Modeling of photometric variations for matching correspondences, its validity and its application to tracking

Michèle Gouiffès, Christophe Collewet

Modeling of photometric variations for matching correspondences, its validity and its application to tracking.

November 2007

Michèle Gouiffès[1] and Christophe Collewet [2]

(1) IEF, Institut d'Électronique Fondamentale, UMR 8622, Orsay, France
michele.gouiffes@ief.u-psud.fr
(2) Cemagref and INRIA-IRISA, Rennes, France

Abstract

Since modeling reflections in image processing is a difficult task, most computer vision algorithms assume that objects are Lambertian and that no lighting change occurs. Some photometric models can partly answer this issue by computing the illumination changes in small areas of the image, but they often assume that the lighting changes are the same in each point of a window of interest. Through a study based on specular reflection models, such as the Phong and the Torrance-Sparrow ones, we explain explicitly the assumptions on which these models are implicitly based and therefore the situations in which they fail.

In this report, we propose two photometric models, which compensate for specular highlights and lighting variations. They are based on the assumption that illumination changes vary smoothly on the window of interest. The first one is more suitable when specular highlights occur and when small windows of interest are used, as in feature points tracking. The second model compensates for more comprehensive changes such as specular highlights and lighting changes, and can be used on larger areas of the image. Contrary to existing models, the characteristics of the surface of the object and the lighting changes can vary in the area being observed. A part of this report deals with the study on the validity of these modelings with respect to the acquisition configuration: relative locations between the lighting source, the camera and the object, properties of the surface (curvatures and roughness). These models are used to improve feature points tracking in image sequences, by computing simultaneously the photometric and geometric changes. The proposed methods are compared to tracking methods with photometric normalization [34] and the technique proposed by Jin *et al.* [31]. Both of them compensate for affine photometric changes. Since our approach corrects spatial photometric variations, the robustness and the accuracy of the tracking are improved. Experimental results on specular objects demonstrate the robustness of our approaches to specular highlights and lighting changes, without increasing computation times. These procedures provide a good accuracy of the points location during the sequence.

Keywords

Illumination changes, lighting, specular reflection, photometric models, tracking.

Résumé

Puisque la modélisation précise des réflexions dans des images est une tâche difficile, la plupart des algorithmes de vision par ordinateur suppose que les objets sont lambertiens et qu'aucun changement d'éclairage ne se produit. Des modèles photométriques répondent partiellement à ce problème en calculant les changements d'illumination dans de petites fenêtres d'intérêt de l'image, mais ils font généralement l'hypothèse que les changements d'intensité sont identiques en tout point de la fenêtre. A partir d'une étude basée sur des modèles de réflexion spéculaires, comme les modèles de Phong ou de Torrance-Sparrow, nous décrivons explicitement les hypothèses sur lesquelles ces modèles sont implicitement basés, et donc les situations pour lesquelles ils échouent.

Nous proposons ensuite de nouveaux modèles photométriques locaux, qui peuvent compenser différents types de changements d'illumination, tels que des variations de réflexion spéculaire et des changements d'éclairage. Ils sont basés sur l'hypothèse selon laquelle les changements d'illumination varient doucement dans la fenêtre d'intérêt considérée. Le premier s'avère le plus adapté aux variations spéculaires sur de petites fenêtres d'intérêt, comme celles utilisées dans le cadre du suivi de points d'intérêt. Par contre, le second s'avère approprié à la fois pour les changements spéculaires et les variations d'éclairage.

Nous nous attachons à analyser la validité de ces modélisations, en fonction de la configuration d'acquisition : positions relatives entre la source d'éclairage, le capteur et la surface de l'objet, ainsi que les propriétés de la surface. Ces modèles sont ensuite mis en oeuvre pour améliorer le suivi de points caractéristiques et de zones d'intérêt dans des séquences d'images.

Les méthodes proposées sont comparées à la méthode de suivi avec normalisation photométrique et la technique proposée par Jin et al. [31], qui sont robustes aux variations d'illumination affines. Du fait que la modélisation photométrique proposée prend correctement en compte les variations spatiales d'illumination, la robustesse du suivi et le calcul du modèle de mouvement sont améliorés. Des résultats expérimentaux sur des objets spéculaires montrent la bonne robustesse de ces approches vis-à-vis des réflexions et des changements d'éclairage. Elles assurent également une bonne précision de la localisation des points au cours du suivi, sans augmenter de manière significative les temps de calcul.

Mots clé : Changements d'illumination, éclairage, réflexion spéculaire, modèles photométriques, suivi de points et de zones d'intérêt.

Contents

1 Introduction

Computer vision has recently emerged in many fields such as mobile robotics [9], visual inspection, in surgical, agricultural, spatial or underwater domains [11], i.e in various natural environments. For such practical applications, one of the crucial problems lies in the robustness of the low level algorithms with respect to some critical acquisition conditions: blurred images, acquisition noise, illumination changes, reflections. High level algorithms such as 3D reconstruction, active vision or visual servoing for example can be efficiently improved by increasing the robustness of spatial and temporal matching process.

This paper addresses more precisely the problem of robust feature tracking with respect to lighting changes and specular highlights.

When it is possible, the robustness of this procedure can be improved by extracting salient features in the image, such as edges [36], corners [16], lines [6] since they almost only depend on the objects shape or on the luminance gradients. It becomes far more complicated when no mark can be extracted from the observed object, such as in natural environment. In such a context, only points, among possible features, are likely to be easily detectable. However, tracking a point in an image is not a trivial task since the only available information is the luminance of the point and of its neighboring pixels. In such a context, the illumination variations are problematic, since they often make processings fail.

The seminal works in the domain of feature points tracking are due to Lucas and Kanade [23, 33] who assume the conservation of the point luminance during the image sequence [19]. The measure of a correlation function between two successive frames provides the translation motion undergone by the point to track. This motion model theoretically assumes that each point in the window centered around the point to track moves parallel to the image sensor at constant depth. Therefore, this differential tracker assumes a high acquisition frequency and a small motion between two successive frames. However, this technique is still considered to be powerful [32].

Thereafter, the robustness of this tracking approach has been improved, by using some more powerful motion models. For example, the literature has proposed several motion models: affine [29], quadratic [26] and homographic [7, 8]. More recently, [2] has compared several implementations of the differential trackers. Since these formalims are quite more realistic than the translational one, the correlation can be measured between the first and the current frame, so that the tracking errors are not cumulated during the sequence. The accuracy of the tracking can also be verified *a posteriori*, by detecting and rejecting outliers points automatically [34]. Moreover, it is possible to use a robust estimator [26], in order to weight the measurements by an influence function and give less confidence to outliers. This type of methods has proved to be efficient to overcome the problem of occlusions, and to avoid taking noise into account in the correlation measure [28]. Using statistical filters [1, 24] can also improve the robustness of the process, when points trajectories are complex.

The tracking of planes can also be implemented by an efficient second order minimization (ESM) [4].

However, these methods assume that the luminance remains constant between two successive

frames, which is not true. Indeed, most surfaces are not Lambertian and lighting conditions are mostly variable during an image sequence. When color sensors are available, the measure of the correlation function has proved to be more efficient by using color invariants, as in [13]. Under a few assumptions, these attributes do not depend neither on the intensity of lighting, nor on its direction. Nevertheless, their computation requires images with highly saturated colors.

Hager and Belhumeur [15] propose to acquire an image data base of the scene under several illuminations and to use these data to improve the tracking. This method is efficient and no salient feature is needed. Nevertheless, it requires a prior learning step, which can be seen as too restrictive. Very often, one can prefer to achieve a simple local photometric normalization as in [34].

Illumination changes can also be compensated by computing a photometric model which properly fits the luminance variations in restricted areas of the image. Such models have been used in several applications such as optical flow computation [5, 17, 21, 25], object recognition [12], image matching and indexing [14]. For instance in [31], the feature points tracking procedure compensates for affine illumination changes by computing the contrast and illumination variations during the image sequence. Recently in [30], the authors have computed arbitrary illumination changes on a large planar patch in a tracking context, by using an ESM algorithm. However, the main difficulty of the illumination compensation is to balance the trade-off between complexity, and thus computational cost, and adequacy of the model with the real illumination changes.

Moreover, these illumination models are based on several assumptions, about the scene geometry and the surface roughness, which have not been clearly defined yet. In general, the spatial variations of illumination changes, such as contrast and intensity changes, are not taken into account. However, the luminance results from a collection of interaction mechanisms between the light, the matter and the sensor, which are difficult to compute in a computer vision application. The first contribution of this report is to clearly explain the modeling of illumination changes occurring when the relative position between the objects, the lighting and the camera are modified or when the lighting conditions are changed. This analysis is based on some widely used specular reflection models, such as the Phong [27] and the Torrance-Sparrow ones [35]. In particular, we focus on two specific illumination models. The first one, which uses three photometric parameters, is particularly well adapted to compensate for specular highlights and lighting changes when small areas are concerned. The second model, based on six parameters, is more relevant for larger windows. In particular, these models can compensate for spatial variations of illumination changes. They correctly fit the real photometric changes, while requiring a low algorithmic complexity.

Obviously, the validity of these models depends on the object surfaces (orientation, reflectance and roughness), on the location of the lighting sources but also on the viewer direction. Therefore, the second contribution of the paper consists in studying the validity of the proposed models, by considering several simplified acquisition geometries. Finally, we compare our approach to the most commonly used in the literature.

This report is organized as follows. Section 2 focuses on the general modeling of luminance changes, especially in the case of specular reflections and lighting variations. Then, Section 3 deals first with the local illumination models which are used in temporal correspondences

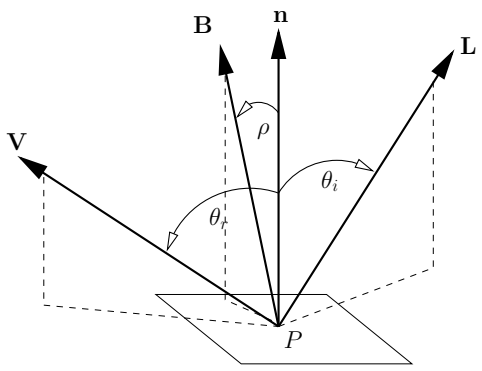


Figure 1: Vectors and angles involved in the reflections description.

matching, then details the two photometric models we propose.

The theoretical validity of the photometric models, and consequently of the tracking procedures, is studied by considering several specific configurations on the viewing geometry and the surface properties. This study is the aim of Section 4. Section 5 details some of the existing trackers, regarding to the illumination model on which they are based. Then, the two proposed trackers are detailed in Section 5.3.

The relevance of our approaches is proved through experimental results, in Section 6. Moreover, a comparison with the standard tracking techniques is also performed, in terms of robustness, location accuracy and convergence of the tracking.

2 Modeling of luminance changes

In this section, we detail the description of the luminance, while referring to physical models largely used in image synthesis and image analysis. Then, starting from this modeling, we focus particularly on the luminance changes occurring between two images of the same scene, acquired for example during an image sequence. Let us notice that we do not consider the modeling of luminance changes caused by the acquisition process (for instance distortion due to the objective, blur), but only on those due to illumination changes.

Let us first introduce our notations (see figure 1 which sketches the vectors and the angles). Let be P a point of the object. $\mathbf{V}$ and $\mathbf{L}$ are respectively the viewing and the lighting directions, which form the angles θ_r and θ_i with the normal $\mathbf{n}$ in P. $\mathbf{B}$ is the bisecting line between $\mathbf{V}$ and $\mathbf{L}$, it forms an angle ρ with the normal $\mathbf{n}$. Let f and f' be respectively the images of an object acquired at two different times. A point P of this object projects in image f in p of coordinates (x_p, y_p) and in p' of coordinates (x'_p, y'_p) in the image f' after a relative motion between the camera and the scene. We call δ the vectorial function which links p' to p such that $\delta(p, \boldsymbol{\mu}) = p'$ according to a parameterization described by $\boldsymbol{\mu}$.

2.1 The luminance in the CCD plane

The relationship between the radiance $\mathcal{L}$ of the observed object and the irradiance received by the sensor $\mathcal{I}_c$, is given by [18]

$$\mathcal{I}_c(\lambda) = K_c \mathcal{L}(\lambda). \tag{1}$$

K_c is a scalar which does not depend on the wavelength λ but only on the geometry of the camera such as the focal distance and the aperture. It is generally considered as a constant scalar. Then, the luminance $f(p)$ depends on the spectral sensitivity $\mathcal{S}(\lambda)$ of the sensor

$$f(p) = \int_{\lambda_{\min}}^{\lambda_{\max}} \mathcal{S}(\lambda)\mathcal{I}_c(\lambda, P)d\lambda = K_c \int_{\lambda_{\min}}^{\lambda_{\max}} \mathcal{S}(\lambda)\mathcal{L}(\lambda, P)d\lambda = K_c \int_{\lambda_{\min}}^{\lambda_{\max}} \mathcal{S}(\lambda)\mathcal{E}(\lambda, P)\mathcal{R}(\lambda, P)d\lambda \tag{2}$$

where $\mathcal{R}(\lambda, P)$ is the reflectance of the material and $\mathcal{E}(\lambda, P)$ the illuminant spectrum.

Several expressions of the radiance $\mathcal{L}(\lambda, P)$ have been proposed according to the physical properties of the material and to the scene geometry. Among them, the Lambertian model [22] is undoubtedly the most widely used because of its simplicity and its relevance.

Lambertian model. It expresses the radiance as

$$\mathcal{L}^L(\lambda, P) = \begin{cases} K_d(P)\mathcal{E}(\lambda, P)\mathcal{R}_b(\lambda, P)\cos\theta_i(P) & \text{if } \theta_i(P) \in [-\frac{\pi}{2}, \frac{\pi}{2}] \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

In other words, the radiance in P is expressed as a function of the incident angle $\theta_i(P)$, the diffuse reflectance $\mathcal{R}_b(\lambda, P)$, most often called *body reflection* or *albedo*, and the illumination spectrum $\mathcal{E}(\lambda, P)$ in P.

Most surfaces also reflect light in a specular manner, not only in a diffuse one, and several functions can be used to model this luminance. We describe here the most interesting one according to our problem.

The Phong model. Phong [27] has described the radiance of specular surfaces in a heuristic way. However, this model is simple to use. The radiance is given by

$$\mathcal{L}^P(\lambda, P) = \begin{cases} K_d(P)\mathcal{E}(\lambda, P)\mathcal{R}_b(\lambda, P)\cos\theta_i(P) + K_s(\lambda, P)\cos^n(\rho(P)) + K_a(\lambda, P) & \text{if } \theta_i(P) \in [-\frac{\pi}{2}, \frac{\pi}{2}] \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

It is composed of a diffuse and a specular component and assumes a point light source. The scalar n is inversely proportional to the roughness of the surface and K_s is the specular coefficient of the direct lighting, depending also on the gain of the camera. K_a is the intensity of ambient lighting in P. It is commonly admitted that it is an empirical model but it proves largely interesting for its simplicity, and because it is appropriate for various types of materials, whether they are rough or smooth.

The Torrance-Sparrow model [35]. Contrary to the previous models, this one is based on the optical geometry. However, since it neglects the electromagnetic characteristics of light, it

is valid only when the surface asperity is larger than the light wavelength. The radiance in P is expressed as

$$\mathcal{L}^T(\lambda,P)=\begin{cases} K_d(P)\mathcal{E}(\lambda,P)\mathcal{R}_b(\lambda,P)\,\cos\theta_i(P)+\dfrac{K_s(\lambda,P)}{\cos(\theta_r(P))}e^{(-\rho^2(P)/2\varsigma^2)} & \text{if } \theta_i(P)\in[-\frac{\pi}{2},\frac{\pi}{2}] \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

where ς is the roughness parameter of the model. The Torrance-Sparrow model is viewed as an interesting model because of the good adequacy between simplicity and accuracy compared to physical reality. Let us remark in both cases, for Phong or Torrance models, that the specular term reaches its maximum value for $\rho(P)=0$, that is when $\mathbf{B}$ coincides with $\mathbf{n}$. In the remainder of the paper, we call h this specular term.

Some more advanced formalisms, such as the Beckmann model [3] based on the electromagnetic waves theory can be found in the literature. Nevertheless, this model is difficult to use in practice in computer vision because of the large number of parameters.

2.2 The luminance modeling in an image

Let us note $\mathcal{M}(\lambda)=\mathcal{S}(\lambda)\mathcal{E}(\lambda,P)$ in (2). When the sensor has a linear response and the color of illuminant is constant during the time, $\mathcal{M}(\lambda)$ can be expressed as the product of a gain K_m, which does not depend of the wavelength, with a spectrum shape $e(\lambda)$. In that case, the luminance becomes

$$f(p)=K_cK_m(p)\int_{\lambda_{min}}^{\lambda_{max}}e(\lambda)\mathcal{R}(\lambda,P)d\lambda. \tag{6}$$

According to the reflection models described previously, $\mathcal{R}(\lambda,P)$ is composed of a diffuse reflectance $\mathcal{R}_b(\lambda,P)$ and a specular term directly related to the illuminant. Let us write $a(p)$ the following term:

$$a(p)=\int_{\lambda_{min}}^{\lambda_{max}}e(\lambda)\mathcal{R}_b(\lambda,P)d\lambda. \tag{7}$$

Since it depends on the *albedo* $\mathcal{R}_b(\lambda,P)$, it is also an intrinsic property of the material. Whatever the photometric model is, the luminance f can be modeled as a sum of three terms which are respectively related to the diffuse, specular and ambiant reflexions:

$$f(p)=K_d(p)a(p)\cos\theta_i(P)+K_s(p)h_f(P)+K_a(p) \tag{8}$$

where $K_d(p)=K_cK_m(p)$ and h_f refers to the specular reflection function which depends on the photometric model (see 2.1): it can be either a cosine fonction (Phong) or an exponential one (Torrance-Sparrow). $K_s(p)$ and $K_a(p)$ are the integration values respectively of $K_s(\lambda,P)$ and $K_a(\lambda,P)$ (see (4) and (5)) according to the wavelengths.

According to (8), the illumination changes occurring between two images of the same scene, can be easily deduced.

2.3 The luminance changes between two images of a sequence

Let us first distinguish between the illumination variations due to specular reflection and the illumination changes related to lighting conditions changes.

Specular reflections. They can occur due to a simple motion of the camera with respect to the surface. Then, the incident angle θ_i is constant in P during the time.

Moreover, if no lighting change occurs, the intensities K_d and K_a are also constant. In the same way, $a(\delta(p, \boldsymbol{\mu})) = a(p)$ since this term depends only on the location of P.

However, the specular component h, which depends on the viewing direction via the angle ρ, varies strongly. In those conditions, the luminance f' is given by

$$f'(\delta(p, \boldsymbol{\mu})) = K_d\, a(p)\, \cos\theta_i(P) + h'(\delta(p, \boldsymbol{\mu})) + K_a \tag{9}$$

where h' is the specular function. By subtracting (9) with (8), it yields to the following relationship between the two images

$$f'(\delta(p, \boldsymbol{\mu})) = f(p) + \psi(p) \tag{10}$$

where

$$\psi(p) = h'(\delta(p, \boldsymbol{\mu})) - h(p). \tag{11}$$

Lighting changes and specular highlights. Now, let us consider that some lighting changes ΔK_a, ΔK_d are produced on K_a and K_d respectively. These variations can be due to a shift of the camera gain or a variation of the lighting intensity. Moreover, the incident angle θ_i changes in P according to a function that we call $\Delta\theta_i$. Such variations occur when the object moves according to the light source or when the light source moves. Then, the relative motion between the camera, the surface and the lighting can make the specular term $h'(\delta(p, \boldsymbol{\mu}))$ vary. Thus, the luminance in image f' is expressed as

$$f'(\delta(p, \boldsymbol{\mu})) = K_d^{'}(\delta(p, \boldsymbol{\mu}))\, a(p)\, \cos\theta_i{}'(P) + h'(\delta(p, \boldsymbol{\mu})) + K_a^{'} \tag{12}$$

with:

$$\left\{ \begin{array}{lcl} K_d^{'}(\delta(p, \boldsymbol{\mu})) & = & K_d(\delta(p, \boldsymbol{\mu})) + \Delta K_d(p) \\ \theta_i^{'}(P) & = & \theta_i(P) + \Delta\theta_i(P) \\ K_a^{'} & = & K_a + \Delta K_a. \end{array} \right. \tag{13}$$

The specular term $h'(\delta(p, \boldsymbol{\mu}))$ includes the intensity change of the specular coefficient K_s if necessary.

Therefore, by using equations (8) and (12), the relationship between two images of the same scene can be described by two different expressions.

First, it can be written as (10), where the function ψ is given by the following relationship:

$$\begin{array}{rcl} \psi(p) & = & a(p)(K_d'(\delta(p, \boldsymbol{\mu}))\cos(\theta_i(p) + \Delta\theta_i(p)) - K_d\cos\theta_i(p)) + \\ & & h'(\delta(p, \boldsymbol{\mu})) - h(p) + \Delta K_a \end{array} \tag{14}$$

In that case, the function $\psi(p)$ depends on $a(p)$ and thus on the *albedo* of the material, closely related to its reflectance.

Second, the luminance change can be expressed by the following relationship

$$f'(\delta(p,\boldsymbol{\mu})) = \lambda(p)f(p) + \eta(p) \tag{15}$$

where:

$$\begin{cases} \lambda(p) &= -\dfrac{(K_d(\delta(p,\boldsymbol{\mu})) + \Delta K_d(\delta(p,\boldsymbol{\mu})))\cos(\theta_i(P) + \Delta\theta_i(P))}{K_d \cos\theta_i(P)} \\ \\ \eta(p) &= -(h(p) + K_a)\lambda(p) + h'(\delta(p,\boldsymbol{\mu})) + K_a + \Delta K_a. \end{cases} \tag{16}$$

In the remainder of the paper, it is important to notice that both functions $\lambda(p)$ and $\eta(p)$ do not depend on $a(p)$, but only on the geometric parameters. Nevertheless, since this modeling refers to a large number of parameters, their use in computer vision is not straightforward. Indeed, since it depends on the material properties (the roughness of the surface by the means of the specular terms), the functions $\lambda(p)$ and $\eta(p)$ are not easy to compute. Therefore, some simpler models are used in computer vision.

3 Local modeling of illumination changes

Generally speaking, the simplified photometric models rely on the local modeling of luminance changes in small areas of the image, seldom in the whole image. Therefore they are available for image matching or feature points tracking procedures. Let us see from (15), on which assumptions these models are based. We will refer to $\mathcal{W}$ as a window of interest centered in p. We call m an other point belonging to $\mathcal{W}$.

3.1 The luminance constancy

In a large number of applications, it is assumed that the luminance of images from the same scene remains constant during the time [19]. From the radiance models given in Section 2.1, it can be true only for Lambertian objects under constant lighting. In that case, we simply have:

$$f'(\delta(m,\boldsymbol{\mu})) = f(m) \text{ for any } m \in \mathcal{W}. \tag{17}$$

3.2 The affine model

The affine model assumes that $\lambda(p) = \lambda$ and $\eta(p) = \eta$ leading to:

$$f'(\delta(m,\boldsymbol{\mu})) = \lambda f(m) + \eta. \tag{18}$$

According to (16), this model assumes that the incident angles θ_i and $\Delta\theta_i$ are constant in each point of the window of interest. This statement is rigorously true only if the normal $\mathbf{n}$ is the same in each point of $\mathcal{W}$, i.e if the surface is locally planar.

Moreover, both objects and lighting must be motionless. Second, the specular terms h' and h must be constant in $\mathcal{W}$. According to the specular reflection models (4) or (5), this statement is true if the angle ρ is the same in each point and the roughness is constant in $\mathcal{W}$.

This statement is correct for all m, if the specular functions h and h' are equal to zero in each point of $\mathcal{W}$, that is for Lambertian surfaces only.

Now, let us show that the affine model based on the photometric normalization [34] does not depend on the affine photometric changes. Let us recall that it is defined through the following transformation of luminance f

$$\frac{f(m)-\mu_f}{\sigma_f}, \tag{19}$$

where μ_f and σ_f are respectively the average and standard deviation of the luminance in a window of interest $\mathcal{W}$, of size $\mathcal{N}\times\mathcal{N}$.

Indeed, from the affine photometric model, given by (18): we easily deduce a relationship between the average of f' in $\mathcal{W}$ and the average of f:

$$\begin{array}{rcl} \mu_{f'} & = & \frac{1}{\mathcal{N}^2}\sum_{m\in\mathcal{W}}(\lambda f(m)+\eta) \quad = \quad \lambda\left(\frac{1}{\mathcal{N}^2}\sum_{m\in\mathcal{W}} f(m)\right)+\eta \\ \mu_{f'} & = & \lambda\mu_f+\eta \end{array} \tag{20}$$

The standard deviation of f' in $\mathcal{W}$ is also related to the standard deviation of f:

$$\begin{array}{rcl} \sigma_{f'} & = & \sum_{m\in\mathcal{W}}\left(\lambda f(m)+\eta-(\lambda\mu_f+\eta)\right)^2 \\ \sigma_{f'} & = & \lambda\sigma_f \end{array} \tag{21}$$

Therefore, the photometric normalization given by $\frac{f(\delta(m))-\mu_f}{\sigma_f}$ and the use of (18) yields:

$$\frac{f'(\delta(m,\boldsymbol{\mu}))-\mu_{f'}}{\sigma_{f'}}=\frac{\lambda f(m)+\eta-(\lambda\mu_f+\eta)}{\lambda\sigma_f}=\frac{f(m)-\mu_f}{\sigma_f} \tag{22}$$

This ratio does not depend on the affine photometric changes, under the different assumptions that this model requires. Actually, by writing f' as a function of f in (22), we obtain:

$$f'(\delta(m,\boldsymbol{\mu}))=\frac{\sigma_{f'}}{\sigma_f}f(m)+\mu_{f'}-\frac{\sigma_{f'}\mu_f}{\sigma_f} \tag{23}$$

and therefore the photometric normalization model is an affine model with:

$$\left\{\begin{array}{rcl} \lambda = & \frac{\sigma_{f'}}{\sigma_f} \\ \eta = & \mu_{f'}-\frac{\sigma_{f'}\mu_f}{\sigma_f} \end{array}\right. \tag{24}$$

Remark: each ratio of luminance difference only depends on the albedo. Let us consider two points m_0 and m_1 in $\mathcal{W}$. If the lighting parameters K_a, ΔK_a, θ_i, $\Delta\theta_i$, and the specular term h' are constant on $\mathcal{W}$, we can state from (12) that the difference between the luminance of two points m_0 and m_1 in $\mathcal{W}$ does not depend on specular highlights variations:

$$f'(\delta(m_0,\boldsymbol{\mu}))-f'(\delta(m_1,\boldsymbol{\mu}))=K'_d(a(m_0)-a(m_1))\cos(\theta_i+\Delta\theta_i) \tag{25}$$

but still involves the intensity (or camera gain) and the direction of the lighting. Let us now consider a third point m_2 in $\mathcal{W}$. The following ratio is invariant to every kind of illumination change:

$$\frac{f'(\delta(m_0,\boldsymbol{\mu})) - f'(\delta(m_1,\boldsymbol{\mu}))}{f'(\delta(m_0,\boldsymbol{\mu})) - f'(\delta(m_2,\boldsymbol{\mu}))} = \frac{a(m_0) - a(m_1)}{a(m_0) - a(m_2)}. \tag{26}$$

since the ratio of luminance differences only depends on the albedo, which is an intrinsic characteristic of the material. In the same way, any ratio of luminance differences in $\mathcal{W}$ is invariant to illumination changes but depends on the albedo only. $f(m) - \mu_f$, $a(m) - \mu_a$ and $f'(\delta(m)) - \mu_{f'}$ are invariant to highlights occurrence.

As a conclusion, the photometric properties of (18) are true and the relationships (24) are correct only if the specular reflection and the lighting changes are the same in each point of $\mathcal{W}$, as mentioned above. In some cases, these assumptions are not realistic, particularly when $\mathcal{W}$ is the projection of a large and non planar surface of the scene. In addition, the normalization may get noisy for low standard deviation at denominator, that is when the intensities almost saturate or more generally when they are almost homogeneous in $\mathcal{W}$.

In order to reduce those limitations, we propose and validate two photometric models which compensate for spatial illumination variations in $\mathcal{W}$.

3.3 Some illumination models adapted for specular highlights occurrence and lighting changes

The previous illumination models rely on several restricting assumptions that are incorrect for non-planar objets, for instance the constancy of the angle values. Here, we propose two models, where illumination variations are assumed to be varying in the window of interest. The first one is available for small windows of interest, whereas the second one can be used for larger ones.

3.3.1 An illumination model adapted for small areas

It has been shown in section 2 how each kind of illumination changes can be expressed. When only specular highlights occur, the luminance variations between two frames can properly be described by (10).

According to the most widely used reflection models (see (4) and (5)), the function ψ, given by (11) or (14), is not constant in $\mathcal{W}$ since it depends on the viewing and lighting angles and therefore on the normal $\mathbf{n}$ in each point of $\mathcal{W}$. It also depends on the characteristics of the material, such as the roughness of the surface. We admit that ψ can be correctly approximated on $\mathcal{W}$ by a $\mathcal{C}^K$, $K > 1$ function, that we call ψ_{mod}. In that case, ψ_{mod} can be approximated by a Taylor series expansion, performed in a point m of coordinates (x, y), belonging to the neighborhood of p and being the projection of a point M of the scene:

$$\psi(m) \simeq \psi_{mod}(p) + \left.\frac{\partial \psi_{mod}}{\partial x}\right|_p (x - x_p) + \left.\frac{\partial \psi_{mod}}{\partial y}\right|_p (y - y_p). \tag{27}$$

Let us call $\alpha = \left.\frac{\partial \psi_{mod}}{\partial x}\right|_p$, $\beta = \left.\frac{\partial \psi_{mod}}{\partial y}\right|_p$ and $\gamma = \psi_{mod}(p)$. We write $\boldsymbol{\alpha} = (\alpha, \beta, \gamma)$ and $\mathbf{u} = (x - x_p, y - y_p, 1)$. By injecting (27) in (10) we obtain

$$f'(\delta(m, \boldsymbol{\mu})) = f(m) + \boldsymbol{\alpha}^\top \mathbf{u} \tag{28}$$

Compared to the simpler illumination models described previously, this one relies on lower assumptions about the scene. The surface projected onto $\mathcal{W}$ is not assumed to be planar, the parameters K_s and n (or ς) can vary smoothly in the window of interest. Therefore, specular highlights can be different in each point of $\mathcal{W}$.

Nevertheless, this model is more appropriate to deal with specular highlights than to cope with lighting changes. Indeed, when lighting changes are caused (equation (14)) the albedo may vary strongly in $\mathcal{W}$ according to the reflectance of the object, and thus (27) is not true. The approximation of the albedo by a first order polynomial becomes more and more crude for large and very textured surfaces. Therefore, the next section proposes a model which copes with this issue.

3.3.2 An illumination model adapted for large areas

According to (16), the function λ depends on the incident angle, which can highly vary when $\mathcal{W}$ is large or when the object surface is not planar. Likewise, the function η depends on the specular highlights variations, on the intensities and on the incident angle values. Thus, these functions are not constant in each point of $\mathcal{W}$.

However, it is possible to assume that these functions are continuous and derivable in each point m. This statement implies that the surface varies in a smooth way. In addition, the specular terms have to be continuous and derivable, so that the roughness of the material must be continuous and derivable in $\mathcal{W}$. Then, λ and η can be expanded in Taylor series around p. By neglecting the coefficients of high order, these equations become

$$\lambda(m) = \boldsymbol{\lambda}^\top \mathbf{u} \text{ with } \boldsymbol{\lambda} = \left(\left.\frac{\partial \lambda}{\partial x}\right|_p, \left.\frac{\partial \lambda}{\partial y}\right|_p, \lambda(p) \right) \tag{29}$$

$$\eta(m) = \boldsymbol{\eta}^\top \mathbf{u} \text{ with } \boldsymbol{\eta} = \left(\left.\frac{\partial \eta}{\partial x}\right|_p, \left.\frac{\partial \eta}{\partial y}\right|_p, \eta(p) \right) \tag{30}$$

leading to

$$f'(\delta(m, \boldsymbol{\mu})) = \boldsymbol{\lambda}^\top \mathbf{u} f(m) + \boldsymbol{\eta}^\top \mathbf{u} \tag{31}$$

This model can take many kinds of illumination changes into account, due either to highlights or lighting changes. In contrast to the previous models, it supposes that these changes can be different on the same window of interest $\mathcal{W}$. Particularly, the surface involved in the window of interest is not assumed to be planar, the parameters K_d, K_s and the roughness n (or ς) can also vary. Therefore, specular highlights and lighting changes can be different in each point of the window of interest. Nevertheless, the number of parameters which have to be computed is increased. Now, let us study the conditions of validity of (31).

4 Validity of the photometric model

The purpose of this section is to analyze the validity of the photometric model described by (31). First, we consider a quadratic object, of which the local shape is known. We assume that this object is viewed under one lighting source of known location. We compute the real corresponding photometric changes obtained when the lighting source has moved (η and λ given by (16)), for different acquisition conditions:

- the pose of the camera with regard to the object;
- the pose of the lighting source with regard to the object;
- the shape of the surface (value of the curvatures of a quadratic surface);
- the material properties of the object, that is to say its roughness parameter.

Second, we achieve a local approximation of these photometric changes by computing the Taylor series at second order of η and λ. Our photometric model, which is a first order approximation, will be the most adequate when the coefficients of second order of this latter approximation will be null or approximately null. So this study consists in finding the configurations for which these second order coefficient vanish.

4.1 Modeling of the scene geometry

We consider a frame $\mathcal{F}_c$, linked to the camera. A point P of coordinates (X_p, Y_p, Z_P) is located at the center of a region of interest on the object. Let us also consider a point M, of coordinates (X, Y, Z), which is located in the neighborhood of P (see the figure 2). We assume that the surface in P can be described as a fonction of class $\mathcal{C}^2$ leading to the following approximation of the depth in M

$$Z = Z_P + D_X(X - X_P) + D_Y(Y - Y_P) + \frac{1}{2}D_{XX}(X - X_P)^2 + \frac{1}{2}D_{YY}(Y - Y_P)^2 + D_{XY}(X - X_P)(Y - Y_P) \tag{32}$$

where D_X, D_Y are the first derivatives of the surface at the point P. These parameters describe the orientation between the tangent plane of the surface at the point and the CCD plane:

$$D_X = \left.\frac{\partial Z}{\partial X}\right|_P \quad D_Y = \left.\frac{\partial Z}{\partial Y}\right|_P \tag{33}$$

The values D_{XX}, D_{YY} and D_{XY} refer to the second order derivatives of the surface in P

$$D_{XX} = \left.\frac{\partial^2 Z}{\partial X^2}\right|_p \quad D_{YY} = \left.\frac{\partial^2 Z}{\partial Y^2}\right|_p \quad D_{XY} = \left.\frac{\partial^2 Z}{\partial X \partial Y}\right|_p \tag{34}$$

From (32), we obtain the normal vector in P

$$\mathbf{n} = \left(\frac{\partial Z}{\partial X}, \frac{\partial Z}{\partial Y}, -1\right) \tag{35}$$

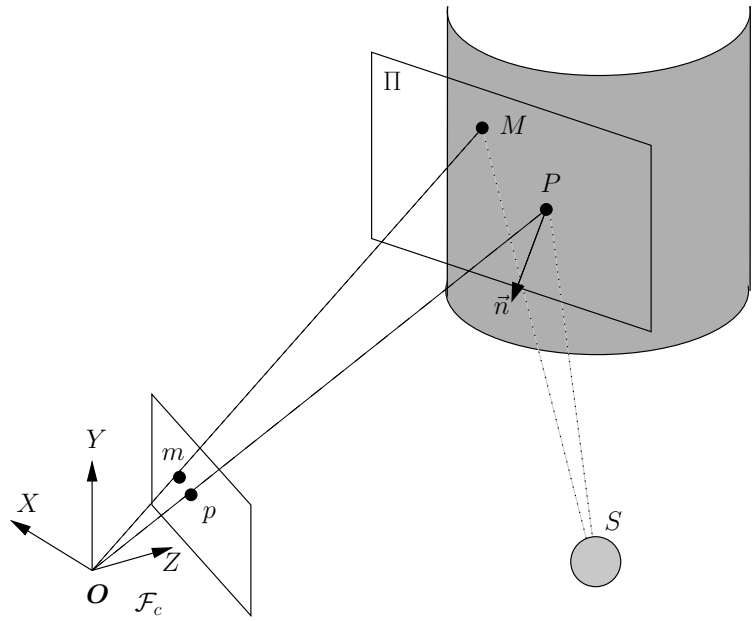


Figure 2: *Modeling of the scene geometry.*

In addition, we suppose that (32) is valid in every point of $\mathcal{W}$.

Given $S = (S_x, S_y, S_z)$ (in the frame $\mathcal{F}_c$) the location of the lighting source, we write $\mathbf{L} = (X - S_x, Y - S_y, Z)$ the vector linking the lighting source S to the point M. Then, the cosinus of the angle formed by S and $\mathbf{n}$ (i.e. $\cos\theta_i$) is written as the scalar product between S and $\mathbf{n}$. By perspective projection and by using Z given by (32), all the geometrical terms (the angle θ_i for example) and the real parameters λ and η given by (16) can be expressed with respect to the pixels coordinates m.

Therefore some approximations and Taylor series expansions are achieved according to the acquisition configurations. In a first step, we study the validity of the approximation of the function λ by (29), which depends on the intensity level and the incident angle of the lighting. The proposed model approximates the variation of this function on $\mathcal{W}$ by a first order polynomial. However, as soon as ψ is concerned, we have seen in 3.3.1 that, when lighting changes are considered, ψ depends on the *albedo*. In this section, we do not take this configuration into account. In addition, in order to simplify this study, we focus on small windows of interest $\mathcal{W}$ which are located near the optical axis of the camera.

4.2 Validity of the modeling of λ

Let be $u = x - x_P$ and $v = y - y_P$. We consider the approximation of λ (see equation (16)) at second order:

$$\lambda(m) = \lambda_1 u + \lambda_2 v + \lambda_3 + \lambda_4 u^2 + \lambda_5 v^2 + \lambda_6 uv. \tag{36}$$

In order to analyze the validity of (31), we study the configurations for which the terms of second order $(\lambda_4, \lambda_5, \lambda_6)$ vanish. The lighting conditions for which they can be neglected are those for which the photometric model fits the illumination changes at best.

We restrict the study to the case of a moving camera which observes a motionless object. A small motion of the direct lighting source $\mathbf{dS} = (dS_X, dS_Y, dS_Z)$ is caused with respect to its initial position S. Several viewing and lighting locations as well as various surfaces curvatures are also considered. Indeed, the only motion of the lighting source causes variations on both terms λ and η. The motion of the lighting source is assumed to be small so that the coefficients λ_i can be expanded in Taylor series around (dS_X, dS_Y, dS_Z). The study is limited to the first order to obtain some useful expressions. Moreover, the following most interesting cases are studied

- the lighting vector coincides with the normal of the surface;
- the lighting source is close to the camera;
- the lighting source is close to the surface.

4.2.1 The lighting vector coincides with the normal of the surface

In this case, we assume that $\mathbf{L} = \tau\mathbf{n}$. For small variations of the lighting angle around the normal, one can show that λ_4, λ_5 and λ_6 are null (their expansion in Taylor series according to $\mathbf{dS}$ yields to null coefficients). Consequently, the approximation of the illumination changes given by (31) is relevant.

4.2.2 The lighting source is close to the camera

In this case, we simply have $S = O$ and thus $\mathbf{V} = \mathbf{L}$. First of all, we consider a planar object, then a non-planar one.

Planar object. When the object is planar, the second order coefficients λ_i become:

$$\left\{ \begin{array}{rcl} \lambda_{4(planar)} & = & -\frac{1}{Z_P}\left(2dS_Z + 2D_X dS_X\right) \\ \lambda_{5(planar)} & = & -\frac{1}{Z_P}\left(2D_Y dS_Y + 2dS_Z\right) \\ \lambda_{6(planar)} & = & -\frac{1}{Z_P}\left(D_X dS_Y + D_Y dS_X\right). \end{array} \right. \tag{37}$$

They are directly related to the error obtained between the photometric model (31) and a more comprehensive approximation of the illumination changes by a second order approximation.

Therefore, these terms vanish when the surface is nearly parallel to the CCD plane. This remark is validated by the example given in the figure 3. This figure shows the real variations of λ (given by (16)) caused locally in a small area $\mathcal{W}$ of a planar surface, without any approximation. The lighting source is moving along dS_X and dS_Y while $dS_Z = 0$. As previously, when the surface of the object is parallel to the camera, that is when $D_X = D_Y = 0$ (figure $3a$), the illumination model is well adapted since the terms $\lambda_{i(planar)}$ vanish. We clearly notice that the shape of the photometric changes is almost planar. In the other hand, when $D_X \neq 0$ and $D_Y \neq 0$ (see figure $3b$), the illumination changes can not be totally compensated by the photometric model and the shape (figure $3b$) is not planar anymore. Only a motion dS_Z of the lighting source along the optical axis (a backward or a forward motion of the camera with respect to the object) yields inevitably some illumination changes which are not compensated by the model. In addition, these latter changes are higher when the camera is close to the surface, as it is shown by the presence of Z_P at the denominator in (37).

Non planar object. In the case of a non-planar object for which (32) is valid, the second order terms of the surface appear in the second-order coefficients:

$$\left\{\begin{array}{rcl} \lambda_4 & = & \lambda_{4(plan)} + 2(D_Y D_{XX} + D_{XY} D_X) dS_Y - 4 D_{XX} dS_Z + 6 D_X D_{XX} dS_X \\ \lambda_5 & = & \lambda_{5(plan)} + 2(D_Y D_{XY} + D_X . D_{YY}) dS_X + 6 D_Y D_{YY} dS_Y - 4 D_{YY} dS_Z \\ \lambda_6 & = & \lambda_{6(plan)} + 2(D_{YY} D_X + D_Y D_{XY}) dS_Y - 2 D_{XY} dS_Z + 2(D_{XX} D_Y + D_X D_{XY}) dSX. \end{array}\right. \tag{38}$$

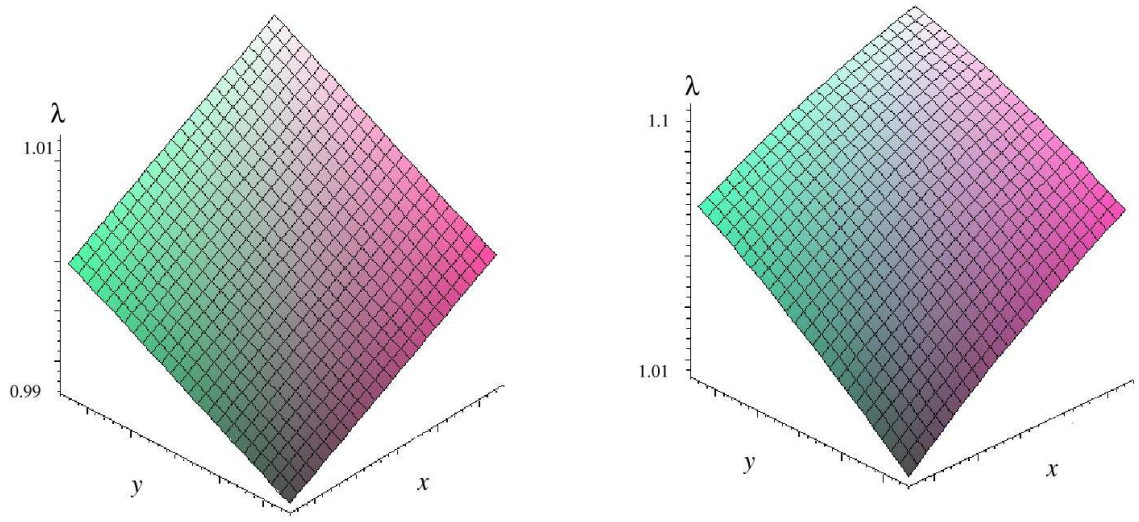
The higher the terms (D_{XX}, D_{XY}, D_{YY}) are, the more the coefficients λ_i vary with respect to a motion $\mathbf{dS}$ of the lighting source. Let us also point out that when the orientation of the tangent plane of the surface in P is parallel to the sensor (D_X and D_Y are close to zero), the motion of the lighting source (dS_X and dS_Y) has a weak influence. In contrast, motions of the camera along the optical axis always cause an error on the second-order coefficients.

4.2.3 The lighting source is close to the surface

Now, let us consider that the lighting source is initially located at a small depth ϵ of the surface so that $S = (X_p, Y_p, Z_p - \epsilon)$. We consider that this distance is small enough to expand the relationships around $\epsilon = 0$. Thus, we obtain the following expression of the coefficients λ_i:

$$\left\{\begin{array}{rcl} \lambda_4 & = & \frac{2Z_P(D_{XX} Z_P D_Y + D_{XY} D_X \epsilon)}{\epsilon^2} dS_Y - \frac{2Z_P(Z_P + D_X^2 \epsilon - Z_P D_X^2 + 2 Z_P D_{XX} \epsilon)}{\epsilon^3} dS_Z + \\ & & \frac{2Z_P(D_X(\epsilon - 1) + D_{XX}(2\epsilon^2 + Z_P \epsilon))}{\epsilon^3} dS_X \\ \lambda_5 & = & \frac{2Z_P D_Y (D_{YY} Z_P \epsilon + 2 D_{YY} \epsilon^2 - 2 Z_P + \epsilon)}{\epsilon^3} dS_Y - \frac{2Z_P(2 D_{YY} Z_P \epsilon - D_Y^2 Z_P + Z_P + D_Y^2 \epsilon)}{\epsilon^3} dS_Z \\ & & + \frac{2Z_P(DYY Z_P D_X + D_{XY} D_Y \epsilon)}{\epsilon^2} dS_X \\ \lambda_6 & = & \frac{Z_P(D_Y D_{XY}(Z_P \epsilon + \epsilon^2) + D_X(\epsilon + 2 D_{YY} \epsilon^2 - 2 Z_P))}{\epsilon^3} dS_Y - \frac{2Z_P(D_Y D_X(\epsilon - Z_P) + Z_P D_{XY} \epsilon)}{\epsilon^3} dS_Z + \\ & & \frac{Z_P(D_X D_{XY}(Z_P \epsilon + \epsilon^2) + D_Y(\epsilon + 2 D_{XX} \epsilon^2 - 2 Z_P)))}{\epsilon^3} dS_X. \end{array}\right. \tag{39}$$

Here again, if the orientation between the surface and the sensor vanishes ($D_X = D_Y = 0$), a motion of the lighting source dS_X and dS_Y does not affect the modeling errors. The



(a) (b)

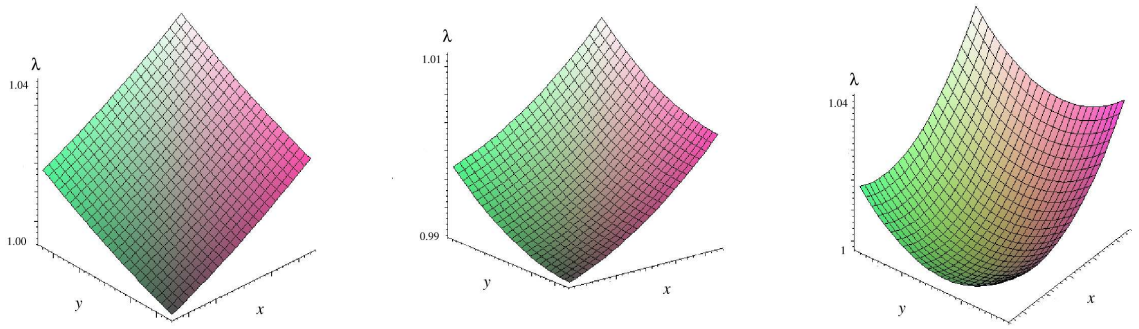
Figure 3: *Examples of illumination changes in $\mathcal{W}$ when the lighting source is close to the camera. (a) $D_X = D_Y = 0$, a motion of the lighting source along X or Y axes is compensated by the photometric model, which forms a plane in $\mathcal{W}$. (b) $D_X = D_Y = 5$ cm, the illumination variations are not perfectly compensated by the model. In each case, the object is planar and the model parameters are the following: $X_P = Y_P = 0$, $Z_P = 100$ cm, $\boldsymbol{dS} = (0.1, 0.1, 0)^\top$.*

approximation of the illumination changes by a first order polynomial is well justified. Moreover, it is more relevant when the depth of the lighting source from the object is higher (high ϵ) than the depth of the camera. In that condition, the contributions of the variations dS_X, dS_Y, dS_Z in the terms λ_i are minimal. However, since the lighting source is considered to be close to the surface, the camera should also be close to the surface. If not, the photometric model is less appropriate. As an example, the figures $4a$, $4b$ and $4c$ show the illumination variations caused by a motion of the lighting source with regard to the surface. In the first case, the depth of the lighting is larger than the depth of the camera. In the second case, the source and the sensor are located at the same distance, and finally in the third case, the source is closer to the surface than the camera is. As a conclusion, the closer the lighting source is with regard to the sensor, the less relevant the proposed photometric model.

To summarize, some conclusions arise from this study about the validity of the estimation of λ by a first order Taylor series expansion.

- It is particularly well adapted when the lighting vector $\mathbf{L}$ coincides with the normal $\mathbf{n}$ in the considered point (see section 4.2.1);

- The approximation is also valid when the orientation of the tangent plane of the surface in P with regard to the sensor plane is low ($\mathbf{V}$ coincides with $\mathbf{n}$), and the second order



(a) (b) (c)

Figure 4: *Examples of illumination variations caused in $\mathcal{W}$ when the lighting source is close to the surface. (a) The lighting source is farther from the surface in comparison with the sensor ($S_Z = 15$, $Z_P = 10$ m). (b) The lighting source and the sensor are located at the same distance to the surface ($S_Z = Z_P = 10m$). (c) The lighting source is closer to the surface in comparison to the sensor ($S_Z = 2$ cm, $Z_P = 10$ m). In the three cases, the parameters used are the following ones: $\epsilon = 0,2m$, $X_P = Y_P = 0$, $D_Y = D_{XY} = 0$, $D_X = D_{XX} = D_{YY} = 0.1$, $\mathbf{dS} = (\mathbf{1}, \mathbf{1}, -\mathbf{0.5})^{\mathbf{T}}$.*

terms of the surface of the object are weak, that is the object is quite planar (see section 4.2.2 for instance).

- When the lighting source is close to the surface, it is more appropriate when the camera is even closer to the surface than the lighting source is (see section 4.2.3).

- The photometric model is more adapted when the depth of camera and lighting source are high (see section 4.2.2).

However, this estimate turns out to be more adequate than an approximation by a constant, which requires the cancellation of the second-order and first-order terms. Obviously, as it is shown by the examples of the previous figures, the illumination changes are not constant.

4.3 Validity of the modeling of η

In order to study the validity of η, expressed by (16), it is necessary to take the specular highlights model into account. Consequently, the material properties of the object have to be considered. For this purpose, we use the specular model of Phong (equation (4) of section 2.1). In order to simplify the equations, we assume a motionless object and constant intensity lighting (K_a and K_d), so that $\lambda(m) = 1$. Consequently, η gets equivalent to the function ψ described by (11). Thus, we study the validity of the following expression:

$$\eta(m') = h_g(M) - h_f(M) \tag{40}$$

After an expansion in Taylor series at second order around p, η is approximated by:

$$\eta(m) = \eta_1 x + \eta_2 y + \eta_3 + \eta_4 x^2 + \eta_5 y^2 + \eta_6 xy \tag{41}$$

where the coefficients η_i depend on the geometry parameters explained in section 4.1. Since the specular highlights function h reaches its maximum when ρ is null, it is interesting to study the validity of the photometric models in this configuration. The initial location of the lighting source is chosen so that the normal $\mathbf{n}$ of the surface coincide with $\mathbf{B}$ (see figure 5).

Similarly to the previous sections, we assume a small motion of the lighting source $\mathbf{dS}$ according to its initial location. This assumption allows us to achieve a Taylor series expansion of (41) around S. Some particular configurations of the scene geometry are studied in order to obtain some simple conclusions about the validity of the models:

- the lighting, the viewing and the normal vectors coincide;
- there is a small orientation between the surface tangent plane and the camera.

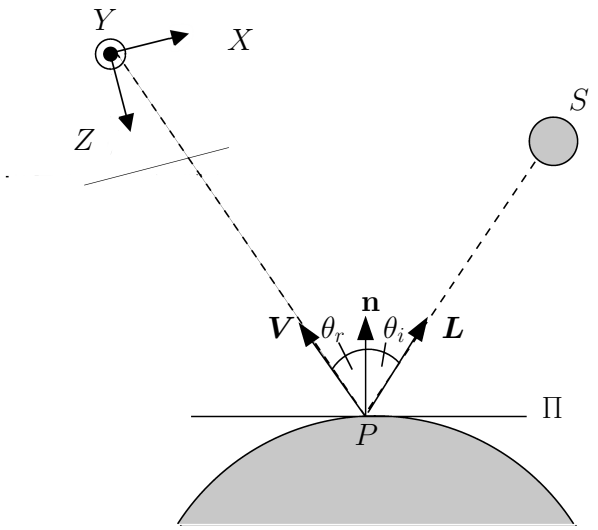


Figure 5: *The normal vector at the surface in point P is the bisecting vector between vectors* $\mathbf{L}$ *and* $\mathbf{V}$.

4.3.1 The lighting, viewing and normals vectors coincide

Initially, before any motion of the lighting source, $\mathbf{L}$, $\mathbf{V}$ and $\mathbf{n}$ are equal. Consequently, the tangent plane at the object surface is parallel to the sensor plane ($D_X = D_Y = 0$) and the lighting angle θ_i is null. Let us consider a non-planar object the surface of which can be described by (32). Unfortunately, even in this simple case, the expressions of coefficients η_4, η_5 and η_6 are far too complicated to deduce any useful information about the validity of the photometric model. In that context we have to focus on some particular configurations, firstly when the lighting source is close to the surface, secondly when the sensor is close to the surface.

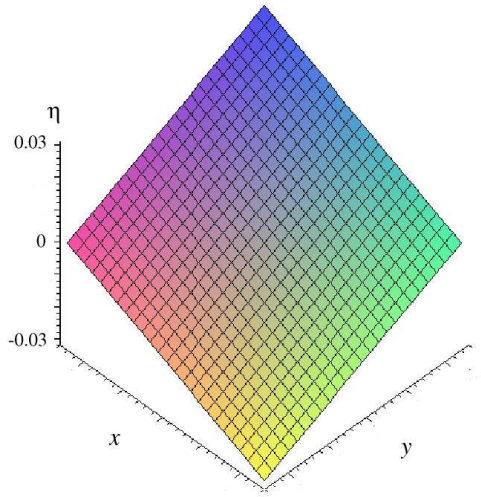


Figure 6: *Example of variation of η when the sensor is close to the surface $Z_P = 10$ cm. The object is not planar, D_X=0, D_Y=0, D_{XX}=0.1 cm, D_{YY}=0.1 cm, D_{XY}=0. The motion of the lighting source is* $\mathbf{dS} = (\mathbf{1}, -\mathbf{1}, -\mathbf{1})^{\mathbf{T}}$ *and* $\epsilon = 100$.

1-The lighting source is close to the surface.

When the lighting source is close to the surface, i.e at a small distance $Z_S = Z_P - \epsilon$, the parameters η_i can be expanded in Taylor series around $\epsilon = 0$. All computations done, the values η_i are expressed as follows:

$$\begin{cases} \eta_4 &= -n\left(2D_{XX} + \frac{1}{Z_P}\right) dSZ \\ \eta_5 &= -n\left(2D_{YY} + \frac{1}{Z_P}\right) dSZ \\ \eta_6 &= -nD_{XY} dSZ \end{cases} \tag{42}$$

When the lighting source is close to the surface, a forward (or a backward) motion dS_Z of the lighting source with respect to the surface always induces some variations of the parameters η_i, whether the surface is planar or not. On the other hand, the parameter $\eta_6 = 0$ when $D_{XY} = 0$, for example for surfaces of revolution (when still assuming that the lighting, viewing and normal vectors coincide). A motion along the Z axis has less influence if the sensor is sufficiently far from the surface and if the surface is rough (in other words when n is low) and planar ($D_{XX} = D_{YY} = D_{XY} = 0$).

2-The sensor is close to the surface (Z_P low). When the sensor is close to the surface, all the coefficients vanish. Consequently, the approximation of the illumination changes by our photometric model is well founded. Figure 6 illustrates this scenario when the surface is not planar. As we can notice, the variations of η are well compensated by a Taylor expansion at first order, since the shape of the function is similar to a plane.

4.3.2 Small orientation of the surface with regard to the sensor plane

In the case of a small orientation between the surface and the sensor (small D_X and D_Y), the coefficients η_i can be expanded in Taylor series around $D_X = D_Y = 0$. We still consider that the normal at the surface $\mathbf{n}$ in P coincide with the bisecting vector $\mathbf{B}$ between $\mathbf{L}$ and $\mathbf{V}$ such that the larger the viewing orientation is, the larger the incident angle θ_i is. We only focus on the case of planar objects, the case of non-planar objects is too complex. As previously, several lighting conditions are analyzed.

1-The lighting source is located near the surface $Z_S = Z_P$. The expressions of η_4, η_5 and η_6 are given by:

$$\left\{ \begin{array}{rcl} \eta_4 & = & -\dfrac{n}{Z_P}\left(\dfrac{D_X}{4}(3n+7)dS_X + \dfrac{D_X}{4}(n+1)\,dS_Y + dS_Z\right) \\ \eta_5 & = & -\dfrac{n}{Z_P}\left(\dfrac{D_Y}{4}(n+1)\,dS_X + \dfrac{D_Y}{4}(3n+7)\,dS_Y + dS_Z\right) \\ \eta_6 & = & \dfrac{n}{Z_P}\dfrac{(n+3)}{4}(D_X dS_X + D_Y dS_Y) \end{array} \right. \tag{43}$$

When D_X and D_Y are not null, a motion of the lighting source (dS_X, dS_Y) causes some variations of the parameters η_i. These changes are higher when the material is smooth (high value of n), when the camera is close (Z_P low) to the surface, and when D_X and D_Y are high.

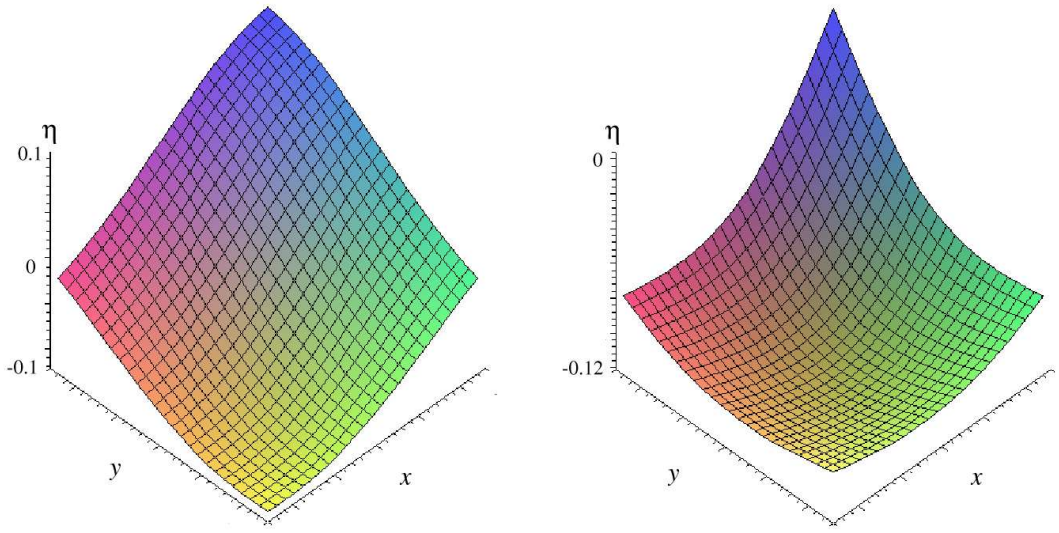
This is illustrated by figures 7*a* and 7*b*, which show respectively two examples of variation of η when the orientation of the tangent plane of the surface in P is low (figure 7*a*) or high (figure 7*b*).

2-The sensor is close to the surface. When the tangent plane and the CCD plane of the sensor are almost parallel and when the sensor is sufficiently close to the surface (low value of Z_P), then the approximation of η by a first order polynomial is perfectly founded. Indeed, the terms η_4, η_5 and η_6 are not significant. This point is illustrated by the figure 8, which shows an example of the variation of η in a window of interest $\mathcal{W}$. Indeed, the shape of the function is clearly a plane.

To summarize, when $\lambda = 1$, the approximation of the term η (equivalent to ψ in this specific case), by a polynomial of first degree is the more appropriate when one or several of the following conditions are observed:

- the second order terms of the surface are small and the tangent plane orientation is low with regard to the sensor plane;
- the surface is rough;
- the sensor is close to the surface.

In those conditions, the specular highlights variations draw up a plane on the window of interest $\mathcal{W}$. Therefore, these photometric changes are well compensated by the proposed illumination model.



(a) (b)

Figure 7: *Examples of variations of η when the lighting source is close to the surface and when the tangent plane at the surface in point P is weakly oriented (figure (a)) or strongly oriented (figure (b)) with respect to the sensor plane.*

4.4 Discussion

Table 1 provides an overview of the configurations for which the proposed photometric model is adapted (+) or not (-), or when the configuration has no influence (=). Let us finally conclude that the approximations of the photometric functions λ and η by a Taylor expansion at second order are adapted at best when the sensor is close to the surface, or when the lighting or the viewing vectors coincide with the normal. On the other hand, the shape of the surface has to be locally continuous and the surface must be rough enough.

However, the photometric model described in Section 3.2 relies on assumptions that are more restrictive in comparision to our model. Indeed, functions λ and η are assumed to be constant at each point of the window of interest $\mathcal{W}$. That means that not only Taylor's coefficients at second order in (36) and (41) are wrong, but also a part of the coefficients of the first order since they are supposed to be null. The few examples of illumination changes (from figure $3a$ to figure 8) have confirmed these remarks. As a conclusion, the photometric model proposed in section 3.3 is theoretically more accurate that the photometric normalization or the affine model with constant parameters.

The different photometric models can be used in applications where temporal correspondences have to be matched, in order to improve some higher level procedures: 3D reconstruction or active vision for example.

In this report, we address the problem of robustifying feature points tracking with respect to illumination changes. The idea is to correctly compensate for the illumination changes by computing the photometric models, in order to obtain more accurately the geometric deforma-

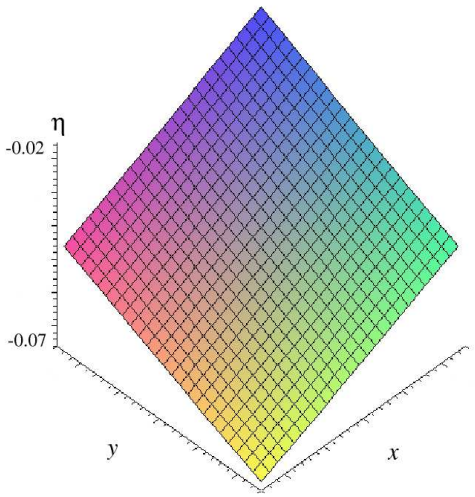


Figure 8: *Example of the variation of η when the tangent plane to the surface and the sensor plane are almost parallel, the camera being close to the surface.*

tions of the windows of interest during the whole sequence. To our knowledge, the two proposed models have not been implemented in such a context. Black *et al.* [5] have used (31) with $\boldsymbol{\eta} = 0$ and (28) in the context of image correction, without any justification.

5 Feature points tracking algorithms

Accurately computing correspondences between two frames or tracking features along an image sequence are two key problems, even though many approaches are available. This section details the tracking techniques involving a photometric model, and proposes two ways to improve them by exploiting the photometric models defined previously in 3.3.1 and 3.3.2.

5.1 Modeling of the geometric deformation

The geometric deformations induced by the relative motion between the camera and the scene are described by a function which models the motion of all the points inside a window of interest $\mathcal{W}$ centered around the point to be tracked p. Therefore, this function is called $\delta(p, \boldsymbol{\mu})$. The feature point tracking procedure consists in computing the parameters $\boldsymbol{\mu}$ such that

$$m' = \delta(m, \boldsymbol{\mu}) \tag{44}$$

according to a photometric model for $m \in \mathcal{W}$. We will show how to compute $\boldsymbol{\mu}$ for the photometric models given in section 3.

Table 1: Overview of the results about the validity of the approximations of λ and η by a Taylor series expansion at first order. +: good approximation. -: bad approximation. = : there is no influence on the validity.

Configuration	λ	η
Lighting vector coincide with the normal	+	+
Viewing vector coincides with the normal	+	+
Rough surface	=	+
Sensor close to the surface and lighting source far from the surface	+	+
Motion of the lighting source along the optical axis	-	-
High values of the second order coefficients of the surface	-	-

5.2 Commonly used tracking methods

5.2.1 The classical approach

The classical feature points tracker, i.e. the KLT technique (for Kanade-Tomasi-Lucas tracker [23,33]) assumes a perfect conservation of luminance at a point during the sequence (see (17)), so we have:

$$f(m) = f'(\delta(m, \boldsymbol{\mu})) \tag{45}$$

However, as seen in section 2, the luminance assumption is not true. Besides, the motion model is also an approximation. Thus, it is more judicious to minimize the following criterion:

$$\epsilon_1(\boldsymbol{\mu}) = \sum_{m \in \mathcal{W}} \left(f(m) - f'(\delta(m, \boldsymbol{\mu}))\right)^2 \tag{46}$$

In order to obtain $\boldsymbol{\mu}$, we suppose that $\boldsymbol{\mu} = \widehat{\boldsymbol{\mu}} + \Delta\boldsymbol{\mu}$, where $\Delta\boldsymbol{\mu}$ expresses a small variation around an estimation $\widehat{\boldsymbol{\mu}}$ of $\boldsymbol{\mu}$. In those conditions, $f'(\delta(m, \boldsymbol{\mu}))$ can be expanded in Taylor series of first order around $\widehat{\boldsymbol{\mu}}$:

$$f'(\delta(m, \boldsymbol{\mu})) = f'(\delta(m, \widehat{\boldsymbol{\mu}})) + \nabla f'^{\top}(\delta(m, \widehat{\boldsymbol{\mu}}))\, J_\delta^{\widehat{\mu}}\, \Delta\boldsymbol{\mu} \tag{47}$$

where $J_\delta^{\widehat{\mu}}$ is the Jacobian of δ according to $\boldsymbol{\mu}$, expressed in $\widehat{\boldsymbol{\mu}}$. We inject (47) in (46), leading to a linear system in $\Delta\boldsymbol{\mu}$, which can be solved iteratively:

$$\left(\sum_{m \in W} \mathbf{v_c}\, \mathbf{v_c}^T\right) \Delta\boldsymbol{\mu} = \sum_{m \in W} \left(f(m) - f'(\delta(m, \widehat{\boldsymbol{\mu}}))\right) \mathbf{v_c} \tag{48}$$

with

$$\mathbf{v_c} = (J_\delta^{\widehat{\mu}})^{\top} \nabla f'(\delta(m, \widehat{\boldsymbol{\mu}})). \tag{49}$$

When considering an affine motion model, $\mathbf{v_c}$ is the vector defined by:

$$\mathbf{v_c} = \left(f'_x, f'_y, x f'_x, x f'_y, y f'_x, y f'_y\right) \tag{50}$$

where f'_x and f'_y are the derivatives of f' with respect to x and y respectively.

5.2.2 Tracking methods robust to affine photometric changes

These approaches are based on the photometric model described in section 3.2. Therefore, instead of minimizing (46), we minimize

$$\epsilon_2(\boldsymbol{\mu}, \lambda, \eta) = \sum_{m \in W} \left(\lambda f(m) + \eta - f'(\delta(m, \boldsymbol{\mu}))\right)^2, \tag{51}$$

where λ and η refer to the parameters of the affine illumination model given by (18). There are two ways to obtain λ and η, either by using (24) or by computing them simultaneously with μ.

The photometric normalization.

Each photometric parameter λ and η is computed from (24). The tracking technique consists in computing $\boldsymbol{\mu}$ as in section 5.2.1 since λ and η are constant. We have to solve:

$$\left(\sum_{m \in W} \mathbf{v}_c \, \mathbf{v}_c{}^\top\right) \Delta\boldsymbol{\mu} = \sum_{m \in \mathcal{W}} \left(\widehat{\lambda} f(m) + \widehat{\eta} - f'(\delta(m, \widehat{\boldsymbol{\mu}}))\right) \mathbf{v}_c \tag{52}$$

Estimation of λ and η : the Jin's technique. In [31], the authors propose to estimate the contrast λ and intensity η simultaneously with the motion model.

Let us call $\boldsymbol{\nu}$ the vector of photometric variations $\boldsymbol{\nu} = (\lambda, \eta)$, and $\mathbf{d}$ the concatenation of $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$. As previously, we suppose a small variation $\Delta\mathbf{d} = (\Delta\boldsymbol{\mu}, \Delta\boldsymbol{\nu})$ of $\mathbf{d}$ around its estimation $\widehat{\mathbf{d}}$ so that $\mathbf{d} = \widehat{\mathbf{d}} + \Delta\mathbf{d}$. Thus, by using (47), we can write (51) as

$$\left(\sum_{m \in W} \mathbf{v}_s \, \mathbf{v}_s{}^\top\right) \Delta\mathbf{d} = \sum_{m \in \mathcal{W}} \left(\widehat{\lambda} f(m) + \widehat{\eta} - f'(\delta(m, \widehat{\boldsymbol{\mu}}))\right) \mathbf{v}_s \tag{53}$$

where $\mathbf{v_s} = (\mathbf{v_c}, \boldsymbol{\nu})$. Unfortunately, as shown in appendix A, the matrix $\sum_{m \in W} \mathbf{v}_s \, \mathbf{v}_s{}^\top$ is ill-conditioned. Therefore, it is required to carry out a preconditioning of this matrix but it depends on the image. That is a drawback of this technique.

On the other hand, this procedure provides a lower computational cost than the photometric normalization, since the averages and standard deviations do not have to be computed in each frame.

In this section, we have presented several feature points tracking techniques; the classical one is based on the luminance constancy, whereas the tracking with normalization and the method proposed by Jin *et al.* are robust to affine illumination variations. In each case, the photometric parameters are supposed to be constant in each window of interest.

In the next section, we propose two tracking procedures which take the spatial variations of illumination changes into account.

5.3 Proposed tracking procedures

The first technique has been defined to compensate for specular highlights and lighting changes on small windows of interest, whereas the second one is its extension to wider windows of interest.

5.3.1 A tracking approach robust to specular highlights

The first tracking method is based on the illumination model given by (28).

Thus, in that case, we have to minimize the following criterion:

$$\epsilon_3(\boldsymbol{\mu}, \boldsymbol{\alpha}) = \sum_{m \in \mathcal{W}} \left(f(m) - f'(\delta(m, \boldsymbol{\mu})) - \mathbf{u}^\top \boldsymbol{\alpha} \right)^2 \tag{54}$$

Be $\mathbf{d} = (\boldsymbol{\mu}, \boldsymbol{\alpha})$. Let us suppose a small displacement $\Delta\mathbf{d} = (\Delta\boldsymbol{\mu}, \Delta\boldsymbol{\alpha})$ around an estimation $\widehat{\mathbf{d}}$ of $\mathbf{d}$, which is the solution of (54). Similarly to the method 5.2.1, $\Delta\mathbf{d}$ is obtained by solving the following linear system:

$$\left(\sum_{m \in \mathcal{W}} \mathbf{v}_p \, {\mathbf{v}_p}^\top \right) \Delta\mathbf{d} = \sum_{m \in \mathcal{W}} \left(f(m) - f'(\delta(m, \widehat{\boldsymbol{\mu}})) - \mathbf{u}^\top \widehat{\boldsymbol{\alpha}} \right) \mathbf{v}_p \tag{55}$$

where the vector $\mathbf{v}_p$ is written as:

$$\mathbf{v_p} = -\left(\mathbf{v_c}, \mathbf{u}\right) \tag{56}$$

for an affine motion model.

Unlike the previous tracker, a preconditioning of the matrix $\left(\sum_{m \in \mathcal{W}} \mathbf{v}_p \, {\mathbf{v}_p}^T \right)$ is not necessary. As shown in appendix A, this matrix is well conditioned.

According to the assumptions of the photometric model (28) described in 3.3.1, this tracking method is appropriate to cope with specular highlights. For small windows of interest, it can also compensate for lighting changes, as soon as the function given by (14) can be approximated by a Taylor series expansion at first order. Since this assumption can be coarse for large windows, the following section proposes a more appropriate algorithm.

5.3.2 A tracking approach robust to specular highlights and lighting changes

Section 3.3.2 has detailed a comprehensive photometric model which compensates for the spatial variations of specular highlights and lighting changes. Let us use this model in order to improve the feature point tracking scheme.

The motion parameter $\boldsymbol{\mu}$ and the reflection parameters $\boldsymbol{\lambda}$ and $\boldsymbol{\eta}$ are obtained by the minimization of the following criterion

$$\epsilon_4(\boldsymbol{\mu}, \boldsymbol{\lambda}, \boldsymbol{\eta}) = \sum_{m \in W} \left(\mathbf{u}^\top \boldsymbol{\lambda} f(m) - f'(\delta(m, \boldsymbol{\mu})) - \mathbf{u}^\top \boldsymbol{\eta} \right)^2 \tag{57}$$

The system can be linearized as in section 5.3.1, with $\mathbf{d} = [\boldsymbol{\mu}, \boldsymbol{\lambda}, \boldsymbol{\eta}]$. Thus, the tracking process consists in solving the following system:

$$\left(\sum_{m \in \mathcal{W}} \mathbf{v}_m \, {\mathbf{v}_m}^\top \right) \Delta\mathbf{d} = \sum_{m \in \mathcal{W}} \left(\mathbf{u}^\top \widehat{\boldsymbol{\lambda}} f(m) - f'(\delta(m, \widehat{\boldsymbol{\mu}})) - \mathbf{u}^\top \widehat{\boldsymbol{\eta}} \right) \mathbf{v}_m \tag{58}$$

where

$$\mathbf{v}_m = (-\mathbf{v}_c, f(m)\mathbf{u}, -\mathbf{u}) \tag{59}$$

The matrix $\sum_{m \in \mathcal{W}} \mathbf{v}_m \, \mathbf{v}_m{}^\top$ can be ill-conditioned (see appendix A), since the values of $\mathbf{v}_m$ are much dissimilar. As for Jin's approach, a preconditioning stage is required.

Moreover, the number of illumination parameters is quite large. Indeed, by using an affine motion model, twelve parameters have to be computed. Obviously, the use of too small windows of interest may alter the accuracy of both photometric and motion models.

The aim of the next section is to validate experimentally our trackers by comparing them with the classical approaches.

6 Validation and experimental results

This section presents some tracking experiments, where the trackers detailed previously are compared through sequences showing geometric and photometric changes simultaneously. First of all, we detail the experimental setup and notations. Second, we analyze the validity of these experimental conditions by comparing experiments on *lab sequences* where ground-truth is available. Finally, the tracking is carried out on *real sequences*.

6.1 Experimental setup

6.1.1 Notations

Throughout this section, we use the following notations:

C	:	the classical tracking approach (section 5.2.1) which assumes that $f'(\delta(m, \boldsymbol{\mu})) = f(m)$
N	:	the tracking with photometric normalization (section 5.2.2) $f'(\delta(m, \boldsymbol{\mu})) = \lambda f(m) + \eta$
J	:	the method proposed by Jin *et al.* (section 5.2.2) $f'(\delta(m, \boldsymbol{\mu})) = \lambda f(m) + \eta$
P_3	:	the tracker which uses three photometric parameters (section 5.3.1) $f'(\delta(m, \boldsymbol{\mu})) = f(m) + \mathbf{u}^\top \boldsymbol{\alpha}$
P_6	:	the tracker which uses six photometric parameters (section 5.3.2) $f'(\delta(m, \boldsymbol{\mu})) = \mathbf{u}^\top \boldsymbol{\lambda} f(m) + \mathbf{u}^\top \boldsymbol{\eta}$

Now, let us detail the setup: the choice of the window's size, the points detection and rejection procedures, the comparison criteria.

6.1.2 Size of the windows of interest

Usually, the choice of the window size $\mathcal{N}$ is based on a trade-off between robustness to noise, computation duration and reliability of the assumptions on which the tracking method is based, such as the planarity of the surface or the constancy of illumination changes. Naturally, it also depends on the application. Here, we consider some sizes from $\mathcal{N} = 9$ to $\mathcal{N} = 35$, since no specific application is concerned.

6.1.3 Rejection process

The points are selected in the first frame of the sequence by the Harris detector [16]. The tracking process computes an affine motion model between the first frame and the current one, as described in section 5.1. They integrate an outliers rejection module, based on the analysis of the convergence of residuals ϵ_i, $i = 1 \dots 4$. A point is rejected as soon as its residuals become greater than a threshold, $\mathcal{S}_{conv} = \mathcal{N}^2 E_{ave}^2$, where E_{ave} is the tolerated luminance variation for each point in $\mathcal{W}$ between f and its modeling. In these experiments, $E_{ave} = 15$.

6.1.4 Comparison criteria

For each image sequence, we can compare the trackers by studying the following criteria:

1. The *robustness* of the tracking, that is to say the number of points that have been tracked during the whole sequence.

2. The temporal evolution of the mean *convergence residues* obtained by the points that are correctly tracked. These two first criteria have to be considered jointly. Indeed, when two methods obtain similar average residues, the more relevant technique is the one which tracks a larger number of points.

3. The temporal evolution of the *reflection parameters* computed by the proposed parametric methods.

   As mentioned in 6.1.3, a point is rejected when its residuals become higher than a threshold. Residuals are commonly used as a comparison criterion, when ground-truth is not available (in [31] or [15] among others). Although some low residues are not an evidence of the tracking correctness (because of potential ambiguities), section 6.2 study their relevance.

4. The *location errors*. In preliminary experiments, where ground-truth is available, a fourth criterion is computed: the average distance (computed on all the points that are correctly tracked by the technique) between the position of the points that is computed by the tracker and the true position. Here again, this criterion has to be considered jointly with the number of points correctly tracked. Indeed, for the same location error, the best technique will be the most robust one.

Next section aims to analyze the relevance of residues as a comparison criterion and gives some first comparison results.

6.2 Validation of the experimental setup on lab sequences

This section studies the validity of our experimental setup by considering lab sequences where ground-truth can be evaluated. We discuss the relevance of criteria 1 and 2.

6.2.1 Computation of the ground-truth

Two techniques of ground-truth extraction are implemented, depending on the shape of the considered object.

Ground-truth for planar objects. When the points to track belong to a planar object, their coordinates in two different images of the sequence are linked together by an homography transformation $\mathbf{H}$, which is described by a 3×3 matrix. Only four points are needed to compute the coefficients of the matrix $\mathbf{H}$ in a linear manner for each frame. However, these points have to be matched accurately between the two frames in order to properly evaluate the tracking techniques, these four points must not depend on the tracking procedure. Thus, they are chosen to be the centers of four white blobs located on the planar surface, which can be easily segmented for each frame of the sequence.

So, on the one hand, the homography matrix is computed between the initial frame and the current one by using four blobs. On the other hand, we estimate the current coordinates of the feature points by applying the homography matrix on the points that have been selected initially in the first frame. Since the homography is known, it becomes easy to obtain the true location of m' from its location m in the first image. Indeed, we have $m' = \mathbf{H}m$ (m and m' are here homogeneous coordinates).

Ground-truth for non-planar objects. In the case of non-planar objects, we use the pose between the camera and the object [10]. This method assumes that we can detect at least four non coplanar points and that we know the 3D location of these points in the object frame. In our context, the four points are four white blobs, which are easy to segment. The whole algorithm is described as follows.

1. Detection of the four non-coplanar blobs $\mathbf{p}^c_{init}$ in the image;

2. Computation of the transformation matrix ${}^c\mathbf{M}_o$ between the object and the camera coordinate frames [10];

3. Intersection of the view line passing through $\mathbf{p}^c_{init}$ with the object in order to obtain $\mathbf{P}^o$; For each experiment, we assume that the object is motionless during the image sequence. Therefore, the coordinates $\mathbf{P}^o$ are constant for each frame;

4. After a motion of the camera, computation of the pose [10] and obtention of the transformation matrix ${}^c\mathbf{M}_o$ between the object and the camera coordinate frames. Consequently, the coordinates $\mathbf{P}^c$ of a point expressed in the camera frame is given by $\mathbf{P}^c = {}^c\mathbf{M}_o\mathbf{P}^o$.

5. Computation of the projection $\mathbf{p}^c$ of $\mathbf{P}^c$ on the CCD plane. Of course, the intrinsic camera parameters are supposed to be known.

6. Comparison between $\mathbf{p}^c$ and the estimate $\widehat{\mathbf{p}^c}$ provided by the considered tracker. Therefore, at each iteration, we compute the euclidean distance between $\mathbf{p}^c$ and $\widehat{\mathbf{p}^c}$, expressed in pixels. When the tracking is perfectly accurate, this distance is null.

6.2.2 Experiments on lab sequences

This section compares the behavior of the methods, in terms of residues, location errors and robustness, for different window sizes, either for planar or non-planar surfaces. Moreover, it is discussed the problem of evaluating the methods in the general case where no ground-truth information is available.

6.2.2.1 Planar surfaces The image sequence depicted on the figure 9 shows a planar surface of size 1×1 meter, on which four blobs have been put. The camera is located approximately 4 meters in front of the object and two lighting sources are located at 2 meters. During the sequence, the camera is motionless and the object is moving with respect to the camera and lightings. The lighting intensity being constant, only specular highlights appear. Because of the low distance between lighting and surface, that is not optimum for the good validity of the photometric models (see section 4).

Robustness. Figure 10 compares the temporal evolution of the number of points that are correctly tracked during the sequence for each approach and for three window sizes: $\mathcal{N} = 9$, $\mathcal{N} = 15$ and $\mathcal{N} = 25$. Simultaneously, table 2*a* page 36 shows the percentage of points correctly tracked until the end of the sequence. Whatever the window size is, C tracks less points than the other approaches. For $\mathcal{N} = 9$, P_3 tracks a larger number of points, which proves its relevance for small window sizes. Unfortunately, for larger ones ($\mathcal{N} = 15$), the performances of P_3 are reduced compared to the other techniques, it tracks correctly one point less than J (see figure 11). On the other hand, for such windows sizes, P_6 is the most competitive method since it tracks around twice more points correctly.

Figure 11 compares more precisely the behavior of each tracking technique, by analyzing both the evolution of the residues during the sequence and the mean location error obtained on the points correctly tracked (the classical method is not taken into account since a too large number of points is lost).

Location errors. The location errors are quite satisfying. They reach only around 1 pixel at the beginning of the sequence and then decrease significantly as soon as outliers points are lost. Indeed, when points of high residuals are lost, the accuracy of the tracking is improved. That shows the reliability of the rejection rule and put in evidence the correlation between residues and accuracy.

Convergence residues. For $\mathcal{N} = 9$, P_3 obtains higher residues that the other approaches. However, these residues are obtained (see figure 10) by averaging the residues of a larger number of points compared to the N and J techniques. To go further in the comparison, figure 12 shows the same criteria as figure 11 while considering only the points tracked simultaneously by each method. Here, for $\mathcal{N} = 9$, P_3 is more accurate. Thus, this method tracks a larger number of points and is more accurate.

For wider windows ($\mathcal{N} = 15$ and $\mathcal{N} = 25$), P_6 obtains the lowest residues, although it tracks a larger number of points. The motion and photometric models are correctly computed.

These results illustrate the fact that the mean residues and the location errors are not always significant without considering the number of points correctly tracked. A method can show low

residues by correctly tracking only few points. This method is although less performant than another one which tracks a hundred of points with slight higher residues.

However, we can also note that, in most cases, the convergence residues evolve roughly in a similar way as location errors. In addition, when residues are low, the location error are also low.

6.2.2.2 Non-planar surfaces In order to study the influence of the surface curvature on the tracking algorithm, a cylinder of radius 7cm has been used. The object and the lighting are motionless and the camera moves. As shown in Section 4, the larger the curvatures are, the less appropriate the photometric models are. The camera is approximately located at 1 meter from the surface and the lighting sources are less close to the surface than the camera is. As noticed in section 4.4, this is one of the favorable configurations for using the photometric models (see table 1). There is no lighting changes but the motion of the camera yields to small specular variations.

Figure 13 shows two images of the cylinder sequence and figure 14 depicts the number of points that are correctly tracked by the procedures versus the frame number (table 2(b) shows the percentage of points which have been correctly tracked until the end of the sequence).

Robustness. Here also, P_3 correctly tracks the largest number of points for small windows (see figure 14). In addition, it remains clearly the most relevant algorithm for $\mathcal{N} = 15$ and $\mathcal{N} = 25$, which was not the case in the previous experiment. Previously the percentage of points tracked was lower.
For $\mathcal{N} = 9$, the J technique obtains very poor performance results, since all the points are lost, whereas even the classical KLT procedure correctly tracks a few points. Obviously, these points can been lost partly because of the ill-conditioning of this technique or because the illumination model is not appropriate (λ and η are constant on $\mathcal{W}$). Since N also tracks a fewer number of points, the latter assumption is quite plausible. These problems will be discussed later in section 6.4. For $\mathcal{N} = 35$, P_6 correctly tracks a larger number of points than the other techniques. Despite the bad conditioning of this method, it is more relevant than J since it better takes the specular changes into account.

Let us now consider the figure 15 which shows the evolution of the convergence residues and location errors.

Location errors. For small windows of interest, (see figure 15 for $\mathcal{N} = 9$) N yields poor accuracy results. In contrast, the use of the P_3 or P_6 photometric models provides an accurate computation of the motion model, i.e a low average of location errors, despite the amount of points tracked (table 2). Thus, the use of an appropriate photometric model improves the computation of the motion model and has yields better accuracy of the points locations during the motion.

Convergence residues. As seen from the results with $\mathcal{N} = 9$, N does not perform well, since its convergence residues are really higher than P_3 and C residues. J loses the whole of

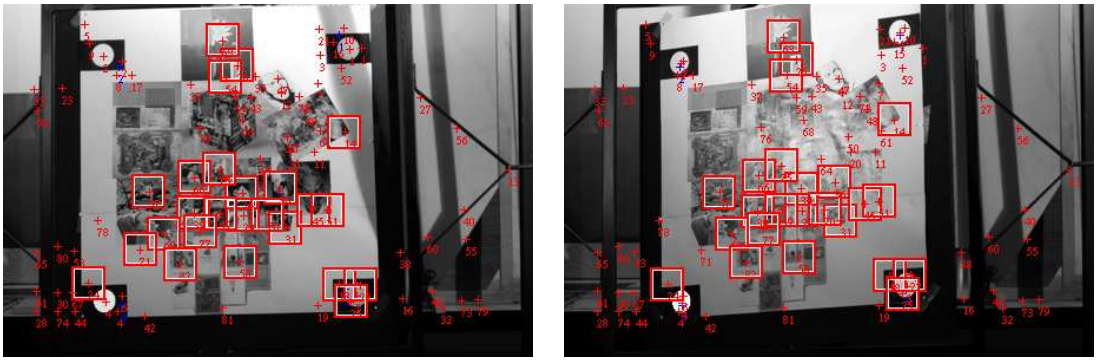


Figure 9: Images of the sequence of a planar surface used to compute the positioning errors.

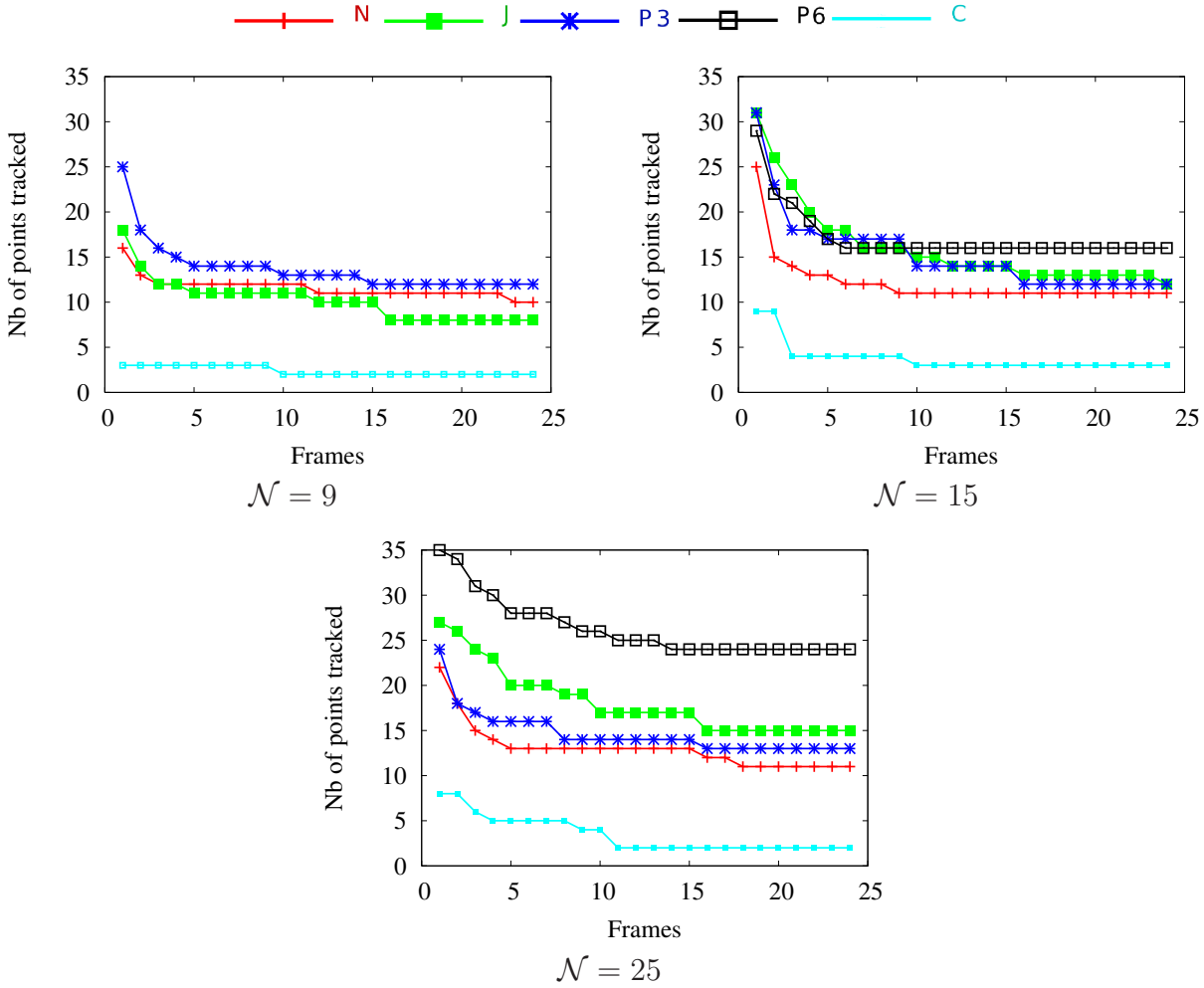


Figure 10: PLANAR SURFACE. Number of tracked points versus the frame number for $\mathcal{N} = 9$ up to $\mathcal{N} = 25$.

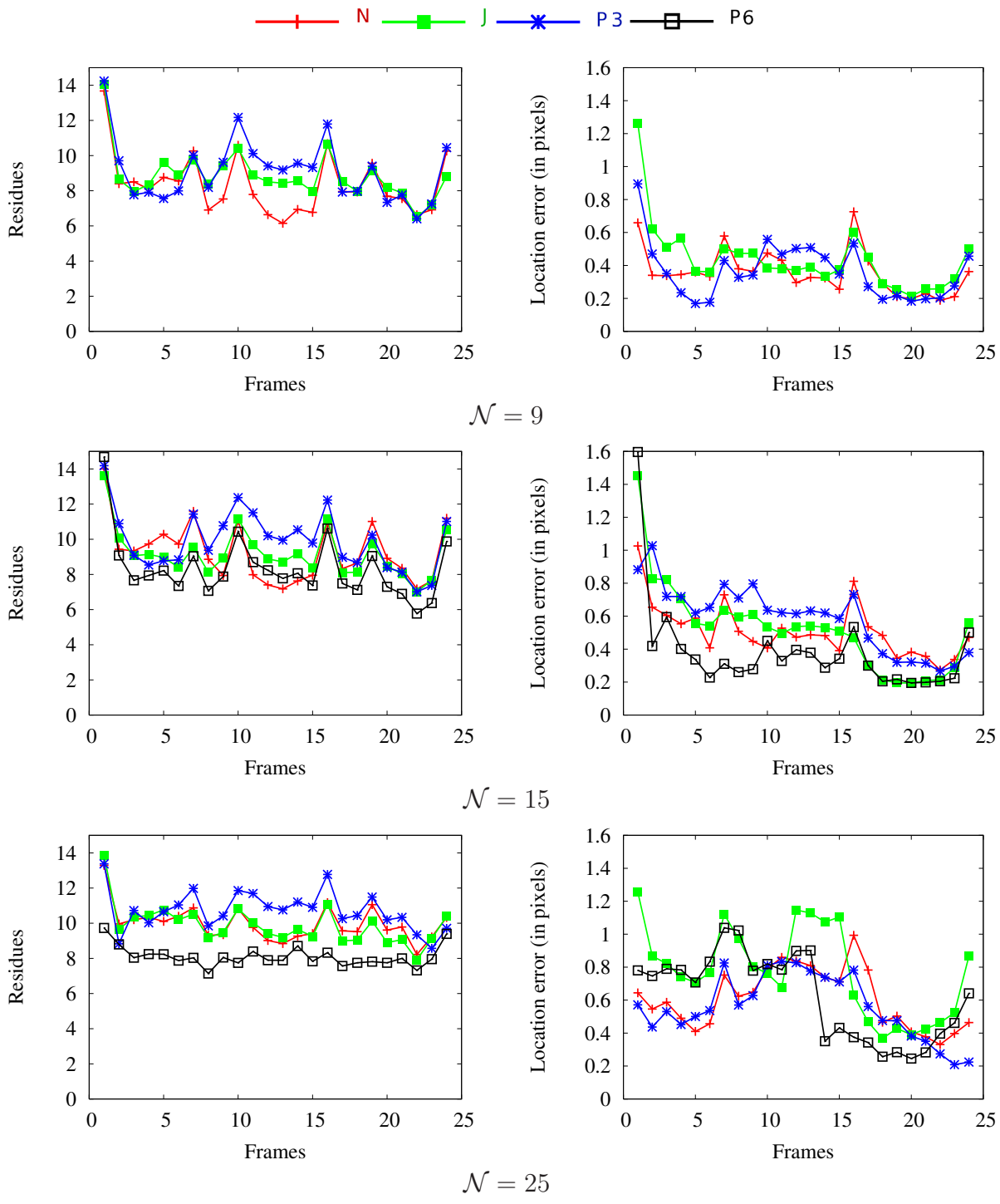


Figure 11: PLANAR SURFACE. Convergence residues and location errors obtained by considering the whole of the points that are tracked by each method.

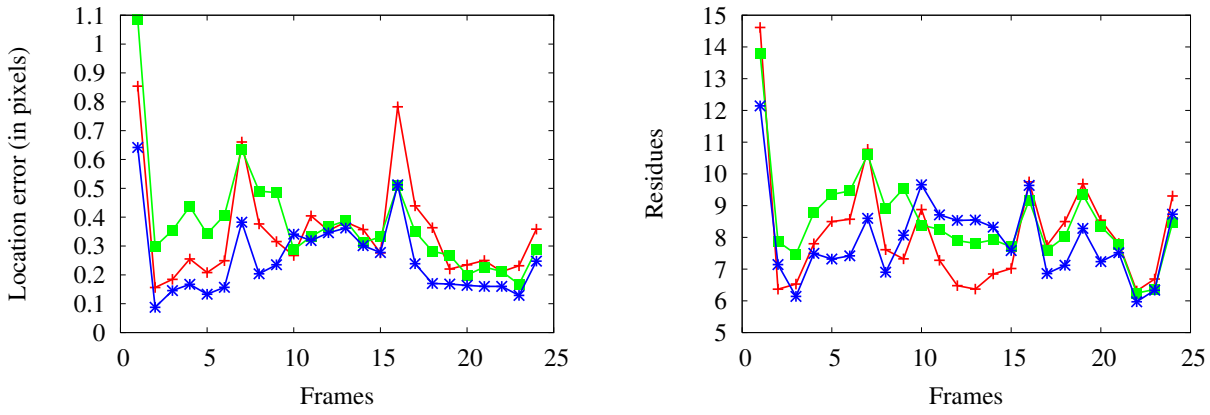


Figure 12: PLANAR SURFACE. Convergence residues and location errors obtained for $\mathcal{N}$=9 by only considering the points that are tracked by each method simultaneously.

Table 2: *Ground-truth is available.* Percentage of points which have been correctly tracked during the sequence (occluded points or points going out of the image are not taken into account).

(*a*) *Planar surface*

$\mathcal{N}$	9	15	25
C	6	9	6
N	29	32	32
J	23	37	43
P_3	**34**	34	37
P_6	0	**46**	**69**

(*b*) *Cylinder*

$\mathcal{N}$	11	15	25	35
C	8	24	14	10
N	2	30	36	40
J	0	8	28	34
P_3	**10**	**40**	**70**	62
P_6	0	0	34	**68**

the points, which explains the vanishing of its residues at the 80^{th} frame. These procedures do not prove to be appropriate for small windows, especially when the considered surface is not planar as it is precisely the case here.

Here also, P_6 provides quite satisfying results on large windows ($\mathcal{N} > 25$). Its convergence residues are globally lower: the geometric and photometric changes are computed more correctly. However, the results of table 2 show that it tracks a lower percentage of points than for planar surfaces, which confirms that the model is more adapted to such kind of surface.

In this sequence, note that the residues assert the results on the location error. Indeed, the lowest residues are obtained for the more accurate tracker (see figure 15).

6.2.2.3 Discussions. As seen in these first experiments, the convergence residues roughly evolve similarly to the mean location error. Generally speaking, as shown on figure 15 for

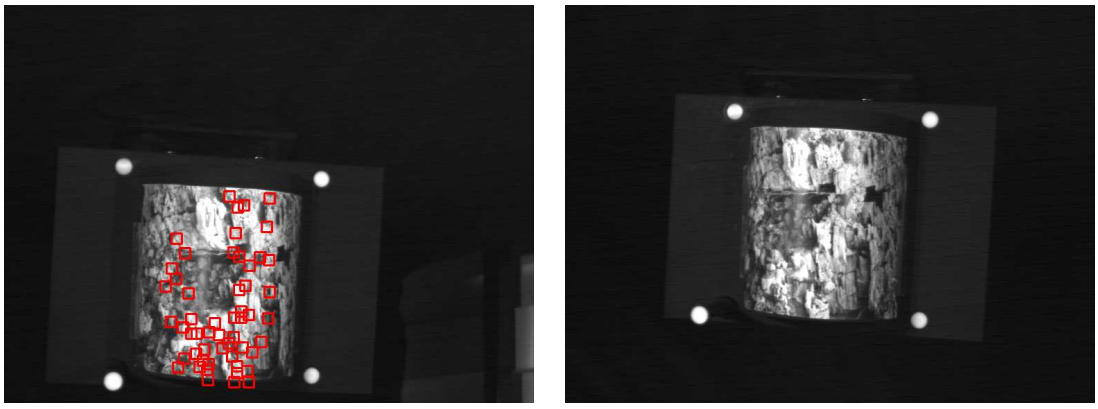


Figure 13: Images of the sequence of a cylinder used to evaluate the accuracy of the tracking procedures.

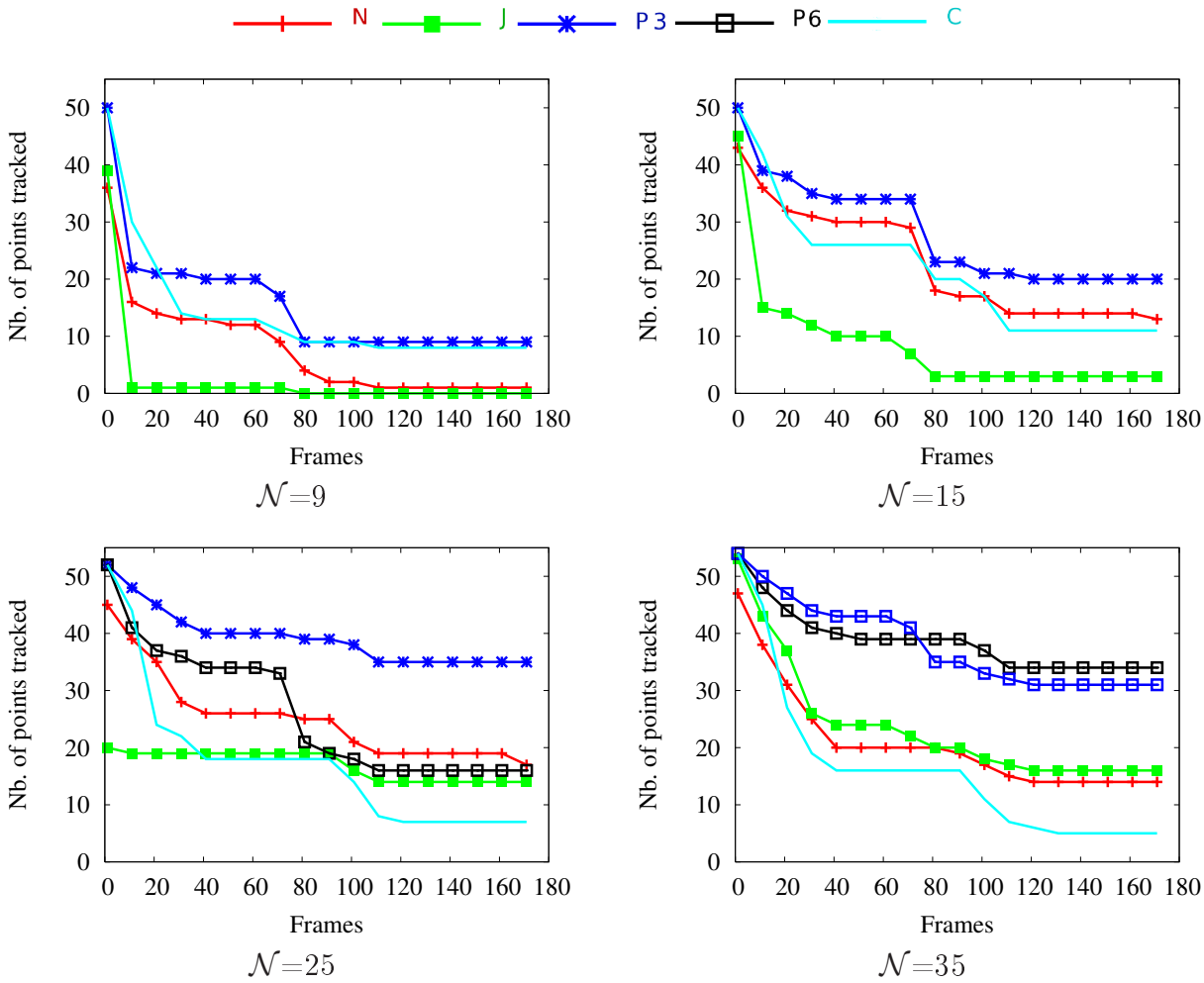


Figure 14: CYLINDER. Number of tracked points versus the frame number, for $\mathcal{N} = 9$ to $\mathcal{N} = 35$

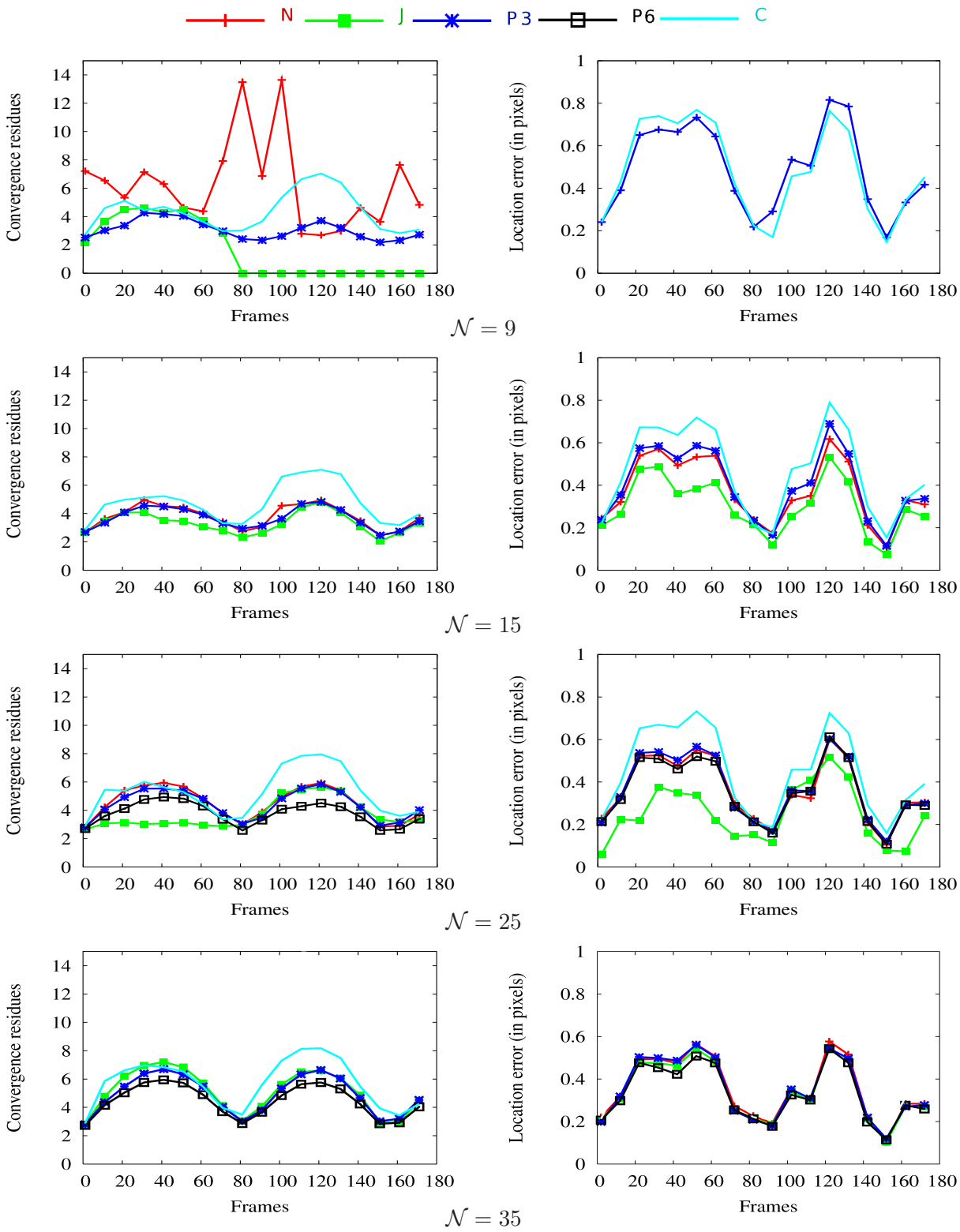


Figure 15: CYLINDER. Convergence residues and location errors obtained by considering the whole of the points that are correctly tracked by each method.

example, the lowest residues are obtained for the most accurate techniques. Although the convergence residues are not exactly an evidence of the good performance of the tracking, they provide reliable information to compare several tracking techniques, especially when ground truth is not available. Of course, this criterion has to be considered jointly with the robustness, i.e the number of points tracked.

In addition, these first experiments have allowed to reach some conclusions concerning the behavior of the tracking experiments.

- For small windows of interest ($\mathcal{N} \leq 15$), P_3 is more performant (lowest convergence residues, more points tracked). In contrast, for larger ones ($\mathcal{N} > 15$), P_6 is the most relevant technique. Obviously, the photometric changes are better compensated for by P_3 for small windows of interest, while they are better taken into account with P_6 on larger windows of interest.

- N and J techniques, which are based on the computation of an affine photometric model are not appropriate for small windows of interest, especially when the surface curvatures are strong (Cylinder for example). This remark confirms the theoretical analysis on the validity of the photometric models (see Section 4) where it has been deduced that the stronger the surface curvatures are, the less efficient these techniques are.

In this section, the ground truth has been obtained, either by computing the homography matrix from markers for planar surfaces or by using the pose and a modeling of the object. Unfortunately, these two approaches cannot be implemented when some real images sequences are considered, residues and robustness are the only available criteria.

6.3 Experiments

Here, experiments are considered on images sequences where no ground-truth is available. However, we have seen in the previous experiments that the convergence residues vary quite similarly to the location error. In conjunction with the number of points correctly tracked, they represent a satisfying way to compare the tracking methods.

Besides, the image sequences are played from the first frame to the last one and then from the last one to the first one in order to qualitatively evaluate the behavior of the techniques by verifying the symmetry of the computed parameters. We analyze successively the robustness, the accuracy of the modeling and the evolution of the photometric parameters. The computation times are also provided.

In order to properly compare the behavior of the trackers on real image sequences, each of them is first tested on sequences where only specular highlights occur. Then, the lighting changes are also taken into account.

6.3.1 Specular highlight occurrence

The two first sequences, *Book* (200 images) and *Cylinder* (150 images) (which are shown respectively on figures 16a and 17a) refer to specular objects, respectively a planar surface and a

cylinder. In each case, the scene is motionless and the lighting conditions do not change but the camera moves, which causes some specular highlights variations at the surface of the objects. In the *Book* sequence, specular highlights disappear during the motion, whereas some other ones appear in different regions of the object. In the case of the cylinder, (figure 17*a*), two lighting sources are taken into account, which causes the appearance of two highly saturated areas. During the motion of the camera, the specular highlights variations are particularly strong in the neighborhood of these two regions.

Robustness. A total number of 97 points is selected initially in the sequence *Book* and 137 in the sequence *Cylinder*. Tables 3*a* and 3*b* report respectively the percentage of points correctly tracked (with respect to the points that are not occluded or those which go out of the image) with respect to $\mathcal{N}$, for each tracking technique.

In sequence *Book*, P_3 tracks the largest number of points for $\mathcal{N} \leq 15$, and for sequence *Cylinder*, it remains the most robust up to $\mathcal{N} = 25$. Consequently, the contribution of P_3 is more significant for non-planar surfaces. Indeed, that is the condition when the illumination changes are the most likely to be different in each point of the window of interest. Besides, this result corroborates the theoretical study of the modelings of section 4 page 15.

P_6 does not converge for small windows of interest, the number of parameters to estimate is too large with respect to the pixels available in $\mathcal{W}$ and the amount of noise.

In the two sequences, J is not more robust than C for $\mathcal{N} \leq 13$. Consequently, this procedure is not adapted for small window sizes, especially when sequences show specular highlights variations.

N is more robust than the classical technique C when the object is planar (sequence *Book*). If not, C provides better results for $\mathcal{N} < 25$. Thus, using an affine photometric model is more appropriate for planar surface, as mentioned in section 4.

Convergence residues. Figures 16*b* and 17*b* compare the average convergence residues obtained respectively during the sequences *Book* and *Cylinder*, for small windows of interest ($\mathcal{N} = 9$). When the object is planar (sequence *Book*), N obtains lower convergence residues than P_3 (figure 16*b*). Nevertheless, the comparison is not fair since the average of the residues is computed on 68 points when P_3 is used and only 33 points when N is carried out. Therefore, in order to obtain a more accurate analysis, figure 16*c* compares the average residues obtained on the same points, i.e the points which have been correctly tracked by N and P_3 simultaneously. These residues are lower for P_3, which confirms the relevance of the photometric model in case of specular highlights occurrence.

In the *Cylinder* sequence, P_3 obtains the lowest residues for $\mathcal{N} = 9$. The residues of N are high at the beginning of the sequence (before the 50^{th} frame) and then decrease when the outliers points are lost. Then, figures 16*d* and 17*c* show the convergence residues obtained with $\mathcal{N}$=35. In such a context, P_6 proves to be the most accurate technique. Indeed, its mean residues (computed on 71 points for the sequence *Book* and 105 points for the sequence *Cylinder*) are lower than those obtained by P_3 (respectively computed on 65 and on 92 points). Not only P_6 tracks a larger number of points, but their location is more accurate.

Photometric model. In order to analyze the illumination variations, we have selected the point A (see figure 16*a*) since it is located on an area of high specular highlights. The first row of

figure 16e shows the luminance values in the window of interest centered around A, whereas the second row refers to the same window after a photometric compensation by the use of P_6, with $\mathcal{N} = 35$. Particularly, let us notice on the first row, that the last image is less luminous than the previous ones, whereas the use of the illumination model has compensated for these changes (the geometric correction has not been achieved here). The spatial changes of the illumination model ($\boldsymbol{\eta}\mathbf{u}^\top$ and $\boldsymbol{\lambda}\mathbf{u}^\top$) are shown on the figure 16f, where the intensity level is proportional to the correction. We notice that the illumination changes are not constant on $\mathcal{W}$. In order to understand the temporal evolution of the photometric models, let us refer to the figure 16g, which displays the parameters λ_i and η_i for $i = 1..3$. The symmetry of the curves attests that for two occurrences of the same image (let us recall that the sequence is played from the first image to the last one and then from the last one to the first one), the photometric parameters remain the same, which asserts the correct convergence of the algorithm. In the sequence *Cylinder*, the evolution of the parameters computed on the point A (which is visible on the image 17a) is shown on the figure 17d. Similarly to the previous experiment, the curves obtained prove the good convergence of the approach. Note that the parameters $\boldsymbol{\lambda}$ also compensates for a part of the specular variations. That may be due to the weakness of the modeling of photometric changes by a first order polynomial.

Computation times. Let us consider a point which is correctly tracked by each procedure for different values of $\mathcal{N}$. The computation times of the techniques are reported in table 4, for $\mathcal{N}$=9, 15 and 35. N and P_6 are the most time-consuming techniques (either because of the computation of the photometric normalization or because of the large number of parameters which have to be approximated). These high values can also be explained by the bad convergence of these techniques when small windows of interest are used. Since the algorithms are iterative, they require a larger number of iterations to converge. For $\mathcal{N}$=9 and 15, the techniques C, J and P_3 obtain some similar computation times.

Up to now, the experiments have been achieved on sequences where specular highlights occur. The next section deals with the comparison of the tracking procedures when lighting changes are also involved.

6.3.2 Lighting variations and specular highlights changes

Let us consider several image sequences showing lighting changes, and for some of them, specular highlights changes.

The sequences *Planar object* and *Marylin* show several textured objects (see the figures 18a and 19a) consisting of several materials (glossy paper, ceramic, metal, cardboard, glass) and lighted by an ambient lightning (the daylight and the fluorescent lamps located at the ceiling) and a direct light source. Then, the sequences *Hill* (figure 20a) and *Corner* [1](figure 21a) show two outdoor scenes *a priori* acquired at different moments of the day. In each case, the camera is moving and the scene is motionless.

[1]These sequences can be found in the image data base CMU/VASC : http://vasc.ri.cmu.edu/idb/html/-motion/index.html

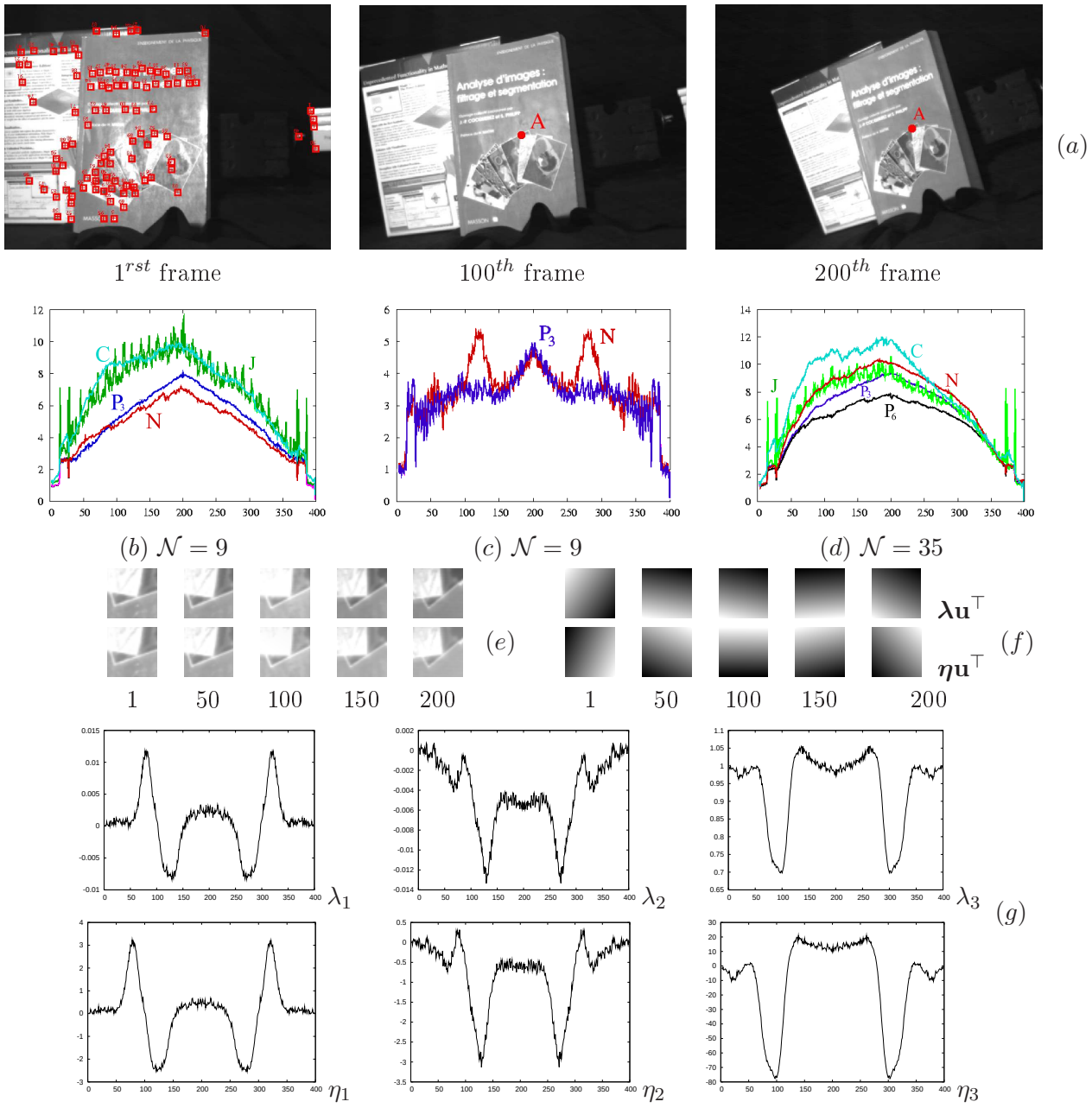


Figure 16: Experiment *Book* (97 points are selected). (a) Frames 1, 100 and 200 in the sequence. (b) Average convergence residues for $\mathcal{N} = 9$. (c) Comparison of the convergence residues obtained on the points that have been tracked simultaneously by N and P_3, for $\mathcal{N} = 9$. (d) Average convergence residues for $\mathcal{N}$=35. (e) Images of the windows of interest centered around the point A : before (first row) and after (second row) a photometric correction by P_6. (f) Illumination parameters of P_3. (g) Evolution of the illumination parameters of P_6.

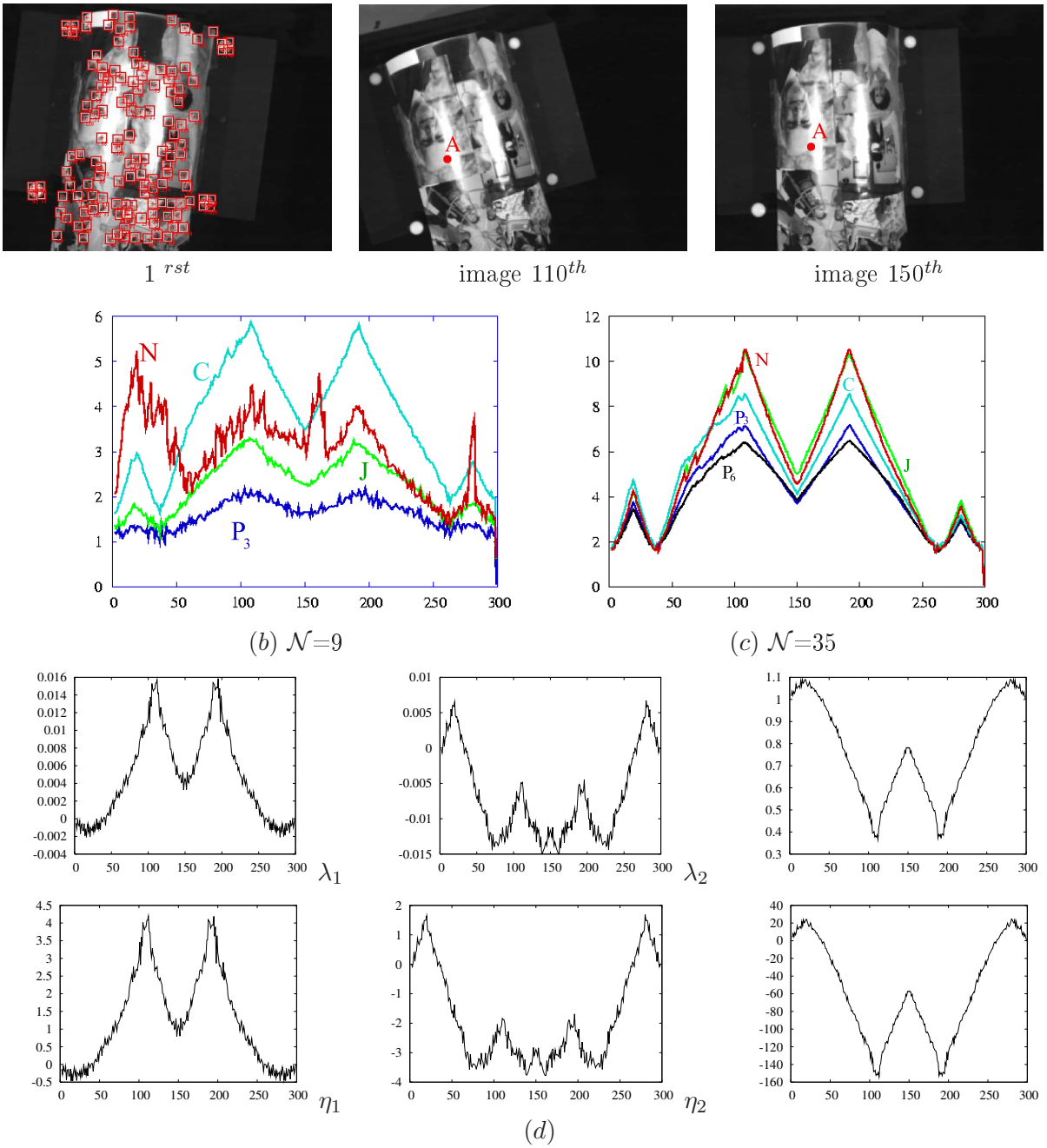


Figure 17: Sequence *Cylinder*. (a) Images of the sequence (137 points have been selected). (b) Evolution of the average of the residues for $\mathcal{N} = 9$. (c) Evolution of the average residues for $\mathcal{N} = 35$. (d) Evolution of the photometric parameters computed at point A by P_6.

Table 3: Percentage of the points that have been tracked up to the end of the sequence with respect to the points which were initially selected, with regard to $\mathcal{N}$, in the case of specular highlights occurrence.

(a) *Book* (97 points selected)

$\mathcal{N}$	9	11	13	15	25	35
C	27.3	30.7	23	18.6	11.8	9.4
N	37.5	53.4	60.9	65.1	63.5	50.6
J	17	28.4	55.2	65.1	77.6	81.2
P_3	**77.3**	**77.3**	**79.3**	**77.9**	78.8	76.5
P_6	-	-	34.5	54.7	**90.6**	**83.5**

(b) *Cylinder* (137 points selected)

$\mathcal{N}$	9	11	13	15	25	35
C	86.5	83.5	81.1	80.2	71.4	60.5
N	40.6	54.9	62.1	69.5	83.3	71.4
J	76.7	80.5	81.8	85.5	85.7	73.9
P_3	**96.2**	**94**	**93.9**	**93.1**	**87.3**	77.3
P_6	-	-	51.5	64.8	81.7	**88.2**

Table 4: Computation times (in ms) of the tracking of one point in the sequence *Book*, for $\mathcal{N}$=9, 15 et 35.

Method	$\mathcal{N}$=9	$\mathcal{N}$=15	$\mathcal{N}$=35
C	1.3	2.7	21
N	4.6	6.8	31.2
J	1.7	3.1	21.7
P_3	1.4	3.2	12.3
P_6	9	8.5	25

In the sequence *Planar object*, the intensity of the direct lighting varies strongly and periodically, with a period of about 20 frames, from a maximum value to a minimal one. The sequence *Marylin* is particularly complicated because of the large motion of the camera and some occlusions. In addition, some intensity variations of the lighting source are deliberately caused: around the iteration 135, the direct light is switched off, which induces some strong illumination changes. To finish, the sequences *Hill* and *Corner* show some lighting changes which are not homogeneous in the image. Here also, we focus both on the robustness of the tracking and on the accuracy of the modeling.

Robustness. We select 58 points in the first frame of the sequence *Planar object*, 156 in *Hill*, 56 points in *Marylin* and 44 in *Corner*. These values are reported respectively in the tables 5a, 5b, 5c, 5d. For small windows of interest ($\mathcal{N}$<15) and whatever the image sequence is, P_3 loses less points than the other techniques. For larger windows, most technique are robust (C is an exception). Nevertheless, for $\mathcal{N} \geq 25$ P_6 is the most robust.

Compared to the previous experiments where only specular highlights changes were caused,

methods N and J appear to be more robust. Indeed, they are always more relevant than C, whereas it was not the case for small windows of interest when only highlights appeared. As a conclusion, these techniques are more appropriate to compensate for lighting changes than to take specular highlights into account.

Convergence residues. The convergence residues obtained with $\mathcal{N} = 9$ by the trackers are shown on figures 18*b*, 19*b*, 20*b* and 21*b*. Here again, these residues evolve in the same way as the illumination changes. This can be clearly seen on figure 18*b*, where they vary periodically with the same frequency as the intensity changes that have been caused.

In the sequence *Marylin*, P_3 tracks a larger number of points (figure 19*b*) from $\mathcal{N} = 9$ up to $\mathcal{N} = 13$. As regards the other sequences, the average residues of J, N and P_3 are comparable, although P_3 residues are computed with a larger number of points than J and N (refer to tables 5*a*, 5*c* and 5*d*).

For sake of clarity, figures 18*c* and 20*c* compare the residues obtained by P_3, N and J on the few points that are correctly tracked by each of these three trackers simultaneously. For the sequence *Planar Object*, P_3 obtains lower residues than N. However, in *Hill*, it is more difficult to reach a conclusion since the residues are almost similar. In *Marylin* sequence, with $\mathcal{N} = 15$ (see the figure 19*c*), the residues of P_3 are the highest. As a first conclusion, P_3 is more appropriate for tracking small windows of interest, especially when only specular highlights are caused (see the previous experiments).

Now, for wider windows (from $\mathcal{N} = 15$ to $\mathcal{N} = 35$), whatever the image sequence is (see the evolution of the residues on the figures 18*d*, 19*d*, 20*d* and 21*c*) the procedure P_6 yields the lowest convergence residues and tracks the largest number of points.

In contrast, for sequences *Planar Object*, *Marylin* and *Corner*, P_3 obtains worse residues than J, N and P_6. On the other hand, it yields the lowest residues in the *Hill* sequence. Indeed, the three first sequences represent scenes which are strongly structured, and where reflectance is likely to show strong edges, whereas the *Hill* sequence is more textured and shows few strong variations of reflectance. Yet, when lighting changes are caused, the performances of P_3 depend on the reflectance changes of the considered surface. Because of the assumption formulated about the smooth illumination changes (in section 3.3.1), the more the reflectance varies the less the proposed modeling compensates for these changes.

Consequently, P_3 proves to be more relevant to take specular highlights into account than to compensate for lighting intensity variations, since the model must compensate for variations which depend on the parameter a depending on the reflectance. More precisely, P_3 approximates a by a plane. Unlike P_3, P_6 procedure does not have to compensate for reflectance changes and can deal with the spatial variations of the illumination changes. However, it is relevant for large windows of interest since a higher number of parameters has to be taken into account.

Photometric parameters. The first row of the figure 18*e* shows the intensities of the window of interest centered around one of the tracked points, which has been selected in an area of high illumination change (point A is visible in the first image of figure 18*a*). The coarse lighting changes are noticeable. The second row is associated to the intensities that have been corrected by the photometric model of P_6. The illumination changes are not visible anymore.

The components of $\boldsymbol{\eta}\mathbf{u}^\top$ and $\boldsymbol{\lambda}\mathbf{u}^\top$ are shown on figure 18f, respectively on the first and second rows. They compensate for the spatial variations of the contrast and for the specular reflection changes (as it can be seen, they are not constant). The evolution of these parameters during the sequence is shown on the figure 18g. It really corresponds to the lighting variations which have been caused. Indeed, we recognize the frequency of 20 iterations between a maximum intensity value and a minimum one.

Computation time. According to the computation times written in the table 6, the techniques N and P_6 are the most time-consuming. For small windows of interest, the computation times of P_6 are high, since this technique does not converge efficiently on small windows of interest. Let us also notice that P_3 obtains larger computation times than in the case of specular highlights (see table 4). This fact shows that, even if this approach is robust it is more adapted to specular highlights than to lighting changes.

6.3.3 Tracking experiments on large windows of interest.

Due to the adequacy of the considered motion model in the differential techniques, such approaches can be extended to track wider windows of interest, as it is done for instance in [4,8,15,20]. Let us notice that the lighting changes are taken into account in [15], where the authors use an image data base acquired offline under various illumination conditions, in order to cope with each possible appearance change. This technique is quite efficient but it requires a prior learning stage, which can be seen as too restrictive.

In section 3.3.2, a comprehensive illumination model has been introduced, which compensates spatial variations of specular and lighting variations. We use this model to track large windows of interest.

Specular reflection. Figure 22a shows an image sequence of a non-planar specular object. An area of the image has been selected by hand in the neighborhood of the areas of high saturation (specular highlights). With $E_{ave} = 25$, only P_3 et P_6 are able to track the window of interest from the beginning to the end of the sequence. In addition, figure 22b, which displays the convergence residues, shows that P_6 models more accurately the specular changes compared to P_3, since it yields lower residues. Let us also notice that C loses the area of interest very quickly compared to N and J.

Lighting changes (Sequence Planar Object). Figure 23a represents an image sequence with shows the scene of the *Planar object* sequence. An area of the image is selected, and the tracking is achieved with $E_{ave} = 15$. Figure 23b refers to the convergence residues obtained. They show that P_6 models more accurately the specular changes which have been caused in comparison to the other techniques, since it yields lowest convergence residues. We have not displayed the residues obtained by C, since this technique was not able to track the area during the whole sequence. P_6 is more adapted to compensate the illumination changes on wide areas of the image.

Tracking of a road sign (lighting changes). The sequence of figure 24(a) has been acquired from a moving car[2]. This sequence is of poor quality, because of noise, gain changes,

[2]This sequence is available on http://vasc.ri.cmu.edu/idb/html/jisct/index.html.

Table 5: *Lighting changes.* Percentage of points which have been correctly tracked during the sequence (the occluded points or points which go out of the image are not taken into account).

(a) *Planar object* (58 points are selected)

$\mathcal{N}$	9	11	13	15	25	35
C	63.8	50	39.7	36.2	8.6	6.9
N	77.6	87.9	91.4	91.4	96.6	93.1
J	67.2	82.8	87.9	51	87.9	89.7
P_3	**100**	**100**	**100**	**100**	96.6	96.6
P_6	48.3	75.9	87.9	94.8	**100**	**100**

(b) *Hill* (156 points are selected)

$\mathcal{N}$	9	11	13	15	25	35
C	49.6	31.1	23.7	20	11.2	9.6
N	45.9	60	63.7	63.7	73.7	75
J	55.6	56.3	63.7	70.4	85.6	93.3
P_3	**74.8**	**74.8**	**74.8**	75.6	86.4	95.2
P_6	-	67.4	70.4	**77.8**	**89**	**97.1**

(c) *Marylin* (56 points are selected)

$\mathcal{N}$	9	11	13	15	25	35
C	0	0	0	0	0	0
N	0	3.6	3.6	21.4	17.9	17.9
J	0	3.6	3.6	7.2	10.7	17.9
P_3	**46.4**	**28.6**	**21.4**	14.3	7.2	3.6
P_6	-	-	-	14.3	**42.9**	**39.3**

(d) *Corner* (44 points are selected)

$\mathcal{N}$	9	11	13	15	25	35
C	90.9	86.4	88.6	88.6	84.1	67.4
N	72.7	63.6	90.9	95.5	86.4	81.8
J	**100**	**100**	**100**	**100**	97.7	88.6
P_3	**100**	**100**	**100**	**100**	**100**	**100**
P_6	34.1	50	72.7	86.4	**100**	**100**

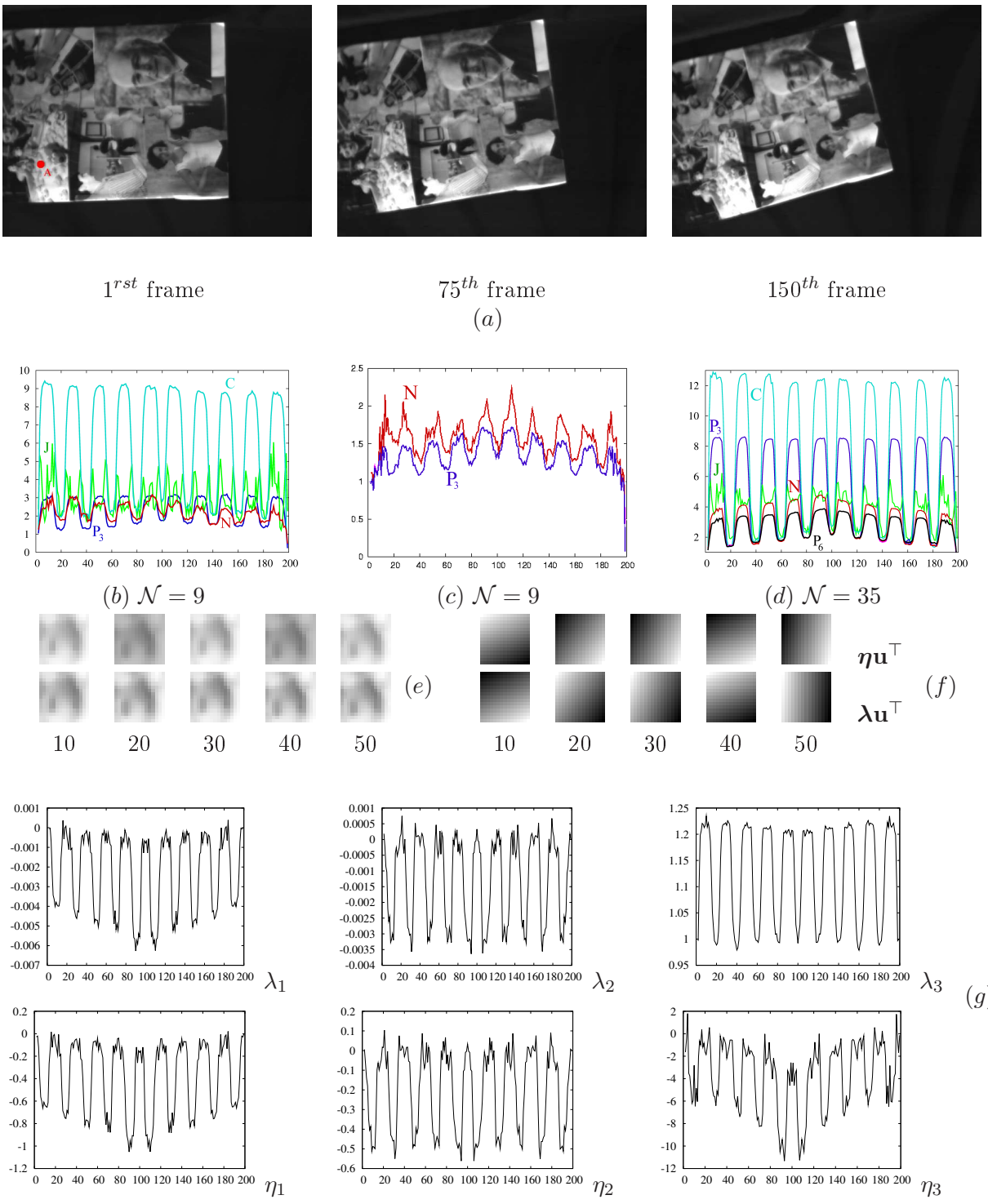


Figure 18: Experiment *Planar object*. (a) Three images of the sequence. (b) Average residues obtained with $\mathcal{N}$=9. (c) Residues obtained with $\mathcal{N}$=9 on the points which are tracked simultaneously by N and P_3. (d) Average residues obtained with $\mathcal{N}$=35. (e) Images of the window of interest centered on A: before (first row) and after (second row) correction by the six parameters of the photometric model P_6. (f) Illumination parameters computed with P_6 on the window of interest. (g) Evolution of the photometric parameters computed with P_6.

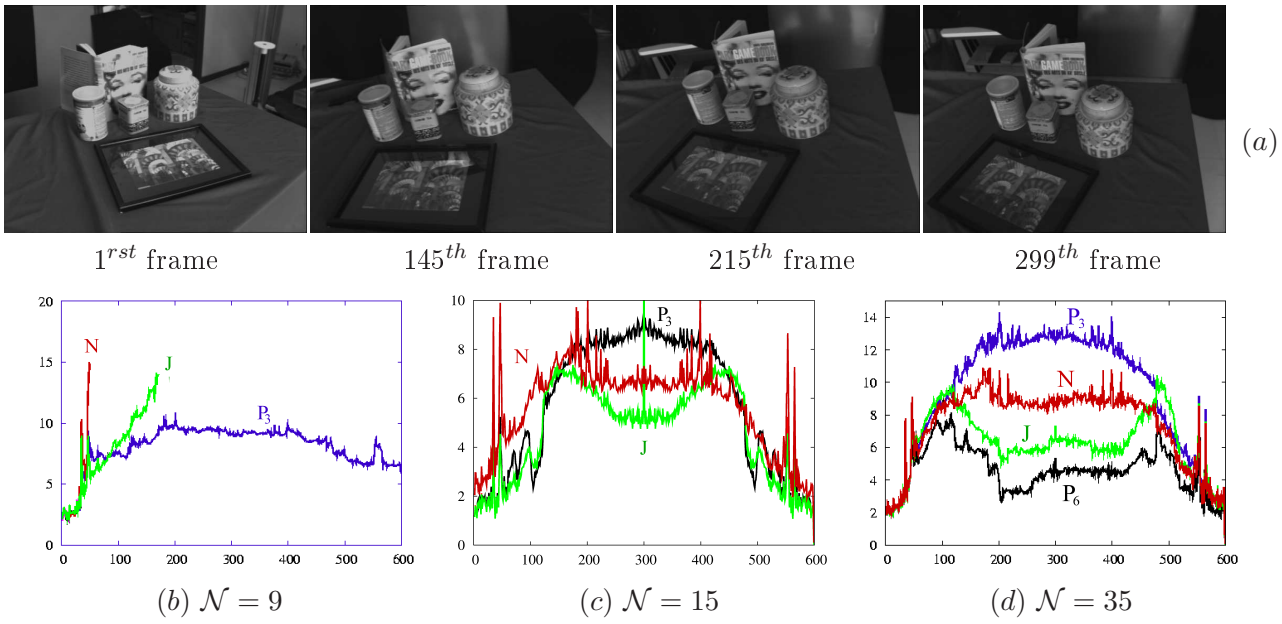


Figure 19: Experiment *Marilyn*. (a) Images of the sequence. (b) Average residues obtained with $\mathcal{N}$=9. (c) Average residues obtained with $\mathcal{N}$=15. (d) Average residues obtained with $\mathcal{N}$=29.

Table 6: Computation times (in ms) used to track one point in the *Planar object* sequence, with $\mathcal{N}$=9, 15 and 35.

Method	$\mathcal{N}$=9	$\mathcal{N}$=15	$\mathcal{N}$=35
C	1.3	2.9	11.4
N	4.3	3.5	14.1
J	1.6	3.2	11.5
P_3	2	3.5	13.8
P_6	32	5.9	18.4

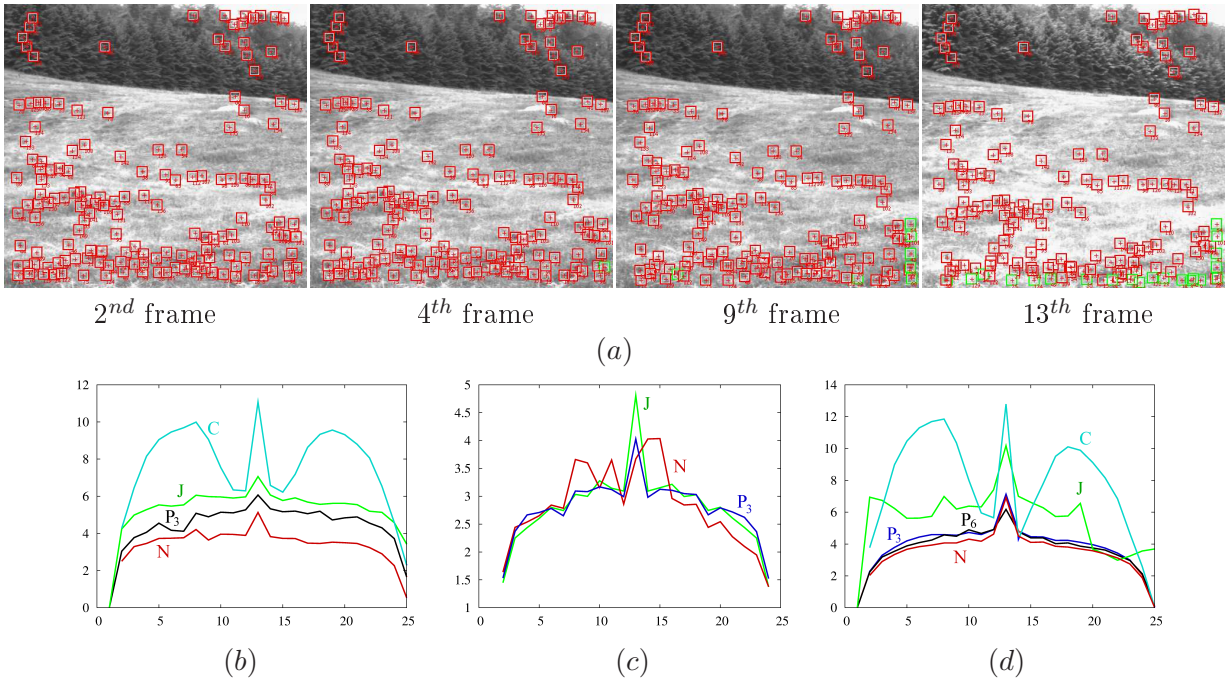


Figure 20: Experiment *Hill*. (a) Images of the sequence. (b) Average residues with $\mathcal{N} = 9$. (c) : residues with $\mathcal{N} = 15$ obtained on the points that have been tracked simultaneously with P_3, N and J. (c) Average residues with $\mathcal{N} = 35$.

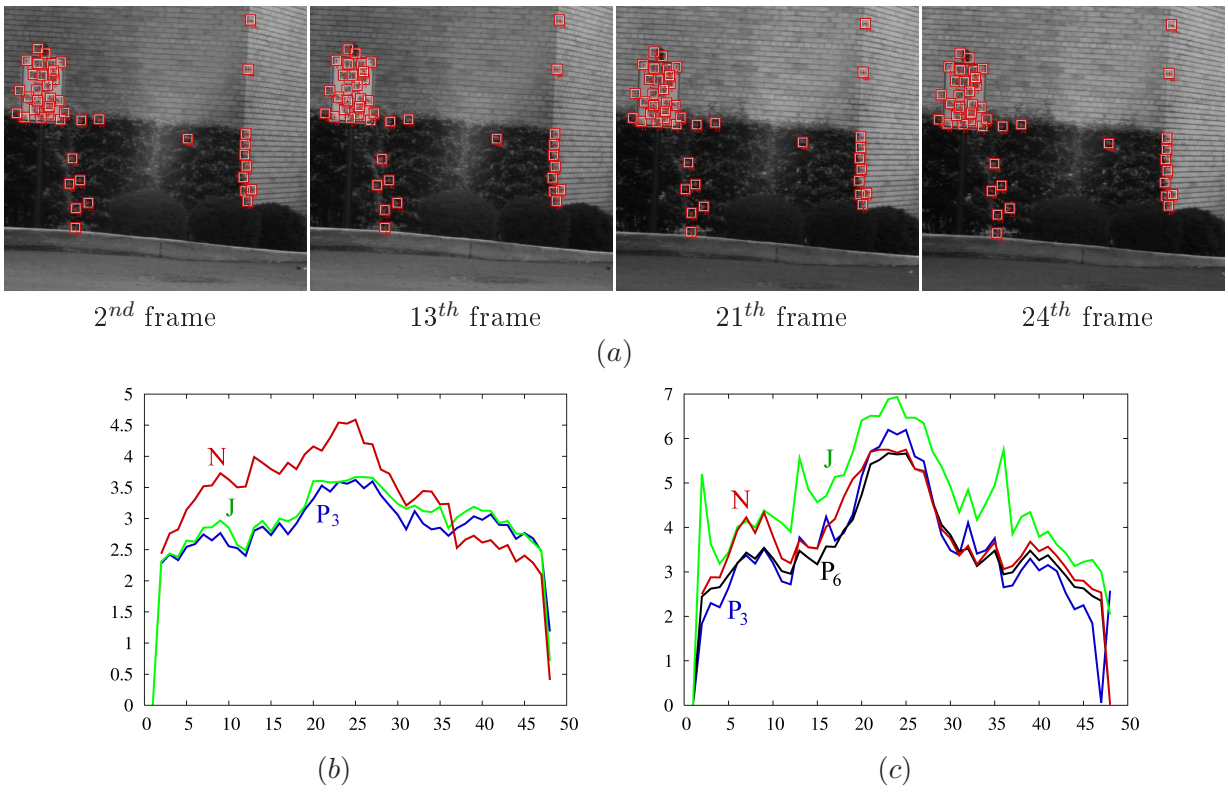
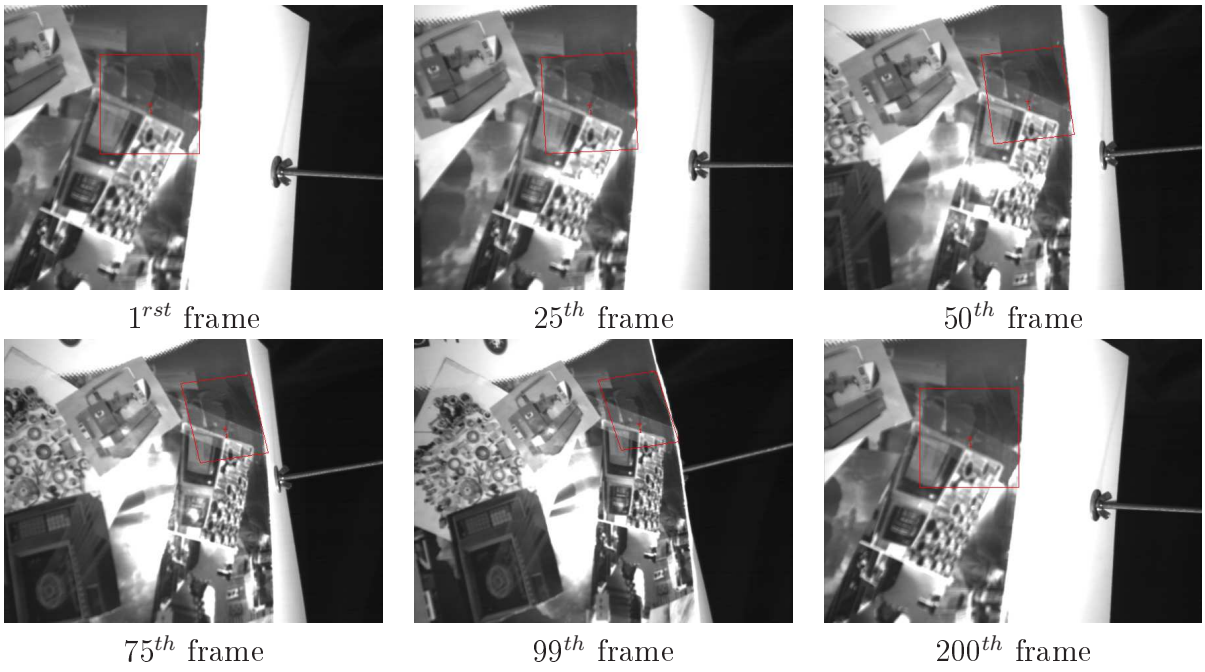
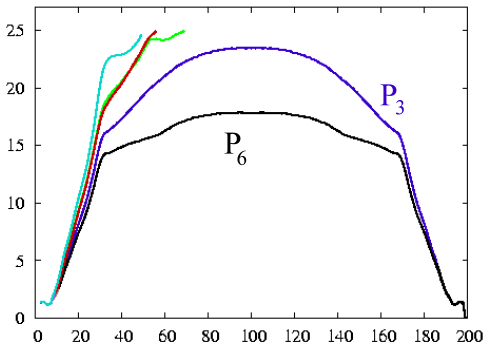


Figure 21: Experiment *Corner*. (a) A few images of the sequence. (b) Average residues for $\mathcal{N} = 9$. (c) Average residues for $\mathcal{N} = 35$.

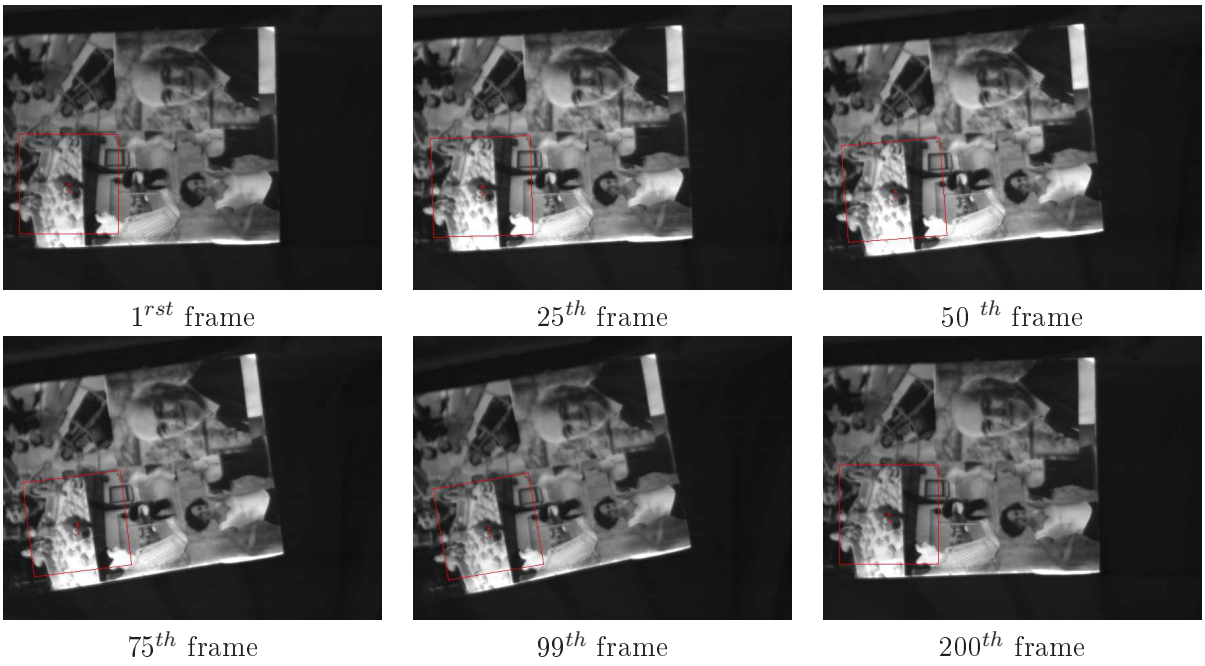


(a)

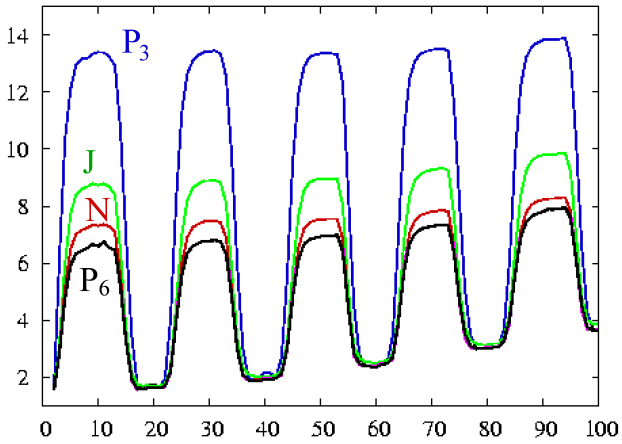


(b)

Figure 22: Tracking of large regions of interest ($\mathcal{N} = 151$), specular highlights occur. (a) Images of the sequence and region tracked with P_6. (b) Evolution of the convergence residues.



(*a*)

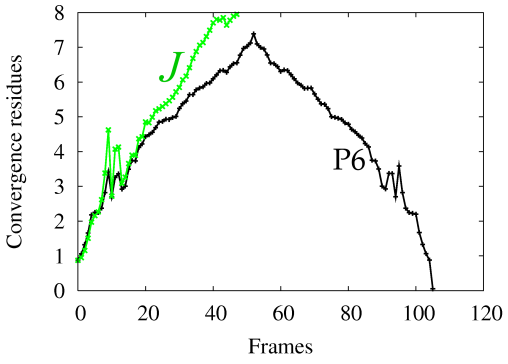


(*b*)

Figure 23: Tracking of large regions of interest ($\mathcal{N} = 151$): lighting and specular highlights changes. (a) Images of the sequence with the region tracked by P_6. (b) Evolution of the convergence residues versus the number of the frame.

1^{rst} frame 11^{th} frame 54^{th} frame

(a)



(b)

Figure 24: *Tracking of a road sign.* (a) Images of the sequence. (b) Convergence residues versus the number of frame.

darkness. Around the 11^{th} frame, a wide motion is caused, because of the vehicle vibrations. The road sign is selected by hand in the first frame with a window size 81×81. The tracking of this road sign has been achieved correctly and the convergence residues of figure 24(b) show that, here again, P_6 compensates more comprehensively for the illumination changes, in comparison to J.

6.4 Discussion

First, the experimental results have shown that the classical tracking technique C is not robust neither to the specular highlights variations nor to ilumination changes since it is based on the assumption of luminance constancy.

In contrast, using an affine photometric model (methods J or N) provides a better robustness, except when the window of interest is small. It can be partly explained by their sensivity to noise. Indeed, when a pixel is noisy in $\mathcal{W}$, the values of μ_f, σ_f, μ_g, σ_g, and λ become also noisy, since they depend on each luminance in $\mathcal{W}$. For the J approach, λ is multiplied by each value of f. Consequently, an error caused on λ can have a huge influence. The minimization of ϵ_2 can finally lead to an incorrect value of $\boldsymbol{\mu}$. On the other hand, for wider windows of interest, the contribution of one noisy pixel in the computation of these parameters becomes

less significant. Consequently, the computation of μ_f, μ_g, σ_f, σ_g, λ is more accurate leading to a more precise value of $\boldsymbol{\mu}$. This remark has been illustrated by the convergence residues obtained by these approaches on small windows of interest.

For small windows of interest, P_3 tracks a larger number of points than N and J. It correctly compensates for the specular highlights and lighting changes on $\mathcal{W}$ and is quite accurate. On the other hand, its performances are reduced when lighting changes have to be modelled, particularly on very large windows of interest. Indeed, in such a context, the modeling has to approximate the *albedo* of the object by a first order polynomial on $\mathcal{W}$. This assumption can be seen as a strong assumption on large windows of interest, where the reflectance may vary drastically. On the other hand, N and J cannot cope correctly with the photometric modeling on non-planar surfaces.

Let us also notice that, from the computation time point of view, even if P_3 requires the computation of an additional parameter with regard to J and consequently the inversion of a wider matrix, the computation times of these techniques are similar, due to a better convergence of P_3.

On the other hand, P_6 is more accurate for large windows, whatever the illumination changes are. Indeed, using a comprehensive photometric model improve the estimation of the motion model during the sequence. In contrast, using it on small windows does not allow the computation of the true photometric and motion parameters.

7 Conclusions

Since the use of specular reflectance models implies the handling of a large number of parameters, most computer vision algorithms assume that the objects in the scene are Lambertian and that no lighting change occurs. However, that is a coarse assumption.

Nevertheless, the use of local simplified photometric models can significantly robustify the processings, by considering the luminance changes occurring between images. Through the analysis of specular reflection models, we have explained explicitly on which assumptions the most widely used photometric models are implicitly based. Then, we propose some new photometric models, which rely on the precise analysis of the reflection, and on the assumption that each kind of illumination change can be approximated by a continuous and derivable function in a local are of the image. The first model, which uses three parameters, is well appropriate to compensate for specular highlights occurrence. The second one uses six parameters and takes each kind of illumination changes into account: specular highlights occurrence, lighting variations or changes of the gain of the camera.

The validity of these photometric models has been theoretically studied, by considering some particular configurations of the scene. First of all, it appears that the photometric models are more appropriate than the affine photometric model and the photometric normalization, since they allow some spatial variations of illumination changes. Our models are quite close to the real illumination changes when surfaces projected in the windows of interest show some low curvature discontinuities, and when the surface is rough enough. Moreover, the photometric

models are more relevant when the sensor is sufficiently close to the surface, and when the lights are sufficiently far from the surfaces.

The two proposed photometric models can be useful in many computer vision applications, where lightning is not perfectly controlled, especially in outdoor experiments.

In this paper, we implemented them in two feature points tracking procedures. The aim of these approaches was to efficiently compensate for the photometric changes caused during an image sequence, in order to obtain a more accurate estimation of the motion model. These procedures have been compared, theoretically and experimentally, to some widely used feature points tracking methods: the classical approach, the tracking with photometric normalization and the tracker proposed by Jin *et al.*, in their simple implementation.

By compensating for the spatial variations of illumination changes, the proposed methods have proved to be more robust than the existing approaches. The first tracker is well adapted for small windows of interest, whereas the second one is applicable for larger windows of interest. Experimental results obtained from several images sequences have shown a good convergence and a good accuracy of these procedures. In this paper, we have deliberately focused on the specific problem of illumination changes, without considering the problem of occlusions, which is an other difficult subject.

A Conditioning

The conditioning of the trackers detailed in that report can be compared. Indeed, each linear equation system envolved in the tracking procedure (see equations (48), (52), (53), (55) and (58)) can be written as $\mathbf{Ax} = \mathbf{y}$ and more precisely

$$\begin{pmatrix} \mathbf{A_{11}} & \mathbf{A_{12}} \\ \mathbf{A_{12}}^\top & \mathbf{A_{22}} \end{pmatrix} \begin{pmatrix} \mathbf{x_1} \\ \mathbf{x_2} \end{pmatrix} = \begin{pmatrix} \mathbf{y_1} \\ \mathbf{y_2} \end{pmatrix} \tag{60}$$

Consequently, the inversion of $\mathbf{A}$ is given by

$$\mathbf{A}^{-1} = \begin{pmatrix} \mathbf{I_3} & \mathbf{0} \\ -\mathbf{A_{22}}^{-1}\mathbf{A_{12}}^\top & \mathbf{I_3} \end{pmatrix} \begin{pmatrix} \boldsymbol{\nabla}^{-1} & 0 \\ \mathbf{0} & \mathbf{A_{22}}^{-1} \end{pmatrix} = \begin{pmatrix} \mathbf{I_3} & -\mathbf{A_{12}}\mathbf{A_{22}}^{-1} \\ \mathbf{0} & \mathbf{I_3} \end{pmatrix} \tag{61}$$

where $\boldsymbol{\nabla}$ is the Schur complement $\boldsymbol{\nabla} = \mathbf{A_{11}} - \mathbf{A_{12}}\mathbf{A_{22}}^{-1}\mathbf{A_{12}}^\top$. The inversion of $\boldsymbol{\nabla}$ can be achieved in the following way:

$$\boldsymbol{\nabla}^{-1} = \mathbf{A_{11}}^{-1} + \mathbf{A_{11}}^{-1}\mathbf{A_{12}}(\mathbf{A_{22}} - \mathbf{A_{12}}^\top\mathbf{A_{11}}^{-1}\mathbf{A_{12}})^{-1}\mathbf{A_{12}}^\top\mathbf{A_{11}}^{-1} \tag{62}$$

Since $(\mathbf{A_{12}}\mathbf{A_{22}}^{-1})^\top = \mathbf{A_{22}}^{-1^\top}\mathbf{A_{12}}^\top$ we can introduce :

$$\mathbf{M} = \begin{pmatrix} \mathbf{I_3} & -\mathbf{A_{12}}\mathbf{A_{22}}^{-1} \\ \mathbf{0} & \mathbf{I_3} \end{pmatrix}$$

so that

$$\mathbf{A}^{-1} = \mathbf{M}^\top \begin{pmatrix} \boldsymbol{\nabla}^{-1} & 0 \\ 0 & \mathbf{A_{22}}^{-1} \end{pmatrix} \mathbf{M}$$

Consequently, the inversion of $\mathbf{A}$ succeeds if $\mathbf{A_{11}}$ and $\mathbf{A_{22}}$ are well-conditioned and can be correctly inverted.

In the tracking techniques, for each approach the matrix $\mathbf{A_{11}}$ is the same. Therefore, the comparison of the conditioning of the method only depends on the conditioning of $\mathbf{A_{22}}$. The matrices associated to the methods which approximate the photometric parameters ${\mathbf{A_{22}}}^J$ (for the Jin's technique), ${\mathbf{A_{22}}}^{P_3}$ and ${\mathbf{A_{22}}}^{P_6}$ are written as:

$$\begin{array}{lrcl} J & {\mathbf{A_{22}}}^J & = & \sum_m (f(m),1)(f(m),1)^\top \\ P_3: & {\mathbf{A_{22}}}^{P_3} & = & \sum_m \mathbf{u}\mathbf{u}^\top \\ P_6: & {\mathbf{A_{22}}}^{P_6} & = & \sum_m (\mathbf{u}f(m),-\mathbf{u})(\mathbf{u}f(m),-\mathbf{u})^\top \end{array} \tag{63}$$

The matrix ${\mathbf{A_{22}}}^{P_3}$ is the best well-conditioned. In addition, its terms are constant, therefore they can be computed off-line. On the contrary, the matrix ${\mathbf{A_{22}}}^J$ and ${\mathbf{A_{22}}}^{P_6}$ are ill-conditioned and their terms depend on the image.

References

[1] E. Arnaud, E. Memin, and B. Cernuschi-Frias. Conditional filters for image sequence based tracking - application to point tracking. *IEEE Transactions on Image Processing*, 14(1), 2005.

[2] S. Baker. Lucas-kanade 20 years on : a unifying framework. *International Journal of Computer Vision*, 56(3):221–255, 2004.

[3] P. Beckmann and A. Spizzichino. *The scattering of electromagnetic waves from rough surfaces*. Artech House Inc, 2 edition, 1987.

[4] S. Benhimane and E. Malis. Homography-based 2d visual tracking and servoing. *The International Journal of Robotics Research*, 26(7):661–676, 2007.

[5] M.J. Black, D.J. Fleet, and Y.Yacoob. Robustly estimating changes in image appearance. *Computer Vision and Image Understanding*, 78:8 – 31, 2000.

[6] S. Bouchafa and B. Zavidovique. Efficient cumulative matching for image registration. *Image and Vision Computing*, 24(1):70–79, 2006.

[7] J.M. Buenaposada and L. Baumela. Real-time tracking and estimation of plane pose. In *IEEE International Conference on Pattern Recognition*, volume 2, pages 697–700, Quebec, Canada, August 2002.

[8] C. Collewet, A. Alhaj, and F. Chaumette. Model-free visual servoing on complex images based on 3d reconstruction. In *IEEE International Conference on Robotics and Automation*, New Orleans, USA, April 26-May 1, 2004.

[9] A.I. Comport, E. Malis, and P. Rives. Accurate quadrifocal tracking for robust 3d visual odometry. *IEEE International Conference on Robotics and Automation*, April 2007.

[10] D.F. Dementhon and L. S. Davis. Model-based object pose in 25 lines of code. *International Journal of Computer Vision*, 15:123–141, 1995.

[11] F.X. Espiau, E. Malis, and P. Rives. Robust features tracking for robotic applications: towards 2d1/2 visual servoing with natural. In *IEEE Int. Conf. on Robotics and Automation, ICRA'2002*, Washington, USA, May 2002.

[12] A. S. Georghiades, P. Belhumeur, and D. J. Kriegman. From few to many: generative models for recognition under variable pose and illumination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(8):643 – 660, June 2001.

[13] M. Gouiffès, C. Collewet, C. Fernandez-Maloigne, and A. Trémeau. Feature points tracking by using a photometric model and color invariants. In *IS&T Third European Conference on Graphics, Image and Vision*, Leeds, UK, June 2006.

[14] P. Gros. Color illumination models for image matching and indexing. In *International Conference on Pattern Recognition*, Barcelona, Spain, September 2000.

[15] G. D. Hager and P. N. Belhumeur. Efficient region tracking with parametric models of geometry and illumination. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20(10):1025–1039, 1998.

[16] C.G. Harris and M. Stephens. A combined corner and edge detector. In *4th Alvey Vision Conference*, pages 147–151, 1988.

[17] H.W. Hausseker and D.J. Fleet. Computing optical flow with physical models of brightness variation. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 760–767, Hilton Head, 2000.

[18] B.K.P. Horn. Robot vision. *The MIT Press*, 1986.

[19] K.P. Horn and B. G. Schunck. Determinig optical flow. *Artificial Intelligence*, 7:185–203, 1981.

[20] F. Jurie and M. Dhome. Hyperplane approximation for template matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):996–1000, 2002.

[21] S-H Lai. Computation of optical flow under non-uniform brightness variations. *Pattern Recognition Letters*, 25(8):885 – 892, June 2004.

[22] J.H. Lambert. *Photometria sive de mensura de gratibus luminis, colorum et ombrae*. Ebrhard Klett, Augsburg, 1760.

[23] B.D. Lucas and T. Kanade. An iterative image registration technique. In *IJCAI'81*, pages 674–679, Vancouver, British Columbia, August 1981.

[24] D. Marimon, Y. Abdeljaoued, B. Palacios, and T. Ebrahimi. Feature point tracking combining the interacting multiple model filter and an efficient assignment algorithm. In *Proceedings of the IS&T/SPIE Electronic Imaging Conference on Visual Communications and Image Processing*, San Jose, California, January 2007.

[25] S. Negahdaripour. Revised definition of optical flow: integration of radiometric and geometric cues for dynamic scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(9):961 – 979, September 1998.

[26] J.M. Odobez and P. Bouthemy. Robust multiresolution estimation of parametric motion models. *Int. Jour. of Visual Communication and Image Representation*, 6(4):348–365, December 1995.

[27] B-T Phong. Illumination for computer generated images. *Communications of the ACM*, 18(6):311–317, June 1975.

[28] M. Pressigout and E. Marchand. Real-time 3d model-based tracking: Combining edge and texture information. In *IEEE Int. Conf on Robotics and Automation, ICRA'06*, Orlando, Florida, May 2006.

[29] J. Shi and C. Tomasi. Good features to track. In *IEEE Int. Conf. on Computer Vision and Pattern Recognition, CVPR'94*, pages 593–600, Seattle, Washington, USA, June 1994.

[30] G. Silveira and E. Malis. Real-time visual tracking under arbitrary illumination changes. *IEEE Computer Vision and Pattern Recognition*, June 2007.

[31] S. Soatto, H. Jin, and P. Favaro. Real-time feature tracking and outlier rejection with changes in illumination. In *IEEE International Conference on Computer Vision*, pages 684–689, Vancouver, Canada, July 9-12, 2001.

[32] P. Tissainayagam and D. Suter. Assessing the performance of corner detectors for point feature tracking applications. *Image and Vision Processing*, 22(6):663–679, 2004.

[33] C. Tomasi and T. Kanade. Detection and tracking of point features. Technical report CMU-CS-91-132, Carnegie Mellon University, April 1991.

[34] T. Tommasini, A. Fusiello, E. Trucco, and V. Roberto. Improving feature tracking with robust statistics. *Pattern Analysis & Applications*, 2:312–320, 1999.

[35] K.E. Torrance and E.M. Sparrow. Theory for off-specular reflection from roughened surfaces. *Journal of the Optical Society of America*, 57(9), September 1967.

[36] D. Ziou and S. Tabbone. Edge detection techniques - an overview. *International Journal of Pattern Recognition and Image Analysis*, 8:537–559, 1998.