Application of SfM methods for the photogrammetric processing of historical aerial VHS videos

Grzegorz Jóźków, Maurycy Hechmann

Wrocław University of Environmental and Life Sciences, 25 Norwida, 50-375 Wrocław, Poland – grzegorz.jozkow@upwr.edu.pl, 120770@student.upwr.edu.pl

Keywords: Historical video, 1997 Central European Flood, bundle adjustment, image matching, orthomosaic.

Abstract

Archival images or videos collected for various purposes may be a valuable data source for photogrammetric processing. Many such examples are reported in the literature. However, the examples of the VHS aerial video photogrammetric processing are not mentioned. This investigation aims to evaluate whether such a type of data collected in a challenging scenario, i.e., an ad-hoc helicopter flight over a flooded area, can be processed with the SfM software. The analysis is focused on image matching and bundle block adjustment. The test data consists of 6 image sequences extracted from VHS videos collected with two different cameras: RGB and near infrared (NIR) from two different heights along the Odra river near Wroclaw during the 1997 Central European Flood. In addition to the images, a low accuracy GPS trajectory logs were used to extract approximated image positions to support tie-point extraction. Archival orthophotomap and digital surface model were used to create ground control points (GCPs) and check points. Each image sequence was processed separately in Agisoft Metashape Professional software. Results showed that the matching of VHS images is possible. Although the trajectory data was of very low accuracy, it was essential to perform tie-point extraction. GCPs extracted from other geospatial data allowed for image georeferencing and creation of orthomosaics, though the geometrical accuracy of the created products is rather low. Nevertheless, executed experiments proved the usefulness of such images, especially NIR, in the flooded area mapping, in which the photogrammetric processing can be executed with the commercial SfM software.

1. Introduction

Structure from motion (SfM) was developed by the Computer Vision Society to reconstruct geometry from a series of overlapping images collected in motion. These series can be a collection of still images or a sequence of frames extracted from a video. Although the quality of still images is usually higher, the video data benefits from the high image overlap. There are many examples of the use of SfM algorithms for the photogrammetric processing of different kinds of videos or historical aerial images. For example, SfM processing is often applied for a 3D reconstruction from video data collected with modern cameras, including those mounted on drones (Qu et al., 2018), but in the past, low-resolution VHS camcorder videos were used for indoor scene reconstruction (Zisserman et al., 1999). SfM algorithms have also proven their usefulness in processing historical images collected with digital consumer-grade cameras (Peppa et al., 2018) or historical photos collected with analogue photogrammetric cameras (Bakker and Lane, 2017). However, the examples of processing historical aerial VHS videos with the SfM algorithms are not reported in the literature.

This investigation aims at the analysis of the performance of the SfM algorithm implemented in commercial software (Agisoft Metashape, 2025) for the processing of historical (almost 3 decades old) aerial VHS videos collected with RGB and near infrared (NIR) cameras over the Odra river during the 1997 Central European Flood. Despite the unusual type of image data processed, other challenges in the photogrammetric processing result from:

- lack of field measurements for georeferencing,
- poor quality of video data (non-metric cameras used, vignetting, low resolution, noise created during copying S-VHS and Betacam videos to VHS tapes),
- unavailability of the source data (only a few image sequences with relatively low image overlap are available)
- variation of the image scale and overlap (helicopter flights were executed for ad-hoc inspection, not designed for photogrammetric data acquisition),
- corridor mapping without image side overlap,
- difficult object (a significant part of some images is the water or forest).

The contribution of this investigation is to prove that SfM algorithms are suitable to process such VHS data. The expected outcome of successful processing is the georeferenced orthomosaic created in an automatic manner with minimum user interaction. It means that intermediate processing, including feature matching for tie-point (TP) extraction as well as dense image matching for point cloud creation, will be executed using algorithms implemented in the used software.

2. Data

2.1 Imagery data

The videos were collected by the Military University of Technology (Warsaw, Poland) along the Odra river in the vicinity of Wrocław (Poland) on 22-31.07.1997 before the second flood wave. The first flood wave occurred in the analyzed area on July 13th, while the second one occurred on July 30th. In total, 35 helicopter flights were executed and 40 hours of video were recorded. The purpose of the flights was to evaluate the state of the embankments after the first flood wave and to assess the range of the flood. For that reason, flights at two different heights were performed.

During the flights, two camcorders (Table 1) were used, recording videos in the visual (RGB) and near NIR spectrum. However, the spectral parameters of the recorded images are unknown. The camcorders were oriented to picture a similar area (Figure 1) with a slightly larger field of view for the RGB videos. During the flight, videos were recorded on the S-VHS and Betacam tapes, and then subsequently copied to the VHS tapes. In 2002, for part of the videos (~65 minutes), sequences of images were extracted from VHS tapes for interpretation purposes (Jóźków, 2003). These images were used in this study. The extraction was performed every 3 s for part of 3 flights executed from 2 different heights (Figure 2, Table 2). The size of extracted images is equal to 768 × 576 pixels. The detailed information of the video recording, as well as other parameters (e.g., accurate focal length) is unknown. However, the given values should be sufficient to determine approximated inner orientation (IO) parameters and further photogrammetric processing. In addition, the VHS tapes are not available currently, thus the second extraction of image frames with possible higher image overlap is impossible.

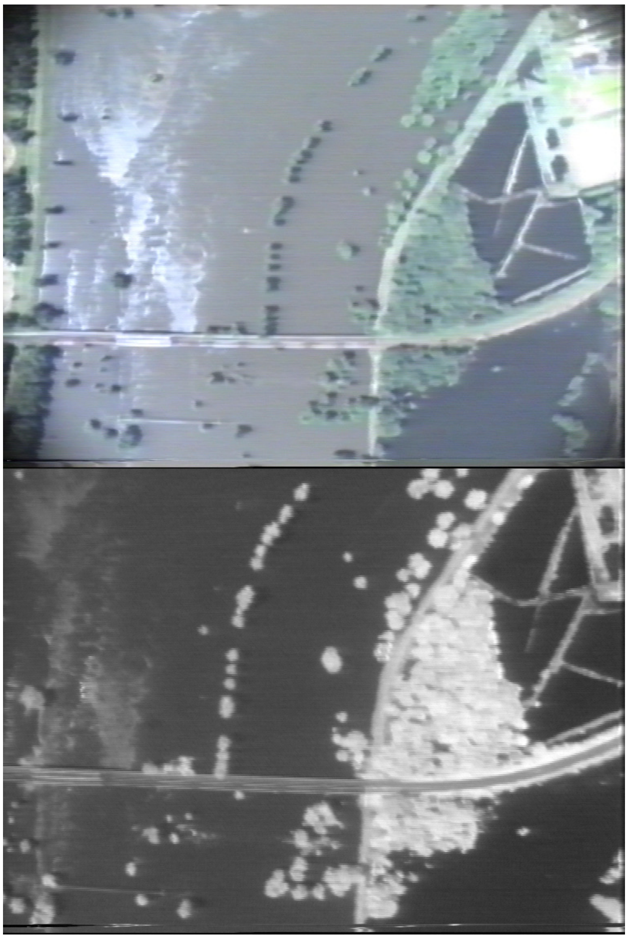

Figure 1. Example of extracted RGB (top) and NIR (bottom) images with visible interlacing and 2 noise lines.

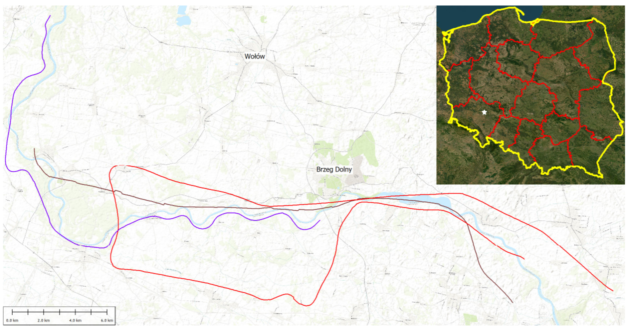

Figure 2. Visualization of test site location and trajectories for the used data: flight 1 (brown), flight 2 (red), flight 3 (violet).

Since the flying height for some data differs significantly (e.g., more than 600 m for flight 1, see Table 2), the image scale varies a lot, posing possible challenges in image matching. However, these changes occur in different parts of the trajectories, causing overlapping images to have similar scale. For that reason, no issues in image matching due to varying image scale are expected.

Camcorder	Sony DXC-D30P	Elemis K15
Video type	RGB	NIR
Sensor	CCD 2/3"	CCD 1/2"
Sensor size	8.8 mm × 6.6 mm	6.4 mm × 4.8 mm
Lens focal length	6 mm	6 mm

Table 1. Parameters of the used camcorders.

Flight no	1	2	3
Flight date	29.07.1997	29.07.1997	30.07.1997
No of images	300	580	420
Trajectory length	36 km	72 km	36 km
Average GSD RGB/NIR [m]	1.10 0.88	2.93 2.17	0.94 0.70
Average image overlap RGB/NIR	~80% ~65%	>90% ~90%	~80% ~65%
Absolute flying heights min/max [m]	339 1025	1613 1886	396 697
Flight speed min/max [km/h]	104 217	99 190	61 135

Table 2. Basic characteristics of the used data.

2.2 Trajectory data

Additional available data was the helicopter trajectory measured with the Trimble Pathfinder Basic GPS receiver. This trajectory may be used to estimate the approximated image exterior position. The available data were the trajectory logs containing WGS-84 latitude and longitude (10E-9 deg resolution), ellipsoidal height (1 m resolution), UTC, and video timestamps (1 s resolution). It should be emphasized that the accuracy of the trajectory data does not match the data resolution and is much worse, especially in horizontal position, because neither the used equipment nor the positioning methods are of surveying grade. In addition, at that time, the GPS Selective Availability had not been terminated yet, which affected the accuracy of GPS measurements. Since the positioning accuracy is rather tens of meters, the lack of the camcorder-to-antenna offsets is irrelevant. The video timestamp given in the log was possibly the timestamp of the original S-VHS or Betacam video, because it did not match any timestamp of the video copies on VHS tapes.

2.3 Other archival data

For the image georeferencing, the oldest available in the public domain orthophotomaps and Digital Surface Models (DSMs) were used. The oldest data was selected because it has the best similarity to the terrain surface in 1997. This data was obtained from the Polish geoportal (https://www.geoportal.gov.pl/). The oldest data available for the investigated area was the aerial B/W orthophotomap from 2004 (GSD = 0.5 m) created from film-based photos, and DSM (GSD = 1 m) created from airborne laser scanning data collected in 2011.

The search for other geospatial data from the 1997 Central European Flood resulted in the information about some processing of satellite radar (ERS-2 mission) images (https://earth.esa.int/eogateway/gallery/flood-in-europe-poland-1997), and 1:20000 aerial photogrammetric color photos collected with a film-based camera on 14.07.1997 (PHARE project) (Jóźków, 2003). The satellite data were not useful for comparison purposes due to very low spatial resolution. The aerial photos, due to their higher resolution and suitable acquisition time (day after the first flood wave) seems to be very useful, however, only paper copies of a few images were available at the resources of the authors' university. In the previous work (Jóźków, 2003) two paper copies of these photos were scanned to convert them to digital form and then transformed to ground coordinates by a simple projective transformation with manual selection of corresponding points. These transformed images were used in this investigation for visual analysis and interpretation. According to our best knowledge, no photogrammetric processing of these images was executed. It is also unknown where the films with images are deployed, if the films were digitized (scanned), or if any field data for the georeferencing was collected.

The ranges of some historical floods are also reported in the other databases and are available in the Polish Hydroportal (https://wody.isok.gov.pl/imap_kzgw/index.html). The metadata of these ranges do not contain any information about their position accuracy, the estimation method (rather modelling than inventory), or what exactly the border of the floor range shows (maximal range of the water visible on the ground surface, some threshold of soil moisture, etc.). For that reason, the data obtained in this manner may be less reliable. In addition, this data does not show any changes over the flood period.

All the available archival data have some issues and cannot be taken as a ground truth for empirical evaluation of the accuracy of products obtained from the photogrammetric processing of VHS images. It may even occur that the results obtained in this investigation are actually more reliable than the information included in other sources mentioned above, but the lack of reliable ground truth data prevents from confirming this statement.

3. Methodology

As mentioned earlier, the investigation aims to evaluate whether the SfM algorithms implemented in the commercial software are suitable to process VHS images. One of the most popular software used for photogrammetric processing of images collected with non-photogrammetric cameras is Agisoft Metashape Professional (Agisoft, 2025). It offers the complete workflow of the processing, including feature matching and TP extraction, Bundle Adjustment (BA), dense image matching, DSM, and orthomosaic creation. In contrast to some other similar software (e.g., DJI Terra), Agisoft Metashape Professional gives the user quite a lot of flexibility regarding the input data and adjusting processing parameters. Agisoft Metashape Professional v 2.1.3 was used in this investigation. The general processing workflow is shown in Figure 3 (numbers in the figure are the subsequent steps), and explained in more detail in subsequent sections.

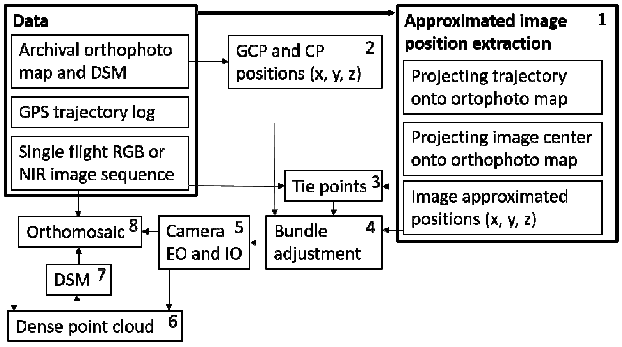

Figure 3. Processing workflow.

Image sequences for each fight and each camera were processed separately (i.e., 6 image blocks were created). Processing of images from different flights was not tested due to several reasons, including very limited image overlap between flights, and possible failures in matching due to different GSD (especially for flight 2). Similarly, the matching between RGB and NIR images is challenging due to the different interactions of different wavelengths with mapped objects. For that reason, often new matching algorithms are developed (Soroush and Baleghi, 2023). The matching of RGB and NIR VHS images may be investigated in the feature if the matching of single type images shows promising results.

3.1 Extraction of approximated image positions

The approximated image exterior orientation (EO) parameters, or only the approximated position of images supports photogrammetric processing workflow in a few steps. Together with IO parameters, they can be used to speed up image feature matching to check only neighboring images that overlap. They can also be used during Aero Triangulation (AT) as the first approximation of unknowns, and serve in AT as observations in the form of air control points that allow for image georeferencing without Ground Control Points (GCPs), though with lower accuracy.

In this work, the approximated image positions were determined semi-automatically. Firstly, the WGS-84 coordinates of the trajectory points were projected on the archival orthophotomap (Figure 4). For the first and last image in the sequence, the location of the image center was found manually on the orthopohotomap and projected perpendicularly on the trajectory. This allowed to interpolate an approximated timestamp for these two images. These timestamps were required to determine the time scale, because the actual time difference between extracted video frames could be affected by the inaccuracy of the playback speed of the videocassette recorder, causing the real time difference between image is different from 3 s. Knowing the GPS timestamps for the first image, timescale, and video sampling rate (3 s), the GPS timestamps of the remaining images were linearly interpolated, and allowed to subsequently interpolate 3D approximated coordinates of image centers. Mentioned earlier inaccuracy of GPS positioning meant that it was not necessary to consider any GPS antenna-to-camera offsets. The interpolation procedure also assumes that the image is exactly vertical and the principal point is in the center of the image.

Figure 4. Example of timestamp and position determination for the first and last image in the sequence. The top view shows an image with a marked center, the bottom view shows the image center in the archival orthophotomap projected on the trajectory (yellow line) with marked time stamps (red dots show trajectory points, green dots are 1 s interpolated timestamps).

3.2 Ground control points and check points extraction

Although the coordinates of image centers are sufficient to perform AT (despite their low accuracy), we decided to use GCPs for georeferencing and Check Points (CPs) for bundle adjustment evaluation. Because no field surveying was performed during the flood, relatively small natural objects were used as GCP and CP markers. Selected objects had to be visible in VHS images and in the archival orthophotomap, clearly distinguishable from their surroundings and having some symmetry that allows for determining their center. An additional condition was that these objects are unchanged (constant) with time. We named them as landmark constants. The selection of these objects was performed manually along the entire path with possibly uniform distribution (depending of landmark constant presence). The coordinates of these points were determined from orthophotomap and the DSM using GIS software.

3.3 Extraction of tie-points

TPs were created fully automatically using feature matching algorithm built into the Agisoft Metashape software. Software developers do not provide details about the algorithms used for this purpose, but it performs a key (feature) point extraction in each image and then performs their matching. Matched key points are then TPs. We decided to perform key point extraction in images upsampled 2 times (upsampling executed automatically by the software) to find possibly more TPs. Parameters used for key point and TP extraction were set to software default, i.e., maximal numbers of key points and TPs per image were set to 40000 and 4000, respectively. In the case of more points, weak points are automatically removed, however, a rather lower number of key points and TPs is expected due to the low resolution of images. The software automatically verifies matches in image pairs and removes outlier correspondences from the set of TPs, thus usually no additional verification is needed, but we decided to visually verify the ground coordinates of TPs created automatically after the first adjustment and remove manually outlier TPs created outside image coverage or significantly lower than the ground surface or higher than its cover. After this removal, the final bundle adjustment was executed.

Since the TP extraction is the core of this investigation, three variants of their extraction were tested. The variants differ by the manner of finding overlapping images: (1) without any knowledge about the image sequence or position, (2) knowing the image sequence only, (3) knowing image approximated positions, thus their sequence. All three variants are possible to execute in the Agisoft Maetashape software.

3.4 Bundle adjustment

The bundle adjustment was executed automatically in the software used. The observations in the adjustments were TPs, GCPs, and approximated coordinates of image centers. In addition, the information about the camera approximated IO parameters (i.e., focal length and pixel size) was used. Approximated position of the principal point and distortion coefficients were set to 0, meaning that the principal point was in the center of the image and the lens had no distortion.

Agisoft Metashape allows to set accuracies of observations to be used for a weighting in the BA. These parameters were selected manually based on the analysis of the source data quality and our experience. The accuracy of GCPs was assumed to be 2 and 5 m for the horizontal and vertical components, respectively. These values are a few times larger than the accuracy specified by the archival data provider, however, we decided to use larger errors for GCPs due to uncertainty in estimating the centers of the landmark constants (see section 3.2) or their possible changes in time. Note that in some areas covered by the water, the available landmark constants were only the tree tops. This affected the accuracy of the GCP position. The accuracy of identifying GCPs in VHS images was set to 3 pixels. The reason was the low quality of these images (e.g., line interlacing) and some difficulties in manual identification of landmark constant centers. All the factors affecting the accuracy of the approximated position of images (e.g., quality of GPS positioning, interpolation method, lack of antenna-to-camera offsets, assumption on strictly vertical images, etc.) were considered together and resulted in the assumption on the accuracy of image positions equal to 30 and 50 m for horizontal and vertical components, respectively. The accuracy of automatic TP extraction was left default and equal to 1 pixel.

During the bundle adjustment, a self-calibration is also performed. Besides the focal length, principal point position, Agisoft Metashape can additionally determine four radial distortion coefficients, two tangential distortion coefficients, and two other coefficients representing affine image distortion

(Agisoft, 2025). Together, 11 camera parameters were determined during BA. The software does not allow setting any accuracy values for initial IO parameters, thus they are probably not treated as observations in the BA process and are used only as the first approximation. Agisoft Metashape can also determine parameters of the rolling shutter distortion, however, this was not used in this investigation since the cameras used have CCD sensors that read all pixels at the same time, thus they do not cause this distortion.

3.5 Creation of subsequent products

In the next step, a dense point cloud was created. The process is done automatically with some options regarding the resolution of the matched images and the strength of depth map filtering. We decided to upsample images 2 times to possibly get more dense point cloud and use the mild (default) filtering of the depth maps that removes some noise. The developer does not provide any details about the used algorithm. The dense point cloud was then manually verified for any obvious low or high points or points created outside the area overlapped by images.

In the subsequent step, the dense point cloud was used to interpolate a GRID DSM. According to the developer information, an inverse distance weighting interpolation method is used, but the details are unknown. The resolution of the GRID model is determined automatically based on the dense point cloud density.

The orthomosaic was also created in the Agisoft Metashape software. The orthorectification was executed automatically based on the created DSM. The seamlines for the orthorectification were created automatically and involve DSM, but details of the algorithm are not published by the developer, similarly as the algorithm for image mosaicking. The resolution of the created orthomosaic was set to the average pixel size.

The analysis of created products was not the aim of this investigation, thus no details about the results are provided. The final product, i.e., orthomosaic was used in this investigation for the interpretation and some comparison with other available data.

3.6 Evaluation of the processing results

The accuracy of the bundle adjustment was evaluated mainly based on the GCP and CP residuals, however, it should be noted that the main purpose of this investigation is not the accurate mapping with such images, just verification of the SfM algorithm's usefulness for photogrammetric processing of VHS images. In addition, the accuracy of GCP and CP ground coordinates is also quite low. The second aspect of the evaluation is the analysis of the trajectories obtained from the image EO parameters calculated in the BA. In particular, any unexpected shifts are identified, and the comparison with the approximated values (obtained from the trajectory, see section 3.1) was performed. In this case, it is difficult to estimate any accuracy parameters, because there is no reference data – the trajectory data quality is too low to treat it as the reference. In fact, the image position determined during BA might be more accurate. The analysis of the self-calibration process was also performed, in particular by comparing the obtained IO parameters for the same camera but for different flights (different image blocks) and by analyzing image residuals. Finally, the visual comparison with other available data was executed for one section of the mapped area.

4. Results and discussion

Among 3 variants of the executed feature matching and TP extraction, only the variant (3) with a known approximated position was successful. The variant (1), without any knowledge of image position or the image sequence, failed. For all image blocks it allowed to find TPs only for several small groups of images. The variant (2) brought slightly better results, in which the number of image groups with TPs was smaller, with a larger number of images in each group. However, the number of images without TPs was still significant, making large gaps in the mapped area. In the first two variants, each image block resulted in an insufficient number of adjusted images (with known EO) covering the mapped area. The best solution was obtained for the variant (3) in which 4 out of 6 image blocks got the EO parameters for 100% of the used images. For the RGB image block of flight 2, one group of 13 images could not be adjusted. The second issue occurred in the NIR image block of flight 3, in which only 3 images (in one group) could not be adjusted. These issues caused a small gap in the mapped area in the mentioned 2 image blocks. Results of variant (3) are discussed in more detail in subsequent sections.

4.1 Image matching and tie-point extraction

The essential part of the processing (feature matching and TP extraction) was generally successful. Key points were created for all images, but the image parts showing the water surface in both types of images remain free of key points. This proves that the feature extraction algorithm implemented in Agisoft Metashape in such areas works in the same manner for VHS videos as for other images. The matching of the key points needs to be evaluated positively since the ground coordinates of the created TPs show a relatively low number of outliers (points outside the mapped area or low and high point noise). Most of the issues with key point matching and TP extraction occurred in the RGB image block for flight 2 (large GSD, dark images). For the images showing mostly the forest, corresponding points did not pass the relative orientation criteria, resulting in a lack of TPs (Figure 5), and consequently, several images without EO. The NIR images for this flight were less affected by the sun illumination, thus the matching of key points was generally successful. Table 3 shows statistics of extracted TPs. The comparison between RGB and NIR image blocks shows that the performance of TP extraction for NIR images is better than for RGB. The TP density is larger for NIR images, while the TP average reprojection error is smaller. This can be explained by lower noise in NIR images, which could be an effect of their lower dependency on the illumination or cloud shadows. For flight 2 (larger GSD), the average reprojection error is larger.

	Flight no	1	2	3
RGB	TPs	94467	133678	182286
	Projections	253238	449647	562327
	Avg. image count	2.68	3.36	3.08
	Avg. TPs per image	844	796	1339
	Avg. TP density per [ha]	30	8	64
	Reprojection error [pix]	0.67	1.15	0.66
NIR	TPs	99477	223943	245306
	Projections	265103	831781	719273
	Avg. image count	2.66	3.71	2.93
	Avg. TPs per image	884	1434	1725
	Avg. TP density per [ha]	42	18	117
	Reprojection error [pix]	0.59	0.77	0.44

Table 3. TP statistics.

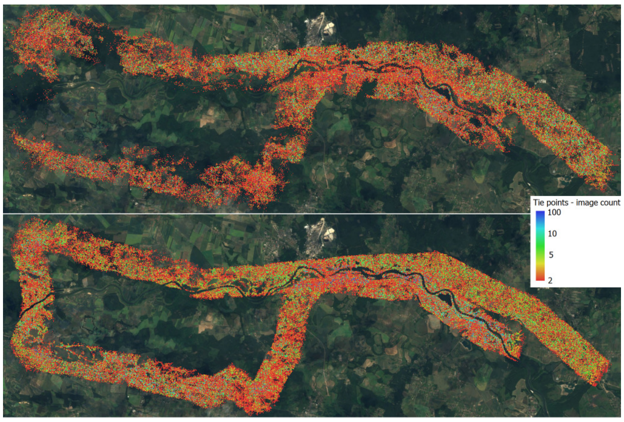

Figure 5. Visualization of TPs extracted for flight 2 (top – RGB image block, bottom – NIR image block).

4.2 Bundle adjustment accuracy

The BA was executed initially with both types of controls (air – approximated image positions, and ground – GCPs). Their weights were assigned according to their accuracy (see section 3.4). Since the residuals for some GCPs and CPs were very large, we decided to perform the bundle adjustment only with GCPs. The large residuals for some ground points may be the result of the unreliable approximated image position caused by the large trajectory errors in some areas. The number of used GCPs was intentionally large (Table 4) to perform as good as possible georeferencing, and it was sufficient to execute BA without the air controls. RMSE calculated for both GCPs and CPs based on the residuals shows that, in general, flight 1 resulted in the best accuracy, and the blocks with gaps (RGB images for flight 2, NIR images for flight 3) resulted in the largest RMSE, exceeding 20 m in 3D. The largest residuals occurred for the points close to the gap and for the areas with a small number of TPs, indicating errors in matching. The large residuals for some points (see row MAX in Table 4 for maximal absolute residuals) caused that for flight 3 with the smallest average GSD, the RMSE for NIR images is larger than for flight 2 NIR images with much larger GSD.

		GCPs			CPs		
Flight no		1	2	3	1	2	3
Points (RGB)		75	63	35	11	10	15
RMSE	XY [m]	1.5	12.5	3.8	5.4	17.5	9.8
	Z [m]	1.7	15.8	3.2	3.2	22.1	13.3
	3D [m]	2.3	20.1	5.0	6.2	28.2	16.5
MAX	XY [m]	4.8	39.7	10.4	9.1	38.6	20.7
	Z [m]	5.1	39.3	7.3	5.1	44.6	41.9
	3D [m]	5.9	46.4	12.0	9.4	47.4	46.7
Points (NIR)		59	89	32	9	10	15
RMSE	XY [m]	3.4	9.3	7.1	8.9	13.2	21.2
	Z [m]	1.7	6.9	5.7	9.2	8.7	10.1
	3D [m]	3.8	11.5	9.1	12.8	15.8	23.4
MAX	XY [m]	15.1	26.2	31.1	15.9	22.5	67.0
	Z [m]	3.9	26.8	24.2	19.5	13.3	20.7
	3D [m]	15.5	37.2	39.4	22.2	24.5	69.0

Table 4. Accuracy of BA calculated from GCPs and CPs residuals.

The BA results show no clear difference between RGB and NIR images. Sometimes, vertical (Z) residuals and RMSE are smaller than horizontal, which may indicate some issues during camera self-calibration included in the BA process.

The accuracy of BA for all blocks is rather poor and much worse than the image GSD size, thus it prevents the use of such images for accurate mapping purposes. However, the approximate georeferencing for all image blocks should be considered as a success despite all the issues and challenges (see section 1). The lack of other reliable data from the flood causes the outputs of such processing to be still a valuable source of information, though their low geometrical accuracy.

4.3 Image exterior orientation and trajectory

The analysis of the image position obtained from the BA may give an answer about the quality of the GPS positioning used and indicate other problems in the BA. Both of these issues should be especially visible in height profiles. Although the flight was performed with an ad-hoc helicopter flight, any sudden changes in the trajectory should not occur since the GPS log does not show any significant (close to 0) changes in the speed (see flight speed in Table 2). The analysis of the vertical profiles of the trajectories (Figure 6) clearly shows that in many places, approximated image height, thus GPS positions, are unreliable and show significant jumps. Such vertical jumps may be an effect of cycle slips. These jumps also affected the BA results, e.g., flight 1, NIR images no 135-145. However, other vertical jumps in the BA results (e.g., flight 2, RGB images no 133-134) are caused by incorrect image matching. The detailed analysis of this pair of images indicates that the built-in algorithm resulted in the wrong selection of inlier key points. It found 10 pairs of corresponding points that include a few outliers. The adjustment was possible thanks to the GCPs, but was executed with such errors. The same reason caused a few other similar jumps for flight 2 RGB images, and one in flight 3 NIR images. In addition, for flight 3 from about image no 300, the heights obtained from BA for RGB and NIR images differ a lot, with the RGB trajectory far from the GPS trajectory. The reason for this issue also comes from the matching errors and the low number of TPs. The low number of correct TPs is the result of the forested area along the river (see Figure 2, SN part of the trajectory for flight 3). Significant differences in trajectory height obtained from RGB and NIR image block BA are also visible for flight 2, while for flight 1, trajectories obtained from BA of image blocks of both types show much more similarity (except mentioned earlier one section for NIR images). This similarity coincides with the lowest RMSEs obtained for flight 1 (Table 4).

		Standard deviation			Max. abs. differences		
	Flight no	1	2	3	1	2	3
RGB	XY [m]	60	76	50	118	207	121
	Z [m]	119	47	58	276	146	198
	3D [m]	134	89	76	295	208	230
NIR	XY [m]	43	35	36	121	108	86
	Z [m]	75	39	53	220	110	210
	3D [m]	87	52	65	229	151	223

Table 5. Statistics for image position differences (between approximated and calculated from BA).

Besides the height profiles, the statistics for the image position differences (approximated from GPS trajectory and calculated from BA) were calculated (Table 5). Since none of the used data is ground truth, these values cannot be considered as errors, but still can confirm the low quality of the GPS positioning used. The differences in the positions are more than tenths of meters on average (standard deviation), and reach for some images almost 300 m.

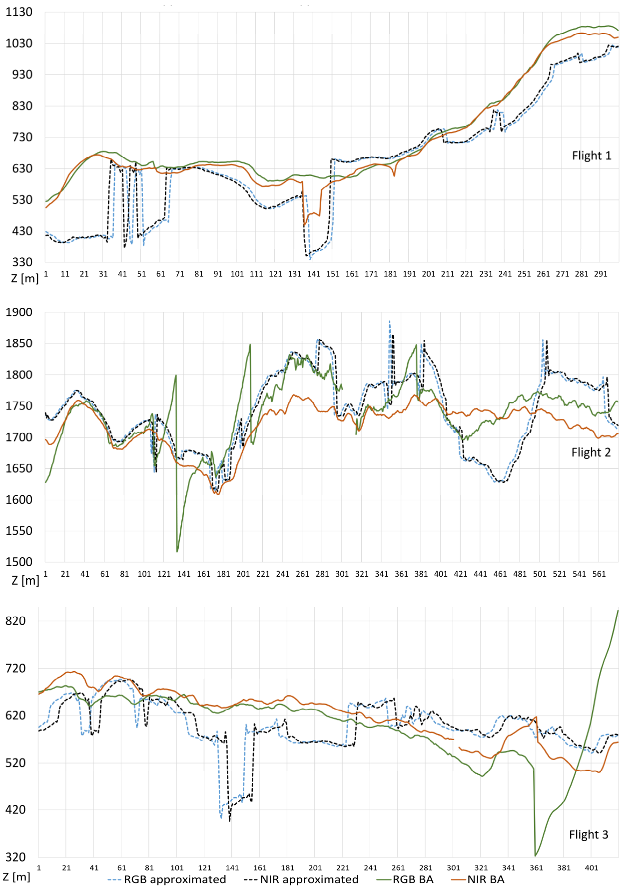

Figure 6. Height profiles of image sequences. The vertical axis shows image height, while the horizontal axis shows the image number in the sequence.

4.4 Camera self-calibrations

Another challenge in the processing was unknown camera IO and lens distortion. As mentioned earlier, this issue was addressed by the camera self-calibration included in the BA process, similar to UAV photogrammetry. The results of the obtained IO parameters are given in Table 6. The differences in focal length of the same camera but different flight is clearly visible for the RGB camera. The incorrect self-calibration of this camera likely occurs due to the mentioned issues in the image matching for flights 2 and 3. For the NIR camera, the self-calibration is more consistent, though the difference in focal length reaching almost 10 pixels seems to be significant. However, the method of camera focusing is unknown (adjusted between flights or between days). The consistency in the principal point position for the NIR camera is also higher.

	Flight no	1	2	3
RGB	f [mm]	6.50	6.30	6.11
	x_0 [pix]	-1.5	0.8	3.0
	y_0 [pix]	-2.3	26.0	-2.5
NIR	f [mm]	6.18	6.15	6.11
	x_0 [pix]	-46.3	-40.2	-31.1
	y_0 [pix]	48.2	44.2	43.1

Table 6. IO parameters calculated by self-calibration.

The software used allows for showing average image residuals after applying distortion parameters (Figure 7). Although the residuals are relatively low (RMSE below 0.2 pixels for all datasets), it is clear that VHS line interlacing affected the calibration because, for the particular lines, significant residuals are aligned with the line direction. A similar pattern is shown for both cameras and all flights, which means that this issue was probably caused by the other equipment, i.e., a videocassette recorder used to copy VHS tapes or extract images from these tapes.

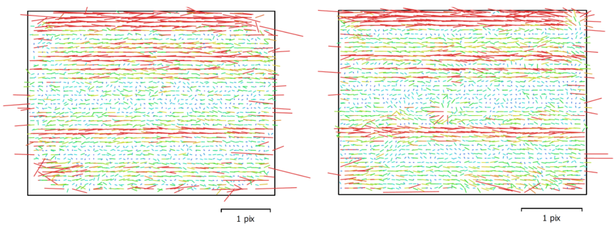

Figure 7. Visualization of image average residuals after applying distortion parameters for the RGB camera, flight 1 (left), and the NIR camera, flight 2 (right).

4.5 Comparison with available data

The flood in 1997 in Poland took more than one month and had two major waves (in the analyzed area around July 13th and July 30th). The second wave was significantly smaller and was not as destructive as the first one due to the executed preparations. The area covered by the water has the largest extent during and directly after the first wave and decreases over time. For that reason, the water range was dynamic, and in general, not mapped because of more important needs. In that time, the satellite remote sensing systems were not so developed in terms of the number of different satellites and the resolution. Mentioned above (section 2.3) results from the radar system have too low resolution to compare them with the results from this study. Airborne data, besides those used in this study, were 1:20000 film-based photos, which for the part of the analyzed area were collected just after the first flood wave, thus documenting almost the farthest water extent. According to the best author’s knowledge, these photos were not used in any photogrammetric processing to extract the water range, most likely due to the lack of GCPs. Processing of such images goes beyond this study and requires additional effort (e.g. fining the location of original images, obtaining necessary agreements, scanning the images, etc.), thus the results of the previous work (Jóźków, 2003) were used in this study. However, due to the different time (more than 2 weeks) of the acquisition of these photos and VHS videos, it is not possible to compare them in terms of the water range. The last data used for comparison is the flood range obtained from the Hydroportal (see section 2.3) As is clearly seen in the given example (Figure 8), the flood range reported in this data is not reliable, since it shows a flooded area behind the southern river embankment, while the aerial photo and VHS videos show that this area was untouched by the flood. This simple comparison shows that although the accuracy of VHS image georeferencing is quite low, it brings more reliable information than the data available in other public sources.

Figure 8. Visual comparison of 1997 Central European Flood range near Brzeg Dolny for different datasets. From top: Polish Hydroportal, aerial photos, RGB VHS orthoimosaic (flight 2, increased brightness), NIR VHS orthomosaic (flight 2) with the 2025 orthophotomap in the background.

5. Conclusions

Executed experiments proved that the SfM algorithms implemented in Agisoft Metashape software allowed for photogrammetric processing of historical aerial RGB and NIR images extracted from VHS videos. The usefulness of the key SfM algorithm (image feature matching) was proven on very challenging datasets, though some errors in RGB image matching occurred for forested areas. The trajectory data, although very inaccurate, was essential for TP extraction. In addition, the position of landmark constants extracted from another historical geospatial data allowed for the georeferencing of VHS images. Despite the low geometrical accuracy of the BA, the orthomosaics created bring useful information about the flood range and can be used for an approximate measurement of the water extent between two flood waves. Among the two image types, the image quality, processing accuracy, and information about the flood range are better for the NIR image sequence than for the RGB.

Executed experiments showed that ad-hoc fights over flooded areas with a low-quality camera may be used for reliable documentation of the flood dynamics. In comparison to regular photogrammetric flights with a mapping camera, they are much easier to execute, cheaper, and can be repeated to map flood development. The use of modern technological achievements, e.g., 8k digital videos recorded with full-frame sensors, centimeter-level GNSS positioning, unmanned aerial vehicles, efficient processing algorithms, etc., will likely result in much better georeferencing accuracy and product quality.

References

Agisoft Metashape, 2025. https://www.agisoft.com/ (20 October 2025).

Bakker, M., and Lane, S. N., 2017. Archival photogrammetric analysis of river–floodplain systems using Structure from Motion (SfM) methods. *Earth Surf. Process. Landforms*, 42: 1274–1286. doi.org/10.1002/esp.4085.

Jóźków, G., 2003. Analiza porównawcza treści zdjęć lotniczych i cyfrowych lotniczych obrazów video w podczerwieni z powodzi w 1997 roku na wybranym odcinku rzeki Odry [Comparative analysis of the content of aerial photographs and digital infrared aerial videos from the 1997 Central European Flood on a selected section of the Odra River]. *Master thesis*, Agricultural University of Wroclaw, Poland.

Peppa, M. V., Mills, J. P., Fieber, K. D., Haynes, I., Turner, S., Turner, A., Douglas, M., and Bryan, P. G., 2018. Archeological Feature Detection from Archive Aerial Photography with a SFM-MVS and Image Enhancement Pipeline. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2, 869–875, doi.org/10.5194/isprs-archives-XLII-2-869-2018.

Qu, Y., Huang, J., and Zhang, X., 2018. Rapid 3D Reconstruction for Image Sequence Acquired from UAV Camera. *Sensors*, 18(1), 225. doi.org/10.3390/s18010225.

Soroush, R., and Baleghi, Y., (2023). NIR/RGB image fusion for scene classification using deep neural networks. *The visual computer*, 39(7), 2725-2739. doi.org/10.1007/s00371-022-02488-0.

Zisserman, A., Fitzgibbon, A. and Cross, G., 1999. VHS to VRML: 3D graphical models from video sequences, *Proceedings IEEE International Conference on Multimedia Computing and Systems*, Florence, Italy, pp. 51-57 vol.1, doi.org/10.1109/MMCS.1999.779119