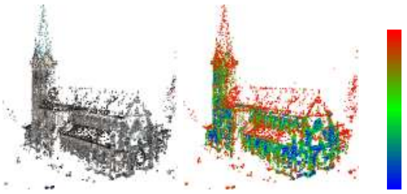
A unified framework for content-aware view selection and planning through view importance

Massimo Mauro[1]
m.mauro001@unibs.it
Hayko Riemenschneider[2]
http://www.vision.ee.ethz.ch/~rhayko/
Alberto Signoroni[1]
http://www.ing.unibs.it/~signoron/
Riccardo Leonardi[1]
http://www.ing.unibs.it/~leon/
Luc Van Gool[2]
http://www.vision.ee.ethz.ch/~vangool/

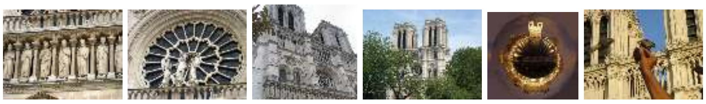
[1] Department of Information Engineering
University of Brescia
Brescia, Italia

[2] Computer Vision Lab
Swiss Federal Institute of Technology
Zurich, Switzerland



(a) SFM (b) E_{agg}

Figure 1: View importance as energy heatmap (the more red, the more salient and hence important) as example on *Fraumunster* SfM cloud.



(a) Best1 (b) Best2 (c) Best3 (d) Worst1 (e) Worst2 (f) Worst3

Figure 2: Best and worst views on *Notre Dame* dataset.

Take home message: Reduction and selection of views through structure analysis of 3D point clouds. Our importance measure is much more effective without loosing salient structures.

Introduction: The great and unordered deal of images available on the Internet leads to two challenging problems for image-based 3D reconstructions: completeness and scalability. On one side, photographs are only taken from "popular" viewpoints, leading to incomplete 3D models. On the other side, the collected images are redundant. Next-Best-View (NBV) and Image Selection (IS) algorithms are thus needed to propose new and select from redundant viewpoints for efficient reconstruction.

In this work we propose two methods for IS and NBV, based on the idea of *view importance*: how important is a given viewpoint for a 3D reconstruction? Our answer is a unified framework for search of important viewsbased on a set of content-aware *quality features* extracted on the Structure-from-Motion (SfM) point cloud.

Quality Features. For every 3D point, we extract the following:

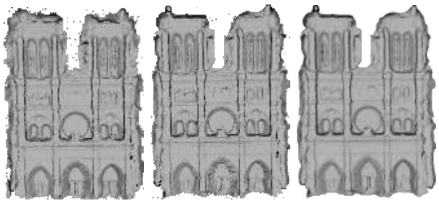
- *Density* is defined as the number of points contained in a sphere around the point.
- *Uncertainty* considers the maximum angle between the viewing directions of the evaluated point.
- *2D saliency* evaluates the meaningfulness of the 2D content around the point. It is estimated by reprojecting the point in the original images and measuring the gradient in the neighbourhood.
- *3D saliency* measures the geometric complexity around a point. It is estimated by the Difference of Normals (DoN) operator [2].

Feature aggregation. All the features have different ranges. We rescale them in the range [0,1] using a logistic function and we call *normalized energies* the obtained values. We note them as E_D, E_U, E_{2D} and E_{3D} respectively. The *aggregate energy* (example in Figure 1) is then defined as a linear combination

$$E_{agg} = w_D E_D + w_U E_U + w_{2D} E_{2D} + w_{3D} E_{3D} \tag{1}$$

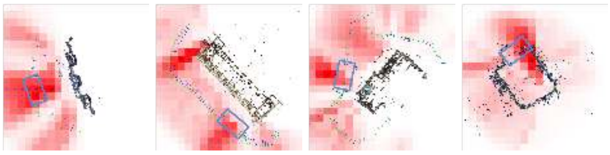
View importance. The key concept behind both our IS and NBV algorithms is the *view importance*. Given a point cloud $\mathcal{P}$, the view importance I of a camera C is defined as the mean energy E_{agg} combined over all its visible points:

$$I(C,\mathcal{P}) = \frac{\sum_{p_i \in V_C} E_{agg}(p_i)}{|V_C|} \tag{2}$$



(a) 715 img (all) (b) 167 (CMVS) (c) 106 (ours)

Figure 3: Similar 3D mesh results on *Notre Dame* for much smaller image sets. Our method effectively reduces yet keeps the salient 3D structures.



(a) Herz-Jesu-P25 (b) Hall (c) Fraumunster (d) Castle-P30

Figure 4: Next-Best-View grids. Importance is high in regions (blue rectangle) where cameras have been artificially removed.

where V_C is the set of points in $\mathcal{P}$ visible from camera C. We use this basic definition in two variants I_{IS} and I_{NBV} (for IS and NBV respectively) to better adapt to the problem at hand. See paper for details.

View selection. The aim of image selection (IS) is to remove redundant images. We use an "importance-guided" approach: at every step our algorithm cuts out the *worst view* in terms of *view importance*, for an example see Figure 2. The worst view satisfies the relation:

$$C_{IS} = \arg\min_{C} I_{IS}(C,\mathcal{P}) \tag{3}$$

Next-Best-View planning. The goal of a Next-Best-View algorithm is to find the camera C_{NBV} with the largest view importance

$$C_{NBV} = \arg\max_{C} I_{NBV}(C,\mathcal{P}) \tag{4}$$

Since a great deal of images are collected by humans, we simplify the NBV search by fitting a plane primitive to the SfM camera centers. We then define a rectangular region around the point cloud and divide it in cells. We position a camera in every grid cell we evaluate the view importance for a given number of evenly spaced orientations, obtaining *view importance grids* as in Figure 4.

Experiments. The experiments show the effectiveness of the proposed content-aware methods. Our NBV planning effectively finds regions where viewpoints are missing. Our IS method reduces the number of images without losing salient regions of the scene, comparing favorably with the state-of-the-art image selection in CMVS [1]. E.g., For *Notre Dame*, we can remove more than 90% of the images (609/715), reducing the runtime of the reconstruction to 1/20th of the time without causing significant differences in reconstruction quality (Figure 3).

Acknowledgments. This work was supported by the European Research Council (ERC) under the project VarCity (#273940) and by the Italian Ministry of Education, University and Research under the PRIN project BHIMM (Built Heritage Information Modeling and Management).

[1] Y. Furukawa, B. Curless, S. Seitz, and R. Szeliski. Towards internet-scale multi-view stereo. In *CVPR*, 2010.

[2] Y. Ioanou, B. Taati, R. Harrap, and M. Greenspan. Difference of normals as a multi-scale operator in unorganized point clouds. In *3DPVT*, 2012.