Geo-positional Image Forensics through Scene-terrain Registration

P. Chippendale[1], M. Zanin[1] and M. Dalla Mura[2]

[1]*Technologies of Vision Research Unit, Fondazione Bruno Kessler, Trento, Italy*
[2]*GIPSA-Lab, Signal and Image Department, Grenoble Institute of Technology, Saint Martin d'Hères, France*

Keywords: Media Forensics, Visual Authentication, 3D Terrain-model Registration, Augmented Reality, Machine Vision, Geo-informatics.

Abstract: In this paper, we explore the topic of geo-tagged photo authentication and introduce a novel forensic tool created to semi-automate the process. We will demonstrate how a photo's location and time can be corroborated through the correlation of geo-modellable features to embedded visual content. Unlike previous approaches, a machine-vision processing engine iteratively guides users through the photo registration process, building upon available meta-data evidence. By integrating state-of-the-art visual-feature to 3D-model correlation algorithms, camera intrinsic and extrinsic calibration parameters can also be derived in an automatic or semi-supervised interactive manner. Experimental results, considering forensic scenarios, demonstrate the validity of the system introduced.

1 INTRODUCTION

Digital photographs and videos have proven to be crucial sources of evidence in forensic science; they can capture a snapshot of a scene, or its evolution through time (Casey, 2004); (Boehme et al., 2009). Geo-tagging (Luo et al., 2011), i.e. the collocation of geo-spatial information to media objects, is a relative newcomer to the field of data annotation, but is growing rapidly. Concurrently, the availability of easy-to-use image processing tools and meta-data editors is leading to a diffusion of fake geo-tagged content throughout the digital world. As geo-tagged media can be used to corroborate a person's or an object's presence at a given location at a given time, it can be highly persuasive in nature. Therefore, it is essential that the content be authenticated and the associated geo meta-data be proved trustworthy.

The addition of location information has been fuelled in recent years thanks to the embedding of geo-deriving hardware, such as Global Positioning System (GPS), in many consumer-level imaging devices. Nowadays, the most common way in which photographs are geo-tagged is through the automatic insertion of spatial coordinates into the EXIF meta-data fields of JPEG images; however, a reported location can easily be tampered with, and varies in precision according to its means of derivation. For example, in urban or forested environments, GPS signals suffer from attenuation and reflection which leads to inexact, or the lack of, triangulation of position as was illustrated by (Paek et al., 2010), commonly referred to as the 'Urban Canyon' problem in dense cities (Cui and Ge, 2003).

Standard geo-tagged photos contain three non-independent pieces of information that provide valuable location indicative clues:

- Time when the media object was captured;
- Positional information (some devices also provide orientation data);
- Embedded visual content of the scene.

Although these three indicators are derived from independent sources and sensors, they are closely intertwined since they all spatiotemporally describe a particular scene. These interdependences can be exploited to derive or validate one piece of information against the others. (Hays and Efros, 2008) showed that in a natural scene observed from an arbitrary position, the geometry of solar shadows cast by objects can provide clues about the time and orientation of the camera. Analogously, (Chippendale et al., 2009) illustrated how a captured location can also be confirmed or hypothesized by comparing the image content of the real scene with expected geo-content, through synthetically generated terrain correlation.

Three elements must be examined in order to prove, beyond a reasonable doubt, that a geo-tagged

photograph is genuine:

1) Photo is unmodified at a pixel level; 2) The visual content of the photo is consistent with that suggested by its location; 3) The observation of time-dependent events embedded in the photo is consistent with the location and time suggested.
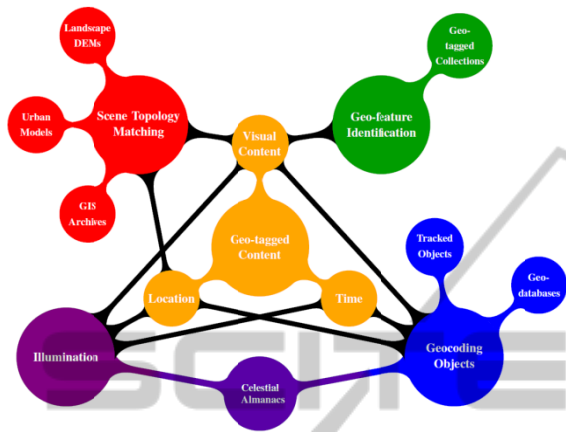

Figure 1: Mind map of geo-positional image forensics interconnections.

The interrelationships between visual content, location and acquisition time have been generalized in the mind map seen in Figure 1. The yellow core of the diagram highlights the three key elements that describe a geo-tagged media object. The red region encompasses scene-topology matching; green relates to geo-feature identification tools that aim to match portions of photos to similar ones in pre-tagged images; purple shows illumination relationships, e.g. shadows or reflections; and the blue region relates to geocoding based on the matching of machine identifiable content to geo-databases, e.g. scene text or logos.

In this paper, we will concentrate on the second of the three elements, relating to the authentication of spatiotemporal meta-data associated to a photograph, through visual geo-content consistency checks.

1.1 Related Work

Prior operational approaches in the geo-positional image forensics field have been very laborious and expert-dependent, using ad-hoc methods tailored towards a specific case. Tools like World Construction Set[1] or Visual Nature Studio[2] can be used to provide visual evidence by virtually recreating a scene, or, visual correlations can be made to nearby geo-tagged photos (taken from massive socially generated geo-photo databases like

Panoramio[3]).

The automatic matching of photo content to 3D synthetic models has been well explored in literature in recent years, spanning many fields of research including robotic navigation (Leotta and Mundy, 2011), Augmented Reality, photogrammetry (Guidi et al., 2003), computer vision and remote sensing. In the outdoor environment, (Baatz et al., 2012), (Baboud et al., 2011) and (Hays and Efros, 2008) have recently explored terrain-feature identification to align visual terrain profiles to geo-specific features. This approach involves the matching of machine-extractable features from a photo to similar features in a synthetic 3D-model, essentially pairing 2D pixel-region features to 3D locations. The synthetic models used, rendered from a wide variety of 3D geo-located data (e.g. Digital Elevation Models (DEMs)), yield silhouette profiles which should correspond to the terrain visible at a stated photo location.

1.2 Contribution

In this paper, we illustrate how we have integrated state-of-the-art visual-feature to 3D-model correlation algorithms to create an interactive pixel-to-world mapping tool (seen in Figure 2) which can be used to validate the authenticity of terrain-evident, geo-tagged photographs.
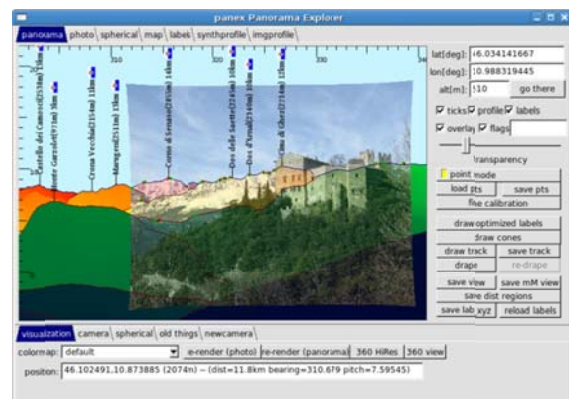

Figure 2: Screenshot of the photo registration tool.

2 SYSTEM OVERVIEW

In forensic investigations, it is essential that human operators are placed inside the visual content analysis loop. Often, occlusions, difficult lighting conditions, and tampering impede automatic terrain feature detection, hence manual guidance can be utilised to reduce detection ambiguities. Results

from the computer-vision powered registration process are ranked and offered to the user in the form of an Augmented Reality visualization, together with a correlation metric to indicate match 'goodness'. In situations where ambiguities arise, the user can influence subsequent refinement stages by assigning new correlation pairs using a point and click interface.

The automatic geo-registration process is based on the algorithm proposed by (Baboud et al., 2011) and (Kostelec and Rockmore, 2008), which attempts to match visibly-evident geographical features to corresponding synthetic features in a virtual model. When dealing with natural outdoor scenes, one of the most relevant features is terrain morphology. In our approach, the virtual terrain models are generated from DEMs, which represent the Earth's surface as a regular grid of elevation values (Li et al., 2005). These renderings are generated by systematically projecting rays from a photo's location onto the inside surface of a sphere recording their angular intersections with the DEM. This spherical image is then projected onto an equiangular planar grid, in which each point in the plane corresponds to a certain latitude, longitude, altitude and depth (i.e. distance from the photo's location).

To derive camera pose, i.e. the pan, tilt and roll of the camera that took the original photograph, the image is firstly scaled according to the field-of-view reported (usually derived from the focal length meta-data in the EXIF). If focal-length information is absent (or incorrect), then the system can estimate it through a user-assisted scaling strategy forming part of the alignment stage.

Photographs are aligned to the equiangular model in four phases:

1. extraction of salient depth profiles from the synthetic image;

2. extraction of salient profiles from the photo;

3. a registration algorithm searches for correspondences between the two profile sets;

4. apply an optimization algorithm to derive the best alignment parameters that minimize re-projection error.

The processing pipeline of our system is visualized below in Figure 3.

We generate salient depth profiles from the synthetic image by marking abrupt distance changes between adjacent points lying on a virtual ray originating from the photo's position.

The photo's salient depth profiles are extracted using a variety of means: edge detection, colour

appearance, blue shift and texture discontinuities, following on from the work by (He et al. 2009). The profiles generated from the real world observation, unlike those in the synthetic image, are noisy and incomplete due to the presence of occlusions (clouds or foreground objects), uneven illumination (which could lead to strong shadows) and poor visibility (e.g. mist or fog).
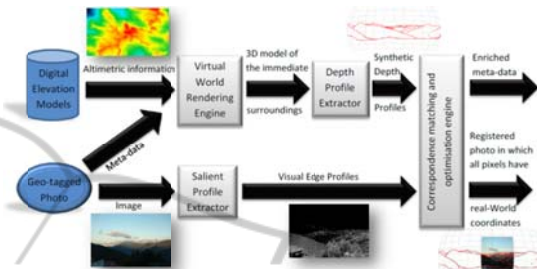

Figure 3: Flowchart of the alignment process.

Figure 4: Tree coverage GIS data taken from the 25m Corine Land cover 2000 (purple) has been rendered onto the 3D terrain model (indicated by blue and red depth profile lines). In this image the 3D model has been made transparent to illustrate photo-model registration.

To increase robustness, region boundary profiles can also be included, e.g. sky-rock, vegetation (see Figure 4), urban-forest, or combination thereof, extracted with appropriate image processing and machine learning approaches.

To provide optimal alignment, even in difficult cases, a visual registration algorithm has been implemented inside an iterative procedure to evaluate correspondences on both local and global scales through an optimization procedure. This optimization algorithm aims to find the camera lens parameters that produce the most accurate alignment. This procedure can be performed by exhaustively evaluating the full range of camera parameters (sampled on a finite grid). This is, however, very computationally intensive. Consequently, we perform profile alignment in a

similar manner to that proposed by (Baboud et al., 2011), and compute a spherical cross correlation based on the method demonstrated by (Kostelec and Rockmore, 2008), using features sampled on an equally spaced grid in the spherical reference system. This approach can be explained more clearly through example:

Let $f(\omega)$ and $g(\omega)$ represent the two spherical functions, where $\omega \in S^2$ is a position in the 2D spherical domain. f refers to the synthetic features, whilst g refers to features extracted from the real image. We are searching for an optimal rotation $R(\alpha, \beta, \gamma) \in SO(3)$, obtained by the composition of three rotations with Euler angles α, β, γ, that, once applied to g, maximizes the matching f. Equation (1) shows how the 'goodness' of the match can be estimated by computing the cross correlation between f and g after the rotation.

$$C(R) = \int_{S^2} f(\omega) g(R^T \omega)\, d\omega \qquad (1)$$
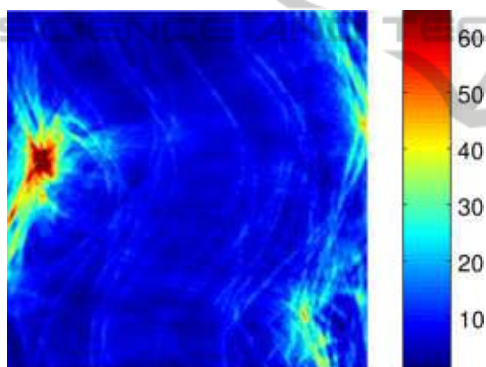

Figure 5: Spherical cross correlation between features sampled on a equally spaced grid in the spherical reference system. The red region represents the best match and blue the worst; the x and y axes relate to pan and tilt angles.

The maximum value of such a correlation in spherical coordinates corresponds to the optimal alignment that can be achieved (in terms of cross-correlation) between the two spherical functions, and the results from a typical alignment can be seen in Figure 5.

3 BENEFITS OF PIXEL-GEO MAPPING

As a direct consequence of profile registration, non-sky pixels within a photo can be assigned to real 3D locations on the planet (i.e. latitude, longitude, altitude). This mapping enables:

- 3D depth perception from 2D images (e.g. retrieve world co-ordinates from an image region),

- re-photography (Bae et al., 2010),

- retrospective augmentation (i.e. facilitating a forensics team to insert and/or extract 3D information; thus, corroborating embedded evidence or potentially revealing manipulation.

The possibility to automatically enrich a photograph with geographically accurate information empowers a user with the ability to discover elusive pieces of evidence that would have otherwise been difficult to retrieve or prove.

Figure 6: Substitution of pixels with geo-data, based on distance from observer and altitude.

Apart from annotating a photo with toponyms or relevant geo-referenced points of interest, we can also drape GIS layer inside it, both in the form of raster (e.g. maps, other geo-registered photos, satellite data) and vector data (e.g. road, rivers, paths, regions, boundaries). An example of such an overlay can be seen in the composite image shown in Figure 6. In this image, two different types of geo-referenced layers have been visualized simultaneously (pixels mapping to locations further than 20km away from the camera were replaced by a synthetic colour representing depth (dark green closer and red farther away), and pixels mapping to an altitude of less than 200m were replaced by geo-corresponding pixels from a cartographic map.

4 FORENSIC EXPERIMENTS

To illustrate the uses and type of problems which can be tackled using scene-model matching, this section provides a few real-life cases where geo-

verification has revealed interesting results. Further examples of registered images can be found on our results Blog[4].

4.1 Bogus K2 Summiting

With a peak elevation of 8611m, K2 is the second-highest mountain on Earth. In this case study, the geo-positional image forensic analysis is focused on Christian Stangl, a famous Austrian climber, whose ambition is to be the first man to complete the Triple Seven Summits challenge[5], climbing the three highest peaks on each of the seven continents.

Conditions were harsh when Stangl arrived at the K2 Base Camp in the summer of 2010 for his third attempt. Stangl left Base Camp on 10 August for a solo attempt. After a 70-hour long summit push, he returned, claiming that he had topped-out at 10am on 12 August. Given the prohibitive conditions and other suspicious incoherencies, his claim was received with scepticism among fellow climbers.

To validate his claim, Stangl quickly submitted a self-portrait photograph supposedly taken on the summit (see Figure 7), *but devoid of meta-data,* to specialized magazines and websites[6].

Figure 7: Self portrait of Christian Stangl, submitted to ExplorersWeb as proof of climbing K2.

As can be seen, there is a small portion of a glacier just visible over his right shoulder, offering an ideal case for a geo-positional image forensic study.

Figure 8 illustrates a visual summary of our findings using the forensic tool we developed and described in this paper.

The upper two images show an excerpt from an Internet sourced photo known to have been taken at Camp 3 on K2 (approximate location Lat: 35.875°N, Lon: 076.531°E, Alt: 7250m) and a synthetic rendering from the same view point. The lower two images are an excerpt from a photo taken at the peak

of K2 (Lat: 35.881°N, Lon: 076.514°E, Alt: 8611m) by Czech climber Libor Uher[7], and likewise a synthetic rendering from the same view point is shown. The central image is a crop, and contrast stretched, excerpt from the same photo submitted by Stangl, shown in Figure 7.
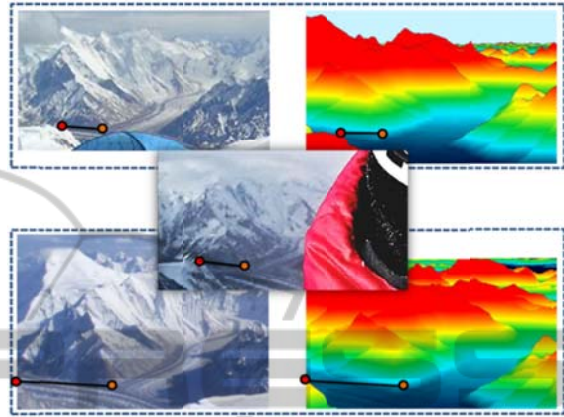

Figure 8: (top row) Photo crop and synthetic rendering from Camp 3; (centre) crop from Stangl's supposed summit photo; (bottom row) crop and synthetic rendering from K2 summit. *Note: same 3D coordinate pairs drawn in all images. Synthetic colouring relative to height.*

The three photographs were registered using our tool and into each, a pair of 3D points with a connecting vector were inserted into the scenes (Lat: 35.802°N, Lon: 076.541°E, 6234 m and Lat: 35.704°N, Lon: 076.542°E, 4688 m). As is evident from this analysis, the photo in question was in fact taken from a location very close to that of Camp 3 and not from the summit.

At the time of the original investigation, the editors of ExplorersWeb[8] used a more basic technique: direct image comparison[9]. They found a photo in a book that had exactly the same composition and managed to overlay Stangl's (fake) summit shot onto it. This photograph was known to have been taken from Camp 3 so thus confirmed their doubts. Inevitably, when Stangl was presented with ExplorersWeb's analysis, he decided to confess that he had faked the climb, generating a huge scandal in the mountaineering community.

4.2 Fake Moon

Another interesting example of a geo-positional image forensic study using the Moon as a non-static but predictable geo-referenceable object is shown in Figure 9.

Figure 9: Conjunction of the Moon, Jupiter and Venus, Palermo, Italy. (left) Photo from Flickr with artificially enlarged Moon; (right) adjusted version of original photo, with Moon resized and repositioned according to EXIF location and time data.

In this popular Flickr image[10], the size of the Moon looks suspiciously large, therefore scene topology matching methods were applied to understand if it was authentic. The stated location of the geo-tagged Flickr image was Lat: 38.1713°N, Lon: 013.3439°E and the time reported in the EXIF was 2008:11:30 18:32:45.

Given these constraints, our tool was used to register the photograph to a 3D synthetic model for that region. As the registration process delivers the relative distances and thus camera calibration parameters, we can determine that the Moon in this photo appears to span 5.6° of the sky. In reality, the apparent diameter of the Moon as viewed from any point on Earth, is always approximately 0.5°, hence it was over 10 times too large. Interestingly, the proportion of the Lunar surface bathed in light was correct, at about 7.9%, so it is suspected that two photos from the same evening had been merged. The location of the Moon in the sky was also incorrect, as it should have been present at 234.04° azimuth and 2.07° altitude (derived from web-based celestial almanacs and the EXIF time).

Based on these findings, a new image (see right of Figure 9) was generated using Photoshop to illustrate the correct size and correct location of the Moon in the sky, based on the original EXIF meta-data; the visible mountains and their relative distances from the observer have also been labeled using GeoNames[11] toponym database. As is evident, the Moon's real location should have been just above the rightmost peak, *Pizzo Vuturo*, producing a less provocative image. Incidentally, the planets Venus and Jupiter are also visible, and had likewise

been subjected to the same up-scaling and repositioning for visual effect.

5 CONCLUSIONS

In this paper, we have presented a system for geo-forensic analysis using computer vision and graphics techniques. The power of such a cross-modal correlation approach has been exemplified through three case-studies, in which claims were disproved, truths revealed or doubts confirmed.

The relative novelty of geo-tagging photos together with the scale and diversity of urban and natural landscapes means that the approaches detailed herein are not suitable for all scenarios. Images containing nondescript content, e.g. indoors, gently rolling countryside and deserts, cannot provide sufficient clues to uniquely pinpoint location or time. However, as more sources of geo-referenced material, e.g. Points of Interest, geo-tagged photos and accurate 3D urban models (like those being created in GoogleSketchUp[12] or OpenStreetMap[13]) become publicly available, the potential to exploit the methods described here will increase correspondingly

ACKNOWLEDGEMENTS

REFERENCES

Baatz, G., Saurer, O., Koeser K. and Pollefeys M. (2012). Large Scale Visual Geo-Localization of Images in Mountainous Terrain. *In Proc. European Conference on Computer Vision.*

Baboud, L., Cadik, M., Eisemann, E. and Seidel, H. (2011). Automatic photo-to-terrain alignment for the annotation of mountain pictures. *In IEEE Conference on Computer Vision and Pattern Recognition*, 41-48.

Bae, S., Agarwala, A. and Durand, F. (2010). Computational rephotography. *In ACM Trans. Graph 29(3)*, 1-15.

Boehme, R., Freiling, F., Gloe T. and Kirchner, M. (2009). Multimedia forensics is not computer forensics. *In Computational Forensics, Lecture Notes in Computer Science,* vol. 5718, 90-103.

Casey, E. (2004). Digital evidence and computer crime. *In Forensic science, computers and the Internet, Academic Press.*

Chippendale, P., Zanin, M. and Andreatta, C. (2009). Collective photography. *In Conference for Visual Media Production,* 188-194.

Cui, Y. and Ge, S. (2003). Autonomous vehicle positioning with GPS in urban canyon environments. *In IEEE Transactions on Robotics and Automation 19(1),* 15-25.

Guidi, G., Beraldin, J., Ciofi, S. and Atzeni, C. (2003). Fusion of range camera and photogrammetry: a systematic procedure for improving 3-d models metric accuracy". *In IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 33(4),* 667-676.

Hays, J. and Efros, A. (2008). Im2gps: estimating geographic information from a single image. *In IEEE Conference on Computer Vision and Pattern Recognition,* 1-8.

He, K., Sun, J., Tang, X. (2009). Single image haze removal using dark channel prior. *In Computer Vision and Pattern Recognition,* 1956–1963.

Karpischek, S., Marforio, C., Godenzi, M., Heuel, S. and Michahelles, F. (2009). Swisspeaks–mobile augmented reality to identify mountains. *In Workshop at the International Symposium on Mixed and Augmented Reality.*

Kostelec, P. and Rockmore, D. (2008). FFTs on the rotation group. *In Journal of Fourier Analysis and Applications 14(2),* 145-179.

Leotta, M. and Mundy, J. (2011). Vehicle surveillance with a generic, adaptive, 3d vehicle model. *In IEEE Transactions on Pattern Analysis and Machine Intelligence 33(7),* 1457-1469.

Li, Z., Zhu, Q. and Gold C. (2005). Digital terrain modeling: principles and methodology. *In CRC Press.*

Luo, J., Joshi, D., Yu, J. and Gallagher A. (2011). Geotagging in multimedia and computer vision – a survey. *In Multimedia Tools and Applications 51,* 187-211.

Paek, J., Kim, J. and Govindan, P. (2010). Energy-e cient rate-adaptive GPS-based positioning for smartphones. *In Proceedings of the 8th international conference on Mobile systems, applications, and services, MobiSys '10,* 299-314.

[1] http://3dnature.com/wcs6info.html

[2] http://3dnature.com/vnsinfo.html

[3] http://www.panoramio.com/

[4] http://tev.fbk.eu/marmota/blog/mosaic/

[5] http://skyrunning.at/en

[6] http://www.explorersweb.com/everest_k2/news.php?id=19634

[7] http://cs.wikipedia.org/wiki/Soubor:Liban_K2_sumit_1_resize.JPG

[8] http://www.explorersweb.com/

[9] http://www.explorersweb.com/everest_k2/news.php?id=19634

[10] http://www.flickr.com/photos/lorca/3074227829/

[11] http://www.geonames.org/

[12] http://sketchup.google.com/

[13] http://www.openstreetmap.org/

[14] https://venturi.fbk.eu