POSE ESTIMATION OF WEB-SHARED LANDSCAPE PICTURES

Timothée Produit[1]*, Devis Tuia[1], Vincent Lepetit[2], François Golay[1]

[1]LaSIG laboratory, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland
[2]Institute for Computer Graphics and Vision, Graz University of Technology, 8010 Graz, Austria

Commission III, WG III/1

KEY WORDS: photosharing, landscape images, pose estimation, georeferencing, horizon, DTW, DEM, GIS

ABSTRACT:

We propose a robust method for registering high oblique images of landscapes. Typically, an input image can be registered by matching it against a set of registered images of the same location. While this has been shown to work very well for images of popular urban landmarks, registering landscape images remains a very challenging task: For a given place, only a very small amount of registered images is generally already available on photo-sharing platforms. Moreover, the appearance of landscapes can vary drastically depending on the season and the weather conditions. For these two reasons, matching the input images with registered images in a reliable way remains a challenging task. Our contribution is two-fold: first, we show how to estimate the camera orientation for images with GPS data using a novel algorithm for horizon matching based on Dynamic Time Warping. The proposed algorithm exploits an elevation model. Each image is processed independently from the others, there is therefore no need neither for image matching or for a large set of images. This step provides a set of reliable, fully registered images. Second, and in odrer to register new images with no GPS data available, we first ask the user to provide an approximate image localization on a 2D map. Then, we exploit this prior on the camera location to efficiently and robustly constrain and guide the matching process used to register the query image. We apply our method to a case study from the Zermatt area in Southern Switzerland, and show that the method provides registrations, which are accurate enough to map each pixel to an aerial map.

1 INTRODUCTION

Many recent works have shown that 3D reconstruction of famous monuments or of entire urban areas using tourist photos only is possible (Snavely et al., 2006, Strecha et al., 2010, Agarwal et al., 2011). Such reconstructions are usually computed in an arbitrary coordinate system and are then georeferenced afterwards, either with Ground Control Points (GCP) or with some reference GIS data, such as a 3D city model or roads vector layers (Strecha et al., 2010). By contrast, much fewer works have considered landscape areas imaged with oblique photography. This is mainly because in natural landscapes the images are much more sparsely distributed and prone to illumination and seasonal variations (Baboud et al., 2011). With these constraints in mind, applying bundle adjustment becomes difficult, if at all possible (Chippendale et al., 2008). Nonetheless the possibility of exploiting tourist images is very appealing: precisely georeferenced photo images could become extremely valuable alternative data sources for Earth sciences, which could be used in the estimation of the snow melting rate, of the displacements of glaciers, or of land cover changes.
A second motivation is related to web sharing platforms and landscape image database management. Web sharing applications usually provide 2D maps with pictures location. However, images with associated 3D pose would open new opportunities. First, make 3D navigation between images possible in a virtual globe. Second, a 3D pose is required to augment images with georeferenced data (rivers, trails, toponyms etc.). Those attributes will also ensure that correct tags are associated to the images for more advanced querying.

This paper aims at providing a semi-automatic 2D-3D registration methodology for landscape pictures issued from the Web: we present an automatic pose estimation workflow for sets of landscape tourist images. The georeference of each photography is estimated using two data sources: a small set of GPS-registered images and a Digital Elevation Model (DEM). To estimate the camera orientations for these GPS-located images, we propose an algorithm that matches the horizon silhouette in the images with the DEM. Once oriented, the 3D coordinates of each GPS image is computed. To register a query image, we start from a coarse location provided by the user by clicking on a 2D map. We use this location as a prior information in a our proposed algorithm that simultaneously estimates the camera pose and establishes SIFT 2D-3D point correspondences between the query image and the GPS-registered images. Finally, the horizon line is further used to fine-tune the camera orientation.

The proposed workflow is applied to a set of pictures of one of the most attractive places in the Swiss Alps: the region of Zermatt and the Matterhorn. Query images are downloaded from the web-sharing service Panoramio. To assess the accuracy of the resulting orthorectifed images, we compare them with a state-of-the-art orthoimage of the area. This comparison shows an 50 m accuracy on the average pixel localization.

The remainder of the paper is as follows: Section 2 presents related work. Section 3 introduces the pose estimation problem and the ingredients of the proposed workflow, which is detailed in Section 6. Section 7 presents the dataset studied and the experimental results, which are discussed in Section 8. Section 9 concludes the paper.

2 RELATED WORK

Oblique terrestrial images are often used to assess the changes of a landscape (Roush et al., 2007, Kull, 2005, Debussche et al., 1999). Oblique images have several advantages compared to aerial images. First, fixed terrestrial cameras such as webcams can observe landscapes continuously, from the same location and with high temporal and spatial resolutions. Second, these cameras are cheap and do not require skilled operators running them. Finally, for studies considering past images (∼1850-1920), mostly only terrestrial images are available.

*Corresponding author: timothee.produit@epfl.ch

Despite all these advantages, georeferencing routines for oblique terrestrial images are less widespread than those dedicated to sets of aerial or satellite images. Among the papers proposing georeferencing solution for oblique photography, (Corripio, 2004) assess movements of glaciers, while methods in (Bozzini et al., 2011, Produit et al., 2013) georeference historic images. These papers propose tools to compute the pose of an image from user-defined GCP and facilitate the interaction between the oblique image and the map. Once the images have been registered, one can measure objects displacements of augment the image with vector information.

Currently, large efforts are done in the automatic pose estimation of images, either at the global or local scale (Li et al., 2012, Crandall et al., 2009, Hays and Efros, 2008, Friedland et al., 2011, or the MediaEval Benchmarking Initiative[1]). At local scale (and specifically for landscape images), authors showed the potential of using GIS data as landmark features or models (see (Jacobs et al., 2007, Hammoud et al., 2013, Produit et al., 2014)). A straightforward landmark element in mountainous area is the horizon, which can be easily extracted from a DEM. It has been used in (Baatz et al., 2012a) to recover an image location and orientation at the scale of the Switzerland. A DEM provides alos other morphologic edges: authors in (Baboud et al., 2011, Chippendale et al., 2008) use this source to orient the images. Finally, the DEM can also be warped with landcover maps (Baatz et al., 2012b) or orthoimages to render more realistic views that can in turn be used as reference for matching query images (Produit et al., 2012). Both these methods allow to avoid explicit matching of the horizon, but the first requires precise knowledge about the camera position and the second is specific to glacier areas with high contrast geomorphological features.
Among the works above, the closest to ours is the one of (Baatz et al., 2012a). The authors build a database of horizon contourlets for the whole territory of Switzerland. This way, they locate 88% of the query images within 1km of their real location. This result is very impressive and promising, but has some drawbacks. First, the method relies only on the horizon, which in the query images can be hidden behind clouds or other foreground objects. Second, the camera orientation is obtained from the horizon silhouette matched with ICP (Iterative Closest Point): the matching is done in 2D, while the horizon extracted from the DEM has 3D coordinates, which can be used to compute a finer pose. Finally, the horizon segmentation in (Baatz et al., 2012a) requires user intervention for 49% of the images, thus reducing the degree of automation of the system. For all these reasons, we believe that complementary methods still need to be developed.

Landscape images collections differ from urban images mostly used in comparable studies in several ways. First, images density is usually lower in rural area and their overlap is small. Pictures are generally shot from some easily accessible locations and only few pictures can be found in between these popular locations. Second, landscape images show large illumination variations due to daylight changes and seasonal effects. Third, colors, textures and land cover change during the year. For these reasons, traditional bundle adjustment is difficult. Hence, and at the best of our knowledge, there is no workflow able to estimate the pose (location and orientation) of landscape images, if structure-from-motion can't be applied.
In this paper, we describe a workflow to address these problems and provide the georeferencing of a set of landscape images with limited user interaction. Such a process is well adapted for current image databases, which often contain only a few precisely located images (GPS acquired) and a wider set of roughly located ones (toponym or click on a map). In this paper, we focus on landscape images shared on the web. In the workflow proposed, the pose of each image is estimated sequentially using the 3D model of the area and some pose priors.

[1] http://www.multimediaeval.org

3 PROBLEM SETUP

In this section, we briefly present the basic concepts and equations used in the proposed workflow.
We use the collinearity equations to describe the relation between the 3D world coordinates $\mathbf{X} = [\text{Easting, Northing, Height}]$ and the corresponding image coordinates $\mathbf{u}$. First, a translation and a rotation transform the world coordinates $\mathbf{X}$ in the camera frame with coordinates $\mathbf{x}$:

$$\mathbf{x} = R(\mathbf{X} - T) \tag{1}$$

where R is a rotation matrix parameterized by three Euler angles and $T = [E_0, N_0, Z_0]$ is the location of the camera. The camera coordinates frame has two axes parallel to the image sides and the third one is pointing in the viewing direction. Then, image coordinates $\mathbf{u}$ are obtained with:

$$\mathbf{u} = \begin{bmatrix} u \\ v \end{bmatrix} = -c \begin{bmatrix} x_1/x_3 \\ x_2/x_3 \end{bmatrix} \tag{2}$$

In this formulation, the image pose $\mathbf{p}$ is made of the three Euler rotation angles forming the rotation matrix R and the three camera location coordinates in T. Thus, 6 parameters are unknown. The constant c corresponds to the focal length:

$$\mathbf{u} = \begin{bmatrix} f_1(p, \mathbf{X}) \\ f_2(p, \mathbf{X}) \end{bmatrix} = F(p, \mathbf{X}) \tag{3}$$

The pose estimation from n 2D-3D correspondences is defined as a least squares problem:

$$\min_{p \in \mathbb{R}^6} \sum \left\| F(\hat{p}, \mathbf{X}) - \mathbf{u} \right\|^2 . \tag{4}$$

4 DYNAMIC TIME WARPING FOR HORIZON MATCHING

To estimate the camera orientation when its location is available, we match the horizon in the image with the DEM. This is done with a novel algorithm based on Dynamic Time Warping (Berndt and Clifford, 1994), a technique matching sequences of different length based on the warping and distance measurements.
Typically, the horizon silhouette is extracted from a rendered view or a synthetic panorama and assuming the camera roll and tilt equal to zero. In this specific setup, DTW shows some desirable properties. First, the warping is used to be robust to image or panorama distortions, inaccurate focal estimation and rendering artifacts.

DTW was originally developed to match two time series with time distortions, acceleration and deceleration, sampled at equal time steps. In our problem of horizon matching, time is replaced by the image coordinates on the horizontal axis u_0. The first series is the set of N query horizon features $h^Q = \{\mathbf{u}_0, ..., \mathbf{u}_N\}$ and the second is composed of M reference horizon features $h^R = \{\mathbf{u}_0, ..., \mathbf{u}_M\}$, both ordered by increasing horizontal coordinate. DTW computes a set of correspondence pairs $C_{i \in [0,N], j \in [0,M]} = \{..., (i,j), ...\}$, which minimizes the global performance measure between the matched h^Q and h^R. The global performance is the sum of the Euclidean distances measured between corresponding features in C. Moreover, each horizon feature in the reference image is linked to a 3D coordinate $H^R = \{\mathbf{X}_0, ..., \mathbf{X}_M\}$.

Matches computed with DTW are iteratively inserted in the least-square problem of Eq. (4) in order to re-estimate the orientation and compute a more accurate reference horizon, by fixing the

translation to the GPS data and optimizing the rotation only. Iterations are stopped once the orientation is stable or after a predefined number of iterations. In this way, the global measure is used to detect the best azimuth and 2D-3D correspondences are used in an iterative way to retrieve accurately the azimuth, roll and tilt and thus ensure that each pixel will be accurately georeferenced.

5 FULL POSE ESTIMATION VIA KALMAN FILTERING

If a pose p and its variance Σ^p are known, variance propagation is used to estimate the covariance ellipses of a 3D correspondence X_i projected in the image frame :

$$\Sigma_i^u = A_i \Sigma^p A_i^T \tag{5}$$

where A_i is the jacobian of $F(p, \mathbf{X}_i)$. In our implementation, J sets of 2D-3D correspondences are found iteratively. The problem can be reformulated in the following form to be used in a Kalman filter:

$$\min_{p \in \mathbb{R}} \sum_{j=1}^{J} \|A_j(p, \mathbf{X}_j) - \mathbf{u}_j\|^2 \tag{6}$$

Following the Kalman filter algorithm, if a pose p_j is associated with a covariance matrix Σ_j^p and the noise covariance of the measurement is Σ_j^u. Then the gain K_j is:

$$K_j = \Sigma_j^p A_j^T (A_j \Sigma_j^p A_j^T + \Sigma_j^u)^{-1} \tag{7}$$

The gain is then used to update the pose:

$$p_{j+1} = p_j + K_j(\mathbf{u}_j - A_j p_j) \tag{8}$$

And to update the pose covariance:

$$\Sigma_{j+1}^p = (I - K_j A_j)\Sigma_j^p \tag{9}$$

In the proposed workflow, a recursive pose estimation process is needed to refine the 2D-3D correspondences extraction for each estimation of an image pose. The Kalman filter has the advantages of using prior information and of keeping a memory of the process, which limits the influence of false positives. Moreover, it is able to deal with several types of measurements, for example an update of the camera height with the DEM. However, since in this case study our focus is the minimization of the reprojection error and not the exact camera location computation, we did not include the height update.

6 PROPOSED WORKFLOW

In the previous section, we defined the tools required for the pose estimation of landscape images. In this section, we first explain how images having known location (e.g. acquired with a GPS-enabled camera) are oriented. Then, we present how these images are used as references for the pose estimation of the remaining images in the database.

6.1 Orientation of the GPS images using the horizon

A small part of images found in photosharing platforms are geolocated with a GPS device. GPS mounted on cameras and cellphones have a sufficient accuracy for our purposes ($< 20m$). In particular, our study area is open and above the forests limit, thus no obstructions should disturb the GPS measurements. For this small set of images, we assume that the camera location T is known, while the orientation is unknown. It has been shown that

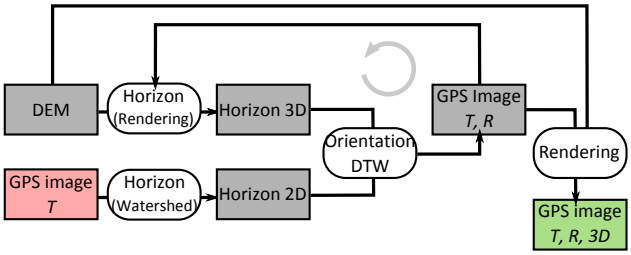
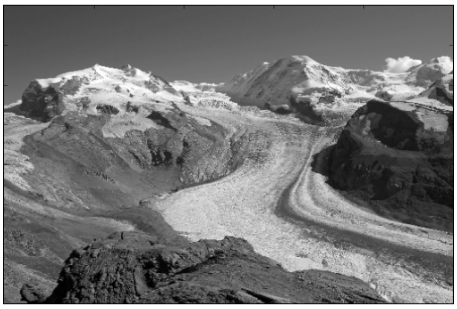
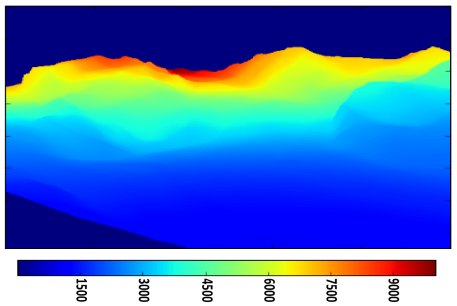


Figure 1: Processing of the GPS images, T refers to the camera location measured with the GPS, R is the orientation of the camera. At the end of this step, each GPS-image pixel owns a 3D coordinate.



(a)



(b)

Figure 2: (a) One of the reference GPS images; (b) Corresponding distance matrix: each pixel is coloured as a function of its distance to the camera (in meters).

the camera orientation in mountainous area can be automatically retrieved (e.g. (Baboud et al., 2011)). Nevertheless, as in (Baatz et al., 2012a) we choose to involve the user in the extraction of the horizon in order to ensure an accurate delineation and at the same time avoid heavy pose computation. We propose to extract it using a watershed segmentation, where the user provides an initialization region. Indeed, a precise horizon delineation is essential for a precise camera orientation: the user involvement ensures a trustable horizon detection, especially in cloudy images and ensures that those images used as reference later are well oriented.

Following the process presented previously for DTW horizon matching, 2D-3D correspondences in the horizon are used to compute the image orientation. Ultimately, 3D coordinates of each one of the GPS image pixels are computed by projecting the DEM in the image plane via the z-buffer method. Figure 1 summarizes the workflow applied to get the orientation of the GPS images, while Figure 2 illustrates an example of 3D coordinates of an oriented image obtained with the proposed method.

6.2 Pose estimation of the remaining images

The remaining images only have inaccurate geotags provided by the user with a click on a map. The proposed workflow is inspired by the one of (Moreno-Noguer et al., 2008) in which camera pose

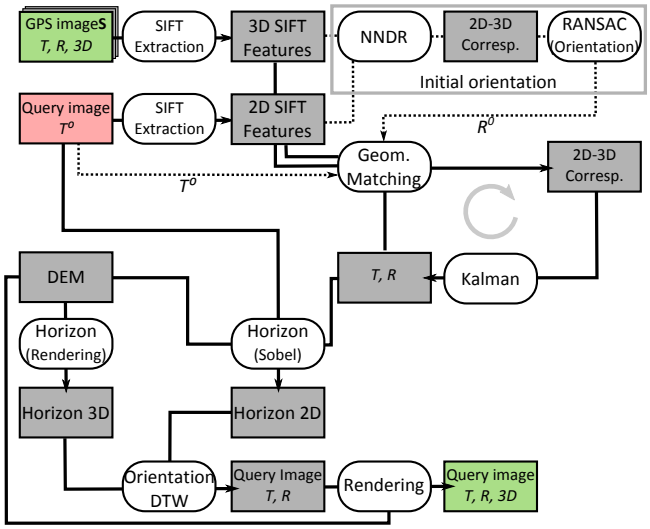


Figure 3: Processing of the query images, T^0 refers to the geotag, R is the orientation of the camera.

prior are known and 2D-3D correspondences are detected solely based on the geometric location of features. In our case the prior is the geotag and orientation extracted from reference images. We add feature descriptions to the process, since they are also available. The involvement of the 3D coordinates of the features detected in the reference images has two effects: on the one hand, 3D coordinates are used to restrict possible matches and to adapt the SIFT matching threshold, while, on the other hand, 3D coordinates in a global coordinates system allows to take advantage of small overlaps in non-densely photographed areas.

The workflow of this second phase is summarized in Figure 3. In the first part, the geometric constraint requires an estimation of the camera location T^0 (which is obtained with the geotag) and an estimation of the orientation R^0. To compute the latter, SIFT features are extracted from the collection of reference GPS images. Features from every reference image are merged into one set of n features R_i with corresponding 3D coordinates $\mathbf{X}_i^R$ and SIFT descriptions $\mathbf{d}_i^R$. Given a query image, a set of m features Q_i are computed and each of them has an image location $\mathbf{u}_i^Q$ and a description $\mathbf{d}_i^Q$. SIFT features are firstly matched according to the ratio of the two closest features descriptions found in the other set (Nearest Neighbour Distance Ratio, NNDR). Using the NNDR, a set of 2D-3D correspondences is extracted. Those first matches are used to compute the initial camera orientation with RANSAC in conjunction with the camera orientation model (the camera is initially fixed at the geotag location T^0, figure 4(a). Typically, few and poorly distributed matches are found, and this is why we do not use RANSAC in association with the pose estimation model. The pose p_0^Q is associated with a high standard deviation Σ_0^p and the matches are associated with noise Σ_0^u.

During the second part, corresponding to the geometric matching, the reference features $\mathbf{X}_i^R$ are projected in the query image: $\mathbf{u}_i^R = f(p^Q, \mathbf{X}_i^R)$ and the variance propagation in Eq. (5) is used to compute the corresponding covariance matrix Σ_u^R. Based on the covariances, ellipses are drawn in the query images and used to constrain the SIFT feature matching, as illustrated in Figure 4(b-c). The Kalman filter prerequisites are met and with each new bloc of 2D-3D correspondences, a new pose is computed. During the iterations, the SIFT distance threshold between two features descriptions is relaxed to take into account texture and illumination variations.

At this stage, the estimated pose is quite accurate, but the alignment with the 3D model is still not exact because of the pose inaccuracy of some reference images and of accumulated errors. We propose to update the pose with an additional horizon alignment. This time, the current pose and its covariance are used to delineate the region in the query image where the horizon should be located. Then, a Sobel edge extractor is applied inside this region and the DTW algorithm is used to refine the orientation, as shown in Figure 4(d).
At the end of the process, the query image is draped on the DEM to generate an orthorectified image that can be used in a GIS (Figure 4(f)).

7 RESULTS

In this section, we apply the proposed workflow to a real collection of landscape images shared on the web. Particularly, with limited user interaction, our method can be used to compute the pose of the oblique image and orthrectify it.

7.1 Data

The area of interest is located in the southern Swiss Alps in the surroundings of the famous Zermatt ski resort. From Zermatt, a train brings people to the *Gornergrat*, a ski and hiking region in the middle of the highest Alps peaks. From this area, there is a great view on the *Matterhorn* and the *Dufourspitze*, which are among the most famous Swiss peaks. A large amount of pictures is shot in the area representing one of those two peaks. Figure 5 shows the area and the location of the images available.

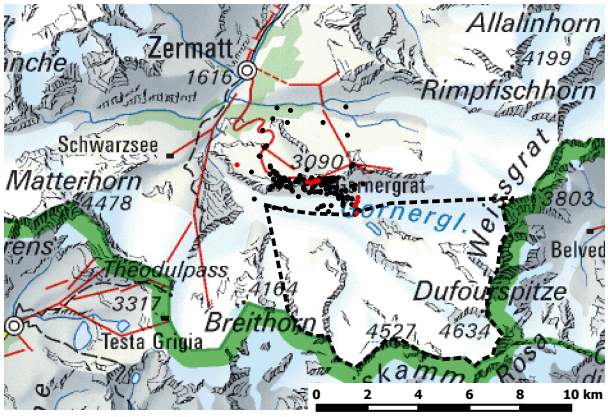
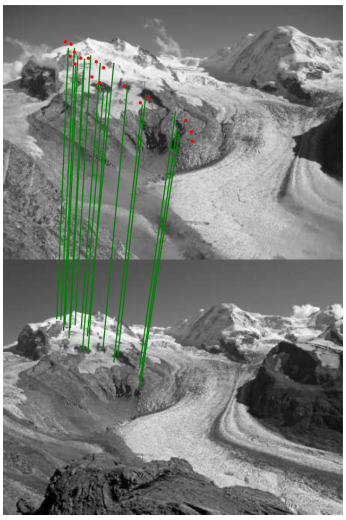


Figure 5: Zermatt area. The red dots represent the GPS images location, while black dots are the images with geotags only. The dashed line encloses the area of interest. Most of the images are shot from the Gornergrat ridge.

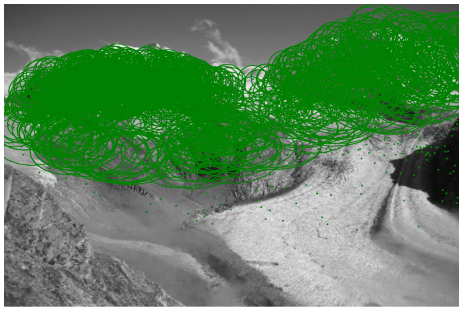
A set of images is extracted from the photosharing platform *Panoramio*. This platform is mainly dedicated to *"Photos of the world"* and thus oriented to landscape scenery. Once they connect to this platform, users can assign a map location to the pictures by the GPS record, a click on a map or by entering an image location. Especially because of this last technique, some geotags are very inaccurate. For instance, in the map of Figure 5, some inaccurate geotags are those located on the very North.

In order to compute the camera pose correctly, the focal length must be estimated. A good estimate can be computed from the focal which is stored in the image metadata and from the camera sensor size, which is found in online camera databases. If the image metadata is incomplete, a normal lens is assumed (focal length equal to the diagonal of the image). In total, our set of images pointing to the area of interest is composed of 198 images, among which 10 have a GPS location and 118 have the focal stored in the metadata.

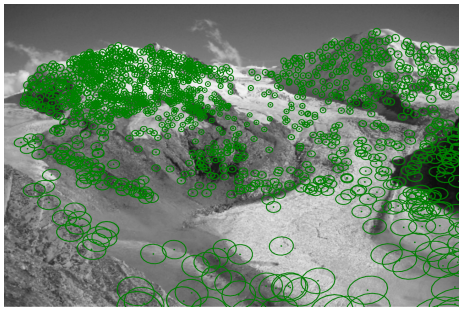
To compute both the horizon and the pixels world coordinates, a DEM is needed. We use the one provided by the Swiss Office of



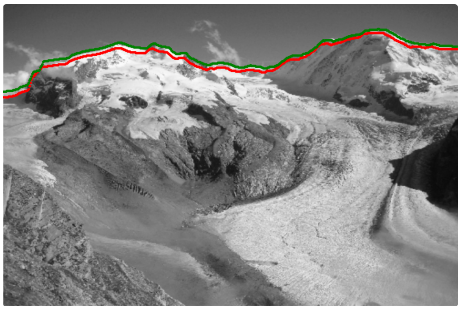
(a) Initial matches found after the NNDR matching and RANSAC orientation.



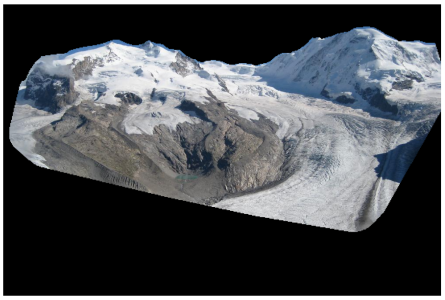
(b) Covariance ellipses at the first iteration.



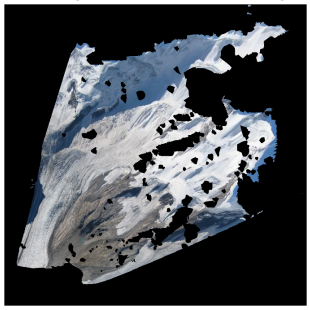
(c) Covariance ellipses after 10 iterations.



(d) Horizon projected in the image before (red) and after (green) DTW matching.



(e) Query picture masked with DEM and the SIFT features (only the area within the convex hull of the matched SIFT is conserved).



(f) Query picture orthorectified.

Figure 4: Representation of the query images pose estimation workflow.

Topography[2], which has a 25m pixel size. Indeed, our focus is on objects far from the camera (beyond 1km), which makes the use of a finer DEM unnecessary. Some height inaccuracies are expected in glacier areas, since glaciers move and melt.

7.2 Reference GPS image orientation

To validate the success of the proposed workflow, we first measure the accuracy of the pixel localization. Indeed, our method minimizes the reprojection error, but because of distortions not taken into account explicitly (focal, principal point location, image distortion) and false positive matches, the estimated pose may differ from the real one, while still providing good pixel localization close to the correspondences. To measure the accuracy of the pixel localization, we applied the following scheme: the images are orthorectified with the DEM (i.e. according to the pose each pixel is projected on a triangulation of the DEM to retrieve its corresponding map coordinates), thus resulting in orthorectified images. Then, distances are measured between similar landmarks found on the rectified images and on an official orthoimage (accuracy $<$ 1m). For each image, between 5 and 15 planar distances are measured and statistics are presented in Table 1. Two orientation methods are compared: horizon matching with DTW and user defined GCP. The GCP digitization accuracy depends on the resolution of the oblique image. To provide a fair comparison, we used images with 500 pixel width for both methods.

First, we comment the user interaction required for both methods. It is much easier for an user to provide the initialization region of the horizon than to detect precise GCP. Indeed, and specifically for this study case, it is not obvious to find similar rocks or cracks in oblique images and orthoimages.

Second, we compare their accuracy: from Table 1, we observe

[2]http://www.swisstopo.admin.ch/

Table 1: Distances measured between recognizable landmarks found both in an orthoimage and in rectified **reference GPS** images. Reference images were oriented either with DTW (left) or GCP (right) and they are presented in Figure 6.

	DTW/GCP		
Picture ID	**Min. [m]**	**Max. [m]**	**Avg. [m]**
24694380	62 / 32	289 / 447	149 / 156
47903616	6 / 6	73 / 76	44 / 40
58907261	10 / 25	76 / 96	44 / 58
65319402	13 / 10	57 / 70	36 / 41

Table 2: Distances measured between recognisable landmarks found on an orthoimage and on **rectified query images**. Those images are also those presented in Figure 6.

Picture ID	**Min. dist. [m]**	**Max. dist. [m]**	**Av. dist. [m]**
14653410	21	343	139
39408227	24	406	148
42359812	28	223	84
54833218	18	239	81
78060652	58	207	98
87063176	14	102	39
96861787	44	538	196

no sensible difference in the accuracy of the methods. On the average, the accuracy of the pixel localization is better than 50m, which is good according to the image resolution and the geometry of the problem (Kraus, 2007). Indeed, and unlike vertical images, the angle between pixel rays and the DEM in oblique images is usually large. In particular, the ray is almost tangent in flat areas and close to the horizon. In these regions of the pictures, small inaccuracy of the pose generates large distortions. The worst accuracy observed, for both DTW and GCP, is measured for the same image (24694380), which may indicate that either the GPS location or the focal estimation is incorrect. Images presented in Table 1 have easily extractable horizon (sky without clouds): the performances tend to decrease when clouds occlude the horizon.

7.3 Query image pose estimation

Once the pose and the 3D coordinates have been estimated for all the GPS images, we can use them as references for the remaining query images. A pose is computed for the query images, for which RANSAC can found initial matches (100 images over the 188 query images). However, some poses are clearly incorrect (10 images): these incorrect poses are associated to images, for which RANSAC returns only false positives. At the end of the pose estimation process, inexact poses are also computed for images with few correspondences irregularly distributed in the image. These poorly distributed correspondences generate large uncertainty in the areas without correspondences. However, by erasing image areas without correspondences, we somehow avoid very large distortions. Finally, the horizon matching step applied to refine the pose is useful for images without clouds, but may suffer of undesirable effects in presence of clouds (see the discussion in the previous section). Statistics for 7 query images are presented in Table 2. As expected the accuracy decrease compared to the reference images. The geometry effects presented above generate even larger distortions in some regions of the image close to silhouette break lines.

To assess the accuracy of the pose itself, we computed as reference the orientation of one image with GCP (the image is presented in Figure 6). Then, we started the pose estimation using a geotag randomly located on circle centered on the location estimated with GCP and with increasing radius. For each distance, 10 pose estimations are conducted. The mean and standard deviation of the absolute difference between the computed parameters and the one obtained with GCP are presented in Table 3.

Table 3: Impact of the geotag on the pose accuracy. For a same image, the initialisation location is generated at different distances. Distances are measured in meters, angles in degrees.

Mean	**Dist.**	**Δ XY (# $<$ 100m)**	**Δ Z**	**Δ head.**	**Δ tilt**	**Δ roll**
	50	85.5 (9)	14.7	0.4	0.5	0.5
	200	72.2 (10)	10.8	0.2	0.2	0.1
	500	85.3 (10)	35.9	0.4	0.7	0.5
	1000	232.7 (3)	95.6	1.7	1.3	1
	1500	227 (7)	132.6	1.7	1.4	0.8
	2000	253.2 (4)	144.4	2.1	1.9	2
	3000	2992.3 (0)	367.5	21.9	7.4	2.4
Std. dev.	**Dist.**	**Δ XY**	**Δ Z**	**Δ head.**	**Δ tilt**	**Δ roll**
	50	15.6	7.2	0.3	0.6	0.8
	200	18.2	1.7	0.2	0.1	0.1
	500	14.9	46	0.3	0.9	1.
	1000	129.2	51.4	1.5	0.7	1.
	1500	275.5	179.2	3.4	1.7	1.
	2000	280.1	189.3	2.8	2.4	0.4
	3000	1877.1	540.9	13.5	7.8	0.4

Between parenthesis, we summed the number of pose locations within a 100m radius of the reference location. It appears, that most of the poses in a 500m radius reach a local minima. We can see it also from the small standard deviations in ΔXY. This minima is not centered on the real location but 100m away, this shift can be explained by the registration errors of the reference images and some false positives detected. Beyond this threshold, for distances from 500m to 2km, the variance increases, and some poses do not converge to the minima. Beyond 2km, computed poses hardly converge.

A video is available on the following link[3]. In this video, the 100 images for which the pose was estimated are rendered on a shaded 3D model of the area, including those with poor matching. Holes in the reconstructed surface correspond to regions of the map which are hidden from the camera position. Quite often, patches of sky are visible at the proximity of the horizon; this illustrates the misregistration problem due to tangential geometry. Attentive viewers will also notice foregrounds projected on the background (a man with a hat, a bird on a fence, a lake). However at the scale represented in this video, the registration is usually of good quality and at least represents a great improvement compared to an unique geotag.

8 DISCUSSION

The proposed workflow is composed of four stages: For the first stage (orientation of the GPS images), we propose a DTW-based horizon matching. The area studied in this application is indeed perfect for horizon matching thanks to the mountains that provide specific silhouettes. However, in presence of other kinds of silhouettes (flatter or with repetitive shapes) or in presence of foreground objects perturbing the horizon (trees, buildings) DTW may fail. In this case, the user can constrain the matching with an azimuth range or provide some GCP. In term of time spent and skills, less than 20 seconds are required to provide a very good watershed initialization, while the digitization of GCP will require a skilled operator, a GIS and for each GCP at least the same amount of time as the one spent for the sky segmentation. Currently, the segmentation is led by the user and automatic sky segmentation (if robust enough) would be a great improvement (see a recent review in (Boroujeni et al., 2012)).

Our workflow is strongly dependent on the *a priori* orientation, which gives a lot of emphasis to SIFT. This step supposes that images, which are similar (season, illumination) to the query images

[3] http://youtu.be/87dHVDdlPSs

are present in the reference database. At this stage, more than 50% of the images were correctly georeferenced. We could also increase this percentage by running a second pass of the process (using not only GPS images as reference, but also the newly georeferenced ones). To overcome the use of SIFT, rendered views could also be used as reference images, but the challenge then becomes to find a set of descriptors able to match synthetic and real images (see (Produit et al., 2012)).
The location part of the *a priori* orientation is based on the geotag provided by the user. Some web applications store the zoom level applied on the map by the user when clicking as a measure of the localization accuracy. However, the relation between the zoom and the accuracy is not straightforward and here we assumed for every geotag a standard deviation of 1000m. (Zielstra and Hochmair, 2013) studied image geotag accuracies for several type of landscapes and several areas of the world. We can learn from this paper that in natural landscape users provide location much better than that.
In general, the accuracy around detected correspondences is coherent to an *a priori* expected accuracy (<50m). For this case study, we used low resolution images, and improvement may still be observable by considering full resolution images. Accuracy tends to decrease in image regions where no or false correspondences are found (for instance close to the image borders) and in sinuous parts of the DEM, which generate large distortions during the orthorectification of the query images. To digitize the GCP used to compute the statistics, we choose the best quality images, which often correspond to those more reliably oriented. We can then imagine that the accuracy within the whole set could slightly decrease. Considering our setup, which involves very differing images shot with uncalibrated camera and rough pose, the results meet the expected accuracy.

One of the motivations for computing the pose of landscape images is related to environmental monitoring. To be effective, such monitoring requires a very high accuracy of the measurements of natural objects position and movements. According to statistics presented in Tables 1 and 2, the accuracy is not sufficient for environmental studies in most part of the images. However, by providing our estimation as an initial pose, the task of manually georeferencing images with GCP becomes strongly facilitated. Moreover, once some GCP are provided by the user, we can take advantage of the Kalman filter and get more restrictive error ellipses, in order to propose more correspondences and a final better pose.
Database of landscape images are currently list of images, sometimes associated with geotags, sometimes linked to a 2D map. The orientation computed with ourworkflow could be useful in several way to create more user-friendly image database browsers. First, the location and heading measured are accurate enough to detect visible points of interest and link the images with appropriate tags. Then, computed poses can also be used to overlay toponyms and other geographic layers in the images (as the mountain names). Finally, the visual matching of the image and the landscape model is quite good and thus the images could be inserted in a virtual globe.

9 CONCLUSION

In this paper, we presented a workflow to estimate the pose for a set of landscape images downloaded from a photosharing platform. Such a workflow is necessary to answer the problem of finding the pose of the images and extracting geoinformation from non-photogrammetric sets of images whose geolocation is sometimes very approximative and for which only sparse images, covering the area with low density, are generally available. This setup is not limited to web-shared oblique landscape images, since it also corresponds, for example, to historic image databases.

It has been shown by several other authors (Baboud et al., 2011, Chippendale et al., 2008) that the orientation of landscape images can be computed from a landscape model if the location of the camera is provided. Their approaches are indeed automatic, but the processing involved is very demanding, reason for which we propose an alternative method using horizon line matching. The proposed method is less computationally expensive, but involves the user in its initialization.

In our proposition, reference images are used to recover automatically the full pose (orientation and location) of the other images belonging to the same collection, but without precise location. To reach this goal, we propose an original workflow based on a Kalman filter and use the landscape model to add more robustness to the SIFT matching. To the best of our knowledge, our method is the first to recover orientation and location of tourist images collections in rural area. Since the user is involved only at the beginning of the process, i.e. during the orientation of the GPS located images, it remains reasonably close to an automatic routine.

The achieved accuracy is not comparable to the one of orthoimages generated via a classic acquisition and processing of photogrammetric images, which remains a limitation for the usage of our workflow at its current state for environmental studies. To improve the accuracy further, an increased involvement of the user would be necessary. Nonetheless, the pose is correct enough to open interesting opportunities for images database management, for example for advanced querying or augmented reality purposes.
Our workflow can only be applied in area of interest where quite large collections of landscape pictures are available. The application of such a workflow in other areas of the world where only a landscape model is available as reference remain an open challenge, which is to date only partially addressed (Baatz et al., 2012a). Nevertheless, with the popularization of GPS enabled camera connected to the internet, we can reasonably think that the image databases will continue their growth.

Further developments will make the estimation of the orientation of the GPS images automatic and the initial estimate of the orientation for the query images independent of SIFT. Moreover, the process is also designed to be extendible to images without geotags, if a rough assumption of their location is provided. For instance, we showed in (Produit et al., 2014) that landscape models can be used to discard unlikely shooting locations and detect preferred ones.

ACKNOWLEDGEMENTS

We would like to thank the reviewers who helped us to step back and improve this paper. This project has been partially funded by the Swiss National Science Foundation (grant PZ00P2-136827).

REFERENCES

Agarwal, S., Furukawa, Y., Snavely, N., Simon, I., Curless, B., Seitz, S. M. and Szeliski, R., 2011. Building Rome in a day. Communications of the ACM.

Baatz, G., Saurer, O., Köser, K. and Pollefeys, M., 2012a. Large scale visual geo-localization of images in mountainous terrain. In: ECCV.

Baatz, G., Saurer, O., Koser, K. and Pollefeys, M., 2012b. Leveraging topographic maps for image to terrain alignment. In: 3DIM/3DPVT.

Baboud, L., Cadík, M., Eisemann, E. and Seidel, H.-P., 2011. Automatic photo-to-terrain alignment for the annotation of mountain pictures. In: CVPR.

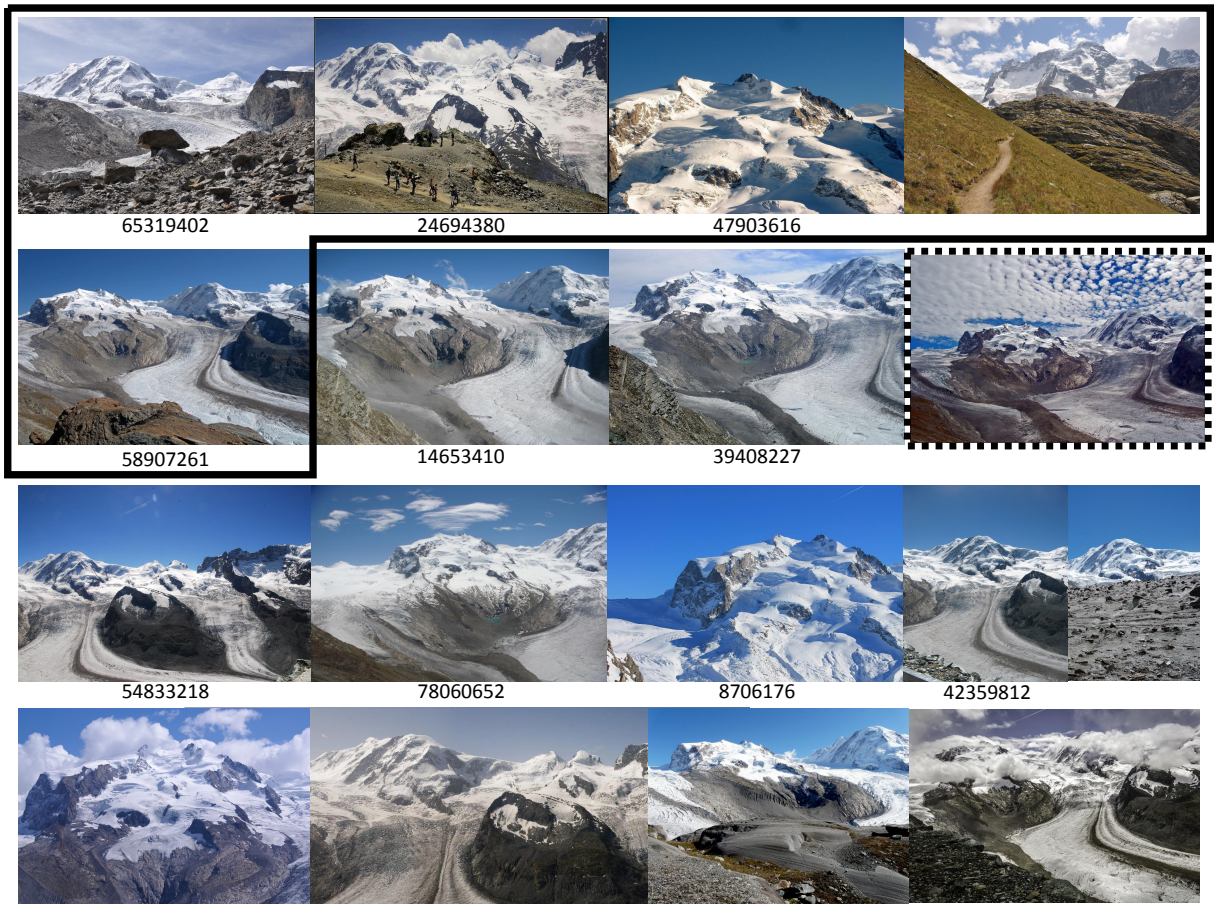


Figure 6: Images used to compute statistics in Tables 1 and 2. GPS images are within the box, numbers correspond to the ones within the tables. Thei Image within the dashed box, is used for the statistic in Table 3 Unlabelled images are randomly selected within the dataset.

Berndt, D. J. and Clifford, J., 1994. Using dynamic time warping to find patterns in time series. In: KDD workshop.

Boroujeni, N. S., Etemad, S. A. and Whitehead, A., 2012. Robust horizon detection using segmentation for UAV applications. In: CRV.

Bozzini, C., Conedera, M. and Krebs, P., 2011. A new tool for obtaining cartographic georeferenced data from single oblique photos. CIPA Symposium.

Chippendale, P., Zanin, M. and Andreatta, C., 2008. Spatial and temporal attractiveness analysis through geo-referenced photo alignment. In: IGARSS.

Corripio, J., 2004. Snow surface albedo estimation using terrestrial photography. International Journal of Remote Sensing.

Crandall, D. J., Backstrom, L., Huttenlocher, D. and Kleinberg, J., 2009. Mapping the world's photos. In: WWW.

Debussche, M., Lepart, J. and Dervieux, A., 1999. Mediterranean landscape changes: evidence from old postcards. Global Ecology and Biogeography.

Friedland, G., Choi, J., Lei, H. and Janin, A., 2011. Multimodal location estimation on flickr videos. In: IWSM.

Hammoud, R., Kuzdeba, S., Berard, B., Tom, V., Ivey, R., Bostwick, R., HandUber, J., Vinciguerra, L., Shnidman, N. and Smiley, B., 2013. Overhead-based image and video geo-localization framework. In: CVPR Workshop.

Hays, J. and Efros, A. A., 2008. Im2gps: estimating geographic information from a single image. In: CVPR.

Jacobs, N., Satkin, S., Roman, N., Speyer, R. and Pless, R., 2007. Geolocating static cameras. In: ICCV.

Kraus, K., 2007. Photogrammetry: geometry from images and laser scans. Walter de Gruyter.

Kull, C., 2005. Historical landscape repeat photography as a tool for land use change research. Norwegian Journal of Geography.

Li, Y., Snavely, N., Huttenlocher, D. and Fua, P., 2012. Worldwide pose estimation using 3d point clouds. In: ECCV.

Moreno-Noguer, F., Lepetit, V. and Fua, P., 2008. Pose priors for simultaneously solving alignment and correspondence. In: ECCV.

Produit, T., Tuia, D., De Morsier, F. and Golay, F., 2014. Do geographic features impact pictures location shared on the web? modeling photographic suitabiliy in the swiss alps. In: ICMR.

Produit, T., Tuia, D., Golay, F. and Strecha, C., 2012. Pose estimation of landscape images using DEM and orthophotos. In: ICCVRS.

Produit, T., Tuia, D., Strecha, C. and Golay, F., 2013. An open tool to register landscape oblique images and generate their synthetic model. In: OGRS.

Roush, W., Munroe, J. and Fagre, D., 2007. Development of a spatial analysis method using ground-based repeat photography to detect changes in the alpine treeline ecotone. Arctic, Antarctic, and Alpine Research.

Snavely, N., Seitz, S. and Szeliski, R., 2006. Photo tourism: exploring photo collections in 3D. ACM Transactions on Graphics.

Strecha, C., Pylvanainen, T. and Fua, P., 2010. Dynamic and scalable large scale image reconstruction. In: CVPR.

Zielstra, D. and Hochmair, H. H., 2013. Positional accuracy analysis of flickr and panoramio images for selected world regions. Journal of Spatial Science.