

Abstract

Aspects concern a method for estimating the shape and pose of a body comprising obtaining an image of the body, extracting image features from the image, and transforming the image features to a set of parameter values describing the shape and pose of the body by means of a transformer comprising a cascade of transformer units, wherein each transformer unit operates on a representation of a respective version of the set of parameter values and comprises a cross-attention layer modifying the representation of the respective version of the set of parameter values by the image features weighted by an attention of the respective version of the set of parameter values to the image features in sequence with a self-attention layer modifying the representation of the respective version of the set of parameter values by components of the representation of the respective version of the set of parameter values weighted by an attention between the components of the representation of the respective version of the set of parameter values to themselves.

FIG. 6