UNIVERSITY OF TWENTE.

EFFICIENT USE OF VIDEO FOR 3D MODELLING OF CULTURAL HERITAGE OBJECTS

BASHAR ALSADIK, MARKUS GERKE, GEORGE VOSSELMAN

PIA 15 : PHOTOGRAMMETRIC IMAGE ANALYSIS
MUNICH 2015

ITC FACULTY OF GEO-INFORMATION SCIENCE AND EARTH OBSERVATION

Introduction

For image based modeling IBM:

Still imaging:

Pros: high resolution – better quality and accuracy – reasonable number of shots to process

Cons: need proficiency - wide baseline (difficult to match images)

Introduction

For image based modeling IBM:

Still imaging:

Pros: high resolution – better quality and accuracy – reasonable number of shots to process

Cons: need proficiency - wide baseline (difficult to match images)

Video imaging:

Pros: much easier to take – high redundancy (short baseline)

Cons : low resolution - large number of images, possibly blurred

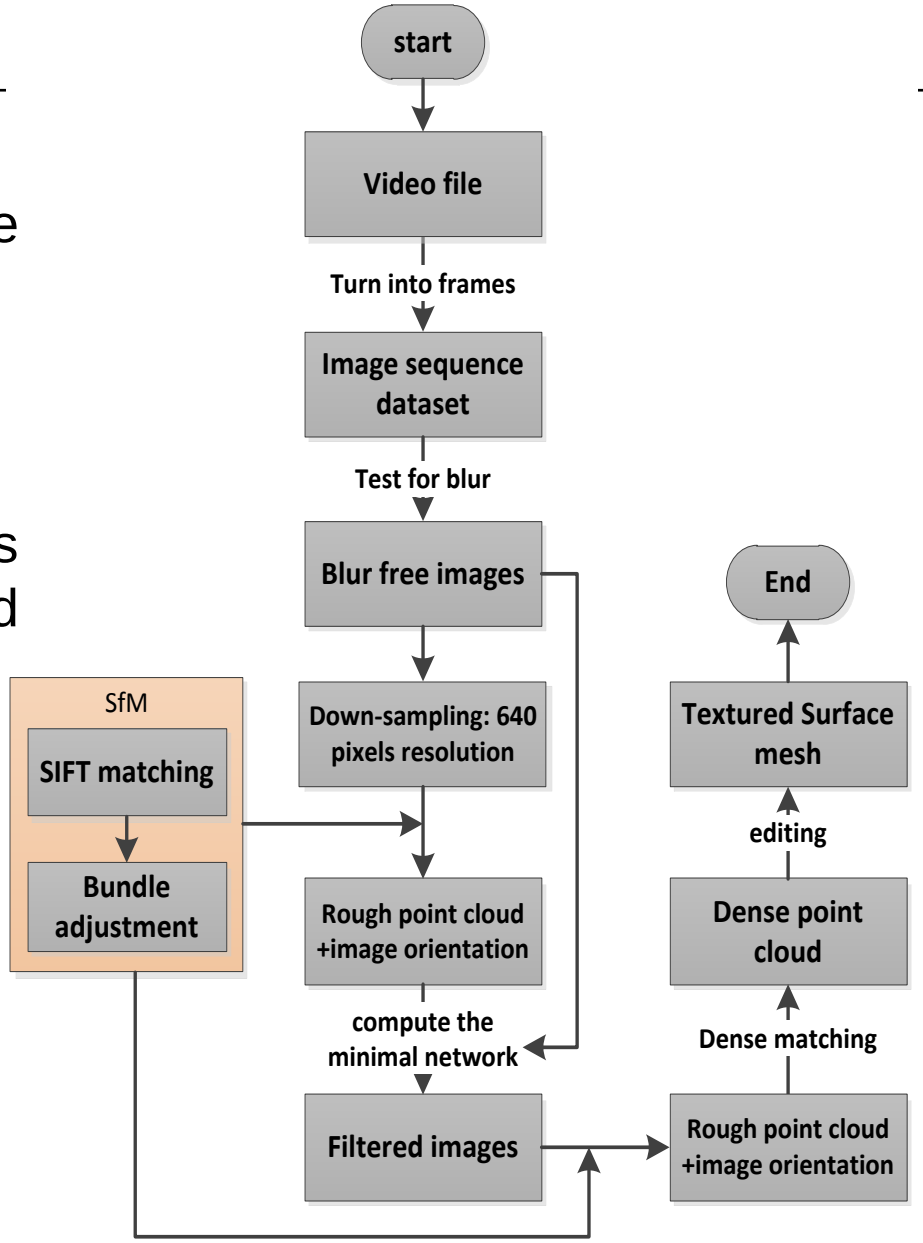
This paper presents a method to create 3D models from the minimum number of video images that guarantees both:

- **Full coverage.**
- **Limited blur.**
- **Faster implementation because of reduced image number**

Method

The key idea of an efficient use of the video image sequence in modelling:

- By removing blurry video images.
- Filter out redundant image frames according to some criteria based on coverage and B/D ratio.



Removal of Blurred Images

Use the Crete et al. 2007-approach to compute a blur metric and select only sharp images.

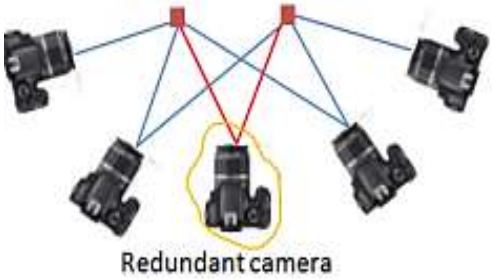
Blur metric= 0.29

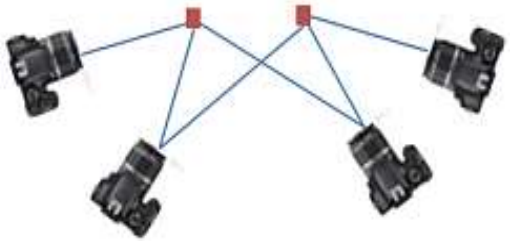
Blur metric= 0.46

Crete, F., Dolmiere, T., Ladret, P., Nicolas, M., 2007. The Blur Effect: Perception and Estimation with a New No-Reference Perceptual Blur Metric, SPIE Electronic Imaging Symposium Conf Human Vision and Electronic Imaging, San Jose : États-Unis d'Amérique

Minimal camera network

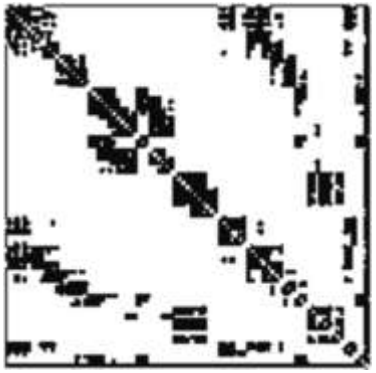
- Concept: at least three cameras viewing each object point.
- Cameras are redundant if they only add the 4th or more view, but B/D ratio considered, as well!





B/D < threshold

- Needed: sparse point cloud and approximate image orientation: Thus apply SfM on the blur-free full set of the downsampled video images.
- After filtering: use full resolution images and approximate orientations and matching graph to guide tie point matching



Alsadik, B., Gerke, M., Vosselman, G., 2013. Automated camera network design for 3D modeling of cultural heritage objects. Journal of Cultural Heritage 14, 515-526.

Experimental Tests

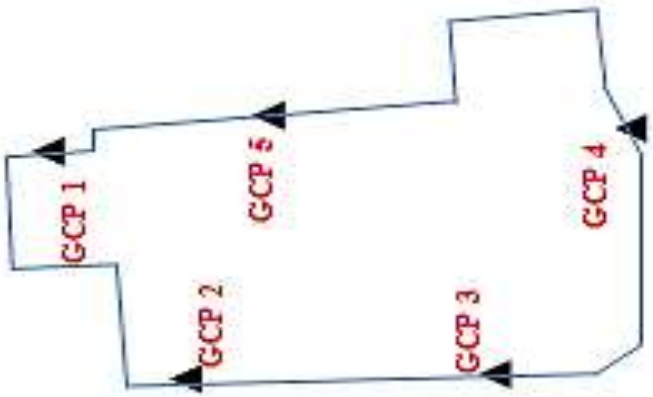
Church building

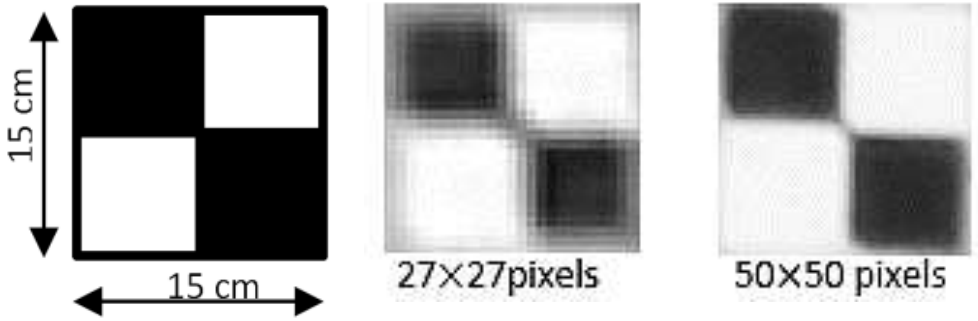
Monument

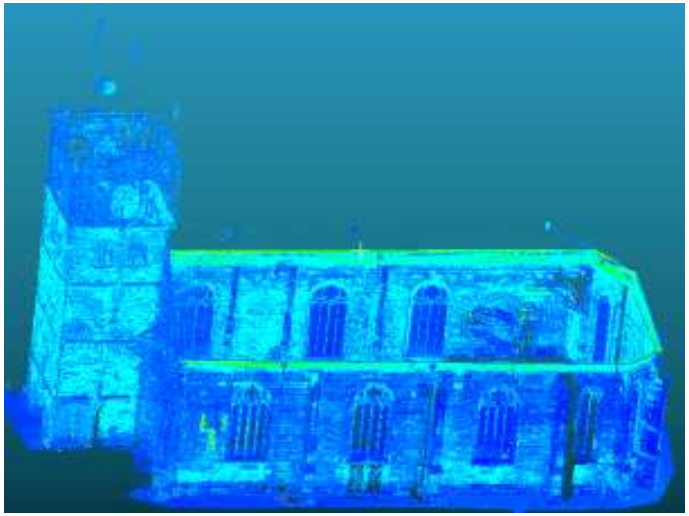
- Canon EOS 500D with 1920×1080 pixels in .MOV format with a frame rate of 20 fps.
- Dell Latitude E6540 Core i7
- Agisoft Photoscan software
- Terrestrial laser scanning (TLS) “Trimble CX scanner” where the manufacturer single point accuracy standards were: 4.5 mm @ 30 m

Church building experiment – referencing

- Five ground control points GCPs were fixed on the church facades to register the created video based point clouds into the TLS point cloud (23 million points)

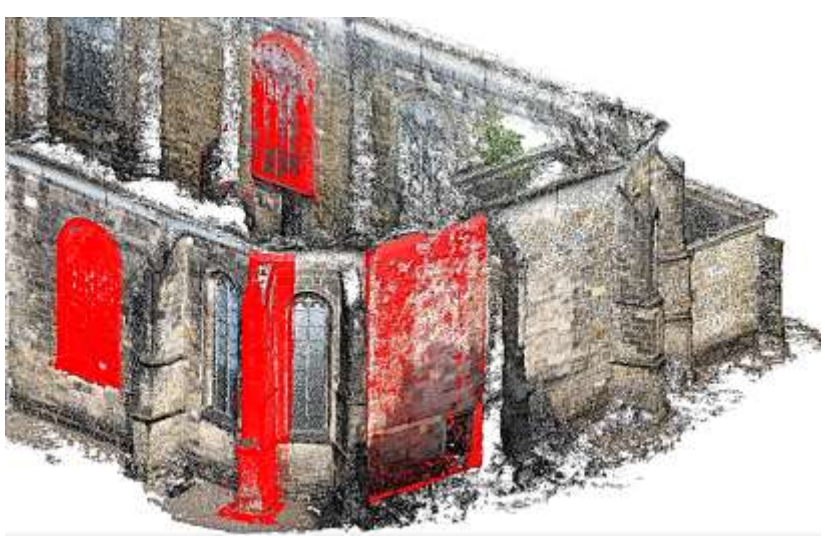






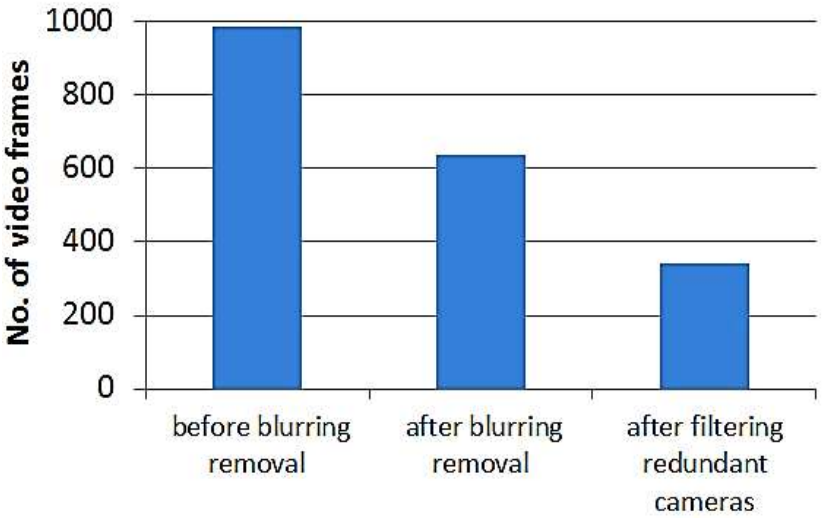
Imaging network [pixels]	Av. scale	Pixel size [mm]	GSD [mm]
Video : 1920×1080	1/300	0.020	6
Still images: 4752×3168	1/600	0.005	3

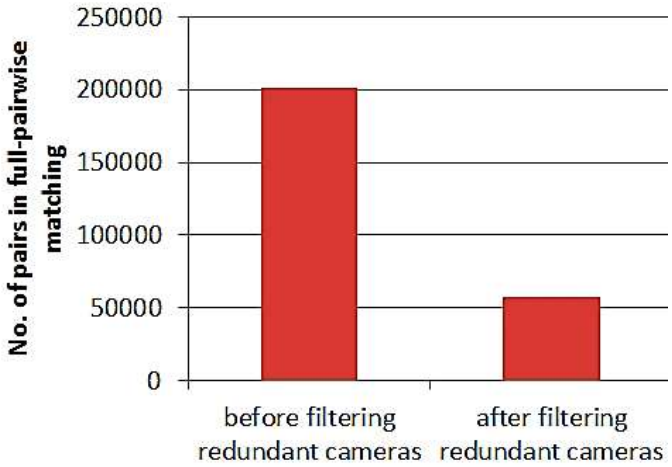
Church building experiment - validation



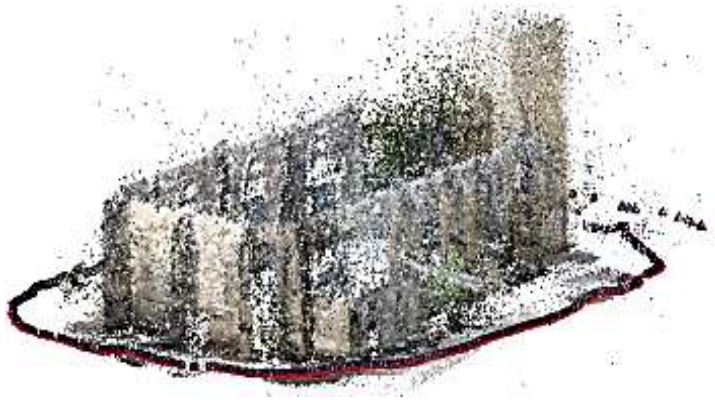
Evaluation: cloud to cloud distance C2C is computed for a randomly selected four elements of the whole church building.

Church building experiment - processing





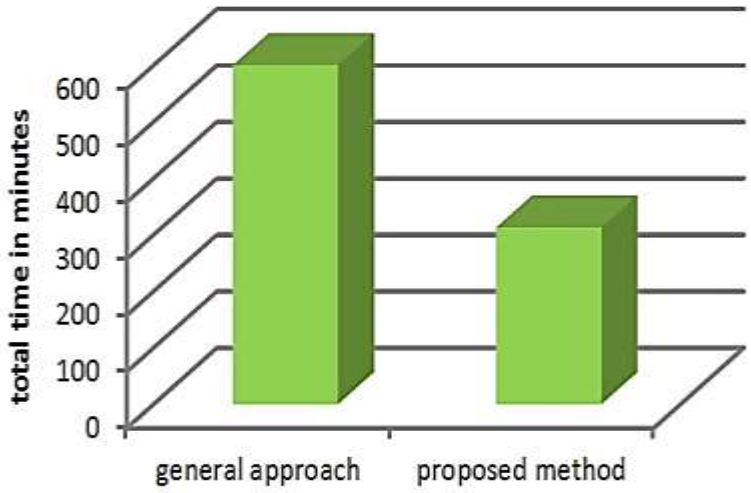
635 image



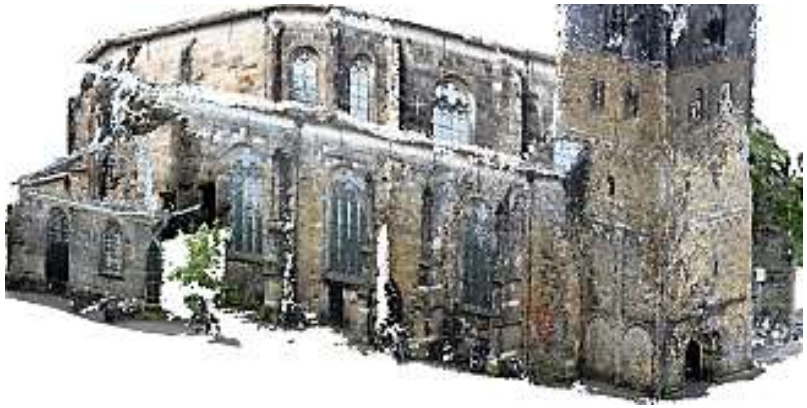
347 image

Church building experiment

Time consumption for SfM and dense matching



Point cloud of unfiltered sequence



Point cloud of filtered sequence

Church building experiment - C2C distance

Before filtering

Point cloud comparison	No. of points	Mean shift [cm]	Std. deviation [cm]
Window I	221294	2.0	±1.5
Window III	174690	3.6	±2.4
Column	337692	4.9	±2.7
Facade	494347	7.9	±2.0

After filtering

Point cloud comparison	No. of points	Mean shift [cm]	Std. deviation [cm]
Window I	205811	2.7	±2.1
Window II	144556	3.0	±2.1
Column	314312	5.3	±3.1
Facade	487977	6.6	±1.9

Church building experiment - still imaging

118 images

- Compare to still imaging model.
- evaluate the amount of details and visualization acquired from video imaging.

Video - based

Still images - based

Video - based

Still image - based

$\cong 200000\ points$ $\cong 850000\ points$

Monument experiment

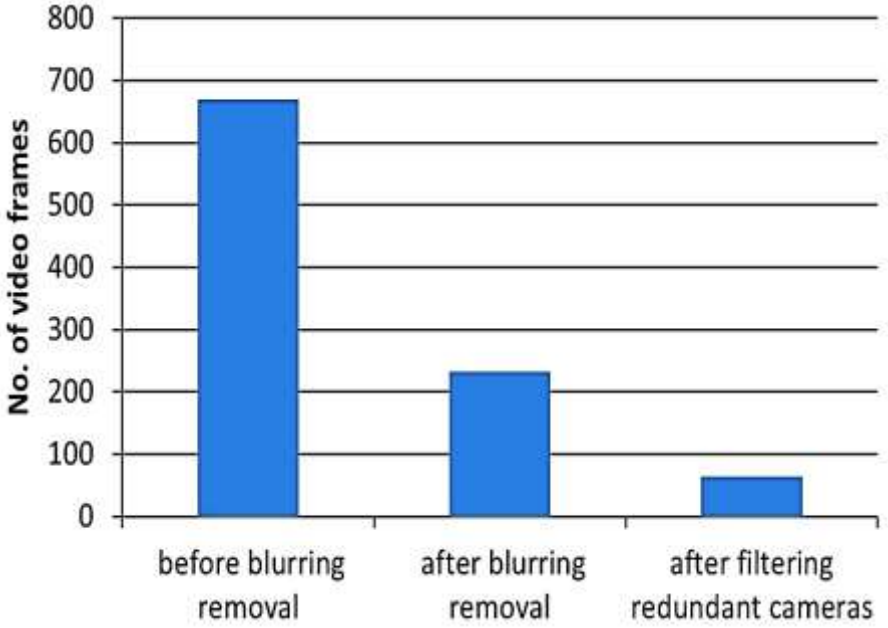
The second experiment is applied to a monument in the old city of Enschede which is built in 1912 to commemorate the disaster of the city fire in 1863.

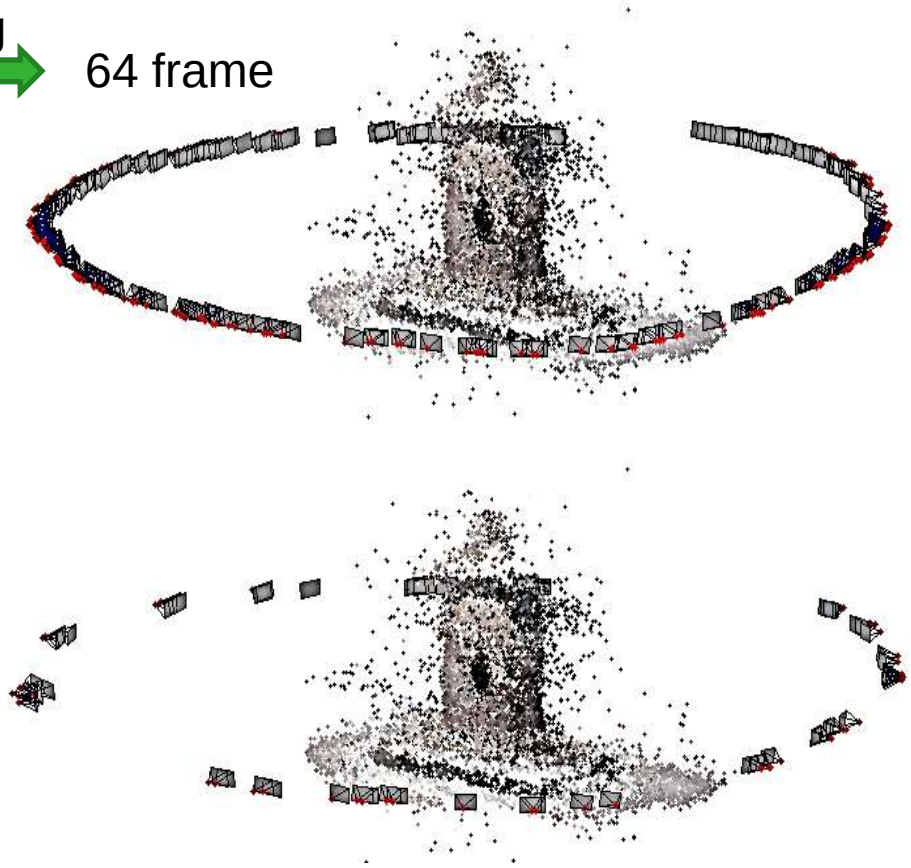
The point cloud acquired by TLS consisting of 1 million points

- A video imaging with a scale of 1/250.
- 3 GCPs for referencing.
- The pixel size was 0.02mm and the GSD was 5mm.

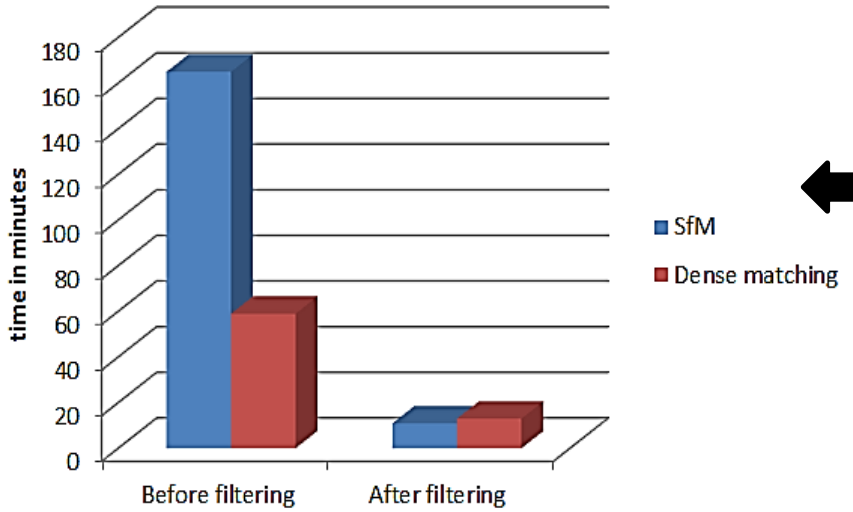
Monument experiment – filtering

670 frame → Blur free → 233 frame → filtering → 64 frame





Monument experiment



The time consumed for the SfM and dense matching before filtering (233 images) and after filtering (64 images).

A dense point cloud after filtering was created and resulted in $\cong 9 \times 10^5$ points.

Monument experiment - validation

For validation, two patch clusters of points were selected to check the accuracy.

The tests resulted in mean distances of 4.7±1.2cm and 1.0 ± 0.6 cm respectively.

Conclusions

- It is possible to have a reliable video (1920×1080 pixels) based 3D models of objects for a low or midrange applications accuracy (≈5cm error) and visualization.

- The proposed method is efficient to reduce the computations for processing video frames with no significant loss of model accuracy and reconstructed model completeness.

- The proposed filtering will significantly reduce the processing time compared to the conventional approach.

Thank you