Corner Matching Refinement for Monocular Pose Estimation

Dinesh Gamage
dinesh.gamage@moansh.edu

Tom Drummond
tom.drummond@monash.edu

Monash University
Australia

Accurate 3D structure calculation requires reliable feature extraction and matching. To improve the accuracy of the generated hypothesis, descriptor based feature matching which works at pixel level may not be adequate and sub-pixel level information may be needed. In this paper we propose a frequency domain optimisation technique, which yields improved results and a fast convergence rate. The fast convergence is a result of the multi-resolution nature of the solution.

We make use of the affine theorem in the frequency domain [1]. Given two image patches $I_0(\bar{x})$ and $I_1(\bar{x})$ surrounding two corresponding corners, which are related by an affine coordinate transformation $I_1(\bar{x}) = I_0\left[A^{-1}\left(\bar{x}-\bar{b}\right)\right]$, where A and b are the four non translational affine parameters and two translational parameters respectively, their 2-D Fourier transforms are related by:

$$\hat{I}_1(\bar{u}) = |det\,(A)|\, e^{-j\bar{u}\cdot\bar{b}} \hat{I}_0\left(A^T\bar{u}\right) \tag{1}$$

Here we use the six parameter affine model with an additional parameter. The seventh parameter compensates for energy changes caused by different local illumination conditions. If we select $\bar{\beta} = \{\beta_1 \ldots \beta_7\}$ to be the parameter set and absorb the $|det\,(A)|$ of the equation 1 into β_7 we have:

$$\beta_7\hat{I}_1(\bar{u}) = e^{-j\bar{u}\cdot\bar{b}}\hat{I}_0\left(A^T\bar{u}\right) \qquad \text{where} \quad A = \begin{pmatrix}\beta_1 & \beta_2\\ \beta_3 & \beta_4\end{pmatrix} \quad \text{and} \tag{2}$$

$$\bar{b} = \begin{pmatrix}\beta_5\\ \beta_6\end{pmatrix} \tag{3}$$

$$\text{so} \quad \beta_7 e^{j\bar{u}\cdot\bar{b}}\hat{I}_1(\bar{u}) = \hat{I}_0\left(A^T\bar{u}\right) \tag{4}$$

Thus the error r, for a frequency $\bar{u}$ can be written as,

$$r\left(\bar{u},\bar{\beta}\right) = \beta_7 e^{j\bar{u}\cdot\bar{b}}\hat{I}_1(\bar{u}) - \hat{I}_0\left(A^T\bar{u}\right) \tag{5}$$

The above equation enables us to model the affine transformation as a phase change of $\hat{I}_1$ and a warp of $\hat{I}_0$ with respect to matrix A. The Jacobian J_i of partial derivatives of r with respect to β_i can then be computed:

$$\bar{J} = \left[-\frac{\partial I_0}{\partial u}u,\ -\frac{\partial I_0}{\partial u}v,\ -\frac{\partial I_0}{\partial v}u,\ -\frac{\partial I_0}{\partial v}v,\ -\beta_7\tilde{I}_1 u,\ -\beta_7\tilde{I}_1 v,\ \tilde{I}_1\right] \tag{6}$$

Given a set of frequencies $\{u_j\}$, the errors $r(u_j)$ and the Jacobian J_{ij} can be used to obtain the parameters $\bar{\beta}$ that minimise $E = \sum_j \|r(u_j)\|^2$ using Gauss-Newton.

FFT requires a periodic signal. So each patch has to be compensated for edge effects at the border. Secondly, the presence of a large DC component in the signal corrupts low frequency components of the signal in the frequency domain (those where $\|u\|$ is small). To remove edge effects from the image patches, we multiply it by a Gaussian weighting window ($G(x,y)$) centered at the detected landmark before taking the Fourier transform. Before doing that, the DC component of the patch which appears as a large spike at $u = 0$ in the frequency domain can be removed by subtracting the average, which gives a new patch. After alleviating both of these effects we get a new patch $I'(x,y)$ defined as:

$$I'(x,y) = G(x,y)\left(I(x,y) - \frac{\sum_{x,y}G(x,y)I(x,y)}{\sum_{x,y}G(x,y)}\right) \tag{7}$$

This gives a patch with a 0 DC coefficient. The frequency response of the Gaussian multiplied patch, $\mathscr{F}[I']$ has a direct relationship with the Gabor filter with an identical Gaussian support. We use this relationship to select the useful frequency range (in order to eliminate possible aliasing effects) for the optimisation in a multi resolution manner.

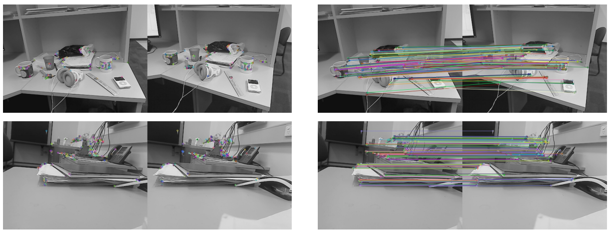

Figure 1: Pose estimation after sub-pixel refinement

Because the phase of a particular Gabor filter response changes linearly under spatial translations of the signal, it can be used to estimate spatial disparity of two instances of a the same signal with a relative shift [2]. This phase disparity is useful only if the displacement is smaller than a half a wavelength of the tuning frequency [2] of the Gabor filter, i.e the domain of convergence for the phase is $\pm\pi$. This imposes an upper frequency limit for the useful frequency range. If we assume a maximum displacement of d pixels for a 1-D signal this criteria suggests a frequency f such that $f \leq 1/2d$. In 2-D this requirement can be met by limiting the useful frequency range radially to a maximum of $1/2d$ radius. After estimating the translation (and other parameters) using small frequencies for large displacements finer refinements can be done gradually by increasing the radius, incorporating higher frequency responses to the optimisation.

Though subtracting the DC component spatially as in equation 7 can mostly reduce its effect, for better results we have to impose a lower frequency limit as well. We select the minimum frequency using the one octave bandwidth criteria which has been suggested in the literature for Gabor filter based disparity estimations. The one octave bandwidth in the frequency domain for the Gabor filter shows the spatial support to be:

$$\sigma = \frac{1}{2\pi f}\left(\frac{2^\alpha+1}{2^\alpha-1}\right) \tag{8}$$

If we select above frequency f keeping the spatial support σ a constant, in order to eliminate any DC distortion the minimum frequency should be:

$$f \geq \frac{1}{2\pi\sigma}\left(\frac{2^\alpha+1}{2^\alpha-1}\right) \tag{9}$$

Combining the minimum and the maximum criteria for frequency selection gives:

$$\frac{1}{2d} \geq f \geq \frac{1}{2\pi\sigma}\left(\frac{2^\alpha+1}{2^\alpha-1}\right) \tag{10}$$

At the end of each iteration we can expect the displacement d to reduce, increasing the useful frequency range. Higher frequencies carry finer details about the translation, which improves the final solution. This naturally enables a multi-resolution framework for refinement without any additional computations.

[1] RN Bracewell, K.Y. Chang, AK Jha, and Y.H. Wang. Affine theorem for two-dimensional fourier transform. *Electronics Letters*, 29(3): 304, 1993.

[2] D.J. Fleet, A.D. Jepson, and M.R.M. Jenkin. Phase-based disparity measurement. *CVGIP: Image understanding*, 53(2):198–210, 1991.