3.3 Image sequence processing

Anette Eltner, Salvatore Manfreda and Borbala Hortobagyi

Measuring object displacement and deformation in image sequences is an important task in remote sensing, photogrammetry and computer vision and a vast number of approaches have been introduced (Leprince et al., 2007; Alba et al., 2008; Debella-Gilo & Kääb, 2011). In the field of environmental sciences, applications are, for instance, in the studies of landslides, tectonic displacements, glaciers, and river flows (Manfreda et al., 2018). Tracking algorithms are vastly utilized for monitoring purposes in terrestrial settings and in satellite remote sensing, which need to be adapted for the application with UAV imagery because resolution, frequency and perspective are different. For instance, geometric and radiometric distortion need to be minimal

for successful feature tracking, which can be a large issue for UAV imagery in contrast to satellite imagery with much smaller image scales (Gruen, 2012).

Using UAV systems for multi-temporal data acquisition as well as capturing images with high frequencies during single flights enables lateral change-detection of moving objects. And if the topography is known, a full recovery of the 3D motion vector is possible. The underlying idea is the detection or definition of points or areas of interest, which are tracked through consecutive images or frames considering the similarity measures.

In this chapter, pre-processing steps to successful image tracking and vector scaling are introduced. Afterwards, two possible strategies of tracking, i.e. feature-based and patch-based, are explained. Furthermore, different choices of tracking in image sequences are discussed. And finally, examples are given in different fields.

3.3.1 Image pre-processing

UAV image sequences can be either acquired during multiple flight campaigns to observe phenomena evolving at slow rates, e.g. landslide monitoring or during a single campaign focusing on faster change rates, e.g. lava or river flows. In both cases, information about the terrain has to be considered to calculate scaled motion vectors (chapter 3.3.1.1). Thereafter, frame co-registration is necessary for precise tracking of objects. This step becomes more critical when image sequences of high frequencies are captured (chapter 3.3.1.2). Finally, image filtering may be required to increase the robustness of image tracking (chapter 3.3.1.3).

3.3.1.1 Image ortho-rectification

It is important to account for impacts of camera perspective and relief to avoid false scaling of tracking vectors. The objective is the projection of the original image, which might be captured from oblique viewing angles looking at unlevelled terrain, into an image plane to calculate a distortion-free photo where the scale remains constant (Figure 3.3-1). Without this transformation, correct measurements would solely be possible if a planar terrain is captured from nadir view. To achieve the conversion from central projection, i.e. lines of projection intersect at one point (projection centre), to parallel projection, i.e. lines of projection are orthogonal to the projection plane, knowledge about the interior camera geometry, the camera position and orientation during the moment of capture, and the topography is required. This information can be retrieved, capturing overlapping images and using SfM photogrammetry. The result is an orthophoto al-

lowing for distance and angle measurements. You can find more details regarding the process of calculating an orthophoto in chapter 2.2.

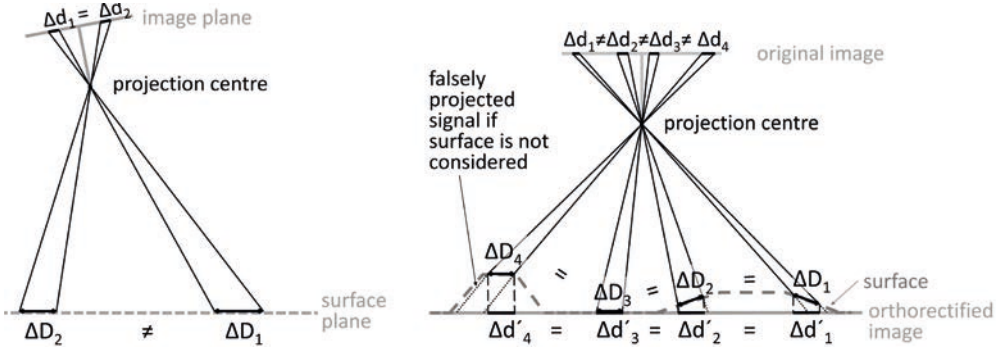

Figure 3.3-1: Captured scene can be distorted due to the influence of camera perspective and relief hindering scaled measurements. Oblique view at a planar terrain leads to increased scale overestimation with increasing distance to the camera projection centre. Terrain deviating from a plane leads to increased scale underestimation with decreased projection centre to object distance. Information about the relief has to be implemented for correct transformation of central projection to parallel projection. All figures were prepared by the authors for this chapter.

3.3.1.2 Image co-registration

To track the displacement of fast-moving objects, such as particles on water, it becomes necessary to capture images in a fast sequence, for instance, using videos. In most circumstances, UAVs are not able to capture the entire event from a stable position and orientation among others due to vehicle drifts and tilts caused by wind and due to vibrations of the sensor. If these movements are not mitigated, they will affect the calculation of correct flow velocity vectors. Therefore, image sequences need to be stabilized exploiting fixed targets, which can be identified in the image sequence.

Image stabilization can be achieved by identifying manually tie points or performing an automatic detection and matching of points of interest (chapter 3.3.2 and 2.2). The information of the corresponding points is used to retrieve the parameters of a transformation matrix between the two images. Usually, either an affine transformation with six parameters (two scales, two shifts, one rotation, and on shear) is considered (Figure 3.3-2b) or a homography with eight parameters is estimated, where lines between both images still remain straight lines after the transformation (Figure 3.3-2c). With the retrieved transformation matrix, the source image will

be converted requiring the interpolation of a new image. In the end, the co-registered image sequence has to be ortho-rectified for correct scaling of tracks (chapter 3.3.1.1) applying the same transformation to all images.

It has to be noted that the approach via tie points assumes that the surface is a plane, which can be a suitable approximation for higher flying heights and/or relatively flat terrain. Another requirement is that the UAV imagery captures stable areas distributed around the area of interest. This is not possible in all scenarios, for instance, if large areas are affected by movements. In such cases, other possibilities need to be considered. One option can be direct referencing (chapter 2.1). However, accuracy demands regarding position estimation with dGNSS, orientation reconstruction with the IMU, and camera synchronisation are very high, and future research has to reveal whether such an approach will be possible.

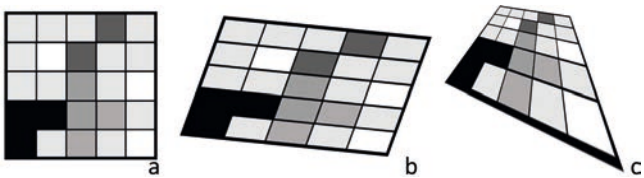

Figure 3.3-2: Distortion of the image due to off-nadir image acquisition and/or sloping terrain. (a) Un-distorted image. (b) Distorted image describable with affine transformation. (c) Distorted image describable with perspective transformation (homography).

3.3.1.3 Image filtering

Tracking objects in image sequences can be sensitive to noise and low signal strength leading to ambiguities. Especially in environmental applications difficulties due to lighting conditions (e.g. glares and shadow) or water turbidity (e.g. transparent, clear water) have to be mitigated. Therefore, different image processing approaches might be considered to increase the robustness of data analysis.

Applying a low-pass filter is a possible method to decrease image noise. An option of image smoothing is convolution. A kernel or window with a specific size is applied to the original image (Figure 3.3-6). Possible kernels are a Gaussian kernel (Figure 3.3-3b), where the weight of the pixel decreases with distance to the centre pixel, a median kernel, which is especially suitable for salt and pepper noises, or a bilateral kernel, where the noise is reduced, but the edges are preserved. Further image improvements are possible via contrast enhancement (Dellenback et al., 2000), gamma correction (Tauro et al., 2017), histogram equalization (Dal Sasso et al., 2018) or intensity threshold criterion (Jodeau et al., 2008).

Another option to increase the robustness of image sequence analysis is the calculation of image derivatives, for instance, considering edges applying a Laplace operator (Figure 3.3-3c). To improve the signal strength, the histogram of the radiometric pixel values of an image can be modified. An example is the adaptive histogram equalization that amplifies the contrast in distinct image regions instead of applying a global histogram change (Pizer et al., 1987). Another approach to improve the signal for tracking is the calculation of derivatives from SfM (chapter 2.2), or Lidar (chapter 2.6) derived digital elevation models (chapter 3.4), e.g. considering hillshades to identify traceable features in the terrain.

Figure 3.3-3: Different options of image filtering to reduce the impact of image noise or to increase the tracking robustness. (a) Original image. (b) Gaussian filtered image for smoothing. (c) Laplace filtered image to keep edges only for tracking.

3.3.2 Feature-based tracking

Feature-based tracking in image sequences can be separated into three processing steps: feature detection, feature description, and feature matching. These steps are similar to the image matching approach during SfM, which was introduced in chapter 2.2. The result of feature-based matching is in most scenarios a sparse set of correspondences. To find distinct and traceable image points, assumptions about the required feature shape are made. The feature has to reveal a large contrast to its neighbourhood, and the strong intensity changes have to occur in at least two directions. First- or second-order derivatives of the image can be calculated to assess the radiometric gradients and their orientation. In flat areas, no changes in all directions are measurable. Along edges, intensity changes occur solely in one direction resulting in ambiguous feature matches. Thus, blobs or corners are the interest operators of choice (Figure 3.3-4). As blob features were already introduced in detail in chapter 2.2, the focus lies on corner features.

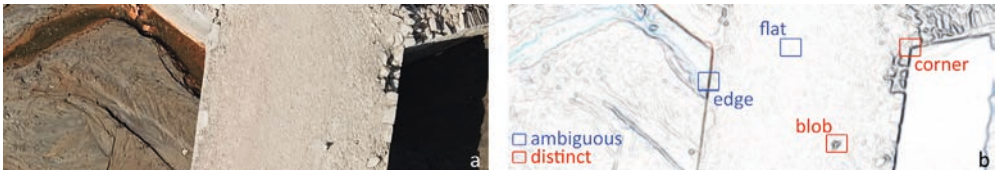

Figure 3.3-4: Examples of unsuitable features as well as corners and blobs as suitable features for tracking. (a) Unfiltered, raw image. (b) Radiometric gradient filtered image.

An example of a corner feature detector is the Harris feature (Harris & Stephens, 1988). Image gradients are calculated via convolution using the Sobel operator. Thus, first derivatives are estimated for both image directions. Within local neighbourhoods, the distribution of the retrieved gradient intensities is assessed, and corresponding eigenvalues are calculated, making the feature detector rotation invariant. Finally, a score is computed from the eigenvalues. Both eigenvalues are high for corners. If they are only high for one eigenvector or low for both eigenvectors, an edge or flat area has been detected, respectively. Another corner feature is the Shi-Tomasi feature (Shi & Tomasi, 1994), which is especially designed for tracking tasks. The approach is similar to the Harris detector, however, the score function is different as both eigenvalues solely have to be above a minimum threshold.

Another possibility to extract features can be simply performed through the binarization of the images and identifying a threshold value, which allows to separate the background from the particles represented by brighter colours. Thus, the pixels at a higher intensity than the threshold will keep their value unaltered and pixels at lower intensities will be assigned a black colour (Figure 3.3-5). The procedure described above is called global threshold, but there are also other methods in the literature, such as: i) local threshold, which overcomes the limits of the global approach, varying the value of the threshold within the image depending on the light intensity, or ii) Otsu's method (Otsu, 1979) which performs clustering-based image thresholding.

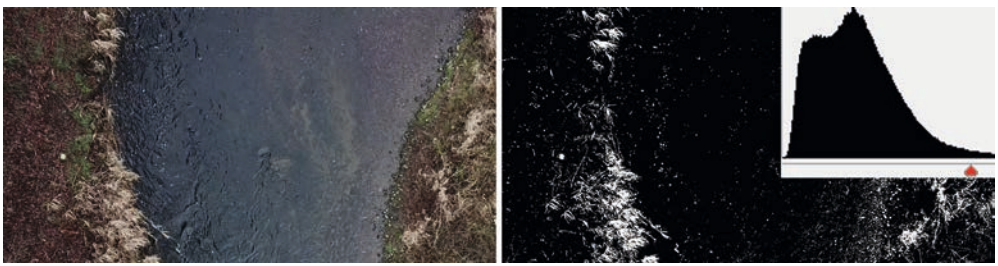

Figure 3.3-5: Binarization of radiometric information to apply a threshold (histogram) to keep points of interest, in this case, floating particles at the water surface.

The extracted features can be either used to estimate descriptors considering their local neighbourhood and subsequently matching these features or the features can be considered as points of interest for a subsequent patch-based matching approach.

3.3.3 Patch-based tracking

Patch-based tracking approaches define areas or patches, which are then tracked by searching for the corresponding location of the highest similarity in the next image. The areas to track can be chosen manually, defining regular grids, or considering the locations of detected features (chapter 3.3.2) to create templates. Dense sets of correspondences are possible, e.g. in the case of the definition of grids with high resolution. In patch-based tracking techniques correspondences are found at locations where matching costs are minimal. Tracking can either be performed in the spatial or the frequency domain.

3.3.3.1 Tracking in the spatial domain

The most common approaches in the spatial domain are represented by the similarity and optimization algorithms. In the case of similarity estimates kernels of finite size, with radiometric information extracted from the source image, are searched for in the target image. Thus, the kernel is moved across the search image to find the position, where the kernel information and the overlapping local target information are most similar (Figure 3.3-6). Different kernel functions can be applied in the convolution, e.g. considering the sum of squared differences (SSD). Another frequently used template matching function is the normalized cross-correlation (NCC), which accounts for brightness and contrast changes to increase the matching robustness. The results of the kernel applications are similarity maps, where the similarity peak (e.g. for SSD and NCC negative and positive, respectively) corresponds to the final position of the tracked feature.

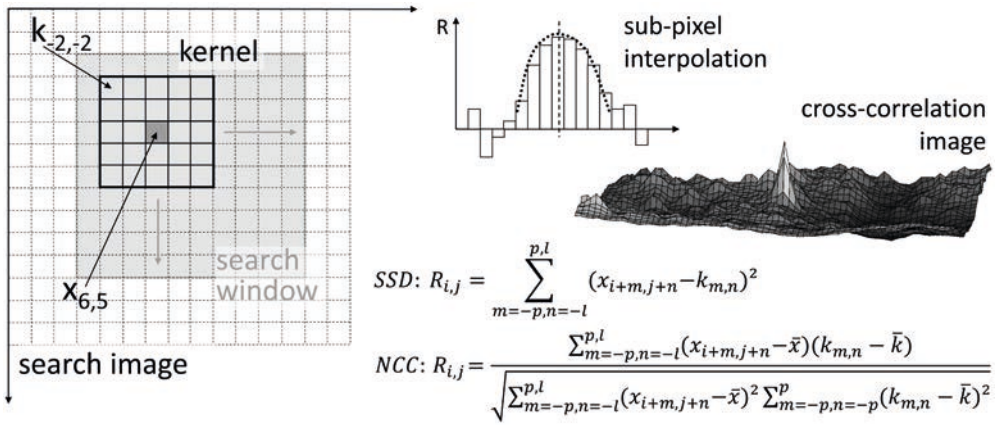

Figure 3.3-6: Patch-based tracking approaches. Kernel k with information of image x-1 (source image) sliding across search image x (target image). At each pixel position $x_{i,j}$ in the extracted patch of the search image, corresponding to the overlapping area of the kernel, is computed with the kernel applying different functions. Different similarity measures R can be considered, e.g. SSD (sum of squared differences) or NCC (normalized cross-correlation). Image displays a cross-correlation map, where NCC values were computed using a moving window over the search area. Diagram illustrates a 1D representation of sub-pixel interpolation by estimating the extreme value for a Gaussian fitted curve to NCC values along the x-axis of similarity image.

SSD and NCC have the disadvantage that both measures are sensitive to rotation, scale changes and shear. However, other patch-based matching such as optimization algorithms can overcome these constraints. An example is represented by the least-square-matching (LSM; Ackermann, 1984; Förstner, 1982). LSM searches for the transformation matrix between two image patches such that the square of sums of grey value differences is minimized. For instance, if it is assumed that the corresponding patches are located in a plane, six parameters of an affine transformation are estimated (Figure 5.3-2b). This enables the tracking of distorted features, e.g. at stretching landslides, buckling glaciers, or rotating particles on rivers. The optical flow algorithm Lucas-Kanade (Lucas & Kanade, 1981), increasingly used in hydrological tracking tasks, is another optimization approach fitting an affine model to the motion field. Sub-pixel accurate measurements are possible, and the statistical output of the adjustment can be used to assess the matching quality. Due to the non-linearity of the adjustment, approximation values are required, which can be provided assuming solely minimal changes between images (e.g. in the case of high-speed imagery or very slow-moving objects), using the results of other matching approaches (e.g. NCC) as first estimates, or considering hierarchical approaches (chapter 3.3.3.3).

3.3.3.2 Tracking in the frequency domain

To find the position of highest similarity, it is also possible to estimate displacements in the frequency domain using the Fourier transformation. The phase correlation approach (e.g. De Castro & Morandi, 1987) calculates the cross-correlation between the Fourier transformed search and kernel patch to retrieve the phase shift in the frequency domain and thus lateral shift between both image patches in the spatial domain (Figure 3.3-7). Finding matches in the frequency domain is significantly faster than measuring in the spatial domain.

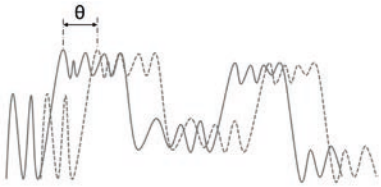

Figure 3.3-7: Simplified 1D representation of measuring phase shift θ between search (target) and source object in the frequency domain to retrieve displacement.

3.3.3.3 Improving robustness and accuracy

In most cases of patch-based tracking, the feature to track will not be located at the pixel centre in the search image due to signal discretization, i.e. the conversion of a continuous signal to a discrete (integer) value during the image capture process. Thus, to improve the matching accuracy sub-pixel estimation can be necessary. One approach is the fitting of a paraboloid (Figure 3.3-6) at the position of the highest score in the similarity map and then extracting the coordinates at the local extreme value. The advantage of that method is that also the strength of the match can be evaluated considering the steepness of the paraboloid. Further parameters for quality assessment of the similarity measure are height and uniqueness of the estimated values.

Patch-based matching approaches can be further improved regarding their robustness and accuracy with hierarchical methods, which build image pyramids made off increasingly downsampled images to incrementally decrease image resolution (Figure 3.3-8). The tracking will start at the highest pyramid level, thus at the image with the lowest resolution. The search area can cover nearly the entire image. The position of the matching result is used as an approximation to confine the search area in the next pyramid level. These steps are repeated until the last level with the full image resolution, where the final location of the match is extracted. The hierarchical approach enables to mitigate the impact of choosing the right kernel and search window sizes. The larger the kernel is chosen, the less sensitive it is to ambiguities due to repeating patterns and

the smaller it is chosen, the higher the accuracy will be because more details are captured. And the larger and smaller the search window is chosen, the larger displacements can be captured and the faster processing times are achieved, respectively. Therefore, applying image pyramids allows for processing from the stage of high robustness at the first low-resolution levels to the stage of high accuracies at the last high-resolution levels.

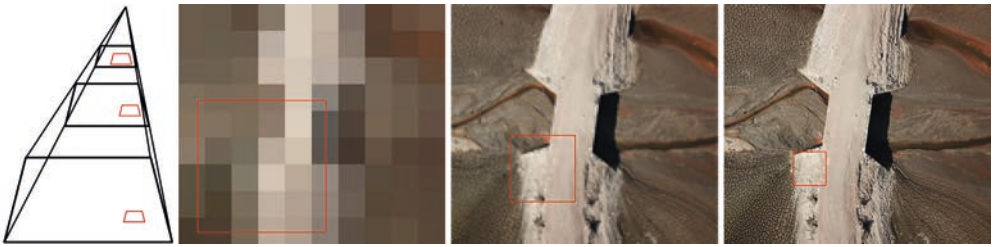

Figure 3.3-8: Applying image pyramids to improve the tracking robustness and accuracy. The highest level corresponds to the image of the lowest resolution (first image), and the base level corresponds to the image of the highest resolution (last image). The matching result at each level serves as an approximation for the next level. The kernel has the same number of pixels in each level, and therefore different areas of the scenery are covered. Note that kernel size and downsampling are not scaled accordingly in this example to enhance the visibility of changes at different levels.

A further option to increase the accuracy of the tracking is the application of filtering algorithms to the final tracks. These can be either used globally, considering, e.g. the average and standard deviation of all measured displacements to identify outliers, or locally, considering, e.g. displacement statistics only within a specified neighbourhood. The latter approach is especially useful for objects with complex movement patterns.

3.3.4 Tracking strategies

Different spatial tracking strategies are possible for successful estimation of velocities and direction of moving objects in UAV image sequences. First of all, it has to be considered if tracking is performed in stationary image sectors, thus where in each subsequent image tracking starts again at the same image coordinate, i.e. Euler approach, or if the track of a specific target in the image sequence is searched for, i.e. Lagrangian approach. The Euler method is generally computationally more efficient with respect to the Lagrangian method. In return, the latter approach is able to perform measures also with low tracer density, whereas the former relies on abun-

dant seeding density. To identify matching regions or features, the concept of similarity between groups of particles in two consecutive images is used, but it is also possible to use multi-frame algorithms that use three or more consecutive frames to solve the problem of correspondences.

Once the particle positions are identified, the velocity is estimated by dividing the displacement of particles between consecutive frames by the time interval between the pair of images. A finite difference scheme is applied implicitly for calculating the velocity. Therefore, the temporal accuracy is directly correlated to the image frequency. Sampling frequency must be identified properly in order to avoid over- or undersampling that may lead to missed features or high velocity uncertainties if displacements are happening at the sub-pixel range, respectively. Different temporal tracking strategies are possible with different temporal bases, overlap and resolutions (Schwalbe, 2013, Figure 3.3-9). For instance, in a scenario of very slow-moving particles captured with high framerates, instead of tracking consecutive frames illustrated by strategy two in Figure 3.3-9, it might be suitable to skip frames and track features subsampling frames at a lower frequency. This may help to enhance the visibility of shifts and movements of objects within each frame. Thereby, features or patches might be detected, e.g. every frame or every second frame (strategy four and three in Figure 3.3-9, respectively).

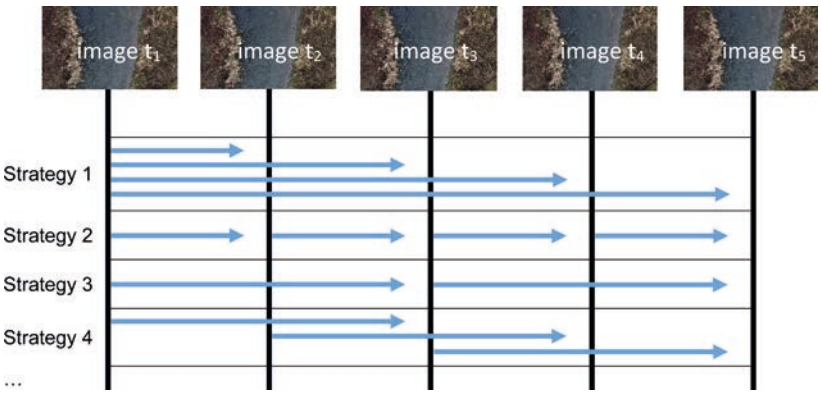

Figure 3.3-9: Temporal matching strategies (after Schwalbe, 2013).

To transform the measurements within the image sequences into displacements in a scaled coordinate system and correspondingly to metric velocity values, it is necessary to reference the tracking result (chapter 3.3.1.1). Referencing can be either performed prior to the tracking processing or afterwards. Executing the tracking in the original image, and thus transforming the image measurement afterwards, only considering the coordinates of the tracked particles, entails the advantage that interpolation errors, especially in strongly tilted images, are avoided.

3.3.5 UAV monitoring applications

The applications of tracking approaches to UAV data are vast and therefore entail very case-specific challenges. Therefore, we display three common fields of application – hydrology (chapter 4.3), geomorphology (chapter 4.2) and glaciology (chapter 4.5) – to highlight different advantages, challenges and limits of image sequence analysis of UAV-based data.

3.3.5.1 Streamflow

Image-based flow velocity measurement with UAV imagery is a valuable emerging flow gauging technique, which can also be applied to terrestrial images captured by fixed station or mobile stations (Eltner et al., 2020). The advantage of using UAVs is the possibility for greater coverage of the river surface at multiple locations, including potentially inaccessible sites. Furthermore, they tend to fail less at high flow conditions compared to classical monitoring systems.

A vast number of methodological approaches are available to compute water surface velocities. The most frequently adopted algorithms are large scale particle velocimetry (LSPIV, Le Coz et al., 2010), belonging to the Euler tracking strategy, and particle tracking velocimetry (PTV, Tauro & Grimaldi, 2017), belonging to the Lagrangian tracking strategy. LSPIV is an adaption of particle image velocimetry (PIV, Creutin et al., 2003). In contrast to PIV, LSPIV can be used for a wider range of physical phenomena due to its capacity to cover larger areas and to adopt low-cost cameras. Regardless of the specific algorithm considered for tracking, the estimated velocity is recovered from the information of tracing features on the water surface, i.e. natural foam, seeds, woody debris, and turbulence-driven pattern.

Accuracy assessments of UAV image velocimetry revealed that stationary UAV measurements are in strong agreement with established flow gauging approaches. To better understand the complexity of 2D river flow structures, following major points have to be respected: i) the stability of the camera, ii) a good compromise between flight altitude, camera resolution, tracer particle size and river width (Lewis & Rhoads, 2018), iii) the potential necessity of non-oblique UAV imagery at wider rivers to enable the coverage of the entire cross-section, and iv) the presence of a traceable pattern on the water surface. Seeding density is one of the most relevant parameters in the determination of reliable velocity fields. When facing low seeding density conditions, the number of analysed frames should be increased for more accurate results (Dal Sasso et al., 2018).

3.3.5.2 Landslide

UAVs offer a cost-effective, time-efficient, flexible and safe data collection solution to improve the spatio-temporal resolution of landslide movement maps (chapter 4.2), e.g. through the comparison of SfM-derived co-registered digital surface models (DSM) or using multi-temporal orthophotos. Landslide tracking techniques applied to satellite, airborne or terrestrial data cannot be easily transferred to UAV-imagery, due to the different monitoring scales. Therefore, Lucieer et al. (2014) applied the COSI-Corr (co-registration of optically sensed images and correlation) algorithm (Ayoub et al., 2009) to hill shaded DSMs, instead of RGB imagery, to measure landslide movements. In a further step, other UAV-derived morphological attributes, such as slope, openness and curvature, can be considered (Peppa et al., 2017). Furthermore, feature tracking approaches based on terrain break-lines can be more suitable to detect landslide movements with important surface deformation, whereas NCC-based correlation can be more appropriate when targeting small landscape elements.

The presence of vegetation can become an important challenge. For instance, image cross-correlation performance decreases when terrain surface is covered with grass. And vegetation's negative effect on correlation is even more pronounced when images were produced in different seasons (e.g. spring and winter). Although some errors are expected, especially over regions with rotational failures, UAV-based methods offer a reliable quantification of translational earth-flow activity, in particular, movement of ground material pieces, vegetation patches and landslide toes (Lucieer et al., 2014; Peppa et al., 2017).

3.3.5.3 Glacier

Similarly to landslide monitoring, UAV-acquired data can be beneficial to better understand glacial dynamics (chapter 4.5). However, applying UAV image-based processing can be particularly challenging in these landscapes due to large uniform surfaces, but whose texture can be enhanced by the presence of dust or debris. One of the challenges when quantifying glacier velocity is isolating ice movement from other surface displacements (e.g. debris slope collapse or falling blocks from the moraine on the ice surface) (Rossini et al., 2018). Application of a multi-scale mode, implemented in COSI-Corr, allowed for the exclusion of the majority of these noises. The best results involved a trade-off between limited noise, when using larger correlation windows, and fine-scale details. Besides orthomosaic, hillshaded DSMs and DSM derivates, e.g. detected edges, can also provide a globally coherent output. Feature-tracking algorithms used to compute glacier surface velocity can perform similarly compared to manual digitalization, and they enable fine spatio-temporal displacement quantification of debris-covered glaciers (Rossini et al., 2018).

References for further reading

Lucas, B. & Kanade, T. (1981): An iterative image registration technique with an application to stereo vision, in: Proceedings of the 7th International Joint Conference on Artificial Intelligence, 121–130.

Schwalbe, E. & Maas, H.-G. (2017): The determination of high-resolution spatio-temporal glacier motion fields from time-lapse sequences, in: Earth Surface Dynamics, 5, 861–879.

Szeliski, R. (2010): Computer Vision: Algorithms and Applications. New York: Springer-Verlag New York Inc.

Thielicke, W. & Stamhuis, E. (2014): PIVlab – Towards User-friendly, Affordable and Accurate Digital Particle Image Velocimetry in MATLAB, in: J. Open Res. Softw., 2, e30.