INCORPORATING THE EPIPOLAR CONSTRAINT INTO A MULTIRESOLUTION ALGORITHM FOR STEREO IMAGE MATCHING

JULIAN MAGAREY, ANTHONY DICK, MIKE BROOKS,
GARRY NEWSAM, and ANTON VAN DEN HENGEL
Cooperative Research Centre for Sensor Signal and Information Processing
Mawson Lakes, Adelaide, SA 5095, Australia

ABSTRACT

We present a new algorithm for matching calibrated stereo image pairs. Matching is achieved within a multiresolution framework that utilises a complex wavelet transform of each image. Integrated within this coarse-to-fine approach is a regularisation step at each resolution. Calibration information, in the form of the epipolar constraint, is incorporated into the regularisation step. Uncertainty in the calibration parameters can be accommodated. This composite matching scheme avoids the need for expensive prior resampling of one image along epipolar lines and yields excellent results in various tests with real images.

Keywords: Wavelets, matching, stereo vision, epipolar constraint.

1 INTRODUCTION

A major challenge in computer vision is the so-called *stereo vision* task of reliably reconstructing a 3D representation of a scene given two perspective views. The key sub-problem is to accurately locate a dense set of *corresponding pairs*, these being image points that depict a common scene point. Many automated stereo systems adopt the strategy of locating a feature in one image and selecting the location of the most similar feature in an appropriate sub-region of the other image. However, areas of an image often lack distinct features such as corners, instead exhibiting smooth untextured brightness values. An alternative approach is to carry out dense matching, whereby correspondences are sought for every pixel in one image. This approach uses more information than merely a set of isolated features, thus reducing the potential for ambiguous matching. Preferably this should be carried out within a multiresolution framework whereby matching initially takes place at a coarse level, with the results being passed down to successively finer levels for refinement.

If the calibration and relative orientation parameters of two cameras are known, then the *epipolar constraint* [1] may be inferred. In theory, given a point in one image, this constraint specifies a line in the other image along which the corresponding point must lie. The constraint therefore serves to confine the matching problem to a one-dimensional search, lessening the chances of mismatching.

One method of incorporating the epipolar constraint into a matching strategy is to *rectify* one image with respect to the other. After this resampling, corresponding features are guaranteed to be displaced by a purely horizontal disparity [1]. This reduces the matching to a very simple 1D problem, which may be handled efficiently using, for example, dynamic programming [2]. However, such an approach does not easily take into account any uncertainty in the calibration parameters, and the initial rectification step is expensive to compute.

The algorithm for calibrated stereo matching described in this paper mitigates the aforementioned difficulties. We make use of a Bayesian formulation of stereo matching [3] and incorporate the epipolar constraint as an *a priori* probability distribution on the disparity field. Furthermore our chosen feature space, derived from the Complex Discrete Wavelet Transform [4], is dense and relatively robust to global and local intensity perturbations. It is also amenable to a multiresolution or pyramidal approach.

This new composite matching scheme generates dense disparity fields well approximating the epipolar constraint. The correspondences generated tend to be locally consistent over textureless surfaces and to provide close matches for salient features of the scene. The algorithm is demonstrated by reconstructing human faces from calibrated stereo image pairs, a challenging application due to the scarcity of distinctive feature points, and the sensitivity of the human eye to distortions or abnormalities in the appearance of a reconstructed face.

2 THE MATCHING ALGORITHM

2.1 Feature matching

We have previously described our algorithm for uncalibrated stereo matching [5]. It is based on similarity in a *multiresolution* feature space. Complex-valued feature vectors $\mathbf{D}_l$ and $\mathbf{D}_r$, for left and right images, respectively, embody the amplitudes and phases of inner products with 2D Gabor functions over a variety of different orientations and a *dyadic* (powers-of-2) set of scales. Feature extraction is achieved efficiently by means of the Complex Discrete Wavelet Transform (CDWT) [4]. The matching strategy is coarse-to-fine, in that searches at each scale are guided by the results propagated down from the previous larger scale. The matching criterion at a given scale is defined so as to be primarily dependent on phase, and hence robust to minor variations in local intensity which might result from fluctuating lighting conditions. Feature vectors have a known behaviour between the values on a fixed subsampling grid, so a matching *surface* may be defined on a compact 2D interval around each grid point or *subpixel*.

Bayesian estimation is expressed in terms of probability density functions (pdfs). The *likelihood* function is based on feature similarity, and may be approximated at each level m subpixel $\mathbf{x}_r$ in the right image by a bivariate pdf in the disparity $\mathbf{f}$:

$$p(\mathbf{D}_l(\mathbf{x}_r)|\mathbf{D}_r(\mathbf{x}_r),\mathbf{f}) \propto \exp\{-\frac{1}{2\sigma^2}SD^{(m)}(\mathbf{x}_r,\mathbf{f})\} \quad (1)$$

where $SD^{(m)}(\mathbf{x}_r,\mathbf{f})$ is the *squared difference* surface based on the level m feature vectors $\mathbf{D}_l^{(m)}$ and $\mathbf{D}_r^{(m)}$ at $\mathbf{x}_r$:

$$SD^{(m)}(\mathbf{x}_r,\mathbf{f}) = \left\|\mathbf{D}_l^{(m)}(\mathbf{x}_r+\mathbf{f}) - \mathbf{D}_r^{(m)}(\mathbf{x}_r)\right\|^2. \quad (2)$$

The maximum likelihood estimate of the disparity is therefore that which minimises the squared difference. (The constant σ^2 is the variance of the Gaussian pixel intensity noise, assumed to be independently identically distributed.)

2.2 Incorporating calibration information

The calibration information is given in the form of the *perspective projection* matrices $\tilde{P}_l$ and $\tilde{P}_r$ for the left and right images respectively. These relate homogeneous scene coordinates $\tilde{\mathbf{X}}$ to the corresponding homogeneous image coordinates $\tilde{\mathbf{x}}$ for each camera as follows [1]:

$$\tilde{\mathbf{x}}_l = \tilde{P}_l\tilde{\mathbf{X}} \quad (3)$$

$$\tilde{\mathbf{x}}_r = \tilde{P}_r\tilde{\mathbf{X}} \quad (4)$$

It may be shown that the corresponding left and right image coordinates (i.e. those representing the same scene point $\mathbf{X}$) are related as follows:

$$\tilde{\mathbf{x}}_l^T F \tilde{\mathbf{x}}_r = 0, \quad (5)$$

where F is the 3×3 *fundamental matrix* obtainable from $\tilde{P}_l$ and $\tilde{P}_r$. Equation (5), known as the *epipolar constraint*, defines, for every pixel $\mathbf{x}_r$ in the right image, a line $l(\mathbf{x}_r)$ (the *epipolar line*) in the left image along which its correspondent $\mathbf{x}_l$ must lie.

At level m of the CDWT feature pyramid, write

$$\mathbf{x}_l = \mathbf{x}_r + 2^m\mathbf{f}_r, \quad (6)$$

where $\mathbf{f}_r = \left[\begin{array}{cc} f_{1r} & f_{2r}\end{array}\right]$ is the true right-left disparity, normalised to the level m grid spacing of 2^m. In terms of $\mathbf{f}_r$, the epipolar line may be written as

$$f_{1r} + Qf_{2r} + R = 0, \quad (7)$$

where

$$Q = \frac{F_2\tilde{\mathbf{x}}_r}{F_1\tilde{\mathbf{x}}_r}, \quad (8)$$

$$R = \frac{\tilde{\mathbf{x}}_r^T F\tilde{\mathbf{x}}_r}{2^m F_1\tilde{\mathbf{x}}_r} \quad (9)$$

$$F = \left[\begin{array}{ccc} F_1^T & F_2^T & F_3^T\end{array}\right]^T. \quad (10)$$

To incorporate this constraint deterministically (i.e. as a *hard constraint*), we would simply maximise the likelihood function (1) over the epipolar line, which is a 1D search problem. However, in a Bayesian estimation framework, we may incorporate it *probabilistically* (i.e. as a *soft constraint*), allowing for the possibility of uncertainty in the exact location of the epipolar line (due to errors in calibration). To do this, as with any feature-independent information about the disparity, it should be expressed as a separate pdf $p(\mathbf{f})$ called the *prior*. This is multiplied by the likelihood function to form the *a posteriori* pdf:

$$p(\mathbf{f}|\mathbf{D}_l(\mathbf{x}),\mathbf{D}_r(\mathbf{x})) \propto p(\mathbf{D}_l(\mathbf{x})|\mathbf{D}_r(\mathbf{x}),\mathbf{f}) \times p(\mathbf{f}) \quad (11)$$

The epipolar prior at the right subpixel $\mathbf{x}_r$ may be written as

$$p(\mathbf{f}) \propto \exp\{-\frac{1}{(\tau/2^m)^2}S_{ep}(\mathbf{x}_r,\mathbf{f})\}, \quad (12)$$

where

$$S_{ep}(\mathbf{x}_r,\mathbf{f}) = \frac{f_1^2 + Q^2f_2^2 + 2Qf_1f_2 + 2Rf_1 + 2QRf_2 + R^2}{2(1+Q^2)}.$$

Here, S_{ep} is a trough-shaped surface which reaches its minimum value of zero all along the line $l(\mathbf{x}_r)$. It has a parabolic cross-section with unit curvature. The constant τ captures the uncertainty (in pixels) in the location of the epipolar line, with scaling by 2^{-m} to allow for the varying definition of $\mathbf{f}$ with m.

To multiply this by the likelihood, we simply add the exponents to obtain a new, epipolar-constrained squared difference surface:

$$SD_{ep}^{(m)}(\mathbf{x}_r,\mathbf{f}) = SD^{(m)}(\mathbf{x}_r,\mathbf{f}) + \frac{2.4^m}{\rho^2}S_{ep}(\mathbf{x}_r,\mathbf{f}) \quad (13)$$

It is clear that the ratio $\rho = \tau/\sigma$ controls the relative weighting of the epipolar constraint. If it is large, because of significant uncertainty in the calibration parameters, the epipolar constraint is very weak. On the other hand, for very noisy images, ρ will be small, thus increasing the effect of the constraint. In the limiting case of $\rho \to 0$ the constraint becomes "hard" as defined above.

2.3 Regularisation

Previously, we showed that incorporating a regularisation step at each level significantly improves the robustness of the matching in the uncalibrated case [5]. This we retain in the calibrated case. We note that the constrained squared difference surface

$SD_{ep}^{(m)}$, like the unconstrained $SD^{(m)}$, may be well approximated by a quadratic provided $\mathbf{f}$ is not too large, viz:

$$SD_{ep}^{(m)}(\mathbf{x}_r, \mathbf{f}) \approx \frac{1}{2}(\mathbf{f} - \mathbf{f}_0)^T \mathcal{K}(\mathbf{f} - \mathbf{f}_0) + \delta \qquad (14)$$

The surface minimum location $\mathbf{f}_0$ forms our initial estimate of the disparity at $\mathbf{x}_r$, while the symmetric 2×2 *curvature* matrix $\mathcal{K}$ encodes the shape of the surface, in particular how rapidly it increases as $\mathbf{f}$ moves away from $\mathbf{f}_0$ in each direction. The regularisation procedure, which makes use of this shape information, is based on that of Anandan [6]. Given the field $\{\mathbf{f}_0\}$, we aim to obtain an optimal compromise between feature similarity and disparity field *smoothness* by finding a field $\{\mathbf{u}_0\}$ which minimises the functional

$$E\left(\{\mathbf{u}\}\right) = E_{sm}\left(\{\mathbf{u}\}\right) + \lambda E_{ap}\left(\{\mathbf{u}\}\right). \qquad (15)$$

The term $E_{ap}\left(\{\mathbf{u}\}\right)$ is the global "approximation error" energy, which is defined as the normalised sum of the differences between $SD_{ep}^{(m)}(\mathbf{x}_r, \mathbf{u})$ and the feature-space minimum value δ. (Note that $E_{ap}\left(\{\mathbf{f}_0\}\right) = 0$). The smoothness term $E_{sm}\left(\{\mathbf{u}\}\right)$ is defined as the energy of the gradients of the components (u_x, u_y) of $\mathbf{u}$ over the image.

An approximate minimum of the functional E is found by a Gauss-Seidel iterative procedure using the eigenvalues and eigenvectors of $\mathcal{K}$ [5]. The scaling factor λ in (15) controls the relative influence of field smoothness with respect to feature similarity. The smaller the value of λ, the greater the smoothing effect of the regularisation.

2.4 Coarse-to-fine refinement

Matching begins at the coarsest level m_{max}. The smoothed disparity field $\{\mathbf{u}_0\}$ is used as an initial value for estimation at the next finer level $m_{max} - 1$, after scaling up by 2 to allow for the denser spacing of subpixels. We use the *overlapping pyramid* scheme of Anandan [6], in which the disparities at the four level $m + 1$ subpixels directly subtending the level m subpixel under scrutiny are all used as candidate starting guesses. The candidate which leads to the best match (lowest δ value) is selected. Estimation proceeds in this fashion, decrementing m from coarse to fine, with matching followed by regularisation, until the finest level of detail ($m = m_{min}$) is reached. The set of corresponding pairs is then given by $\{\mathbf{x}_r + \mathbf{u}_0(\mathbf{x}_r), \mathbf{x}_r\}$. Finally, we use the projection matrices $\tilde{P}_l$ and $\tilde{P}_r$ to find the corresponding 3D scene point $\mathbf{X}$ by simultaneously solving (3) and (4) in a procedure known as *triangulation* [1]. In general, the system is overdetermined, so a least-squares solution is found. It may be shown that if $\mathbf{x}_l$ lies on $l(\mathbf{x}_r)$, i.e. the pair satisfies the epipolar constraint, the residual error of the solution is zero.

3 EXPERIMENTAL RESULTS

To demonstrate first of all that the unconstrained algorithm works well in most circumstances, we include some results from the stereo test pair (each 768 by 572 pixels) shown in Figure 1. Two views of the surface reconstructed from the unconstrained disparities are also shown in Figure 1. The original (right) image has been texture-mapped to the surface.

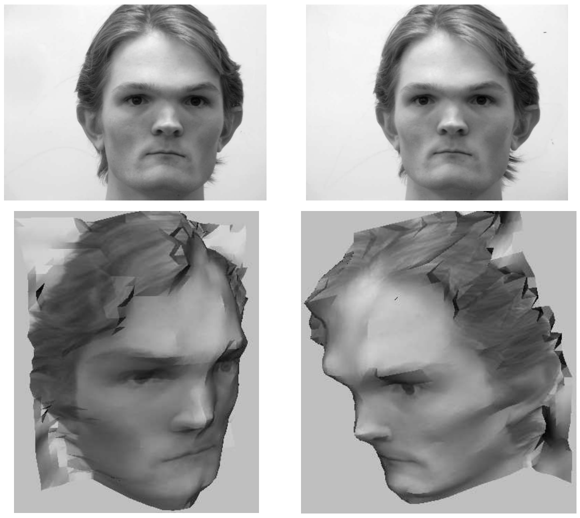

Figure 1: Stereo image pair, and two views of reconstructed surface with image texture-mapped.

We expect the application of the epipolar constraint to improve the matching in two distinct ways. Firstly, at the topmost level, where there is no starting guess to guide estimation, we need to perform a search over a number of candidate left subpixels for each right image subpixel; the one providing the lowest value for δ is chosen for further refinement. If a "false" candidate match is chosen, it is likely this error will persist down the pyramid, giving rise to catastrophic errors in reconstruction. The epipolar constraint guides the top-level search to the most likely set of correspondents, i.e. those lying on or near the epipolar line. This greatly reduces the possibility that a "false" candidate match will be selected. Thus the epipolar constraint should greatly reduce the catastrophic errors, which show up as large spikes or holes in the reconstructed surface.

Secondly, the epipolar constraint should produce an overall improvement in the accuracy of the reconstructed surface by providing more accurate correspondences in the matching stage. This effect should show up as a significant reduction in the mean squared error in the triangulation stage (which for a "hard" epipolar constrained disparity field should be zero).

To demonstrate these two effects, we carried out constrained and unconstrained matching on the calibrated stereo pair of Figures 2(a) and (b), captured using a stereo camera head. The reconstructions obtained without and with the epipolar constraint are shown in Figures 2(c) and (d). (Algorithm parameters: $m_{max} = 6, m_{min} = 1, \lambda = 0.2$; $\rho = \infty$ for unconstrained, $\rho = 1$ for constrained matching.) The original image has been texture-mapped to the surface in each case.

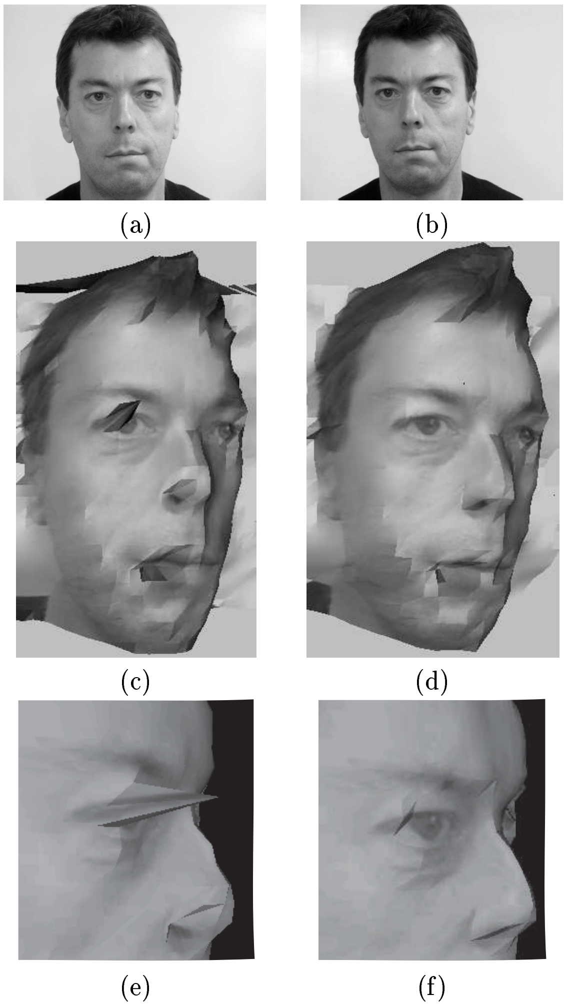

Figure 2: (a) Left image. (b) Right image. (c) Reconstructed surface without epipolar constraint. (d) Reconstructed surface with epipolar constraint. (ce) Detail from right eye area, of reconstruction without epipolar constraint. (f) Same area of reconstruction with epipolar constraint.

The catastrophic error in the unconstrained reconstruction is a spike on the right eye. This has been eliminated in the constrained reconstruction, as the enlargements in Figures 2(e) and (f) show.

To show the overall improvement in accuracy over reasonably well-matched regions, we selected a few points around the left eye and drew the corresponding epipolar lines in the other image. The corresponding points found by the unconstrained and constrained matching algorithms were superimposed on the line image. The results are shown in Figure 3. It may be seen that the corresponding points lie closer to the epipolar lines in the constrained matching case. This improvement is observable over the entire image, but most particularly in regions of sparse detail. Indeed the RMS residual triangulation error over the face is reduced by a ratio of approximately 10 after the application of the epipolar constraint.

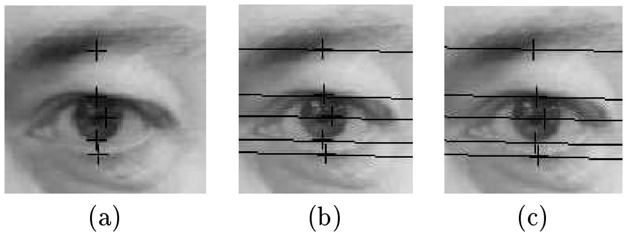

Figure 3: (a) Points in region of left eye. (b) The corresponding points found by unconstrained matching, together with the corresponding epipolar lines. (c) The corresponding points found by constrained matching, together with the corresponding epipolar lines.

4 CONCLUSION

We have described a method for incorporating the epipolar constraint into our algorithm for multiresolution stereo image matching. The constraint, derived from the projection matrices which are the output of the calibration procedure, is incorporated in a "soft" or probabilistic fashion into the complex-wavelet-domain algorithm. This allows the balancing (by means of an adjustable parameter ρ) of the relative errors in the feature matching and the calibration parameters. We have demonstrated two benefits of the additional constraint: preventing catastrophic errors caused by mismatches at the topmost pyramid level, and reducing the RMS triangulation error over the reconstructed surface.

References

[1] O. Faugeras. *Three-dimensional computer vision.* MIT Press, 1993.

[2] Y. Ohta and T. Kanade. Stereo by intra- and inter-scanline search. *IEEE Trans. Pattern. Anal. Mach. Intell.*, 7(2):139–154, 1985.

[3] H. Pan and J.F.A. Magarey. Multiresolution phase-based bidirectional stereo matching with provision for discontinuity and occlusion. *Digital Signal Processing—A Review Journal*, 1999. *(To appear)*.

[4] J.F.A. Magarey and N.G. Kingsbury. Motion estimation using a complex-valued wavelet transform. *IEEE Transactions on Signal Processing*, 46(4):1069–1084, April 1998.

[5] J.F.A. Magarey and A. Dick. Multiresolution stereo image matching using complex wavelets. In *Proc. 14th Int. Conf. on Pattern Recognition (ICPR)*, volume I, pages 4–7, August 1998.

[6] P. Anandan. A computational framework and an algorithm for the measurement of visual motion. *Intern. J. Comput. Vis.*, 2:283–310, 1989.