Scene Flow Estimation using Intelligent Cost Functions

Simon Hadfield
S.Hadfield@surrey.ac.uk

Richard Bowden
R.Bowden@surrey.ac.uk

Centre for Vision Speech and Signal Processing
University of Surrey
Surrey, UK

Scene flow is the 3D counterpart to optical flow, describing the 3D motion field of a scene, independent of the cameras which view it. Motion estimation techniques (both scene flow and optical flow) is a fundamental tool in computer vision. It forms the basis or pre-processing step for many other algorithms, and is included in many vision libraries. These techniques are typically based upon the assumption of brightness constancy, or related assumptions such as gradient constancy and filter response constancy.

A lot of previous work has been dedicated to accurately modelling the behaviour of these consistency assumptions, in the motion fields of real scenes. This helps handling scene artifacts such as non lambertian surfaces, illumination changes and occlusions. In this paper we extend this analysis further, and examine the behaviour of visual consistency assumptions, in cases where the motion field has been incorrectly estimated.

Distinguishing truth from errors

Intuitively, the accurate modelling of visual consistency for ground truth motion fields helps ensure that correct motion fields are always recognised as such (reducing "False negatives"). However, it tells us nothing about the metrics ability to reject erroneous motion fields ("False positives").

For our analysis we examine a range of common visual constancy assumptions. These include the Optical Flow Constraint (OFC), L_2 brightness constancy (SQ), and gradient constancy counterparts OFC_g and SQ_g.

Ideally these metrics should provide a low cost for true motions and a high cost for incorrect motions, as illustrated by the PDF in fig. 1(a). In this case the true motion field registers no violation in the underlying assumption (the PDF contains all responses at 0), while the incorrect motions strongly violate the assumption (PDF is concentrated at 1).
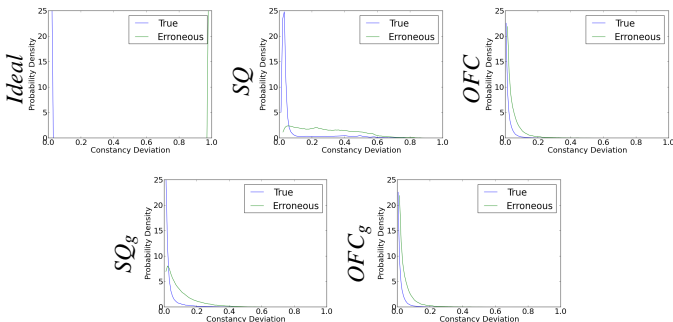

Figure 1: Responses for various motion estimation metrics (including an ideal example), applied to the ground truth and error motion fields of a real scene.

The actual response distributions found for the examined metrics tell an unfortunate story. Most ground truth motions are assigned to the lower 20% of the responses, with the occlusion and specularity effects seen previously being the minority. However, similar responses are produced, even for the significantly erroneous motions. Indeed the linearised brightness constancy metric OFC shows an 80% overlap between the two PDFs. Attempting to minimize these metric responses across the scene, will result in almost as many correct motions being discarded, as incorrect. In the full paper, further analysis is performed to examine how the metric response changes as the amount of error in the motion field varies (i.e. does the response smoothly decrease, as the error is reduced).

Intelligent Cost Functions

We propose a simple solution to this problem; Explicitly finding discriminatory metrics, using machine learning techniques. These "Intelligent cost functions" (ICFs) are able to embody more complex behaviours. As an example, it may be expected that in very light or dark parts of the scene, image contrast would be reduced. In this case, little variation may be expected naturally, and any appearance deviations may be more significant. Alternatively, specular effects may cause a large change in appearance across all colour channels, while a change in appearance for only one channel is more likely to relate to an erroneous motion.
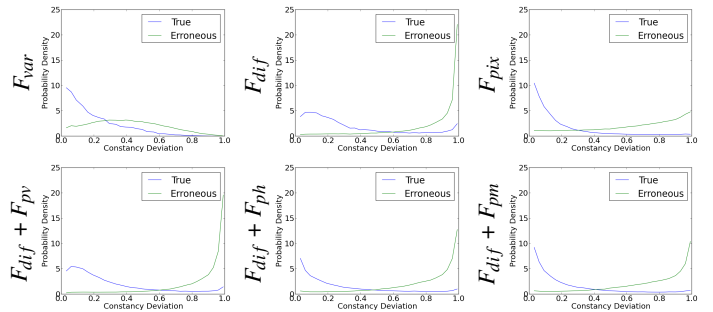

Figure 2: Distribution of responses for various ICFs, applied to ground truth and erroneous motions. Top row: pixel features. Bottom row: local context features.

To learn these Intelligent Cost Functions (ICFs) Gaussian Processes (GP) were trained to model the relationship between various input features and the level of motion error. The GP provides a non-parametric means for fitting these complex relationships, by estimating a distribution over the infinite set of possible cost functions.

Fig. 2 shows the performance of ICFs, based on various visual features (see supplementary material[1] for results on a range of additional sequences). The simplest (F_{var}) provides little additional separation. However, in the case of F_{dif} encoding, and the local context features, the ICF exploits richer infrmation to greatly improve separation.

Motion Estimation with ICFs

We've seen that standard motion estimation cost functions have some significant flaws, and that greater robustness may be obtained via ICFs. However, much work in motion estimation (particularly for optical flow where there are no problems with differing sensor responses) has looked at producing specialised subsystems to mitigate, rather than correct, these issues. As such, it is important to examine whether the use of ICFs does in fact translate to more accurate motion estimates.

Metric	ε_{of}	ε_{sf}	ε_{st}	ε_{ae}	Runtime (secs)
SQ	0.173	0.010	1.52	1.66	352
F_{var}	0.164	0.021	1.53	1.63	389
F_{dif}	0.111	0.009	**1.04**	1.41	363
F_{pix}	0.142	0.012	1.17	1.47	**340**
$F_{dif} + F_{pv}$	0.100	**0.005**	1.10	1.50	440
$F_{dif} + F_{ph}$	0.134	0.008	1.14	1.59	560
$F_{dif} + F_{pm}$	**0.098**	0.014	1.06	**1.23**	430

Table 1: Performance for scene flow estimation, based on the original SQ metric, and a range of ICFs.

To this end, a recent, publicly available, algorithm for scene flow estimation (based on the SQ cost function) is modified to exploit ICFs. Results in tab. 1 are averaged over all sequences from the Middlebury dataset. The results show an almost universal improvement in motion estimation accuracy, with $F_{dif} + F_{pm}$ providing improvements to magnitude, directional and structural accuracies of 44%, 20% and 30% respectively.

We also examine the behaviour of ICFs in optical flow scenarios. In this case we discover a more modest 20% improvement in accuracy.
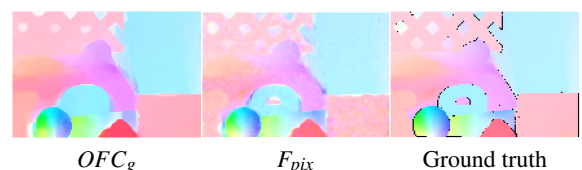

| OFC_g | F_{pix} | Ground truth |

Figure 3: Example motion fields for one of the Middlebury sequences, comparing a standard metric and an ICF against the ground truth.