

Visual servoing from robust direct color image registration

Geraldo Silveira and Ezio Malis

Abstract—To date, there exist only few works on the use of color images for visual servoing. Perhaps, this is due to the difficulties usually found to cope with illumination changes in these images. This paper presents new parametric models and optimization methods for robustly and directly registering color images. Direct methods refer to those that exploit the pixel intensities, without resorting to image features. We then show how a robust and generic visual servoing scheme can be constructed using the obtained optimal parameters. The proposed models ensure robustness to arbitrary illumination changes in color images, do not require prior knowledge (including the spectral ones) of the object, illuminants or camera, and naturally encompass gray-level images. Furthermore, the exploitation of all information within the images, even from areas where no features exist, allow the algorithm to achieve high levels of accuracy. Various results are reported to show that visual servoing can indeed be highly accurate and robust despite unknown objects and unknown imaging conditions.

I. INTRODUCTION

Visual tracking of an object of interest can be formulated as an image registration problem. Image registration consists in estimating the transformations that best align a reference image to a second one. Generally, they can be classified into feature-based methods or direct methods [1]. Feature-based methods require extracting and matching a set of features (e.g., points, lines) from the two images. Since they may afford relatively larger displacements of the object in the field-of-view, feature-based methods are suitable when the two images are taken under disparate viewpoints. In turn, direct methods exploit the pixel intensities without having to rely on image features. They can then be highly accurate mainly owing to the exploitation of all possible image information, even from areas where no features exist. On the other hand, direct methods assume that the two images of the object have a sufficient overlapping [2]. Since this paper considers real-time vision-based robot control [3], we can suppose that the frame rate is sufficiently high such that only relatively small inter-frame displacements of the object are observed. Moreover, high accuracy is often needed for robot positioning applications. Thus, we focus in this article on direct registration methods of color images and their integration in visual servoing schemes, e.g., [4]. Note however that the parameters estimated by image registration methods can in fact be used in a variety of visual servoing techniques, e.g., [5].

Geraldo Silveira is with CTI Renato Archer – Division DRVC, Rod. Dom Pedro I, km 143,6, Amarais, CEP 13069-901, Campinas/SP, Brazil, Geraldo.Silveira@cti.gov.br

Ezio Malis is with INRIA Sophia-Antipolis – Project ARobAS, 2004 Route des Lucioles, BP 93, 06902 Sophia-Antipolis Cedex, France, Ezio.Malis@sophia.inria.fr

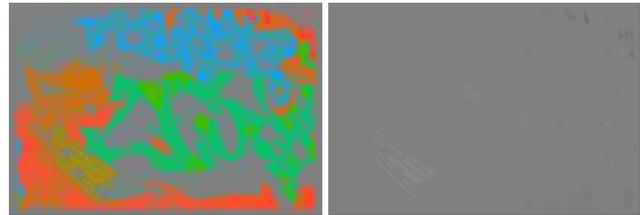


Fig. 1. (a) Original color image and (b) after its conversion to gray-scale. Almost all information has been lost in this example. This illustrates the need to work with the color image directly. Please print in color so as to see how rich the original image is!

To our knowledge, only few techniques on the use of color images in a visual servoing scheme have been proposed to date. Perhaps, this is due to the difficulties usually found to adequately cope with illumination changes in color images. Another possible reason is that one may think that the use of color images do not contribute so much to the final precision of the servoing. This is not always true, and extreme cases exist where all visual information is lost when gray-scale cameras are used (see Fig. 1). Even if this is an unlikely situation in practice, we can conjecture that in many cases color cameras provide much richer information than their gray-scale counterparts. Therefore, their application should be studied in more depth.

Color cameras, like the human eye, are generally (but not always) trichromatic. In this case each pixel of a color image is a three-vector, one component per sensor channel. An active research topic concerns color constancy, which seeks illuminant-invariant color descriptors. A closely related problem is to find illuminant-invariant relationships between color vectors. Given two images of a Mondrian world¹ under specific conditions,² the results presented in [6] claim that a multiplication of each tristimulus value (in an appropriate basis) by a scale factor is sufficient to support color constancy in practice. This framework has been exploited in color-based point tracking, e.g., [7], and has also been applied in [8] to the control of a pan-tilt (i.e., 2 dofs) by finding the centroid of a red object. An effective technique also to find the centroid of an object in color images is through mean-shift [9]. However, these methods are not enough to our purposes since we are interested in accurately and robustly controlling all 6 dofs of a robot end-effector.

In this paper, we propose new models and methods to overcome the limitations of both the Mondrian world¹ and

¹A Mondrian is a planar surface composed of only Lambertian patches, and is after Piet Mondrian (1872-1944) whose paintings are similar.

²For example, the light that strikes the surface has to be of uniform intensity and spectrally unchanging, no inter-reflections, etc.

of those working conditions,² when directly registering color images. Indeed, the proposed transformation models ensure robustness to arbitrary illumination changes in color images, do not require prior knowledge (including the spectral ones) of the light sources (e.g., type, power, pose), of the object (which can be non-Lambertian and of unknown shape) or of the camera sensors, and naturally encompass the gray-level case [10]. Obviously, as any direct method, sufficient gradient along different directions must still be present in the images. Then, we show how to construct a robust and generic visual servoing scheme that directly exploits the obtained optimal parameters. Another important aspect is that the exploitation of all information within the images allows both the registration and the resulting visual servoing scheme to achieve high levels of accuracy. Experimental and simulation results show that the proposed techniques can indeed be highly accurate and robust despite unknown objects and unknown imaging conditions.

II. THEORETICAL BACKGROUND

A. Two-view Epipolar Geometry

Consider the pinhole camera model. The epipolar geometry establishes the relations between corresponding image points $\mathbf{p} \leftrightarrow \mathbf{p}^*$ [11]. In the general uncalibrated case, this relation is given by

$$\mathbf{p} \propto \mathbf{G}\mathbf{p}^* + \rho^* \mathbf{e}, \quad (1)$$

where $\mathbf{G} \in \mathbb{SL}(3)$ is a homography, $\mathbf{e} \in \mathbb{R}^3$ denotes the epipole, and $\rho^* \in \mathbb{R}$ is the parallax of the 3D point projected in the reference image \mathcal{I}^* as \mathbf{p}^* . This parallax also encodes the inverse of the depth of this 3D point. Note that (1) also encompasses the particular situations of a pure rotation motion and that of a planar object. In these both cases, $\rho^* \mathbf{e} = \mathbf{0}$. In the calibrated case, the relation between corresponding image points $\mathbf{p} \leftrightarrow \mathbf{p}^*$ is given by

$$\mathbf{p} \propto \mathbf{K}\mathbf{R}\mathbf{K}^{-1}\mathbf{p}^* + (z^*)^{-1}\mathbf{K}\mathbf{t}, \quad (2)$$

where $\mathbf{R} \in \mathbb{SO}(3)$ and $\mathbf{t} \in \mathbb{R}^3$ respectively denote the relative rotational and translational displacements between the camera frames \mathcal{F} and \mathcal{F}^* , $z^* > 0$ is the depth of the 3D point in the reference camera frame, and the matrix $\mathbf{K} \in \mathbb{R}^{3 \times 3}$ gathers the perfect camera intrinsic parameters.

B. Geometric Direct Image Registration Problem

The geometric direct image registration problem can be formulated as the seek of the geometric parameters that warp the reference image such that each pixel intensity is matched as closely as possible to the corresponding one in the current image. This non-linear optimization problem can be solved by iteratively performing three main steps: image warping, computation of the incremental displacement, and the update of these parameters. Considering that the global minimum can be attained, the final amount of iterations depends on the desired precision.

Therefore, a first step consists in choosing an appropriate warping (i.e., mapping) function

$$\mathbf{p} = \mathbf{w}(\mathbf{g}, \mathbf{p}^*), \quad (3)$$

which can be based on a parametric model. In the case of a purely geometric model one may choose between (1) or (2), depending on the considered framework. If the problem is formulated in the uncalibrated framework, the parameters can be $\mathbf{g} = \{\mathbf{G}, \mathbf{e}, \rho^*\}$ as in [12]. Many applications exploit this set, given its intrinsic robustness characteristics. This will be further discussed in Section IV. If one considers the calibrated case, then the natural set of parameters is $\mathbf{g} = \{\mathbf{R}, \mathbf{t}, (z^*)^{-1}\}$ as in [13]. That is, the Location and the Map. The visual SLAM problem can indeed be formulated as a direct image registration task.

The second step in the iterative procedure consists in computing the incremental displacement $\tilde{\mathbf{g}}$. To this end, a suitable discrepancy measure (i.e., residual), e.g.,

$$\sum_i \left[\mathcal{I}(\mathbf{w}(\tilde{\mathbf{g}} \circ \mathbf{g}, \mathbf{p}_i^*)) - \mathcal{I}^*(\mathbf{p}_i^*) \right]^2, \quad (4)$$

and the best direction of descent are first determined. Finally, a linear least squares problem in terms of $\tilde{\mathbf{g}}$ can be obtained by applying a necessary condition of optimality. Throughout this article, let $\mathcal{I}(\mathbf{p})$ represent the intensity value of pixel \mathbf{p} .

The third step can then be performed once $\tilde{\mathbf{g}}$ has been computed. It consists in updating the transformation parameters through the related composition rule

$$\mathbf{g} \leftarrow \tilde{\mathbf{g}} \circ \mathbf{g}, \quad (5)$$

where the arrow “ \leftarrow ” denotes the update assignment within the iterations, and the operator “ \circ ” depends on the involved Lie group [14]. For example, if one considers a matrix Lie group then the operation to be performed is the matrix multiplication. This three-step procedure is iterated until convergence. The convergence to the optimal \mathbf{g} can be established when the incremental displacement $\tilde{\mathbf{g}}$ is arbitrarily close to the identity element of the involved group.

C. Modeling Arbitrary Illumination Changes

In regard to purely photometric registration tasks, the interest concerns the recovery of which intensity modifications have to be applied to the image \mathcal{I} in order to obtain a transformed \mathcal{I}' such that it matches as closely as possible to the reference one \mathcal{I}^* . Inspired by major illumination models [15], [16], a model of illumination changes has been proposed in [10] to cope with arbitrary lighting variations:

$$\mathcal{I}'(\mathcal{S}, \beta) = \mathcal{S} \cdot \mathcal{I} + \beta, \quad (6)$$

where the elementwise multiplicative lighting variation \mathcal{S} is viewed as a surface that evolves with time. Note that, whilst $\beta \in \mathbb{R}$ captures only global variations, the surface \mathcal{S} also models local illumination changes. For instance, those produced by specular reflections. Very importantly, this model allows the registration to be performed without prior knowledge of the object’s attributes (e.g., albedos, shape) or the illuminants’ characteristics (e.g., number, type, pose).

Nevertheless, if the alignment involves only two images, an under-constrained system is obtained (more unknowns than equations). Surface reconstruction algorithms classically solve this problem through a regularization of the surface.

The basic idea is to prevent pixel intensities from changing independently of each other. Given that the model of the illumination changes is viewed as an evolving surface, the same technique can be applied to the registration at hand. Indeed, \mathcal{S} is supposed to be described by a parametric surface

$$\mathcal{S} \approx f(\boldsymbol{\gamma}, \mathbf{p}), \quad \forall \mathbf{p} \in \mathcal{I}, \quad (7)$$

where the real-valued vector $\boldsymbol{\gamma}$ contains less parameters than the available equations. Then, one has to choose an appropriate finite-dimensional approximation $f(\boldsymbol{\gamma}, \mathbf{p})$ of the actual surface. Afterward, the optimal image \mathcal{I}' in (6) can be obtained by directly estimating the set of parameters $\{\boldsymbol{\gamma}, \beta\}$. To this end, an optimization procedure is also proposed in [10] to simultaneously obtain all of these photometric and of some geometric parameters.

III. PROPOSED PHOTO-GEOMETRIC DIRECT COLOR IMAGE REGISTRATION

This section presents a new transformation model and an optimization method for robustly and efficiently aligning two color images of the same unknown object under unknown imaging conditions. Let \mathcal{I} represent a color image, which is obtained by stacking the channels \mathcal{I}_k , $k = 1, 2, \dots, n$.

A. A Generic Model of Illumination Changes

Let us describe here how to extend the model of lighting variations reviewed in Subsection II-C to the case of color images. The main idea consists in respecting any intrinsic coupling that may be present in the channels so as to be as generic as possible.

Indeed, we propose to obtain the transformed color image \mathcal{I}' that best matches the reference \mathcal{I}^* through

$$\mathcal{I}'(\mathbf{h}) = \mathcal{S} \bullet \mathcal{I} + \beta, \quad (8)$$

where the full set of photometric variables \mathbf{h} comprises all surfaces related to the illumination changes,

$$\mathcal{S} = \begin{bmatrix} \mathcal{S}_{11} & \mathcal{S}_{12} & \cdots & \mathcal{S}_{1n} \\ \mathcal{S}_{21} & \mathcal{S}_{22} & \cdots & \mathcal{S}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{S}_{n1} & \mathcal{S}_{n2} & \cdots & \mathcal{S}_{nn} \end{bmatrix}, \quad (9)$$

and the per-channel shift, $\beta \in \mathbb{R}^n$, which captures the variations both in the ambient lighting changes and in the camera bias. The operator “ \bullet ” stands for a linear combination of the color channels, elementwise multiplied by the corresponding surface. That is, Equation (8) can be rewritten using the operator for elementwise multiplication “ \cdot ” by stacking each

$$\mathcal{I}'_k(\mathbf{h}) = \sum_{j=1}^n \mathcal{S}_{kj} \cdot \mathcal{I}_j + \beta_k, \quad k = 1, 2, \dots, n. \quad (10)$$

The proposed fully coupling model (8) allows the registration to be performed without prior knowledge (including the spectral ones) of the light sources, of the object (which can be non-Lambertian and of unknown shape) or of the camera sensors. Nonetheless, if available they can be easily applied to that generic model. For example, prior knowledge of the

spectral response of the camera sensors (e.g., from its data-sheet) allows for suitably uncoupling the lighting variation \mathcal{S} , at an eventual expense of robustness. Indeed, consider RGB (red-green-blue) images in the following example. Given that at least the red and the blue channels are only weakly coupled in many color cameras, one may set $\mathcal{S}_{13} = \mathcal{S}_{31} = \mathbf{0}$, or even

$$\mathcal{S} \approx \text{diag}(\mathcal{S}_{11}, \mathcal{S}_{22}, \mathcal{S}_{33}) \quad (11)$$

if the camera sensors are narrow-band. Other particular models can also be derived from the generic one (8). For example, if a symmetry between a particular coupling is present then one may set $\mathcal{S}_{12} = \mathcal{S}_{21}$ and/or $\mathcal{S}_{23} = \mathcal{S}_{32}$.

Independently of the choice, if the alignment involves only two images, an under-constrained system is still obtained. Thus, following the same technique for the gray-level case, we suppose that \mathcal{S} can be described by parametric surfaces:

$$\mathcal{S} \approx f(\boldsymbol{\Gamma}, \mathbf{p}), \quad \forall \mathbf{p} \in \mathcal{I}, \quad (12)$$

where $\boldsymbol{\Gamma} = \{\boldsymbol{\gamma}_{kj}\}$. One then has to choose an appropriate finite-dimensional approximation $f(\boldsymbol{\Gamma}, \mathbf{p})$ of the actual \mathcal{S} . For example, through a discretization of the space followed by a suitable interpolation. This choice also plays an important role in the computational efficiency of the algorithm. Another possible solution to reduce the burden is to focus on the dimension of $\boldsymbol{\Gamma}$.

B. The Full System

Hence, our parametric generative model is composed of both the proposed photometric model (8) and a chosen geometric model (see Subsection II-B):

$$\mathcal{I}'(\mathbf{g}, \mathbf{h}, \mathbf{p}^*) = \mathcal{S}(\boldsymbol{\Gamma}, \mathbf{p}^*) \bullet \mathcal{I}(\mathbf{w}(\mathbf{g}, \mathbf{p}^*)) + \beta, \quad (13)$$

with the set of both global and local geometric and photometric parameters, respectively, \mathbf{g} and $\mathbf{h} = \{\boldsymbol{\Gamma}, \beta\}$. Since it must be applied within a fast iterative procedure, the proposed model (13) is transformed into:

$$\begin{aligned} \mathcal{I}'(\tilde{\mathbf{g}} \circ \mathbf{g}, \tilde{\mathbf{h}} \circ \mathbf{h}, \mathbf{p}^*) = \\ = \mathcal{S}(\tilde{\boldsymbol{\Gamma}} \circ \boldsymbol{\Gamma}, \mathbf{p}^*) \bullet \mathcal{I}(\mathbf{w}(\tilde{\mathbf{g}} \circ \mathbf{g}, \mathbf{p}^*)) + \tilde{\beta} \circ \beta. \end{aligned} \quad (14)$$

The robust direct color image registration task can thus be formally posed as

$$\min_{\tilde{\mathbf{x}}} \frac{1}{2} \sum_i \left[\underbrace{\mathcal{I}'(\tilde{\mathbf{x}} \circ \mathbf{x}, \mathbf{p}_i^*) - \mathcal{I}^*(\mathbf{p}_i^*)}_{d_i(\tilde{\mathbf{x}})} \right]^2 \quad (15)$$

using (14) with $\mathbf{x} = \{\mathbf{g}, \mathbf{h}\}$.

C. The Minimization Procedure

The full system (15) can be concisely rewritten as the seek of the incremental parameters $\tilde{\mathbf{x}} = \{\tilde{\mathbf{g}}, \tilde{\mathbf{h}}\}$ such that the image discrepancy is minimized, i.e., as

$$\min_{\tilde{\mathbf{x}}} \frac{1}{2} \|\mathbf{d}(\tilde{\mathbf{x}})\|^2, \quad (16)$$

with $\mathbf{d}(\tilde{\mathbf{x}}) = \{d_i(\tilde{\mathbf{x}})\}$. To iteratively solve this non-linear optimization problem, an expansion in Taylor series is firstly performed. For this, another key technique to achieve nice

convergence properties and high accuracy is to perform an efficient second-order approximation of $\mathbf{d}(\tilde{\mathbf{x}})$ [12]. It is computationally efficient because the Hessians are never computed explicitly. Indeed, it can be shown that, neglecting the third-order remainder, such an approximation of $\mathbf{d}(\tilde{\mathbf{x}})$ around $\tilde{\mathbf{x}} = \mathbf{0}$ is given by

$$\mathbf{d}(\tilde{\mathbf{x}}) = \mathbf{d}(\mathbf{0}) + \frac{1}{2}(\mathbf{J}_{\mathbf{x}}(\mathbf{0}) + \mathbf{J}_{\mathbf{x}}(\bar{\mathbf{x}})) \tilde{\mathbf{x}}, \quad (17)$$

where $\mathbf{J}_{\mathbf{x}}(\mathbf{0})$ and $\mathbf{J}_{\mathbf{x}}(\bar{\mathbf{x}})$ respectively represent the Jacobian at the current state and at the (unknown) solution. Obviously, they both depend on the parametrization of \mathbf{x} . Their analytical expressions for the calibrated framework with gray-level images can be found in [13]. Similarly, the general uncalibrated case with gray-level images is derived by applying the model of illumination changes proposed in [10] to [12]. Their extensions to color images are devised from the proposed photo-geometric generative model (14).

Applying a necessary condition of optimality so that $\tilde{\mathbf{x}}$ is an extremum of (16) gives

$$\frac{1}{2}(\mathbf{J}_{\mathbf{x}}(\mathbf{0}) + \mathbf{J}_{\mathbf{x}}(\bar{\mathbf{x}})) \tilde{\mathbf{x}} = -\mathbf{d}(\mathbf{0}), \quad (18)$$

where $\mathbf{J}_{\mathbf{x}}(\mathbf{0})$ is completely computed from image data. This system of equations (18) is not linear in $\tilde{\mathbf{x}}$ because of $\mathbf{J}_{\mathbf{x}}(\bar{\mathbf{x}})$. However, if the Lie algebra is used to describe motion (either in the calibrated or uncalibrated case), then the corresponding Jacobian part within $\mathbf{J}_{\mathbf{x}}(\bar{\mathbf{x}})$ satisfies the left-invariance property. Thus, it can also be completely computed from image data. This does not hold for the parameters related to surfaces (either in the calibrated domain or in the uncalibrated case). Thus, an approximation of the corresponding Jacobian part has to be used, e.g., the one at the current state.

Afterward, a linear least squares problem can finally be written from (18):

$$\mathbf{J}'_{\tilde{\mathbf{x}}} \tilde{\mathbf{x}} = -\mathbf{d}(\mathbf{0}), \quad (19)$$

where $\mathbf{J}'_{\tilde{\mathbf{x}}}$ represents our proposed direction of descent. The solution $\tilde{\mathbf{x}}$ to the linear system of equations (19) is obtained by solving its normal equations. The parameters are then updated through the related composition rule:

$$\mathbf{x} \leftarrow \tilde{\mathbf{x}} \circ \mathbf{x}. \quad (20)$$

The procedure is iterated until the convergence to the optimal \mathbf{x} , as outlined in Subsection II-B. As a remark, while having an equivalent computational cost to the Gauss-Newton method, it exploits all available information from both the current and the reference images. This contributes to achieve large domains and rates of convergence.

IV. APPLICATION TO VISUAL SERVOING

The image registration method proposed in Section III simultaneously recovers the optimal set of parameters $\mathbf{x} = \{\mathbf{g}, \mathbf{h}\}$. The photometric parameters \mathbf{h} are estimated so as to achieve effective robustness to illumination changes. On the other hand, the geometric parameters \mathbf{g} can be used for visual servoing purposes, as described in this section.

A variety of visual servoing techniques can be applied using \mathbf{g} , either in the calibrated or uncalibrated setting. However, let us focus here on the uncalibrated case, which is intrinsically more robust to errors in the camera parameters \mathbf{K} (when controlling all 6 dofs, at least a coarse estimate $\hat{\mathbf{K}}$ is always needed). In this setting, the reference pose is given by means of a reference image, a framework also called teach-by-showing. Considering then the uncalibrated case, the related set of geometric parameters is $\mathbf{g} = \{\mathbf{G}, \mathbf{e}, \rho^*\}$ (see Section II). Existing visual servoing techniques that exploit this set for controlling all 6 dofs of a robot mainly differ in the required prior knowledge, e.g., scene structure, camera motion, normal to a plane. Recently, the Direct Visual Servoing (DVS) has been proposed [4] as a general technique where neither priors nor decompositions are required.

The DVS uses that projective geometric information as follows. Firstly, normalized entities are obtained:

$$\mathbf{m}^* = \hat{\mathbf{K}}^{-1} \mathbf{p}^* \quad (21)$$

$$\mathbf{e}' = \hat{\mathbf{K}}^{-1} \mathbf{e} \quad (22)$$

$$\mathbf{H} = \hat{\mathbf{K}}^{-1} \mathbf{G} \hat{\mathbf{K}}, \quad (23)$$

where \mathbf{p}^* corresponds to a chosen point (not necessarily an interest point) of the object. The control error

$$\boldsymbol{\varepsilon} = \begin{bmatrix} (\mathbf{H} - \mathbf{I}) \mathbf{m}^* + \rho^* \mathbf{e}' \\ \boldsymbol{\mu}(\mathbf{H}) \vartheta(\mathbf{H}) \end{bmatrix} \quad (24)$$

is theoretically proved to be locally diffeomorphic to the camera pose. In (24), $[\boldsymbol{\mu}(\mathbf{H})]_{\times} = \mathbf{H} - \mathbf{H}^T$, where $[\boldsymbol{\mu}]_{\times}$ denotes the anti-symmetric matrix associated to the 3-vector $\boldsymbol{\mu}$, and $\vartheta(\mathbf{H}) = 1$ (or another function that allows for path planning). Provided the diffeomorphism, the control law

$$\mathbf{v} = \lambda \boldsymbol{\varepsilon} \in \mathbb{R}^6, \quad \lambda > 0, \quad (25)$$

is also theoretically proved to ensure local asymptotic stability. Furthermore, it is shown that a very large domain of convergence is obtained if path planning is performed. The control signal \mathbf{v} comprises both translational and rotational velocities. Using this technique, no prior knowledge of the object's attributes or of the camera's motion are required, both for registration and servoing.

V. RESULTS

This section reports some representative sets of experiments (they can also be found as multimedia material) using unknown objects under challenging imaging conditions, both for image registration and for full 6-dof positioning tasks.

A. Direct Color Image Registration

The image registration algorithm immediately starts after selecting the area of interest, also called template. This template is then considered as the reference one. Assuming relatively small inter-frame displacements, it is automatically aligned to successive frames of the sequence.

The first set of results is shown in Fig. 2. The unknown light source and camera perform unknown motions in space. Despite severe specularities, shadows and instantaneous



Fig. 2. (First row) Image registration of a reference image (the first one) to successive frames of a video sequence. (Bottom row) Aligned (warped) images are shown to demonstrate the stability of the tracker. The sequence contains severe changes in the specular, diffuse and ambient reflections.

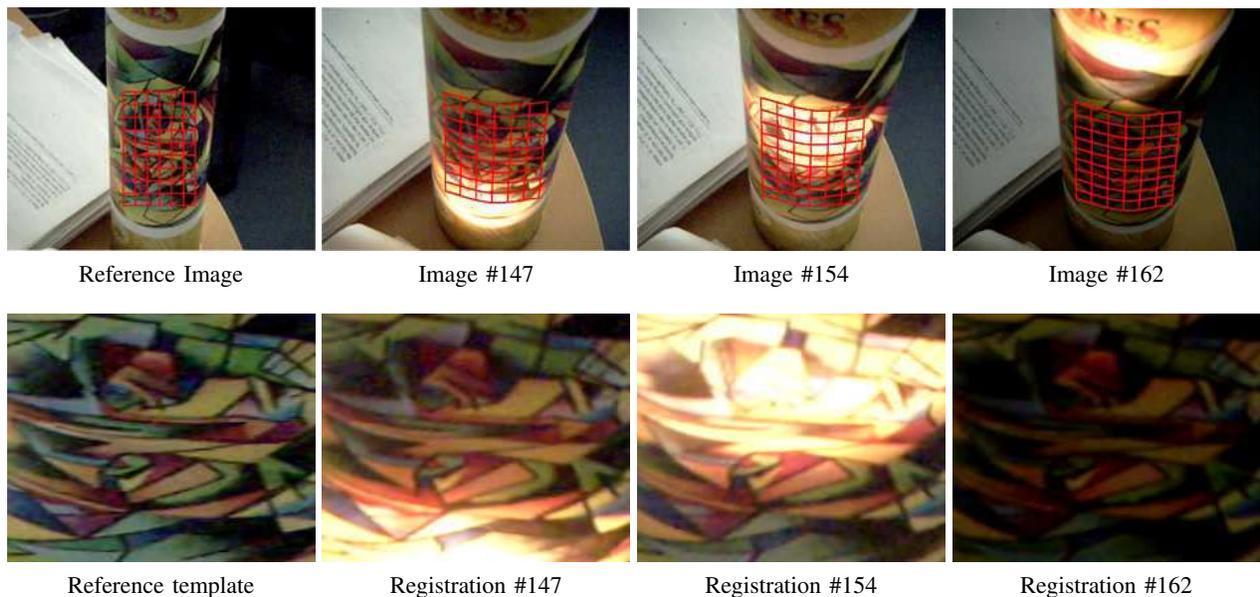


Fig. 3. (First row) Image registration of a reference image to successive frames of a video sequence. (Bottom row) Aligned (warped) images are shown to demonstrate the stability of the tracker along the sequence. The unknown light source and camera perform unknown motions in space. No prior knowledge of the object's attributes (e.g., shape, albedos) is exploited.

changes in the diffuse and ambient reflections, all images are accurately registered, with a median RMS error of 15.73 levels of gray-scale along the sequence, performing a median of 9 iterations per image.

The second set of results is shown in Fig. 3. We have used the same pattern as in the preceding sequence, although with an object of different shape (in this case, a cylinder). Of course, this knowledge is not a-priori provided to the algorithm. Once again, a challenging scenario is set up with very disparate types of lighting variations, and the images are successfully aligned with a median RMS error of 16.76 levels of gray-scale along the sequence, performing a median of 7 iterations per image.

B. Direct Visual Servoing

The optimal set of parameters computed from each image registration can be used for visually servoing a robot. We provide in this subsection the results obtained by the DVS technique [4], given its attractive properties (e.g., no prior knowledge required, convergence properties, etc.) as discussed in Section IV.

To have a ground truth, we constructed a synthetic object (a sphere) and created light sources. In particular, the specular reflections are due to an illuminant rigidly attached to the virtual camera. It points towards the object with a slightly different direction with respect to the camera's optical axis.

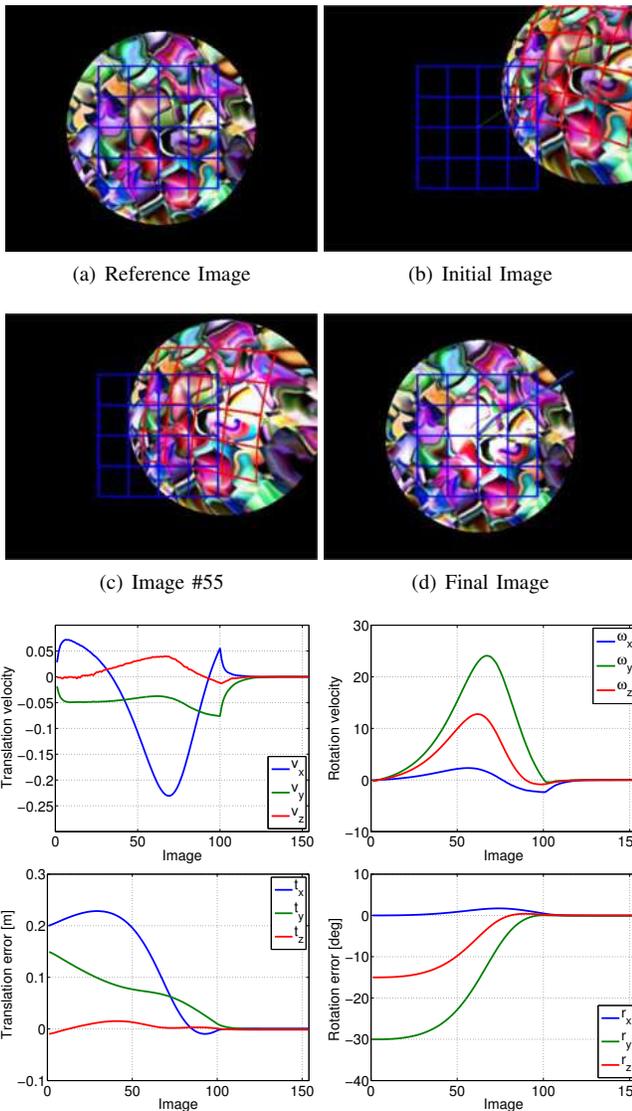


Fig. 4. Direct visual servoing with respect to a sphere (a-priori unknown) using a pinhole color camera. Note that the servoing is successfully performed despite large specular reflections, even at the final image. Compare (d) the final image with (a) the reference one.

This simulates a misalignment between the camera and the carrying light. Then, a textured image is mapped onto the object. Finally, we model the control system as a pinhole color camera mounted on the end-effector of a classical manipulator robot.

A visual servoing result is depicted in Fig. 4. The control law is stable: both translational and rotational velocities converge to zero. At the convergence, the camera is positioned at the desired pose very accurately. The norm of the final Cartesian error is around 1 mm for the translation, and 0.1° for the rotation. We remark that high accuracy is obtained despite large specular reflections even at the final image (compare it with the reference one). See Fig. 5 for both the synthetic reflection present in the image at the convergence, and a particular surface related to the illumination changes reconstructed by the image registration method.

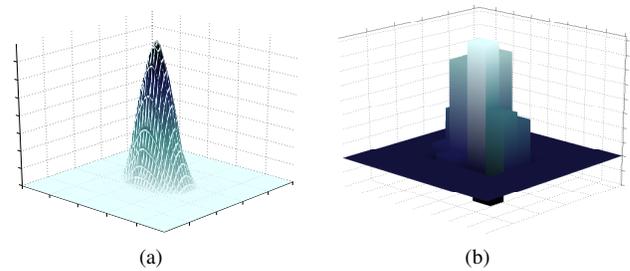


Fig. 5. (a) The synthetic specular reflection at the convergence for the visual servoing task showed in Fig. 4. (b) A particular reconstructed surface (S_{22}) to counterbalance the lighting variations. All surfaces are modeled through discretization into blocks for computational efficiency.

VI. CONCLUSIONS

In this paper, we have investigated how to improve the robustness and the accuracy of visual servoing methods through appropriate color image registration techniques. In particular, we have focused on general photo-geometric parametric transformation models to perform direct image alignment. These models can cope with generic illumination changes, e.g., specularities and shadows, even in color images. The registration approach is then integrated into a visual servoing technique that directly uses the estimated parameters. Experimental results with both real-world and simulated sequences show that the registration and the resulting visual servoing scheme can be highly robust and accurate despite unknown objects and unknown imaging conditions.

ACKNOWLEDGMENTS

This work is also partially supported by the Brazilian CAPES Foundation.

REFERENCES

- [1] R. Szeliski, *Handbook of Mathematical Models in Computer Vision*. Springer, 2006, pp. 273–292.
- [2] B. Lucas and T. Kanade, “An iterative image registration technique with an application to stereo vision,” in *IJCAI*, 1981, pp. 674–679.
- [3] F. Chaumette and S. Hutchinson, “Visual servo control part I: Basic approaches,” *IEEE Rob. & Autom. Magazine*, pp. 82–90, 2006.
- [4] G. Silveira and E. Malis, “Direct visual servoing with respect to rigid objects,” in *IEEE/RSJ IROS, USA, 2007*.
- [5] E. Malis, F. Chaumette, and S. Boudet, “2D 1/2 visual servoing,” *IEEE Trans. on Robotics and Automation*, vol. 15, no. 2, pp. 238–250, 1999.
- [6] G. Finlayson, M. Drew, and B. Funt, “Color constancy: Generalized diagonal transforms suffice,” *J. Opt. Soc. Am. A*, vol. 11, no. 11, pp. 3011–3020, 1994.
- [7] P. Montesinos, V. Gouet, R. Deriche, and D. Pele, “Matching color uncalibrated images using differential invariants,” *Image and Vision Computing*, vol. 18, no. 9, pp. 659–671, 1999.
- [8] G. de Cubber, S. A. Berrabah, and H. Sahli, “Color-based visual servoing under varying illumination conditions,” *Robotics and Autonomous Systems*, vol. 47, no. 4, pp. 225–249, 2004.
- [9] D. Comaniciu, V. Ramesh, and P. Meer, “Real-time tracking of non-rigid objects using mean-shift,” in *IEEE CVPR*, 2000.
- [10] G. Silveira and E. Malis, “Real-time visual tracking under arbitrary illumination changes,” in *IEEE CVPR*, 2007.
- [11] O. Faugeras, Q.-T. Luong, and T. Papadopoulos, *The geometry of multiple images*. The MIT Press, 2001.
- [12] E. Malis, “An efficient unified approach to direct visual tracking of rigid and deformable surfaces,” in *IEEE/RSJ IROS, USA, 2007*.
- [13] G. Silveira, E. Malis, and P. Rives, “An efficient direct approach to visual SLAM,” *IEEE Transactions on Robotics*, vol. 24, no. 5, pp. 969–979, 2008, special issue on Visual SLAM.
- [14] F. W. Warner, *Foundations of differential manifolds and Lie groups*. Springer Verlag, 1987.
- [15] J. F. Blinn, “Models of light reflection for computer synthesized pictures,” in *SIGGRAPH*, 1977, pp. 192–198.
- [16] R. Cook and K. Torrance, “A reflectance model for computer graphics,” *ACM Trans. on Graphics 1*, pp. 7–24, 1982.