Real-time Dense Visual Tracking under Large Lighting Variations

Maxime Meilland[1]
maxime.meilland@inria.fr

Andrew Ian Comport[2]
comport@i3s.unice.fr

Patrick Rives[1]
patrick.rives@inria.fr

[1] INRIA Sophia Antipolis Méditerranée
2004 Route des Lucioles BP 93
Sophia Antipolis, France

[2] CNRS, I3S Laboratory,
Université Nice Sophia Antipolis
2000 Route des Lucioles BP 121
Sophia Antipolis, France

This paper proposes a model for large illumination variations to improve *direct* 3D tracking techniques since they are highly prone to illumination changes. Within this context dense monocular and multi-camera tracking techniques are presented which each perform in real-time (45Hz). The existing real-time *direct* tracking techniques which consider light change only tackle simple geometric scene models: planar homographies [3, 5] and only small regions (patches) are tracked. Alternatively, the method proposed here handles general 3D scenes, and many additional undesirable effects are present: occlusions, self-shadowing, objects interactions, specular reflections, camera saturation and also uncertainties on the scene geometry (e.g due to reconstruction). The central purpose of this paper is to address these illumination problems whilst maintaining a *dense* direct approach and considering real-time aspects.

In the case of direct *model-based* tracking, photometric models are usually acquired under significantly greater lighting differences than those observed by the current camera view, however, model-based approaches avoid drift. Incremental *visual odometry* [1], on the other hand, has relatively less lighting variation but integrates drift. The novelty is to propose a hybrid model-based/visual-odometry method which is robust to a large set of illumination changes whilst minimising drift with respect to a global reference frame (3D model).

Firstly, a model-based (MB) approach is considered here to be one that minimises the error between a known model (3D+photometric) and the warped current image. In terms of illumination, large changes can be expected for the Model-Based approach (see Fig. 2) since there is a large temporal difference between the moment the model was acquired and the current image.

Secondly, a visual odometry (VO) approach is defined as minimising the error between the image acquired at time $t-1$ and the warped current image at time t (see Fig. 1). In this case relatively small changes can be expected (even if they must still be modelled).

Direct 6 dof tracking is performed by an accurate method, which directly minimizes dense image measurements iteratively, using a robust non-linear optimisation [2, 4]. A stereo technique for automatically acquiring the 3D photometric model has also been optimised for the purpose of this paper.

Real experiments are shown on complex 3D scenes for a hand-held monocular camera undergoing fast 3D movement and various illumination changes including daylight, artificial-lights, significant shadows, non-Lambertian reflections, occlusions and saturations (see Fig. 2).

[1] A.I. Comport, E. Malis, and P. Rives. Real-time quadrifocal visual odometry. *In The International Journal of Robotics Research*, 29(2-3):245–266, February 2010.

[2] T. Gonçalves and A.I Comport. Real-time direct tracking of color images in the presence of illumination variation. In *IEEE International Conference on Robotics and Automation*, May 2011.

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

[4] P.J. Huber. *Robust Statistics*. New york, Wiley, 1981.

[5] G. Silveira and E. Malis. Real-time visual tracking under arbitrary illumination changes. In *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, pages 1 –6, 2007.

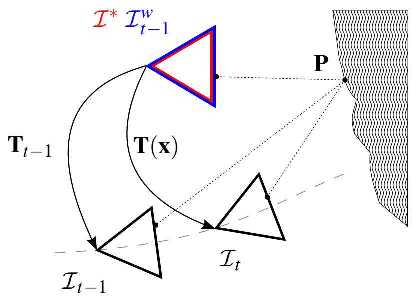
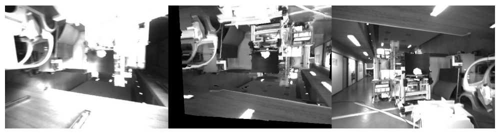
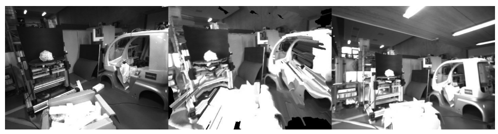


Figure 1: Cameras position in the scene. In red, the augmented reference image $\mathcal{I}^*$. In blue, the image $\mathcal{I}^w_{t-1}$ from time $t-1$ warped onto the reference. The current camera image $\mathcal{I}_t$ at time t. For the MB configuration the error $\mathcal{I}_t - \mathcal{I}^*$ is minimised and for the VO configuration $\mathcal{I}_t - \mathcal{I}^w_{t-1}$ is minimised.

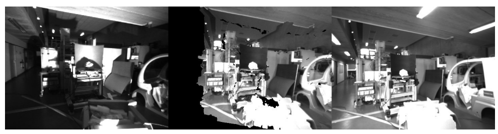
(a) Full 6 dof motion with local saturations.



(b) 180 degree rotation with local saturations.



(c) Night illumination with global illumination variation.



(d) Spot illumination with shadows, and local reflections.

(e) Spot illumination with occlusions.

Figure 2: Different tracking results. For each row of the image from left to right: Current image; Synthetic 3D view rendered at the estimated pose (corresponding to the current image pose); Original reference image