

Real-time Visual Tracking under Arbitrary Illumination Changes

Geraldo Silveira^{1,2}

¹ CenPRA Research Center – DRVC Division
Rod. Dom Pedro I, km 143,6, Amarais
CEP 13069-901, Campinas/SP, Brazil
Geraldo.Silveira@cenpra.gov.br

Ezio Malis²

² INRIA Sophia-Antipolis – Project ICARE
2004 Route des Lucioles, BP 93
06902 Sophia-Antipolis Cedex, France
FirstName.LastName@sophia.inria.fr

Abstract

In this paper, we investigate how to improve the robustness of visual tracking methods with respect to generic lighting changes. We propose a new approach to the direct image alignment of either Lambertian or non-Lambertian objects under shadows, inter-reflections, glints as well as ambient, diffuse and specular reflections which may vary in power, type, number and space. The method is based on a proposed model of illumination changes together with an appropriate geometric model of image motion. The parameters related to these models are obtained through an efficient second-order optimization technique which minimizes directly the intensity discrepancies. Comparison results with existing direct methods show significant improvements in the tracking performance. Extensive experiments confirm the robustness and reliability of our method.

1. Introduction

This work considers visual tracking as the problem of estimating the incremental transformations which align a reference image with successive frames of a video sequence. This is a fundamental task to many applications and numerous techniques exist in the literature. Generally, they can be classified into feature-based and direct (appearance-based) methods [10]. The former class extracts a sparse set of salient features to compute an appropriate tensor which will allow to perform the alignment. The latter uses directly the intensity value of all pixels of interest. This work focuses only on efficient direct visual tracking algorithms. Hence, techniques which perform bundle adjustment are not considered here. Also, we consider applications where it is not possible to execute an off-line learning stage, such as in [6, 9]. Moreover, the solution to our problem must support all classes of image transformations up to perspective deformations. Thus, the tracking technique proposed e.g. in [5], although effective, is not sufficient since it provides up to a similarity transformation.

Specifically, we tackle here an important issue to all vision-based algorithms: the robustness to generic lighting

changes. Indeed, we address the efficient tracking of either Lambertian or non-Lambertian objects under unknown imaging conditions. To this end, a possible scheme to increase the robustness to variable illumination is by performing a photometric normalization. For example, the images may be normalized by using the mean and the standard deviation. However, this method provides inferior performance, especially when the inter-frame displacements (geometric and/or photometric) are large [1]. Another widely used technique is to model the change in illumination as an affine transformation, e.g. [1, 2, 8]. Despite the fact that improved results are obtained, only global changes are modeled and thus specular reflections, for example, are not taken into consideration. A possible strategy to deal with the specularities is to use a robust error function [7]. Nevertheless, they are proved to be inefficient in the case of appearance-based tracking [1]. The reasons are twofold. First, in this case there is an ambiguity in the interpretation of the intensity differences between those caused by motion and those caused by lighting changes [9]. For example, strong differences caused by motion may be discarded though weak differences (e.g. produced by shadows) may be interpreted as motion. Secondly, they may discard important, pertinent information that could be easily modeled and thus exploited. Hence, the convergence rate of the algorithm is slowed down or, even worse, the tracking may fail. On the other hand, those robust functions might be applied to handle unknown occlusions since their realistic modeling is difficult, if not impossible.

This work proposes a new direct visual tracking approach where the robustness to lighting changes is assured by using a proposed model of illumination changes together with an appropriate geometric model of image motion. The former is generic: it does not require either the characteristics of the light sources (e.g. number, power, pose), nor about the reflectance properties of the surface. Also, the model is devised such that the real-time constraint can be respected. Furthermore, we extend the efficient second-order minimization technique [3] to simultaneously obtain the optimal global and local parameters related to those models. Hence, large rates and domains of convergence are achieved. In addition, it is computationally efficient

because the Hessians are never computed explicitly. As a means to compare the method with existing ones, a planar surface is considered in this work. Extensions to deal with non-planar scenes, rigid or not, as well as with color images can be made. Results are provided from real-world data under ambient, diffuse and specular reflections which vary in power, type, number and space. Another complication that can arise concerns the occurrence of off-specular peaks (glints) and inter-reflections. Results demonstrate that the proposed approach also accommodates them without making any additional change. For the experiments, representative sample surfaces were chosen which range from smooth to rough, and including metal and dielectric objects. To the authors' knowledge, no existing efficient direct techniques are capable of coping with such a challenging scenario, especially when the object is not near-Lambertian and/or large displacements are carried out. Whenever a comparison was indeed possible, the results show significant improvements in the tracking performance.

2. Theoretical Background

2.1. Notations

Let \mathcal{I}^* denote the reference image and $\mathbf{p} = [u, v, 1]^\top \in \mathbb{P}^2$ pixel coordinates. In turn, $\mathcal{I}^*(\mathbf{p}^*)$ denotes the image intensity of a pixel \mathbf{p}^* . The usual notations $\hat{\mathbf{u}}$, $\tilde{\mathbf{u}}$, $\bar{\mathbf{u}}$, \mathbf{u}' to represent respectively the estimate of a variable \mathbf{u} , an increment to be found, an augmented version and a modified one of \mathbf{u} are also followed here.

2.2. Plane-based two-view geometry

Consider that an appropriate planar region $\mathcal{R}^* \subseteq \mathcal{I}^*$ has been defined. The coordinates of a pixel \mathbf{p}^* in \mathcal{R}^* are related to its corresponding \mathbf{p} in the current image \mathcal{I} by a projective homography \mathbf{H} . Thus, given \mathbf{H} , it is possible to define a warping function on the pixel coordinates:

$$\mathbf{w}(\cdot; \mathbf{H}) : \mathbb{P}^2 \mapsto \mathbb{P}^2; \quad \mathbf{p}^* \mapsto \mathbf{p} = \mathbf{w}(\mathbf{p}^*; \mathbf{H}). \quad (1)$$

This geometric relation was chosen since it defines the most general class of projective transformations for planar objects, which include translations, rotations and affinities. It has 8 degrees of freedom (since it is defined up to scale) and can be parameterized in different ways. In [3], the constraint $\det(\mathbf{H}) = 1$, i.e. $\mathbf{H} \in \mathbb{SL}(3)$, is imposed. This is well-justified since $\det(\mathbf{H}) = 0$ occurs only in the degenerate case of the plane, i.e. when it is projected as a line.

2.3. Parameterizing incremental displacements

Let \mathbf{A}_i , $i = 1, 2, \dots, 8$, be the canonical basis of the Lie algebra $\mathfrak{sl}(3)$ [11]. The basis are linearly independent constant (3×3) matrices such that $\text{trace}(\mathbf{A}_i) = 0$. Any $\mathbf{A} \in \mathfrak{sl}(3)$ can be written as a linear combination of the \mathbf{A}_i :

$$\mathbf{A}(\mathbf{x}) = \sum_{i=1}^8 x_i \mathbf{A}_i \in \mathfrak{sl}(3), \quad (2)$$

where $\mathbf{x} = [x_1, x_2, \dots, x_8]^\top \in \mathbb{R}^8$, and x_i is the i -th element of the base field. Such an algebra is related to its Lie group $\mathbb{SL}(3)$ via the exponential map

$$\exp : \mathfrak{sl}(3) \mapsto \mathbb{SL}(3); \quad \mathbf{A}(\mathbf{x}) \mapsto e^{\mathbf{A}(\mathbf{x})}. \quad (3)$$

Notice that $\det(e^{\mathbf{A}(\mathbf{x})}) = e^{\text{trace}(\mathbf{A}(\mathbf{x}))} = 1$. Furthermore, the transformation (3) is smooth and one-to-one onto, with a smooth inverse, within a very large neighborhood of $\mathbf{0} \in \mathfrak{sl}(3)$ and the identity element $\mathbf{I}_3 \in \mathbb{SL}(3)$.

2.4. The geometric direct visual tracking

This problem can be formulated as a search for the geometric parameters to warp all the pixels in the region $\mathcal{R}^* \subseteq \mathcal{I}^*$ so that their intensities match as closely as possible to their corresponding ones in \mathcal{I} . These geometric parameters are encoded in \mathbf{H} in the case of planar surfaces under generic motion. For that, a non-linear minimization procedure has to be derived since the pixel intensity $\mathcal{I}(\mathbf{p})$ are, in general, non-linear in \mathbf{p} . A standard technique to solve this problem consists in performing an expansion of the cost function in Taylor series and after applying a necessary condition of optimality. The solution of the obtained linear least squares problem iteratively updates an initial guess until convergence. Hence, given an estimate $\hat{\mathbf{H}}$ of \mathbf{H} , the problem here is to find the optimal incremental $\tilde{\mathbf{H}} = \mathbf{H}(\tilde{\mathbf{x}})$ through an iterative method which solves

$$\min_{\tilde{\mathbf{x}} \in \mathbb{R}^8} \frac{1}{2} \sum_{\mathbf{p}_i^* \in \mathcal{R}^*} \left[\mathcal{I}(\mathbf{w}(\mathbf{p}_i^*; \hat{\mathbf{H}}\mathbf{H}(\tilde{\mathbf{x}}))) - \mathcal{I}^*(\mathbf{p}_i^*) \right]^2. \quad (4)$$

The update of the homography is then obtained within its iterations through the exponential map (3):

$$\hat{\mathbf{H}} \longleftarrow \hat{\mathbf{H}}\mathbf{H}(\tilde{\mathbf{x}}) = \hat{\mathbf{H}} e^{\mathbf{A}(\tilde{\mathbf{x}})}. \quad (5)$$

We remark that the resulting $\hat{\mathbf{H}}$ is always in the group, and no approximation is performed. Hence, the local parameterization (2) improves stability and accuracy. The convergence to the optimal solution may be established when the displacement is arbitrarily small, i.e. $\|\tilde{\mathbf{x}}\| < \epsilon$. This estimate may then be used for initializing the same procedure when a new image is available.

3. The Proposed Visual Tracking Approach

3.1. Modeling the illumination changes

According to major illumination models (e.g. Blinn-Phong, Cook-Torrance), the luminance at a particular pixel \mathbf{p}_i is due to diffuse, specular and ambient reflections:

$$\mathcal{I}(\mathbf{p}_i; \cdot) = \mathcal{I}_d(\mathbf{p}_i; \cdot) + \mathcal{I}_s(\mathbf{p}_i; \cdot) + \mathcal{I}_a, \quad (6)$$

where each model has their own set of parameters. For example, with the Blinn-Phong model Eq. (6) can be written

$$\begin{aligned} \mathcal{I}(\mathbf{p}_i; \boldsymbol{\pi}, \mathcal{F}, k_c, \mathcal{L}, k_{d_i}, k_{s_i}, m) &= \mathcal{I}_d(\mathbf{p}_i; \boldsymbol{\pi}, k_c, \mathcal{L}, k_{d_i}) + \\ &+ \mathcal{I}_s(\mathbf{p}_i; \boldsymbol{\pi}, \mathcal{F}, k_c, \mathcal{L}, k_{s_i}, m) + \mathcal{I}_a, \end{aligned} \quad (7)$$

where the plane π is defined w.r.t. the camera frame \mathcal{F} , \mathcal{L} comprises the set of light sources with unit direction vector and corresponding radiance (from a wavelength λ) from the surface with Lambertian and specular albedos $k_{di}(\lambda)$ and $k_{si}(\lambda)$, respectively. Also, k_c denotes the camera gain, while m is the specular exponent. In the case of the Cook-Torrance model, other parameters include the Fresnel reflectance and the surface roughness.

For visual tracking purposes, however, the interest concerns the recovery of which lighting variations has to be applied to the current image \mathcal{I} (6) in order to obtain an image \mathcal{I}' whose illumination conditions are as closely as possible to those at the time of acquiring \mathcal{I}^* :

$$\mathcal{I}'(\mathbf{p}_i; \alpha_i, \eta_i, \beta) = \alpha_i \mathcal{I}_d(\mathbf{p}_i) + \eta_i \mathcal{I}_s(\mathbf{p}_i) + \beta, \quad (8)$$

where $\alpha_i, \eta_i, \beta \in \mathbb{R}$ capture respectively the variations caused by diffuse, specular and global lighting changes. The latter also includes the shift in the camera bias. Notice that the first two variations depend on the albedos of each point on the surface, as well as its shape, the camera parameters and the imaging conditions. This is then a difficult, computationally intensive problem where many images and priors are required to consistently recover those parameters. Indeed, two assumptions are commonly adopted by visual tracking algorithms, e.g. [1, 2, 8]. The first assumption is to consider that the surface is perfectly Lambertian so that $\eta_i = 0, \forall \mathbf{p}_i$, and consequently, $\mathcal{I} = \mathcal{I}_d$. Secondly, they assume that the entire surface holds exactly the same reflectance properties so that $\alpha = \alpha_i, \forall \mathbf{p}_i$. Although suited to some applications, both assumptions are obviously violated in many cases.

Since we do not make any assumption about either the imaging conditions or about the materials, we propose to reformulate the problem (8) so as to find an elementwise multiplicative lighting variation $\tilde{\mathcal{I}}$ over the current \mathcal{I} , and a global β , such that \mathcal{I}' matches as closely as possible to \mathcal{I}^* :

$$\mathcal{I}' = \tilde{\mathcal{I}} \cdot \mathcal{I} + \beta. \quad (9)$$

In any case, if only two images are concerned, we have an observability problem if we consider that the intensity of each pixel can change independently (more unknowns than equations). To solve this problem, we suppose that $\tilde{\mathcal{I}}$ can be modeled by a parametric surface $\tilde{\mathcal{I}} = f(\mathbf{p}_i; \gamma), \forall \mathbf{p}_i$, where the vector γ contains less parameters than the available equations and varies with time. Then, one has to choose an appropriate approximation of such a surface. We have tested various techniques. First of all we have used Radial Basis Functions (RBF) [4]. The surface can then be approximated by using e.g. the thin-plate spline for defining $\phi(\|\cdot\|)$ and a first-degree polynomial:

$$\tilde{\mathcal{I}}(\mathbf{p}_i; \gamma) \approx [\gamma_{q+1}, \gamma_{q+2}, \gamma_{q+3}]^\top \mathbf{p}_i + \sum_{k=1}^q \gamma_k \phi(\|\mathbf{p}_i - \mathbf{q}_k\|), \quad (10)$$

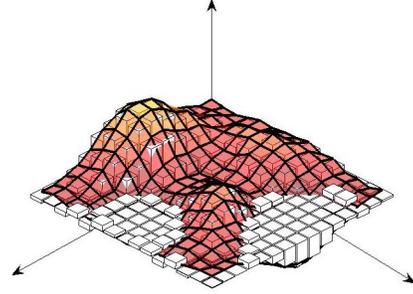


Figure 1. Discretized surface (represented by boxes) for approximating the lighting changes $\tilde{\mathcal{I}}$ (colored).

where $\{\mathbf{q}_k\}_{k=1}^q$ are the image points (called centers) that can be selected on a regular grid or correspond to interest points of the image. The side conditions can be easily imposed by solving a linear system while the interpolation conditions will be indirectly imposed by minimizing a similarity measure. The use of RBFs allows to regularize the surface but they may fail to accurately capture discontinuities. Thus, we have also tested another suitable strategy for dealing with discontinuous surfaces. In this case, $\tilde{\mathcal{I}}$ is approximated by a discretized surface which evolves with time:

$$\tilde{\mathcal{I}}(\mathbf{p}_{ij}; \gamma) \approx \begin{cases} \gamma_j, & \forall \mathbf{p}_{ij} \in \Delta \mathcal{R}_j, \\ 0, & \text{otherwise,} \end{cases} \quad (11)$$

where \mathbf{p}_{ij} denotes the i -th pixel of the j -th subregion $\Delta \mathcal{R}_j$. In practice, the image region \mathcal{R} is discretized into n ($\Delta u \times \Delta v$) blocks $\mathcal{B}_j, j = 1, 2, \dots, n$, of sufficiently small size. Thus, Eq. (11) satisfies

$$\iint_{\mathcal{R}} \tilde{\mathcal{I}}(u, v) du dv \approx \sum_{j=1}^n \tilde{\mathcal{I}}(\mathbf{p}_{ij}; \gamma) \Delta u \Delta v. \quad (12)$$

See Fig. 1 for an illustration. This discretization amounts to the reasonable assumption that the surface possesses piecewise similar reflectance properties. Nevertheless, the discretization exhibits important strengths. First, an accurate, compact and computationally efficient representation of the illumination changes is obtained. In fact, a piecewise linear system in those parameters as well as a sparse Jacobian matrix are both achieved (see Section 3.2). These are highly desirable if frame-rate performance is a concern. Also, this model allows the algorithm to operate without any a priori knowledge about the specific reflectance properties of the materials as well as about the number, color, power and the types of the light sources, including their pose in space. Last, but not least it represents a generalization of the previous approaches where priors can be easily applied.

Saturations due to highlights and shadows. These particular effects are here interpreted as well-structured types of occluders. This characterization justifies since, wherever

they are present, all information which are useful for tracking purposes are hidden. Moreover, they are well-structured because a saturation pattern is exhibited either to the highest or to the lowest intensity levels. Therefore, they can be filtered suitably: one only needs to check whether or not those homogeneous patterns appear in each warped image block.

3.2. The minimization procedure

Due to paper length restrictions, we present the framework for the discretization case. A similar one can be derived for the RBF case. Hence, using (11) and incorporating the geometric parameters (1), Eq. (9) is then translated to

$$\mathcal{I}'(\mathbf{w}(\mathbf{p}_{ij}^*; \mathbf{H}); \gamma_j, \beta) = \gamma_j \mathcal{I}(\mathbf{w}(\mathbf{p}_{ij}^*; \mathbf{H})) + \beta \quad (13)$$

for the particular pixel \mathbf{p}_{ij}^* . We remark that such a transformation uses the set of parameters $\{\mathbf{H}, \gamma, \beta\}$, which is homeomorphic to a Lie group defined over $\mathbb{S}\mathbb{L}(3) \times \mathbb{R}^{n+1}$, with $\gamma = [\gamma_1, \gamma_2, \dots, \gamma_n]^\top$. Given that the optimal solution is found iteratively, Eq. (13) is then transformed into

$$\mathcal{I}'_{ij} = (\hat{\gamma}_j + \tilde{\gamma}_j) \mathcal{I}(\mathbf{w}(\mathbf{p}_{ij}^*; \hat{\mathbf{H}}\mathbf{H}(\tilde{\mathbf{x}}))) + \hat{\beta} + \tilde{\beta}, \quad (14)$$

since the update rule for the photometric parameters is

$$\begin{cases} \hat{\gamma} \leftarrow \hat{\gamma} + \tilde{\gamma} \\ \hat{\beta} \leftarrow \hat{\beta} + \tilde{\beta}, \end{cases} \quad (15)$$

while the update rule for the geometric parameters is given in (5). Therefore, injecting (14) into the original direct geometric visual tracking problem (4), it is then translated to

$$\min_{\theta \in \mathbb{R}^{9+n}} \frac{1}{2} \sum_{j=1}^n \sum_{\mathbf{p}_{ij}^* \in \mathcal{B}_j^*} \underbrace{[\mathcal{I}'_{ij} - \mathcal{I}^*(\mathbf{p}_{ij}^*)]^2}_{d_{ij}} = \frac{1}{2} \|\mathbf{d}(\theta)\|^2. \quad (16)$$

Thus, our visual tracking system consists in estimating the incremental parameters $\theta = [\tilde{\mathbf{x}}^\top, \tilde{\gamma}^\top, \tilde{\beta}]^\top \in \mathbb{R}^{9+n}$ so that $\|\mathbf{d}(\theta)\|$ is minimized. In order to iteratively solve this non-linear optimization problem, an expansion in Taylor series is firstly performed. For that, another key technique to achieve nice convergence properties and higher accuracy of the tracker is to perform an efficient second-order approximation of $\mathbf{d}(\theta)$ [3]. 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 $\mathbf{d}(\theta)$ around $\theta = \mathbf{0}$ is given by

$$\mathbf{d}(\theta) = \mathbf{d}(\mathbf{0}) + \frac{1}{2}(\mathbf{J}(\mathbf{0}) + \mathbf{J}(\theta))\theta. \quad (17)$$

The Jacobians can be found in [3] where this approximation was used to solve the problem (4). Here, we extend such a technique to estimate the illumination parameters.

Consider the j -th image block. The current Jacobian $\mathbf{J}(\mathbf{0})$ can be written as

$$\mathbf{J}(\mathbf{0}) = [\mathbf{J}_x(\mathbf{0}), \mathbf{J}_\gamma(\mathbf{0}), \mathbf{J}_\beta] = [\hat{\gamma} \mathbf{J}_{\mathcal{I}} \mathbf{J}_w \mathbf{J}_x(\mathbf{0}), \mathcal{I}, 1] \quad (18)$$

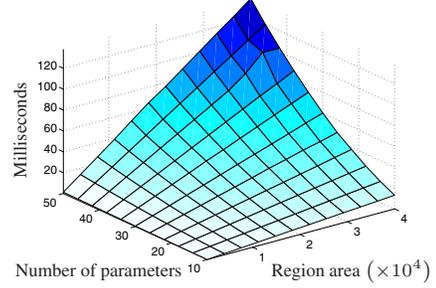


Figure 2. Processing time/iteration for a non-optimized implementation of our method in C.

by applying the chain rule. In this case, $\mathbf{J}_{\mathcal{I}}$ is the image gradient of \mathcal{I} , \mathbf{J}_w is the Jacobian due to a pixel warping and $\mathbf{J}_x(\mathbf{0})$ is the derivative of the warping w.r.t. its parameterization. Correspondingly, the Jacobian $\mathbf{J}(\theta)$ is

$$\mathbf{J}(\theta) = [\mathbf{J}_x(\theta), \mathbf{J}_\gamma(\theta), \mathbf{J}_\beta] = [\hat{\gamma} \mathbf{J}_{\mathcal{I}} \mathbf{J}_w \mathbf{J}_x(\theta), \mathcal{I}^*, 1]. \quad (19)$$

Then, applying a necessary condition for $\theta = \theta^\circ$ to be an extremum of (16) gives

$$\frac{1}{2}(\mathbf{J}(\mathbf{0}) + \mathbf{J}(\theta^\circ))\theta^\circ = -\mathbf{d}(\mathbf{0}). \quad (20)$$

This is not a linear system in θ° because of $\mathbf{J}(\theta^\circ)$. However, due to the suitable parameterization of the homography, the left-invariance property of the vector fields on $\mathbb{S}\mathbb{L}(3)$ [11] is exploited. Therefore, $\mathbf{J}_x(\theta^\circ)\theta^\circ = \mathbf{J}_x(\mathbf{0})\theta^\circ$ and the left hand side of (20) can be written

$$\mathbf{J}'\theta^\circ = \frac{1}{2}[\hat{\gamma}(\mathbf{J}_{\mathcal{I}} + \mathbf{J}_{\mathcal{I}^*})\mathbf{J}_w\mathbf{J}_x(\mathbf{0}), \mathcal{I} + \mathcal{I}^*, 2]\theta^\circ. \quad (21)$$

As a result, the following linear system is obtained:

$$\bar{\mathbf{J}}'\theta^\circ = -\mathbf{d}(\mathbf{0}), \quad (22)$$

whose solution¹ θ° updates iteratively the parameters until convergence. The augmented $\bar{\mathbf{J}}'$ is obtained by stacking appropriately each $\mathbf{J}'_j = \mathbf{J}'$ given in (21), in order to take into consideration all image blocks, $j = 1, 2, \dots, n$. The matrix $\bar{\mathbf{J}}'$ is then sparse in the columns corresponding to γ . As stated, the Hessians were never computed explicitly.

4. Experimental Results

First of all, we emphasize that the proposed algorithm does not require any off-line training phase and also that bundle adjustment is not performed in any case. In the sequel, the photometric error is defined as the RMS of the difference image between \mathcal{I}' and \mathcal{I}^* .

To the authors' knowledge, the existing efficient direct image alignment techniques tackle essentially affine lighting variations. In order to show the generality of the proposed method we compared it with DIRT [2],

¹obtained in the least-squares sense by solving its normal equations $\bar{\mathbf{J}}'^\top \bar{\mathbf{J}}'\theta^\circ = -\bar{\mathbf{J}}'^\top \mathbf{d}(\mathbf{0})$.

which is designed for that particular context. The non-optimized implementation of our method in C runs at about 2.4ms/iteration for an image region of 100×100 and for this affine case (10 parameters to be estimated). See Fig. 2 for the processing times when varying those parameters. These timings are obtained, as for all the results presented in this paper, without applying any filtering technique (e.g. coarse-to-fine strategy, Kalman filter, etc.). The comparison results of an image registration task is shown in Fig. 3. The two images to be aligned present a large displacement in the geometric and photometric parameters, and is adequate to illustrate the improvements gained by the method. Two conclusions can be drawn directly. First, the error obtained by the our technique is always smaller through iterations. Second, the DIRT got stuck in a local minimum and thus it obtained a higher error at the convergence. With respect to other existing strategies, it is showed in [2] the improvements w.r.t. the well-known Simultaneous Inverse Compositional (SIC) [1]. The strategy [8] did not converge after 100 iterations and was not included in the figure.

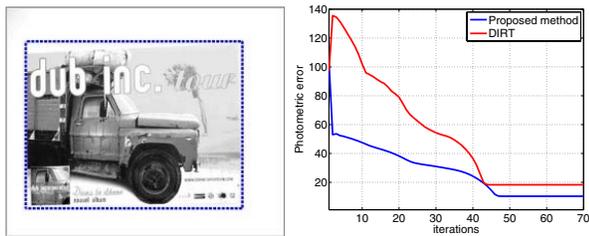


Figure 3. Results of an image alignment task where there are only affine lighting variations. Only the reference image is shown.

With respect to arbitrary illumination changes, we have used the algorithm on several real-world sequences. We report the results from 5 representative sample videos (they are all submitted as supplemental material). See Figs. 4–8. We tested both the discretization approximation and by using a RBF. The obtained results are similar. However, the former is more adequate to real-time systems since it yields a sparse Jacobian matrix. We have set for all the experiments $\epsilon = 10^{-7}$ and image blocks of size 50×50 , whose parameter can be viewed as a compromise between modeling error and computational complexity. The video sequences present severe changes in ambient, diffuse and specular reflections as well as shadows, inter-reflections and glints. In addition, it comprises large geometric displacements and surfaces with unknown reflectance properties. The surfaces ranged from smooth to rough, and including metal and dielectric objects. The unknown light sources are varied in power, type, number and moved in space. Some statistical measures obtained by applying our method to those difficult sequences are given in Fig. 9. To the authors’ knowledge, no existing efficient direct techniques are capable of coping with those challenging scenarios, especially when the object is not near-Lambertian and/or large displacements are carried out. In all case, we tried DIRT, SIC and [8] but they have failed. This includes their variants, e.g. by performing

a photometric normalization with/or a robust error function (a M-estimator). In fact, the experiments showed that, when the latter function leads to a convergence for a given image, it takes at least 2 times more iterations.

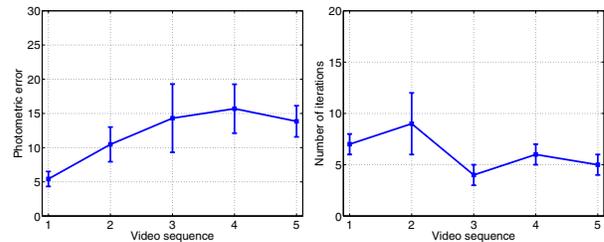


Figure 9. The median error obtained and number of iterations taken by our method for such challenging video sequences.

5. Conclusions

We have presented a new image alignment algorithm to cope with generic illumination changes. Neither the characteristics of the light sources (e.g. number, power, pose) nor about the reflectance properties of the surface (which can be non-Lambertian) are required. We model the lighting changes as a surface which evolves with time. We have tested two methods for approximating this surface and similar results were obtained. An efficient second-order optimization procedure is then applied to jointly obtain the photometric and geometric parameters. Hence, the object may undergo large displacements and the algorithm does not get stuck in irrelevant minima. Extensive experiments confirm the robustness and reliability of our method.

References

- [1] S. Baker, R. Gross, and I. Matthews. Lucas-kanade 20 years on: A unifying framework: Part 3. Technical Report CMU-RI-TR-03-35, Carnegie Mellon Univ., USA, 2003.
- [2] A. Bartoli. Groupwise geometric and photometric direct image registration. In *Proc. of BMVC*, UK, 2006.
- [3] S. Benhimane and E. Malis. Real-time image-based tracking of planes using Efficient Second-order Minimization. In *Proc. of the IEEE/RSJ IROS*, Japan, October 2004.
- [4] J. Carr, W. Fright, and R. Beatson. Surface interpolation with radial basis functions for medical imaging. *IEEE Transactions Med. Imag.*, 16(1), February 1997.
- [5] D. Comaniciu, V. Ramesh, and P. Meer. Real-time tracking of non-rigid objects using mean-shift. In *CVPR*, 2000.
- [6] G. Hager and P. Belhumeur. Efficient region tracking with parametric models of geometry and illumination. *IEEE PAMI*, 20(10):1025–1039, 1998.
- [7] P. J. Huber. *Robust Statistics*. John Wiley & Sons, 1981.
- [8] H. Jin, P. Favaro, and S. Soatto. A semi-direct approach to structure from motion. *The Visual Computer*, 6:377–394, 2003.
- [9] F. Jurie and M. Dhome. Real time robust template matching. In *Proc. of the BMVC*, pages 123–131, UK, 2002.
- [10] R. Szeliski. *Handbook of Mathematical Models in Computer Vision*, pages 273–292. Springer, 2006.
- [11] F. Warner. *Foundations of Differentiable Manifolds and Lie Groups*. Springer, 1994.

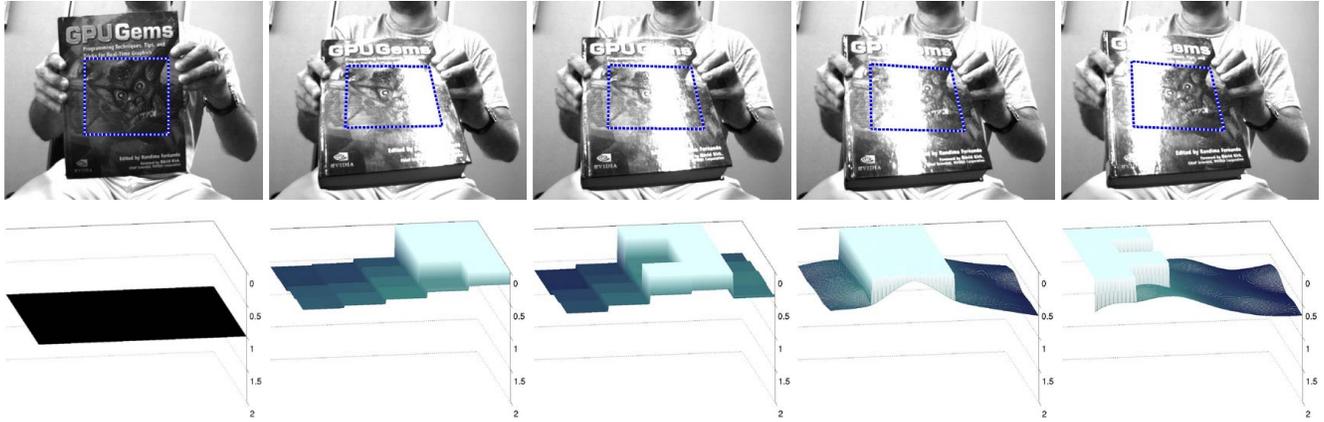


Figure 4. Video 1: (Top) Visual tracking under variable specular reflections produced by a line source, albeit no assumption about its type is made. (Bottom) The reconstructed surface corresponds to the lighting changes for each image. The first three surfaces are obtained if the discretization is applied whilst the last two are obtained if a RBF is used instead. Similar results are obtained for the other videos.



Figure 5. Video 2: Visual tracking under variable illumination produced by different types of light sources and shadows. Notice also the irregular specular reflection.

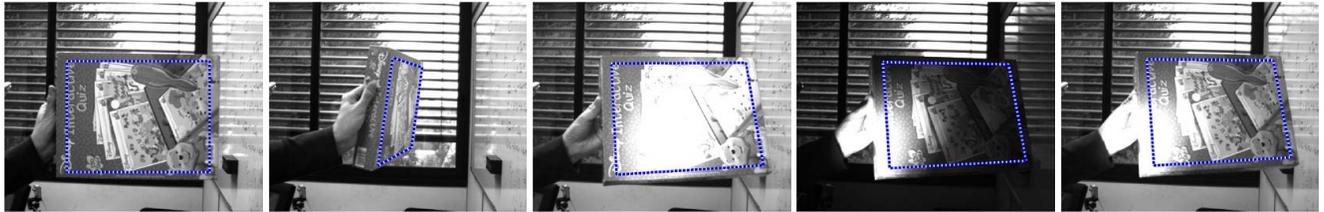


Figure 6. Video 3: Sequence with large surface obliquity and instantaneous changes in lighting. During the visual tracking a large part of the region is occluded by the highlight.



Figure 7. Video 4: Visual tracking under severe specular reflections produced by an unknown number of light sources. The sources as well as the surface perform unknown motions in space.



Figure 8. Video 5: A metallic box is tracked under large changes in rotation and scale while experiencing high specular reflections.