

## Chapter 9

# Image Clustering and Video Summarization for Efficient 3D Modelling and Reconstruction

**Athanasios Voulodimos**

*University of West Attica, Athens, Greece*

**Eftychios Protopapadakis**

*National Technical University of Athens, Athens, Greece*

**Nikolaos Doulamis**

*National Technical University of Athens, Greece*

**Anastasios Doulamis**

*National Technical University of Athens, Greece*

### ABSTRACT

*Although high quality 3D representations of important cultural landmarks can be obtained via sophisticated photogrammetric techniques, their demands in terms of resources and expertise pose limitations on the scale at which such approaches are used. In parallel, the proliferation of multimedia content posted online creates new possibilities in terms of the ways that such rich content can be leveraged, but only after addressing the significant challenges associated with this content, including its massive volume, unstructured nature, and noise. In this chapter, two strategies are proposed for using multimedia content for 3D reconstruction: an image-based approach that employs clustering techniques to eliminate outliers and a video-based approach that extracts key frames via a summarization technique. In both cases, the reduced and outlier-free image data set is used as input to a structure from motion framework for 3D reconstruction. The presented techniques are evaluated on the basis of the reconstruction of two world-class cultural heritage monuments.*

DOI: 10.4018/978-1-5225-5294-9.ch009

## INTRODUCTION

The free and online availability of large collections of images and videos located on distributed and heterogeneous platforms over the Web is one of the prominent characteristics of today's digital era, reigned by the Internet, social media and powerful mobile devices. The abundance of shared photographs spurred the emergence of new image retrieval techniques based not only on images' visual information, but also on geo-location tags and camera exif data. These massive visual collections provide a unique opportunity for urban areas and cultural heritage documentation and 3D reconstruction. The main challenge, nevertheless, is that Internet image datasets are unstructured containing many outliers. Therefore, content-based image filtering is necessary to discard image outliers that either confuse or significantly delay the employed 3D reconstruction frameworks, such as, for example, Structure from Motion (e.g. VisualSFM).

In contrast with sophisticated airborne and close range photogrammetric approaches, where 3D data acquisition is accomplished in a constrained environment using specialized equipment and sophisticated techniques, Web-based collections can be exploited for a much easier and more "user-friendly" cultural heritage e-documentation. However, the main difficulty in implementing a precise 3D reconstruction of an object from unstructured internet image collections (being captured for personal use instead of reconstruction purposes), is that there are usually several outliers in the set of retrieved data deteriorating both performance and computational cost. Consider, for example, a query containing the keywords "Acropolis, Parthenon." As a response to that query, a large set of images are retrieved, which depict not only the Parthenon monument itself, but also the view of the city of Athens from the Acropolis hill or of people being photographed in the environment of the monument. These image outliers confuse any e-documentation algorithm. Although auto-generated geo-location tags can improve visual content characterization and therefore the retrieval performance, they suffer from low precision since geo-information does not correctly describe what is actually depicted. While there exist 3D reconstruction algorithms, such as Structure from Motion (SfM), which demonstrate robustness against noisy data, their computational complexity significantly increases with respect to the number of input data. This makes direct implementation of such methods for large image volumes practically impossible. Therefore, in the cases where retrieved images are used as input, content-based filtering algorithms are necessary for an effective and computationally efficient 3D reconstruction exploiting distributed Web based image collections. Content-based filtering algorithms, apart from discarding image outliers, also organize the retrieved unstructured content into well-structured forms to optimize both 3D reconstruction performance and computational cost.

In more detail, the accuracy of 3D reconstruction over a given image dataset is inherently dependent on the number of images that will be fed as input to the Structure from Motion (SfM) scheme. Given an image dataset the best 3D reconstruction accuracy is achieved when all visually similar images are fed into a SfM method. One way to exploit all visually similar images is to give as input to the SfM method the entire image dataset. However, the time complexity for a typical incremental SfM method is of order where  $N$  stands for the number of images. This complexity makes SfM not scalable to large image collections. In order to decrease computational cost associated with 3D reconstruction, the initial image dataset can be pruned by removing outliers. When outliers' removal process is very precise, dataset reduction does not affect reconstruction accuracy, since the relevant (reduced) dataset will contain only all visually similar images. Therefore, the metric that can be associated with reconstruction accuracy is the metric of recall. On the other hand, the reduction of SfM computational time is dependent on the

20 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/image-clustering-and-video-summarization-for-efficient-3d-modelling-and-reconstruction/254592](http://www.igi-global.com/chapter/image-clustering-and-video-summarization-for-efficient-3d-modelling-and-reconstruction/254592)

## Related Content

---

### Development of a Surface EMG-Based Control System for Controlling Assistive Devices: A Study on Robotic Vehicle

Uvanesh K., Suraj Kumar Nayak, Biswajeet Champaty, Goutam Thakur, Biswajit Mohapatra, D. N. Tibarewala and Kunal Pal (2016). *Classification and Clustering in Biomedical Signal Processing* (pp. 335-355).

[www.irma-international.org/chapter/development-of-a-surface-emg-based-control-system-for-controlling-assistive-devices/149398](http://www.irma-international.org/chapter/development-of-a-surface-emg-based-control-system-for-controlling-assistive-devices/149398)

### Projective Geometry for 3D Modeling of Objects

Rimon Elias (2013). *Image Processing: Concepts, Methodologies, Tools, and Applications* (pp. 125-144).

[www.irma-international.org/chapter/projective-geometry-modeling-objects/77542](http://www.irma-international.org/chapter/projective-geometry-modeling-objects/77542)

### Elastic Matching Techniques for Handwritten Character Recognition

Seiichi Uchida (2008). *Pattern Recognition Technologies and Applications: Recent Advances* (pp. 17-38).

[www.irma-international.org/chapter/elastic-matching-techniques-handwritten-character/28024](http://www.irma-international.org/chapter/elastic-matching-techniques-handwritten-character/28024)

### Augmented Small-Scale Database to Improve the Performance of Eigenface Recognition Technique

Allam Mousa, Rana Salameh and Rawan Abu-Shmais (2012). *International Journal of Computer Vision and Image Processing* (pp. 59-70).

[www.irma-international.org/article/augmented-small-scale-database-improve/72316](http://www.irma-international.org/article/augmented-small-scale-database-improve/72316)

### Image Compression Technique for Low Bit Rate Transmission

Shaimaa A. El-said, Khalid F. A. Hussein and Mohamed M. Fouad (2011). *International Journal of Computer Vision and Image Processing* (pp. 1-18).

[www.irma-international.org/article/image-compression-technique-low-bit/64182](http://www.irma-international.org/article/image-compression-technique-low-bit/64182)