Structure-from-motion with varying principal point

W. A. P. Smith, P. Lewińska, M. A. Cooper, E. R. Hancock, *Fellow, IEEE*, J. A. Dowdeswell, and D. M. Rippin

***Abstract*—We consider the problem of structure-from-motion (SfM) for images with fixed calibration but varying principal point. This scenario occurs for archival imagery taken using historic glass plate and film cameras without fiducial markers, when images have been inconsistently cropped or when image plates are broken into multiple fragments. We derive initialisation and pose estimation methods and regularisation penalties tuned specifically for this scenario leading to a complete archival imagery SfM pipeline. We illustrate the performance of our methods on challenging real world examples from image archives. Specifically, we use archival images of the East coast of Greenland from the British Arctic Air Route Expedition (BAARE). This is of particular glaciological interest for measuring historic ice loss. We use a modern digital elevation model (ArcticDEM), masked to stable regions, as ground truth to evaluate our method.**

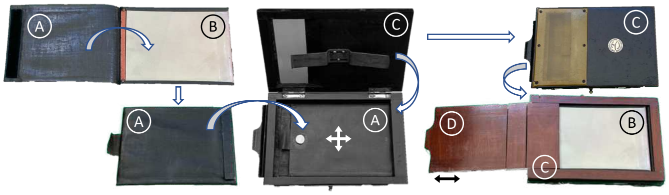

Fig. 1. Motivating scenario. To enable changing plates in daylight, a glass plate (B) is placed in a fabric envelope (A). The envelope is placed inside an envelope adaptor (C). When the adaptor is attached to the camera, the draw-slide (D) is pulled open, allowing the plate to be exposed. The plate can move within the adaptor (middle) and the draw-slide may not be pulled fully open. The exposed area is therefore bounded by the envelope cut-out on three sides and the draw-slide on the other, none of which are fixed.

I. Introduction

Photogrammetry has long provided valuable data for Earth surface analysis. Examples include landslide and snow cover monitoring, and glacier terminus and surface evolution [1], [2]. Aerial photogrammetry is usually favoured due to the relatively large areas that can be covered in a short time, thus allowing the study of dynamic changes [2]. Especially for environmental studies, a geolocalisation is required, and this proves to be difficult with historic images since the terrain has often eroded and there is not enough high quality data available for the creation of reliable ground control points (GCPs) or reference model for alignment [2]–[4]. Also, the change of capture medium from glass plates to film and then to digital and changes in the construction of the camera itself brings many challenges for accurate 3D reconstruction.

Historical images, archival, analogue photographs present several challenges such as low quality, various state of original medium and complex and unknown processing pipeline from capture to digitisation that could alter the geometry. Images were taken with or without fiducials (markers on the image used for the calculation of internal camera calibration parameters [5]). Where available, fiducials can be used to correct for film shrinkage or linear deformation from image scanning. Alternatively, if the intrinsics are not known but all images were taken by the same camera, then the scanned images can be cropped and moved so that they retain the same size after the scanning procedure [5]. Even though some approaches rely on removing fiducials [5], images that do not have any intrinsic information are harder to process in a traditional way [3]. State-of-the-art SfM methods [6] perform self-calibration, but in practice this procedure can be unstable [7]. Making use of known camera calibration parameters is thus to be preferred.

Pose estimation plays a key role in SfM for the purpose of aligning images to the reconstructed model. Perspective-n-point (PnP) methods differ in the minimum number of points, n, required and whether they can self-calibrate for focal length, f, or principal point (PP), uv. In the context of our scenario, some relevant uncalibrated pose estimation methods have been proposed. An uncalibrated version of EPnP [8], known as UPnP [9], computes both pose and focal length. Minimal methods exist for P4Pf with unknown focal length [10] and P4.5Pfuv with both unknown focal length and PP [11].

Scenario. We consider a camera with fixed parameters that captures multiview images of a scene. However, each image is subject to a different unknown, image plane transformation and, or cropping. This amounts to assuming fixed focal length and distortion parameters through the sequence but varying PP. This occurs when images are arbitrarily cropped or with archival images when the physical medium moves within the camera (Fig. 1) or scanner and there are no fixed reference fiducial markers that can be used to compensate.

Contributions First, we identify a new problem with wide applicability (Sec. II). Second, we introduce the PnPuv problem and propose a first solution (Sec. III). Third, we propose the first SfM pipeline for this scenario, including a novel exposed area constraint that makes PP estimates between images non-independent (Sec. III). Fourth, we describe a variant of our pipeline that can handle fragmented images and implicitly reassemble the fragments (Sec. IV). Finally, we find a real world application scenario in which our method is applicable and outperforms standard pipelines. To the best of our knowledge the case of unknown PP with known focal length has not been previously studied. Of course, modern SfM pipelines such as COLMAP [12] can optimise PP position during bundle adjustment. However, no existing tools provide the option to share all intrinsic parameters across images while allowing PP to vary between images.

W. Smith, P. Lewińska and E. Hancock are with the Department of Computer Science and M. Cooper and D. Rippin with the Department of Environment and Geography, University of York, UK e-mail: william.smith@york.ac.uk. J. Dowdeswell is with the Scott Polar Research Institute, University of Cambridge, Cambridge, UK

Manuscript received April 19, 2005; revised September 17, 2014.

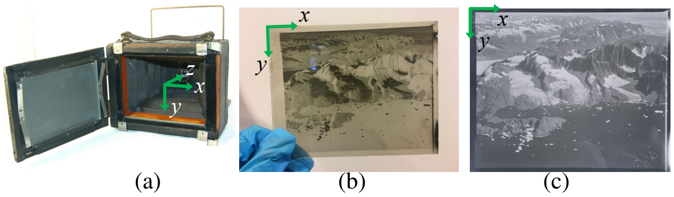

Fig. 2. Coordinate systems: (a) camera, (b) photographic medium, (c) cropped digital image.

II. Problem scenario

We model the physical camera as a classical pinhole with fixed intrinsic parameters (we additionally model nonlinear distortion during bundle adjustment):

$$\lambda \begin{bmatrix} \mathbf{u}_{ij}^{\text{cam}} \\ 1 \end{bmatrix} = \begin{bmatrix} f_{\text{mm}} & 0 & 0 \\ 0 & f_{\text{mm}} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{R}_i & \mathbf{t}_i \end{bmatrix} \begin{bmatrix} \mathbf{x}_j \\ 1 \end{bmatrix}, \quad (1)$$

where λ is a scale, f_{mm} is the physical focal length in world units (i.e. millimetres), $\mathbf{R}_i \in SO(3)$ and $\mathbf{t}_i \in \mathbb{R}^3$ the pose of the ith camera that transforms from world to camera coordinates, $\mathbf{x}_j = [x_j, y_j, z_j]^\top$ the jth world point and $\mathbf{u}_{ij}^{\text{cam}} = [u_{ij}, v_{ij}]^\top$ the corresponding projection into the ith image, i.e. camera coordinates in world units (see Fig. 2(a)).

The image is captured on a photographic medium which can move between exposures (we assume a 2D rigid transformation). The coordinate system with respect to the medium (e.g. measured from the upper left corner of the plate - see Fig. 2(b)) is different for each image and defined as: $\mathbf{u}_{ij}^{\text{media}} = \begin{bmatrix} \mathbf{R}_i^{\text{media}} & \mathbf{t}_i^{\text{media}} \end{bmatrix} [\mathbf{u}_{ij}^{\text{cam}\top}, 1]^\top$, where $\mathbf{R}_i^{\text{media}} \in SO(2)$ and $\mathbf{t}_i^{\text{media}} \in \mathbb{R}^2$ define a 2D rigid transformation of the medium within the camera. This coordinate system is still in world units, e.g. millimetres.

Finally, physical media is digitised via scanning and manually cropped (Fig. 2(c)). This change in coordinate system amounts to a scale (world units to pixels) and translation (a rotation could also be included here without changing the final model but usually the medium can be aligned to the edge of the scanner in a consistent manner): $\mathbf{u}_{ij} = s_i(\mathbf{u}_{ij}^{\text{media}} + \mathbf{t}_i^{\text{scan}})$, with $\mathbf{u}_{ij}$ now in units of pixels and s_i is the scanning resolution in units of pixels per millimetre, known from metadata.

Combining the above series of transformations, we note that $\mathbf{R}_i^{\text{media}}$ can be factored into $\mathbf{R}_i$, i.e. we compensate for image plane rotation via camera pose, and combine the translations to form the effective centre of projection:$\lambda[s_i^{-1}\mathbf{u}_{ij}^\top, 1]^\top = \mathbf{K}_i \begin{bmatrix} \mathbf{R}_i & \mathbf{t}_i \end{bmatrix} [\mathbf{x}_j^\top, 1]^\top$, where $\mathbf{K}_i$ is the intrinsic matrix with focal length f_{mm} and PP $[u_{i0}, v_{i0}]^\top = \mathbf{t}_i^{\text{scan}} + \mathbf{t}_i^{\text{media}}$. With s_i assumed known, this amounts to a scenario of fixed intrinsics (focal length) but varying PP, i.e. we have (u_{i0}, v_{i0}) unknown per image. This model is applicable to scanned physical media or digital camera images with inconsistent crops.

III. Archival image SfM pipeline

We assume that focal length remains fixed throughout the image sequence. Where a geolocated model is required, we assume that sparse GCPs are available. Feature extraction and matching is identical to traditional SfM pipelines. However, for initialisation, pose estimation during incremental SfM and bundle adjustment we propose specialised methods to account for per-image varying PP.

Feature extraction and matching For all images we extract SIFT features [13], [14]. We perform brute force matching and discard ambiguous matches using Lowe's ratio test [13] with a ratio of 1.5. Finally, we perform geometric verification by fitting a fundamental matrix with MLESAC [15] and discarding image pairs with fewer than 30 inlying matches.

Initialisation In modern SfM pipelines, intrinsics are usually initialised by reading focal length in millimetres and the camera model from image meta data, allowing a good initial estimate for the focal length in pixels. In the archival scenario, contemporary logs can play the same role: often the focal length was recorded. In addition, either the physical size of the plate or film is known or, if a flat bed scanner was used, the scanning resolution in real world units is known. Again, this enables a good initialisation for the focal length in pixels.

We initialise camera pose for the first two images using GCPs. We select the pair of images with maximum mutually visible GCPs and then solve for pose using the known focal length and the given 2D/3D GCP correspondences with our PnPuv method in Sec. III. We triangulate matched features between the initial pair, initialise distortion parameters to zero and run an initial bundle adjustment (see Sec. III).

PnPuv A key component of an SfM pipeline is an absolute pose solver to align new views to the current reconstruction. With unknown PP this *PnPuv problem* contains 8 unknowns (6 for pose, 2 for PP) and so, in principal, a minimal P4Puv solution could be derived, e.g. by adapting [11]. However, a good initialisation/prior is available for the PP (the image centre) which would not normally be true for focal length in PnPf. Also, in the archival scenario, GCPs may be available, manual outlier removal may be viable and speed is not a concern. Hence, we do not seek a minimal solution but instead a least squares solution over all points which can optionally incorporate regularisation of the PP according to a prior.

We show how to write PnPuv as separable nonlinear least squares [16], i.e. a form that is linear in some parameters and nonlinear in the rest. This means that the optimal PP can be implicitly solved for using linear least squares and only requires nonlinear optimisation over pose. We denote by: $\mathbf{u}(\mathbf{x}, f, \mathbf{R}, \mathbf{t}) = \begin{bmatrix} \frac{f(\mathbf{r}_1\mathbf{x}+t_1)}{\mathbf{r}_3\mathbf{x}+t_3} & \frac{f(\mathbf{r}_2\mathbf{x}+t_2)}{\mathbf{r}_3\mathbf{x}+t_3} \end{bmatrix}^\top$, the perspective projection of $\mathbf{x}$ without accounting for the PP. Each correspondence $(\mathbf{u}_j, \mathbf{x}_j)$ gives us a pair of equations of the form: $\mathbf{u}(\mathbf{x}_j, f, \mathbf{R}, \mathbf{t}) + \mathbf{u}_0 = \mathbf{u}_j$. Note that this equation is linear in $\mathbf{u}_0$, but nonlinear in $\mathbf{t}$ and any parameterisation of $\mathbf{R}$. Hence, stacking the equations for all points and introducing a PP prior, we can write it in the form $\mathbf{A}\mathbf{u}_0 = \mathbf{d}(\mathbf{R}, \mathbf{t})$, where:

$$\mathbf{A} = \begin{bmatrix} \mathbf{1}_{n\times 1} \otimes \mathbf{I}_2 \\ w_{\text{PP}}\mathbf{I}_2 \end{bmatrix}, \quad \mathbf{d}(\mathbf{R}, \mathbf{t}) = \begin{bmatrix} \mathbf{u}_1 - \mathbf{u}(\mathbf{x}_1, f, \mathbf{R}, \mathbf{t}) \\ \vdots \\ \mathbf{u}_n - \mathbf{u}(\mathbf{x}_n, f, \mathbf{R}, \mathbf{t}) \\ w_{\text{PP}}\mathbf{u}_0^{\text{init}} \end{bmatrix}, \quad (2)$$

$\otimes$ is the Kronecker product, $\mathbf{u}_0^{\text{init}}$ is the PP prior (usually set to the image centre) and w_{PP} is the prior weight (set to

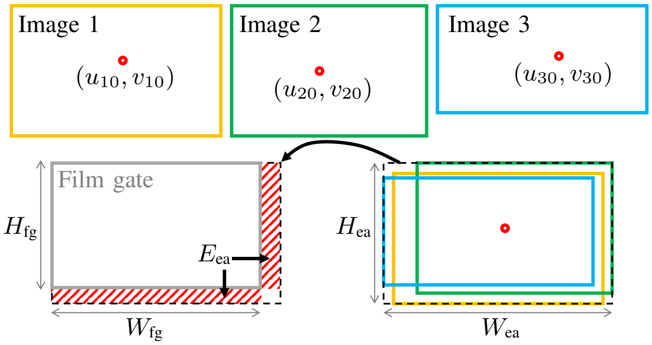

Fig. 3. Exposed area constraint. Aligning images (top) by their estimated centres of projection defines a bounding box (bottom right, dashed) which should lie within the physical film gate (bottom left, gray) and we penalise the additional area (red stripes).

zero for an unregularised solution). The linear least squares solution for $\mathbf{u}_0$ is given by $\mathbf{A}^+\mathbf{d}$ where $\mathbf{A}^+$ is the pseudoinverse of $\mathbf{A}$ which does not depend on any optimisation variables and has a very simple closed form: $\mathbf{A}^+ = 1/(w_{\text{PP}}^2 + n)\begin{bmatrix}\mathbf{1}_{1\times n}\otimes\mathbf{I}_2 & w_{\text{PP}}\mathbf{I}_2\end{bmatrix}$. Substituting the optimal PP into (2) we can write a nonlinear least squares problem in terms of only pose: $\min_{\mathbf{R},\mathbf{t}} \|\mathbf{A}\mathbf{A}^+\mathbf{d}(\mathbf{R},\mathbf{t}) - \mathbf{d}(\mathbf{R},\mathbf{t})\|_2^2$. We solve this optimisation using Levenberg-Marquardt and parameterise $\mathbf{R}$ using axis-angle. We initialise using EPnP [8] with the PP assumed to be at the image centre.

Exposed area constraint Variations in PP position between images imply an alignment of the images by translating to align their PPs. This provides a constraint. The *film gate* is the frame in front of the film or glass plate which has the dual purpose of holding the film in place and letting light through. The film gate has fixed area. Hence, once the images have been aligned, the total area of the exposed pixels cannot exceed the film gate area. Alignment by per-image PP implies an exposed area (see Fig. 3) with width and height given by: $W_{\text{ea}} = \max_i(u_{i0}) + \max_i(W_i - u_{i0})$ and $H_{\text{ea}} = \max_i(v_{i0}) + \max_i(H_i - v_{i0})$, where W_i and H_i are the width and height of the ith image. For film gate with dimensions $W_{\text{fg}} \times H_{\text{fg}}$ we define an exposed area loss which penalises the exposed area having larger dimensions than the measured film gate:

$$E_{\text{ea}} = \max(0, W_{\text{ea}} - W_{\text{fg}})H_{\text{ea}} + \max(0, H_{\text{ea}} - H_{\text{fg}})W_{\text{ea}}. \quad (3)$$

Bundle adjustment and incremental SfM We perform incremental SfM. For each new view we initialise by solving PnPuv using any keypoints that have been reconstructed previously. We initialise new 3D scene points by triangulating any keypoints seen for the second time. Finally, we perform bundle adjustment over all views. We repeat this process until finally performing bundle adjustment over the entire dataset.

Our bundle adjustment procedure solves the following optimisation problem: $\min_\Theta E_{\text{reproj}} + w_{\text{ea}}E_{\text{ea}} + w_{\text{f\_prior}}E_{\text{f\_prior}}$, where $\Theta = (\theta, \{\theta_i\}_i, \{\mathbf{x}_j\}_j)$, with per-dataset parameters $\theta = (f_{\text{mm}}, k_{1,2,3}, p_{1,2})$ (in which $k_{1,2,3}$ and $p_{1,2}$ are nonlinear distortion parameters), per-image parameters $\theta_i = (\mathbf{R}_i, \mathbf{t}_i, \mathbf{u}_{i0})$ and 3D scene points $\mathbf{x}_j$. All the loss functions in our objective are sums of squared quantities and we solve the resulting nonlinear least squares problem using the Levenberg-Marquardt algorithm. The reprojection error is given by:

$$E_{\text{reproj}} = \sum_i \sum_{j\in\mathcal{V}_i} \|\mathbf{u}_{ij} - \boldsymbol{\pi}(\theta, \theta_i, \mathbf{x}_j)\|_2^2, \quad (4)$$

where $\mathcal{V}_i$ is the set of keypoints visible in image i and $\boldsymbol{\pi}(\theta, \theta_i, \mathbf{x}_j)$ performs perspective projection (with radial and tangential distortion) of $\mathbf{x}_j$. We normalise the reprojection error by the number of keypoints in each image. The focal length prior encourages the estimated focal length to stay close to the initial physical focal length $f_{\text{mm}}^{\text{init}}$: $E_{\text{f\_prior}} = \left\|f_{\text{mm}} - f_{\text{mm}}^{\text{init}}\right\|_2^2$. We set $w_{\text{ea}} = 0.1$ and $w_{\text{f\_prior}} = 0.001$ empirically and use these weights in all experiments.

Implementation Our pipeline is implemented in Matlab. Our sparse reconstruction, estimated camera intrinsics and extrinsics and normalised images with embedded camera meta data can be exported to existing dense reconstruction tools [17].

IV. Image Fragments

Our pipeline can be applied, with slight modification, to a scenario in which one or more of the images in the sequence has been fragmented into parts. This could occur, for example, due to a photographic glass plate being broken or a photographic print being torn into pieces. In this case, the correct arrangement of the fragments into a single image requires estimation of a 2D rigid transformation for each fragment relative to one chosen reference fragment. With real fragmented images there is usually some image missing along the fragment boundaries and so they cannot be used to reassemble the fragments. Instead, we use 3D geometric scene information to resolve the 2D placement of fragments.

We assume that the ith image has been fragmented into n_i parts where $n_i = 1$ for images that are complete. We partition the 2D keypoints associated with the ith image into non-overlapping sets of keypoints, one for each fragment: $\mathcal{F}_1^i, \dots, \mathcal{F}_{n_i}^i$. Image matching and geometric verification is applied independently to each fragment. We choose as the reference fragment the one with the highest number of matches to previously reconstructed points. To initialise, we solve PnPuv for keypoints in this fragment alone using the method in Section III. Then, for each additional fragment we keep pose and intrinsics fixed and compute the optimal 2D rigid transformation for that fragment in closed form.

During bundle adjustment, we use a modified reprojection error that incorporates per-fragment transformation:

$$E_{\text{reproj}} = \sum_{k=1}^{n_i} \sum_{j\in\mathcal{F}_k^i} \left\|\mathbf{R}_{\mathcal{F}_k^i}\mathbf{u}_{ij} + \mathbf{t}_{\mathcal{F}_k^i} - \boldsymbol{\pi}(\theta, \theta_i, \mathbf{x}_j)\right\|_2^2,$$

where $\mathbf{R}_{\mathcal{F}_k^i} \in SO(2)$ and $\mathbf{t}_{\mathcal{F}_k^i} \in \mathbb{R}^2$ are a 2D rotation and translation respectively. Assuming the first fragment is the reference to which other fragments are aligned then $\mathbf{R}_{\mathcal{F}_1^i} = \mathbf{I}_3$ and $\mathbf{t}_{\mathcal{F}_1^i} = \mathbf{0}$. Hence, each additional fragment in each image adds 3 unknowns to the optimisation (rotation angle and 2D translation). Finally, we require an additional constraint to avoid fragments overlapping each other. We assume that each fragment is scanned separately and that a per-pixel mask indicates which image pixels belong to the fragment (Fig. 6)

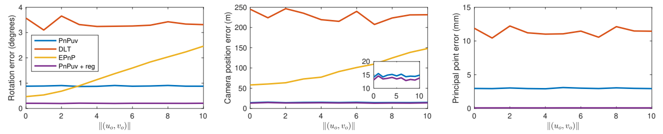

Fig. 4. Quantitative PnP results with unknown PP and ablation of PP prior. We show median errors over 1000 runs. EPnP does not estimate PP so is not included in right hand plot.

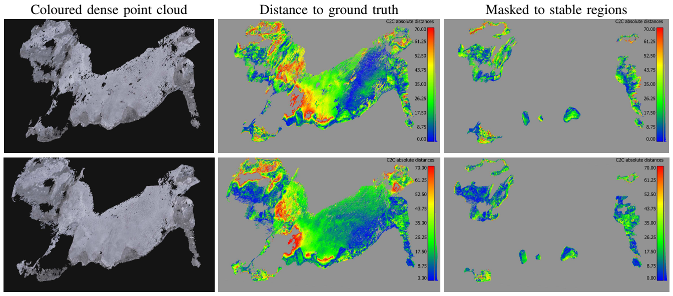

Fig. 5. Qualitative results for region 1. Top row: baseline method, bottom row: proposed pipeline. The middle column shows the point to point distance to a modern reference DEM. The right column shows the same but cropped to 'stable regions'. Errors over 70m are clipped.

Method	Input	Geoloc.	E_{geo} (m) mean / max	E_{pp} (pixels)
[17]	Orig	GCP	0.046 / 0.324	0.00*
[12]	Orig	"GPS"	0.236 / 1.100	47.9
Ours	Crop	GCP	0.125 / 1.24	73.2 ± 6.5
Ours w/o E_{ea}	Crop	GCP	0.134 / 1.100	194.9 ± 20.2
[17]	Crop	N/A	Failed to reconstruct	
[12]	Crop	N/A	Failed to reconstruct	
Ours + [17]	Crop	"GPS"	0.091 / 0.999	N/A†
Ours + [12]	Crop	"GPS"	0.218 / 0.837	N/A†

TABLE I
QUANTITATIVE RESULTS FOR ZAGÓRZ DATASET.

	Method	Change (m)	Error (m)
Region 1	Baseline	19.4 ± 41.5	12.2 ± 33.3
	Proposed	15.6 ± 39.6	6.65 ± 27.6
Region 2	Baseline	30.4 ± 66.3	26.5 ± 64.3
	Proposed	10.6 ± 41.1	8.37 ± 38.0

TABLE II
QUANTITATIVE RESULTS FOR ARCHIVAL IMAGE DATA.

such that $\mathcal{M}_k^i$ is the set of pixels in the mask for the k-th fragment in the i-th image. For each fragment mask, we compute the distance transform of the compliment of the mask (i.e. it has zero value outside the fragment and increases with distance inside the fragment) see Fig. 6 for an example. This distance map, $d_k^i(\mathbf{u})$, can be evaluated at any (non-integer) position $\mathbf{u}$ via differentiable bilinear interpolation. We can now measure interpenetration between aligned fragments:

$$E_{\text{overlap}} = \sum_i \sum_{k=1}^{n_i} \sum_{l=k+1}^{n_i} \sum_{\mathbf{u} \in \mathcal{M}_l^i} d_k^i \left[\mathbf{R}_{\mathcal{F}_k^i}^{-1} \left(\mathbf{R}_{\mathcal{F}_l^i} \mathbf{u} + \mathbf{t}_{\mathcal{F}_l^i} - \mathbf{t}_{\mathcal{F}_k^i} \right) \right]^2 .$$

This additional cost pushes fragments apart when they overlap and is added to the overall bundle adjustment objective.

V. Experiments

PnPuv We begin by quantitatively evaluating proposed PnPuv method on synthetic data using the same parameters as our real data (focal length 7in, scene scale ranging 1km-10km, image size 5in × 4in). We generate 8 random 3D points in the field of view of the camera and project to 2D before adding Gaussian noise of standard deviation 1mm. We examine how pose and PP position estimation behaves as the true PP is varied up to 10mm from the centre of the image. While no existing methods solve the PnPuv problem, the classical direct linear transform (DLT) can be used for this purpose. We use the DLT to estimate the camera matrix then decompose to $\mathbf{P} = \mathbf{K}[\mathbf{R}\ \mathbf{t}]$ via the RQ decomposition [18], providing a PP estimate in $\mathbf{K}$. We also compare against EPnP [8] when assuming the PP is the image centre. We perform an ablation on the PP prior, comparing $w_{\text{PP}} = 0$ and $w_{\text{PP}} = 2$. For our method pose accuracy does not degrade as the PP moves away from the image centre while the PP itself is accurately estimated (Fig. 4). Using the PP prior always improves performance.

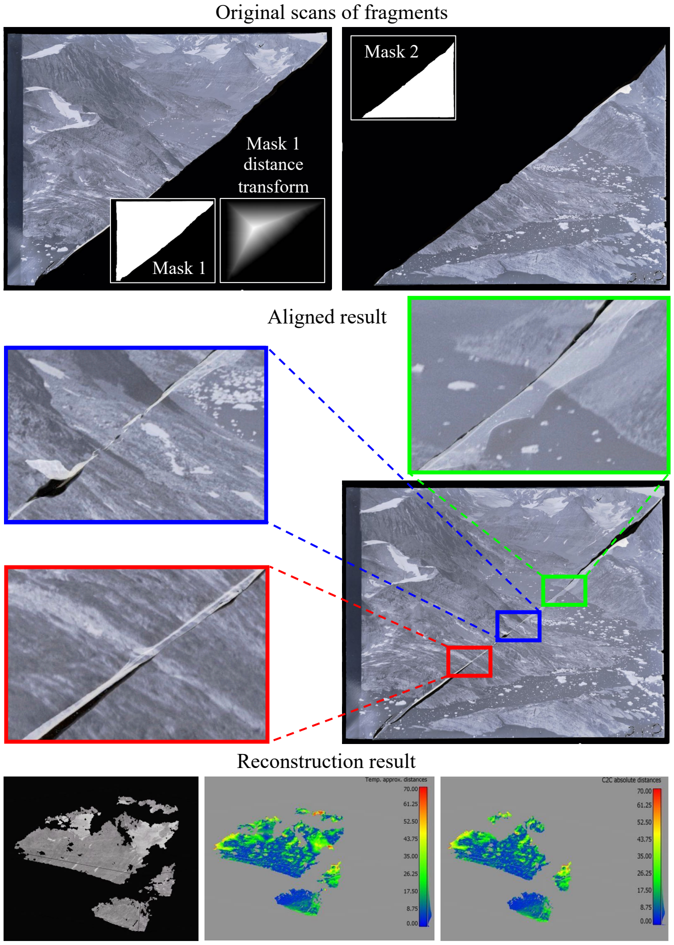

Fig. 6. Top: original scans of image fragments and (inset) pixel-wise masks and distance transform for mask 1. Middle: alignment using the estimated 2D rigid transformation for the second fragment with zoomed regions illustrating good alignment of features that cross the crack. Bottom: Results for three image sequence containing cracked image. Heat maps show distance to ground truth for whole model (middle) and stable regions only (right).

SfM In Table I we begin with a quantitative evaluation on the Zagórz dataset [19] (see supplementary material) using either the original uncropped images (*Orig*) or cropped images (*Crop*) as shown in supplementary material. To geolocate the point clouds to ground truth, we either use GCPs (*GCP*) or the camera positions estimated by our method as pseudo GPS locations (*"GPS"*). We compare against Agisoft metashape [17] and COLMAP [12]. We treat the [17] result on uncropped images as pseudo ground truth for PP (*). Results labelled † use the same PP estimates as *Ours*. We also provide an ablation study removing the exposed area constraint (*Ours w/o* E_{ea}).

Next we use the 1931 British Arctic Air Route Expedition (BAARE) dataset consisting of aerial oblique images taken along the East coast of Greenland using a Williamson P14 camera with a 7in (177.8mm) focal length lens and 5in × 4in glass plates. For ground truth we use the modern (2015-2018) ArcticDEM [20]. We process image sequences covering two regions: the Hutchinson glacier (region 1, 11 images) and an unnamed glacier at W31°49′18.84″, N68°5′41.64″ (region 2, 13 images). For comparison, our baseline method is to run Agisoft Metashape [17], allowing all camera parameters to vary between images. We show results in Fig. 5 and Table II, where we show mean ± standard deviation of point to point distance between our reconstructed models and ground truth DEM. *Change* denotes distances over the whole model giving the overall degree of change while *Error* is calculated only over stable regions and can be used to quantify the accuracy of the model. When using stable regions (areas that do not include changeable glaciers or sea level) to compare against ground truth, we achieve a reduction in mean error of 45% and 68% respectively. Our result for non-stable regions, that will be used in the future for evaluating glacier change between 1931 and current times, show more uniform change across the ice covered areas and better highlights the change in ice covered areas in region 2. The estimated camera locations place the aeroplane altitude at an average of 2,372m and 2,851m for the two regions - plausible given the typical 10,000ft (3,048m) reported in contemporary logs [21].

Finally, we show a three image sequence in which the glass plate for the middle image was broken in half. Fig. 6 top and middle shows the processing pipeline and illustrates the result of aligning the bottom fragment to the top. Note that linear features on the terrain that cross the crack boundary appear well aligned in the zoomed regions. Fig. 6 bottom shows a reconstruction result from a three image sequence including the cracked image. This is highly challenging due to the extreme viewpoint change, sparse regions of correspondence (approximately indicated by bounding boxes) and the cracked middle image. We are still able to achieve average error of 5.13 ± 14.8m over stable regions. The baseline method fails on this sequence. With the fragments not precisely aligned, geometric verification fails to find inliers in both fragments. Treating each fragment as a separate image also fails.

VI. Conclusions

In this paper we have presented a SfM pipeline that is specifically adapted to work with archival photographs. Most importantly, this deals with the motion of photographic media within the camera but also can handle images that have been fragmented into parts. We demonstrated that our approach yields significantly more accurate reconstructions on challenging real world data of glaciological interest. Besides application to other historical datasets, there are several interesting avenues for future technical work. For cracked images, a textured dense model may provide a model-based approach to inpainting the missing regions in the cracked image. The laborious task of labelling GCPs could be automated by solving a very challenging matching task between modern and archival images.

Acknowledgement This work is funded by Leverhulme Trust Research Grant RPG-2017-346.

References

[1] D. Li, O. Wigmore, M. T. Durand, B. Vander-Jagt, S. A. Margulis, N. P. Molotch, and R. C. Bales, "Potential of balloon photogrammetry for spatially continuous snow depth measurements," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 10, pp. 1667–1671, 2020.

[2] N. Midgley and T. Tonkin, "Reconstruction of former glacier surface topography from archive oblique aerial images," *Geomorphology*, vol. 282, pp. 18–26, 01 2017.

[3] M. Kunz, J. Mills, P. Miller, M. King, A. Fox, and S. Marsh, "Application of surface matching for improved measurements of historic glacier volume change in the Antarctic peninsula," *Int. Arch. Photogramm. Remote Sens.*, vol. XXXIX-B8, pp. 579–584, 07 2012.

[4] J. R. Mertes, J. D. Gulley, D. I. Benn, S. S. Thompson, and L. I. Nicholson, "Using structure-from-motion to create glacier dems and orthoimagery from historical terrestrial and oblique aerial imagery," *Earth Surface Processes and Landforms*, vol. 42, no. 14, 2017.

[5] A. Riquelme, M. Del Soldato, R. Tomás, M. Cano, L. Jorda, and S. Moretti, "Digital landform reconstruction using old and recent open access digital aerial photos," *Geomorphology*, vol. 329, 03 2019.

[6] K. Fieber, J. Mills, P. Miller, L. Clarke, L. Ireland, and A. Fox, "Rigorous 3D change determination in Antarctic peninsula glaciers from stereo worldview-2 and archival aerial imagery," *Remote Sensing of Environment*, vol. 205, pp. 18–31, 02 2018.

[7] S. Bougnoux, "From projective to Euclidean space under any practical situation, a criticism of self-calibration," in *Proc. ICCV*, 1998.

[8] V. Lepetit, F. Moreno-Noguer, and P. Fua, "EPnP: An accurate O(n) solution to the PnP problem," *Int. J. Comput. Vis.*, vol. 81, no. 2, 2009.

[9] A. Penate-Sanchez, J. Andrade-Cetto, and F. Moreno-Noguer, "Exhaustive linearization for robust camera pose and focal length estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 10, 2013.

[10] M. Bujnak, Z. Kukelova, and T. Pajdla, "A general solution to the P4P problem for camera with unknown focal length," in *Proc. CVPR*, 2008.

[11] V. Larsson, Z. Kukelova, and Y. Zheng, "Camera pose estimation with unknown principal point," in *Proc. CVPR*, 2018, pp. 2984–2992.

[12] J. L. Schönberger and J.-M. Frahm, "Structure-from-motion revisited," in *Proc. CVPR*, 2016.

[13] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[14] H. Aliakbarpour, K. Palaniappan, and G. Seetharaman, "Robust camera pose refinement and rapid sfm for multiview aerial imagery—without ransac," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 11, pp. 2203–2207, 2015.

[15] P. H. Torr and A. Zisserman, "MLESAC: A new robust estimator with application to estimating image geometry," *Computer vision and image understanding*, vol. 78, no. 1, pp. 138–156, 2000.

[16] G. Golub and V. Pereyra, "Separable nonlinear least squares: the variable projection method and its applications," *Inverse problems*, vol. 19, 2003.

[17] Agisoft, LLC, St Petersburg, Russia, "Agisoft metashape," vol. 7, 2019.

[18] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. USA: Cambridge University Press, 2003.

[19] P. Lewińska, A. Żadło, M. Róg, and S. Szombara, "To save from oblivion: comparative analysis of remote sensing means of documenting forgotten architectural treasures - Zagórz monastery complex, Poland," *Measurement*, 2022.

[20] P. Morin, C. Porter, M. Cloutier, I. Howat, M.-J. Noh, M. Willis, B. Bates, C. Willamson, and K. Peterman, "ArcticDEM; a publically available, high resolution elevation model of the Arctic," in *EGU general assembly conference abstracts*, 2016, pp. EPSC2016–8396.

[21] G. Watkins, N. D'Aeth, Q. Riley, L. Wager, A. Stephenson, and F. Chapman, "The British Arctic Air Route Expedition (continued)," *The Geographical Journal*, vol. 79, no. 6, pp. 466–496, 1932.