OSCOR: An Orientation Sensor Data Correction System for Mobile Generated Contents

Guanfeng Wang[†], Beomjoo Seo[†], Yifang Yin[†], Roger Zimmermann[†], Zhijie Shen[‡]
[†]School of Computing, National University of Singapore. [‡]Hortonworks Inc.
[†]{wanggf,seobj,yifang,rogerz}@comp.nus.edu.sg [‡]zshen@hortonworks.com

ABSTRACT

In addition to positioning data, other sensor information – such as orientation data, have become a useful and powerful contextual feature. Such auxiliary information can facilitate higher-level semantic description inferences in many multimedia applications, *e.g.*, video tagging and video summarization. However, sensor data collected from current mobile devices is often not accurate enough for upstream multimedia analysis. An effective orientation data correction system for mobile multimedia content has been an elusive goal so far. Here we present a system, termed *Oscor*, which aims to improve the accuracy of noisy orientation sensor measurements generated by mobile devices during image and video recording. We provide a user-friendly camera interface to facilitate the gathering of additional information, which enables the correction process on the server-side. Geographic field-of-view (FOV) visualizations based on the original and corrected sensor data help users understand the corrected contextual information and how the erroneous data possibly may affect further processes.

Categories and Subject Descriptors

H.3.4 [**Information Storage and Retrieval**]: Systems and Software; I.4.8 [**Image Processing and Computer Vision**]: Scene Analysis - Sensor Fusion

Keywords

Mobile multimedia, camera orientation, data correction, digital compass

1. INTRODUCTION

With today's prevalence of camera-equipped mobile devices and their convenience of worldwide sharing, the multimedia content generated from mobile platforms, *e.g.*, smartphones and tablets, have become one of the primary contributors to the media-rich web. The top three most popular cameras in the Flickr community are smartphone models [1].

MM'13, October 21–25, 2013, Barcelona, Spain.
ACM 978-1-4503-2404-5/13/10.
http://dx.doi.org/10.1145/2502081.2502259.

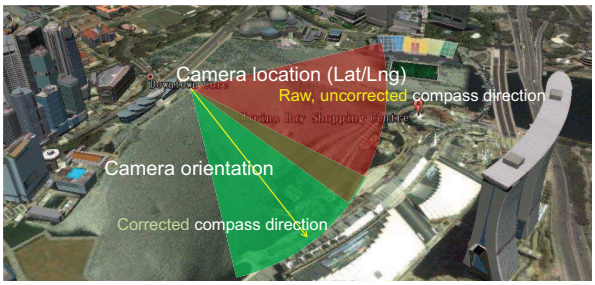

Figure 2: Illustration of inaccurate raw camera orientation data.

Considering the increasing number of new sensors integrated into these devices, a large amount of auxiliary sensor information other than locations is available for both research and commercial utilization. Very useful are the orientation data measured by a digital compass in conjunction with still images and video frames [3].

Researchers have emphasized the importance of rich metadata that surrounds a multimedia object. The orientation information has so far been employed in video encoding complexity reduction, photo organization, video indexing, tagging and summarization. From our experiments, these research studies and related applications are mostly not error resilient with respect to incorrect sensor data input. However, unfortunately most sensor information collected from phones or tablets is not highly accurate due to varying surrounding environmental conditions during data acquisition, and the use of low-cost, consumer-grade sensors. As illustrated in Figure 2, the red pie-shaped slice represents the raw, uncorrected orientation measurement and the green slice indicates the corrected orientation data. As shown, the raw data directly from the smartphone may be extremely incorrect. In some cases the discrepancy is more than 90 degrees from the ground-truth value.

To solve this inaccuracy issue effectively we introduce *Oscor*, which corrects orientation measurements for still images and video frames based on geographic analysis and image processing. Our system collects visual landmark information and matches it against GIS data sources to infer a target landmark's real geo-location. By knowing the geographic coordinates of the captured landmark and the camera, we are able to calculate the corrected orientation measurements. With regards to video recording, in order to minimize the user interaction, we compute the horizontal motion flows and perform landmark matching between sampled frames to interpolate highly accurate orientation data for every frame.

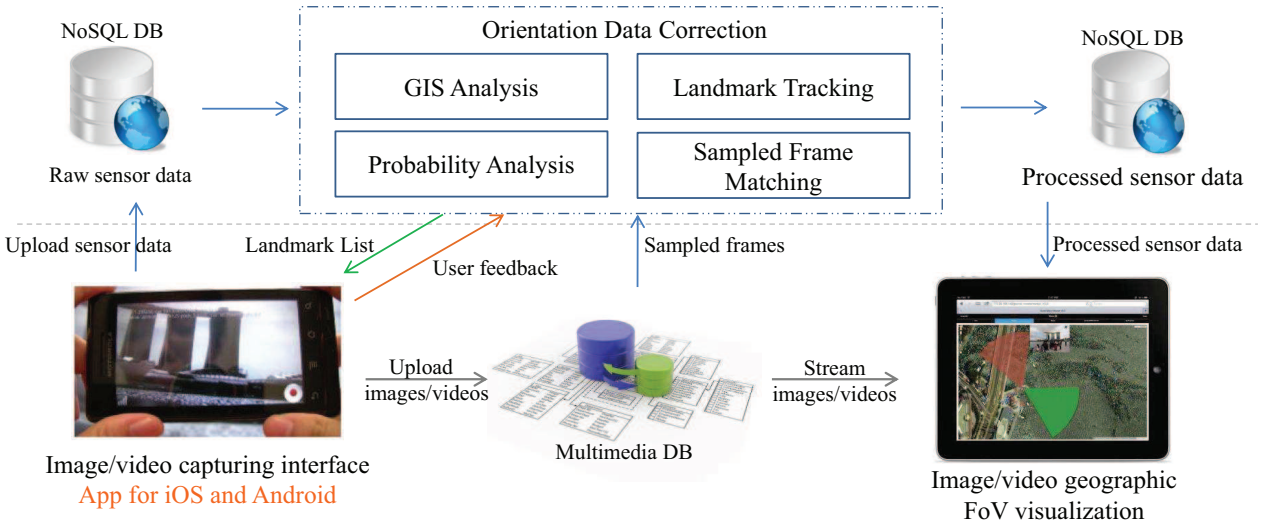

Figure 1: Overview of the *Oscor* system.

2. SYSTEM ARCHITECTURE

Figure 1 illustrates *Oscor*'s overall architecture. For media capture, we provide a camera app which acquires location and orientation information concurrently from multiple sensors while taking images or recording videos [2]. Moreover, our app presents a convenient interface to gather extra information, *e.g.*, a landmark scene position and its name, without requiring cumbersome input from users. For still image capture, we allow users to indicate the width boundary of a close and conspicuous structure by moving their fingers along a building's vertical edges on the just-taken picture. For the video, users can also indicate a pair of perpendicular lines with the same touch-and-move operation during recording. After receiving this edge information on the server side, we convert the users' input of a landmark's position from the pixel index domain into the geospatial domain, and subsequently perform GIS analysis.

We employ OpenStreetMap (OSM) as our GIS reference database and query related building locations and their 2D shape polygons within a certain distance range to the camera position [3]. From the 2D shape information, our system computes each building's width value in the image's x-dimension from the camera's point of view. Afterwards, we utilize the building width information extracted from pixel level, the building distance to the camera from the GIS analysis, and the initial orientation measurement collected from the mobile-embedded digital compass, to build a 3D Gaussian model calculating each building's probability of being the landmark in the user's still picture or video frame. Subsequently, Our system returns the top K landmark names back to the user for selection of the correct one. With the known building reference in both the geospatial and pixel domains, we can accurately estimate the camera pose on a 2D map, and hence output the corrected camera orientation data for the corresponding image/frame.

When recording video, *Oscor* continues to track the interesting feature points detected around the indicated landmark to calculate the positions of this building in the next several seconds of frames. We use an affine model to estimate a referenced landmark's 2D transformation and extract motion vector information in the horizontal axis to compute orientation values for this portion of frames. Since the camera may be moving as time progresses, we perform landmark object matching between sampled frames. For every new matched frame, our system updates the camera coordinates and landmark positions and carries out the previous three steps of analysis again to correct the orientation data.

3. DEMONSTRATION

On the app we have added a transparent overlay on top of the camera interface and leverage multi-touch gestures to collect the building position information indicated by users. After a user uploads the raw sensor data via an HTTP link, the server efficiently stores and indexes this information into a NoSQL MongoDB database.

The *Oscor* user interface visualizes the static or moving field-of-views of images and videos, which allows users to experience fused video browsing based on geographic properties. On a Google Maps canvas multiple images/videos are presented as pins. When a user clicks or touches a pin, a map-overlay image viewer/video player is launched and the video is rendered from the designated starting location. During video playback, the camera's current location and viewable scenes are animated along the corresponding GPS trajectory. To help users visualize the corrected contextual information and how the erroneous data possibly effects further processes, two viewable scenes based on asynchronously retrieved raw and corrected camera orientation data are rendered on the same interface.

4. ACKNOWLEDGMENT

This research has been supported by the Singapore National Research Foundation under its International Research Centre @ Singapore Funding Initiative and administered by the IDM Programme Office.

5. REFERENCES

[1] Flickr: Camera Finder. `www.flickr.com/cameras`. [Online; accessed April-2013].

[2] B. Seo, J. Hao, and G. Wang. Sensor-rich Video Exploration on A Map Interface. In *ACM Multimedia*, 2011.

[3] Z. Shen, S. Arslan Ay, S. H. Kim, and R. Zimmermann. Automatic Tag Generation and Ranking for Sensor-rich Outdoor Videos. In *ACM Multimedia*, 2011.