Multiresolution Stereo Image Matching Using Complex Wavelets

Julian Magarey and Anthony Dick
Cooperative Research Centre for Sensor Signal and Information Processing
SPRI Building, Technology Park Adelaide, SA 5095, Australia
Email: jfam@cssip.edu.au, anthony@cs.adelaide.edu.au

Abstract

This paper describes a multiresolution image-matching strategy, based on the Complex Discrete Wavelet Transform (CDWT), to derive a dense disparity field with hierarchical (coarse-to-fine) refinement. The CDWT feature space efficiently provides fractionally accurate matching results which are robust to typical image formation perturbations such as offsets, global scaling, and additive noise. At each level of the hierarchy, the disparity field is regularised to provide a global compromise between feature similarity and disparity field continuity, resulting in feature-sensitive smoothing. The algorithm is well suited to analysing facial images, for which we demonstrate striking reconstruction results.

1. Introduction

This paper describes a depth reconstruction algorithm for binocular stereo which uses a dense multiresolution matching strategy. Multiresolution strategies have been shown to be useful in resolving the ambiguities of dense matching smooth surfaces which lack clear feature points and contain few if any occlusions [5]. The matching is fully two-dimensional as we do not make use of the epipolar constraint to reduce the dimensionality of the problem. The key feature of our algorithm is the feature space in which matching takes place, which is based on the Complex Discrete Wavelet Transform [6]. It has been demonstrated that this feature space yields matching results which are more robust than original-pixel-intensity-domain matching to typical distortions of binocular stereo pairs such as level shifts, scaling, and additive noise [6]. Even so, various sources of error can corrupt the disparity field, particularly over broad expanses with little texture or detail. For this reason, following Anandan [1], we insert a regularisation step after matching at each level, which trades off feature space similarity for disparity field smoothness. Our matching procedure provides a directional confidence measure which assists this regularisation step, preserving sharp features while smoothing over untextured regions.

We chose to demonstrate the performance of our algorithm on faces for several reasons. Firstly, face recognition is a significant application area in its own right, and 3-d reconstructions can be a key part of this. Secondly, as faces consist mostly of smooth untextured surfaces with relatively few strong features, they present severe problems for feature-based matching algorithms, but are well suited to our dense matching algorithm. Thirdly, the human visual system is strongly adapted to analysing facial structure, drawing the eye to any areas which do not “look right”. Thus we have a ready-made mechanism for qualitatively assessing the performance of our algorithm in the absence of ground truth values.

2. Outline of our algorithm

Most previous matching strategies have been based on cross-correlation in the intensity domain [3]. While this achieves satisfactory results in certain cases, it is usually implemented using a computationally intensive exhaustive search. It has been shown that phase-based methods of disparity measurement [4] can overcome some of the weaknesses of intensity-domain correlation in stereo matching. The principle is based on the the Fourier shift theorem, which relates global translations in the spatial domain to phase rotations in the frequency domain. Application of this theorem to the coefficients of a local Fourier basis expansion can yield a spatially varying field of disparity estimates which is robust to certain image formation perturbations including level shifts and intensity scaling. Gabor basis functions (Gaussian-windowed complex exponentials) are often used because of their optimum combined compactness in time and frequency domains [2].

2.1. The Complex Discrete Wavelet Transform

Our aim was to implement a multiresolution phase-based matching algorithm. The separable discrete wavelet trans-
2.2. Matching in the complex wavelet domain

The CDWT is applied to both left and right images. We define our CDWT-domain matching criterion between level \( m \) left and right subpixels \( \mathbf{x} \) and \( \mathbf{x}' \) as follows:

\[
SD^{(m)}(\mathbf{x}, \mathbf{x}') = \sum_{n=1}^{6} \left| D^{(n,m)}_l(\mathbf{x}) - D^{(n,m)}_r(\mathbf{x}') \right|^2
\]

The correspondent for the left subpixel \( \mathbf{x} \) is taken to be the right subpixel \( \mathbf{x}' \) which minimises this quantity. The level \( m \) left-right disparity \( d_{mn} \) is \( \mathbf{x}' - \mathbf{x} \). (Because of the downsampling, the disparity over the region of pixels subtended by \( \mathbf{x} \) is \( 2^m d_{mn} \).

In a coarse-to-fine strategy, it is important that the matching subpixel is located to fractional accuracy. The Gabor-like nature of the wavelet filters allows us to accurately estimate coefficient values in between the subpixel grid, by interpolation from the surrounding grid coefficients [6]. Hence, given a starting subpixel \( \mathbf{x}' \) for each \( \mathbf{x} \), we can use the coefficient \( D^{(n,m)}_l(\mathbf{x}) \) and those surrounding \( D^{(n,m)}_r(\mathbf{x}') \) to estimate a matching surface consisting of the values of \( SD^{(m)}(\mathbf{x}, \mathbf{x}'+\mathbf{f}) \) over a real 2-d interval of offsets \( \mathbf{f} \). Over small offsets, a 2-d quadratic model fits the surface well:

\[
SD^{(m)}(\mathbf{x}, \mathbf{x}'+\mathbf{f}) \approx \frac{1}{2}(\mathbf{f} - \mathbf{f}_0)^T K (\mathbf{f} - \mathbf{f}_0) + \delta
\]

The surface parameters \( \{\mathbf{f}_0, K, \delta\} \) (the surface minimum location, curvature matrix, and minimum value) may be computed directly from the coefficients \( D^{(n,m)}_l(\mathbf{x}) \) and \( D^{(n,m)}_r(\mathbf{x}') \) and the spatial frequencies \( \Omega_{n,m} \). It turns out that \( \mathbf{f}_0 \) and hence the level \( m \) disparity, depends heavily on the phase of the two complex coefficient vectors, while the magnitudes have comparatively little influence. The dependence of disparity on phase and insensitivity to magnitude gives the algorithm immunity to global perturbations such as level shifts and intensity scaling. This means that the matching is based on a generalised level-and-scaling-independent area similarity, increasing the robustness of the algorithm.

The shape of the matching surface \( SD^{(m)} \), encapsulated by the symmetric \( 2 \times 2 \) matrix \( K \), acts as our directional measure of confidence. It may be shown [6] that the curvature of the surface in a given direction indicates our relative confidence in the component of \( \mathbf{f}_0 \) in that direction.

2.3. Regularisation of the disparity field

Even with its robustness, the disparity field is subject to random errors. If these were allowed to propagate down the pyramid, the final reconstructed surface would be extremely jagged. To minimise this effect, we use a regularisation procedure based on that proposed by Anandan [1]. Given the field \( \{\mathbf{f}_0\} \) of feature space minimum location offsets, we aim to obtain an optimal compromise between feature similarity and disparity field smoothness by finding a field \( \{\mathbf{u}\} \) which minimises the functional

\[
E(\{\mathbf{u}\}) = E_{sm}(\{\mathbf{u}\}) + \lambda E_{ap}(\{\mathbf{u}\})
\]

The term \( E_{ap}(\{\mathbf{u}\}) \) is the global “approximation error” energy, which we define as the normalised sum of the differences between \( SD(\mathbf{x}, \mathbf{x}'+\mathbf{u}) \) and the feature-space minimum value \( \delta \):

\[
E_{ap}(\{\mathbf{u}\}) = \sum_{\mathbf{x}} \frac{SD^{(m)}(\mathbf{x}, \mathbf{x}'+\mathbf{u}) - \delta}{\delta}
\]

The smoothness term \( E_{sm}(\{\mathbf{u}\}) \) is defined in terms of the gradients of the components \( (u_x, u_y) \) of \( \mathbf{u} \):

\[
E_{sm}(\{\mathbf{u}\}) = \sum_{\mathbf{x}} ||u_x||^2 + ||u_y||^2
\]

The scaling factor \( \lambda \) in (3) controls the relative influence of each term on the resultant disparity field. The smaller it is made, the greater the smoothing effect of the regularisation.

An exact global minimisation of \( E(\{\mathbf{u}\}) \) is computationally expensive to compute. A good approximate solution my be found using only the local information surrounding each subpixel, by solving

\[
(\mathbf{u} - \mathbf{m}) + \lambda c_{nx} \Delta x + \lambda c_{ny} \Delta y = 0
\]
shown in Figure 1. They are given a realistic appearance
workstation.
processing time was approximately ten minutes on a Sun Ultra
ascertained camera calibration matrices in a triangulation
field (of size 286 by 384) was combined with the previously
obtained in a manner robust to global intensity perturbations such as
low efficient fractionally-accurate matching at each level,
in a manner robust to global intensity perturbations such as
scaling, offsets, and additive noise. Disparity estimates are
accompanied by an intuitively meaningful directional confidence
measure. The confidence measure is utilised in a regularisation
which is suitable for producing dense depth maps for
smooth and relatively featureless surfaces from stereo image
pairs. The algorithm is multiresolution in structure
and is based on similarity in the Complex Discrete Wavelet Transform domain. The properties of this feature space allow
efficient fractionally-accurate matching at each level,
in a manner robust to global intensity perturbations such as
scaling, offsets, and additive noise. Disparity estimates are
accompanied by an intuitively meaningful directional confidence
measure. The confidence measure is utilised in a reg-
ularisation which is carried out at each level of
processing to smooth away noise-induced disparity errors
in relatively featureless regions while preserving disparities
based on clear features. The effectiveness of the algorithm
has been demonstrated in 3-d reconstruction of faces from
calibrated stereo image pairs.

4. Conclusion

In this paper we have described a matching algorithm
which is suitable for producing dense depth maps for
smooth and relatively featureless surfaces from stereo image
pairs. The algorithm is multiresolution in structure
and is based on similarity in the Complex Discrete Wavelet Transform domain. The properties of this feature space allow
efficient fractionally-accurate matching at each level,
in a manner robust to global intensity perturbations such as
scaling, offsets, and additive noise. Disparity estimates are
accompanied by an intuitively meaningful directional confidence
measure. The confidence measure is utilised in a reg-
ularisation which is carried out at each level of
processing to smooth away noise-induced disparity errors
in relatively featureless regions while preserving disparities
based on clear features. The effectiveness of the algorithm
has been demonstrated in 3-d reconstruction of faces from
calibrated stereo image pairs.

References

for the measurement of visual motion. Intern. J. Comput. Vis.,
tial frequency, and orientation optimized by two-dimensional
1993.
using visual front-end operations with automatic scale selec-
tion. In Proc. 5th International Conference on Computer Vi-
complex-valued wavelet transform. IEEE Transactions on
[7] S. Mallat. A theory for multiresolution signal decomposi-
tion: The wavelet representation. IEEE Trans. on PAMI,
Figure 1. Facial image pairs and 3-d reconstructions viewed from two angles.