International Cartographic Association
Association Cartographique Internationale

Inferring Implicit 3D Representations from Human Figures on Pictorial Maps

Raimund Schnürer [a, *], A. Cengiz Öztireli [b], Magnus Heitzler [a], René Sieber [a], Lorenz Hurni [a]

[a] *Institute of Cartography and Geoinformation, ETH Zurich, {schnuerer|hmagnus|sieberr|lhurni}@ethz.ch*
[b] *Department of Computer Science and Technology, University of Cambridge, aco41@cam.ac.uk*

* Corresponding author

Keywords: Artificial Neural Networks, Single-view 3D Reconstruction, Pictorial Maps, Human Parsing

Abstract:

Human figures frequently occur on pictorial maps besides other illustrative entities. In this work, we present how to automatically derive 3D depictions from these 2D human figures. Previous research has shown that silhouettes, body parts, and joints of 2D human figures in common poses can be detected on pictorial maps by artificial neural networks (Schnürer et al., 2019). Architectures for these networks have been also developed to reconstruct 3D models of real persons from photos in good accuracy (Varol et al., 2018). Single-view methods are particularly suited for our use case since pictorial figures are usually drawn from one perspective only. Furthermore, a trend can be observed to represent the recovered 3D models by implicit surfaces, expressed by level sets of functions (Saito et al., 2019) or signed distance functions (Wang et al., 2019). Compared to other 3D structures, implicit geometries are memory-efficient, but they require special ray tracing algorithms like marching cubes or sphere tracing to be rendered.

We examine two approaches: (1) A convolutional neural network, consisting of a feature extractor and a head network, shall learn to directly predict body parts and joints of a 3D model from a 2D image. For this approach, a large amount of training data is essential, for instance, body scans from real persons (e.g. Human3.6M[1]) or synthetically created persons (e.g. SURREAL[2]). For our case, these 3D models may be additionally distorted or enriched by rigged human characters from computer games. After converting the geometries from explicit into implicit forms (e.g. mesh-to-sdf[3]), the network is trained to estimate the resulting values of sample points. (2) Implicit function parameters can be stepwise optimized, for example by Stochastic Gradient Descent, to reduce differences between the target image and its approximation. The latter is a projection of 3D primitives which are combined, transformed, morphed, or deformed by mathematical operations (Pasko et al., 1995). This approach facilitates to formulate constraints such as the connectivity of body parts or rotation angles of joints, but it requires more iterations and eventually ends in a local minimum.

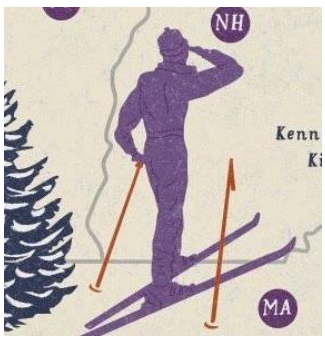
The following challenges exist for both approaches: Due to occlusions, multiple reconstruction outputs are plausible. Perhaps, a generative model such as a variational autoencoder or generative adversarial network needs to be introduced to reflect the variety of poses by latent codes. Moreover, a certain strategy may be pursued to sample equally points near the surface, within the body, and in the surrounding space so that local details and thin parts (e.g. fingers) can be preserved (Paschalidou et al., 2020). To speed up the training or optimization process, possibly a meta-learning algorithm may help to find good initialization parameters (Sitzmann et al., 2020). Since human figures on maps are mostly hand-drawn or manually created with graphic software, the camera perspective or lighting conditions may not be fully consistent. It is not clear yet whether this has an impact on differentiable rendering methods (Niemeyer et al., 2020), which may be applied in our networks. Lastly, the texture needs to be mapped to the 3D model and estimated for the hidden parts, which can be achieved by a subnetwork (Saito et al., 2019).

We will evaluate the two approaches according to their effectiveness and efficiency. Based on the outcomes of related works and the proposed methods to overcome the challenges, we are optimistic to create meaningful representations. When being successful, the inferred 3D figures could emerge from the original map by augmented reality devices. The figures could then be animated and act as guides on touristic maps or storytellers on historic maps in museums. Due to their attractiveness, the generated 3D figures may raise the interest of people, especially children, in maps and may also serve educative purposes.

[1] http://vision.imar.ro/human3.6m/
[2] https://www.di.ens.fr/willow/research/surreal/data/
[3] https://github.com/marian42/mesh_to_sdf

https://doi.org/10.5194/ica-abs-3-262-2021

Figure 1. Example of a pictorial map[4]



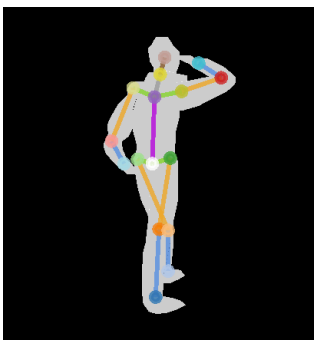


Figure 2. A pictorial human figure[5], its silhouette and pose keypoints, and a manual 3D reconstruction[6]

Figure 3. Split 3D body parts[6]

[4] https://i.pinimg.com/736x/0f/77/85/0f77857d2a528bf65ff6639874b03b3e--italy-illustration-map-illustrations.jpg
[5] https://i.pinimg.com/736x/02/29/4f/02294fff4335751b88fe57b87cc03dd9--illustrated-maps-map-design.jpg
[6] https://smpl-x.is.tue.mpg.de

https://doi.org/10.5194/ica-abs-3-262-2021