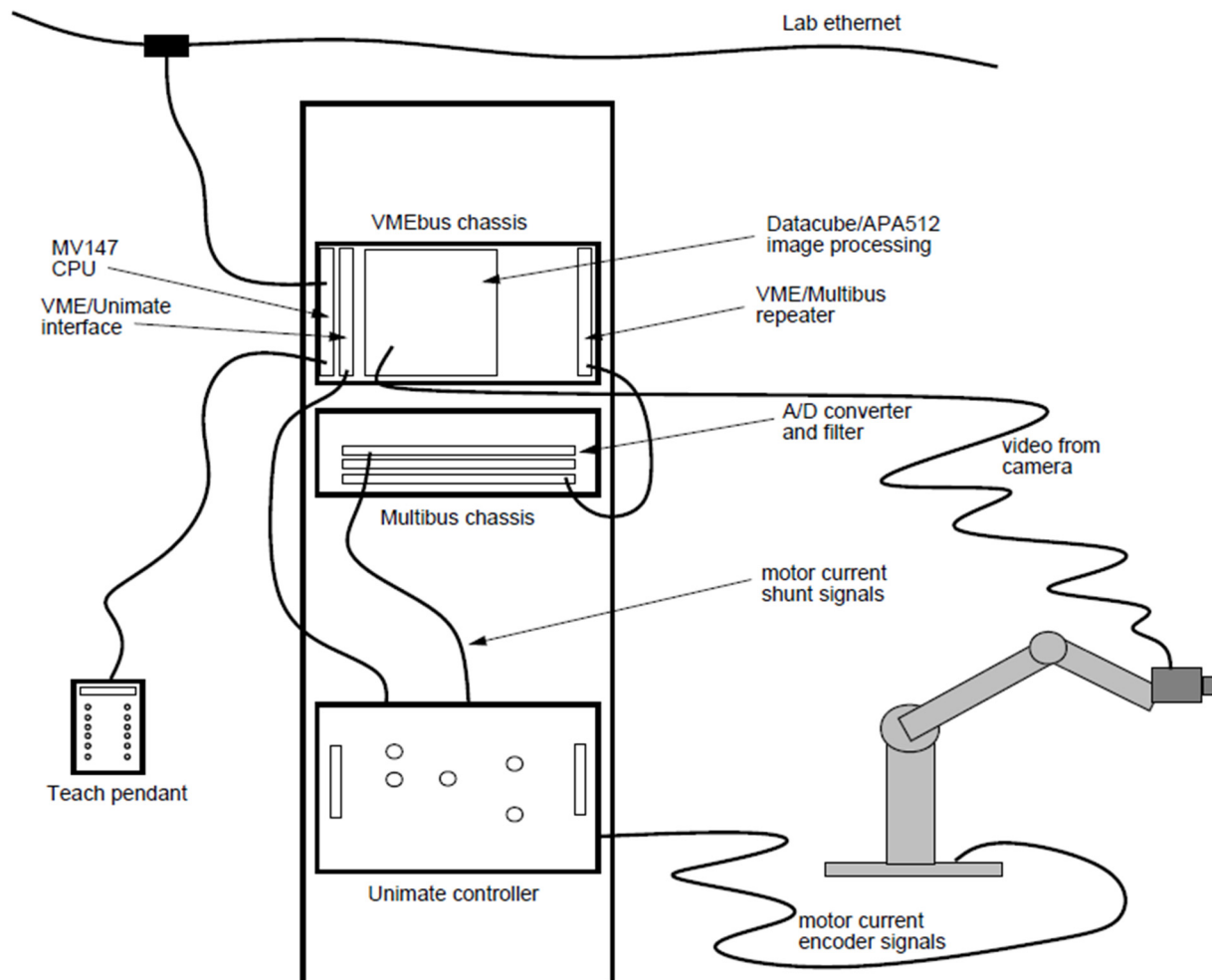


Figure 6.2: Overall view of the system used for visual servo experiments.

EX1005, p.176; EX1003, ¶271. The computer in Figure 6.2 processes information received from the robot and uses it to calculate calibration parameters and determine pose. EX1003, ¶272. A POSITA would have understood that the Datacube image processing device in Figure 6.2 is a computer that processes images obtained from the camera. EX1004, pp.175-177; EX1003, ¶272. Corke also describes the processor and operating system used in the system shown in Figure 6.2, “[t]he CPU is a Motorola MV147 68030 processor card,” which a POSITA would have understood to be a computer, and the operating system as VxWorks, a “real-time multi-tasking operating system is used to support software running in the target processor.” EX1004, pp.175-177; EX1003, ¶272. A POSITA would have therefore understood that this operating system requires a computer system that includes a computer-readable medium for storing instructions. EX1003, ¶272.

Corke does not specifically disclose use of a teaching object, but as described above, it discloses use of a target object, and there is no discernable difference between a teaching and target object. EX1003, ¶273. For the same reasons stated above, to the extent it would not have been understood from Corke to use a “teaching object of the type a robot will manipulate,” it would have been obvious to do so in view of Wei-I. EX1003, ¶¶273-274. Wei-I discloses the use of a computer processor—specifically a Datacube MaxVideo 200 image processing system—to determine the calibration parameters. EX1005, p.13.

Corke in view of Wei-I discloses Claims 22 and 23. EX1003, ¶¶269-275.

XII. DISCRETIONARY DENIAL UNDER § 314 OR § 325(d) IS NOT WARRANTED

A. § 314

The Petition should not be denied under the discretionary factors set out in *Apple Inc. v. Fintiv, Inc.*, IPR2020-00019, Paper 11, at 6 (PTAB March 20, 2020) (“*Fintiv I*”). Petitioner addresses each factor below:

1. Stay. No stay has been requested in the Related Litigation. Given that, the Board should not infer the outcome of any such motion. *Apple Inc. v. Fintiv, Inc.*, IPR2020-00019, Paper 15, at 12 (PTAB May 13, 2020) (“*Fintiv II*”). This factor is neutral on discretionary denial.
2. Trial Date. Pursuant to 35 U.S.C. §316(a)(11), a Final Written Decision will likely issue in February 2025. By contrast, jury selection in the Related Litigation is scheduled to begin, at the earliest, on April 21, 2025. EX1013, p.13. Accordingly, the trial court proceedings would not be completed until several months after a Final Written Decision would be due in this IPR. *Fintiv I* states that this factor does not implicate the decision whether to institute even “if the trial date is at or around the same time as the projected statutory deadline.” IPR2020-00019, Paper 11, at 6. This factor weighs against discretionary denial.

3. Investment. The Related Litigation is in its initial stages, and the investment in it has been minimal. Indeed, fact discovery in the matter is in its early stages.

EX1011. This factor weighs against discretionary denial.

4. Overlap. Instituting a proceeding will allow the Board to address the art and the issues will be narrowed in the Related Litigation through the estoppel provisions of 35 U.S.C. §315(e)(2). The most efficient course is for the Board to address the issues before it, and the district court to address the issues that remain in view of the estoppel provisions of 35 U.S.C. §315(e)(2). This factor weighs against discretionary denial.

5. Same party. Petitioner is the defendant in the Related Litigation filed by Patent Owner. EX1011. That is true of most Petitioners in IPR proceedings. This factor should not be a basis for denying institution.

6. Other circumstances. As discussed above, the prior art strongly demonstrates that Claims 1-10 and 12-28 are unpatentable as obvious. The merits of Petitioner's arguments are strong and the evidence in support of them is substantial. Thus, this factor weighs against discretionary denial. *See Sand Revolution II, LLC v. Continental Intermodal Group – Trucking LLC*, IPR2019-01393, Paper 24 at 13 (PTAB June 16, 2020).

B. § 325(d)

Noninstitution under § 325(d) also is unwarranted based on the factors set forth in *Becton, Dickinson & Co. v. B. Braun Melsungen AG*, IPR2017-01586, Paper 8 at 17-18 (P.T.A.B. Dec. 15, 2017) (precedential). Neither Corke nor Wei-I, the prior art references relied upon in this Petition, were relied upon during the prosecution in a rejection. Furthermore, Corke does not appear in the record for prosecution history. Wei-I is expressly referenced in the specification of the '237 Patent (EX1001, 6:6-7:2), but was not relied upon by the Examiner during prosecution. The Examiner did rely upon Wei-II, another reference authored by the same author as Wei-I, to reject the claims, but these references disclose completely different methods. *See* fn. 1. In fact, Patent Owner overcame Wei-II on the basis that “Wei[-II] ...is directed to the use of a stereo *pair of cameras*” but the claimed invention is “directed to methods and apparatus that employ *single camera* three-dimensional (3-D) vision for robotic guidance.” EX1002, p.129. Indeed, this Petition relies upon prior art to supply the limitations that Patent Owner argued were missing from the prior art, which the Examiner relied upon to allow the claims. Furthermore, neither Corke nor Wei-I are cumulative of the art relied upon by the Examiner during prosecution, and the proposed challenges in this Petition do not overlap with the arguments made during prosecution.

Accordingly, this Petition should not be denied on discretionary factors.

IPR2023-01426
U.S. Patent No. 8,095,237

Dated: September 22, 2023

Respectfully submitted,

/s/Andrew R. Sommer

Andrew R. Sommer
Reg. No. 53,932

CERTIFICATE OF COMPLIANCE

This petition complies with the word count limits set forth in 37 C.F.R. § 42.24(a)(i), effective May 2, 2016, because this Petition contains 13,874 words, excluding the parts of the petition exempted by 37 C.F.R. § 42.24(a), as determined using the word count provided by Microsoft Word, which was used to prepare this Petition.

Dated: September 22, 2023

Respectfully submitted,

/s/Andrew R. Sommer

Andrew R. Sommer
Reg. No. 53,932

CERTIFICATE OF SERVICE

Pursuant to 37 C.F.R. §§ 42.6(e) and 42.105(a), the undersigned certifies that on September 22, 2023, I caused a true and correct copy of the Petition for *Inter Partes* Review, including Exhibit List; Power of Attorney; and Exhibits 1001 through 1013, by UPS on the Patent Owner at the correspondence address of record for U.S. Patent No. 8,095,237 as follows:

Seed Intellectual Property Law Group LLP
701 Fifth Avenue
Suite 5400
Seattle, WA 98104

I further certify that a courtesy copy of the above-identified materials were served on litigation counsel by electronic means upon the foregoing individuals:

John Shaw – jshaw@shawkeller.com
Nathan Hoeschen – nhoeschen@shawkeller.com
Anna Phillips – aphillips@sternekessler.com
John Rozendaal – jcrozendaal@sternekessler.com
Kristina Caggiano Kelly – kckelly@sternekessler.com
Michael Joffre – mjoffre@sternekessler.com
William Milliken – wmilliken@sternekessler.com

/s/Andrew R. Sommer

Andrew R. Sommer
Reg. No. 53,932