Smoothly Varying Affine Stitching

Wen-Yan Lin[1] Siying Liu[1] Yasuyuki Matsushita[2] Tian-Tsong Ng[1] Loong-Fah Cheong[3]

[1]Institute for Infocomm Research [2]Microsoft Research Asia [3]National University of Singapore

{wdlin, sliu, ttngg}@i2r.a-star.edu.sg yasumat@microsoft.com eleclf@nus.edu.sg

Abstract

Traditional image stitching using parametric transforms such as homography, only produces perceptually correct composites for planar scenes or parallax free camera motion between source frames. This limits mosaicing to source images taken from the same physical location. In this paper, we introduce a smoothly varying affine stitching field which is flexible enough to handle parallax while retaining the good extrapolation and occlusion handling properties of parametric transforms. Our algorithm which jointly estimates both the stitching field and correspondence, permits the stitching of general motion source images, provided the scenes do not contain abrupt protrusions.

1. Introduction

Image stitching has long been of interest in graphics and vision. Its primary goal is the integration of multiple images into a single seamless mosaic. This serves many purposes, such as increasing the effective field of view, motion summarization and clean plate photography. Typically, stitching relies on an underlying transform which warps pixels from one coordinate frame to another. As the transformation must ensure perceptually accurate alignment of large (often quarter image width or greater) non-overlapping image regions, it must be robust to large view point changes and be able to generalize (interpolate and extrapolate) the motion over significant occlusion. To handle uncontrolled outdoor environments, the transform must also accommodate illumination changes and independent motion. For robust warping, mosaicing algorithms have traditionally sought to parameterize the warping field using a sparse set of global transformation parameters, such as the 3×3 affine or homographic matrix. This sparse parametrization ensures robustness at the expense of flexibility and is only accurate for a limited set of scenes and motions. For example, the commonly used homographic transforms are only accurate for planar scenes or parallax free camera motion between source frames i.e. the photographer's physical location must be fixed and only rotational motion is permitted.

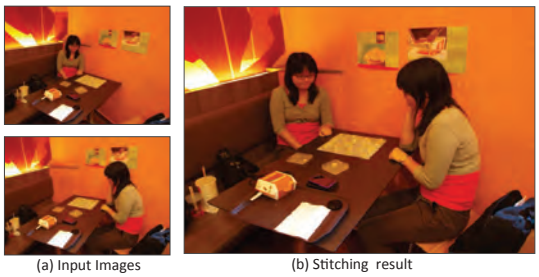

Figure 1. A girl passing the time by playing chess with herself, an example of our image stitching algorithm.

Dornaika et al. [8] highlighted that ideally, an image stitching algorithm would allow for both general motion and scene structure. As affine and homographic stitching can be considered a special case of a 3-D world's re-projection, a number of works [8, 21, 16] have proposed combining image stitching with 3-D reconstruction to enable the handling of parallax in general motion source images. However, using pre-computed 3-D points has a number of drawbacks. Firstly, 3-D reconstruction is only defined on the overlapping sections of the source image, making it difficult to integrate the non-overlapping regions, which is the primary objective of image stitching. Secondly, as noted by Liu et al. [16], the 3-D reconstruction pipeline is brittle, with its main components, accurate camera pose recovery and outlier free matching, still being active research issues. Thirdly, camera pose computation deteriorates if the motion contains too strong a rotational element or if the overlapping image regions are of inadequate size, both of which occur frequently in image stitching.

To achieve flexibility, we look to the 2D non-rigid warping approaches such as thin plate spline [4, 23], as-rigid-as-possible warping [13] and motion coherence [19]. They eschew the sparsity of parametric warping in favor of considering warping as a general matching problem with a smoothness constraint. This provides the flexibility needed to handle most motion types but at the expense of motion generalization over occluded regions. Hence, while warp-

ing algorithms may be used as a form of interpolation, they are seldom directly employed to solve traditional two view stitching problems. In this paper, we seek to adapt the warping framework to take advantage of the fact that many scenes can be modeled as having smoothly varying depth. This permits a general stitching algorithm which does not require an explicit 3-D reconstruction.

Our formulation is based upon the affine transform. A global affine transformation defines a set of shear, rotation, scaling and translational parameters, which preserve collinearity and the ratios of distances along a line. The general shape preservation property of the affine transform means that even when image pairs are not strictly related by a global affine matrix, it can still capture the gist of the camera motion induced deformation, such as whether the scene is translating upwards or side ways. Thus, provided motion discontinuities are not too extreme, a global affine transformation can be considered a parametric warping [12] which approximates the motion field and is utilized in a number of applications as a first approximation [26, 11]. This suggests that we can relax the affine constraint while retaining much of its strong motion generalization properties.

In this paper, we replace the global affine transformation with a smoothly varying affine stitching field which is defined over the entire coordinate frame. Every point has an associated affine parameter which is biased towards a pre-computed global affine transform and smoothness is enforced on the deviation of each affine parameters from the global affine. This model has two major advantages. Firstly, it is flexible enough to handle most kinds of motions, provided the scene contains no major protrusions Secondly, affine parameters can generalize the motion of a image region. Hence, a region of rather un-smooth 2D motion flow (such as a strong sheer, or forward translation) can become smooth if described using an affine stitching field, since all pixels in that region can be assigned a single, constant affine parameter. This makes the affine stitching field very smooth and thus, easily extrapolated over the non-overlapping regions.

To robustly compute the desired affine stitching field over large displacements, illumination change and occlusion noise, we utilize local view invariant feature descriptors like SIFT [17]. Unfortunately, dense feature descriptors such as those used in SIFT flow [15], introduce a lot of localization error as neighboring pixels are likely to share similar feature descriptors, thus making accurate dense descriptor based matching difficult. Instead, we rely on a sparse set of corner features (with associated SIFT descriptors) to compute the stitching field, which has an additional advantage in terms of computation time. While one can extrapolate a stitching field from pre-computed point matches, this is extremely vulnerable to outlier matches and a varying stitching field does not permit RANSAC based outlier rejection. Instead, we observe that a good stitching field can help validate existing correspondence and determine additional ones. These correspondences can in turn refine the stitching field. We exploit the inter-connectedness of these problems by jointly estimating both the matching and the stitching field. This prevents outlier matches, provides significantly more matches and yields a better stitching field.

To summarize our contribution:

1) We introduce a flexible image stitching algorithm that retains much of the motion generalization properties associated with global parametric transforms like affine/ homography. This permits the handling of general scenes and motions provided there are no abrupt protrusions. While our results do not always conform to the ground truth, it provides a good approximate which enables the creation of a perceptually correct composite.

2) We explore a range of applications made possible by this flexibility. These include novel scene generation illustrated in fig 1, computation of point correspondence and mosaicing of panoramas from translational motion.

1.1. Related Works

Our stitching field is related to the affinely over-parameterized optical flow algorithm of Tal et al. [20]. However, it is unclear how the framework of [20] can be adapted to the utilization of sparse high dimensional features and a bias towards a pre-defined affine. Instead, we utilize the motion coherence framework of Yuille et al. [27] and Myronenko et al. [19] to fit the affine stitching field.

Our work is also related to the 3-D reconstruction based image stitching methods mentioned in the introduction. These techniques have difficulty integrating the non-overlapping image regions. While this is not important for applications like Liu et al.'s [16] work on 3-D video stabilization, it is the central issue in forming large panoramas. A simple solution is to utilize an additional image [8], so that regions viewed by at least two images increase. However, this approach also increases the non-overlapping regions which cannot be mosaiced. An alternative is offered by Qi et.al. [21], where the 3-D reconstruction is used to generate virtual cameras from which strips are cut to ensure a smooth transition between the non-overlapping regions. This averages the error over the mosaic rather than attempting to align the images and is unsatisfactory because the error is incurred in the constrained overlapping region, rather than the unconstrained non-overlapping region. The lack of an underlying warping field also makes handling of occlusions and illumination change difficult and limits the algorithms applicability to other image editing task.

There are also works which seek to attain a perceptually accurate large field of view image through inputs other than conventional image stills. An interesting work is that of Kopf et al. [14] who generated virtual cameras from a series

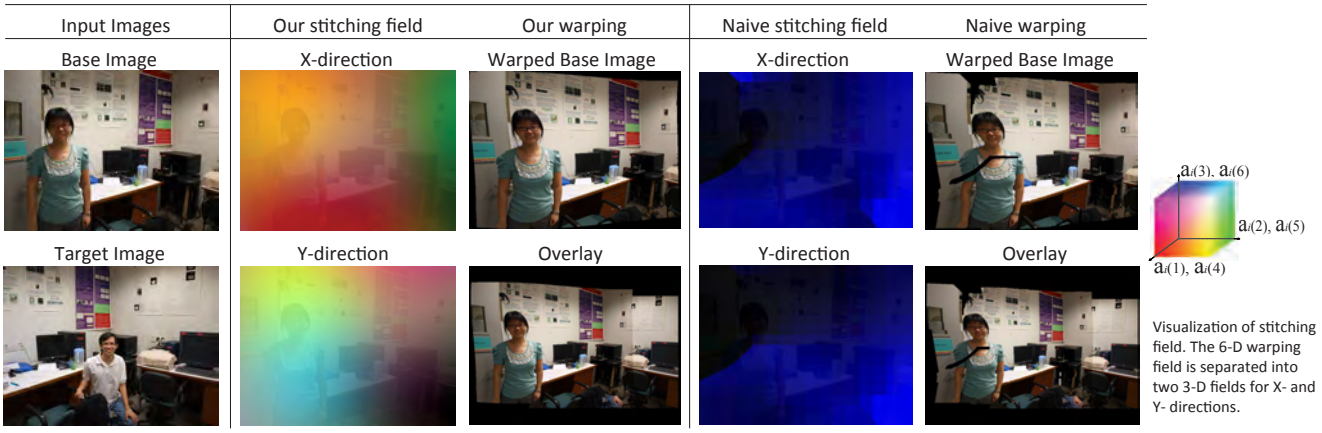

Figure 2. The affine stitching field transfers the base image to the target. The color coded deviation of each point's affine parameters from the global affine is overlaid on the base image. Affine parameters are divided into 2 groups according to the axis they operate on. Parameters in each group are assigned to one RGB color channel. The deviation from the global affine is proportional to color brightness. We present both our method and a naive method where the stitching field is computed by averaging the affine parameters computed form correspondences within a window. Observe that the naive method's stitching field is strongly biased towards regions where there are many correspondences. This makes it difficult to extrapolate the field to occluded regions such as the girl. Our algorithm can create a smooth field (seen in the color transitions) over the right angled corner, and has better extrapolation ability.

of "bubbles" (360^o panoramas), thus creating a long street view. Carrol et al. [6] introduced a warping which enables the un-distortion of a very wide angled image if the user defines a number of straight lines. For video sequences, Rav-Acha et al. [22] showed it is possible to leverage on the trackability and redundancy present in closely spaced video frames to incrementally stitch a large mosaic from general motion. However, it does not extend to the large displacement two view stitching considered in this paper.

There are also a large number of works which seek to refine conventional parametric image stitching. Interested readers can refer to the comprehensive tutorial by Szeliski [24] for an overview of such image stitching and blending techniques.

Finally, the field of image stitching is also related to various large displacement matching works such as those by Bhat et al. [3] and Fraundorfer et al. [10] and Brox et al. [25]. However, the focus of these works is on matching, rather than interpolation and extrapolation and as such, are not directly applicable to the image stitching domain.

2. Our Approach

A naive method of computing an affine stitching field would be to compute local affine parameters from SIFT correspondences within a sliding window. These affine parameters could then be averaged to give a smooth, dense affine stitching field with the parameters for non-overlapping regions obtained by extrapolation. This method produce a good, smooth stitching field in regions where the point correspondence is fairly plentiful. However, the performance declines significantly for regions where there are few/ no

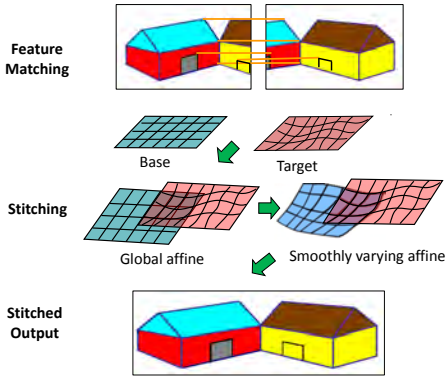

Figure 3. System overview. Point correspondence is used to obtain a global affine which our relaxes to form a smooth affine stitching field. The images are warped together and their overlapping regions blended to form a composite.

point correspondences and the extrapolation is generally poor. The reason is as follows. For any point (or small patch), there are many possible affines that can approximate its motion. Pre-computing an affine from correspondence forces us to choose one of the possibilities. While this choice may be locally optimal, it may not extrapolate well over the rest of the scene. The result is an affine stitching field that fits the regions of dense correspondence very well but does not give due weight to the sparser correspondences from the outlying regions. This problem is illustrated in fig 2, where even though we use a fairly large window (which helps avoid local over-fitting) with a length of a quarter image width, the affine stitching field computed still has difficulty extrapolating over regions with few correspondence.

In contrast, our problem is formulated as finding the

smoothest stitching field which can align the feature points of both images. This avoids a hard pre-assignment of the affine parameters, while the choice of stitching field carries within it an implicit extrapolation. An overview of our system is given in figure 3.

Our formulation's primary constraint consist of two sets of unmatched SIFT feature points. We denote the M features from the first, base image as $\mathbf{b}_{0i}$, while the N features from the second, target image, are denoted as $\mathbf{t}_{0j}$. The first two entries of the $\mathbf{b}_{0i}, \mathbf{t}_{0j}$ vectors represent image coordinates, with the remaining entries containing SIFT feature descriptors (This concatenation is done for notational simplicity and the necessary associated normalization is discussed in section 3). The stitching of the $\mathbf{b}_{0i}$'s to $\mathbf{t}_{0j}$'s is defined using a continuous affine stitching field $v(\mathbf{z}_{2\times 1}) : \mathbb{R}^2 \rightarrow \mathbb{R}^6$, which represents the deviation from a global affine $\mathbf{a}_{global}$. Mathematically, this is expressed as

$$(\Delta \mathbf{a}_i)_{6\times 1} = v\left(\left[\mathbf{b}_{0i(1)}; \mathbf{b}_{0i(2)}\right]\right), \tag{1}$$

where $\Delta \mathbf{a}_i$ is the deviation of feature i's affine $\mathbf{a}_i$ term from the global affine, given by $\mathbf{a}_i = \mathbf{a}_{global} + \Delta \mathbf{a}_i$.

We use $\mathbf{b}_i$ to represent the stitched feature points. $\mathbf{b}_i$ value depends only on the affine $\mathbf{a}_i$ term associated with the stitching field $v(.)$ and their original position $\mathbf{b}_{0i}$. This relationship can be expressed by the affine transform

$$\mathbf{b}_i = \left[\begin{array}{cc|c} \mathbf{a}_{i(1)} & \mathbf{a}_{i(2)} & \mathbf{0}_{2\times S} \\ \mathbf{a}_{i(4)} & \mathbf{a}_{i(5)} & \\ \hline \mathbf{0}_{S\times 2} & & \mathbf{I}_{S\times S} \end{array}\right] \mathbf{b}_{0i} + \left[\begin{array}{c} \mathbf{a}_{i(3)} \\ \mathbf{a}_{i(6)} \\ \mathbf{0}_{S\times 1} \end{array}\right]. \tag{2}$$

To facilitate easy reference to these affine parameters, we also define matrices

$$\mathbf{A}_{M\times 6} = [\mathbf{a}_1, ..., \mathbf{a}_M]^T, \Delta\mathbf{A}_{M\times 6} = [\Delta\mathbf{a}_1, ..., \Delta\mathbf{a}_M]^T.$$

We relate the base point set's alignment to the target point set, using the conditional probability based on a robust gaussian mixture

$$P(\mathbf{t}_{01:N}|\mathbf{b}_{1:M}) = \prod_{j=1}^{N}\left(\left(\sum_{i=1}^{M} g(\mathbf{t}_{0j} - \mathbf{b}_i, \sigma_t)\right) + 2\kappa\pi\sigma_t^2\right), \tag{3}$$

where $g(\mathbf{z}, \sigma) = e^{-\frac{\|\mathbf{z}\|^2}{2\sigma^2}}$ is a gaussian function and κ controls the strength of a uniform function which thickens the gaussian mixtures tails. κ is usually set to 0.5.

Apart from SIFT features, we also desire to incorporate a number of soft constraints. As mentioned earlier, we assume that the stitching field is a relaxation of a single global affine. Hence, we impose a smoothness constraint on the deviation of each points affine parameters from the global affine. We incorporate these soft constraints into a smoothing regularization term

$$\int_{\mathbb{R}^2} \frac{|v'(\omega)|^2}{g'(\omega)} d\omega, \tag{4}$$

where $v'(\omega)$ denotes the fourier transform of the continuous stitching field $v(.)$ and $g'(\omega)$ represents the fourier transform of a gaussian with spatial distribution γ. The regularization term biases the affine stitching field towards the global affine and ensures smooth transition between the constrained stitching field in the overlapping regions and the extrapolated stitching field in the occluded regions. While $\mathbf{A}$ is a discrete quantity and the stitching field $v(.)$ is continuous, the regularization term can be re-expressed in terms of $\mathbf{A}$ by choosing the smoothest stitching field which satisfies eqn (1). This yields

$$\Psi(\mathbf{A}) = \min_{v'(\omega)}\left(\int_{\mathbb{R}^2} \frac{|v'(\omega)|^2}{g'(\omega)} d\omega\right), \tag{5}$$

We combine the negative log of eqn (3) with the regularization term in eqn (5) and a λ weighting term, to form a single cost function,

$$E(\mathbf{A}) = -\sum_{j=1}^{N} log\left(\left(\sum_{i=1}^{M} g(\mathbf{t}_{0j} - \mathbf{b}_i, \sigma_t)\right) + 2\kappa\pi\sigma_t^2\right) + \lambda\Psi(\mathbf{A}) \tag{6}$$

which can then be minimized with respect to the variables in $\mathbf{A}$. Our minimization employs an EM style formulation, successfully used in [19].

2.1. Minimization

We follow a minimization procedure which computes an $\mathbf{A}^{k+1}$, using the $M \times 6$ linear equations defined by $\mathbf{A}^k$ in eqn (8). Compared to $\mathbf{A}^k$, $\mathbf{A}^{k+1}$ lowers the overall cost cost defined in eqn (6). The process is iterated until convergence. We define

$$\begin{aligned} \phi_{ij}(\mathbf{b}_i, \mathbf{t}_{0j}) &= g(\mathbf{t}_{0j} - \mathbf{b}_i, \sigma_t), \\ \overline{\phi_{ij}}(\mathbf{A}, \mathbf{t}_{0j}) &= \frac{\phi_{ij}(\mathbf{b}_i, \mathbf{t}_{0j})}{\sum_l \phi_{lj}(\mathbf{b}_l, \mathbf{t}_{0j}) + 2\kappa\pi\sigma_t^2}. \end{aligned} \tag{7}$$

Note that the second function's argument is given as $\mathbf{A}$ because, as can be seen from eqn (2), the $\mathbf{b}_i$ are base features after being warped by $\mathbf{A}$ and are wholly dependent on the $\mathbf{A}$.

Using Jensen's inequality, we can write

$$\begin{aligned} & E(\mathbf{A}^{k+1}) - E(\mathbf{A}^k) \\ \leq & -\sum_{j=1}^{N}\sum_{i=1}^{M} \overline{\phi_{ij}}(\mathbf{A}^k, \mathbf{t}_{0j}) log \frac{\phi_{ij}(\mathbf{b}_i^{k+1}, \mathbf{t}_{0j})}{\phi_{ij}(\mathbf{b}_i^k, \mathbf{t}_{0j})} \\ & + \lambda\left(\Psi(\mathbf{A}^{k+1}) - \Psi(\mathbf{A}^k)\right) \\ = & \Delta E(\mathbf{A}^{k+1}, \mathbf{A}^k). \end{aligned}$$

From the above, we know $\Delta E(\mathbf{A}^k, \mathbf{A}^k) = 0$. Hence, an $\mathbf{A}^{k+1}$ which minimizes $\Delta E(\mathbf{A}^{k+1}, \mathbf{A}^k)$ will ensure $E(\mathbf{A}^{k+1}) \leq E(\mathbf{A}^k)$.
Dropping all the terms in $\Delta E(\mathbf{A}^{k+1}, \mathbf{A}^k)$ which are independent of $\mathbf{A}^{k+1}$, we obtain a simplified cost function

$$Q = \frac{1}{2}\sum_{j=1}^{N}\sum_{i=1}^{M} \overline{\phi_{ij}}(\mathbf{A}^k, \mathbf{t}_{0j}) \frac{\left\|\mathbf{t}_{0j} - \mathbf{b}_i^{k+1}\right\|^2}{\sigma_t^2} + \lambda\Psi(\mathbf{A}^{k+1}).$$

Using a proof similar to that in Myronenko et al.[19], we show in the appendix that the regularization term $\Psi(\mathbf{A})$ has the simplified form $\Psi(\mathbf{A}) = tr(\Delta\mathbf{A}^T\mathbf{G}^{-1}\Delta\mathbf{A})$ where $\mathbf{G}$ is a $M \times M$ matrix, whose elements are given by $\mathbf{G}(i,j) = g(\mathbf{b}_{0i(1:2)} - \mathbf{b}_{0j(1:2)}, \gamma)$. Substitute this definition of $\Psi(\mathbf{A})$ into Q, take partial differentiation of Q with respect to $\mathbf{A}^{k+1}$ and post multiply $\mathbf{G}$ throughout, we have

$$\begin{aligned}\frac{\delta Q}{\delta \mathbf{A}^{k+1}} &= \begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \dots & \mathbf{c}_M \end{bmatrix} + 2\lambda \left(\Delta\mathbf{A}^{k+1}\right)^T \mathbf{G}^{-1} \\ &= \mathbf{C} + 2\lambda \left(\Delta\mathbf{A}^{k+1}\right)^T \mathbf{G}^{-1} = \mathbf{0}_{6\times M} \\ \Rightarrow &\mathbf{C}\mathbf{G} + 2\lambda \left(\Delta\mathbf{A}^{k+1}\right)^T = \mathbf{0}_{6\times M}, \\ \mathbf{c}_i &= \sum_{j=1}^{N} \frac{\overline{\phi_{ij}}(\mathbf{A}^k, \mathbf{t}_{0j})}{\sigma_t^2} \mathbb{D}(\mathbf{b}_i^{k+1} - \mathbf{t}_{0j}) \mathbb{V}(\mathbf{b}_{0i}).\end{aligned} \tag{8}$$

$\mathbb{D}(.), \mathbb{V}(.)$ are simultaneous truncation and tiling operators. They re-arrange **only the first two entries** of an input vector $\mathbf{z}$ (where $\mathbf{z}$ must have a length greater or equal to 2) to form the respective output matrices

$$\mathbb{D}(\mathbf{z})_{6\times 6} = \left[\begin{array}{c|c} \mathbf{z}_{(1)}\mathbf{I}_{3\times 3} & \mathbf{0}_{3\times 3} \\ \hline \mathbf{0}_{3\times 3} & \mathbf{z}_{(2)}\mathbf{I}_{3\times 3} \end{array}\right]$$

$$\mathbb{V}(\mathbf{z})_{6\times 1} = \begin{bmatrix} \mathbf{z}_{(1)} & \mathbf{z}_{(2)} & 1 & \mathbf{z}_{(1)} & \mathbf{z}_{(2)} & 1 \end{bmatrix}^T$$

From the definition of $\mathbf{A}$ in (2), we know that $\mathbf{b}_i^{k+1}$ can be expressed as a linear combination of the entries of $\mathbf{A}^{k+1}$. Hence, eqn (8) produces $M \times 6$ linear equations which can be used to estimate $\mathbf{A}^{k+1}$. $\mathbf{A}^{k+1}$ is used to estimate $\mathbf{A}^{k+2}$ and the process is repeated until convergence.

After convergence, the continuous stitching field $v(.)$ at any point $\mathbf{z}_{2\times 1}$ can be obtained from $\mathbf{A}$ using a weighted sum of gaussian given by

$$\begin{aligned}\mathbf{W}_{M\times 6} &= [\mathbf{w}_1, \dots \mathbf{w}_M]^T = \mathbf{G}^+\Delta\mathbf{A}, \\ v(\mathbf{z}_{2\times 1}) &= \sum_{i=1}^{M} \mathbf{w}_i g(\mathbf{z} - \begin{bmatrix} \mathbf{b}_{0i(1)} & \mathbf{b}_{0i(2)} \end{bmatrix}^T, \gamma),\end{aligned} \tag{9}$$

where $\mathbf{G}^+$ is the pseudo-inverse of $\mathbf{G}$ and the $6 \times 1, \mathbf{w}_i$ vectors can be considered weights for the gaussians. The detailed proof is given in the appendix.

3. Implementation

We now discuss our system implementation. A process overview is given in fig 3, with stitching field computation algorithmized in fig 4. In the formulation section, we have a global affine, $\mathbf{a}_{global}$ acting as a regularization term. $\mathbf{a}_{global}$ is computed from SIFT correspondences using RANSAC [9] for outliers removal. As $\mathbf{a}_{global}$'s regularization role lies in ensuring a smoother stitching field, its precise value is not important. All the $\mathbf{a}_i$ vectors in $\mathbf{A}$ are initially set to $\mathbf{a}_{global}$. The affine stitching field is then computed by repeatedly minimizing the cost in eqn (6) with increasingly

Input: Base image features $\mathbf{b}_i$, target image features $\mathbf{t}_j$, global affine matrix a_{global}

while σ_t *above threshold* **do**
  while *No convergence* **do**
    Use eqn (7) to evaluate $\phi_{ij}(\mathbf{b}_i^k, \mathbf{t}_{0j})$ from $\mathbf{A}^k$;
    Use eqn (8) to determine $\mathbf{A}^{k+1}$ from $\phi_{ij}(\mathbf{b}_i^k, \mathbf{t}_{0j})$
  end
  Anneal $\sigma_t = \alpha\sigma_t$, where $\alpha < 1$.
end

Output: $\mathbf{A}^{converged}$

Figure 4. Algorithm to compute stitching field.

smaller values of σ_t. Each step in this annealing process uses the previously calculated stitching field is as an initialization. We begin with $\sigma_t = 1$ and decrement it by a factor of 0.97, until $\sigma_t = 0.1$. The progressively smaller σ_t values increase the penalty for deviation between the target and base point sets, forcing the stitching field to evolve until the base point coordinates register onto the target points, resulting in a "match".Using an i7 computer running matlab, the algorithm can typically compute the registration of 1200 SIFT features in 8 minutes.

For notational simplicity, SIFT descriptors and point coordinates are condensed into a single vector. This implies a need for normalization. The point coordinates for the target and base points are normalized to have zero mean, unit variance, thus making the remaining parameter settings invariant to image size. We normalize the SIFT descriptors to have magnitudes of $10\sigma_t$, which gives good empirical results. The smoothing weight λ and outlier handling term κ are assigned values of $10, 0.5$ respectively. The γ term which penalizes un-smooth flow, is set to 1. Finally, we blend the images into a single mosaic, using the poisson blending with optimal seam finding algorithm of [7].

4. Analysis

The computed smooth affine stitching field is a "sparse" representation of the true warping function and errors will be incurred by smoothing over depth boundaries and generalization from a small set of feature points. In figure 5, we use two simulated scenes. The first is a simple "V" scene, while the second contains significant depth discontinuities. The scenes are projected onto 500×500 pixel images. Our computed warping has an avergae error of $1.92, 4.57$ pixels. This is small, considering we generalized the motion of 0.25 megapixels using 625 feature points, a ratio of $1:400$.

In fig 6, we provide a qualitative error analysis. The stitched images are overlayed. In the overlapping region, the green color channel is from the base image while the red and blue channels come from the target image. This

allows a visualization of alignment errors. While our algorithm incurs some errors along depth boundaries, they can be removed by blending.

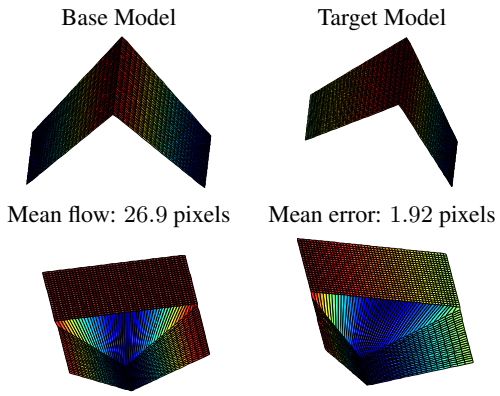

Figure 5. Quantitative analysis of our algorithm's motion generalization ability. The camera rotates 0.3 radians about the object. Using 625 uniformly distributed, unique features, we generalize the motion of a 500×500 image (0.25 megapixels), a 1:400 ratio.

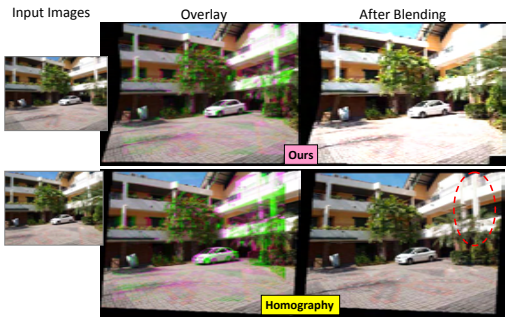

Figure 6. Results before and after poisson blending. For the pre-blended images, the overlapping regions take the green color channel from the base image and the red, blue channels from the target image. This enables visualization of alignment errors. Our algorithm incurs some errors along the depth boundaries. However, after blending, the results are perceptually accurate. The homographic mosaicing, incurs much larger errors and even after applying the same blending, clear artifacts remain.

5. Applications

Our algorithm's flexibility means that it can stitch images even when the photographer does not maintain a fixed position. This opens up a range of different possibilities.

5.1. Re-shoot

Bae et al.[2], noted that if the photographer has moved away from the original location, it is difficult to recover the exact view point. Our algorithm's good motion generalization and flexible stitching capability mean we can "re-shoot" a scene to incorporate information from different time instances. Observe that image editing using "re-shoot" differs from a "cut and paste" method of overlaying an object onto a background image. In "cut and paste", the overlay must be a discrete object such as a man or a car, with no attached background. As our stitching algorithm automatically warps the appended region to fit smoothly with the target image, "re-shooting" allows the overlay of an entire region, including the complex background and the subjects' interactions with it.

Figure 7. "Re-shoot" permits the integration of image pairs to create novel composites where the subject is interacting with the environment. This is not possible using conventional "cut and paste" image fusing methodology. **(a)** A girl passing the time by playing chess with herself. **(b)** Two people alternating as photographers to obtain a group photo. **(c)** The changing seasons at the famous Kiyomizu temple in Japan. **(d)** The passage of time at the Archbishop's palace in Prague.

In the first two scenes of fig 7, we insert a person into an image where he/she was not originally present and conversely, remove a person from the image. This allows interesting compositions such as a girl playing chess with herself in a cafe and permits two people to alternate as photographers to obtain a group photo. The following two scenes of fig 7 test our algorithm's limit by using internet images. These are much more challenging because photometric changes affect the SIFT feature invariance our al-

gorithm depends on. However, it permits more dramatic effects such as integration of summer time vista with the spectacular autumn foliage at Kiyomizu temple in Japan, as well as an image of a young couple walking from Prague's present into its past. We believe that our algorithm can be adapted to permit changes in the SIFT feature which would significantly improve its performance on internet images.

Technical discussion: "Re-shoot" is more challenging than panorama formation as the available blending region is narrow and the amount of occlusion typically very large. To ensure image consistency, we normalize the image colors. Poisson blending with optimal cut is employed [7] and followed by an additional alpha blending to merge the colors. It is carried out on a 25 pixel wide boundary along a user defined transfer region. For the shots using internet images, the blending boundary is set to be 50 pixels wide to accommodate the photometric variations and color normalization is discarded. In the Prague scene, the global affine was not pre-computed (due to a shortage of reliable matches) but set to an identity matrix. A more sophisticated blending for "re-shoot" can by obtained from [1].

5.2. Panoramic stitching

Our algorithm can be used for panorama creation. Its ability to handle general motion allows image stitching from un-conventional sequences, such as a series of images taken from different windows of a high-rise flat. As most windows are set back from the facade, this is not possible with homographic mosaicing [5] which requires a large un-occluded rotational field of view from a single window. Results are shown in fig 8. Observe that many of the views have only limited overlap, making camera pose recovery and hence mosaicing via 3-D reconstruction difficult.

5.3. Matching

Our algorithm can serve as a matcher across two views that can be related by a smoothly varying affine field (it will not match independent motion). As it matches features as a set, rather than individually, there is reduced dependency on feature descriptor uniqueness. In fig 9 we show that applying our algorithm with traditional SIFT descriptors [17], we can obtain 40% more matches. This is more than using a nearest neighbor matcher with more sophisticated A-SIFT [18] descriptors.

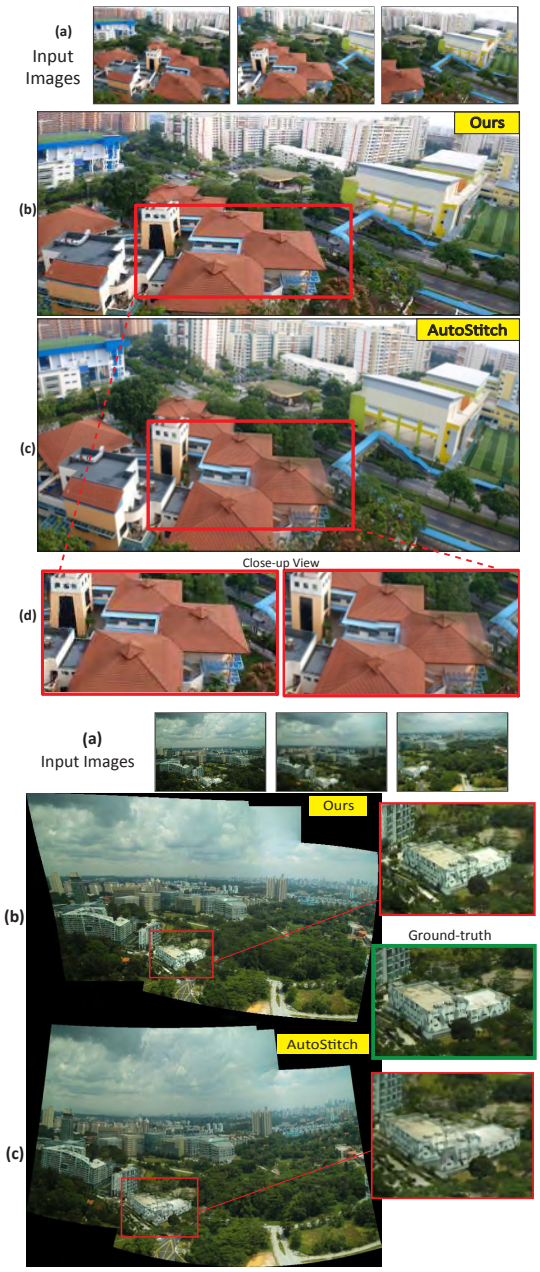

Figure 8. Results of panoramic stitching. Input images in **(a)** are taken from a series of windows. Our mosaic in **(b)** is perceptually accurate while homographic mosaicing using AutoStitch [5] in **(c)** has difficulty merging the fore-ground buildings.

6. Conclusion

We present an image stitching algorithm based on a smoothly varying affine field. It is significantly more tolerant to parallax than traditional homographic stitching but retains much of homographies ability to generalize motion over occlusion. Its flexibility enables integration of views taken from different physical locations, permitting a number of interesting applications like panorama creation from a translating camera or integration of images taken at different times. It can also accommodate other heuristics such as requiring straight lines to warp to straight lines, which may be considered in future work. Our algorithm's primary limitation is the violation of affine coherence at depth boundaries. While our results show these errors are often small enough to be blended over, explicit detection and handling would be better. In this regard, our results provide an excellent starting point for further refinement.

Figure 9. We show the results obtained by using our algorithm as a matcher, compared against conventional nearest neighbor SIFT [17] and A-SIFT [18] feature matching. Although we use traditional SIFT [17] descriptors, we can obtain more matches than applying nearest neighbor matching to the more sophisticated A-SIFT descriptor. The above figures show that the additional matches do not come at the expense of accuracy and the matching is stable to significant occlusion.

Acknowledgement This work is partially supported by project grant NRF2007IDM-IDM002-069 on Life Spaces from IDM Project Office, Media Development Authority of Singapore

References

[1] A. Agarwala, M. Dontcheva, M. Agrawal, S. Drucker, A. Colburn, B. Curless, D. Salesin, and M. Cohen. Interactive digital photomontage. *ACM Trans. Graph.*, 2004.

[2] S. Bae, A. Agarwala, and F. Durand. Computational rephotography. *ACM Trans. Graph.*, 29(3):1–15, 2010.

[3] P. Bhat, K. C. Zheng, N. Snavely, and A. Agarwala. Piecewise image registeration in the presence of multiple large motions. *CVPR*, 2006.

[4] F. L. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *PAMI*, 11(6):567–585, 1989.

[5] M. Brown and D. Lowe. Automatic panoramic image stitching using invariant features. *IJCV*, 1(74):59–73, 2007.

[6] R. Carroll, M. Agrawala, and A. Agarwala. Optimizing content-preserving projections for wide-angle images. *ACM Trans. Graph.*, 2009.

[7] L. H. Chan. and A. A. Efros. Automatic generation of an infinite panorama. *Technical Report*, 2007.

[8] F. Dornaika and R. Chung. Mosaicking images with parallax. *Signal Processing: Image Communication*, 2004.

[9] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Comm. of the ACM*, 24:381–395, 1981.

[10] F. Fraundorfer, K. Schindles, and H. Bischof. Piecewise planar scene reconstruction from sparse correspondance. *Image and Vision Computing*, 24(4), 2006.

[11] G. L. Gimel'farb and J. Q. Zhang. Initial matching of multiple-view images by affine approximation of relative distortions. *Proc of International Workshops on Advances in Pattern Recognition*, 2000.

[12] C. Glasbey and K. Mardia. A review of image warping methods. *In Journal of Applied Statistics*, 1998.

[13] T. Igarashi, T. Moscovich, and J. F. Hughes. As-rigid-as-possible shape manipulation. *ACM Trans. Graph.*, 24:1134–1141, July 2005.

[14] J. Kopf, B. Chen, R. Szeliski, and M. Cohen. Street slide: Browsing street level imagery. *ACM Trans. Graph.*, 2010.

[15] C. Liu, J. Yue, A. Torralba, J. Sivic, and W. T. Freeman. Sift flow: Dense correspondence across different scenes. *ECCV*, 2008.

[16] F. Liu, M. Gleicher, H. Jin, and A. Agarwala. Content-preserving warps for 3d video stabilization. *ACM Trans. Graph.*, 2009.

[17] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.

[18] J. Morel and G. Yu. Asift: A new framework for fully affine invariant image comparison. *SIAM Journal on Imaging Sciences*, 2(2):438–469, 2009.

[19] A. Myronenko, X. Song, and M. Carreira-Perpinan. Non-rigid point set registration: Coherent point drift. *NIPS*, 2007.

[20] T. Nir, A. M. Bruckstein, and R. Kimmel. Over-parameterized variational optical flow. *IJCV*, 2007.

[21] Z. Qi and J. Cooperstock. Overcoming parallax and sampling density issues in image mosaicing of non-planar scenes. *BMVC*, 2007.

[22] A. Rav-Acha, P. Kohli, C. Rother, and A. Fitzgibbon. Unwrap mosaics: a new representation for video editing. *ACM Trans. Graph.*, 27(3):1–11, 2008.

[23] R. Sprengel, K. Rohr, and H. S. Stiehl. Thin-plate spline approximation for image registration. *In Proc. of Engineering in Medicine and Biology Society*, 1996.

[24] R. Szeliski. Image alignment and stitching a tutorial. 2005.

[25] T.Brox and J.Malik. Large displacement optical flow: Descriptor matching in variational motion estimation. *PAMI*, 2010.

[26] K. Uno and H. Miiket. A stereo vision through creating a virtual image using affine transformation. *MVA*, 1996.

[27] A. L. Yuille and N. M. Grywacz. The motion coherence theory. *ICCV*, 1988.

7. Appendix A: Affine Coherence

This appendix deals with how the smoothness function can be simplified into a more computationally tractable form.

At the minima, the derivative of the energy term in (6) with respect to the stitching field $v'(.)$, must be zero. Hence, utilizing the fourier transform relation, $(\Delta \mathbf{a}_i)_{6\times 1} = v(\mu_i) = \int_{\mathbb{R}^2} v'(\omega) e^{2\pi\iota<\mu_i,\omega>} d\omega$, where $\mu_i = \begin{bmatrix} \mathbf{b}_{0i(1)} & \mathbf{b}_{0i(2)} \end{bmatrix}^T$, we obtain the constraint

$$
\begin{aligned}
&\frac{\delta E(v')}{\delta v'(\mathbf{z})} = \mathbf{0}_{6\times 1}, \forall z \in \mathbb{R}^2 \\
=> &-\sum_{j=1}^{N} \frac{\sum_{i=1}^{M} \left(\frac{g(\mathbf{t}_{0j} - \mathbf{b}_i, \sigma_t)}{\sigma_t^2} \right) diag\left(\mathbb{D}(\mathbf{b}_i - \mathbf{t}_{0j}) \mathbb{V}(\mathbf{b}_{0i}) \right) \int_{\mathbb{R}^2} \frac{\delta v'(\omega)}{\delta v'(\mathbf{z})} e^{2\pi\iota<\mu_i,\omega>} d\omega}{\sum_{i=1}^{M} g(\mathbf{t}_{0j} - \mathbf{b}_i, \sigma_t) + 2\kappa\pi\sigma_t^2} + \lambda \int_{\mathbb{R}^2} \frac{\delta}{\delta v'(\mathbf{z})} \frac{|v'(\omega)|^2}{g'(\omega)} d\omega = \mathbf{0}_{6\times 1} \\
=> &-\sum_{j=1}^{N} \frac{\sum_{i=1}^{M} \left(\frac{g(\mathbf{t}_{0j} - \mathbf{b}_i, \sigma_t)}{\sigma_t^2} \right) diag\left(\mathbb{D}(\mathbf{b}_i - \mathbf{t}_{0j}) \mathbb{V}(\mathbf{b}_{0i}) \right) e^{2\pi\iota<\mu_i,\mathbf{z}>}}{\sum_{i=1}^{M} g(\mathbf{t}_{0j} - \mathbf{b}_i, \sigma_t) + 2\kappa\pi\sigma_t^2} + 2\lambda \frac{v'(-\mathbf{z})}{g'(\mathbf{z})} = \mathbf{0}_{6\times 1}
\end{aligned}
\tag{10}
$$

$\mathbb{D}(.), \mathbb{V}(.)$ are simultaneous truncation and tiling operators. They re-arrange **only the first two entries** of an input vector $\mathbf{z}$ (where $\mathbf{z}$ must have a length greater or equal to 2) to respectively form the 6×6 and 6×1 output matrices

$$
\begin{aligned}
\mathbb{D}(z)_{6\times 6} &= \left[\begin{array}{c|c} \mathbf{z}_{(1)}\mathbf{I}_{3\times 3} & \mathbf{0}_{3\times 3} \\ \hline \mathbf{0}_{3\times 3} & \mathbf{z}_{(2)}\mathbf{I}_{3\times 3} \end{array} \right] \\
\mathbb{V}(z)_{6\times 1} &= \begin{bmatrix} \mathbf{z}_{(1)} & \mathbf{z}_{(2)} & 1 & \mathbf{z}_{(1)} & \mathbf{z}_{(2)} & 1 \end{bmatrix}^T
\end{aligned}
$$

$diag(.)$ is a diagonalization operator which converts a k dimensional vector $\mathbf{z}$ into a diagonal matrix, such that

$$
diag(\mathbf{z}_{k\times 1}) = \begin{bmatrix} \mathbf{z}_{(1)} & 0 & \cdots & 0 \\ 0 & \mathbf{z}_{(2)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{z}_{(k)} \end{bmatrix}_{k\times k}.
$$

Simplifying eqn (10), we obtain

$$
-2\lambda \sum_{i=1}^{M} \mathbf{w}_i e^{2\pi\iota<\mu_i,\mathbf{z}>} + 2\lambda \frac{v'(-\mathbf{z})}{g'(\mathbf{z})} = 0
$$

where the six dimensional vectors $\mathbf{w}_i$ act as placeholders for the more complicated terms in (10).

Substituting $\mathbf{z}$ with $-\mathbf{z}$ into the preceding equation and making some minor rearrangements, we have

$$
v'(\mathbf{z}) = g'(-\mathbf{z}) \sum_{i=1}^{M} \mathbf{w}_i e^{-2\pi\iota<\mu_i,\mathbf{z}>}. \tag{11}
$$

where the six dimensional vectors, $\mathbf{w}_i$, can be considered as weights which parameterize the stitching field.

Using the inverse Fourier transform relation

$$
\int_{\mathbb{R}^2} \mathbf{w}_i^T \mathbf{w}_j g'(\mathbf{z}) e^{+2\pi\iota<\mu_j-\mu_i,\mathbf{z}>} d\mathbf{z} = \mathbf{w}_i^T \mathbf{w}_j g(\mu_j - \mu_i, \gamma),
$$

and eqn (11), we can rewrite the regularization term of eqn (6) as

$$\begin{aligned}\Psi(\mathbf{A}) &= \int_{\mathbb{R}^2} \frac{(v'(\mathbf{z}))^T (v'(\mathbf{z}))^*}{g'(\mathbf{z})} d\mathbf{z} \\ &= \int_{\mathbb{R}^2} \frac{g'(\mathbf{z})^2 \sum_{i=1}^{M} \sum_{j=1}^{M} \mathbf{w}_i^T \mathbf{w}_j e^{+2\pi\iota<\mu_j-\mu_i,\mathbf{z}>}}{g'(\mathbf{z})} d\mathbf{z} \\ &= \sum_{i=1}^{M} \sum_{j=1}^{M} \int_{\mathbb{R}^2} \mathbf{w}_i^T \mathbf{w}_j g'(\mathbf{z}) e^{+2\pi\iota<\mu_j-\mu_i,\mathbf{z}>} d\mathbf{z} \\ &= tr(\mathbf{W^T G W}),\end{aligned} \tag{12}$$

where

$$\begin{aligned}\mathbf{W}_{M\times 6} &= [\mathbf{w}_1, ..., \mathbf{w}_M]^T, \\ \mathbf{G}(i,j) &= g(\mu_i - \mu_j, \gamma).\end{aligned}$$

Taking the inverse Fourier transform of eqn (11), we obtain

$$v(\mathbf{z}) = g(\mathbf{z}, \gamma) * \sum_{i=1}^{M} \mathbf{w}_i \delta(\mathbf{z} - \mu_i) = \sum_{i=1}^{M} \mathbf{w}_i g(\mathbf{z} - \mu_i, \gamma). \tag{13}$$

As $\Delta\mathbf{a}_j = v(\mu_j)$,

$$\Delta\mathbf{A} = \mathbf{GW}. \tag{14}$$

Substituting eqn (14) into (12), we see that the regularization term $\Psi(\mathbf{A})$, has the simplified form used in the main body

$$\Psi(\mathbf{A}) = tr(\mathbf{W^T G W}) = tr(\Delta\mathbf{A}^T \mathbf{G}^{-1} \Delta\mathbf{A}). \tag{15}$$

It can also be seen from eqn (14) that the stitching field $v(.)$ can be defined in terms of $\mathbf{A}$. This is done by using the matrices $\Delta\mathbf{A}, \mathbf{G}$ to define define the weighting matrix $\mathbf{W}$ via,

$$\mathbf{W} = \mathbf{G}^{+} \Delta\mathbf{A}. \tag{16}$$

This allows the definition of the stitching field at any point $\mathbf{z}_{2\times 1}$ using equation (13).

8. Appendix B: Additional Results

8.1. Panorama Creation

It is possible to stitch together fairly large panoramas using our algorithm. In figure 10, we show an extended version of the panorama used in the main body.

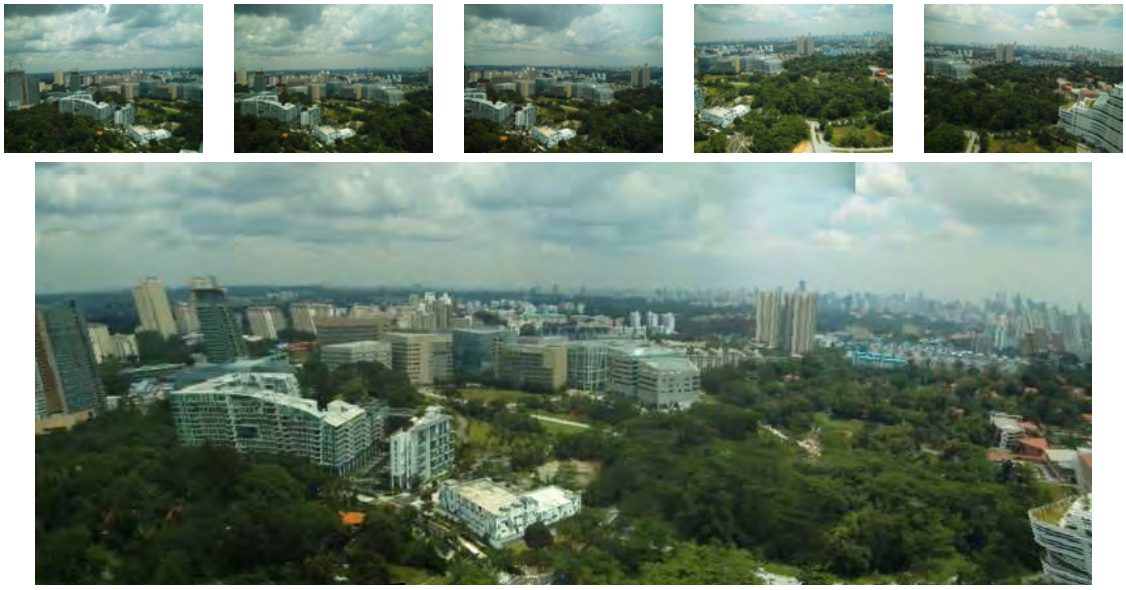

Figure 10. Long panoramic view

8.2. Protrusion Handling

While strong depth deviation is our algorithm's weak point, it is often possible for the post-blending results to be visually pleasing. This is illustrated in figure 11, where the tree is quite far in front than the store.

8.3. Re-Shoot

"Re-shoot" is quite stable to changes in the environment. In figure 12, we integrate images taken a few months apart, both before and after the workspace was occupied. Observe that there have been many modifications to the equipment in the room after it was occupied by a new worker. To re-assure the readers that the introductory image does not consist of a pure rotation, we also include the result of stitching the images using auto-stitch in fig 13.

9. Comparison with other warping techniques

Finally, we compare our results with some of those obtained using other warping techniques. In figure 14, we compare our affine based relaxation of a global affine, to a motion coherence based relaxation. Observe that our approach gives better extrapolation ability. In figure 15, we compare against SIFT flow [15] and large displacement optical flow [25], two recently developed dense matching techniques. These algorithms are not designed for image stitching and do not give good extrapolation.

Figure 11. While we incur clear errors, the mixture of our warping and a blending algorithm could still recover a somewhat pleasing mosiac from an image pair where the frontal tree represents a major protrusion.

Figure 12. Combing images taken a few months apart, before and after a office worker occupied his work station.

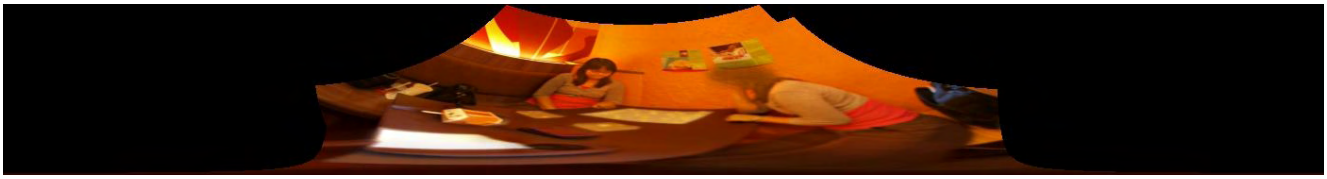

Figure 13. Auto-stitch result on the introductory image.

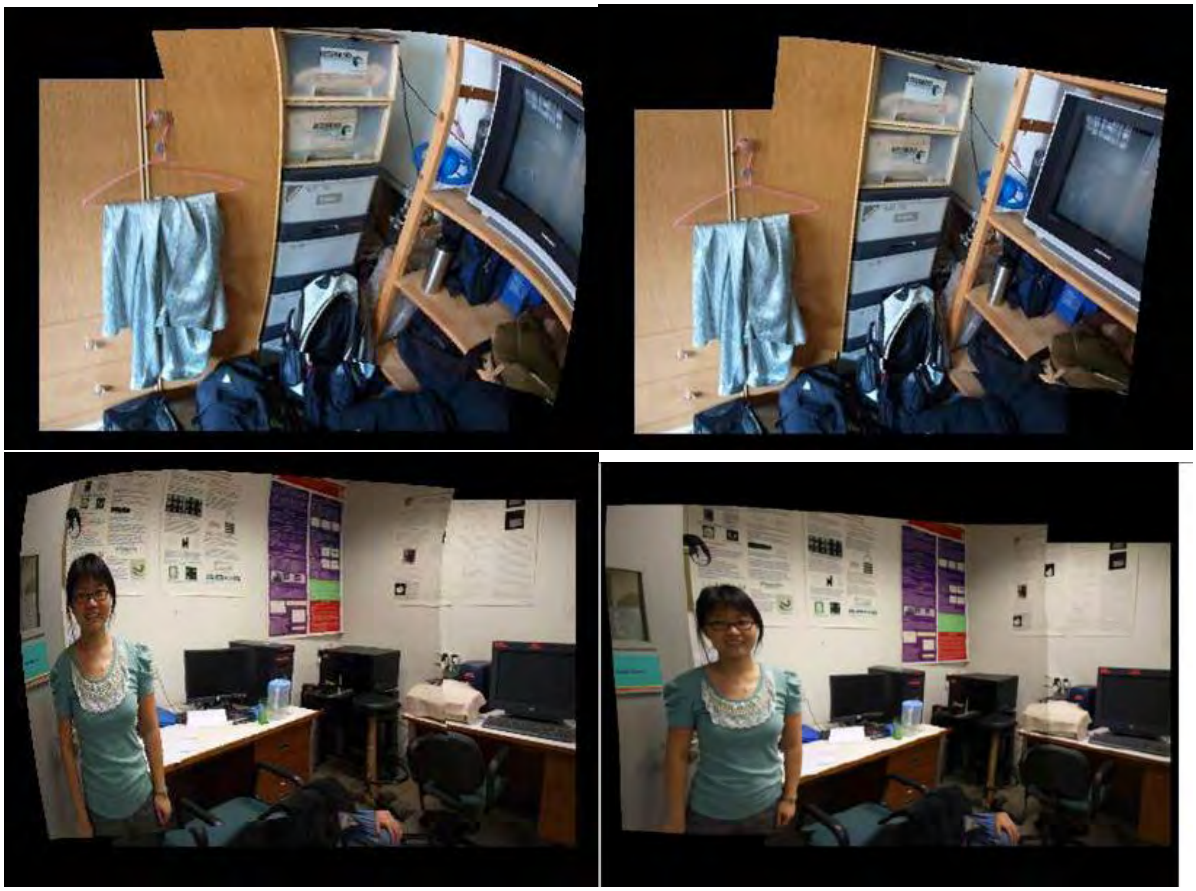

Figure 14. Left: Relaxing a global affine using normal coherence. Right: Our affine coherence relaxation.

SIFT flow [15] Large displacement flow [25] Our Sparse SIFT warping

Figure 15. Computing the warping using dense SIFT features in the SIFT flow algorithm of [15], large displacement optical flow [25] and our algorithm. Our results are perceptually more pleasing and easier to mosaic. Our warping also extrapolates the motion for occluded regions. These results are for the image pair shown in figure 1