

Multiscale Gigapixel Video: A Cross Resolution Image Matching and Warping Approach

Xiaoyun Yuan^{1,2}, Lu Fang^{1*,2}, Qionghai Dai¹, David J. Brady³, Yebin Liu^{1*}

¹Tsinghua University, ²Hong Kong University of Science and Technology, ³Duke University

<http://media.au.tsinghua.edu.cn/gigavideo.html>

Abstract

We present a multi-scale camera array to capture and synthesize gigapixel videos in an efficient way. Our acquisition setup contains a reference camera with a short-focus lens to get a large field-of-view video and a number of unstructured long-focus cameras to capture local-view details. Based on this new design, we propose an iterative feature matching and image warping method to independently warp each local-view video to the reference video. The key feature of the proposed algorithm is its robustness to and high accuracy for the huge resolution gap (more than 8× resolution gap between the reference and the local-view videos), camera parallaxes, complex scene appearances and color inconsistency among cameras. Experimental results show that the proposed multi-scale camera array and cross resolution video warping scheme is capable of generating seamless gigapixel video without the need of camera calibration and large overlapping area constraints between the local-view cameras.

1. Introduction

Traditionally, video systems have assumed that the resolution of the camera matches the resolution of the display, i.e., HD video uses HD cameras and displays, 4K video uses 4K cameras and displays, etc. The recent development of gigapixel [18, 1, 10, 4] and VR video systems [20] has illustrated the potential and need for systems in which the camera captures substantially more image information than the display can show. These systems use tiled multiscale image structures to enable viewers to interactively explore the captured image stream.

Size, weight, power and cost are central challenges in gigapixel video. The multiscale optical design of the Duke AWARE cameras [4] substantially reduced the size and weight of gigapixel scale optical systems, but in video op-

eration the volume and weight of camera electronics was more than 10× larger than the optics [27]. Moreover, compressive sampling strategies are essential to practical implementation of gigapixel scale broadcasting systems. Typical scenes are highly compressible, often with large sections of sky, foliage and background and with limited corridors of action such as roadways or playing fields. The ideal gigapixel system would have variable frame rates across the scene and adaptively capture regions base on activity levels.

To this end, this paper presents an efficient method, in terms of budget, sensor bandwidth and set up labors, for gigapixel videography using a novel multiscale camera array. Our capture system owns a reference camera with a short-focus lens to capture a reference video with comparably large field-of-view, and a parallel of local-view cameras each with a long-focus lens to obtain high-definition local-view videos. Such setting enables gigapixel videography by independently warping each local-view video to the reference video, and allows flexible, adaptive and moveable local-view camera setting during data capture. However, it also poses new challenges for the video warping algorithm, including parallaxes, abundant and various complex scene appearances, color inconsistency among cameras, and, most importantly, a more than 8× resolution gap that exists between the reference and the local-view videos.

We therefore contribute a cross resolution matching and warping approach to automatically and iteratively embed each local-view video in the reference video. Our algorithm jointly leverages patch-based matching, pixel-based variational warping, mesh-based homography and temporal stabilization, to form a coarse-to-fine optimization framework with several iteration updates between feature matching and video warping. In each iteration, both the feature matching and the video warping modules gradually update the information to benefit each other, finally leading to a robust and precise warping result.

With the proposed algorithms, our system enables several unique capabilities over the prior works: 1) it bypasses the need for careful camera alignment, tedious geometry and color calibration [38, 13], as well as the requirement

⁰*Corresponding authors:
fanglu@sz.tsinghua.edu.cn, liuyebin@tsinghua.edu.cn

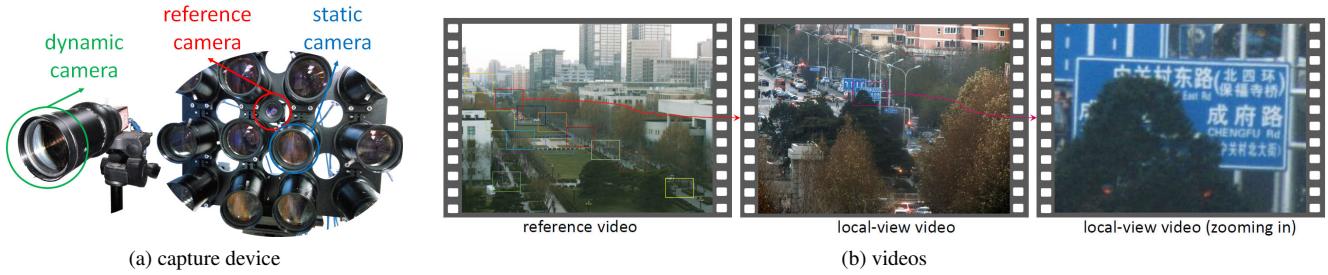


Figure 1: Capture device. (a) our capturing device with hybrid cameras. (b) example data captured by our camera array.

for image overlaps among local-view cameras in available methods [4, 18]; 2) it allows local-view camera movements such that the gigapixel video is captured in an adaptive and efficient way by allocating more sensor resources to the regions of interest; 3) it enables parallel stitching of the final high resolution video and may be optimized for real time synthesis of gigapixel video in the future.

We demonstrate a system with 14 local-view video cameras and successfully stitch them with another reference camera. Experimental results validate that our proposed cross resolution matching and warping algorithm is the only algorithm that is competent in outdoor scene warping in the face of more than $8\times$ resolution gap. We believe that our system and algorithm will open up new research on more adaptive and efficient gigapixel video capture and synthesis using capture setups with smaller size and lower cost.

2. Related Work

The contributions of this paper are two-fold, i.e., the system contribution and the algorithm contribution. In this section, we review these two aspects respectively.

2.1. Video Stitching Systems

High resolution videography [41] or 360° panoramic video [20] has been achieved by stitching a large amount of regularly spaced cameras. The main challenges in these kind of array are the requirement of manually and precisely structured and aligned with enough overlap of regions in neighboring views. Parallax between cameras may cause disturbing artifacts such as video misalignment and discontinuity [36]. Video stitching systems using unstructured camera arrays [28] bypassed the need to delicately layout and calibration, but require large area of overlapping regions for sparse feature matching and consume substantial amount of computation resources in spatial-temporal optimization to reduce artifacts caused by parallax.

The real gigapixel video camera [4] adopted a multiscale imaging design, with a spherical objective lens producing an image of the whole scene and 98 micro-optic cameras each relaying a portion of the image onto its own sensor. However, although relay cameras can be selected for imag-

ing regions of interest, the whole size of hardware framework is fix. Moreover, because the image targets are outdoor scenes in a large scale, it presents great challenges in optical design and evaluation [16], camera calibration and mechanical testing [17].

Kopf et al. [18] demonstrated a motor-controlled camera mount with a long-focus lens for capturing and stitching gigapixel images. Pirk et al. [29] further proposed to add video clips captured by the same camera and embed them into the gigapixel image. Whilst our system shares a similar embedding procedure, we face a more challenging problem where the reference is an unsynchronized video captured by a different camera. Moreover, it is of much lower resolution with parallax and with different color settings. We therefore contribute the first method for matching and warping of high resolution videos to the extremely low resolution video reference.

Our system is also related with multi-camera systems using hybrid resolution setup. Sawhney et al. [34] proposed the hybrid stereo camera system with $4\times$ resolution gap between the high and the low resolution video. In our system, to form a gigapixel video using a 4k reference video, the resolution gap between the reference and the local-view should be larger than $8\times$. Under this resolution gap, most of the features in the high resolution one will disappear in the low resolution one. Even though, $8\times$ resolution gap has been considered in the light field super-resolution systems such as [3, 39], the data capture in their works are under controlled indoor setup. In contrast, our system captures wide range wild scene with large depth of field, resulting in spatially nonuniform quality gap between the reference video and the local view videos in our system. As demonstrated in Sec 6, the PatchMatch method used in [3, 39] fails with our data.

2.2. Matching and Warping Algorithms

For unstructured multi-view inputs, an image and video stitching algorithm involves two steps: the feature matching, i.e., building of sparse feature correspondences on overlapping regions, and the dense estimation of a per-pixel warping field for the whole images. For the first step, a

number of feature descriptors such as SIFT [26][5], SURF [2], DASIFY [37], multi-scale oriented patches [6] and outlier elimination algorithm RANSAC [11] are widely used. While these algorithms work quite well for images with similar quality and resolution, their performances on large cross-scale cases have not been examined.

Warping field estimation is a crucial step to get the dense per-pixel matching. Existing warping field estimations can be divided into two categories: mesh-based homography warping [24, 25] and pixel-based variational warping [32, 7]. For homography warping, since parallax exists between different images, a uniform transformation is not powerful enough and the whole image is divided into many blocks or mesh grids to estimate transformation parameters for each grids, resulting in a non-uniform warping fields. As-projective-as possible warping [44], as-similar-as possible shape-preserving image warping [25], combined projective and similarity transform [9] and spatial-temporal content-preserving warping [15] have been proposed for stitching of images or videos. Different optimization problems are proposed for solving these transformations, where the data terms are very similar (based on matched feature points), and the difference lies in smooth terms. Weak smooth terms make it possible to handle large parallax but reduce the robustness to outlier matches or lack-of-correspondence regions.

Most of the pixel-based variational warping methods are based on the optical flow framework of Horn and Schunck [14]. This kind of variational optimization is a local optimization on an under-determined problem, which can be easily trapped into local minimum. Although robust regularization such as L1 norm [43], sparse descriptor matching term [8, 40], feature correspondences [32], coarse-to-ne mechanism [42] have been proposed to handle large motions, the variational framework is still incompetent in preserving scene structures.

3. System Overview

This paper proposes a multiscale camera system which enables high quality and efficient gigapixel video capture. Our system is integrated with one reference camera and 14 local-view cameras. All of them are PointGrey FL3-U3-88S2C-C rolling-shutter cameras without hardware synchronization requirement, and work at 4000×3000 spatial resolution and 15 fps frame rate. In particular, the local-view cameras share a focal length $f_l = 135mm$ to capture the local high resolution videos, and the focal length of the reference camera is $f_r = 16mm$ so as to cover a large field-of-view of an outdoor scene. Both of the two kinds of lenses cost only \$150 for each. Each local-view camera can be static or moving (see Fig. 1(a) left) during data capture. Fig. 1(b) shows the reference video embedded with all the local-view videos, and one of the local-view videos as well

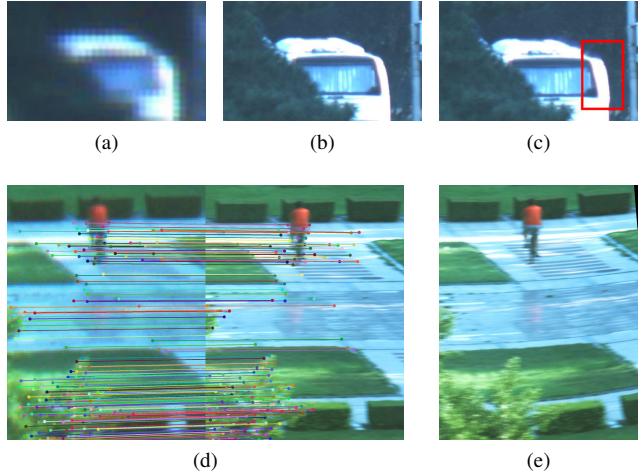


Figure 2: Artifacts caused by pixel-based warping (a,b,c) and mesh-based warping (d,e). (a) A crop from the reference block, (b) corresponding region in the local-view video, (c) warping result of typical optical flow [7], where the boundary frame of the bus is seriously deformed, (d) non-uniformly distributed matching points, and (e) mesh based warping result using (d) as input, where the shape of the ground is severely curved.

as its zoom-in detail. With the reference video, the construction of a higher resolution video is similar to the case of finishing a jigsaw puzzle with a provided low resolution guide map, and all the local-view videos are enabled to form accurately. Note that our system requires neither color nor geometry calibration before data capture.

Preprocessing We first scale down each image in the local-view video using the ratio f_r/f_l . A zero-mean normalized cross-correlation (ZNCC) [22] matching is then applied on the whole scaled-down image to find its position on the corresponding reference frame. For moving cameras, we use Kalman filter [33] to predict the moving paths for detecting and removal of erroneous matched positions. Finally, we crop the best matched block out from the reference image and up scale it to the same size as the local-view videos, denoted as *reference block*. The following warping threads will operate on each local-view video and its corresponding reference block in parallel, which will be described in Section 5.

4. Algorithm Analysis : Challenges

While our system enables many novel advantages as described above, it brings the following new challenges.

Complex Warping Field Because of camera parallax, scene depth variations, diversity of object appearances (pedestrians, cars, trees, buildings, reflective regions, etc.) and camera rolling shutter effects, the ideal warping field

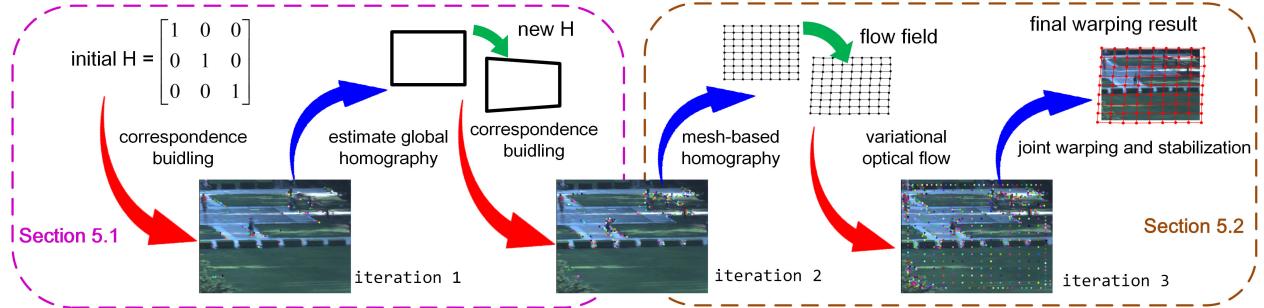


Figure 3: Pipeline of our cross resolution matching and warping algorithm. The red arrows denote the feature correspondence building process and the blue arrows denote the warping process.

between local-view video and reference block is highly non-uniform. In addition, the temporal warping field should guarantee the stability of the final warped video.

Large Resolution Gap Because the warping field between the local-view video and the reference block is highly non-uniform, it is essential to build as-many-as-possible and as-uniform-as-possible feature correspondences. However, there exists a f_l/f_r times, i.e., more than $8\times$, resolution gap between the local-view video and the reference block, and available feature detectors such as SIFT, SURF, DAISY and LATCH are incapable in finding enough effective feature points in the reference blocks.

Color Inconsistency Two factors lead to the color inconsistency: 1) vignetting in the local-view videos, and 2) different color sensitivities among local-view videos.

As reviewed in Section 2 and shown in Fig. 2, existing warping field algorithms are vulnerable to the above challenges. Pixel-based variational warping can handle large non-uniformity, but fails in maintaining scene structures that are sensitive to human eyes (Fig. 2(a,b,c)). Mesh-based warping approaches can preserve scene structures since pixels in the same mesh quad follow a projective transformation. However, it is still challenging to handle a serious nonuniform warping field because artifacts may appear near boundaries of neighbor quads with a deficient number of feature correspondences, as shown in Fig. 2(d,e).

5. Cross Resolution Matching and Warping

To solve the above key challenges introduced by the presented system, we propose a novel non-rigid warping algorithm to warp each local-view video to its reference block. Inspired by the iterative closest point (ICP) method [23] widely used in non-rigid 3D geometry registration, we design the algorithm as iteration updates between feature correspondence construction and image warping. The basic idea is to start with a strict threshold with some highly confident feature correspondences, and then gradually relax the threshold with the increase of iterations. Warping and correspondence are beneficial to each other, and with the increasing

of iterations, warping becomes more and more accurate and the correspondences become more and more dense and uniform. Note that in each iteration, we use different correspondence and warping strategies. Fig. 3 shows our 3-iteration cross resolution matching and warping pipeline. Overall, these iterations can be divided into two main parts, i.e., global homography warping and mesh-based warping.

5.1. Global Warping Iterations

The former part of our iteration pipeline aims at finding robust feature correspondences between each local-view image and its reference block. As shown in the left dash window of Fig. 3, initial feature correspondences are measured by structure similarity using zero-mean normalized cross correlation (ZNCC) between patch p_l in the local view image I_l and patch p_r in reference block I_r . Based on the matched correspondences, we then calculate a global homography \mathbf{H} to warp the local-view image on the reference block. In the second iteration, given the estimated homography \mathbf{H} , the algorithm detects as-many-as-possible reliable matching points, still based on the ZNCC metric.

The ZNCC matching scheme in these two iterations can be formulated as

$$\begin{aligned} p_r^* &= \arg \max_{p_r} \text{ZNCC}(I_r(p_r, w), I_l(p_l, w)) \\ \text{s.t. } & \|\pi(\mathbf{H}\tilde{p}_l) - p_r\|_2 < \epsilon, \end{aligned} \quad (1)$$

where \tilde{p}_l is the homogenization of p_l and function π performs perspective projection and dehomogenization based on homography matrix \mathbf{H} . In the first iteration, \mathbf{H} is initialized to be an identity matrix, and both the patch size w and the search range ϵ are set to be 256. In the second iteration, we refine the matching, and both w and ϵ are set to be 128.

We choose the ZNCC metric because of its robustness to blur and luminance. Moreover, ZNCC can be accelerated by FFT [22]. We compare our proposed feature matching algorithm with the state-of-the-arts including SIFT [26], SURF [2], DAISY [37] and LATCH [21]. As shown in Fig. 4, none of these feature detectors can find as many correct correspondences as ours in the second iteration (Fig. 4(c)).

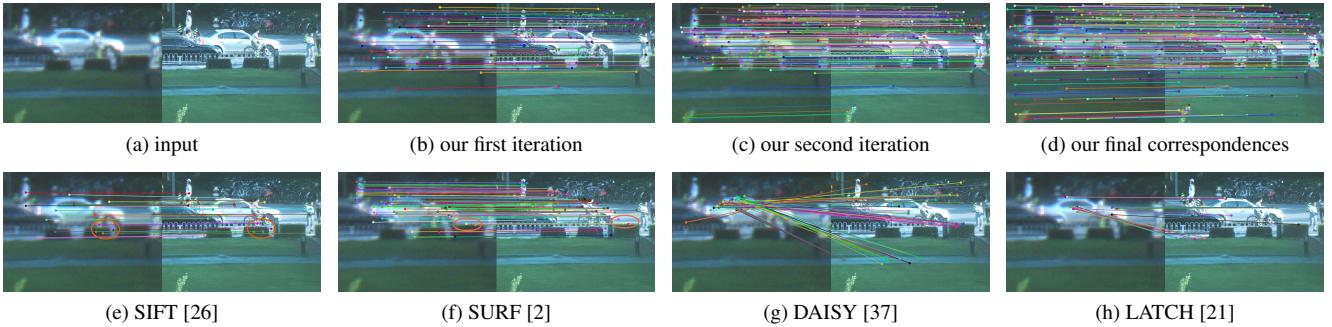


Figure 4: The distributions of feature correspondences between a local-view image and its reference block. The ellipses highlight the error matches.

5.2. Mesh-based Warping Iterations

Since the ideal warping field is non-rigid and complex, a mesh-based warping strategy such as [24, 25] is essential for a successful warping. However, after the global warping step, a large area of smooth texture regions still lacks feature correspondences, e.g., the grassland region in Fig. 4(c). In this situation, available mesh-based homography approaches introduce mismatches on these regions, especially when the lack-of-feature areas appear on the boundary of the image. As analyzed in Section 4.

Complementary optical flow matching. Against this backdrop, we propose a novel mesh-based warping method that takes the advantages of both mesh-based warping and optical flow warping, and apply it in the second and the third iterations. The key insight here is that the mesh-based warping and variational optical flow warping are complementary. The former is excellent at visual salient regions such as structures, while the later performs better on large regions without feature correspondences.

Specifically, we perform a mesh-based warping using feature correspondences detected in the second iteration. Based on the warping result, a variational optical flow algorithm is then applied to refine the per-pixel correspondence between the reference and the local-view image. After this, we obtain an improved per-pixel correspondence for the smooth texture regions. This correspondence can not be obtained by ZNCC matching because of its lack of spatial smoothness prior regularizer. Also, note that the optical flow algorithm can not be applied in the earlier steps because it requires a comparably high quality initialization, which can be better provided by the mesh-based warping step.

Lastly, our algorithm sparsely selects optical flow correspondences as the complementary optical flow features. The positions of these features should be 1) salient features on the local-view image, or 2) on the mesh-grid vertices. We combine them to guide the final mesh-based homography warping.

Joint warping and stabilization. Our mesh-based

warping utilizes the as-similar-as-possible (ASAP) [25] warping framework. To maintain the temporal stability of the final warped video, we introduce a temporal stability term in it. Mathematically, we minimize the energy function

$$E(V) = \lambda_r E_r(V) + \lambda_t E_t(V) + \lambda_s E_s(V). \quad (2)$$

Here, V is the positions of the deformed grid vertices. E_r is the data term to enforce the current warped local-view image to be close to the reference block, and E_t is the temporal stability term. The smooth term E_s regularizes the spatial smoothness deformation on neighboring vertices, and is the same as described in [25].

The data term E_r sums the distances of all the feature correspondences $\{p_l, p_r\}$ on image pair $\{\mathbf{I}_l, \mathbf{I}_r\}$ (we omit time notation t here) after warping V (defined on the grid vertices) is operated. E_r is formulated as

$$E_r(V) = \sum_{p_l} \|\alpha_{p_l} V - p_r\|_2^2, \quad (3)$$

where α_{p_l} is the bilinear interpolation weight of p_l . For more details of ASAP warping, please refer to [24][25].

The stability term E_t is defined as

$$E_t(V) = \sum_{p_l} \mathbf{B}(p_l) \|\alpha_{p_l} V - (\hat{p}'_l + s)\|_2^2, \quad (4)$$

which sums the distances over the feature pair $\{p_l, \hat{p}'_l\}$ on image \mathbf{I}_l and its temporal precedent warped image $\hat{\mathbf{I}}'_l$. Here, notation $\hat{\cdot}$ implies “warped” and $'$ implies “temporal precedent”. Still, the ZNCC metric is used in extracting these correspondences. \mathbf{B} denotes the indicator function and checks if a pixel is on the static background, i.e., $\mathbf{B}(p_l) = 0$ if p_l locates on moving objects. s is the global translation between \mathbf{I}_l and $\hat{\mathbf{I}}'_l$, which can be calculated from the global homography between \mathbf{I}_r and its temporal precedent image \mathbf{I}'_r .

Fig. 5 shows an example of the performance improvement by adding the temporal stability term. In this case, the

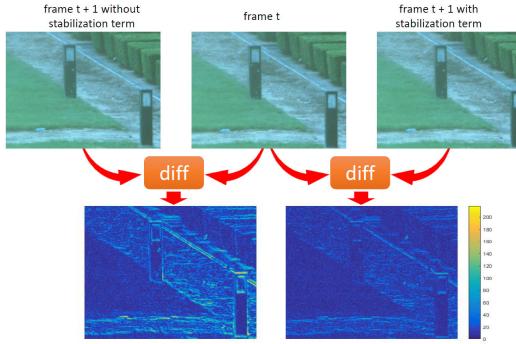


Figure 5: Illustration of the effectiveness of the stabilization term.

local-view camera is static, we calculate the per-pixel difference map between two continuous warped frames to show its stability. With the aid of stability term, the difference becomes much smaller.

Fig. 6 shows the effectiveness of the complementary optical flow correspondences. As shown in Fig. 6(b), because no matching features are found near the boundary, some mis-alignments (red ellipses) appear between the warped detail video and the reference block. After enriching matching points using complementary optical flow, the detail video and reference block are aligned much better, see Fig. 6(c).

5.3. Post Processing

Color Alignment Spatio-temporal coherent color response over the gigapixel space is crucial to the visual perception. After non-rigid warping, a de-vignetting step and a color transfer step are sequentially applied to remove intrinsic and extrinsic color inconsistency. Severe vignetting usually exists in local-view images captured by long-focus lenses, see Fig. 7(b) for an example. For each local view image, we use the 6th order even polynomial vignetting function [12] to compensate the vignetting.

We further take advantage of the fact that all the reference blocks share the same color style, and transfer the color style of each reference block to the local-view image based on the affine color mapping model [31][35]. Because of the extensively appearing and randomly distributed overexposed objects like cars and windows of buildings, we propose a local transfer model based on an edge aware interpolator [32]. First, we uniformly sample N keypoints on each local-view image, then we generate N regular overlapping patches centered on these keypoints on both the reference and the local-view image. For each patch, we use the Monge-Kantorovich solution [30] to estimate an affine mapping model [31]. Finally, an edge aware interpolator is applied to compute a mapping model for each pixel.

Dealing with Overlaps Since our local-view cameras are moveable, overlapping regions inevitably exist among local-view cameras. To cope with possible misalignment

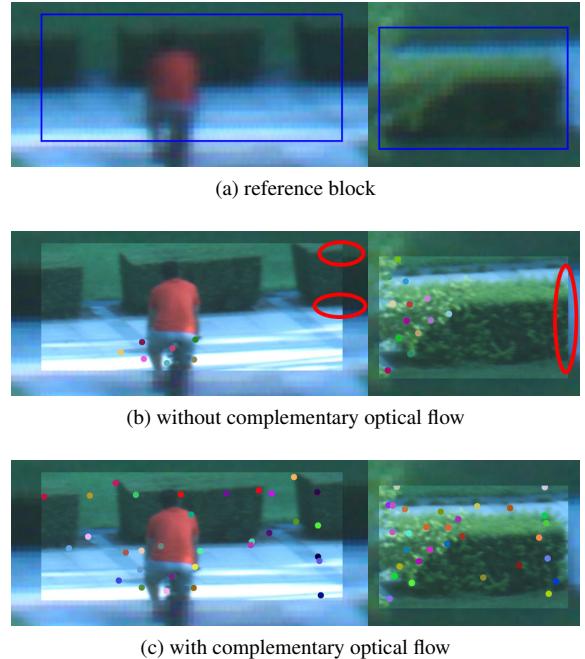


Figure 6: Effectiveness of complementary optical flow. (a) Reference block, (b) warping result without complementary optical flow, (c) warping result with complementary optical flow. Both (b) and (c) are shown by overlaying the local-view part on the reference.

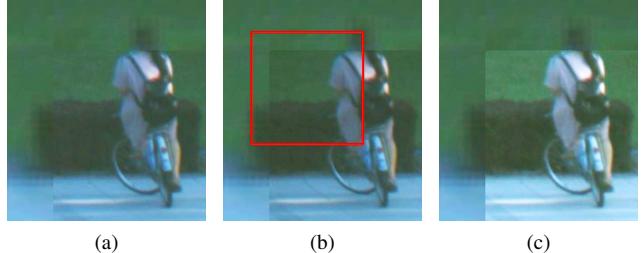


Figure 7: Color alignment. (a) With de-vignetting and color transfer, (b) with color transfer only, and (c) with de-vignetting only.

among warped local-view images, we need to optimize a stitching seam based on Graphcut [19] algorithm. To assure the capability of parallel processing on each local-view camera, we perform Graphcut on the reference blocks, instead of operation on the overlapped local-view images. Fig. 8 demonstrates the seamlessly stitching of this scheme.

6. Experiments

We implement our algorithm in C++ on an Intel Core i7-4790 CPU. Currently, without runtime optimization, the whole pipeline costs about 4 seconds to warp a local-view image on the reference video. We captured two sequences (one in summer and one in winter) with each consists of

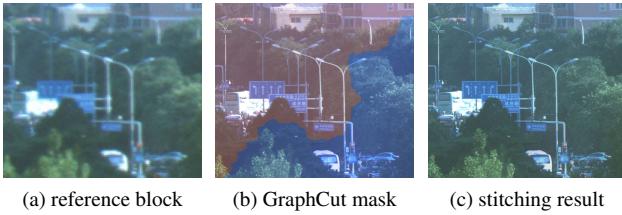


Figure 8: GraphCut on an overlap region and the stitching result. (a) Reference block, (b) Graphcut mask based on the reference block, and (c) stitching result.

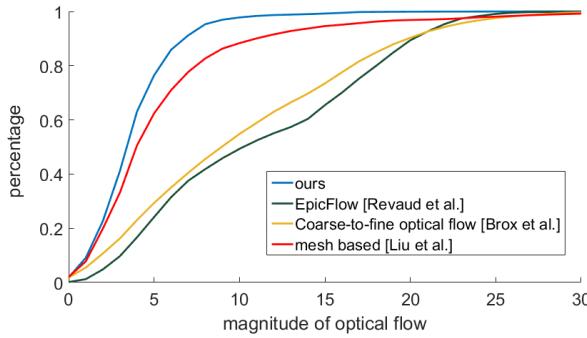


Figure 9: The distribution of pixel misalignment (calculated by optical flow) on the warped results of all the overlapping regions.

200 frames. Fig. 11 demonstrates the final gigapixel video stitching results. Four subimages are zoomed-in twice for visualization of scene details including the traffic board (more than 1 km away), car number plate (about 500 m away) and human bodies and faces (about 200 m away). Please also refer to the supplemental video for visualization of more results and comparisons.

Qualitative Evaluation We compare our warping algorithm with 4 representative warping or super-resolution algorithms, as shown in Fig. 10. From left to right: a) input; b) our result; c) EpicFlow [32] which is one of the state-of-the-art sparse to dense optical flow algorithms; d) typical coarse-to-fine variational optical flow [7]; e) PatchMatch super-resolution [3], which is the state-of-the-art example-based super-resolution method up to $8\times$; and f) global homography warping. Since we mainly care about the warping performances, we use the feature correspondences computed in iteration 1 of our pipeline as the input for f), and iteration 2 of our pipeline as the input for c) and our method.

Fig. 10 shows the comparison results. As discussed above, pixel-based warping strategies including Epicflow [32] and optical flow [7] can not preserve the image structures. Obvious visual artifacts are introduced in Fig. 10(c)(d): the fence in (d) is broken after warping, and the car window and street lamp in (c)(d) are severely distorted. The PatchMatch super-resolution algorithm fails to recover any details since the image quality of the reference

block is too poor. The baseline method global homography is insufficient to fit the non-uniform warping field, which causes misalignments and broken effects such as the walking person and the street lamps in (f). Compared with other methods, our algorithm is able to estimate an accurate warping field while preserving image structure. Qualitative comparisons with typical mesh-based warping strategy [25] are shown in the supplemental video.

Quantitative Evaluation There is no groundtruth for evaluation of the warping accuracy. However, the overlapping regions among local-view cameras provide the opportunity to evaluate the warping consistency. Because the warping of each local-view image are independent, we use optical flow algorithm [7] to compute the mismatches between these independent warping results on all the overlap regions. We calculate the optical flow magnitude distributions of our proposed method, the Epicflow [32], the optical flow [7], and the mesh-based warping [25]. For Epicflow and mesh-based warping, our matching features obtained in iteration 2 is used. Results in Fig. 9 shows that our method outperforms other methods, and has about 70% of pixels with flow vectors less than 5 pixels, which corresponds to only about 0.5 pixel mismatch on the reference video.

7. Limitations and Future Work

This paper makes an important step towards efficient gigapixel video capture. Based on our proposed multiscale hybrid camera array setting, we achieve automatic and high precision stitching of a video that can be zoomed-in for dynamic details. Low budget, small pixel bandwidth consumption, parallel processing capability and without the need of calibration are the main advantages of our system. We contribute the first warping algorithm that is robust to huge resolution gap for outdoor scene appearance. Although the summation number of all the local-view pixels cannot reach Giga level, the moveable camera setting allows us to capture region-of-interest dynamics and form a seamless video that has the feeling of gigapixel roaming.

The low resolution of scene background in the final video can be improved by adding a local-view camera for scanning the background, such that the background can be high resolution and update at a comparatively low frame rate. The exploration of a more adaptive region-of-interest capture scheme is interesting future work. A potential application of this adaptive scheme is in capturing gigapixel football game video, where the players take only small parts of the visualization space while the field remains static but takes up a major part. Undoubtedly, large scale outdoor surveillance is another important application of our system. Finally, our architecture enables parallel processing of each camera, which is promising to be optimized in realtime for online streaming and interacting applications by resorting to GPU computing techniques.

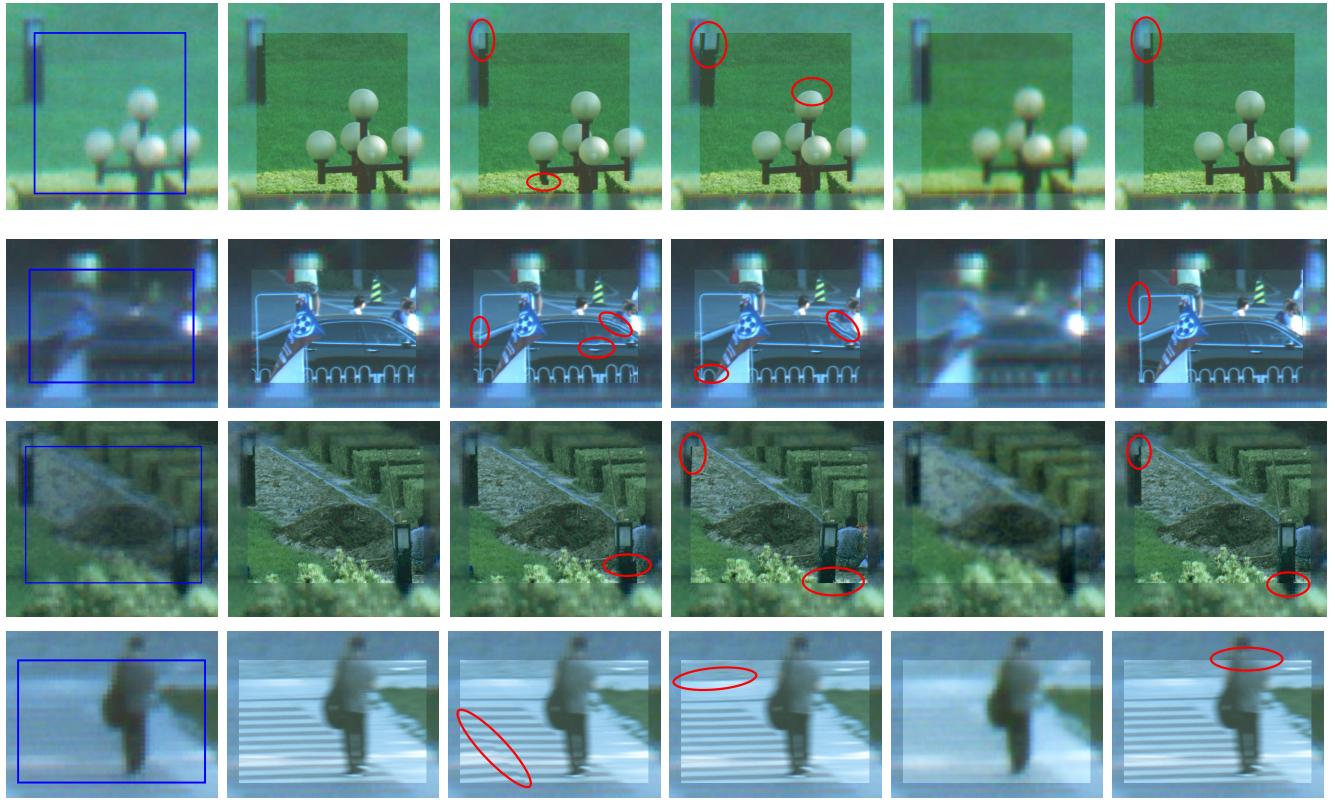


Figure 10: Comparison with available warping algorithms. Elliptical regions mark the artifacts.

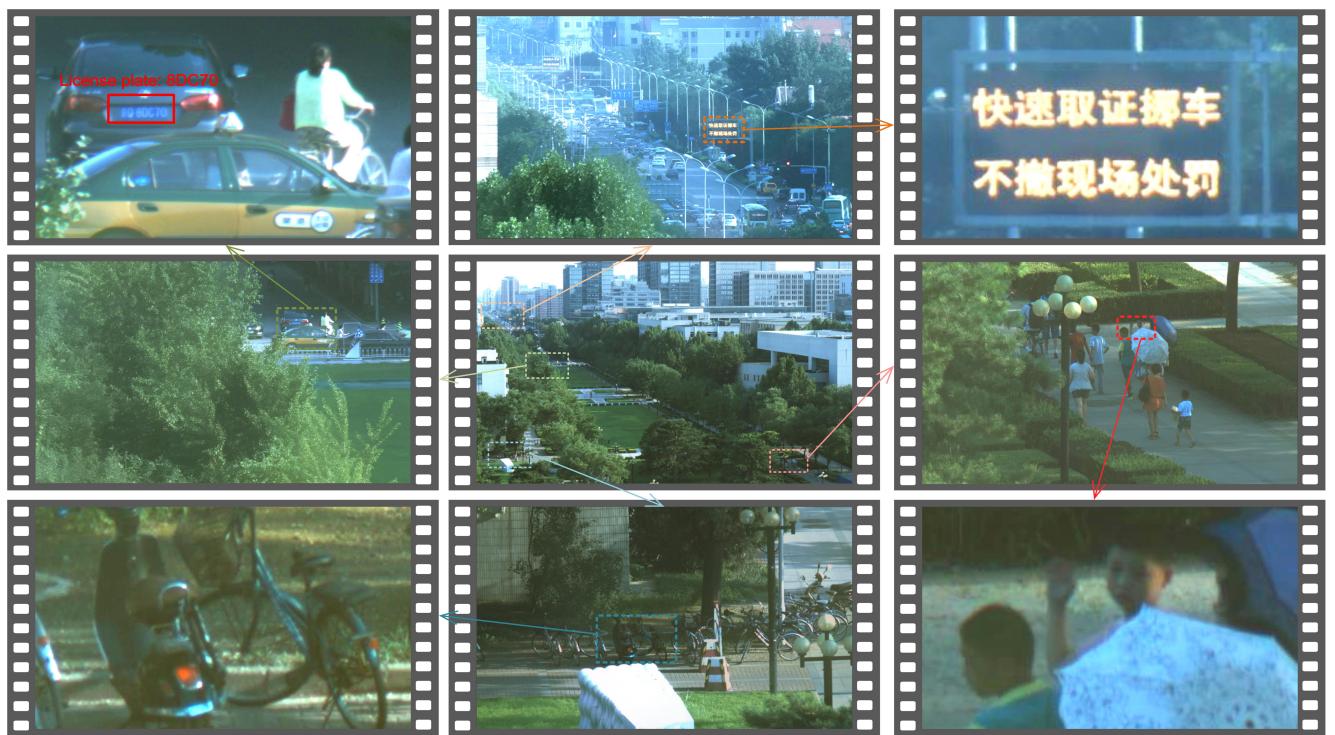


Figure 11: Zoom in the final composite gigapixel video.

8. Acknowledgment

This work was supported by the National key foundation for exploring scientific instrument No.2013YQ140517, the National NSF of China grant No.61522111, No.61531014 and the Research Grants Council (RGC) of the Hong Kong Special Administrative Region, China. (GRF 16211615).

References

- [1] Gigapan. <http://www.gigapan.com/>, 2013.
- [2] H. Bay, T.uytelaars, and L. Van Gool. Surf: Speeded up robust features. In *ECCV*, pages 404–417. Springer, 2006.
- [3] V. Boominathan, K. Mitra, and A. Veeraraghavan. Improving resolution and depth-of-field of light field cameras using a hybrid imaging system. In *IEEE ICCP*, pages 1–10, 2014.
- [4] D. Brady, M. Gehm, R. Stack, D. Marks, D. Kittle, D. Golish, E. Vera, and S. Feller. Multiscale gigapixel photography. *Nature*, 486(7403):386–389, 2012.
- [5] M. Brown and D. G. Lowe. Automatic panoramic image stitching using invariant features. *IJCV*, 74(1):59–73, 2007.
- [6] M. Brown, R. Szeliski, and S. Winder. Multi-image matching using multi-scale oriented patches. In *IEEE CVPR*, 2005.
- [7] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In *ECCV*, pages 25–36. Springer, 2004.
- [8] T. Brox and J. Malik. Large displacement optical flow: Descriptor matching in variational motion estimation. *IEEE TPAMI*, 33(3):500–513, 2011.
- [9] C.-H. Chang, Y. Sato, and Y.-Y. Chuang. Shape-preserving half-projective warps for image stitching. In *IEEE CVPR*, pages 3254–3261, 2014.
- [10] O. S. Cossairt, D. Miao, and S. K. Nayar. Gigapixel computational imaging. In *IEEE ICCP*, pages 1–8, 2011.
- [11] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.
- [12] D. B. Goldman and J.-H. Chen. Vignette and exposure calibration and compensation. In *IEEE ICCV*, 2005.
- [13] D. R. Golish, E. Vera, K. Kelly, Q. Gong, P. Jansen, J. Hughes, D. S. Kittle, D. J. Brady, and M. E. Gehm. Challenges in gigapixel multiscale image formation. In *Computational Optical Sensing and Imaging*, pages JW3A–4, 2012.
- [14] B. K. P. Horn and B. G. Schunck. Determining optical flow. *Artif. Intell.*, 17(1-3):185–203, 1981.
- [15] W. Jiang and J. Gu. Video stitching with spatial-temporal content-preserving warping. In *IEEE CVPR Workshops*, pages 42–48, 2015.
- [16] D. S. Kittle, D. L. Marks, and D. J. Brady. Automated calibration and optical testing of the aware-2 gigapixel multiscale camera. In *IS&T/SPIE Electronic Imaging*, 2013.
- [17] D. S. Kittle, D. L. Marks, H. S. Son, J. Kim, and D. J. Brady. A testbed for wide-field, high-resolution, gigapixel-class cameras. *Review of Scientific Instruments*, 84(5):053107, 2013.
- [18] J. Kopf, M. Uyttendaele, O. Deussen, and M. F. Cohen. Capturing and viewing gigapixel images. *ACM TOG*, 26(3):93, 2007.
- [19] V. Kwatra, A. Schödl, I. Essa, G. Turk, and A. Bobick. Graphcut textures: image and video synthesis using graph cuts. In *ACM ToG*, volume 22, pages 277–286, 2003.
- [20] J. Lee, B. Kim, K. Kim, Y. Kim, and J. Noh. Rich360: optimized spherical representation from structured panoramic camera arrays. *ACM TOG*, 35(4):63, 2016.
- [21] G. Levi and T. Hassner. Latch: learned arrangements of three patch codes. In *IEEE WACV*, pages 1–9, 2016.
- [22] J. Lewis. Fast normalized cross-correlation. In *Vision interface*, volume 10, pages 120–123, 1995.
- [23] H. Li, B. Adams, L. J. Guibas, and M. Pauly. Robust single-view geometry and motion reconstruction. *ACM Trans. Graph.*, 28(5):175:1–175:10, 2009.
- [24] F. Liu, M. Gleicher, H. Jin, and A. Agarwala. Content-preserving warps for 3d video stabilization. In *ACM TOG*, volume 28, page 44, 2009.
- [25] S. Liu, L. Yuan, P. Tan, and J. Sun. Bundled camera paths for video stabilization. *ACM ToG*, 32(4):78, 2013.
- [26] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.
- [27] J. Nichols, K. Judd, C. Olson, K. Novak, J. Waterman, S. Feller, S. McCain, J. Anderson, and D. Brady. Range performance of the darpa aware wide field-of-view visible imager. *Applied Optics*, 55(16):4478–4484, 2016.
- [28] F. Perazzi, A. Sorkine-Hornung, H. Zimmer, P. Kaufmann, O. Wang, S. Watson, and M. Gross. Panoramic video from unstructured camera arrays. In *CGF*, volume 34, pages 57–68, 2015.
- [29] S. Pirk, M. F. Cohen, O. Deussen, M. Uyttendaele, and J. Kopf. Video enhanced gigapixel panoramas. In *ACM SIGGRAPH Asia Technical Briefs*, pages 7:1–7:4, 2012.
- [30] F. Pitié and A. Kokaram. The linear monge-kantorovitch linear colour mapping for example-based colour transfer. In *CVMP*, pages 1–9. IET, 2007.
- [31] F. Pitié, A. Kokaram, and R. Dahyot. Enhancement of digital photographs using color transfer techniques. *Single-Sensor Imaging: Methods and Applications for Digital Cameras (R. Lukac, ed.)*, pages 295–321, 2008.
- [32] J. Revaud, P. Weinzaepfel, Z. Harchaoui, and C. Schmid. Epicflow: Edge-preserving interpolation of correspondences for optical flow. In *IEEE CVPR*, 2015.
- [33] B. Ristic, S. Arulampalam, and N. Gordon. *Beyond the Kalman filter: Particle filters for tracking applications*, volume 685. Artech house Boston, 2004.
- [34] H. S. Sawhney, Y. Guo, K. J. Hanna, R. Kumar, S. Adkins, and S. Zhou. Hybrid stereo camera: an IBR approach for synthesis of very high resolution stereoscopic image sequences. In *SIGGRAPH*, pages 451–460, 2001.
- [35] Y. Shih, S. Paris, F. Durand, and W. T. Freeman. Data-driven hallucination of different times of day from a single outdoor photo. *ACM ToG*, 32(6):200, 2013.
- [36] R. Szeliski. Image alignment and stitching: A tutorial. *Foundations and Trends in Computer Graphics and Vision*, 2006.
- [37] E. Tola, V. Lepetit, and P. Fua. DAISY: An Efficient Dense Descriptor Applied to Wide Baseline Stereo. *IEEE TPAMI*, 32(5):815–830, May 2010.
- [38] A. Tommaselli, J. Marcato Jr, M. Moraes, S. Silva, and A. Artero. Calibration of panoramic cameras with coded targets and a 3d calibration field. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40(3):137, 2014.
- [39] Y. Wang, Y. Liu, W. Heidrich, and Q. Dai. The light field attachment: Turning a dslr into a light field camera using a low budget camera ring. *IEEE Trans. Visualization and Computer Graphics*, 2017.
- [40] P. Weinzaepfel, J. Revaud, Z. Harchaoui, and C. Schmid. Deepflow: Large displacement optical flow with deep matching. In *IEEE ICCV*, pages 1385–1392, 2013.
- [41] B. Wilburn, N. Joshi, V. Vaish, E.-V. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy. High performance imaging using large camera arrays. *ACM TOG*, 24(3):765–776, 2005.
- [42] L. Xu, J. Jia, and Y. Matsushita. Motion detail preserving optical flow estimation. *IEEE TPAMI*, 34(9):1744–1757, 2012.
- [43] C. Zach, T. Pock, and H. Bischof. A duality based approach for realtime tv- L^1 optical flow. In *DAGM*, pages 214–223, 2007.
- [44] J. Zaragoza, T.-J. Chin, M. Brown, and D. Suter. As-projective-as-possible image stitching with moving dlt. In *IEEE VCPR*, pages 2339–2346, 2013.