Proc. of International workshop, VLBV 2003. LNCS 2849, pp. 50-57

Tracking a planar patch by additive image registration

José Miguel Buenaposada, Enrique Muñoz, and Luis Baumela [*]

Departamento de Inteligencia Artificial, Universidad Politécnica de Madrid
Campus de Montegancedo s/n, 28660 Boadilla del Monte, Spain
{jmbuena,kike}@dia.fi.upm.es, lbaumela@fi.upm.es

Abstract. We present a procedure for tracking a planar patch based on a precomputed Jacobian of the target region to be tracked and the sum of squared differences between the image of the patch in the current position and a previously stored image of if. The procedure presented improves previous tracking algorithms for planar patches in that we use a minimal parameterisation for the motion model. In the paper, after a brief presentation of the incremental alignment paradigm for tracking, we present the motion model, the procedure to estimate the image Jacobian and, finally, an experiment in which we compare the gain in accuracy of the new tracker compared to previous approaches to solve the same problem.

1 Introduction

Image registration has traditionally been a fundamental research area among the image processing, photogrammetry and computer vision communities. Registering two images consist of finding a function that deforms one of the images so that it coincides with the other. The result of the registration process is the raw data that is fed to stereo vision procedures [1], optical flow estimation [2] or image mosaicing [3], to name a few.

Image registration techniques have also been used for tracking planar patches in real-time [4,5,6]. Tracking planar patches is a subject of interest in computer vision, with applications in augmented reality [7], mobile robot navigation [8], face tracking [9,6], or the generation of super-resolution images [10].

Traditional approaches to image registration can be broadly classified into feature-based and direct methods. Feature-based methods minimise an error measure based on geometrical constraints between a few corresponding features [11], while direct methods minimise an error measure based on direct image information collected from all pixels in the region of interest, such as image brightness [12]. The tracking method presented in this paper belongs to

[*] This work was supported by the Spanish Ministry for Science and Technology under contract numbers TIC1999-1021 and TIC2002-591

the second group of methods. It is based on minimising the sum-of-squared differences (SSD) between a selected set of pixels obtained from a previously stored image of the tracked patch (image template) and the current image of it. It extends previous approaches to the same problem [4,6] in that it uses a minimal parameterisation, which provides a more accurate tracking procedure.

In the paper, first we will introduce the fundamentals of the incremental image registration procedure, in section 3 we will present the motion model used for tracking and how to estimate the reference template Jacobian and, finally, in section 4 we show some experiments and draw conclusions.

2 Incremental image registration

Let $\mathbf{x}$ represent the location of a point in an image and $I(\mathbf{x}, t)$ represent the brightness value of that location in the image acquired at time t. Let $\mathcal{R} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ be a set of N image points of the object to be tracked (*target region*), whose brightness values are known in a reference image $I(\mathbf{x}, t_0)$. These image points together with their brightness values at the reference image represent the *reference template* to be tracked.

Assuming that the brightness constancy assumption holds, then

$$I(\mathbf{x}, t_0) = I(\mathbf{f}(\mathbf{x}, \bar{\mu}), t) \forall \mathbf{x} \in \mathcal{R}, \tag{1}$$

where $I(\mathbf{f}(\mathbf{x}, \bar{\mu}_t), t)$ is the image acquired at time t rectified with motion model $\mathbf{f}(\mathbf{x}, \bar{\mu})$ and motion parameters $\bar{\mu} = \bar{\mu}_t$.

Tracking the object means recovering the motion parameter vector of the target region for each image in the sequence. This can be achieved by minimising the difference between the template and the rectified pixels of the target region for every image in the sequence

$$\min_{\bar{\mu}} \sum_{\forall \mathbf{x} \in \mathcal{R}} [I(\mathbf{f}(\mathbf{x}, \bar{\mu}), t) - I(\mathbf{x}, t_0)]^2 \tag{2}$$

This minimisation problem has been traditionally solved linearly by computing $\bar{\mu}$ incrementally while tracking. We can achieve this by making a Taylor series expansion of (2) at ($\bar{\mu}$, t_n) and computing the increment in the motion parameters between two time instants. Different solutions to this problem have been proposed in the literature, depending on which term of equation (2) the Taylor expansion is made on and how the motion parameters are updated [13,4,3,5,14].

If we update the model parameters of the first term in equation (2) using an additive procedure, then the minimisation can be rewritten as [5,14]

$$\min_{\delta\bar{\mu}} \sum_{\forall \mathbf{x} \in \mathcal{R}} \left[I(\mathbf{f}(\mathbf{x}, \bar{\mu}_t + \delta\bar{\mu}_)), t + \delta t) - I(\mathbf{x}, t_0)\right]^2, \tag{3}$$

where $\delta\bar{\mu}$ represents the estimated increment in the motion parameters of the target region between time instants t and $t + \delta t$.

The solution to this linear minimisation problem can be aproximated by [14]

$$\delta\bar{\mu} = -\mathbf{H}_0^{-1} \sum_{\forall \mathbf{x} \in \mathcal{R}} \mathbf{M}(\mathbf{x}, \mathbf{0})^\top \mathcal{E}(\mathbf{x}, t + \delta t), \tag{4}$$

where $\mathbf{M}(\mathbf{x}, \mathbf{0})$ is the Jacobian vector of pixel $\mathbf{x}$ with respect to the model parameters $\bar{\mu}$ at time instant t_0 ($\bar{\mu} = \mathbf{0}$):

$$\begin{aligned} \mathbf{M}(\mathbf{x}, \mathbf{0}) = \left. \frac{\partial I(\mathbf{f}(\mathbf{x}, \bar{\mu}), t_0)}{\partial \bar{\mu}} \right|_{\bar{\mu}=\mathbf{0}} = \\ \nabla_{\mathbf{f}} I(\mathbf{f}(\mathbf{x}, \bar{\mu}), t_0)^\top \left[\frac{\partial \mathbf{f}(\mathbf{x}, \bar{\mu})}{\partial \bar{\mu}} \right]_{\bar{\mu}=\mathbf{0}}, \end{aligned}$$

$\mathbf{H}_0$ is the Hessian matrix

$$\mathbf{H}_0 = \sum_{\forall \mathbf{x} \in \mathcal{R}} \mathbf{M}(\mathbf{x}, \mathbf{0})^\top \mathbf{M}(\mathbf{x}, \mathbf{0}),$$

and $\mathcal{E}(\mathbf{x}, t + \delta t)$ is the error in the estimation of the motion of pixel $\mathbf{x}$ of the target region

$$\mathcal{E}(\mathbf{x}, t + \delta t) = I(\mathbf{f}(\mathbf{x}, \bar{\mu}_t), t + \delta t) - I(\mathbf{x}, t_0).$$

The Jacobian of pixel $\mathbf{x}$ with respect to the model parameters in the reference template, $\mathbf{M}(\mathbf{x}, \mathbf{0})$, is a vector whose values are our *a priori* knowledge about target structure, that is, how the brightness value of each pixel in the reference template changes as the object moves infinitesimally. It represents the information provided by each template pixel to the tracking process. Note that when $\mathbf{H}_0 = \sum_{\forall \mathbf{x} \in \mathcal{R}} \mathbf{M}(\mathbf{x}, \mathbf{0})^\top \mathbf{M}(\mathbf{x}, \mathbf{0})$ is singular the motion parameters cannot be recovered, this would be a generalisation of the so called *aperture problem* in the estimation of optical flow.

The steps of this tracking procedure are:

- Offline computations:
  1. Compute and store $\mathbf{M}(\mathbf{x}, \mathbf{0})$.
  2. Compute and store $\mathbf{H}_0$.
- On line:
  1. Warp $I(\mathbf{z}, t + \delta t)$ to compute $I(\mathbf{f}(\mathbf{x}, \bar{\mu}_t), t + \delta t)$.
  2. Compute $\mathcal{E}(\mathbf{x}, t + \delta t)$.
  3. From (4) compute $\delta\bar{\mu}$.
  4. Update $\bar{\mu}_{t+\delta t} = \bar{\mu}_t + \delta\bar{\mu}$.

3 Image formation and motion model

In this section we will introduce the target region motion model, $\mathbf{f}$, and the image Jacobian, $\mathbf{M}$, which are the basic components of our tracking algorithm.

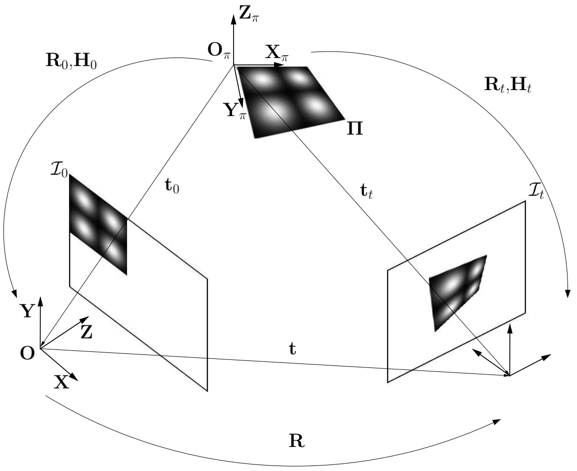

Fig. 1. Geometrical setup of the planar tracking system

3.1 Motion model

Let Π be a plane in 3D space which contains our target region, and let $\mathbf{X}_\pi = (X_\pi, Y_\pi, Z_\pi)^\top$ and $\mathbf{X} = (X, Y, Z)^\top$ be respectively the coordinates of points in the target region expressed, respectively, in a reference system attached to Π, $O_{X_\pi Y_\pi Z_\pi}$, and in a reference system attached to the camera, O_{XYZ} (see Fig. 1). We will assume that the first image in the sequence coincides with our reference template, $I_0(x) \equiv I(\mathbf{x}, t_0)$. The projection of a point $\mathbf{X}_\pi$ of the target region onto image I_i of the sequence is given by

$$\mathbf{x}_i = \mathbf{K}\,\mathbf{R}_i[\,\mathbf{I}\,|-\mathbf{t}_i\,]\begin{bmatrix}\mathbf{X}_\pi\\ 1\end{bmatrix}, \tag{5}$$

where $\mathbf{K}$ is the camera intrinsics matrix, which is assumed to be known for the whole sequence, $\mathbf{I}$ is the 3×3 identity matrix and $\mathbf{R}_i, \mathbf{t}_i$ represent the position of the camera that acquired image I_i with respect to $O_{X_\pi Y_\pi Z_\pi}$. for any point $P_\pi = [X_\pi, Y_\pi, Z_\pi, 1]^\top$ in the plane π.

Equation (5) can be simplified if we consider the fact that all points in the target region belong to the plane $\Pi : Z_\pi = 0$

$$\mathbf{x}_i = \mathbf{K}\,\mathbf{R}_i[\,\mathbf{I}^{12}\,|-\mathbf{t}_i\,]\begin{bmatrix}X_\pi\\ Y_\pi\\ 1\end{bmatrix}, \tag{6}$$

where $\mathbf{I}^{12}$ is the following 3×2 matrix

$$\mathbf{I}^{12} = \begin{bmatrix}1\;0\;0\\ 0\;1\;0\end{bmatrix}^\top .$$

The image motion model that we are seeking, $\mathbf{f}(\mathbf{x}_0, \bar{\mu})$, arises by just considering the relation that equation (6) provides for the projection of a point $\mathbf{X}_\pi \in \mathit{\Pi}$ into images I_0 and I_t of the sequence

$$\mathbf{x}_t = \underbrace{\mathbf{K}\,\mathbf{R}\mathbf{R}_0\,[\mathbf{I}^{12}| - (\mathbf{t}_0 + \mathbf{R}_0^\top \mathbf{t})]}_{\mathbf{H}_t}\,\underbrace{[\mathbf{I}^{12}| - \mathbf{t}_0]^{-1}\mathbf{R}_0^\top \mathbf{K}^{-1}}_{\mathbf{H}_0^{-1}}\,\mathbf{x}_0, \tag{7}$$

where $\mathbf{H}_i$ is the homography that relates $\mathit{\Pi}$ and image I_i, and $\mathbf{R}(\alpha, \beta, \gamma) = \mathbf{R}_t\mathbf{R}_0^\top$ [1] and $\mathbf{t}(t_x, t_y, t_z) = \mathbf{R}_0(\mathbf{t}_t - \mathbf{t}_0)$ are our motion model parameters. Note that vectors $\mathbf{x}_i$ represent a positions of pixels in image I_i and $\bar{\mu}^\top = (\alpha, \beta, \gamma, t_x, t_y, t_z)$ is the minimal parameterisation that represent the relative camera motion between the reference template, I_0, and the current image, I_t.

3.2 The image Jacobian

In order to simplify the notation, we will use projective coordinates, $\mathbf{x} = (r, s, t)^\top$ to represent the position of a point in an image. Let $\mathbf{x}_c = (u, v)^\top$ and $\mathbf{x} = (r, s, t)^\top$ be respectively the Cartesian and Projective coordinates of an image pixel. They are related by:

$$\mathbf{x} = \begin{pmatrix} r \\ s \\ t \end{pmatrix} \rightarrow \mathbf{x}_c = \begin{pmatrix} r/t \\ s/t \end{pmatrix} = \begin{pmatrix} u \\ v \end{pmatrix};\ t \neq 0. \tag{8}$$

Considering this relation, the gradient of the template image is

$$\nabla_{\mathbf{f}} I(\mathbf{f}(\mathbf{x}, \bar{\mu}), t_0)^\top = \left[\frac{\partial I}{\partial u}, \frac{\partial I}{\partial v}, -\left(u\frac{\partial I}{\partial u} + v\frac{\partial I}{\partial v}\right)\right], \tag{9}$$

and the Jacobian of the motion model with respect to the motion parameters

$$\left[\frac{\partial \mathbf{f}(\mathbf{x}, \bar{\mu})}{\partial \bar{\mu}}\right]_{\bar{\mu}=\mathbf{0}} = \left[\frac{\partial \mathbf{f}(\mathbf{x}, \bar{\mu})}{\partial \alpha}, \ldots, \frac{\partial \mathbf{f}(\mathbf{x}, \bar{\mu})}{\partial \mathbf{t}_z}\right]_{\bar{\mu}=\mathbf{0}}, \tag{10}$$

where, for example (for simplicity we assume $\mathbf{f}_x \equiv \frac{\partial \mathbf{f}}{\partial x}$),

$$\mathbf{f}_\alpha(\mathbf{x}, \mathbf{0}) = \mathbf{K}\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}\mathbf{R}_0\,[\mathbf{I}^{12}| - (\mathbf{t}_0 + \mathbf{R}_0^\top \mathbf{t})]\mathbf{H}_0^{-1}\mathbf{x}_0$$

$$\mathbf{f}_{t_x}(\mathbf{x}, \mathbf{0}) = \mathbf{K}\begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}\mathbf{H}_0^{-1}\mathbf{x}_0.$$

[1] Note that R_0 and t_0 can be inmediately computed if, for example, K and four points on $\mathit{\Pi}$ are known

4 Experiments and conclusions

In this section we will describe the experiment conducted in order to test the gain in performance obtained with the new tracking algorithm. We will compare the performance of the tracker presented in this paper, called *minimal ssd-tracker*, which uses a minimal six parameter-based model of the motion of the target region, and a previous tracker [6], called *projective ssd-tracker*, which estimates an eight parameter-based linear projective transformation model.

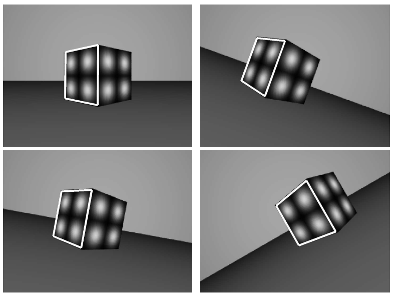

Fig. 2. Images 1, 100, 200 and 300 of the 300 images sequence used for the experiments. In white thick lines is shown the motion estimated by the *minimal ssd-tracker*.

We have used a synthetic image sequence generated using pov-ray[2], in this way we have ground truth data of the motion of our target region to compare the performance of the algorithms. In the sequence we have a planar patch located 4 meters away from a camera, which translates along the X axis (t_x) and rotates around the Z axis (γ) of the reference system associated with the first image of the sequence (see Fig. 1 and Fig. 2).

In Fig. 3 we show the ground truth values and the estimation of the α, γ, t_x, t_y parameters of the motion model for the *minimal* and *projective* ssd-trackers. In a second plot from the same experiment (see Fig. 4) is shown the rms error of the estimation of all parameters in the motion model. As can be seen in all plots, the performance of the *minimal tracker* is always equal or better than the previous *projective tracker*. More concretely, in Fig. 3 we can see that the estimation of the *minimal tracker* is most accurate for α, t_x, and t_y. These are the parameters for which the apparent image motion is smaller.

In conclusion, we have presented a new procedure for tracking a planar patch in real-time, which employs a minimal parameterisation for representing the motion of the patch in 3D space. The accuracy of this new procedure is clearly superior to previous solutions.

[2] A free ray tracer software, `http://www.povray.org`

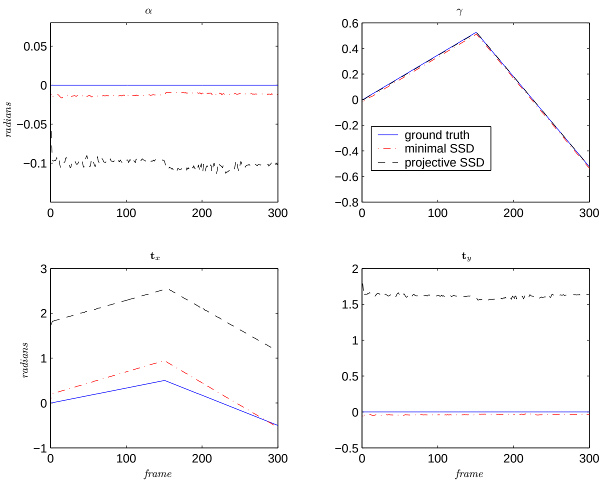

Fig. 3. Estimation of four motion parameters

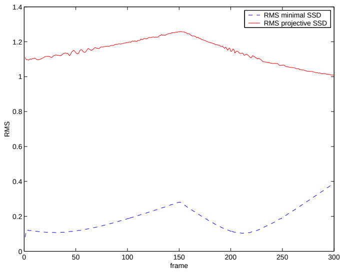

Fig. 4. Rms of the estimation of all six motion parameters.

References

1. G. Bradski and Boult T. F., Eds., *Special Issue in Stereo and Multi-Baseline Vision*, vol. 47. International Journal of Computer Vision, Kluwer, April–June 2002.
2. S. S. Beauchemin and J. L. Barron, "The computation of optical flow," *ACM Computing Surveys*, vol. 27, no. 3, pp. 433–467, 1995.
3. Heung-Yeung Shum and Richard Szeliski, "Construction of panoramic image mosaics with global and local alignment," *International Journal of Computer Vision*, vol. 36, no. 2, pp. 101–130, 2000.
4. Gregory D. Hager and Peter N. Belhumeur, "Efficient region tracking with parametric models of geometry and illumination," *IEEE Transactions on Pattern Analisys and Machine Intelligence*, vol. 20, no. 10, pp. 1025–1039, 1998.
5. Simon Baker and Ian Matthews, "Equivalence and efficiency of image alignment algorithms," in *Proc. of International Conference on Computer Vision and Pattern Recognition*. IEEE, 2001, vol. 1, pp. I–1090–I–1097.

6. José M. Buenaposada and Luis Baumela, "Real-time tracking and estimation of plane pose," in *Proc. of International Conference on Pattern Recognition, ICPR2002*, Quebec, Canada, August 2002, vol. II, pp. 697–700, IEEE.
7. G. Simon, A. Fitzgibbon, and A. Zisserman, "Markerless tracking using planar structures in the scene," in *Proc. International Symposium on Augmented Reality*, October 2000.
8. F. Lerasle V. Ayala, J.B. Hayet and M. Devy, "Visual localization of a mobile robot in indoor environments using planar landmarks," in *Proceedings Intelligent Robots and Systems, 2000.* IEEE, 2000, pp. 275–280.
9. M. J. Black and Y. Yacoob, "Recognizing facial expressions in image sequences using local parameterized models of image motion," *Int. Journal of Computer Vision*, vol. 25, no. 1, pp. 23–48, 1997.
10. C. Thorpe F. Dellaert and S. Thrun, "Super-resolved texture tracking of planar surface patches," in *Proceedings Intelligent Robots and Systems.* IEEE, 1998, pp. 197–203.
11. P. H. S. Torr and A. Zisserman, "Feature based methods for structure and motion estimation," in *Vision Algorithms: Theory and practice*, W. Triggs, A. Zisserman, and R. Szeliski, Eds. Springer-Verlag, 1999, pp. 278–295.
12. M Irani and P. Anandan, "All about direct methods," in *Vision Algorithms: Theory and practice*, W. Triggs, A. Zisserman, and R. Szeliski, Eds. Springer-Verlag, 1999.
13. Bruce D. Lucas and Takeo Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. of Imaging Understanding Workshop*, 1981, pp. 121–130.
14. J.M. Buenaposada, E. Muñoz, and L. Baumela, "Tracking head using piecewise planar models," in *Proc. of Iberian Conference on Pattern Recognition and Image Analysis, IbPRIA2003*, Puerto de Andratx, Mallorca, Spain, June 2003, vol. LNCS 2652, pp. 126–133, Springer.