Mesh saliency detection using Convolutional Neural Networks

Stavros Nousias*, Gerasimos Arvanitis, Aris S. Lalos and Konstantinos Moustakas

Industrial Systems Institute, Athena Research Center
Department of Electrical & Computer Engineering, University of Patras, Greece

June 18, 2020

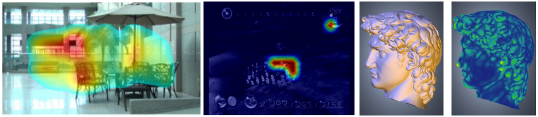
Table of Contents

Table of Contents

Purpose and Trend

Definition of context I

- Visual saliency
  - Significance of a region or point in space, with respect to human visual perception
  - Stimulus-driven process
  - The salient part differs from its neighbouring region
  - Applications
    - Image processing→ 2D
    - **Geometry processing → 3D**



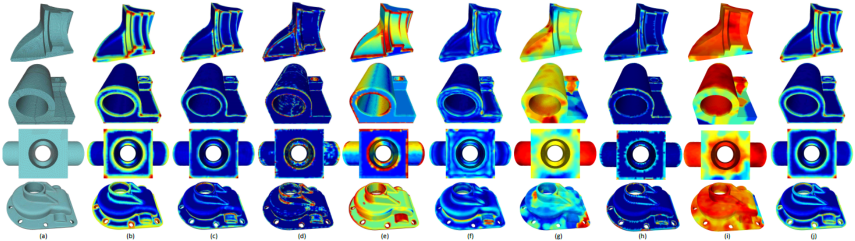
(a) (b) (c) (d)

- **a)** A. Nguyen, J. Kim, H. Oh, H. Kim, W. Lin, and S. Lee,“Deep visual saliency on stereoscopic images,”IEEETransactions on Image Processing, vol. 28, no. 4, pp.1939–1953, April 2019
- **b)** T. Alshawi, Z. Long, and G. AlRegib,“Unsuper-vised uncertainty estimation using spatiotemporal cuesin video saliency detection,”IEEE Transactions on Im-age Processing, vol. 27, no. 6, pp. 2818–2827, June2018
- **c,d)** Ran Song, Yonghuai Liu, Ralph R Martin, and Ka-rina Rodriguez Echavarria,“Local-to-global meshsaliency,”The Visual Computer, vol. 34, no. 3, pp. 323–336, 2018.

Purpose and Trend

Definition of context II

- State of the art on 3D saliency mapping



(a) Original model,**(b)** the eigenvalues of small patches (spectral analysis), **(c)** the Robust PCA approach (geometrical analysis), **(d)** Wei *et al.* 2018, **(e)** Tao *et al.* 2016, **(f)** Lee *et al.* 2005, **(g)** Song *et al.* 2014, **(h)** Guo *et al.* 2018, **(i)** Song *et al.* (CNN) 2019, **(j)** Arvanitis *et al.* 2020

Table of Contents

Motivation and Goal

Motivation

- Convolutional neural networks (CNNs) have been extensively used for the extraction of saliency maps from images
- Recent works have used also employed CNNs to extracts saliency maps for 3D meshes generated by a multi-view setup [a]

[a] MN Favorskaya and LC Jain, "Saliency detection in deep learning era: trends of development," Management Information Systems, , no. 3, pp. 10–36, 2019.

Goal

- Utilize CNNs to automatically extract saliency maps from 3D geometries
- Develop a descriptor to facilitate learning for 3D meshes where
  - The sampling is non-uniform
  - In contrast to images based or voxel based approaches structured lattices are not present
  - Enable the processing of very large and dense meshes
  - Lower complexity
  - Allow parallelization

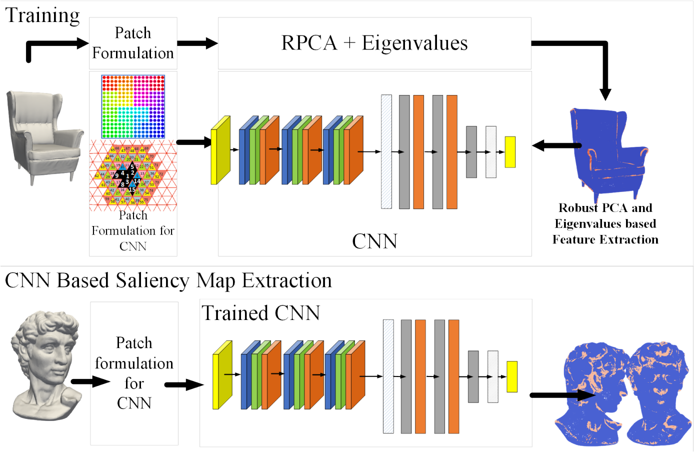
Table of Contents

Contributions

- We employ a method based on fusion of robust principal component analysis and eigenvalues to extract saliency maps of 3D scanned meshes and formulate the training set for the effective training process of a convolutional neural network.
- We formulate a geometric 3D patch descriptor to facilitate the learning process, significantly reducing the required training dataset.
- We provide an experimental evaluation of the proposed methods in terms of execution times and visual inspection.
- We provide use case examples, namely compression and simplification, evaluated with relevant metrics.

Table of Contents

Overview



Training
Patch Formulation
RPCA + Eigenvalues
Patch Formulation for CNN
CNN
Robust PCA and Eigenvalues based Feature Extraction
CNN Based Saliency Map Extraction
Patch formulation for CNN
Trained CNN

Saliency analysis via sparse modelling

Spectral saliency

- For each face f_i , estimate the corresponding patch $\mathbf{N}_i \in \mathbb{R}^{(k+1)\times 3}$, consisting of the $k+1$ neighbouring centroid normals

$$\mathbf{N}_i = [\mathbf{n}_{ci},\ \mathbf{n}_{ci_1},\ \mathbf{n}_{ci_2},\ \cdots,\ \mathbf{n}_{ci_k}]^T\ \forall\ i = 1, \cdots, n_f \tag{1}$$

- Each matrix $\mathbf{N}_i$ is used for the estimation of the covariance matrices

$$\mathbf{R}_i = \mathbf{N}_i^T \mathbf{N}_i \in \mathbb{R}^{3\times 3} \tag{2}$$

- Decomposition

$$\mathbf{R}_i = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T \tag{3}$$

- The spectral saliency s_{1i} of a centroid $\mathbf{c}_i$

$$s_{1i} = \frac{1}{\sqrt{\lambda_{i1}^2 + \lambda_{i2}^2 + \lambda_{i3}^2}}\ \forall\ i = 1, \cdots, n_f \tag{4}$$

- Exploit the sparsity of the centroid normals.
- Construction of three matrices $\mathbf{E}_l \in \mathbb{R}^{n_f \times (k+1)} \ \forall \ l \ \in \{x, y, z\}$
- $\mathbf{E}_l = [\mathbf{N}_{1l}; \mathbf{N}_{2l}; \cdots ; \mathbf{N}_{n_f l};]$
- Decomposition of a matrix $\mathbf{E}$ into a low-rank matrix $\mathbf{L}$ and a sparse matrix $\mathbf{S}$

$$\arg\min_{\mathbf{L},\mathbf{S}} \|\mathbf{L}\|_* + \lambda\|\mathbf{S}\|_1, \quad \text{s.t. } \mathbf{L} + \mathbf{S} = \mathbf{E}, \tag{5}$$

- Estimation of the geometric saliency

$$s_{2i} = \sqrt{S_{i1_x}^2 + S_{i1_y}^2 + S_{i1_z}^2} \ \ \forall \ i = 1, \cdots , n_f \tag{6}$$

Normalization

$$\bar{s}_{1i} = \frac{s_{1i} - \min(s_{1i})}{\max(s_{1i}) - \min(s_{1i})} \; \forall \; i = 1, \cdots, n_f \tag{7}$$

Combination

$$s_{ci} = \frac{\bar{s}_{1i} + \bar{s}_{2i}}{2} \quad \forall \; i = 1, \cdots, n_f \tag{8}$$

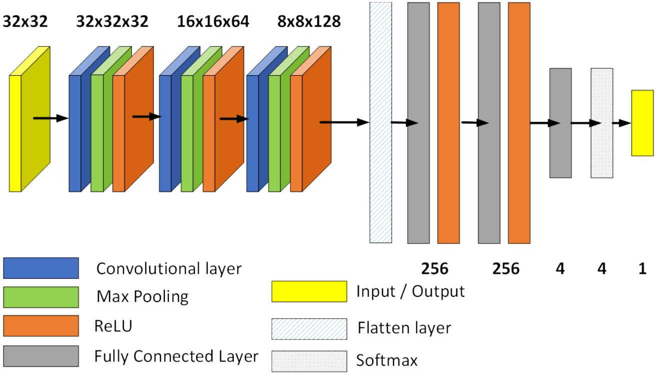
Quantization

$$c_i = \lfloor \frac{s_i}{\Delta} \rfloor \tag{9}$$

- $\lfloor \cdot \rfloor$ is the floor function
- $\Delta = \frac{1}{4}$

CNN architecture, training and saliency map extraction

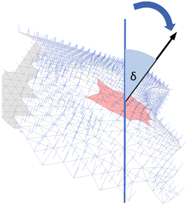
Convolutional neural network

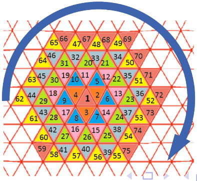


CNN architecture, training and saliency map extraction

Patch descriptor

- Formulate a geometric descriptor to facilitate training
- Requirements
  - Address non-uniformity
  - Achieve invariance in terms of scale, transformation and rotation
- Scale and transformation invariance
  - Utilize face normals instead of positions
- Rotation invariance in x,y,z axes
  - Enumerate faces in ring formations
  - Rotate each patch so that they "look" towards the same direction
    - Rotate by angle δ_{n_i}, around rotation axis $\mathbf{a}_{n_1}$ **so that**,
    - $\mathbf{n}_f = \frac{1}{N}\sum_{i \in \mathcal{N}_i} A_i \cdot \mathbf{n}_{c_i} = \mathbf{c}$





CNN architecture, training and saliency map extraction

Input tensor

- Formulate input tensor $\mathcal{N} \in \mathcal{R}^{32x32x3}$ using the predefined arrangement
- Improve local coherency
    - Utilize **space filling curves** that improve local coherency
    - Hilbert curve

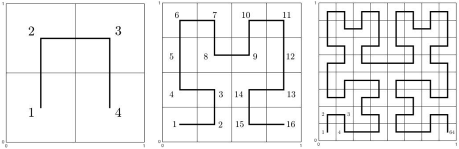
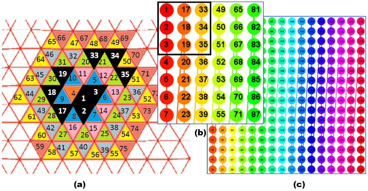


Figure: Hilbert curve.



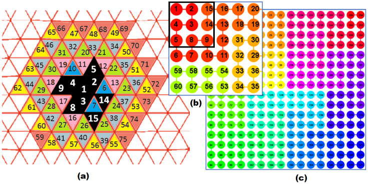
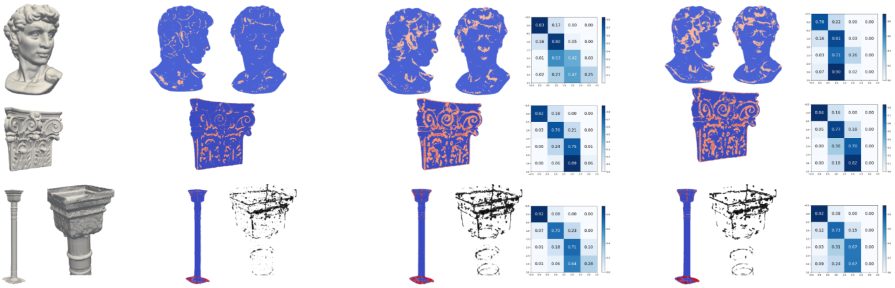


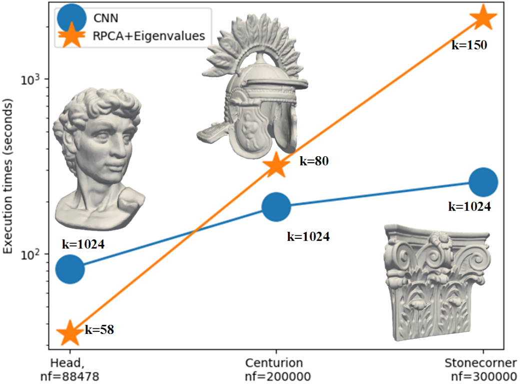
Table of Contents

- Input data
  - Extract patches of face normals from 3D scanned models
- Groundtruth labels
  - Generated from sparse modelling approach
- Training
  - Labeled input patches
  - Softmax cross entropy with logits
  - ADAM Optimizer, $lr = 1e - 4$
  - NVIDIA GeForce GTX 1080 graphics card with 8GB VRAM and compute capability 6.1
- Evaluation
  - CPU only setup, Intel(R) Core(TM) i7-4790 CPU @ 3.60Hz with 32GB of RAM
- Code
  - https://github.com/snousias/mesh-saliency-detection-using-convolutional-neural-networks



(a) (b) (c) (d)

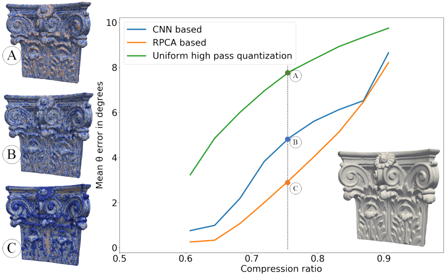
- Saliency maps of 3D scanned geometries.
  - (a) 3D mesh geometry,
  - (b) Sparse modelling based groundtruth estimation,
  - (c) CNN based saliency map extraction employing Hilbert curve arrangement with corresponding confusion matrices
  - (d) CNN based saliency map extraction employing simple reshape based arrangement with corresponding confusion matrices.



- Traditional and CNN-based approaches exhibit similar accuracy
- The CNN approach is much faster in larger and models with higher density

Experimental evaluation

Compression

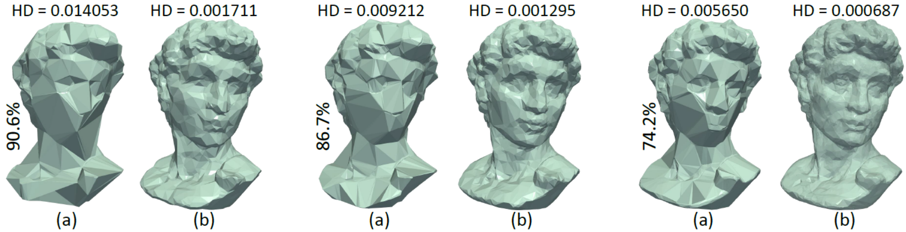


Mean θ error as a function of compression ratio and θ error visualization. Red colors correspond to higher θ.

- Use the saliency map result to order the delta coordinates as follows $\mathcal{D}' = \{d \in \mathcal{D} | v_c = 4\}\{d \in \mathcal{D} | v_c = 3\}\{d \in \mathcal{D} | v_c = 2\}\{d \in \mathcal{D} | v_c = 1\}$
  - $\mathcal{D}$ is initial set of delta coordinates
  - $v_c = 1$ to $v_c = 4$ is the predicted class
- Aris S Lalos, Gerasimos Arvanitis, Aristotelis Spathis-Papadiotis, and Konstantinos Moustakas,"Feature aware 3d mesh compression using robust principal component analysis," in2018 IEEE International Conference on Multimedia and Expo (ICME). IEEE, 2018, pp.1–6.

Experimental evaluation

Simplification



Reconstructed results for different simplification scenarios, using: (a) the feature extraction method, described in [1], (b) our proposed method.

- **[1]** Aris S Lalos, Gerasimos Arvanitis, Aristotelis SpathisPapadiotis and Konstantinos Moustakas, "Feature aware 3d mesh compression using robust principal component analysis", 2018 IEEE International Conference on Multimedia and Expo (ICME), pp. 1-6, 2018.

Table of Contents

Conclusions

- In this work, we presented a CNN-based 3D mesh saliency extraction approach, formulating a training set from spectral and geometrical 3D mesh processing.
- We use the outcome of these methods in order to train a CNN model which can be used for the fast and accurate saliency map extraction of very dense 3D scanned models.
- Our aim is to provide a fast 3D saliency mapping method which could be beneficially used by applications that required the use of very dense 3D models and very fast processing.
- Extensive evaluation studies carried out, utilized a variety of evaluation scenarios, including heatmap visualization for visual perception, simplification and compression use cases.

This work has been partially supported by the EU research and innovation program Horizon2020 Marie Sklodowska WARMEST:loW Altitude Remote sensing for the Monitoring of the state of cultural hEritage Sites: building an inTegrated model for maintenance, under Grant Agreement No. 777981 and the EU research and innovation program Horizon2020 CPSoSaware under Grant Agreement No. 871738.

Thank you for your attention

Contact: lalos@isi.gr

Mesh saliency detection using Convolutional Neural Networks

Stavros Nousias*, Gerasimos Arvanitis, Aris S. Lalos and Konstantinos Moustakas

Industrial Systems Institute, Athena Research Center
Department of Electrical & Computer Engineering, University of Patras, Greece

June 18, 2020