Bilateral Functions for Global Motion Modeling

Wen-Yan Daniel Lin[1], Ming-Ming Cheng[2], Jiangbo Lu[1], Hongsheng Yang[3], Minh N. Do[4], and Philip Torr[2]

[1] Advanced Digital Sciences Center, Singapore
[2] Oxford University, UK
[3] University of North Carolina at Chapel Hill, USA
[4] University of Illinois at Urbana-Champaign, USA

Abstract. This paper proposes modeling motion in a bilateral domain that augments spatial information with the motion itself. We use the bilateral domain to reformulate a piecewise smooth constraint as continuous global modeling constraint. The resultant model can be robustly computed from highly noisy scattered feature points using a global minimization. We demonstrate how the model can reliably obtain large numbers of good quality correspondences over wide baselines, while keeping outliers to a minimum.

1 Introduction

Finding point-to-point correspondence between images is a fundamental vision problem. Applications include recognition, structure from motion, self-localization, warping, etc. For wide baselines, researchers typically focus on matching scattered, feature points which are distinctive and easy to match. Despite substantial success in feature descriptor design [1,2,3], basing correspondence solely on local information remains an unstable, outlier prone process.

Researchers usually handle outliers at the application level through task-specific motion models[1], that are often integrated into a RANSAC [4,5] outlier removal framework. Some of the most successful techniques, are based around identifying specific motion types/aspects that are amenable to global parameterization. Examples include epipolar geometry [6] for different views of a static scene, homography [7,8] for planar or pure rotational motion and non-rigid thin-plate splines [9] for smooth deformations. We believe the high reliance on task-specific global models highlights two issues:

A) Strength of global modeling: Global models have many features important to the correspondence problem. These are:

- Robustness: By defining a global rigidity/ smoothness, model computation can potentially tolerate high noise levels and even handle correlated noise [10].
- Scattered Samples: Models can be computed on a sub-set of available data and the results extrapolated. This approach permits computational efficiency and a natural interface with sparse feature matchers.

*This work was supported by the research grant for the Human Sixth Sense Programme at the Advanced Digital Sciences Center from Singapore's Agency for Science, Technology and Research (A*STAR).

[1] **Motion model:** a finite global parameter set that defines an aspect of a continuous motion field.

D. Fleet et al. (Eds.): ECCV 2014, Part IV, LNCS 8692, pp. 341–356, 2014.

– Generalization: Models can verify existing data and hypothesize new data points.

B) Lack of a general motion model: Creating a general motion model for image correspondence is difficult. Not only are correspondence points scattered and noisy, the primary underlying constraint is piecewise smoothness, with the potential for large motion discontinuities. Thus, modeling requires detection and change of parametrization at motion boundaries; a process that risks preserving outlier clusters and destroys the model's global properties. Yet, exclusively focusing on easily modeled motion aspects discards a great deal of information. This increases the brittleness of more general models like the epipolar constraint, while restricting the flexibility of robust models like homography. Our paper fills the gap by reformulating the piecewise-smoothness as a robust, global constraint. This allows model fitting, outlier removal and matching set expansion to begin before reaching the application level.

Our key concept is a general definition of motion coherence. Traditionally, coherence [11] is equated to spatial (x, y) smoothness, i.e. a motion model is coherent if the motion or its proxy[2] varies smoothly over the spatial domain. In contrast, we suggest a motion model be considered coherent if the motion/proxy varies smoothly in some low dimensional domain.

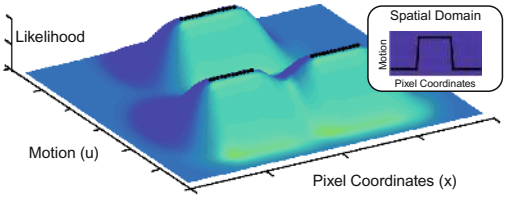


Fig. 1. Inset: a one-dimensional discontinuous set of motion hypothesis Main figure: the same data (black dots) over the bilateral domain with a likelihood motion proxy. Note: the bilateral domain expresses discontinuous data as a smooth field.

In particular, an extended domain that includes the motion itself, x, y, u, v, allows modeling of (spatially) piecewise smooth motions with a smoothly varying motion proxy. A detailed explanation is given in Sec. 3. We term such functions, which achieve smoothness by incorporating the desired output as part of the function domain, *bilateral functions*. Fig. 1 illustrates a bilateral function with a likelihood motion proxy. The bilateral domain makes it possible to fit a global function to (spatially) piecewise smooth motion data via the traditional as-smooth-as-possible data modeling. This can be solved by minimizing a global cost. The global smoothness makes bilateral motion model computation highly robust and it can be computed directly from noisy correspondence without RANSAC's [4] hypothesis and test framework. Once computed, models can robustly validate new matching hypothesis. This provides large numbers of correspondence points over wide baselines, while keeping outliers to a minimum (zero in many cases). Fig. 2 illustrates our performance.

To summarize our paper's contributions:

– We propose a bilateral model as a principled means of imposing a piecewise motion constraint via a global cost minimization. Our model is sufficiently robust to handle high noise levels without RANSAC's hypothesis and test framework.

[2] **Motion proxy:** a value which indicates but does not define the motion (an example is likelihood or a single affine parameter)

(a) (b) (c)

Fig. 2. An example of our global method for finding correspondences. Circles represent feature locations and lines their motion. a) Noisy feature correspondence (inliers highlighted in dark blue). b) Our global model eliminates outliers. The model is directly computed from the noisy matches in (a). c) The global model allows us to robustly expand the set of matches.

– We utilize the bilateral model to remove outliers from feature correspondences and expand the existing matching set. The resultant algorithm procures large numbers of correspondences while keeping outliers to a minimum. over wide baselines.

2 Related Works

Formulation: The piecewise smooth motion constraint has a long history in computer vision. Examples range from piecewise smooth flow estimation [12,13] to plane-fitting stereo [14]. However, to our knowledge, piecewise smoothness has always been enforced with some form of motion boundary discovery. Rather than the usual piecewise smooth approaches, our formulation builds on the global coherence [11,15,16,10] framework which fits a smooth, continuous field over scattered points. By applying the field on the bilateral domain, the global coherence accommodates discontinuities, while retaining its original robustness.

RANSAC: For rigid parametric models, RANSAC [4,17,18] provides a general means of removing outliers from feature correspondence. Of particular relevance are recent piecewise planar RANSAC techniques [18], which handle general motion. Unfortunately, they relinquish the global model, forcing tight thresholding (that removes many inliers). The lack of a global model also makes generalization for matching set expansion difficult. Interestingly, our bilateral coherence constraint is sufficiently robust to directly fit highly noisy data, without RANSAC's hypothesis and test framework. This enables our flexible non-rigid model. If additional parametric constraints are appropriate, RANSAC can be applied as a post-processor to further improve correspondence.

Fig. 3. An image pair with huge motion. The warp projects the second image onto the first. While this is not a difficult feature correspondence scene, the large occlusions makes it difficult for optical flow.

Optical flow [21,22,20], graph matching [23], surface modeling [24]: An alternative to modeling feature correspondences is optical flow or graph matching techniques. These approaches embed the smoothness constraint at the matching step. This reduces correspondence ambiguity and opens the possibility of dense matching. However, wide baselines introduce extensive occlusion which can turn the linkage of every pixel/feature through a neighbor-wise smoothness into a liability as seen in Fig. 3. For such scenes, our integration of bilateral modeling with point correspondence provides a natural means of handling high occlusion. This is discussed in more detail in the supplementary material.

Others: Outlier detection can also be achieved by local mesh techniques developed by Pizarro and Bartoli [25]. However, the lack of a global model reduces its effectiveness at higher noise level.

As our name suggests, bilateral functions are inspired by bilateral filters [26] that use the same domain change technique for edge preserving image de-noising. Bilateral filters have also been applied to optical flow computation [27]. However, they do not deal with the scattered point sets. Further, their window based approach makes identifying outlier clusters difficult. By extending the bilateral concept as a function domain, we overcome both these restrictions.

Interestingly, bilateral functions are related to patch-match [28] and quasi-dense correspondence algorithms [29,30]. These techniques grow seed correspondences by iteratively searching their neighborhood. This growing mechanics is similar manner to minimization of a bilateral likelihood function as elaborated in Sec. 3.2. However, patch-match does not compute an explicit model and it is unclear how its formulation can extend to outlier removal or multi-image matching.

3 Formulation

We begin with an intuitive explanation of our approach. Our underlying function is based on the motion coherence used in [11,15,16,10]. These techniques formulate a non-rigid motion fitting problem in terms of finding the smoothest $f_k(\mathbf{p})$ function that is consistent with given data. In these cases, $\mathbf{p}$ represents pixel coordinates, while the range of $f_k(\mathbf{p})$ represents motion or its proxy.

The smoothness (or infinite differentiability) implies a continuity constraint

$$\lim_{\Delta\mathbf{p}\mapsto 0} f_k(\mathbf{p}+\Delta\mathbf{p}) - f_k(\mathbf{p}) = 0, \quad \mathbf{p} = \begin{bmatrix} x & y \end{bmatrix}^T \in \mathbb{R}^2, \tag{1}$$

which forces the function value in the neighborhood of $\mathbf{p}$ to be similar. This causes the function to incur large errors at discontinuous motion boundaries.

Our formulation changes the domain of $f_k(\mathbf{p})$ to a bilateral one spanning both the spatial and motion dimensions i.e. $\mathbf{p} = \begin{bmatrix} x & y & u & v \end{bmatrix}^T$. This might mean that points with different velocities are no longer near each other. Thus, we can assign very different function values to points with adjacent spatial coordinates, while retaining the constraint that $f_k(\mathbf{p})$ must be smooth. If the motion difference $(\Delta u, \Delta v)$ between two points in the domain $\mathbf{p}$ tends to infinity, the point separation also tends to infinity, reducing their influence on each other. This occurs irrespective of their spatial coordinates and ensures the smoothness penalty in the bilateral domain does not tend to infinity as the magnitude of the motion's spatial discontinuity increases.

The (spatial) piecewise smoothness of the true underlying motion, creates large clusters of inlier points that are similar in both spatial and motion values, while sharing similar motion proxies. These points can be fitted at minimal cost to the smoothing function. In contrast, outliers appear as isolated point clusters requiring their own unique motion proxies. These incur a high smoothness cost if fit. The overall problem can now cast as finding the globally smoothest function consistent with the data. As we do not modify the as-smooth-as possible requirement, the global curve fitting [10] retains all is original robustness. Fig. 1 provides a visualization of the bilateral domain.

Preliminaries: We discuss two different bilateral functions in Sec. 3.1, Sec. 3.2. Sec. 3.3 gives an intuitive explanation of their properties while implementation is discussed in Sec. 3.4. Formally, we denote spatial locations as $\mathbf{x} = [\,x\;y\,]^T$ and motion as $\mathbf{m} = [\,u\;v\,]^T$. The set of N correspondences across two images is $\{\mathbf{x}_j, \mathbf{m}_j, \mathbf{a}_j\}$. $\mathbf{x}_j$ denotes pixel locations, $\mathbf{m}_j$ their corresponding motion hypothesis, and $\mathbf{a}_j$ is a 4×1 vector representing the relative affine orientation derived from local feature's orientation.

3.1 Bilaterally Varying Affine

A bilaterally varying affine is a set of affine parameters which vary smoothly in a bilateral domain. $\mathbf{p} = [\,x\;y\;u\;v\,]^T$ is a point in $D = 4$ dimension bilateral domain and q is a scalar. For simplicity, we focus on the X motion direction first. The given correspondences are observed data:

$$\textit{observed data} = \{\mathbf{p}_j = [\,\mathbf{x}_j; \mathbf{m}_j\,], \hat{q}_{xj} = x_j + u_j\} \tag{2}$$

where j is the correspondence index, and $\{\hat{q}_{xj}\}$ are noisy observations of model $q_x(\mathbf{p})$ evaluated at locations $\{\mathbf{p}_j\}$. We define $q_x(\mathbf{p})$ as a linear sum of smooth functions

$$q_x(\mathbf{p}) = f_1(\mathbf{p})x + f_2(\mathbf{p})y + f_3(\mathbf{p}), \tag{3}$$

where each $f_k(.)$ represents an affine parameter.

Our goal is to fit the smoothest possible, continuous, $f_k(.)$ functions to observed, $\{\hat{q}_{xj}\}$ data. Individual $f_k(.)$ functions are composed of two terms, $f_k(\mathbf{p}) = H_k + \phi_k(\mathbf{p})$.

H_k is an optional scalar offset and $\phi_k(\mathbf{p})$ is a smooth function with attached motion coherence [11,15,10] penalty:

$$\Psi_k = \int_{\mathbb{R}^D} \frac{|\overline{\phi}_k(\omega)|^2}{\overline{g}(\omega)}\, d\omega. \tag{4}$$

$\overline{\phi}_k(.)$ denotes the Fourier transform of a function $\phi_k(.)$, while $\overline{g}(\omega)$ is the Fourier transform of a Gaussian with spatial distribution γ. Hence, (4) achieves smoothness by penalizing high frequency terms.

We seek the $f_k(\mathbf{p})$ functions which minimize the cost

$$E = \sum\nolimits_{j=1}^{N} C(\hat{q}_{xj} - q_x(\mathbf{p}_j)) + \lambda \sum\nolimits_{k=1}^{3} \Psi_k \tag{5}$$

where $C(.)$ represents the Huber cost

$$C(z) = huber(z) = \begin{cases} z^2 & \text{if } |z| \leq \epsilon \\ 2\epsilon|z| - \epsilon^2 & \text{if } |z| > \epsilon \end{cases} \tag{6}$$

that penalizes deviation of the estimated function predictions from given $\hat{q}_{xj}$ observations. λ represents the weight given to the smoothness constraint Ψ_k.

From [10], we know that both the continuous functions $f_k(\mathbf{p})$ and the coherence terms Ψ_k can be re-expressed in terms of a finite number of variables given by the N-dimensional vectors $\mathbf{w}_k$ and scalars H_k:

$$f_k(\mathbf{p}) = H_k + \sum\nolimits_{j=1}^{N} \mathbf{w}_k(j) g(\mathbf{p} - \mathbf{p}_j, \gamma), \quad \Psi_k = \mathbf{w}_k^T G \mathbf{w}_k, \quad k \in \{1,2,3\} \tag{7}$$

where $g(\mathbf{z}, \gamma) = e^{-|\mathbf{z}|^2/\gamma^2}$ and $G(i,j) = g(\mathbf{p}_i - \mathbf{p}_j, \gamma)$. This allows us to minimize the energy in (5) with gradient descent as it is convex in terms of the variables $H_k, \mathbf{w}_k$. The Huber based energy in (5) allows the robust, non-parametric fitting of smooth curves by leveraging the smoothness constraint to ignore individual outliers and outlier clusters [10]. Note that the resultant bilateral affine model (3), is not a one-to-one mapping function. Rather it validates a match location $\mathbf{p}$ by checking the cost $(q_x - (x+u))^2$.

Similarly, for the Y direction, repeating the same steps as X but replacing $k \in \{1,2,3\}$ with $k \in \{4,5,6\}$ respectively and replacing $\hat{q}_{xj}$ with $\hat{q}_{yj} = y_j + v_j$, one obtains the bilateral affine model

$$q_y(\mathbf{p}) = f_4(\mathbf{p})x + f_5(\mathbf{p})y + f_6(\mathbf{p}). \tag{8}$$

Note that the formulation used here is not restricted to bilateral affines, with different choices of $f_k(.)$ domain and range leading to different functions as we demonstrate in the following subsection.

3.2 Likelihood Proxy

Another potential motion proxy is likelihood. For this proxy, we choose an 8 dimensional bilateral domain over spatial position, motion and relative feature affine parameters. A point in the domain given by $\mathbf{p} = [\mathbf{x}; \mathbf{m}; \mathbf{a}]$. The function range is set from

$[\,0\;1\,]$. Each match j hypothesizes a 1 value at location $\mathbf{p}_j$. Thus the *observed data* in (2) takes the form:

$$\textit{observed data} = \{\mathbf{p}_j = [\,\mathbf{x}_j;\,\mathbf{m}_j;\,\mathbf{a}_j\,], \hat{q}_j = 1\}.$$

We can fit a likelihood surface, $f(\mathbf{p})$, to the *observed data* in a manner similar to the bilaterally varying affine of Sec. 3.1. In this case, we have only one smooth function, $q = f(\mathbf{p})$. $f(\mathbf{p})$ is subject to the smoothness constraint

$$\Psi = \int_{\mathbb{R}^8} \frac{|\overline{f}(\omega)|^2}{\overline{g}(\omega)}\, d\omega. \tag{9}$$

Similar to (5), we seek the final $f(\mathbf{p})$ that minimizes the cost

$$E = \sum\nolimits_{j=1}^{N} huber(1 - f(\mathbf{p}_j)) + \lambda\Psi. \tag{10}$$

Without the H_k offset terms in Sec. 3.1, the Fourier smoothness in (9) causes the $f(\mathbf{p})$ function to return to zero unless given correspondence data biases it to 1. The robust fitting provided by both the Huber function and the smoothness requirement ensures that the cost in (10) does not fit the given data blindly but rejects correspondence clusters without sufficient support.

As in (7), from [10] we know that $f(\mathbf{p})$ and Ψ can be re-parametrized in terms of a N-dimensional $\mathbf{w}$ vector.

$$f(\mathbf{p}) = \sum\nolimits_{j=1}^{N} \mathbf{w}(j) g(\mathbf{p} - \mathbf{p}_j, \gamma), \quad \Psi = \mathbf{w}^T G \mathbf{w} \tag{11}$$

This allows gradient descent minimization of cost in (10), which is convex in $\mathbf{w}$.

$f(\mathbf{p})$ in (11) forms a motion proxy which indicates whether matches at a specific $\mathbf{p}$ location should be considered an inlier. An example with a simplified $\mathbf{p} = [\,x\;u\,]^T$ is shown in Fig. 1. Interestingly, algorithms that iteratively grow seed matches [28,29,30] exploit a similar coherence. By biasing the matching search radius towards the surroundings of pre-existing matches, they implicitly update a bilateral field with likelihood potentials similar to Fig. 1. This may explain why, despite having no explicit smoothing function, such techniques provide matching results with a strong sense of overall coherence (albeit with some outliers).

3.3 One-to-Many Mapping

Observe that the bilateral motion models indicate likely motion directions, rather than specify a one-to-one mapping between images. While inconvenient, the one-to-many mapping reflects the reality of multiple motion layers. For a foreground image point, its alternative motion hypothesis reflects the estimated motion of the occluded background. For a background image point, the alternative motion hypothesis suggests its position if it were actually part of the foreground. Due to this ambiguity, bilateral motion models must be integrated with some image measure (such as gray-level value or SIFT descriptors) to provide a final match location. In practice, for independently obtained matching hypothesis, the motion model can reliably distinguish between inliers

and outliers, through a thresholding based on deviation from the model. The discrimination is surprisingly good. Even when the given point correspondences contain many outliers, the computed motion model can preserve a large fraction of the inliers, while eliminating all outliers. An example can be seen in the inlier recall of Fig. 2c) with more rigorous analysis performed in Sec. 4.1. Apart from validating matches, the model can also be used to search for desired features as discussed in Sec. 5.2.

3.4 Implementation

This section gives a broad overview of the computation of bilateral models from feature correspondence. We primarily use A-SIFT [1] for matching. The SIFT threshold is set at $\mathbf{t} = 0.82$. This incurs many more outliers than the typical $\mathbf{t} = 0.66$. However, for wide baselines, it provides nearly an order of magnitude more inlier matches as seen in Fig. 8. The inliers at the relaxed threshold are also more evenly distributed, allowing for better motion modeling. The spatial coordinates of features are Hartley normalized [31] to allow parameters to be invariant to image size. We compute the likelihood field given by $f(\mathbf{p})$ in (11) by finding the $\mathbf{w}$ that minimizes the energy function E in (10). We accept as inliers all matches with likelihood greater than 0.5 (i.e. $f(\mathbf{p}_j) > 0.5$). This thresholding is deliberately weak and there are still outliers remaining. The inlier set obtained from the likelihood function is used to obtain a bilaterally varying affine in (3), (8) by finding the $H_k, \mathbf{w}_k$ values that minimize the energy in (5). A match $\mathbf{p} = [x\ y\ u\ v]^T$ is accepted if $(q_x(\mathbf{p}) - (x+u))^2 + (q_y(\mathbf{p}) - (y+v))^2 < 0.01$. The likelihood cost has lower long range ambiguity (removes extreme outliers) but does not validate matches finely (will accept matches with some localization error). Its simpler formulation also makes it more stable. The bilateral affine penalizes localization error but will occasionally accept gross outliers if they coincide with the affine. While the stages can theoretically be integrated into a single cost, they are computed separately for speed. If there too many points, we perform random sub-sampling to keep computation time constant.

Once the bilateral model is computed, we can use it to validate matching hypothesis. In this paper, our matching hypothesis are SIFT matches with no nearest neighbor thresholding but other methods of generating hypothesis can be considered. Fig. 4 gives a system overview.

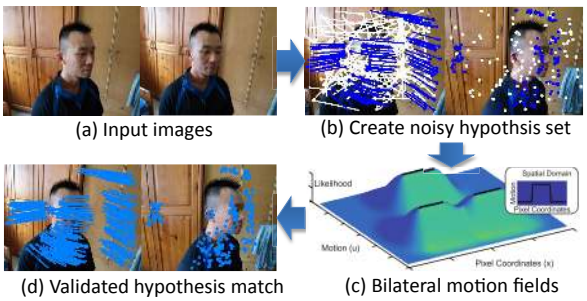


Fig. 4. System overview: to obtain large numbers of high quality matches, we compute bilateral motion fields according to noisy hypothesis set, and use the model to validate matching without nearest neighbor thresholding.

4 Experiments and Discussion

This section focuses on empirical evaluation. Images are evaluated at 480×640 resolution. For data with known camera pose, an inlier is a match whose deviation from epipolar geometry is less than 5 pixels. An outlier is one that deviates more than 40 pixels (due to large scale changes, this is not especially generous). The strict threshold for inliers and outliers, ensures that algorithms are neither penalized nor rewarded based on classification of ambiguous matches. We urge readers to view the supplementary material which contains many images and visualizations of the empirical results.

4.1 Inlier-Outlier Discrimination for Measuring Motion Model Quality

The bilateral motion model does not define a one-to-one correspondence between images, making direct quantitative evaluation difficult. However, the overall correctness of the model and robustness of its computation can be indirectly measured via its ability to discriminate between inlier and outlier correspondences. While many modern outlier removal techniques work robustly at standard 0.66 A-SIFT thresholds, our evaluation uses a more challenging 0.82 threshold[3]. For wide baselines, this also has the practical advantage of providing many more inliers as illustrated in Fig. 8. *Occ* [25] (an excellent outlier detector for general motion) forms the baseline. One measure of inlier-outlier

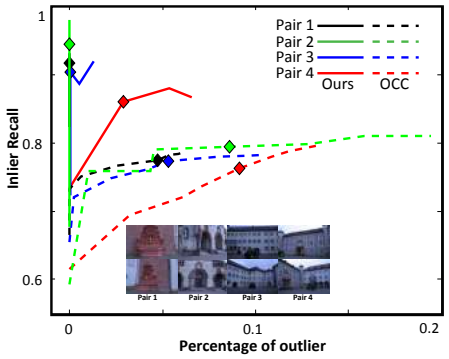


Fig. 5. Recall and % outliers between image pairs, with varying algorithm parameters (diamonds indicate default parameters). Occ [25] is the baseline. The recall value where the curves first intersect outliers $= 0$ represents the maximum number of inliers that can be retained if no outliers are tolerated. Apart from highly repetitive scenes (blue and red curves), our curves are vertical lines on the outliers $= 0$ axis. Our default parameters provide over 90% recall with no outliers for many scenes.

discrimination is the percentage of of inliers sacrificed to eliminate outliers. We perform this analysis on wide baseline image pairs chosen from Strecha's dataset [32] . Results are shown in Fig. 5. Pair $1, 2$ are the first and last images of [33]'s "cannonical" sampling of Strecha's Dataset. Pair $3, 4$ are chosen as difficult cases for our algorithms as they involve both wide baselines and strong image self similarity. Observe that local fitting of *Occ* trades a large percentage of inliers to remove the final few outliers. In contrast, for the "canonical" Pair $1, 2$, our lines are vertical on the outlier $= 0$ axis, indicating little need to trade inlier recall for outlier removal. For repetitive scenes like Pairs $3, 4$, there is indeed a trade-off but performance is still substantially better.

[3] Ideally, we would evaluate directly on matches without thresholding. However, the noise level is too high for all tested algorithms.

We also use the entire viewpoint change section of Hienly's dataset [33] for evaluation. This includes many images with extreme illumination and viewpoint changes. For each set, we use the first image as a base and match the rest to it. We compared our results against epipolar RANSAC [34] and MLESAC [35], implemented by [36]. We also evaluated piecewise homographic RANSAC (RCM homo) [18] and Occ [25], run at their default parameters. Results are shown below

	Ours	epipolar RANSAC [34,35]	RCM homo [18]	Occ [25]
Images with outliers	**5/35**	18/35, 24/35	6/35	25/35
Precision (1-% outliers)	0.987	0.947, 0.963	0.966	0.983
Inlier recall	0.928	0.873, 0.886	0.561	0.733
F-Measure	**0.957**	0.908, 0.923	0.709	0.839

Our bilateral motion models provide good recall and precision compared to other outlier rejection methods. Our recall is even comparable with epipolar RANSAC which is a valid motion model for test images. Apart from precision and recall, the flexibility of bilateral models allow us to tune parameters to a level where recall remains high despite having zeros outliers for many image pairs. As discussed in Fig. 5, for other methods, removing the last few outliers often involves discarding a large percentage of inliers. It is important to bear in mind that these methods are not mutually exclusive. As we do not enforce the epipolar constraint, RANSAC can still be run as a post-processor on our correspondences.

4.2 Independent Motion

We evaluate independent motion on images from AdelideRMF [37]. Every image pair contains large independent motions. Ground truth inlier-outlier labeling is provided for noisy pre-computed correspondences. As the small set of background matches are automatically labeled outliers, the dataset systematically lowers precision statistics of some scenes. Thus, the number of images with no outliers is not directly meaningful. Images used are the same as those in [18] and results for other RANSAC algorithms on the same data can be found in [18]. As feature orientation is not given, we modified our algorithm to use only spatial location information. Fig. 6 shows our performance improvement over multi-fundamental (RCM fund), multi-homographic (RCM homo) RANSAC [18] and Occ.

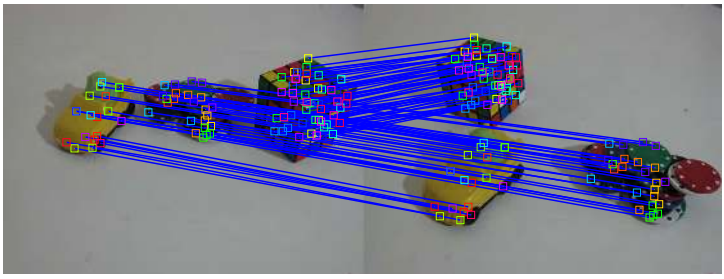


Fig. 6. Example of our performance on independent motion data [37]. Statistics for inlier recall and precision are given below.

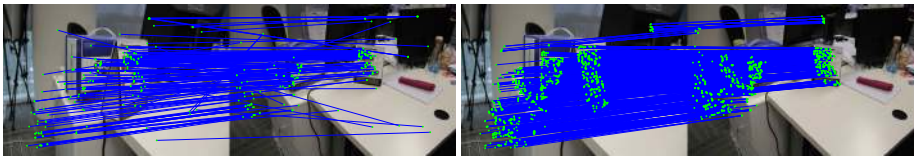
	Ours Precision	Ours Recall	RCM fund Precision	RCM fund Recall	RCM homo Precision	RCM homo Recall	Occ Precision	Occ Recall
dinobooks	0.8062	0.8927	0.7519	0.9902	1.0000	0.1659	0.8672	0.5415
toycubecar	1.0000	0.9063	0.8880	0.8672	1.0000	0.7734	0.9878	0.6328
cubebreadtoychips	0.9712	0.9874	0.9271	0.9582	1.0000	0.3264	0.9943	0.7238
carchipscube	1.0000	0.8762	0.9196	0.9810	1.0000	0.8000	1.0000	0.7429
breadtoycar	0.9892	0.8364	0.8560	0.9727	1.0000	0.5364	1.0000	0.6909
breadcubechips	1.0000	0.8591	0.9013	0.9195	1.0000	0.6443	1.0000	0.6779
breadcartoychips	1.0000	0.9935	0.8284	0.7161	1.0000	0.6516	1.0000	0.5548
biscuitbookbox	1.0000	0.9444	0.8710	1.0000	1.0000	0.6790	0.9926	0.8272
average	0.9708	0.9120	0.8679	0.9256	1.0000	0.5721	0.9802	0.6740
F-Measure	**0.94**		0.90		0.73		0.80	

Limitations Our algorithm assumes piecewise smooth motion and treats small sets of independently moving matches as outliers. This makes it unsuited for tasks like correspondence on pedestrian scenes, as the algorithm will only focus on the background. In extreme scenarios (large viewpoint change and lighting changes), scenes with strong self similarity can cause ambiguities unresolvable by global coherence. For moderate motions, matches are sufficiently well distributed for global coherence to handle self similar images. Visual examples are given in the supplementary material.

5 Applications

As a general motion model, bilateral functions can extend to many correspondence related problems. This section discusses examples such as expanding a correspondence set, drift free multi-image correspondence and template image search.

(a1) A-SIFT: **t**=0.66 (Default) (a2) Ours with $8.9\times$ more matches



(b1) A-SIFT: **t**=0.66 (Default) (b2) Ours with $5.8\times$ more matches

Fig. 7. Top: two views of half-dome, taken a few miles apart, causing parallax in the foreground. Bottom: an office scene with re-arranged stationary. Observe that our algorithm provides many more matches and fewer outliers than standard A-SIFT.

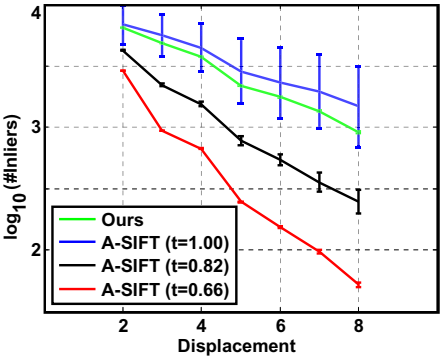


Fig. 8. Inlier numbers (log scale) versus increasing displacement for the *herzjesu* sequence [32]. Vertical bars represents the $-log_{10}$(fraction of inliers). Thresholding varies from 1 (no thresholding) to 0.66 (typical SIFT threshold). A-SIFT controls outlier numbers, at the cost of rejecting many inliers. Our bilateral models retain most inliers, while having no outliers (short vertical bars).

5.1 Additional Correspondence

Applications like Structure from Motion rely on matching algorithms to deliver many correspondences while controlling outlier numbers. Typically, outliers are controlled by a ratio test of nearest neighbor matches. A match is accepted, only if its descriptor matching score (zero is best) is below a certain fraction of the second best match. A typical ratio threshold is $\mathbf{t} = 1/1.5 = 0.66$ [38]. This threshold ensures that as baselines increase, the number of outliers remain restricted to a small handful but at the price of removing an enormous percentage of the inliers (often over 90%) from wide baseline image pairs. A typical SIFT threshold is shown by the red, $\mathbf{t} = 0.66$ curve in Fig. 8, while the maximum number of inliers is shown by the blue $\mathbf{t} = 1$ curve.

Our bilateral models can procure many additional correspondence as seen in Fig. 7). These results rely on two properties of bilateral functions. Firstly, as shown in Sec. 4.1, the functions can be run on SIFT thresholds that are much weaker than usual. This produces substantially more, better distributed potential inlier matches. The accompanying explosion of outliers can be controlled as shown in Sec. 4.1. This makes the correspondences usable. Secondly, once computed, the bilateral model can validate new matching hypothesis, using the steps in Sec. 3.4. These hypothesis can be obtained by setting $\mathbf{t} = 1$. This produces the green curve of Fig. 8.

We applied this algorithm to the 35 image pairs of the Heinly dataset of Sec. 4.1. We incurred no new outlier images compared to our basic outlier rejection technique in Sec. 4.1 (number remained at 5). However, we average of 225% more matches than our basic algorithm. Overall, we have 470% more matches than standard 0.66 A-SIFT. This occurs despite the high noise level at $\mathbf{t} = 1$, with image pairs having an average of only 25% inliers. Note that the averages mask extreme cases. In one example, our model successfully filtered an inlier ratio of below 5%. It had 4100% more matches than standard A-SIFT with no outliers. More modest improvements are recorded when the baseline is narrow.

If correspondence numbers are important, quasi-dense NRDC [30] techniques offer a viable alternative. While NRDC provides many more matches, its average correspondence error of 5.94 pixels is significantly higher than our 3.16, and it had $17/35$ images with outliers, compared to our $5/35$. However, the bilateral model's primary advantage lies in a graceful handling of difficult scenes. Fig. 9 shows NRDC's error distribution has a heavy tail. This is due to large sections of erroneous correspondence on difficult

scenes. Ideally, NRDC should be fused with our bilateral functions, however, this is a research direction we have not yet explored.

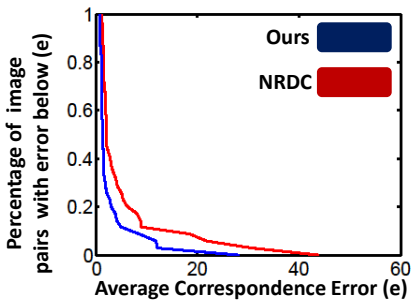
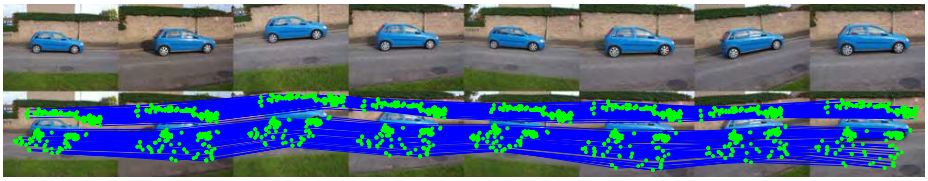


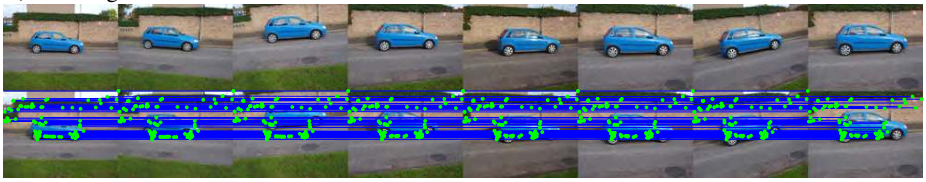
Fig. 9. Tail distribution of average correspondence error for test image pairs. In difficult cases, NRDC gives extremely high average errors (in one case reaching 43 pixels). In contrast, bilateral models have stabler error, with average error exceeding 15 pixels only once.

5.2 Drift-Free Multi-Image Correspondence

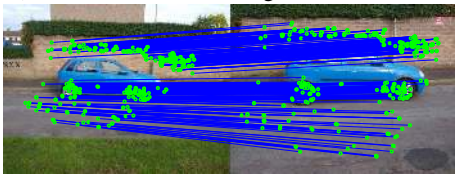
Multi-image correspondence is a perennial computer vision problem. Many algorithms like factorization [39], tri-focal tensor [40] and camera calibration [41] require correspondence across multiple wide-baseline images. However, feature matchers seldom reliably match the same feature across multiple frames, while feature trackers are prone



a) Our long tracks



b) KLT SIFT [42] matching



c) Zoom in on our first and last view

d) Zoom in on KLT's first and last view

Fig. 10. Matching across the car sequence. Observe that KLT-SIFT tracking drifts. A clear example can be seen at the car's front wheel.

to drift. Bilateral motion models can solve this problem as they can find a specified feature's matching location in another image. This is achieved by proposing potential matching locations which are verified using local information. By matching all images to a base frame, the formulation avoids drift. In fact, to save computational time, we automatically choose the widest baseline (quickly detected by using our algorithm to match low resolution images) pair to begin correspondence. Examples are shown in Fig. 10. Implementation details and more results are in the supplementary.

5.3 Needle in the Haystack

When multiple agents operate collaboratively, they need to identify locations of interest to each other. This becomes a needle-in-the-haystack search problem which involves locating a template within a much larger image. Bilateral matching is especially adept at this problem as it produces many matches and few outliers. This creates a large response gap between image pairs from similar and different locations, an ideal situation when comparing sub-images. In practice, we decompose the large image into multiple sub-images. These are ranked based on similarity to the template using gist [43]. Bilateral matching is applied on the top 10 candidates to perform the final selection and matching. Results are shown in Fig. 11. Details and alternative solutions are discussed in the supplementary.

(a) Input Images (b) Matching

Fig. 11. Examples of our algorithm localizing a template in a large image. Note that the template was taken at street level while the target image is from an overlooking roof.

6 Conclusion

We proposed a principled solution for modeling of general motion from noisy, scattered features matches. This allows reliable recovery of large numbers of inlier matches without reliance on situation specific models or RANSAC. Our formulation extends naturally to associated correspondence tasks like drift free multi-image correspondence and template localization. Applying the current formulation for dense correspondence estimation [44] is an interesting future direction.

References

1. Morel, J., Yu, G.: Asift: A new framework for fully affine invariant image comparison. SIAM Journal on Imaging Sciences 2(2), 438–469 (2009)
2. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. IJCV 60(2), 91–110 (2004)

3. Bay, H., Tuytelaars, T., Van Gool, L.: SURF: Speeded up robust features. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006, Part I. LNCS, vol. 3951, pp. 404–417. Springer, Heidelberg (2006)
4. Fischler, M.A., Bolles, R.C.: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Comm. of the ACM 24, 381–395 (1981)
5. Raguram, R., Frahm, J.-M., Pollefeys, M.: A comparative analysis of ransac techniques leading to adaptive real-time random sample consensus. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part II. LNCS, vol. 5303, pp. 500–513. Springer, Heidelberg (2008)
6. Longuet-Higgins, H.C.: A computer algorithm for reconstructing a scene from two projections. Nature, 133–135 (1981)
7. Brown, M., Lowe, D.: Automatic panoramic image stitching using invariant features. IJCV 1(74), 59–73 (2007)
8. Serradell, E., Özuysal, M., Lepetit, V., Fua, P., Moreno-Noguer, F.: Combining geometric and appearance priors for robust homography estimation. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part III. LNCS, vol. 6313, pp. 58–72. Springer, Heidelberg (2010)
9. Sprengel, R., Rohr, K., Stiehl, H.S.: Thin-plate spline approximation for image registration. In: Proc. of Engineering in Medicine and Biology Society (1996)
10. Lin, W.Y., Cheng, M.M., Zheng, S., Lu, J., Crook, N.: Robust non-parametric data fitting for correspondence modeling. In: IEEE ICCV (2013)
11. Yuille, A.L., Grywacz, N.M.: The motion coherence theory. In: IEEE ICCV (1988)
12. Black, M.J., Anandan, P.: The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. Computer Vision and Image Understanding (1996)
13. Ye, M., Haralick, R.M., Shapiro, L.G.: Estimating piecewise-smooth optical flow with global matching and graduated optimization. PAMI (2003)
14. Sinha, S.N., Steedly, D., Szeliski, R.: Piecewise planar stereo for image-based rendering. In: ICCV (2009)
15. Myronenko, A., Song, X., Carreira-Perpinan, M.: Non-rigid point set registration: Coherent point drift. In: NIPS (2007)
16. Lin, W.Y., Liu, S., Matsushita, Y., Ng, T.T., Cheong, L.F.: Smoothly varying affine stitching. In: IEEE CVPR (2011)
17. Raguram, R., Frahm, J.M.: Recon: Scale-adaptive robust estimation via residual consensus. In: ICCV (2011)
18. Pham, T.T., Chin, T.J., Yu, J., Suter, D.: The random cluster model for robust geometric fitting. In: CVPR (2012)
19. Weinzaepfel, P., Revaud, J., Harchaoui, Z., Schmid, C.: Deepflow: Large displacement optical flow with deep matching. In: ICCV (2013)
20. Brox, T., Malik, J.: Large displacement optical flow: Descriptor matching in variational motion estimation. IEEE TPAMI (2010)
21. Horn, B., Schunck, B.: Determining optical flow. Artificial Intelligence (1981)
22. Lucas, B., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: Proceedings of Imaging Understanding Workshop (1981)
23. Torresani, L., Kolmogorov, V., Rother, C.: Feature correspondence via graph matching: Models and global optimization. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part II. LNCS, vol. 5303, pp. 596–609. Springer, Heidelberg (2008)
24. Garg, R., Roussos, A., Agapito, L.: Dense variational reconstruction of non-rigid surfaces from monocular video. In: CVPR (2013)
25. Pizarro, D., Bartoli, A.: Feature-based deformable surface detection with self-occlusion. IJCV (2012)

26. Tomasi, C., Manduch, R.: Bilateral filtering for gray and color images. In: IEEE ICCV (1998)
27. Xiao, J., Cheng, H., Sawhney, H.S., Rao, C., Isnardi, M.: Bilateral filtering-based optical flow estimation with occlusion detection. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006, Part I. LNCS, vol. 3951, pp. 211–224. Springer, Heidelberg (2006)
28. Barnes, C., Shechtman, E., Goldman, D.B., Finkelstein, A.: The generalized PatchMatch correspondence algorithm. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part III. LNCS, vol. 6313, pp. 29–43. Springer, Heidelberg (2010)
29. Lhuiller, M., Quan, L.: A quasi-dense approach to surface reconstruction from uncalibrated images. PAMI (2005)
30. HaCohen, Y., Shechtman, E., Goldman, D.B., Lischinski, D.: Non-rigid dense correspondence with applications for image enhancement. ACM TOG (2011)
31. Hartley, R.I.: In defense of the eight-point algorithm. IEEE TPAMI 19(6), 580–593 (1997)
32. Strecha, C., von Hansen, W., Gool, L.V., Fua, P., Thoennessen, U.: On benchmarking camera calibration and multi-view stereo for high resolution imagery. In: CVPR (2008)
33. Heinly, J., Dunn, E., Frahm, J.-M.: Comparative evaluation of binary features. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part II. LNCS, vol. 7573, pp. 759–773. Springer, Heidelberg (2012)
34. Kovesi, P.D.: MATLAB and Octave functions for computer vision and image processing, `http://www.csse.uwa.edu.au/~pk/research/matlabfns/`
35. Torr, P.H.S., Zisserman, A.: Mlesac: A new robust estimator with application to estimating image geometry. Computer Vision and Image Understanding (2010)
36. Konouchine, A., Gaganov, V., Veznevets, V.: A new maximum likelihood robust estimator. Computer Vision and Image Understanding (2005)
37. Wong, H.S., Chin, T.J., Yu, J., Suter, D.: Dynamic and hierarchical multi-structure geometric model fitting. In: ICCV (2011)
38. Vedaldi, A., Fulkerson, B.: VLFeat: An open and portable library of computer vision algorithms (2008)
39. Tomasi, C., Kanade, T.: Shape and motion from image streams under orthography: a factorization method. IJCV (1992)
40. Torr, P., Zisserman, A.: Robust parameterization and computation of the trifocal tensor. Image and Vision Computing 15, 591–605 (1997)
41. Agrawal, M.: Practical camera auto calibration using semidefinite programming. In: WMVC (2007)
42. Sharma, A.: (2004), `http://www.cs.cmu.edu/abhishek/softwares.html`
43. Oliva, A., Torralba, A.: Modeling the shape of the scene: A holistic representation of the spatial envelope. IJCV (2001)
44. Yang, H., Lin, W.Y., Lu, J.: Daisy filter flow: A generalized discrete approach to dense correspondences. In: CVPR (2014)