RepMatch: Robust Feature Matching and Pose for Reconstructing Modern Cities

Wen-Yan Lin*, Siying Liu[#o], Nianjuan Jiang*,
Minh. N. Do[#] , Ping Tan[t], Jiangbo Lu* ⋆

*Advanced Digital Sciences Center, [o]Institute of Infocomm Research
[#]University of Illinois Urbana-Champagne, [t]Simon Fraser University

Abstract. A perennial problem in recovering 3-D models from images is repeated structures common in modern cities. The problem can be traced to the feature matcher which needs to match less distinctive features (permitting wide-baselines and avoiding broken sequences), while simultaneously avoiding incorrect matching of ambiguous repeated features. To meet this need, we develop *RepMatch*, an epipolar guided (assumes predominately camera motion) feature matcher that accommodates both wide-baselines and repeated structures. *RepMatch* is based on using *RANSAC* to guide the training of match consistency curves for differentiating true and false matches. By considering the set of all nearest-neighbor matches, *RepMatch* can procure very large numbers of matches over wide baselines. This in turn lends stability to pose estimation. *RepMatch*'s performance compares favorably on standard datasets and enables more complete reconstructions of modern architectures.

Keywords: structure from motion, correspondence, RANSAC

1 Introduction

Structure-from-Motion or SfM is the recovery of 3-D structure from image sets. Over the years, SfM has made remarkable progress. Current technology can create impressively large scale reconstructions, a signature achievement being the reconstruction of ancient Rome by leveraging the abundance of Internet images [1]. However, SfM systems have difficulty reconstructing modern buildings from small, user-captured datasets.

The problem stems from SfM's dependence on feature matching from which camera position (pose) and 3-D structure are inferred. Feature matching needs to procure large numbers of wide-baseline matches to prevent image sequences from splintering. Yet, it must also be robust to repetitive structures. Unfortunately, modern urban environments contain both challenges in abundance. Trees and other occluders limit available view-points, necessitating matching widely separated images. At the same time, mass production makes repeated structures ubiquitous in modern cities (e.g. rows of windows).

⋆ This study is supported by the HCCS grant at ADSC from Singapore's Agency for Science, Technology and Research. Corresponding authors: W-Y. Lin, J. Lu

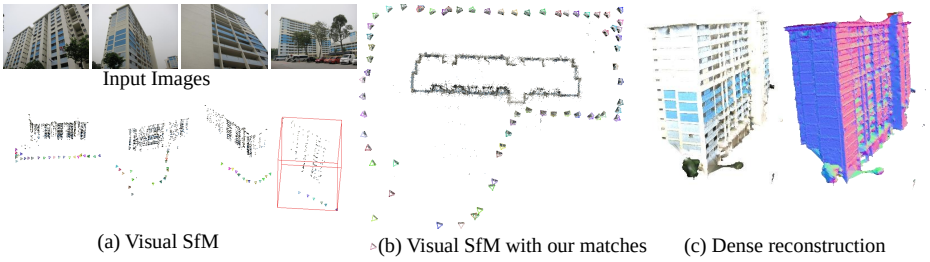


Fig. 1. 3-D reconstruction on modern buildings. (a) Screen shot of Visual SfM [2], a classic 3-D reconstruction system which splinters the sequence into 4 segments. (b) The same reconstruction system with our matches forms a complete loop. (c) The pose estimated in (b) is sufficiently accurate for high quality dense reconstruction.

As matching ambiguous features inevitably results in some errors, vision researchers typically use *RANSAC* [3] for outlier rejection. By using multiple pose estimates from minimal sets of 5 to 8 matches, *RANSAC* is very effective at getting a reasonable pose estimate. However, there are practical limits on the number of wrong matches *RANSAC* can accommodate. As such, *RANSAC* is seldom applied to the set of all nearest-neighbor matches but is itself dependent on preemptive outlier removal. Typically this takes the form of a ratio test [4], which unfortunately discards a large fraction of the true matches [5]. While this framework has brought SfM much success, the quality of feature matches in modern environments is still insufficient. This manifests itself as fragmented reconstructions, with linkages around corners being especially brittle. An illustration is shown in Fig. 1(a). Thus, we propose *RepMatch*, an epipolar guided feature matcher that accommodates wide-baselines and repeated structures.

RepMatch is inspired by the highly successful guided matching framework, where an initial estimate or assumption guides the discovery of more matches. With the right guidance term, such formulations have proven remarkably stable. This is illustrated by the success of SLAM [6, 7], which can reliably propagate pose given some known 3-D points. However, finding a generic guidance term applicable to general two-view pose estimation has proven challenging. Planar based guidance terms have been explored in [8, 9] but are scene specific and often incur significant formulation complexity to define number of planes or demarcate planar boundaries. There are also works [10, 11] which focus on epipolar geometry guidance. While this can give very accurate solutions [10], it is hard to determine the epipolar geometry without good correspondence, which in turn makes performance unpredictable. This paper shows how a core-set of guidance matches can be reliably obtained even under challenging circumstances and explains how they can guide the finding of more matches. The resultant *RepMatch* algorithm is an epipolar guided matcher which, while using pose as a cue, postpones selecting a correct pose to a very late stage (in fact a choice need never be made). This allows *RepMatch* to reliably validate the very large but noisy set of all matches which contains many previously discarded true matches [5].

RepMatch couples *BF* [5] and *RANSAC* outlier rejection schemes. *BF* computes a global match consistency function from very noisy matches. These are subsequently used to separate true and false matches. While usually accurate, *BF* is vulnerable to repeated structures which can induce large sections of consistently wrong matches. However, we observe that repetitive structures often contain micro-textures which make it possible to obtain a (often heavily shrunken) core-set of reliable matches through very strict *BF* parameters. Using *RANSAC* we can procure more local matches that are geometrically consistent with the core-set. These are used to train local *BF* curves with embedded epipolar constraints that verify geometrically consistent matches in the surrounding areas. The resultant *RepMatch* framework breaks the pose and correspondance problem into a sequence of robust steps, giving overall system stability on both wide-baselines and repetive strcutures. On standard datasets, *RepMatch* tolerates up to 45° out-of-plane rotation for all scenes and has over 90° stability on some scenes. On less controlled data, *RepMatch* adds a significant stability margin to existing SfM systems, enabling complete reconstruction of modern buildings from street-level images, something difficult with previous techniques.

1.1 Related Works

RepMatch builds on *BF* [5], a wide-baseline matcher which achieves high precision and recall on challenging scenes. While *BF* recovers many previously discarded true matches, it is vulnerable to repeated structures which can induce large sections of consistently wrong matches. Our *RepMatch* framework retains *BF*'s aggressive match retrieval while avoiding its vulnerability to repeated structures. This results in an effective, general purpose wide-baseline feature matcher.

RepMatch is closely related to *RANSAC* as they are both outlier removal schemes. Many of the concepts on guidance and grouping used in this paper have also been explored within the *RANSAC* framework. However, *RANSAC* has a vulnerability as ill-conditioning of pose estimates [12] makes the minimal set estimation unstable. On the other hand, every match added to the minimal set exponentially increases the likelihood it contains an outlier [13]. Thus despite many refinements such as new sampling schemes [14], local groupings/ranking [15–17] or local pose refinements [18, 10], there are practical limits on the number of outliers *RANSAC* can accommodate. In addition, because epipolar geometry is a point-to-line constraint, even if the ground-truth pose is attained, *RANSAC* often leaves some outliers which are coincident on the epipolar line. While they may not affect the two view pose estimate, such outliers are detrimental to the overall SfM system's stability. *RepMatch* addresses this problem by adding a *BF* estimator to reduce the reliance on potentially erroneous epipolar geometry while also shielding its *RANSAC* model from encountering too many outliers. This allows us to extract the true matches from the very noisy set of all nearest neighbor matches shown in Fig. 4.

Apart from *RepMatch* and *BF*, there are other preemptive outlier removal works [19–23]. Of these, the evaluation in [24] suggests *BF* provides some of the best trade-offs in computational time and match quality. In addition, *BF* has a

guaranteed global minimum. Hence, we build our *RepMatch* framework on *BF*, though other match decision techniques may also be applicable.

There have also been many specialized solutions for repetitive structures, wide-baselines or urban scenes [25–27, 8, 28]. These techniques utilize planar constraints [8] or leverage repetitions [26] to enable high quality pose and correspondence. These solutions will likely provide superior performance on specific scenes but lack generality.

Finally, we wish to acknowledge that *RepMatch* and the results achieved in this paper benefit from many years of research in supporting technologies like feature descriptor design [4, 29–31], bundle adjustment [32–35], dense reconstruction [36, 37], *RANSAC* [3, 16, 17, 38–40], geometric reasoning [41, 42] and motion coherence [43–45]. Improvements in these fields will likely benefit *RepMatch* which we hope will in turn benefit these fields.

2 Background

We approach the feature matching problem as one of reliably partitioning the set of all matching hypotheses into true and false sets. As this is not a main-stream approach, we provide some background to aid understanding.

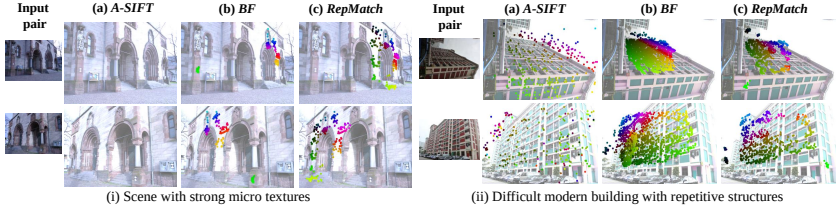


Fig. 2. Comparing three partitioning techniques with the same input matches. Match pairs are shaded with the same color across two views. Results shown are post-*RANSAC*. (a) *A-SIFT*'s ratio-test is very unstable. (b) *BF* retains many previously discarded true matches but incurs many wrong matches on modern buildings. Wrong matches appear as color inconsistencies. (c) Our proposed *RepMatch*.

Using the same input match hypotheses, Fig. 2 shows the impact of different match partitioning schemes. (a) *A-SIFT* [46] uses a ratio test that requires the best match score to be at least 0.6× better than the second best match. (b) *BF* [5] relaxes the 0.6 threshold to attain an initial set of noisy match hypotheses from which it trains a partition function based on match consistency (a joint measure of three attributes, match density, smoothness and spatial coverage). Observe that in both (a) and (b), prior methods either retain too few matches or create too many wrong matches on repetitive structures. (c) Our *RepMatch* framework which integrates *BF* with *RANSAC* significantly enhances match stability. As our formulation makes heavy use of *BF* to estimate match consistency, we will elaborate on both *BF* and match consistency.

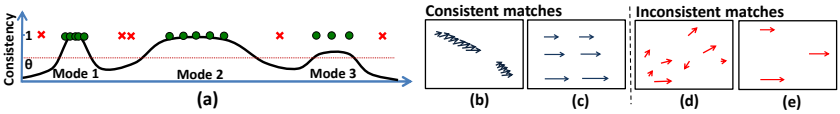


Fig. 3. Match consistency. (a) 1-D illustration of Eq. (2)'s consistency curve, $f(\mathbf{m})$. Consistency measures two basic elements, density and spatial extent, shown in Modes 1 and 2. Note: by incorporating motion into $f(\mathbf{m})$'s domain, density encapsulates motion smoothness. (b,c) show consistent matches. (b) Match is smooth and dense. (c) Match is smooth. While sparse, it has wide spatial extent. (d,e) show inconsistent matches. (d) match is not smooth. (e) Match is smooth but spatial coverage is limited.

2.1 Bilateral Functions and Match Consistency

In *BF* [5], matches are represented on a $D = 8$ dimensional bilateral domain. Each match takes the form $\mathbf{m}_i = [\mathbf{x}_i; \mathbf{v}_i; \mathbf{o}_i]$. Here, $\mathbf{x}_i = \begin{bmatrix} x_i; y_i \end{bmatrix}$ and $\mathbf{v}_i = \begin{bmatrix} u_i; v_i \end{bmatrix}$ are two-dimensional vectors representing a feature point's coordinate (in the first image) and its corresponding motion vector, respectively; $\mathbf{o}_i$ is a 4×1 vector representing the relative affine feature orientation (obtained from the feature's scale and rotation parameters [5]). *BF* learns a match consistency curve (termed likelihood in [5]) from N training matches, $\{\mathbf{m}_i\}$, by minimizing the convex function:

$$\arg\min_{\mathbf{w}} \sum_{i=1}^{N} C(1 - f(\mathbf{m}_i)) + \lambda \mathbf{w}^T G \mathbf{w}, \tag{1}$$

where $C(.)$ is a Huber function, $\mathbf{w}$ is a vector of N unknowns and G is an $N \times N$ matrix with $G(i,j) = \exp(-\|\mathbf{m}_i - \mathbf{m}_j\|^2/\sigma)$. $f(.)$ is the consistency function defined on an 8 dimensional bilateral domain. $f(.)$ is parameterized by $\mathbf{w}$ in (1) and N radial basis functions centered on the training matches:

$$f(\mathbf{m}) = \sum_{i=1}^{N} \mathbf{w}(i) \exp^{-\frac{\|\mathbf{m}-\mathbf{m}_i\|^2}{\sigma}}. \tag{2}$$

Minimizing Eq. (1) with respect to $\mathbf{w}$ (and hence $f(.)$) provides match consistency curve $f(\mathbf{m})$. This allows a set of matches $\{\mathbf{m}_j\}$ to be partitioned into two subsets, $\mathcal{T}$ (true) and $\mathcal{F}$ (false), via thresholding:

$$\mathbf{m}_j \in \begin{cases} \mathcal{T}, \text{ if } f(\mathbf{m}_j) > \theta \\ \mathcal{F}, \text{ otherwise} \end{cases} \tag{3}$$

When minimizing Eq. (1), the local data term draws $f(.)$ to 1 while the regularization term $\mathbf{w}^T G \mathbf{w}$ pulls $f(.)$ to zero and imposes a global smoothness penalty [5]. This is illustrated in Fig. 3(a). The resultant $f(.)$ can be understood as a continuous consistency estimate, where consistency is a joint measure of two elements, (I) Density: If a region has high point density, the data term justifies a sharp spike even if it is not smooth, as shown in Mode 1 of Fig. 3(a); (II) Spatial extent: Alternatively, a large region with sparsely distributed points is

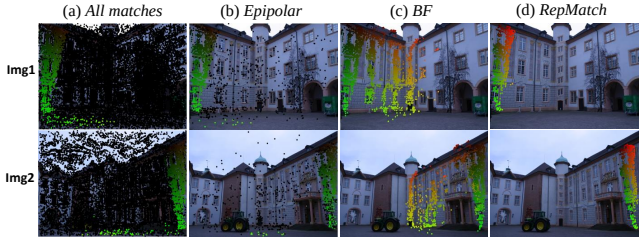


Fig. 4. (a) The set of all matches. (b) Thresholding with ground-truth epipolar geometry still leaves some wrong matches. (c) *BF*'s match consistency based thresholding is unstable as repetitive structures induce consistently wrong matches. (d) *RepMatch* handles such repetitive structures well.

also consistent as a well-rounded hump over a large extent incurs low smoothness penalty, as illustrated by Mode 2. Scattered points are considered inconsistent as their pull cannot overcome the smoothness penalty acting on them (see Mode 3 in Fig. 3(a)). As the bilateral domain encapsulates both spatial and velocity components, consistency on the bilateral domain encapsulates match density and motion smoothness. Match consistency is illustrated in Fig. 3 (b-e).

Explained as match consistency, *BF*'s problem with repetitive structures is clear. Repetitive structures can induce large sections of consistent but wrong matches, creating large falsely consistent match patches shown in Fig. 4(c). Epipolar constraints can also remove many false matches as shown in Fig. 4(b), but it too leaves large numbers of wrong correspondences. This leads to our *RepMatch* framework for integrating epipolar and match consistency curves.

3 RepMatch

RepMatch is based around two innovations. First, *RepMatch* introduces a means to reliably obtain a core-set of matches even for challenging image pairs with significant repetitive structures. Second, *RepMatch* introduces a method to robustly expand this core-set by integrating *BF* with epipolar geometry. Thus, *RepMatch* divides the pose and correspondence problem into three individually robust steps, to give a robust overall system. Fig. 5 gives a general overview, with a match consistency interpretation in Fig. 6.

3.1 Core-set Discovery

Core-set discovery is based on an observation. An image of a visually perfectly repetitive pattern can be matched error-free in the absence of motion (i.e. match the image to itself). This is due to micro gray-level differences captured by descriptors. Thus, we hypothesize that the repetitive structure matching problem is

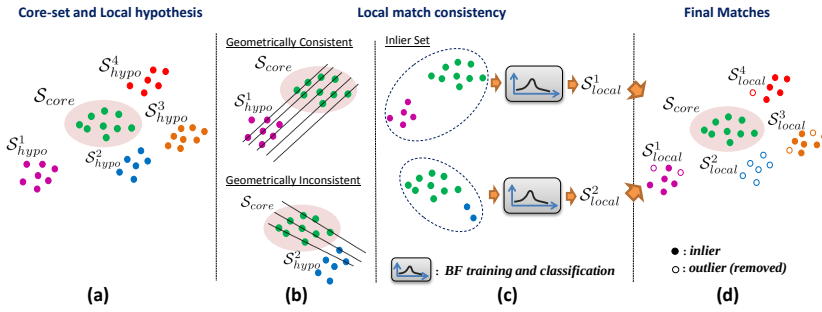


Fig. 5. Overview: *RepMatch* ensures stability by dividing the problem into three individually robust steps. (a) Core matches, S_{core} can be reliably recovered because of strict *BF* thresholds. (b) Geometric verification with epipolar lines (pose). Core matches may be quasi-degenerate and an incorrect pose estimate may discard true positives. Thus, geometric verification uses a *RANSAC* search for common geometry between core matches and each subset. This avoids discarding true positives but may retain some false positives. (c) Local *BF* curves are trained to remove the remaining false positives and discover more matches. (d) All verified matches are consolidated.

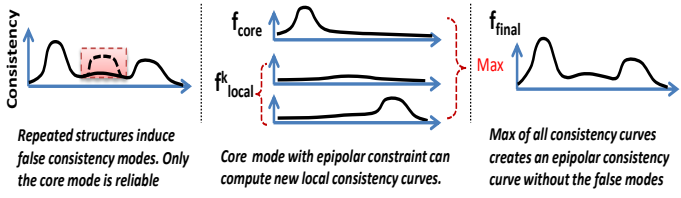


Fig. 6. *RepMatch* algorithm in Fig. 5 explained as match consistency curves.

not due to multiple identical descriptors but the result of subtle descriptor variations being overwhelmed by image noise (induced by motion or other sources). Due to the image's repetitive nature, many mismatches will appear "consistent", causing false modes in Fig. 6. However, even on repetitive scenes, a wrong match can be randomly assigned to many potential alternative positions, making it unlikely that false modes will be more consistent than the original true mode. The difference between true and false modes can be amplified by setting *BF* in Eq. (1) to a very high λ. On curves like Fig. 6 it suppresses weak modes, leaving only the strongest core-mode and its associated core-matches. These are remarkably resilient to noise as shown in Fig. 7.

3.2 Match Expansion Scheme

Once the core-set is discovered, it is theoretically possible to recover more match hypotheses from the epipolar geometry (pose) estimated from the core-set. However, pose estimation is notoriously vulnerable to degeneracies and an incorrect core-set pose may reject many true positives. Instead, core-set matches are merged with clusters of match hypotheses for joint geometric verification. This avoids rejecting true positives but will retain some false positives because of

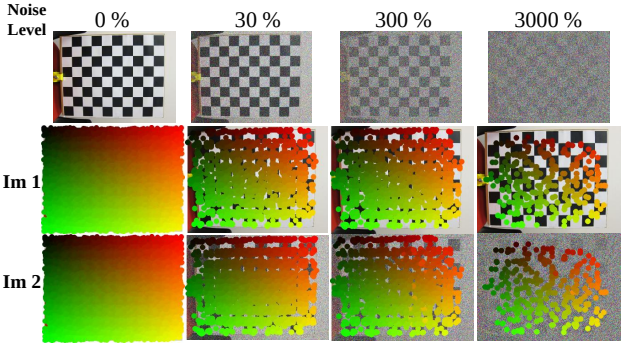


Fig. 7. Core-set recovery on a checker-board image. Image 1 is matched to a noisy version of itself with additive Gaussian noise. Noise variance is a percentage of image contrast. The smooth colors show the core-set estimation on repetitive structures is remarkably resilient to noise.

epipolar geomtry's weak point-to-line relationship. However, the remaining false positives are unlikely to be consistent and can be removed by a final *BF* match consistency step. The resultant framework is intrinsically robust as it avoids both *BF*'s vulnerability to false match consistencies and epipolar geometrie's vulnerability to ill conditioning and false positive rejection.

3.3 Algorithm

Training and classification operators: For later convenience, we first define training and classification operators.

BF training operator is denoted as:

$$f \leftarrow BF^t(\mathcal{S}^t, \Theta^t), \tag{4}$$

where $f(.)$ is the match consistency function defined in Eq. (2). It is learned by minimizing Eq. (1) with training matches $\mathcal{S}^t$ and *BF* parameters Θ^t. To maintain computational tractability, if the training set has more than 1000 matches, a random sample of 1000 are used for computation.

BF finds true matches in a match set $\mathcal{S}^c$, with the classification operator:

$$\mathcal{T} \leftarrow BF^c(f, \mathcal{S}^c, \Theta^c). \tag{5}$$

This partitions $\mathcal{S}^c$ into true and false sets through Eq. (3), with $\theta \in \Theta^c$ acting as classification parameters. Only true matches are returned.

Similar to *BF*, we consider *RANSAC* with parameters α^t as learning a classification function (camera pose) trained from a set of matches, $\mathcal{S}^t$. This is used to find true matches in set $\mathcal{S}^c = \{\mathbf{m}_j\}$. The respective training and classification operators are

$$pose \leftarrow RANSAC^t(\mathcal{S}^t, \alpha^t), \tag{6}$$

$$\mathcal{T} \leftarrow RANSAC^c(pose, \mathcal{S}^c, \alpha^c), \tag{7}$$

where Eq. (7) implements epipolar thresholding:

$$\mathbf{m}_j \in \begin{cases} \mathcal{T}, \text{ if distance from epipolar line} < \alpha^c \\ \mathcal{F}, \text{ otherwise} \end{cases} \tag{8}$$

Core-set, $\mathcal{S}_{core}, f_{core}(.)$: To find the core-set of matches, we threshold the set of all matches $\mathcal{A}$, with a ratio test using threshold 0.82 (this is much weaker than the standard 0.6 to ensure sufficient matches for training [5]) to form $\mathcal{A}_{0.82}$, a set of noisy match hypotheses. *BF* match consistency curves, $f_{core}(.)$, are trained from $\mathcal{A}_{0.82}$ using very strict match consistency parameters. The core-set $\mathcal{S}_{core}$, is defined as all matches consistent with $f_{core}(.)$:

$$f_{core} \leftarrow BF^t(\mathcal{A}_{0.82}, \Theta^t_{strict}), \quad \mathcal{S}_{core} \leftarrow BF^c(f_{core}, \mathcal{A}, \Theta^c) \tag{9}$$

The strict parameters Θ^t_{strict} have a large λ, making $\mathcal{S}_{core}$ remarkably resistant to repeated structures.

Local hypotheses, $\mathcal{S}^k_{hypo}$: A disadvantage of *BF* is that it sub-samples training sets for computation efficiency. This is good for core-set estimation but hurts fine matching. Thus we cluster $\mathcal{A}_{0.82}$ into $K = 20$ disjoint subset using K-means clustering (over-segmentation is fine)

$$\mathcal{A}_{0.82} = \{\mathcal{L}^1, \mathcal{L}^2, \ldots, \mathcal{L}^K\}$$

and compute a local hypothesis set through

$$f^k_{hypo} \leftarrow BF^t(\mathcal{L}^k, \Theta^t_{strict}), \quad \mathcal{S}^k_{hypo} \leftarrow BF^c(f^k_{hypo}, \mathcal{A}, \Theta^c). \tag{10}$$

Local match consistency, $\{f^k_{local}(.)\}$: We next leverage the core-set to robustly estimate local match consistency curves. Each $\mathcal{S}^k_{hypo}$ local hypothesis set is merged with the core-set $\mathcal{S}_{core}$ to form a mixed set $\mathcal{M}^k_{local}$. Core-set matches are forced to make up at least 80% of $\mathcal{M}^k_{local}$ (if there are insufficient core-set matches, they are artificially duplicated). *RANSAC* is performed on $\mathcal{M}^k_{local}$ and a pose hypothesis, $pose^k$ is computed

$$pose^k \leftarrow RANSAC^t(\mathcal{M}^k_{local}, \alpha^t), \tag{11}$$

This preponderance of core-set matches ensures that *RANSAC* need not handle extremely noisy data and prevents it from inadvertently fitting local ambiguities arising from repetitive structures.

Given $pose^k$, we find matches in the local matching sets, $\mathcal{S}^k_{hypo}$ that are geometrically consistent with the core set

$$\widehat{\mathcal{S}}^k_{hypo} \leftarrow RANSAC^c(pose^k, \mathcal{S}^k_{hypo}, \alpha^c). \tag{12}$$

These are used to train locally focused *BF* functions which take into account geometric consistency with the core-set.

$$f^k_{local} \leftarrow BF^t(\widehat{\mathcal{S}}^k_{hypo}, \Theta^t) \tag{13}$$

Wrong local match hypotheses derived from repetitive ambiguities will have many members removed by the epipolar constraint (see Fig. 5 (b)). Correct matches will pass the epipolar constraint and the training step in (13) will create a local match consistency curve, $f^k_{local}(.)$ (see Fig. 6) that describes them, allowing subsequent procurement of more similar matches.

Final output $pose, \widehat{\mathcal{S}}_{out}$**:** Taking the max value of $f_{core}(.)$ and all local $f^k_{local}(.)$ curves for each point in the bilateral domain gives

$$f_{final}(\mathbf{m}) = max(\{f_{core}(\mathbf{m}), f^1_{local}(\mathbf{m}), \ldots, f^K_{local}(\mathbf{m})\}).$$

As shown in Fig. 6, this gives an epipolar-consistency curve without the false match consistency modes of *BF*. However, this is impractical on a continuous domain. An implementation equivalent, is to define the final match consistency derived output, $\mathcal{S}_{out}$, as the union of all matches $\mathcal{A}$, which accord any of the match consistency functions f_{core} and $\{f^k_{local}\}$

$$\mathcal{S}_{out} = BF^c(f_{core}, \mathcal{A}, \Theta^c) \cup \bigcup_{k=1}^{K} BF^c(f^k_{local}, \mathcal{A}, \Theta^c) = \mathcal{S}_{core} \cup \bigcup_{k=1}^{K} \mathcal{S}^k_{local} \quad (14)$$

A final *RANSAC* step is performed on $\mathcal{S}_{out}$ to estimate *pose* and the set of matches geometrically consistent with it, $\widehat{\mathcal{S}}_{out}$

$$pose \leftarrow RANSAC^t(\mathcal{S}_{out}, \alpha^t), \quad \widehat{\mathcal{S}}_{out} \leftarrow RANSAC^c(pose, \mathcal{S}_{out}, \alpha^c). \quad (15)$$

Implementation: This paper uses a basic implementation of *RANSAC* [47] and *BF* (C++ re-implementation of [5] in [24]). Parameters used are detailed below. All feature coordinates are Hartley-normalized and motion vectors multiplied by 10. For core-set discovery, the training parameters $\Theta^t_{strict} = \{\lambda = 10, \sigma = 1, \epsilon = 0.1\}$, where λ and σ refer to the parameters in Eq. (1), and ϵ is the Huber function parameter in [5]. Similarly, in training local match consistency curves, $\Theta^t = \{\lambda = 1,\ \sigma = 1,\ \epsilon = 0.1\}$. For *BF* classification, $\Theta^c = \{\theta = 0.6\}$. In *RANSAC*, $\alpha_t = \alpha_c = 5$ pixels is the threshold for distance to epipolar lines. On a 4-core i7 machine, our mixed MATLAB, C++ implementation of *RepMatch* processes a a few hundred thousand feature matches in approximately 20 seconds. Note: when passing matches to a large-scale SfM systems, skip the final *RANSAC* in Eq. (15). Such systems have inbuilt *RANSAC* and pre-processing matches affects frame selection. Note: *BF* [5] includes a bilateral affine verification step for fine match decisions, which we retain. This verification can also be interpreted as match consistency.

4 Experiments

The experiments focus on two aspects: quantifying the performance and baseline gains of *RepMatch* vs previous algorithms in Sec. 4.1; and integration of *RepMatch* into an overall SfM sysetms in Sec. 4.2.

4.1 Quantitative Evaluation

For quantitative evaluation we use Strecha *et al.*'s dataset [48]. To study performance over a comprehensive range of baselines, we construct a test set by pairing all images with at least 30% overlap from all 4 sequences in the dataset,

giving a total of 619 pairs. To evaluate performance variations with baseline, we subdivide the set according to ground truth rotational baseline.[1]

As pose estimators often give wildly incorrect solutions (or crash) when they fail, average errors are less meaningful. To circumvent this, we propose to measure Success Percentage (SP):

$$\text{Success Percentage}(x) = \frac{\#\text{ pairs with rotation (translation) error} \leq x^\circ}{\text{total number of pairs}} \tag{16}$$

with error in °. By plotting SP against x, we obtain a non-decreasing curve which gives the percentage of two view pose estimates lying below error threshold x°. The success percentage at 1° is an area of interest, as it is a commonly accepted bound for a "good" pose estimate. Finally, as the rotation and translational errors often follow identical trends, in less important cases, we plot SP against Pose Error, a consolidated statistic formed by taking the max of rotational and translational errors.

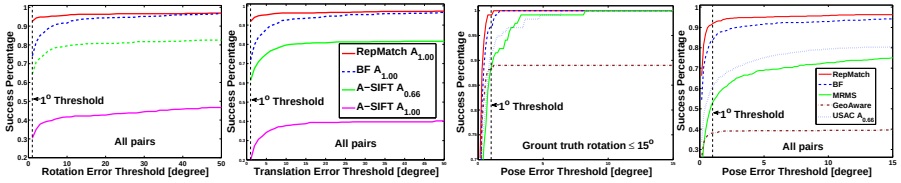
Comparisons: We begin by establishing performance baselines for a "typical" pose estimator with *RANSAC* and non-linear re-projection error refinement step[2]. To represent *RANSAC*, we choose *USAC 1.0* [13], a *RANSAC* variant which integrates many core *RANSAC* innovations. As it has a *PROSAC* component to take in A-SIFT match scores, *USAC* can potentially be applied to the set of all nearest neighbor matches $\mathcal{A}_{1.00}$.

Fig. 8(a) compares *RepMatch* against *USAC* with bundle adjustment. *USAC* was provided with feature matches filtered by three different preemptive outlier removal schemes. The first is $\mathcal{A}_{1.00}$ with no outlier removal. As explained in the introduction, this gives expectedly low scores with a SP of only 20% at 1°. Using a typical ratio-test $\mathcal{A}_{0.66}$ significantly improves pose estimates. Finally applying *BF* [5] match consistency curve to $\mathcal{A}_{1.00}$ and running *USAC* improves results still further. This demonstrates that preemptive outlier removal can significantly impact *RANSAC* performance and explains *RepMatch*'s excellent performance. Fig. 8(b) compares *RepMatch* against other guided matching pose estimators. MRMS [10] has very high pose estimation accuracy while GeoAware [11] is designed to handle repetitive structures. As theses algorithms are tightly coupled to their feature descriptors, we perform system-to-system comparisons. At narrow baselines, with ground truth rotations less than 15°, MRMS and GeoAware have an advantage as they use SIFT [4] rather than the more ambiguous A-SIFT descriptors [46]. Despite this, *RepMatch*'s performance is easily comparable to them. The advantage of *RepMatch*'s full system is clearly evident on the set of all pairs. While the use of different descriptors means comparisons are not strictly fair, it suggests that *RepMatch* has good narrow and wide baseline capability.

Component wise evaluation: Our *RepMatch* framework integrates both *RANSAC* and *BF* outlier removal schemes. In Fig. 10 we compare the performance of *RepMatch* against its individual components. Note that as *RepMatch*

[1] For fine nuances regarding experiment details and comparisons we encourage interested readers to peruse the supplementary material.

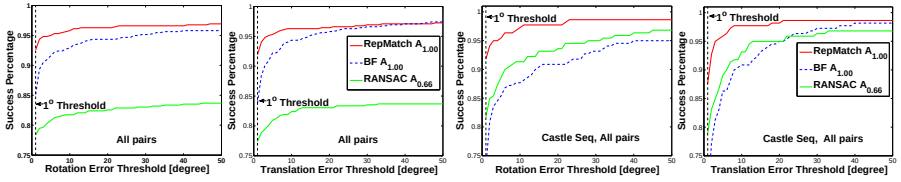
[2] We thank Chin Tat-Jun for his advice on RANSAC comparison.



(a) USAC w. different preemptive outlier removers (b) System-to-system comparisons

Fig. 8. Left: *RepMatch* compared against *USAC* [13] with different preemptive outlier removal schemes. Observe that *USAC*'s performance significantly improves with better preemptive outlier removal. **Right:** System to system comparison. *RepMatch*'s narrow baseline (Ground truth rotation $\leq 15°$) performance is easily comparable to the highly accurate MRMS [10] system. On All pairs, the difference is even (albeit unfairly) greater as *RepMatch* uses wide-baseline A-SIFT features while other systems use narrow baseline SIFT. This demonstrates *RepMatch*'s baseline generality.

shields the *RANSAC* module from outliers, we use a modified *RANSAC* with a larger minimal set of 20. All poses estimates are provided after this *RASANC* and a non-linear refinement step. For comparison to a more conventional *RANSAC* see Fig. 8(a). Fig. 9 shows that *RepMatch* preemptive outlier removal provides consistent performance gain vs both its *BF* and *RANSAC* components. This is especially notable on the castle sequence which has many repeated structures, resulting in *BF* under-performing *RANSAC* with a naive ratio test.



Eval. on all image pairs Eval. on repetitive castle sequence

Fig. 9. Left: Component wise evaluation on all pairs. Observe that *RepMatch* consistently improves its *RANSAC* and *BF* components. **Right:** Castle sequence which has significant repetition. *BF* is too aggresive and actually under-performs *RANSAC* $\mathcal{A}_{0.66}$. *RepMatch* avoids this performance degradation.

Finally, Fig. 10 evaluates *RepMatch*'s performance at different rotational baselines. At narrower baselines (below 45°), *RepMatch* is nearly perfect. It also remains remarkably robust to wide-baselines and maintains a $60-70\%$ pass rate at a 1° threshold for baselines exceeding 90°. Tab. 1 summarizes Fig. 9, Fig. 10 and provides matching statistics. It shows *RepMatch* provides consistent improvements over all scene types and baselines, with especially large gains at wide-baselines and repetitive structures. While spectacularly wide-baselines are

not necessarily useful in themselves, they are an indicator of very high moderate-baseline stability in less controlled environments, investigated in the next section.

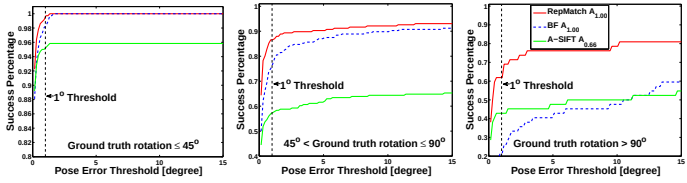


Fig. 10. Dataset divided by rotational baseline. At narrow baselines (below 45°), both *BF* and *RepMatch* are nearly perfect, with close to 100% pass at a 1° threshold. At wider baselines, the gap between *RepMatch* and *BF* widens. Notably, *RepMatch* has a $60-70\%$ SP at a 1° threshold for baselines exceeding 90°.

4.2 Structure-from-Motion Systems

Here we explore *RepMatch*'s performance on less controlled modern city images and its role in an overall SfM pipeline. Fig. 11 shows reconstruction of three modern scenes: (i) indoor; (ii) street; (iii) walking around a block. The sparse reconstruction system used is *Visual SfM*, which we provide with different feature matches[3]. We show performance with *RepMatch*, *BF* and *Visual SfM*'s default matching. While none of the sequences are especially wide baselines, *Visual SfM* has multiple breaks, demonstrating the difficult nature of modern city reconstruction. Using *BF* correspondences reduces the breaks but the reconstructed point clouds show serious errors with stray frames and phantom walls. *RepMatch*

[3] We leverage *RepMatch*'s robust pose estimate to eliminate all triplet poses with relative rotation consistency less than 2°. *BF* was employed with the same scheme.

Table 1. Evaluation on Strecha dataset [48]. We tabulate the match precision, average number of correct matches ("# matches"), Success Percentage (SP) at 1° rotation error and 1° translation error. *RepMatch* algorithm consistently improves on its individual components in terms of pose accuracy and match precision, with the difference increasing with baseline. In terms of match numbers, *RepMatch* has slightly fewer matches than *BF* but still maintains a substantial advantage over standard $\mathcal{A}_{0.66}$.

Algo.	Basline $\leq 45°$				$45° <$ Baseline $\leq 90°$				Baseline $\geq 90°$			
	Preci-sion	# match	$SP(1°)$ Rot.	Trans.	Preci-sion	# match	$SP(1°)$ Rot.	Trans.	Preci-sion	# match	$SP(1°)$ Rot.	Trans.
$\mathcal{A}_{0.66}$	0.915	5845	0.953	0.939	0.556	700	0.574	0.568	0.452	293	0.429	0.452
BF	0.957	19795	0.983	0.956	0.792	5078	0.769	0.759	0.457	2185	0.214	0.287
RepMatch	0.985	17800	0.997	0.975	0.886	4612	0.876	0.870	0.709	1912	0.619	0.714

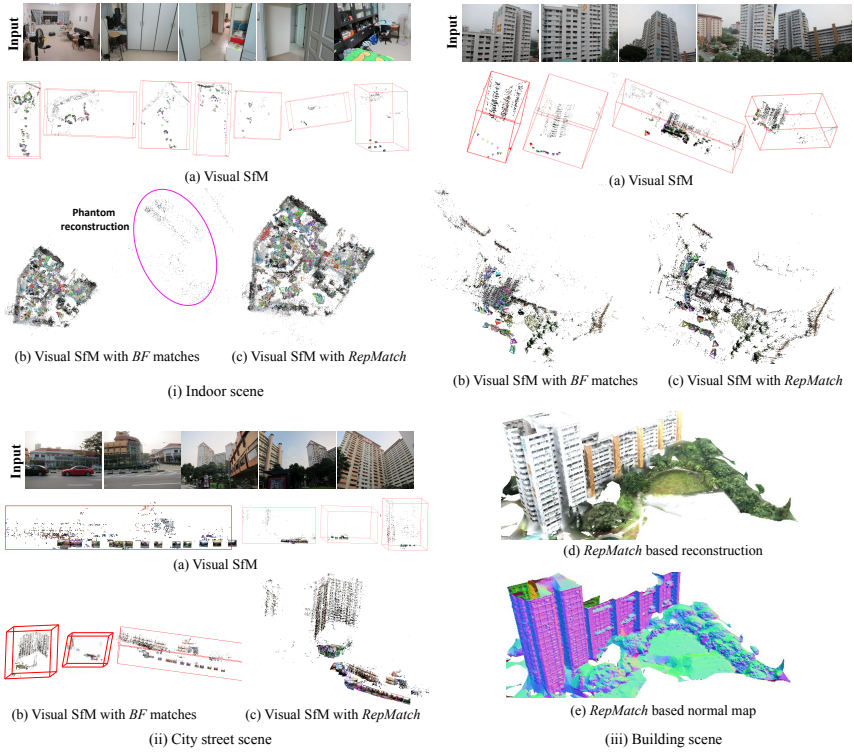


Fig. 11. Three city scenes. (i) Inside of a home. *RepMatch* can link through many weakly textured passages. *BF* also reconstructs the full flat but it creates phantom walls. (ii) A city street. Only *RepMatch* does not fragment the model. (iii) A city block. *RepMatch*'s reconstruction in (c) clearly shows the block outlines. This permits high quality dense reconstruction using [37] in (d) and (e).

permits un-fragmented, high quality reconstruction. Sequence (i) is especially interesting as *RepMatch* improves on *BF* even on indoor environments with few repetitive structures. This opens the possibility image information complementing current depth camera based floor plan recovery [49, 50].

5 Discussion

Apart from modern city reconstructions discussed earlier, two view pose and correspondence estimation are potentially useful in applications like image warping [51], system calibration, photometric estimation, etc. However, the chronic instability of two view pose estimates has limited their practical usefulness and caused a gradual decline in interest. *RepMatch*'s results suggests such pessimism may be unwarranted and the basic pose estimation problem deserves more attention. Perhaps with further research, reliable two view pose and correspondence estimates will be something future vision and robotic systems take for granted.

References

1. Agarwal, S., Snavely, N., Simon, I., Seitz, S.M., Szeliski, R.: Building rome in a day. In: Proc. of Int'l Conf. on Computer Vision. (2009) 72–79
2. Wu, C.: VisualSfM: A visual structure from motion system. URL http://www. cs. washington. edu/homes/ccwu/vsfm (2011)
3. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM **24**(6) (1981) 381–395
4. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int'l Journal of Computer Vision **60**(2) (2004) 91–110
5. Lin, W.Y., Cheng, M.M., Lu, J., Yang, H., Do, M.N., Torr, P.: Bilateral functions for global motion modeling. In: Proc. of European Conf. on Computer Vision. (2014) 341–356
6. Klein, G., Murray, D.: Parallel tracking and mapping for small ar workspaces. In: IEEE and ACM International Symposium on Mixed and Augmented Reality. (2007) 225–234
7. Leonard, J.J., Durrant-Whyte, H.F.: Simultaneous map building and localization for an autonomous mobile robot. In: Proc. of IEEE/RSJ International Workshop on Intelligent Robots and Systems. (1991) 1442–1447
8. Altwaijry, H., Belongie, S.: Ultra-wide baseline aerial imagery matching in urban environments. In: Proc. of British Machine Vision Conference. (2013)
9. Kushnir, M., Shimshoni, I.: Epipolar geometry estimation for urban scenes with repetitive structures. IEEE Trans. on Pattern Analysis and Machine Intelligence **36**(12) (2014) 2381–2395
10. Liu, Z., Monasse, P., Marlet, R.: Match selection and refinement for highly accurate two-view structure from motion. In: Proc. of European Conf. on Computer Vision. (2014) 818–833
11. Shah, R., Srivastava, V., Narayanan, P.: Geometry-aware feature matching for structure from motion applications. In: Proc. of Winter Conf. on Applications of Computer Vision (WACV). (2015) 278–285
12. Xiang, T., Cheong, L.F.: Understanding the behavior of sfm algorithms: A geometric approach. Int'l Journal of Computer Vision **51**(2) (2003) 111–137
13. Raguram, R., Chum, O., Pollefeys, M., Matas, J., Frahm, J.: Usac: A universal framework for random sample consensus. IEEE Trans. on Pattern Analysis and Machine Intelligence **35**(8) (2013) 2022–2038
14. Goshen, L., Shimshoni, I.: Balanced exploration and exploitation model search for efficient epipolar geometry estimation. IEEE Trans. on Pattern Analysis and Machine Intelligence **30**(3) (2008) 1230–1242
15. Chum, O., Matas, J.: Matching with prosac-progressive sample consensus. In: Proc. of Computer Vision and Pattern Recognition. (2005) 220–226
16. Ni, K., Jin, H., Dellaert, F.: Groupsac: Efficient consensus in the presence of groupings. In: Proc. of Int'l Conf. on Computer Vision. (2009) 2193–2200
17. Sattler, T., Leibe, B., Kobbelt, L.: Scramsac: Improving ransac's efficiency with a spatial consistency filter. In: Proc. of Int'l Conf. on Computer Vision. (2009) 2090–2097
18. Chum, O., Matas, J., Kittler, J.: Locally optimized ransac. In: Joint Pattern Recognition Symposium. (2003) 236–243
19. Wang, C., Wang, L., Liu, L.: Progressive mode-seeking on graphs for sparse feature matching. In: Proc. of European Conf. on Computer Vision. (2014) 788–802

20. Wang, C., Wang, L., Liu, L.: Density maximization for improving graph matching with its applications. IEEE Trans. on Image Processing **24**(7) (2015) 2110–2123
21. Lipman, Y., Yagev, S., Poranne, R., Jacobs, D.W., Basri, R.: Feature matching with bounded distortion. ACM Trans. on Graphics **33**(3) (2014) 26
22. Pizarro, D., Bartoli, A.: Feature-based deformable surface detection with self-occlusion. Int'l Journal of Computer Vision **97**(1) (2012) 54–70
23. Ok, D., Marlet, R., Audibert, J.Y.: Efficient and scalable 4-th order match propagation. Proc. of Asian Conf. on Computer Vision (2012) 460–473
24. Lin, W.Y., Wang, F., Cheng, M.M., Yeung, S.K., Torr, P., Do, M., Lu, J.: Code: Coherence based decision boundaires for feature correspondance. http://www.kind-of-works.com (2016)
25. Kamiya, S., Kanazawa, Y.: Accurate image matching in scenes including repetitive patterns. In: Proc. of Int'l Workshop on Robot Vision. (2008)
26. Zhang, Z., Matsushita, Y., Ma, Y.: Camera calibration with lens distortion from low-rank textures. In: Proc. of Computer Vision and Pattern Recognition. (2011)
27. Le Brese, C., Young, C., Zou, J.J.: A robust match filtering algorithm for use with repetitive patterns. In: Proc. of Signal Processing and Communication Systems (ICSPCS). (2013)
28. Lu, X., Manduchi, R.: Wide baseline feature matching using the cross-epipolar ordering constraint. In: Proc. of Computer Vision and Pattern Recognition. (2004)
29. Mortensen, E.N., Deng, H., Shapiro, L.: A sift descriptor with global context. In: Proc. of Computer Vision and Pattern Recognition. (2005)
30. Bay, H., Tuytelaars, T., Gool, L.V.: SURF: Speeded up robust features. In: Proc. of European Conf. on Computer Vision. (2006) 404–417
31. Fan, B., Wu, F., Hu, Z.: Aggregating gradient distributions into intensity orders: A novel local image descriptor. In: Proc. of Computer Vision and Pattern Recognition. (2011)
32. Snavely, N., Seitz, S.M., Szeliski, R.: Photo tourism: Exploring image collections in 3d. In: Proc. of ACM SIGGRAPH. (2006)
33. Chatterjee, A., Govindu, V.M.: Efficient and robust large-scale rotation averaging. In: Proc. of Int'l Conf. on Computer Vision. (2013)
34. Cui, Z., Tan, P.: Global structure-from-motion by similarity averaging. Proc. of Int'l Conf. on Computer Vision (2015)
35. Wu, C., Agarwal, S., Curless, B., Seitz, S.M.: Multicore bundle adjustment. In: Proc. of Computer Vision and Pattern Recognition. (2011) 3057–3064
36. Furukawa, Y., Ponce, J.: Accurate, dense, and robust multi-view stereopsis. In Proc. of Conference on Computer Vision and Pattern Recognition (2007)
37. Jancosek, M., Pajdla, T.: Multi-view reconstruction preserving weakly-supported surfaces. In: Proc. of Computer Vision and Pattern Recognition. (2011) 3121–3128
38. Raguram, R., Frahm, J.M., Pollefeys, M.: Exploiting uncertainty in random sample consensus. In: Proc. of Int'l Conf. on Computer Vision. (2009) 2074–2081
39. Torr, P.H.S., Zisserman, A.: MLESAC: A new robust estimator with application to estimating image geometry. Trans. on Computer Vision and Image Understanding **78** (2000) 138–156
40. Chin, T.J., Purkait, P., Eriksson, A., Suter, D.: Efficient globally optimal consensus maximisation with tree search. In: Proc. of Computer Vision and Pattern Recognition. (2015) 2413–2421
41. Cohen, A., Sattler, T., Pollefeys, M.: Merging the unmatchable: Stitching visually disconnected sfm models. In: Proc. of Int'l Conf. on Computer Vision. (2015) 2129–2137

42. Jiang, N., Tan, P., Cheong, L.F.: Seeing double without confusion: Structure-from-motion in highly ambiguous scenes. In: Proc. of Computer Vision and Pattern Recognition. (2012) 1458–1465
43. Myronenko, A., Song, X., Carreira-Perpinan, M.A.: Non-rigid point set registration: Coherent point drift. In: Proc. of Advances in Neural Information Processing Systems. (2006) 1009–1016
44. Yuille, A.L., Grzywacz, N.M.: The motion coherence theory. In: Proc. of Int'l Conf. on Computer Vision. (1988) 344–353
45. Lin, W.Y., Cheng, M.M., Zheng, S., Lu, J., Crook, N.: Robust non-parametric data fitting for correspondence modeling. In: Proc. of Int'l Conf. on Computer Vision. (2013) 2376–2383
46. Morel, J.M., Yu, G.: Asift: A new framework for fully affine invariant image comparison. SIAM Journal on Imaging Sciences **2**(2) (2009) 438–469
47. Hartley, R., Zisserman, A.: Multiple view geometry in computer vision. Cambridge University Press (2000)
48. 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: Proc. of Computer Vision and Pattern Recognition. (2008) 1–8
49. Yu, L.F., Yeung, S.K., Tang, C.K., Terzopoulos, D., Chan, T.F., Osher, S.J.: Make it home: Automatic optimization of furniture arrangement. ACM Trans. on Graphics **30**(4) (2011)
50. Ikehata, S., Yan, H., Furukawa, Y.: Structured indoor modeling. In: Proc. of Int'l Conf. on Computer Vision. (2015) 1323–1331
51. Lin, W.Y., Liu, S., Matsushita, Y., Ng, T.T., Cheong, L.F.: Smoothly varying affine stitching. In: Proc. of Computer Vision and Pattern Recognition. (2011) 345–352