GENERATION OF A BENCHMARK DATASET USING HISTORICAL PHOTOGRAPHS FOR AN AUTOMATED EVALUATION OF DIFFERENT FEATURE MATCHING METHODS

F. Maiwald[1, *]

[1] Institute of Photogrammetry and Remote Sensing, TU Dresden, Germany – ferdinand.maiwald@tu-dresden.de

Commission II, WG II/8

KEY WORDS: benchmark, image dataset, historical images, image orientation, feature matching

ABSTRACT:

This contribution shows the generation of a benchmark dataset using historical images. The difficulties when working with historical images are pointed out and structured in three categories. Especially large viewpoint differences, image artifacts and radiometric differences lead to weak matching results with classical feature matching approaches. The necessity of publishing an own benchmark dataset is emphasized when comparing to existing datasets which are partly using synthetic data, well-known orientation or strictly categorized image differences. The presented image dataset consists at the moment of 24 images which are oriented in image triples using the properties of the Trifocal Tensor as a more stable image geometry. In the following, three different feature detectors and descriptors that have already been proven well on historical images (MSER, ORB, RIFT) are evaluated using the new benchmark dataset. Then, several outlier removal methods were applied on the detected features. The tests show that for the entirety of image pairs RIFT performs slightly better than the other two methods. Nonetheless, for some image pairs MSER significantly improves the matching score but even so, historical image pairs are difficult to be matched with the presented methods due to challenging outlier removal. Still, the estimated projective relative orientation could be used in an autocalibration approach to place the images in a metric scene.

1. INTRODUCTION

This contribution presents the generation of a benchmark dataset for the evaluation of different feature matching methods on historical images. The work is placed in the context of a 4D web application (3D models and related historical images and data) of the city of Dresden as an alternative media repository for e.g. art historians. Oriented images and methods to match historical images provide the basis for the placement of the images in such a 3D space. The presented images originate from the photo library of the Saxon State and University Library Dresden (SLUB), which contains about 1.8 million images of 80 institutions at this point in time. The majority of images in this archive was taken between 1940 and 1990 (deutschefotothek.de). The images for the benchmark dataset were redigitized for this purpose and show various buildings. While the absolute orientation of these historical photographs is neither given nor easy to define, this approach focuses on the determination of the relative orientation between different historical images.

This leads to diverse issues considering that extrinsic and especially intrinsic camera parameters are mostly unknown. Additionally, the images are taken by different camera types which vary in exposure and acquisition time. Consequently, the presented dataset is relatively oriented in a projective frame using a more stable triple image geometry (Hartley, 1997). The matches between the three images of one building view and additionally the relating Trifocal Tensor T are determined and given. This orientation data can then be used to evaluate different feature detectors, descriptors and feature matching methods on historical images. In the following, it may be possible that an oriented image mosaic can be metrically spatialized in a three-dimensional environment with the appropriate scale using autocalibration (Faugeras et al., 1992). The dataset consists of 24 images (2 image triples respectively for 4 buildings) and could be extended in the future. The images have different properties ranging from small viewpoint and radiometric changes to large differences. These properties can be summarized in the following three categories.

1.1 Image differences based on digitization and image medium

Even if an image would have been taken twice at the same moment in time, some differences concerning the digitized copy could occur during and even before digitization. This happens because historical images are mainly archived on photographic plates or photographic film. Any change on this original data is preserved during digitization. Especially, the conservative emulsion on the glass plates can deteriorate and additionally a glass plate is fragile and any crack will be pictured in the digitized image (Gillet et al., 1986). Scratches, dust and finger-prints may also be visible in the digital copy.

Similarly, photographic film is vulnerable to damage e.g. by mold, photo-oxidation, air pollutants and improper handling (Slate, 2001). All of these image artifacts are transferred using digitization techniques and will interfere with the process of feature detection. One further image difference that may appear and is relevant for photogrammetry is the change of the principal point in the digital copy. It does not have to be necessarily the middle of the digital copy but it can shift, if only a part of the original image is digitized or if the original data has been cropped. It may be even possible that the principal point is not pictured on the digital copy. Additionally, when the digitization information (sensor, resolution, dynamic range, working area, accuracy, filters) is not available every metric data is lost in the process.

* Corresponding author

1.2 Image differences based on different cameras and acquisition technique

When comparing various historical images, the main difference between them is the strongly changing representation of the depicted object. Photographs of the same object are taken in summer and in winter, in daylight and in nighttime and thus, the radiometric properties change. The historical images may be blurred, noisy, under- and overexposed and different light spots, reflections and shadows can appear in the same photographic scene and interfere with the feature detection. Sometimes, people, cars or other objects are in front of the depicted building and influence the feature matching.

Additionally, on the one hand it is possible that there are extreme viewpoint changes between the images and on the other hand sometimes one building is solely photographed from similar perspectives, which makes a 3D reconstruction difficult. Since the camera types are mostly unknown and undocumented the inner orientation important for the reconstruction is not available and has to be estimated.

1.3 Object differences based on different dates of acquisition

A difficult topic is the dealing with object differences shown in the photographs. Building differences can vary between very small changes like on claddings, window frames or small statues to large ones considering destroyed or reconstructed buildings. It is not possible to assume that a historical building that is represented on various images did not change over time. Nonetheless, some valuable orientation information can even be determined using these destroyed or changed buildings. It will be difficult to decide whether an object changed so much that any metric information generated with photogrammetric methods is invalid. Furthermore, it is still discussed how to represent this error-prone data (Apollonio, 2016), (Kensek et al., 2004). It could be possible in a first step to categorize historical images using content-based image-retrieval on a very accurate scale and only use feature matching methods on image pairs of clearly the same building in the same state.

2. RELATED WORK

There exists already a numerous variety of image datasets in computer vision for different purposes like (people-)detection, classification, recognition, tracking, segmentation, multiview and many more. Famous datasets are e.g. the Caltech 256 dataset for classification purposes (Griffin et al., 2007) or the KITTI dataset used in autonomous driving and SLAM research (Geiger et al., 2013). The presented dataset could be integrated in the multiview category and closes a gap between different existing datasets. In contrast to datasets with a lot of images and their inner orientations (Moreels and Perona, 2007) it is not or only hardly possible to provide that many historical images including the proper inner orientation since the camera types are mostly unknown.

Similar to the Affine Covariant Regions dataset (Mikolajczyk et al., 2005) the presented benchmark dataset consists of real data (= not synthetic data) with changes in illumination, viewpoint, blur and rotation. Some of the historical images even have large viewpoint or illumination changes like in the Extreme View Dataset or the Ultra Wide Baseline Dataset (Mishkin et al., 2015). These existing datasets are using the fact that "the images are either of planar scenes or the camera position is fixed during acquisition, so that in all cases the images are related by homographies [..] and this mapping is used to determine ground truth matches [..]." (Mikolajczyk et al., 2005). This is not (always) possible when using historical data, so the presented benchmark dataset is described by the predefined corresponding points and the Trifocal Tensor determining the relative orientation between image triples. At the time of this research no other freely available benchmark dataset with oriented images older than 40 years used for feature detection and matching could be found.

However, many people are working with historical images and further data to reconstruct mostly buildings and sights. This includes e.g. the reconstruction of the great Buddha of Bamiyan (Grün et al., 2004), dinosaur tracks (Falkingham et al., 2014) or the orientation of historical images of Atlanta, GA (Schindler and Dellaert, 2012). But also recent research is done with historical data e.g. in combination with terrestrial laser scanning (Bitelli et al., 2017), using old film negatives (Rodríguez Miranda and Valle Melón, 2017) or aerial images (Giordano et al., 2018). Though, those projects show a developing degree of automation in image processing a lot of work is still done manually in this field of research (Henze et al., 2009), (Gouveia et al., 2015). An oriented historical image dataset could help to improve automated approaches in image classification, image matching and image orientation.

3. THE IMAGE DATASET

Examples for the historical image dataset are shown below (fig. 1). The whole published dataset consists of 24 images with a maximum side length of 3543 pixels. It is mostly unclear, whether the original data is originated from photographic plates or film negatives. The images are grouped in two triplets respectively for 4 buildings ($2 \times 3 \times 4 = 24$). Images were chosen with respect to their possible matching quality. The images show combined differences in illumination, field of view, viewpoints, blurring and slight rotation. Some of the images show building reflections in water or extreme shadowing. Thus, a very challenging dataset when using a single feature matching method is provided.

Since the relative orientation of the image pairs cannot be easily described through a homography as explained before, the first step would be the description of the image pairs using a Fundamental Matrix F calculated out of at least 7 point correspondences, where F is defined by equation 1,

$$x'^{T} F x = 0 \tag{1}$$

where x' and x are at least 7 image correspondences in homogeneous coordinates.

One must say, that this equation can hardly be used to test correspondences determined with feature matching methods because an estimated (e.g. using RANSAC) fundamental matrix F is only a projective map taking a point to a line. That means a point $x(x, y, z)$ in the first image defines a line (the corresponding epipolar line $l' = Fx$) in the second image (Hartley and Zisserman, 2003). Additionally, the point transfer from image 1 to image 2 using the epipolar line can lead to false positives considering matches that lie randomly on the epipolar line but are no true matches.

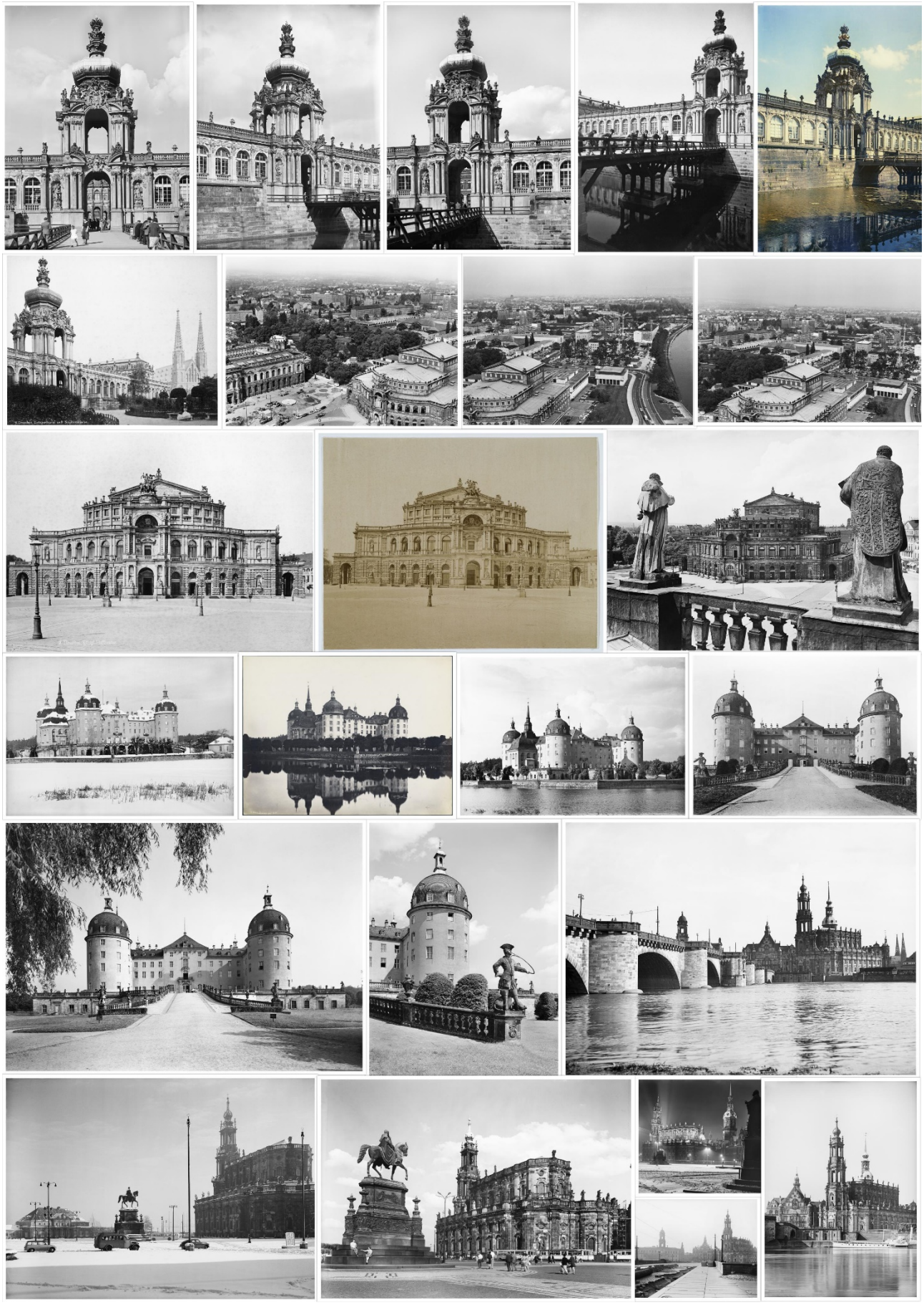

Figure 1. All current images of the benchmark dataset showing the variety of historical images

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLII-2-W13-87-2019

This leads to a more stable image configuration when using three images, because e.g. the epipolar lines from image 1 to image 3 and from image 2 to image 3 of the same feature point intersect in image 3 in the homologue feature point (Maas, 1997). The matching can be simplified using the $3 \times 3 \times 3$ Trifocal Tensor T and its properties for a point-point-point correspondence (eq. 2) (Hartley and Zisserman, 2003).

$$[x']_\times \left(\sum_i x^i T_i \right) [x'']_\times = 0_{3\times3} \tag{2}$$

where $x, x', x'' =$ image coordinates in the three images
$[x]_\times = 3 \times 3$ skew-symmetric matrix of 3-vector
$i =$ number of 3×3 Tensor slice
$T =$ Trifocal Tensor of the three images
$0_{3\times3} = 3 \times 3$ null matrix 3-vector

A point transfer from e.g. the first view to the third view can then be realized using equation 3 and the corrected Fundamental Matrices F_{12}, F_{13} and F_{23} extracted from the Trifocal Tensor (Hartley and Zisserman, 2003).

$$x''^k = x^i l'_j T_i^{jk} \tag{3}$$

where $i, j, k =$ indices that correspond to the entities in the first, second and third views respectively

Since the Trifocal Tensor is not that easy to determine like a homography or the Fundamental Matrix, it is provided for every benchmark image triple. Additionally, the calculation of T is explained in the following. There are various methods that are used for the computation of the Trifocal Tensor namely e.g. the minimal parameterization by Faugeras and Papadopoulo (Faugeras and Papadopoulo, 1998) and by Nordberg (Nordberg, 2009) or the constrained solutions by Ponce and Hebert (Ponce and Hebert, 2014) as well as Ressl (Ressl, 2002). Most approaches have already been tested and the constrained solution by Ressl has shown the most robust results leading to the smallest reprojection errors (Julià and Monasse, 2017). Using this computation method requires approximation values for the Trifocal Tensor and the Projection Matrices of the three images. These can be found by solving $At = 0$ in a linear way. The matrix A is the Jacobian of the trilinearities and consists of the $(n = 4)$ row-wise ordered sub matrices A_c where c is the number of point correspondences (eq. 4) (Ressl, 2003).

$$A_c = \left(S^{red}(x'') \otimes S^{red}(x''') \right) (x'^T \otimes I_9) \tag{4}$$

where $S^{red} =$ reduced axiator for point coordinates
$\otimes =$ Kronecker product for 4 linearly independent equations

Afterwards, the approximation values can be calculated by minimizing the algebraic error using a singular value decompositon (SVD). It is recommended to use at least 10 normalized point correspondences in all three images with a pixel noise of 1 to minimize the reprojection error with the subsequent constrained solution (Ressl, 2003). For the benchmark dataset at least 15 manual point correspondences were used in the image triples and a pixel noise < 1 was targeted. The verified results for the different matching strategies show that this goal could be accomplished. The detailed description, the images, the matched points and the corresponding Trifocal Tensor are available on the website[1].

Since the Trifocal Tensor provides geometric relations between three views only in a projective frame independent of scene structure (Hartley and Zisserman, 2003) the resulting camera matrices P, P', P'' retrieved by eq. 5 could be introduced as a prior relative orientation into an autocalibration algorithm (Heinrich et al., 2011) allowing the estimation of inner and exterior orientation and in the following, the generation of simple structures in euclidean metric 3D space.

$$\begin{aligned} P &= [I|0] \\ P' &= \left[[T_1, T_2, T_3] e'' | e' \right] \\ P'' &= \left[(e'' e''^T - I) \left[T_1^T, T_2^T, T_3^T \right] e' | e'' \right] \end{aligned} \tag{5}$$

where $e', e'' =$ respective normalized epipoles
$T_1, T_2, T_3 =$ Tensor slices

4. COMPARISON OF DIFFERENT FEATURE DETECTION AND DESCRIPTION METHODS

In the following, the different feature detection methods used on the benchmark image dataset are briefly explained. Three distinct algorithms were chosen to process the images in full resolution and find point features. The comparison is done between image pairs but can be evaluated using the Trifocal Tensor. Thus, the number of correct matches in relation to the sum of all matches (= matching score) could be determined. Some of the common methods have already been tested on historical image data and a combination of the ORB (Oriented FAST and Rotated BRIEF) feature detector and the SURF (Speeded-Up Robust Features) feature descriptor produced decent results (Ali and Whitehead, 2014). Another approach that generated a good matching ratio was the MSER (Maximally stable extremal regions) feature detector and descriptor (Wolfe, 2013). Additionally, those results are compared with a newer method called RIFT (radiation-invariant feature transform), that neglects radiometric differences in images and thus, can be a good addition to existing approaches. For the first and second test the standard implementations of ORB, SURF and MSER in OpenCV were used. The third test used the implementation of RIFT in Matlab (Li et al., 2018). The results are presented without outlier removal using brute force matching, outlier removal using a symmetry test and as a third approach outlier removal using Fundamental Matrix calculation with the random sample consensus (RANSAC) (Fischler and Bolles, 1981). Additionally, for RIFT the native calculation using the fast sample consensus (FSAC) (Wu et al., 2015) is shown.

4.1 Oriented FAST and Rotated BRIEF (ORB)

ORB is a common alternative to SIFT and uses an intensity oriented FAST (Rosten and Drummond, 2006) for feature detection and an in-plane rotation invariant version of BRIEF (Calonder et al., 2010) for feature description (Rublee et al., 2011). Since the hybrid version using the ORB detector and the SURF descriptor achieved better results on historical images (Ali and Whitehead, 2014) the presented approach chooses this as a first method for feature detection and description. The oriented FAST detects keypoints using the intensity threshold between the

[1] https://dx.doi.org/10.25532/OPARA-24

Dataset	Mb_1			Mb_2			Zw_1			Zw_2			So_1			So_2			Hk_1			Hk_2		
Imagepair	1_2	1_3	2_3	1_2	1_3	2_3	1_2	1_3	2_3	1_2	1_3	2_3	1_2	1_3	2_3	1_2	1_3	2_3	1_2	1_3	2_3	1_2	1_3	2_3
Serial Number #	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
MSER_bf	0,74	1,17	0,33	0,40	0,58	0,24	9,04	3,01	10,33	0,18	0,39	0,47	3,20	10,24	7,07	0,77	3,30	0,26	0,33	0,26	0,07	1,15	0,61	1,31
MSER_RANSAC	0,00	1,12	0,00	0,00	2,88	16,22	59,82	0,00	54,17	0,00	0,00	0,00	23,85	29,20	63,02	0,00	13,33	0,00	4,95	2,11	0,00	0,00	0,00	0,00
MSER_sym	0,81	1,38	0,40	0,60	0,41	0,20	17,05	5,00	20,89	0,23	0,44	0,77	5,75	18,10	17,79	0,74	8,19	0,62	0,61	0,24	0,06	1,12	1,46	2,06
ORB_bf	2,00	1,28	0,72	1,24	0,52	1,12	6,88	2,64	10,56	0,40	1,28	2,40	2,80	7,88	9,04	1,00	4,44	0,76	0,64	0,08	0,00	4,24	0,44	0,76
ORB_RANSAC	44,83	9,09	0,00	3,45	0,00	0,00	10,71	3,23	57,72	0,00	0,00	22,45	17,31	15,74	42,42	0,00	13,64	5,26	0,00	0,00	0,00	42,55	0,00	0,00
ORB_sym	2,38	1,79	0,58	2,19	1,19	2,37	17,36	6,34	22,34	0,48	2,42	3,89	6,73	14,67	20,97	2,46	7,30	1,39	1,11	0,00	0,00	9,28	0,39	0,70
RIFT_bf	6,44	7,80	4,88	0,92	1,92	1,76	4,92	1,56	16,45	1,40	4,24	4,52	3,12	12,53	9,68	0,80	7,48	0,40	3,72	0,96	0,48	0,84	1,96	3,56
RIFT_RANSAC	0,00	1,01	0,00	0,00	0,00	0,00	3,85	0,00	80,00	0,00	18,18	0,00	0,00	30,95	0,83	0,00	12,50	0,00	0,00	0,00	0,00	0,00	0,00	0,00
RIFT_sym	18,51	26,71	18,18	1,65	5,77	5,15	15,24	4,34	36,49	3,71	10,86	17,80	16,09	27,07	34,30	1,13	16,33	1,15	17,09	2,80	2,94	2,42	10,00	11,46
RIFT_native	77,78	83,33	80,43	0,00	0,00	72,73	47,62	0,00	50,00	15,38	46,94	62,50	42,50	0,00	0,00	0,00	8,82	25,00	66,67	25,00	25,00	0,00	69,23	84,21

Table 1. Results for different feature matching methods for 8 different image triples (=24 image pairs). Matching results are shown respectively for every dataset for the image pairs 1_2, 1_3 and 2_3 as ration in % between all found matches and correct matches (matching score). Good results are highlighted in green whereas bad results are shown in red

Serial Number #	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
MSER_bf	60	95	16	60	86	42	1064	354	1742	24	51	63	707	2266	1904	123	530	43	14	11	9	95	50	102
MSER_RANSAC	0	1	0	0	3	18	329	0	942	0	0	0	150	2239	1554	0	28	0	5	2	0	0	0	0
MSER_sym	7	22	4	11	6	3	282	75	448	4	9	15	213	747	679	15	87	6	5	2	1	11	9	13
ORB_bf	50	32	18	31	13	28	172	66	264	10	32	60	70	197	226	25	111	19	16	2	0	106	11	19
ORB_RANSAC	26	4	0	1	0	0	12	1	142	0	0	11	9	172	154	0	9	2	0	0	0	20	0	0
ORB_sym	15	12	3	12	7	16	117	41	170	3	15	22	44	151	156	14	53	8	7	0	0	57	2	4
RIFT_bf	161	195	122	23	48	44	123	39	411	35	106	113	78	313	242	20	187	10	93	24	12	21	49	89
RIFT_RANSAC	0	1	0	0	0	0	1	0	20	0	2	0	0	39	1	0	1	0	0	0	0	0	0	0
RIFT_sym	87	125	72	4	18	17	80	16	316	15	63	89	60	281	213	3	104	3	61	7	7	7	28	33
RIFT_native	42	35	37	0	0	8	10	0	46	4	23	20	17	0	0	0	3	1	4	2	2	0	9	16

Table 2. Results for different feature matching methods for 8 different image triples (=24 image pairs). The total number of correct matches is shown respectively for every dataset for the image pairs 1_2, 1_3 and 2_3. Good results are highlighted in green whereas bad results are shown in red

center pixel and a circular ring around that center. The orientation of the keypoints is done using an intensity centroid (Rosin, 1999). The standard maximum value for the number of features retained was set from a maximum of 500 to 2500 to allow a better comparison with the other methods. In the following, SURF is used for the description of the features since it outperforms other descriptors by repeatability, distinctiveness and robustness (Bay et al., 2006).

4.2 Maximally Stable Extremal Region Detector (MSER)

As a second method the presented approach uses MSER (Matas et al., 2004). This algorithm is usually applied on image pairs with a wide baseline. Classical feature points are replaced by regions which are closed under projective transformation of image coordinates and monotonic transformation of image intensities (Matas et al., 2004). Those properties can be especially useful for historical images because of the already explained image differences. Regions described by a connected number of pixels are chosen by the property that all pixels inside one extremal region have either a higher or a lower intensity than all the pixels on its outer boundary (Mikolajczyk et al., 2005). Again, SURF is used for the description of the regions consisting of feature point sets.

4.3 Radiation-invariant Feature Transform (RIFT)

The third method used is called RIFT. The radiation-invariant feature transform is chosen because of its invariance to nonlinear radiation distortions (NRD) (Li et al., 2018) and the use of edge features in addition to corner features. Both effects can support the feature detection in historical images. The approach uses the Fourier transform to generate phase congruency maps. Independent maps for each orientation of a 2D log-Gabor filter are created and used for the detection of corner features as well as edge features. In the following, those features are described by a 216-dimensional feature vector calculated through a maximum index map based on a log-Gabor convolution sequence (Li et al., 2018). RIFT is currently not scale invariant and so it should perform bad on large scale-changes. However, it has been observed that feature points in image pairs with small scale-changes can still be matched correctly.

4.4 Feature matching and outlier removal

For the comparison of all methods, the presented approach uses a brute force matching (_bf) for all detected feature points, i.e. all feature points with their particular descriptors are matched (so every descriptor in image 1 is compared with every descriptor in image 2). In the following, two different outlier removal methods are evaluated. The first approach uses a symmetry test (_sym). So matches from image 1 to image 2 are only kept if these are also matches from image 2 to image 1. In the second approach the calculation of a Fundamental Matrix between both images based on the feature matching result using brute force matching is used to eliminate outliers. Therefore, the RANSAC algorithm (_RANSAC) was chosen (Fischler and Bolles, 1981). Additionally, for RIFT the already implemented outlier removal (_native) using the fast sample consensus (FSC) (Wu et al., 2015) is shown.

5. RESULTS

The results of the different feature detection and matching methods are shown in table 1. For every image triple the matches are shown respectively at first between image 1 and image 2, secondly between image 1 and image 3 and at last for image 2 and image 3. Therefore, the number of correct matches (determined using the point transfer with the Trifocal Tensor) is compared with the absolute number of matches and given as ratio in % (also referred to as matching score) with respect to the feature matching method and the applied outlier removal. Good results (> 40 %) are highlighted in green whereas bad results (< 40 %) are highlighted in red. The transition from red to green around 40 % is coloured in white. Additionally, table 2 shows the

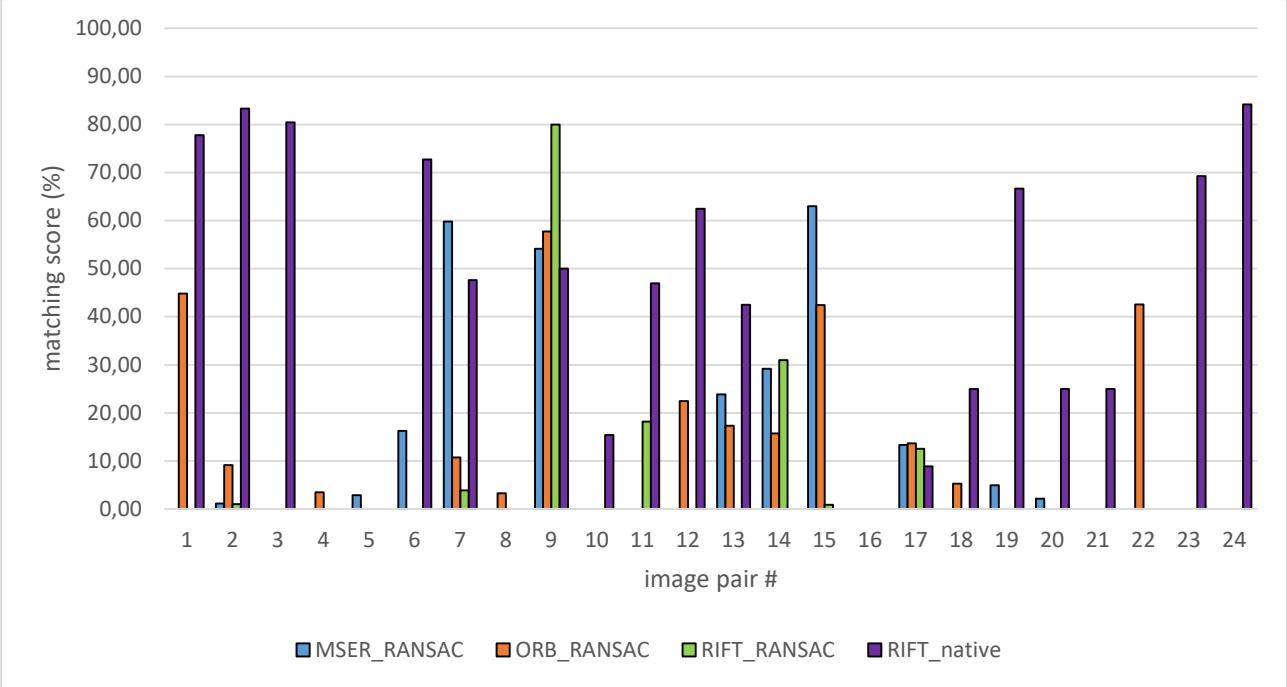

Figure 2. Matching scores of every image pair for the four best performing algorithms MSER_RANSAC, ORB_RANSAC, RIFT_RANSAC, and RIFT_native

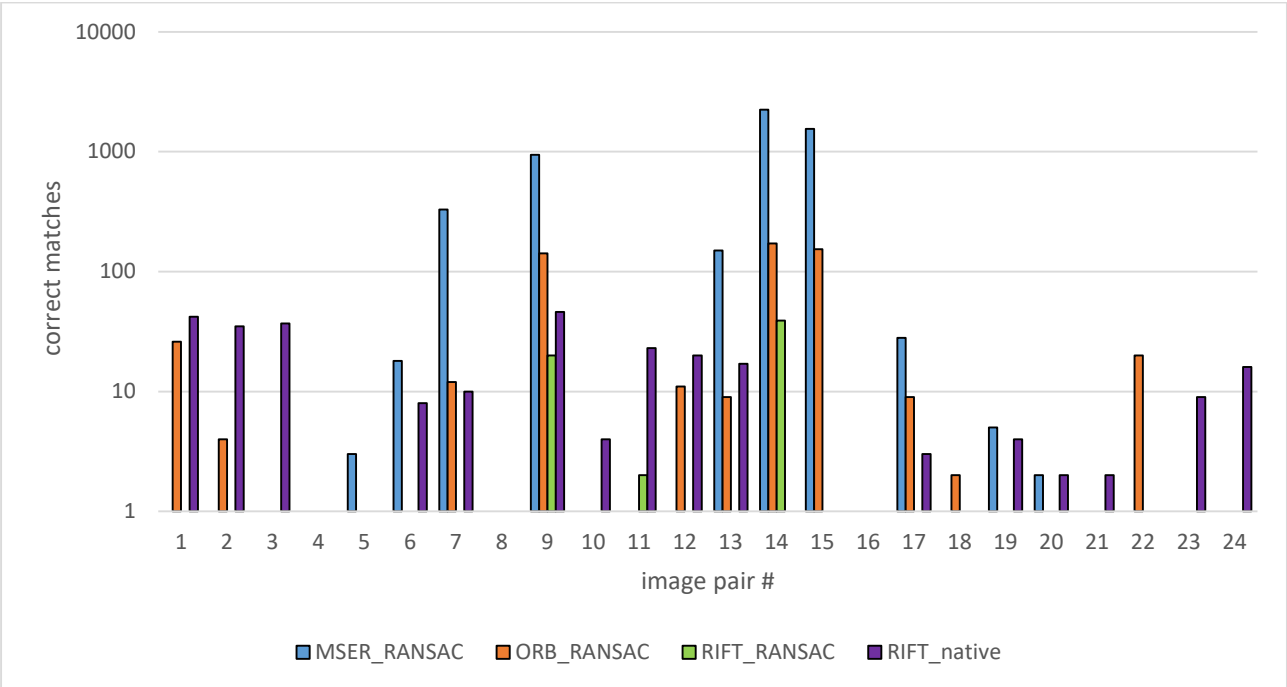

Figure 3. Total number of correct matches shown on a logarithmic scale of every image pair for the four best performing algorithms MSER_RANSAC, ORB_RANSAC, RIFT_RANSAC, and RIFT_native

number of all correct matches determined by the different approaches for the 24 image pairs. The total number of feature points is provided within the dataset. The image pairs are serially numbered from 1 to 24. It is easy to see that for example the image pair 16 could not be matched by any of the algorithms with a good result in opposite to e.g. the image pair 9 where every approach shows better results highlighted in the respective column in green. For an easier comparison the matching scores (ratio) and the number of correct matches of the four best approaches (MSER_RANSAC, ORB_RANSAC, RIFT_RANSAC, RIFT_native) are shown in two different diagrams (fig. 2, 3).

The table as well as the diagrams demonstrate that all three methods fall short of expectations. A small number of correct matches can almost always be found for every image pair with the brute force attempt but it is hardly possible to filter those out. Some exceptions exist like e.g. for image pair 7 and MSER only

around 9 % of the initial matches are correct but with the outlier removal method RANSAC it is achievable to reach a matching score of around 60 %.

However, the native version of RIFT using the fast sample consensus produces better results than the other approaches. For a lot of image pairs, a matching score > 60 % could be attained. But regarding the total number of correct matches (tab. 2, fig. 3) these are very low compared to e.g. MSER but most of the times still enough to perform an estimation of a Fundamental Matrix. The combination of ORB and SURF doesn't outperform any of the other algorithms and is therefore not suitable for the feature matching of the depicted historical images.

100 % correct matches could not be reached with the presented approaches. Consequently, it can be said that the crucial point when working with historical images is the outlier removal step. Since there always is a small number of feature points in the image pairs that could be matched it will be the objective to filter those correctly. The symmetry test only slightly improved the results of the brute force matching. RANSAC performs better but most of the times the exact Fundamental Matrix could not be found. It seems that a refined RANSAC algorithm like FSC that is used in the RIFT approach could improve the matching scores.

A combination of all methods could result in higher scores and will be tested in the future. Multiple iterations when calculating the Fundamental Matrix or improved RANSAC algorithms like FSC, PROSAC (Chum and Matas, 2005) or MSAC (Torr and Zisserman, 2000) could improve the matching scores for all approaches.

Summarizing, for all image triples of the benchmark dataset it is possible to find homologue points and match them almost only using RIFT. MSER generally finds the most feature points but most of the times RIFT shows the highest matching scores in combination with FSC. For some special image constellations, the other approaches could be more appropriate and a combination of methods could lead to better results (Mishkin et al., 2015). Historical images are still a challenge for classical feature detection and matching algorithms, thus a cautious outlier removal is inevitable.

6. CONCLUSIONS AND FUTURE WORK

The contribution shows the generation and evaluation of a dataset consisting of 24 historical images. Difficulties determining the relative orientation of the data arise due to large image differences and unknown camera parameters. Thus, a more stable image configuration using three images described by the Trifocal Tensor T has been established. Therefore, T is given for every image triple in the dataset. The Trifocal Tensor can be used to evaluate different feature detectors and matching methods on historical images and the dataset can be used as a benchmark set. In this research MSER, ORB and RIFT were used since these algorithms have already shown good results in other publications. For the presented dataset RIFT produced better results than the other two methods. FSC performed better in outlier removal than the symmetry test or RANSAC.

It is planned to establish a more reliable workflow for historical image matching using multiple methods consecutively. Also other already developed approaches will be tested on the dataset in the future (Maiwald et al., 2018). Different outlier removal methods could still improve the matching scores. Additional oriented historical images will be added to the dataset to provide a challenging base for other researchers.

Since the images are oriented with the Trifocal Tensor only in a projective space it is planned to use this estimated relative orientation as a base for a metric solution and calculate the inner and exterior orientation of the historical images. In the following, these images could be placed in the 3D/4D web application. Furthermore, simple features like single lines or planes could be generated in 3D space to create generalized historical 3D models.

ACKNOWLEDGEMENTS

The research upon which this paper is based is part of the junior research group UrbanHistory4D's activities which has received funding from the German Federal Ministry of Education and Research under grant agreement No 01UG1630. The image dataset licensed under CC-BY-SA-4.0 is available at https://doi.org/10.25532/OPARA-24.

REFERENCES

Ali, H.K., Whitehead, A., 2014. Feature Matching for Aligning Historical and Modern Images. *IJ Comput. Appl.*, 21(3), pp. 188-201.

Apollonio, F.I., 2016. Classification schemes for visualization of uncertainty in digital hypothetical reconstruction. In, *3D Research Challenges in Cultural Heritage II*. Springer, pp. 173-197.

Bay, H., Tuytelaars, T., Van Gool, L., 2006. SURF: Speeded Up Robust Features. In: *European Conference on Computer Vision*, Berlin, Heidelberg, pp. 404-417.

Bitelli, G., Dellapasqua, M., Girelli, V.A., Sbaraglia, S., Tinia, M.A., 2017. Historical Photogrammetry and Terrestrial Laser Scanning for the 3d Virtual Reconstruction of Destroyed Structures: A Case Study in Italy. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-5/W1, pp. 113-119.

Calonder, M., Lepetit, V., Strecha, C., Fua, P., 2010. BRIEF: Binary Robust Independent Elementary Features. In: *European Conference on Computer Vision*, Berlin, Heidelberg, pp. 778-792.

Chum, O., Matas, J., 2005. Matching with PROSAC-progressive sample consensus. In: *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, pp. 220-226.

Falkingham, P.L., Bates, K.T., Farlow, J.O., 2014. Historical photogrammetry: Bird's Paluxy River dinosaur chase sequence digitally reconstructed as it was prior to excavation 70 years ago. *PLoS One*, 9(4), p. e93247.

Faugeras, O., Papadopoulo, T., 1998. A nonlinear method for estimating the projective geometry of 3 views. In: *Sixth International Conference on Computer Vision, 1998.* , pp. 477-484.

Faugeras, O.D., Luong, Q.-T., Maybank, S.J., 1992. Camera self-calibration: Theory and experiments. In: *European conference on computer vision*, pp. 321-334.

Fischler, M.A., Bolles, R.C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis

and automated cartography. *Communications of the ACM*, 24(6), pp. 381-395.

Geiger, A., Lenz, P., Stiller, C., Urtasun, R., 2013. Vision meets robotics: The KITTI dataset. *The International Journal of Robotics Research*, 32(11), pp. 1231-1237.

Gillet, M., Garnier, C., Flieder, F., 1986. Glass Plate Negatives. Preservation and Restoration. *Restaurator*, 7(2), pp. 49-80.

Giordano, S., Bris, A.L., Mallet, C., 2018. Toward Automatic Georeferencinf of Archival Aerial Photogrammetric Surveys. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 4(2).

Gouveia, J., Branco, F., Rodrigues, A., Correia, N., 2015. Travelling through space and time in Lisbon's religious buildings. In: *Digital Heritage, 2015*, pp. 407-408.

Griffin, G., Holub, A., Perona, P., 2007. Caltech-256 object category dataset.

Grün, A., Remondino, F., Zhang, L., 2004. Photogrammetric reconstruction of the great Buddha of Bamiyan, Afghanistan. *The Photogrammetric Record*, 19(107), pp. 177-199.

Hartley, R., Zisserman, A., 2003. Multiple view geometry in computer vision. Cambridge university press.

Hartley, R.I., 1997. Lines and points in three views and the trifocal tensor. *International Journal of Computer Vision*, 22(2), pp. 125-140.

Heinrich, S.B., Snyder, W.E., Frahm, J.-M., 2011. Maximum likelihood autocalibration. *Image and Vision Computing*, 29(10), pp. 653-665.

Henze, F., Lehmann, H., Bruschke, B., 2009. Nutzung historischer Pläne und Bilder für die Stadtforschungen in Baalbek/Libanon.

Julià, L.F., Monasse, P., 2017. A Critical Review of the Trifocal Tensor Estimation. In: *Pacific-Rim Symposium on Image and Video Technology*, pp. 337-349.

Kensek, K.M., Dodd, L.S., Cipolla, N., 2004. Fantastic reconstructions or reconstructions of the fantastic? Tracking and presenting ambiguity, alternatives, and documentation in virtual worlds. *Automation in construction*, 13(2), pp. 175-186.

Li, J., Hu, Q., Ai, M., 2018. RIFT: Multi-modal Image Matching Based on Radiation-invariant Feature Transform. *arXiv preprint arXiv:1804.09493*.

Maas, H.-G., 1997. Mehrbildtechniken in der digitalen Photogrammetrie. Institut für Geodäsie und Photogrammetrie an der Eidg. Technischen Hochschule Zürich.

Maiwald, F., Schneider, D., Henze, F., Münster, S., Niebling, F., 2018. Feature Matching of Historical Images Based on Geometry of Quadrilaterals. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2, pp. 643-650.

Matas, J., Chum, O., Urban, M., Pajdla, T., 2004. Robust wide-baseline stereo from maximally stable extremal regions. *Image and vision computing*, 22(10), pp. 761-767.

Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., Van Gool, L., 2005. A comparison of affine region detectors. *International journal of computer vision*, 65(1-2), pp. 43-72.

Mishkin, D., Matas, J., Perdoch, M., 2015. MODS: Fast and robust method for two-view matching. *Computer Vision and Image Understanding*, 141, pp. 81-93.

Moreels, P., Perona, P., 2007. Evaluation of features detectors and descriptors based on 3d objects. *International Journal of Computer Vision*, 73(3), pp. 263-284.

Nordberg, K., 2009. A minimal parameterization of the trifocal tensor. In: *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pp. 1224-1230.

Ponce, J., Hebert, M., 2014. Trinocular geometry revisited. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 17-24.

Ressl, C., 2002. A minimal set of constraints and a minimal parameterization for the trifocal tensor. *International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences*, 34(3/A), pp. 277-282.

Ressl, C., 2003. Geometry, constraints and computation of the trifocal tensor. TU Wien.

Rodríguez Miranda, Á., Valle Melón, J.M., 2017. Recovering Old Stereoscopic Negatives and Producing Digital 3d Models of Former Appearances of Historic Buildings. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W3, pp. 601-608.

Rosin, P.L., 1999. Measuring corner properties. *Computer Vision and Image Understanding*, 73(2), pp. 291-307.

Rosten, E., Drummond, T., 2006. Machine learning for high-speed corner detection. In: *European conference on computer vision*, pp. 430-443.

Rublee, E., Rabaud, V., Konolige, K., Bradski, G., 2011. ORB: An efficient alternative to SIFT or SURF. In: *Computer Vision (ICCV), 2011 IEEE international conference on*, pp. 2564-2571.

Schindler, G., Dellaert, F., 2012. 4D Cities: Analyzing, Visualizing, and Interacting with Historical Urban Photo Collections. *Journal of Multimedia*, 7(2), pp. 124-131.

Slate, J.H., 2001. Not fade away: Understanding the definition, preservation and conservation issues of visual ephemera. *Collection management*, 25(4), pp. 51-59.

Torr, P.H., Zisserman, A., 2000. MLESAC: A new robust estimator with application to estimating image geometry. *Computer vision and image understanding*, 78(1), pp. 138-156.

Wolfe, R., 2013. Modern to historical image feature matching. *Published online: http://robbiewolfe. ca/programming/honoursproject/report. pdf*.

Wu, Y., Ma, W., Gong, M., Su, L., Jiao, L., 2015. A Novel Point-Matching Algorithm Based on Fast Sample Consensus for Image Registration. *IEEE Geosci. Remote Sensing Lett.*, 12(1), pp. 43-47.