Re-photography and Environment Monitoring Using a Social Sensor Network

Paul Chippendale, Michele Zanin, and Claudio Andreatta

Fondazione Bruno Kessler, TeV Group,
Via Sommarive, 18, Povo, Trento, 38050 Italy
(chippendale,mizanin,andreatta)@fbk.eu

Abstract. This paper presents a technology capable of enabling the creation of a diffuse, calibrated vision-sensor network from the wealth of socially generated geo-referenced imagery, freely available on the Internet. Through the implementation of an accurate image registration system, based on image processing, terrain modelling and subsequent correlation, we will demonstrate how images taken by the public can potentially be used as a mean to gather environmental information from a unique, ground-level viewpoint normally denied non-terrestrial sensors (consider vertical or overhanging cliffs). Moreover, we will also show this registration technology can be used to synthesize new views using sections of photos taken from a variety of places and times.

Keywords: environmental monitoring, distributed sensors, geo-referencing, re-photography.

1 Introduction

Monitoring the environment, whether it be through vision, radar or a multitude of other technologies, has increased dramatically since the birth of satellite technology. The ability to view large swathes of the Earth in single orbits makes them ideal candidates for monitoring climactic changes in near real-time. The sensing technologies available to climatologists today are many; however the visual spectrum still has an important role to play in environmental observations. One good example is the monitoring of subtle vegetation colour changes over time which can signify variations in the onset of Autumn [1]. Furthermore, due to cloud cover and atmospheric attenuation, the planet's surface can rarely be visually observed with complete clarity from space.

The near ubiquitous ownership of digital cameras, their inclusion in virtually every mobile phone on the market combined with the speed at which an image can be taken and shared with the World (via 3G or WIFI and a multitude of photo websites) together promise a new paradigm in environmental observation. As the integration of GPS receivers into phones (and in the near future cameras) is becoming ever more common, the number of geo-referenced images taken and shared via Internet websites such as www.Flickr.com and www.Panoramio.com, often within minutes of capture, is increasing at a phenomenal rate. For example, the Flickr API tells us that over

P. Foggia, C. Sansone, and M. Vento (Eds.): ICIAP 2009, LNCS 5716, pp. 34–42, 2009.

800,000 geo-referenced photos were taken in the Alps in 2008 (almost 27 millions worldwide). There is a clear trend: geo-tagging photos for the sole purpose of placing one's photos 'on the map', either automatically (via GPS) or manually (via GUIs in GoogleEarth), is becoming increasingly popular.

Geo-tagged photos alone are not a reliable resource to extract spatial information as their orientation and content is unknown. We therefore require the implementation of a system that can take such images and understand precisely their orientation at the time of capture and camera parameters so that registered orthophoto-like images can be generated.

However, image registration in an outdoor environment cannot exploit the established methods developed for indoor use, e.g. magnetic tracking, fiducial markers. Outdoor registration systems traditionally rely on GPS for position measurements combined with magnetic compasses and inertial sensors for orientation as used by Azuma et al. Examples of systems using such sensors include Columbia's Touring machine [2] and MARS [3], the Battlefield Augmented Reality System [4], work by Thomas [5], the Tinmith System [6] and to some extent Sentieri Vivi [7]. Although magnetic compasses, inertial sensors and GPS can be used to obtain a rough estimate of position and orientation, the precision of this registration method (using affordable devices) is insufficient to satisfy many Augmented Reality (AR) overlay applications. Unfortunately, these types of sensors are not yet to be found inside consumer electronics. Undoubtedly, one day they will become commonplace; until that day however, computer vision methods will have to suffice to estimate device orientation through the correlation of visual features with calibrated real-world objects.

In Microsoft's Photosynth [8], highly recognizable man-made features are detected, such as the architecture of Notre Dame de Paris, to align and assemble huge collections of photos. Using this tool, the relative orientation and position of a photo with respect to a calibration object can be calculated due to the scene's unique and unchanging nature. The University of Washington and Microsoft [9] took this idea one step further and created the tool 'PhotoTour' that permits the viewer to travel virtually from one photo into another. Viewfinder by the University of Southern California [10] and [11] are other research projects that aid users to spatially situate their photographs through the creation of alignment tools that interfaces with 3D models visualized inside GoogleEarth. Although their systems are essentially manual, a lot of the hard work of alignment is taken out of the registration process and the University claims that a 10-year-old should be able to find the pose of a photo in less than a minute.

Our approach to the problem of image registration tackles the challenge of the natural environment. We attempt to identify and align evident geographical photo features with similar ones generated in a synthetic terrain model. Extracting feature points from outdoor environments is however challenging, as disturbances such as clouds on mountain tops, foreground objects, large variations in lighting and bad viewing conditions in general such as haze, all inhibit accurate recognition. As a result, great care needs to be taken to overcome such inherent limitations and we attempt to compensate for this by combining different yet complementary methods.

Behringer's approach [12] is similar to ours, based on an edge detector to align the horizon silhouette extracted from Digital Terrain Model (DTM) data with video images

to estimate the orientation of a mobile AR system. They demonstrated that a well-structured terrain could provide information to help other sensors to accurately estimate registration; their solution had problems with lighting and visibility conditions.

Our approach however incorporates an enriched set of feature points and a more accurately rendered terrain model enhanced by additional digital content such as: lake contours, GPS tracks, land-usage GIS layers, etc.

2 Photo Registration

We register photos by correlating their content against a rendered 3D spherical panorama generated about the photo's 'geo-location'. In essence we generate a complete 360° synthetic image of what an observer would see all around them at a given location. To do this we systematically load, scale and re-project DTM data (such as The Shuttle Radar Topography Mission data [13] freely available from NASA) onto the inside of a sphere using ray-tracing techniques. We render up to a distance dictated by the height of the observer above sea-level and the maximum theoretical visible distance due to Earth curvature and possible mountain heights (the maximum rendering distance for a ground-level observer is usually in the 200-550km range). In preparation for photo alignment the sphere is 'unwrapped' into a 360° by 180° rectangular window (a section of which is illustrated in Fig. 1); each synthetic pixel has its own latitude, longitude, altitude and depth.

Fig. 1. 70° wide by 40° tall section of unwrapped synthetic panorama, with intensity proportional to distance

Placing a photograph into this unwrapped space requires a deformation so that photo pixels are correctly mapped onto synthetic ones depending upon estimated camera parameters, such as pan, tilt, lens distortion, etc. The scaling parameter is extracted from the focal length meta-data contained within the EXIF JPEG data of the photo where available; otherwise it is estimated using an iterative strategy from a default value.

Our approach is structured into four phases: 1) extract salient points/profiles from a synthetic rendering, 2) extract salient points/profiles from the photo, 3) search for correspondences aiming to select significant synthetic/photo matches, 4) apply an optimization technique to derive the alignment parameters that minimize re-projection error.

The extraction of profiles from the synthetic panorama is relatively straightforward and is simply a matter of detecting abrupt depth discontinuities between adjacent pixels. In photos we have no prior knowledge of depth therefore we employ various

image processing algorithms which analyze colour, texture and edge content to try to locate region interfaces and thus suspected depth changes in the real world. To place a different emphasis on the various types of profiles in the photo we attempt to locate sky regions (low detail zones with a hue in the blue range located towards the top of the image) and also estimate the presence of foreground objects (sharp and saturated regions starting from the bottom). The contribution of each type of profile is weighted differently; e.g. a land-sky interface is more reliable for correlation than a profile close to the observer.

The third step is hypothesis generation. From the synthetic panorama we generate a grid of match hypotheses by modifying the various camera parameters: the extrinsics such as pan, tilt and roll, and the intrinsics such as focal length and lens distortion. The initial hypotheses are generated with the intrinsic parameters set to a default value typical for that make of camera. The range over which yaw, pitch and roll vary depends upon the findings of a frequency analysis stage which evaluates the photo and synthetics sky/land interfaces. The most prominent/evident geographical peak or valley is used as a starting point for the search for a match.

As can be seen in Fig. 2a, the analysis algorithm follows the synthetic profile (red line) with a sliding window. The relevant features are locally computed inside the window in a multi-resolution manner in order to measure if there is a good correspondence between the synthetic profile and the real one. To do so, we examine the window to see if we have strong profiles in the photo and, moreover, if the profiles are of roughly the same angle. The alignment is measured considering the angle between the normal to the synthetic profile (green arrow) and the normal to the real one (black arrow).

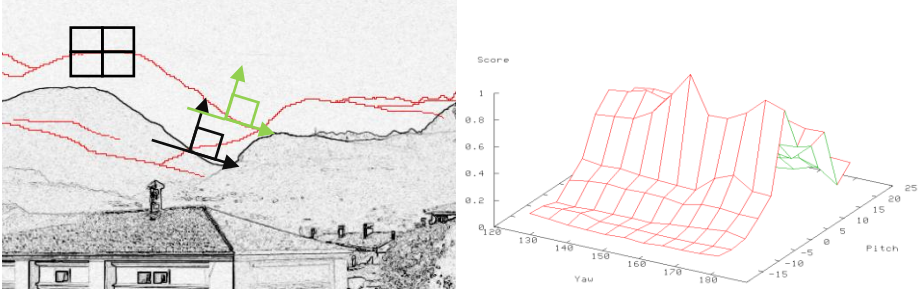

Fig. 2. (a) Correlation hypothesis test (synthetic profile in red) (b) Result of correlation search in the yaw - pitch parameters space

The results obtained from the series of 100 correlation tests can be seen in Fig. 2b. There is a distinctive peak at 150° pan and 10° tilt. The quality of this tentative alignment can be assessed visually in Fig. 3.

This automated brute force approach provides us with an initial 'best guess' alignment. Registration is refined in a following stage where the other parameters such as roll and lens distortion are adjusted using a minimization algorithm. All of the image analysis techniques involved are relatively fast and efficient; a necessity due to the sheer number of hypothesis to analyze.

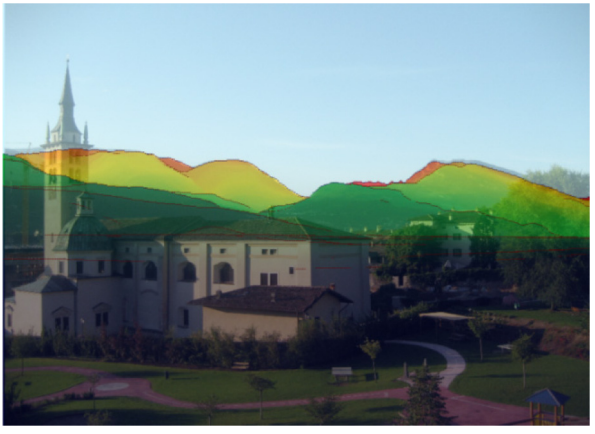

Fig. 3. Result obtained from auto-alignment, with colour proportional to distance

3 Applications and Results

3.1 Creation of a Calibrated Social Sensor Network

Once a photo has been registered, a calibrated sensor network can be created by amassing a large number of similarly aligned images and then systematically draping their photo pixels onto a surface model generated from the DTM. In order to provide adequate spatio-temporal coverage of an area, photos need to be collected from different positions and orientations at various times of the year.

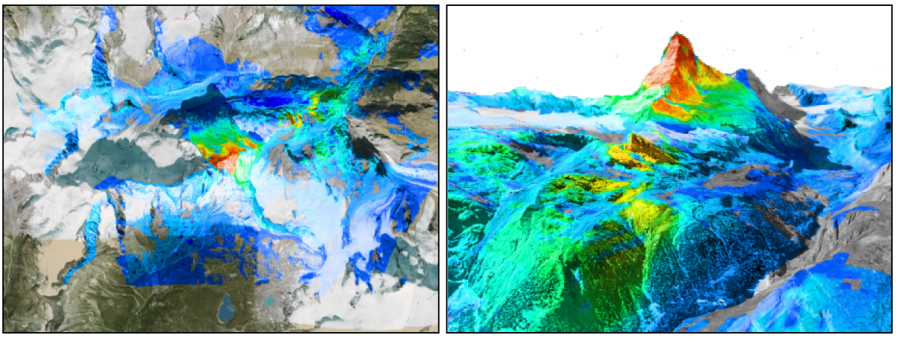

Fig. 4. Heat maps of 100 photo viewsheds combined, taken around the Matterhorn. The spatial resolution of each photo pixel varies in relation to camera-terrain distance, which is not expressed in this composite image.

To demonstrate the coverage obtainable, we downloaded and registered 100 georeferenced images taken in the area of the Matterhorn from Flickr. As can be seen in Fig. 4, (where red represents the terrain that features most often and blue the least),

the coverage is quite extensive. In a previous paper [14], we went a step further with this Matterhorn data and explained how we could combine a viewshed with the social comments made about the photos themselves (number of views, comments, favourites, etc.) in order to estimate the 'attractiveness' of the terrain in question for photographers.

3.2 Generation, Analysis and Visualization of Environmental Content

Once a photo has been accurately registered, each photo pixel is assigned latitude, longitude, altitude and a distance from the camera. This geo-spatial information can be exploited in a variety of ways; for example we can generate new unique views from multiple images taken from a variety of places or we can extract environmental data such as the snowline.

To demonstrate how we can detect the snowline in aligned photos we selected an area of the Brenta Dolomites around Cima di Ghez, bound by the rectangle: 46.077032 < latitude < 46.159568 and 10.84653 < longitude < 10.92353. In this region the terrain altitude goes from 850m to about 3200m. Then we automatically generate a subset from our database of photos that in some part overlaps this area.

We then transform the photos into a top-down view-shed representation through the systematic projection of each terrain pixel from the 2D image plane (when registered these pixels also carry 3D information) into this new viewpoint. The colour of each mapped pixel is inserted into the appropriate position in a 0.2 seconds (about 7 m) resolution grid covering the area described.

In order to understand which pixels/regions from the selected images contain snow, we selected a reference image that adequately covered the region in question and was taken on a clear day in the summer (inferred from the photo's timestamp).

Next, from the DTM data, we generated a 3D representation of the landscape contained within the reference image and draped the photos for comparison onto this model. The results of this can be seen in Fig. 5.

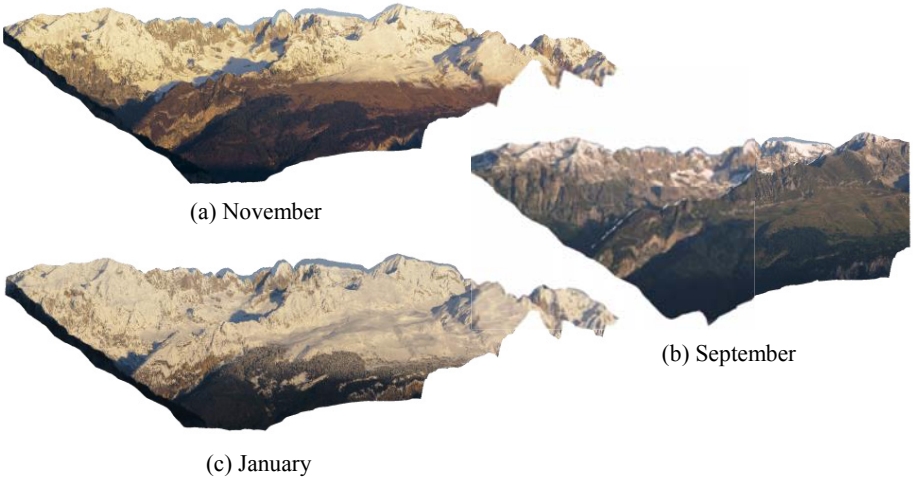

Fig. 5. Re-projection of aligned photos into the same viewpoint as the reference image

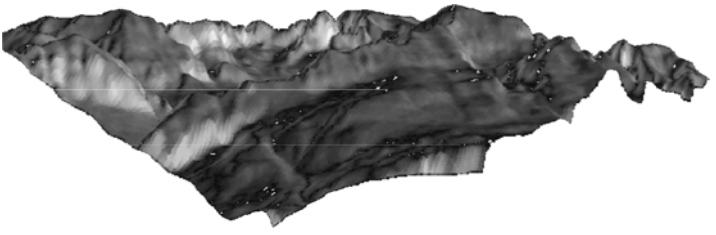

Fig. 6. Local terrain gradient map of the test region, viewed from reference image perspective

Our set of re-projected test photos is then subjected to a sigma filter to reduce noise. This averages the value of each pixel with its similar neighbours by analyzing the standard deviation σ. Thus it only averages those neighbouring pixels whose values lie within 2 σ of the central pixel value; where σ is calculated once for the entire image. This processing smoothes the images in an edge preserving manner.

Next we compare each image with the reference image; aiming to detect local differences using the L2 measure in RGB space with an adaptive local threshold (the threshold is computed in a sliding window considering the mean and variation of the colours among all the test images, thus reducing the impact of changes due to illumination variation).

Once the regions that differ from the reference image have been detected, we try to understand if this is due to the presence of snow. Snow is assumed to have a low saturation and high value in HSV space. As snow is unlikely to be seen on very steep mountain faces, we also take local gradient into account. Fig. 6 graphically illustrates the terrain gradients inside the test region, as seen from the reference perspective (black pixels are horizontal and white vertical).

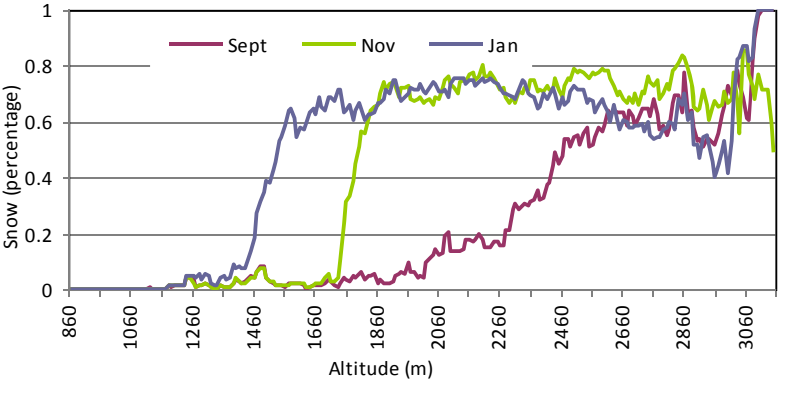

Fig. 7. Snowline detection using a summer image as a reference. The data is noisier at high altitudes: this is mainly caused by the fact that we have fewer pixels representing the highest part of the altitude range.

The saturation, luminance and terrain steepness distribution has been studied on a training set in order to model snow appearance using a multi-dimensional Gaussian distribution.

Fig. 7 shows the percentage of the pixels in the target images that differ from the reference and that have been classified as snow at a given altitude. The evident discontinuity from left to right represents the start of the snowline.

3.3 Re-photography

To further illustrate the re-photography potential of the system we generated two new 'virtual photos' based solely on the content of two images taken over 10km away. We chose an arbitrary point in space (lat. 46.096980, lon. 10.913257, alt. 1526m) and draped the view-sheds of the November and January images into GoogleEarth using the integrated photo overlay function. In this way we could precisely reproduce the same view point compared to GoogleEarth's. The two images, in the centre and on the right of Fig. 8, show the re-photography results.

Fig. 8. (a) GoogleEarth (b) Image November re-projected, (c) Image January re-projected

As Fig. 8 shows, we are able to generate totally new views in GoogleEarth, representing different seasons and weather conditions through the gathering, registration and re-projection of public images shared openly on the Internet.

4 Conclusions

In this paper we have presented some of our current research in the field of photo registration, re-photography and environmental content generation (more details and constantly updated results are available at http://tev.fbk.eu/marmota). We have suggested how a widespread, calibrated network of ground-level vision sensors can be realized through the correlation of geo-referenced photos to the terrain, and have shown specific examples of how this 'composite-sensor' can be exploited. Initial observations suggest that areas close to attractions, honeypot villages, roads or popular footpaths are more heavily covered (down to a few centimetres per pixel) whilst more remote peaks are only covered at lower resolutions.

In the near future we plan to automatically monitor a wide variety of photo sharing websites on a daily basis for a specific area such as Trentino/Alto Adige and observe how extensive the photo coverage actually is.

We will then fuse our alignment results with existing orthophotos in order to investigate how we could improve their resolution and also fill in the many low resolution spots such as vertical cliffs. The user could also visualize such terrain and its changes over time. Registration of historical images taken at known locations will also enrich this temporal record of the environment.

To summarize, we have shown how a widespread vision sensor network can be created in an ad-hoc manner, which will naturally expand in time, improving in resolution and quality (through the availability of new devices) and which is, perhaps best of all, maintenance free and gratis.

References

1. Astola, H., Molinier, M., Mikkola, T., Kubin, E.: Web cameras in automatic autumn colour monitoring. In: IGARSS 2008 (2008)
2. Feiner, S., MacIntyre, B., Hollerer, T., Webster, A.: A touring machine: Prototyping 3D mobile augmented reality systems for exploring the urban environment. In: Proc. ISWC 1997, Cambridge, MA, USA, pp. 74–81 (1997)
3. Hollerer, T., Feiner, S., et al.: Exploring MARS: developing indoor and outdoor user interfaces to a mobile augmented reality system. Computer & Graphics 23, 779–785 (1999)
4. Baillot, Y., Brown, D., Julier, S.: Authoring of physical models using mobile computers. In: Proc. ISWC 2001, pp. 39–46 (2001)
5. Thomas, B., Demczuk, V., Piekarski, et al.: A wearable computer system with augmented reality to support terrestrial navigation. In: ISWC 1998, Pittsburgh, PA, USA, pp. 168–171 (1998)
6. Piekarski, W., Thomas, B.: Tinmith-metro: New outdoor techniques for creating city models with an augmented reality wearable computer. In: IEEE Proc. ISWC 2001, Zurich, Switzerland, pp. 31–38 (2001)
7. Sentieri Vivi (2008), http://www.sentierivivi.com
8. Snavely, N., Seitz, S., Szeliski, R.: Photo tourism: Exploring photo collections in 3D. In: Photo tourism: Exploring photo collections in 3D, SIGGRAPH Proceedings 2006, pp. 835–846 (2006)
9. Snavely, N., Garg, R., Seitz, S., Szeliski, R.: Finding Paths through the World's Photos. In: SIGGRAPH Proceedings 2008 (2008)
10. University of Southern California, Viewfinder - How to seamlessly Flickerize Google Earth (2008), http://interactive.usc.edu/viewfinder/approach.html
11. Chen, B., Ramos, G., Ofek, E., Cohen, M., Drucker, S., Nister, D.: Interactive Techniques for Registering Images to Digital Terrain and Building Models (2008), http://research.microsoft.com/pubs/70622/tr-2008-115.pdf
12. Behringer, R.: Registration for outdoor augmented reality applications using computer vision techniques and hybrid sensors. In: IEEE VR 1999, Houston, Texas, USA (1999)
13. The Shuttle Radar Topography Mission (SRTM) (2000), http://www2.jpl.nasa.gov/srtm/
14. Chippendale, P., Zanin, M., Andreatta, C.: Spatial and Temporal Attractiveness Analysis through Geo-Referenced Photo Alignment. In: International Geoscience and Remote Sensing Symposium (IEEE), Boston, Massachusetts, USA (2008)