HAL open science

Enhancing the point feature tracker by adaptive modelling of the feature support

S. Segvic, A. Remazeilles, François Chaumette

To cite this version:

S. Segvic, A. Remazeilles, François Chaumette. Enhancing the point feature tracker by adaptive modelling of the feature support. European Conf. on Computer Vision, ECCV'2006, 2006, Graz, France. pp.112-124. inria-00350325

HAL Id: inria-00350325

https://inria.hal.science/inria-00350325v1

Submitted on 6 Jan 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Enhancing the Point Feature Tracker by Adaptive Modelling of the Feature Support*

Siniša Šegvić[1], Anthony Remazeilles[2], and François Chaumette[1]

[1] IRISA/INRIA Rennes, Campus de Beaulieu, 35042 Rennes cedex, France
[2] IRISA/INSA Rennes, Campus de Beaulieu, 35042 Rennes cedex, France
{sinisa.segvic, anthony.remazeilles, francois.chaumette}@irisa.fr

Abstract. We consider the problem of tracking a given set of point features over large sequences of image frames. A classic procedure for monitoring the tracking quality consists in requiring that the current features nicely warp towards their reference appearances. The procedure recommends focusing on features projected from planar 3D patches (planar features), by enforcing a conservative threshold on the residual of the difference between the warped current feature and the reference. However, in some important contexts, there are many features for which the planarity assumption is only partially satisfied, while the true planar features are not so abundant. This is especially true when the motion of the camera is mainly translational and parallel to the optical axis (such as when driving a car along straight sections of the road), which induces a permanent increase of the apparent feature size. Tracking features containing occluding boundaries then becomes an interesting goal, for which we propose a multi-scale monitoring solution striving to maximize the lifetime of the feature, while also detecting the tracking failures. The devised technique infers the parts of the reference which are not projected from the same 3D surface as the patch which has been consistently tracked until the present moment. The experiments on real sequences taken from cars driving through urban environments show that the technique is effective in increasing the average feature lifetimes, especially in sequences with occlusions and large photometric variations.

1 Introduction

Tracking point features in a sequence of image frames is an important low-level problem of early computer vision. The quality of the recovered trajectories directly affects the performance of attractive higher level tasks such as structure from motion [1], visual odometry [2], concurrent mapping and localization [3], and visual servoing [4]. However, the priorities of the desired tracking behaviour may differ between the particular contexts, since the former two involve larger numbers of "nameless" features, while the latter ones usually focus on fewer but more important landmarks. Thus, achieving the longest possible contact with

* The presented work has been performed within the french national project Predit Mobivip, and within the project Robea Bodega. The authors would also like to acknowledge Prof. Axel Pinz for a helpful discussion on this subject.

A. Leonardis, H. Bischof, and A. Pinz (Eds.): ECCV 2006, Part II, LNCS 3952, pp. 112–124, 2006.
© Springer-Verlag Berlin Heidelberg 2006

each of the tracked features, being the focus of this paper, is highly desired in the latter tasks, even though the former ones can operate with considerably shorter feature lifetimes. The two main approaches for conceiving a point feature tracker are iterative first-order differential approximation [5, 6], and exhaustive matching [2, 7]. In both approaches, a straightforward implementation based on integrating inter-frame motion is a viable solution only for short-term operation, due to the incontrollable growth of the accumulated drift. It is therefore necessary either to adapt the higher-level task to work only with short feature tracks [2], if applicable, or to devise a monitoring approach which would try to correct the drift by aligning the current appearance of the feature with a previously stored template image or *reference*. The desired alignment is usually performed by minimizing the norm of the *error image*, which is obtained by subtracting the current feature from the reference [8]. Shi and Tomasi [5] have addressed the monitoring over linear deformations of the planar surface, which have been described with a 2D affine transform, under reasonable assumptions of the feature position with respect to the camera. An extension of their work has been proposed by Jin et al. [6] who devised a scheme which additionally compensated for affine photometric deformations of the grey level value in the image.

An important issue in monitored long-term tracking is being able to recognize when a match with the reference can not be confidently established any more, so that the tracking of the feature can be discontinued in order to prevent errors at the higher levels. Previously, this has been accomplished by using criteria based on the RMS (root-mean-square) residual of the error image [5], and normalized cross-correlation score combined with the ratio between the two areas [6]. However, the richer deformation models pose a bigger danger of allowing a warp producing an incorrect match with a low residual [9]. This danger can be mitigated by enlarging the size of a feature window: larger windows provide a better security that a good match score is not due to a chance. On the other hand, large features are more likely to include a 3D surface discontinuity, which usually makes a correct warp towards past appearances impossible. The odds for straddling a discontinuity are especially high if we consider the tracking of features that are initially distant. For a usual horizontal field of view of 30° and a resolution of 320×160 pixels, a 15×15 pixels region corresponds to a perpendicular planar patch of over $1 \times 1\,\mathrm{m}$ at a distance of $50\,\mathrm{m}$. In such a scenario, characteristic for an observer situated in a car moving along a straight road, there may indeed be too few planar features for the needs of a higher task.

A technique is proposed for alleviating the problems with features which are only partly projected from a distinctive quasi-planar 3D surface, while keeping the good behaviour for the true planar features. The well behaved portion of a feature window is termed as *feature support*, while its robust and adaptive detection is the main objective of the paper. The technique is related to robust estimation of the warp parameters [10, 11], but is more suitable for detecting correct feature supports which often contain statistical outliers. Here we do not consider *updating* the reference [12, 13, 11] despite its potential for increasing the tracking flexibility, since it offers less precision while requiring more processing

power. The related research also includes the cumulative similarity transform [14] which is suitable only for tracking homogeneous regions, and the probabilistic filtering of the feature position [15, 13, 11], which has been used for handling temporary total occlusions.

The paper is organized as follows: the theoretical background is briefly summarized in Sect. 2. Sect. 3 describes the two complementary procedures to infer the feature support. Experimental results are described and discussed in Sect. 4, while Sect. 5 contains a short conclusion and the directions for future work.

2 Theoretical Background

2.1 General Differential Tracker with Warp Correction

Let the feature in the current frame is given by $I(\mathbf{x})$, its appearance after a warp with parameters $\mathbf{p}$ by $I_W(\mathbf{x}, \mathbf{p})$, and the corresponding reference by $I_R(\mathbf{x})$. Then the tracking consists in finding $\hat{\mathbf{p}}$ which minimizes the error image norm, or equivalently, the error over the feature window:

$$\hat{\mathbf{p}} = \arg\min_{\mathbf{p}} \sum_{\mathbf{x}} \| I_W(\mathbf{x}, \mathbf{p}) - I_R(\mathbf{x}) \| \ . \tag{1}$$

The minimization is performed in a Gauss-Newton style, by employing a first-order Taylor expansion of the warped feature around the previous approximation of $\hat{\mathbf{p}}$. This can be expressed in different ways [8], and here we present a "forward-additive" formulation with which the best accuracy has been obtained. The current feature warped with a sum of the previous parameter vector $\mathbf{p}$ and an unknown additive improvement $\Delta\mathbf{p}$ is therefore approximated as:

$$I_W(\mathbf{x}, \mathbf{p} + \Delta\mathbf{p}) \approx I_W(\mathbf{x}, \mathbf{p}) + \frac{\partial I_W}{\partial \mathbf{p}} \cdot \Delta\mathbf{p} \ . \tag{2}$$

The scalar residual norm appearing in (1) can now be represented as:

$$\begin{aligned} R(\Delta\mathbf{p}) &= \sum_{\mathbf{x}} \| I_W(\mathbf{x}, \mathbf{p} + \Delta\mathbf{p}) - I_R(\mathbf{x}) \| \\ &\approx \sum_{\mathbf{x}} \| I_W(\mathbf{x}, \mathbf{p}) + \frac{\partial I_W}{\partial \mathbf{p}} \cdot \Delta\mathbf{p} - I_R(\mathbf{x}) \| \ . \end{aligned} \tag{3}$$

For clarity, we omit the arguments, denote the previous error image as e, and introduce $\mathbf{g}$ as the transposed warped feature gradient over the warp parameters:

$$R(\Delta\mathbf{p}) \approx \sum_{\mathbf{x}} \| e + \mathbf{g}^\top \Delta\mathbf{p} \| \ . \tag{4}$$

The requirement (1) can be enforced by finding a $\Delta\hat{\mathbf{p}}$ for which the gradient of the residual vanishes. In case of the L2 norm, this is easy to perform:

$$\frac{\partial R(\Delta\hat{\mathbf{p}})}{\partial \Delta\hat{\mathbf{p}}} \approx \sum_{\mathbf{x}} 2 \cdot (e + \mathbf{g}^\top \Delta\hat{\mathbf{p}}) \cdot \mathbf{g}^\top = \mathbf{0}^\top \ . \tag{5}$$

After transposing both ends of (5), we arrive at the final expression for an iteration in the context of a general warp (note that e is a scalar function):

$$\sum_{\mathbf{x}} (\mathbf{g}e + \mathbf{g}\mathbf{g}^\top \Delta\hat{\mathbf{p}}) = \mathbf{0} \ . \tag{6}$$

Thus, in each iteration, the additive improvement is calculated by solving a linear system of equations. The procedure stops when the norm of the improvement $\|\Delta\hat{\mathbf{p}}\|$ falls below a threshold, or when the new feature position falls outside the image bounds, or when the determinant $|\mathbf{g}\mathbf{g}^\top|$ becomes too small.

2.2 Tracker with Isotropic Scaling and Contrast Compensation

In order to mitigate the danger that a physically unrelated image patch might be well transformed towards the reference, a trade-off between the modelling power and the tracking security should be carefully chosen. For our application, a good balance is obtained by a 5-dimensional warp consisting of a 2-dimensional translational offset ($\mathbf{d}$), an isotropic scaling parameter (m), and the two parameters of the affine contrast compensation model (λ, δ) [6]. It is convenient to express the warp in terms of geometric and photometric components as $\mathbf{p} = (\mathbf{q}, \mathbf{r})$, where $\mathbf{q} = (m, \mathbf{d})$, and $\mathbf{r} = (\lambda, \delta)$. The warped feature is then obtained as:

$$I_W(\mathbf{x}, \mathbf{p}) = \lambda \cdot I(m * \mathbf{x} + \mathbf{d}) + \delta = U(I(T(\mathbf{x}, \mathbf{q})), \mathbf{r}) \ . \tag{7}$$

In order to use the general formulation from 2.1, an expression for $\frac{\partial I_W}{\partial \mathbf{p}} = [\frac{\partial U}{\partial \mathbf{q}} \frac{\partial U}{\partial \mathbf{r}}]$ must be derived using the chain rule. The second term is simpler to obtain:

$$\frac{\partial U}{\partial \mathbf{r}}(I(T(\mathbf{x}, \mathbf{q})), \mathbf{r}) = \begin{bmatrix} I_T & 1 \end{bmatrix} \ , \tag{8}$$

where I_T is the current feature warped with T: $I_T = I(T(\mathbf{x}, \mathbf{q})))$. The derivation of the first term is a little bit more involved:

$$\begin{aligned} \frac{\partial U}{\partial \mathbf{q}}(I(T(\mathbf{x}, \mathbf{q})), \mathbf{r}) &= \frac{\partial U}{\partial I}(I(T(\mathbf{x}, \mathbf{q})), \mathbf{r}) \cdot \frac{\partial I}{\partial T}(T(\mathbf{x}, \mathbf{q})) \cdot \frac{\partial T}{\partial \mathbf{q}}(\mathbf{x}, \mathbf{q}) \\ &= \lambda \cdot I_T^x \cdot \begin{bmatrix} x_1 & 1 & 0 \\ x_2 & 0 & 1 \end{bmatrix} = \lambda \begin{bmatrix} I_T^x \mathbf{x} & I_T^{x1} & I_T^{x2} \end{bmatrix} \ , \end{aligned} \tag{9}$$

where I_T^x is the gradient in the feature warped by T: $I_T^x = \frac{\partial I}{\partial T}(T(\mathbf{x}, \mathbf{q})))$. The combined result, (9) and (8), can be plugged into (6), with $\mathbf{g}$ given by:

$$\mathbf{g}^\top = \begin{bmatrix} I_T^x \mathbf{x} & I_T^{x1} & I_T^{x2} & I_T & 1 \end{bmatrix} \ . \tag{10}$$

2.3 The Running Average Gaussian Estimation

The proposed tracking approach relies on estimating the gray scale value distribution for each single pixel within the feature window. This can be achieved by

a space-efficient approximation of the running average, which has been extensively used in the field of the background subtraction. For each feature pixel x, the current estimate of a distinct normal distribution is updated as follows [16]:

$$\begin{aligned} \mu_{x,t} &= (1-\alpha)\cdot\mu_{x,t-1} + \alpha\cdot x_t \\ \sigma^2_{x,t} &= (1-\alpha)\cdot\sigma^2_{x,t-1} + \alpha\cdot(x_t-\mu_{x,t})^2 \ . \end{aligned} \tag{11}$$

The parameter $\alpha \in \langle 0, 1\rangle$ represents the learning rate, or alternatively, how many previous frames are taken into account for the estimate. Although there are no guarantees that a certain pixel is normally distributed (indeed, the pixels which are interesting in our context may have arbitrary distributions, depending on the scene), the estimates do offer an insight into the pixel mean and variability.

3 The Feature Support Concept

3.1 Assumptions and Basic Notions

The high level application context assumes robot navigation in urban environment, controlled by techniques in which a long term contact with the features from a given set is highly desired. The considerations are therefore focused on tracking over a significant forward motion, as illustrated in Fig. 1. The features which are visible throughout the whole sequence are located quite far from the initial observer location, so that they experience considerable changes of scale and photometry. The 3D surfaces projecting into initial feature windows are quite large (due to the distance), so that many features cross a discontinuity. In fact, since parts of the scenery behind the car (to the left from `#28` and to the right from `#39`) were out of the field of view in some frames of the sequence, `#20` is the only feature in Fig. 1(a), for which the final appearance does not substantially deviate from the affine transformation model. The proposed concept strives to enlarge the application field of a differential tracker with warp correction onto the features for which the initial windows are only partly projected from a plane. The resulting convergence of the feature support provides a valuable shape information allowing the non-rectangular features to be introduced in

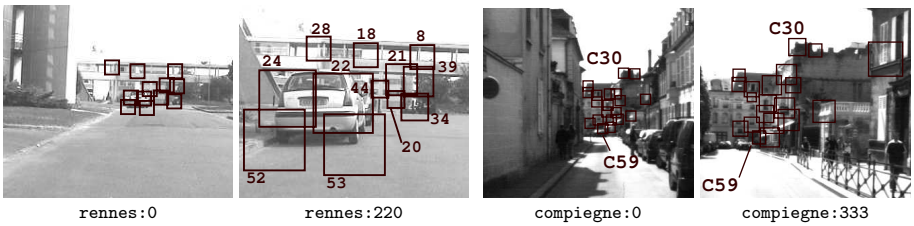


Fig. 1. Illustration of the tracking task: central portions of the first and the last frames of the sequences `rennes` and `compiegne`, with the designated windows of the tracked features

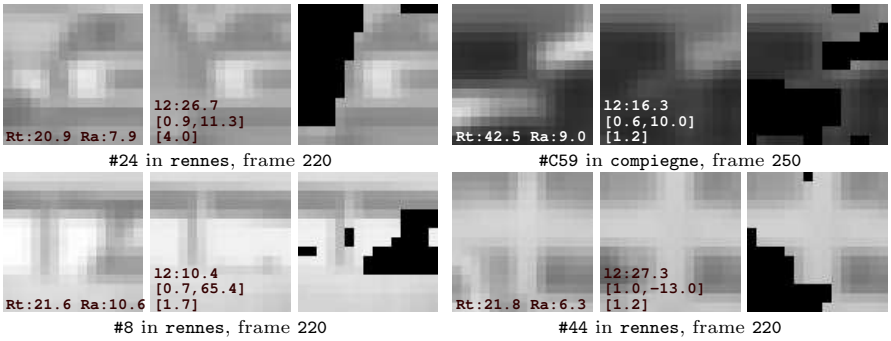


Fig. 2. The enlarged triples of the reference, the warped current feature and the feature support (non-masked areas) for the four features from Fig. 1. The numbers in the reference images indicate the RMS residuals for the whole feature window (`Rt`), and for the feature support only (`Ra`). The numbers in the warped features indicate the smaller eigenvalue of the second-order moment matrix (l_2), the photometric warp (λ, δ) and the isotropic scaling (m).

(6), and in the calculation of the monitoring residual. To illustrate the proposed objectives, the obtained supports for several features from sequences `rennes` and `compiegne` are shown in Fig. 2. The need for feature support arises most often when the feature is on a foreground structure occluding the background, either because the feature is at the boundary (`#24`, `#C59`), or the structure has holes (`#8`). The concept can also be helpful if the feature is situated on a background structure which is at times occluded by the foreground (`#44`), and when there are complex surface radiance variations which can not be counterbalanced by a feature-wide contrast compensation model (`#C59`). The relation between the obtained residuals (`Ra`$\ll$`Rt`, see Fig. 2) illustrates the effectiveness of the technique.

3.2 The Volatile Feature Support Due to a Robust Rejection Rule

In the first investigated approach, the pixels not belonging to the feature support are identified as outliers within the distribution of the squared grey level value within the current error image $\{e_i^2\}$. The outliers are detected following a robust X84 rejection rule, which has also been used to reject the entire features (not the individual pixels), based on the magnitude of their RMS residual [17]. The rule uses the median as an estimator for the distribution location, while the scale of the distribution is estimated by the median absolute deviation (MAD):

$$\begin{aligned} \mathrm{C}_{e^2} &= \mathrm{med}\{e_i^2\} \\ \mathrm{MAD}_{e^2} &= \mathrm{med}\{|e_i^2 - \mathrm{C}_{e^2}|\} \ . \end{aligned} \tag{12}$$

Due to a further noise suppression, much better results are obtained when *temporally smoothed* values are used within (12). This can be achieved by substituting

the error image pixels e_i with a difference between the mean value of the warped feature pixel estimated by (11), and the corresponding reference pixel.

The pixels of the feature support can finally be identified by testing for:

$$(|e_i^2 - \mathrm{C}_{e^2}|) < \max(\mathrm{th}_{\min}, k \cdot \mathrm{MAD}_{e^2}) \ . \tag{13}$$

The choice of $k = 5$ is often appropriate here, since $5 \cdot \mathrm{MAD}$ corresponds to $3.5 \cdot \sigma$ in a Gaussian distribution. Experiments have shown that the threshold $\mathrm{th}_{\min}$ is required for suppressing the bad behaviour when there are no real outliers.

3.3 The Persistent Feature Support Due to Temporal Consistency

Experiments have shown that the previous approach for inferring the feature support is surprisingly effective in increasing the tolerance to the occasional outliers. However, that approach assumes that all the inlier error image pixels come from the same distribution, which is rarely the case. Good features usually have pixels originating from different materials which are likely to generate different error distributions. Thus, the obtained instances of the feature support usually do not resemble the part of the window projected from a continuous surface.

The second approach makes a more explicit check for the temporally consistent feature pixels, by analyzing the standard deviation estimated by (11). During the motion of the observer, the pixels belonging to a different continuous surface than the one which is consistently tracked, will refer to different points of the scene. In the case of natural scenes which are rich in texture, this will be reflected by occasional spikes in the standard deviation. These spikes can be detected by a threshold on the standard deviation σ^{th}, while the corresponding pixels can be persistently excluded from the feature support. An inverse process (adding a pixel to the feature support if it is consistently similar to the reference) could be employed for recovering after temporary occlusions. This has not been performed in our experiments, since for the most frequent foreground features it implies relinquishing the valuable information about the feature shape, which is not always attainable (e.g. when the background is homogenous).

A critical notion in both approaches is controlling the learning rate of the Gaussian estimates in (11). A fixed value would not be acceptable, since it would imply obtaining different results for different dynamics of the same motion. Perhaps the best solution would be to modulate α_0 by a perceived translational displacement with respect to the structure on which the feature resides. However, this would bring a serious increase of the implementation complexity, due to the coupling of the tracker with pose estimation. A simpler solution is therefore proposed, in which the modulating factor is computed from the interframe change of geometric warp parameters $\mathbf{d}$ and m:

$$\alpha = \alpha_0 \cdot \rho(|\Delta m| \cdot w_x + |\Delta d_x|, |\Delta m| \cdot w_y + |\Delta d_y|) \ , \tag{14}$$

where ρ is a 2D metric, and (w_x, w_y) are the feature window dimensions. If the camera motion is strictly translational and the feature occludes the background at infinity, the proposed solution gives each background fraction a fair amount

in the distribution of a feature pixel. The behaviour would be less satisfactory for a chiefly rotational motion and for occlusions of distant features, but these cases do not occur in many realistic situations, as confirmed by experiments.

3.4 Multiscale Considerations

Due to the expected increase in the feature scale, it is suitable to initialize the tracking by the features at the smallest feasible scale. In order to ensure a good behaviour for large features (e.g. the feature `#24` in Fig. 2 is more than 4 times larger than the reference), the tracking is performed at the level of the image pyramid which most closely resembles the previous scale of the feature. This is achieved by a simple scaling of the parameters of the geometrical warp before and after the tracking procedure for each individual feature. However, due to discretization issues, this sometimes causes artificial spikes in the parameters of the pixel Gaussians. The mean estimates for the feature pixels are therefore reinitialized to the corresponding actual values at each change of the pyramid level, in order to avoid the degradation of the feature support.

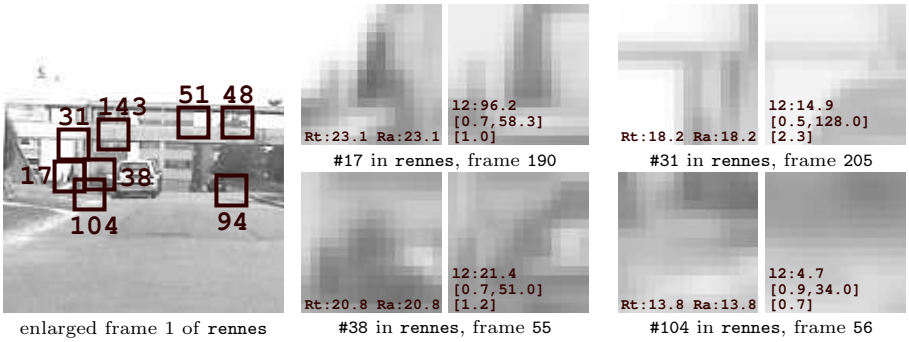
4 Experimental Results

The performed experiments were directed towards three different goals. The first goal was to investigate whether a threshold on the feature RMS residual can be at least partially substituted by other, hopefully more discriminative indicators of bad tracking. The second goal was to obtain a qualitative insight into the benefits of the proposed technique, by analyzing its sensitivity to the change of feature monitoring parameters. The final goal was an objective assessment of the influence of the technique to the measured lifetime of the tracked features.

The provided experimental results were obtained exclusively by the persistent support described in 3.3. The volatile approach described in 3.2 was not evaluated due to the ad-hoc threshold in (13), which undermines the capability to find a right ballance between the feature longevity and the tracking security. The recovered support is used for restricting the area of the feature window both in the tracking equations (6), as well as in the sum for calculating the error image norm (1). In order to be able to deal with large scale changes, a 3-level image pyramid is employed, obtained by successive smoothing and 1:2 subsampling. The switch of the pyramid level occurs whenever the feature window at the current resolution becomes greater than 1.8 times the size of the reference. The initial feature windows are 15×15 pixels wide, while the feature support modelling parameters are: $\alpha_0 = 0.005$, $\sigma^{\mathrm{th}} = 12$. The source code used for performing the experiments is based on a public implementation of the KLT feature tracker [5] (see `http://www.ces.clemson.edu/~stb/klt/`) .

4.1 Criteria for Evaluating the Warp Correction Quality

Knowing when to abandon the tracking is a very important quality of a point tracker. In the previous work [5, 6], this was achieved chiefly by relying on the



enlarged frame 1 of rennes | #17 in rennes, frame 190 | #31 in rennes, frame 205 | #38 in rennes, frame 55 | #104 in rennes, frame 56

Fig. 3. The position of some features from rennes which will be discussed in the further text (left), and the four problematic ones (right). The abrupt magnification change test detects #17 and #31, but not #38. The gradient test detects "dissolved" features such as #104. See Fig. 2 for annotations.

RMS residual. However, the discriminative power of that criterion in real scenes with complex photometric variations leaves to desire, since for a given threshold, there are often both correctly rejected and incorrectly tracked features. For illustration, similar non-masked residuals (Rt) are obtained for the good features in Fig. 2, and for the problematic ones in Fig. 3 (#17, #31, #38, #104). The two most difficult situations for a point tracker are (i) when a foreground structure occludes the feature, which then tends to "jump" onto the foreground, and (ii) when the feature is on a face which is nearly parallel to the motion, when the warp may approach singularity. In both cases, the tracker may diverge from a consistent local minimum, but fortunately, this often can be detected by observing some common divergence symptoms. The latter scenario can be detected by testing for a "blanc wall" condition within the warped feature, by setting a threshold on the smaller eigenvalue of the second-order moment matrix [9]. Naturally, in the proposed context, the test is only performed for the pixels of the feature support. This test is very effective in avoiding tracking errors in low gradient areas, where a bad match often produces a small residual (see #104 in Fig. 3). Despite the efficacy in pruning the bad features, the test is a candidate for refinement because some features can be well tracked in spite of the low gradient (#48, #51 and #94 in Fig. 3).

Although the foreground structure and the background feature may be quite similar, as for features #17 (the car occludes the fence), #38 (the car occludes the bush), and #31 (the car occludes the building) in Fig. 3, the transfer is seldom smooth. This can be detected by an abrupt change of the recovered warp parameters. In particular, a threshold of 10% on the interframe relative magnification change[1] detects many of such situations, while seldom reporting false alarms.

[1] This is also true for other affine degrees of freedom: anisotropic scaling, skew, rotation. These parameters are not allowed since they actually decrease the tracking quality, by providing a way for the tracker to "escape" towards wrong local minima.

Nevertheless, the transfer of the feature #38 (see Fig. 3) involves only a 6% interframe relative magnification change. The proposed technique deals successfully with this situation since the feature support decreases with the occlusion, and the tracking is abandoned when a threshold of 40% is reached. However, as explained in 3.3, this would not work for a very distant feature, since the modulation factor for α would have been zero. Thus, unfortunately, the residuum threshold can not be completely avoided in the current implementation.

4.2 Sensitivity to Threshold Parameters

The choice of the threshold parameters used to detect the bad tracking is a trade-off between the security and the multiplicity of the tracked features. For the case of the RMS residual threshold, this is illustrated in Table 1. The results

Table 1. Count of features tracked until the end of `rennes`, for different thresholds on RMS residual r. For the discussion on feature #38, see 4.1 and Fig. 3.

	$r = 10$	$r = 15$	$r = 20$	$r = 25$
without feature support	1	3	8	12+#38
with feature support	3	11	13	13

suggest that the feature support offers better tracking results, even with a stricter residuum threshold. For example, the basic tracker with $r = 25$ produces a 18% magnification error for #8, while #18 is discontinued due to the abrupt magnification change. Both features are well tracked using the proposed technique, while the development of their supports is shown in Fig. 4.

Similar considerations hold for the threshold on the condition of the second-order moment matrix. If this threshold is released, two more features survive to the last frame in the basic tracker (#94, #143), only one of which is well

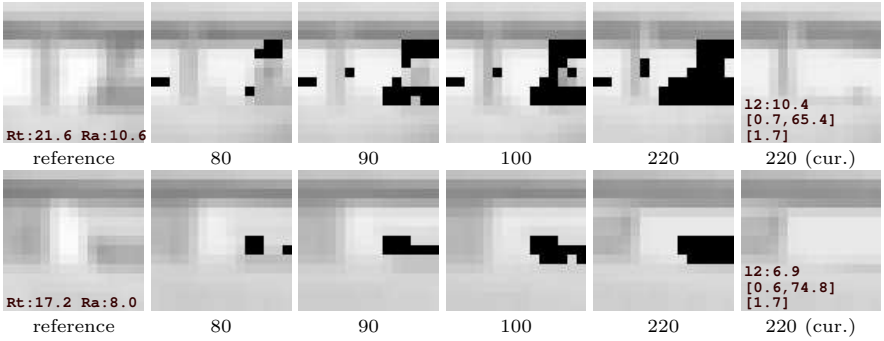


Fig. 4. The development of the support for the two features #8 (up) and #18 (down) from `rennes`, which are not correctly tracked when the feature support is not used. See Fig. 2 for annotations.

tracked. However, when feature support is used, additional two features are tracked without errors (#48, #51, all in Fig. 3).

4.3 Quantitative Experiments

The effects of the proposed technique are quantitatively evaluated on several real sequences taken from cars moving in urban environments. In the experiments, we test whether the proposed technique can provide longer feature lifetimes even with a more restrictive residuum threshold. We consider eight sequences which are briefly described in Table 2, while a more detailed presentation is available at http://www.irisa.fr/lagadic/demo-cv-tracking-eng.html. Each sequence from the table has been acquired during about one minute of mainly translational movement of the vehicle on which the camera was mounted. For each sequence, the tracking procedure was invoked with and without the feature support enabled, for different combinations of the RMS residual threshold. The relation between the two sets of obtained lifetimes (l^i_{FS}) and (l^i_{noFS}), has been analysed exclusively for features in which the tracking was discontinued due to the one of the criteria described in 4.1. In particular, we do not consider the features discontinued after a contact with the image border, which introduces a bias towards shorter-living features. Two different measures of average feature lifetime were used:

1. geometric average of individual lifetime ratios: $M_g = \sqrt[n]{\prod_i l^i_{FS}/l^i_{noFS}}$
2. ratio of the total feature lifetime: $M_a = (\sum_i l^i_{FS})/(\sum_i l^i_{noFS})$

The latter measure is judged as better since it reduces the bias towards short-living features. The obtained results are summarized in Table 2, and they show that the proposed technique favourably influences the feature lifetimes. Besides the occlusions and large photometric variations, the technique also allows to deal with structural changes, affecting the roof silhouettes (see #C30 in Fig. 1), and moderate affine deformations occurring on the pavement signalization. Conversely, the results for compiegne2 and compiegne3 suggest that there is no negative impact if the addressed effects are absent.

Table 2. Quantitative comparison of the total feature lifetime ratio M_a, for different combinations of RMS thresholds $R_{FS} : R_{noFS}$

sequence	description	15:15	20:20	15:20
rennes	approaching a building with holes	1.32	1.14	1.06
rennes2	a tour in the inner court	1.23	1.11	1.01
compiegne	towards a square in the sunlight	1.20	1.23	1.03
compiegne2	traversing a large square into a street	1.09	1.10	0.96
compiegne3	a very narrow street	1.05	1.07	0.93
compiegne4	a street bordered by buildings and trees	1.17	1.18	0.98
antibes1	some trees on the left and far away	1.09	1.13	0.99
antibes2	a narrow downhill winding street	1.07	1.07	1.02

5 Conclusions and the Future Work

A technique for increasing the feature lifetimes in extended real sequences acquired during a mainly translational forward motion of the observer has been presented. The technique addresses "almost good" features, for which the deformations during the tracking can not be completely explained by linear transforms, due to occlusions, photometric variations or small structural developments. The experiments suggest that the technique favourably affects the tracking quality, on both accounts of the correct tracking and the correct rejection.

The future work will be concentrated on applying the technique in the field of the autonomous robot navigation. There we would like to explore the potential of using all geometric warp parameters recovered by the tracking procedure ($\mathbf{d}$,m). Further improvements might be obtained by devising more sophisticated ways to regulate the modulation speed α for estimating the distribution parameters of the warped feature pixels. An eventual faster convergence would allow the monitoring procedure to rely more heavily on the size and the shape of the feature support, and consequently further improve the chances for early detection of ill-conditioned situations, and confident tracking during extended time intervals.

References

1. Faugeras, O.D., Maybank, S.: Motion from point matches: multiplicity of solutions. Int. J. Comput. Vis. **4** (1990) 225–246
2. Nistér, D., Naroditsky, O., Bergen, J.: Visual odometry. In: Proc. of CVPR, Washington, DC, USA, IEEE (2004) 652–659
3. Davison, A.: Real-time simultaneous localisation and mapping with a single camera. In: Proc. of ICCV, Nice, France (2003) 1403–1410
4. Malis, E., Chaumette, F., Boudet, S.: 2 1/2 D visual servoing. IEEE Trans. RA **15** (1999) 234–246
5. Shi, J., Tomasi, C.: Good features to track. In: Proc. of CVPR. (1994) 593–600
6. Jin, H., Favaro, P., Soatto, S.: Real-time feature tracking and outlier rejection with changes in illumination. In: Proc. of ICCV. Volume 1. (2001) 684–689
7. Ma, Y., Soatto, S., Košecká, J., Sastry, S.: An Invitation to 3-D Vision: From Images to Geometric Models. Springer-Verlag, New York, USA (2004)
8. Baker, S., Matthews, I.: Lucas-Kanade 20 years on: A unifying framework. Int. J. Comput. Vis. **56** (2004) 221–255
9. Kenney, C., Manjunath, B., Zuliani, M., Hewer, G., van Nevel, A.: A condition number for point matching with application to registration and postregistration error estimation. IEEE Trans. PAMI **25** (2003) 1437–1454
10. Odobez, J.M., Bouthemy, P.: Robust multiresolution estimation of parametric motion models. J. Vis. Commun. Image R. **6** (1995) 348–365
11. Arnaud, E., Mémin, E., Cernuschi-Frias, B.: Conditional filters for image sequence based tracking - application to point tracking. IEEE Trans. IP **14** (2005) 63–79
12. Matthews, I., Ishikawa, T., Baker, S.: The template update problem. In: Proc. of British Machine Vision Conference. (2003)

13. Nguyen, H.T., Smeulders, A.W.M.: Fast occluded object tracking by a robust appearance filter. IEEE Trans. PAMI **26** (2004) 1099–1104
14. Darrell, T., Covell, M.: Correspondence with cumulative similarity transforms. IEEE Trans. PAMI **23** (2001) 222–227
15. Loutas, E., Diamantaras, K., Pitas, I.: Occlusion resistant object tracking. In: Proc. of ICIP. (2001) II: 65–68
16. Stauffer, C., Grimson, W.: Adaptive background mixture models for real-time tracking. In: Proc. of CVPR, IEEE (1999) II: 246–252
17. Fusiello, A., Trucco, E., Tommasini, T., Roberto, V.: Improving feature tracking with robust statistics. Pattern Anal. Appl. **2** (1999) 312–320