Novel-View Human Action Synthesis Supplementary

Mohamed Ilyes Lakhal[1][0000−0003−4432−9740], Davide Boscaini[2][0000−0003−4887−2038], Fabio Poiesi[2][0000−0002−9769−1279], Oswald Lanz[2][0000−0003−4793−4276], and Andrea Cavallaro[1][0000−0001−5086−7858]

[1] CIS, Queen Mary University of London, London, UK
[2] TeV, Fondazione Bruno Kessler, Trento, Italy
{m.i.lakhal,a.cavallaro}@qmul.ac.uk, {dboscaini,poiesi,lanz}@fbk.eu

1 Texture transfer

In Fig. 1, we show the map **F** computed for different body poses. We can observe that with the geodesic distance we obtain a plausible transfer (see Fig. 2). The geodesic distance encourages pixel transfer within neighboring body regions. The details of each step are presented in Alg. 1.

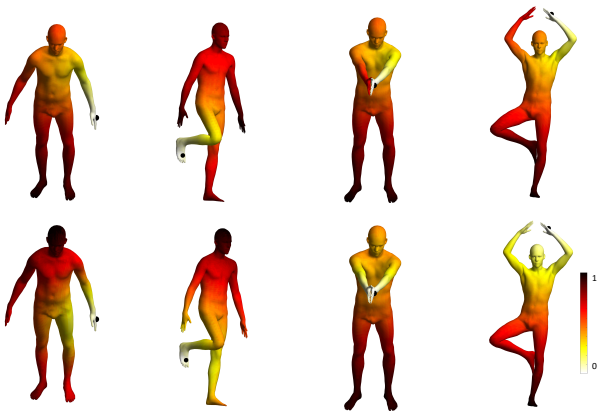

Fig. 1: Illustration of distance between a face in the mesh (black dot) with the rest of the faces. (top row): Geodesic distances; (bottom row): Euclidean distances. White represents small distance while red means large distance.

Algorithm 1 Symmetric texture transfer

Inputs: $n < N_f$ nearest neighbor value; input-view $x^i = \{x_t^i \in \mathbb{R}^{w \times h \times 3}\}_{t=1}^{T}$; time step t; face map $F_t^v \in \mathbb{R}^{w \times h}$ of view $v \in \{i, j\}$; $\mathbf{F} \in \mathbb{R}^{N_f \times N_f}$ pairwise face distances; `face_to_label` dictionary; `sym_parts` hashmap of symmetric body parts.
Output: Rough texture of the target view $\mathcal{T}_{i \to j}^{s}$

function STEP_I
  $\mathbf{F}_{\texttt{dict}} \leftarrow \{\}$
  for $t \in \{1, \ldots, T\}$ **do**
    for $F_{\texttt{id}}, (u_x, u_y) \in F_t^i$ **do**
      $\mathbf{F}_{\texttt{dict}}[F_{\texttt{id}}]$.append$(x_t^i[u_x, u_y])$
    end for
  end for
  for $F_{\texttt{id}} \in \mathbf{F}_{\texttt{dict}}$.keys() **do**
    $\mathbf{F}_{\texttt{dict}}[F_{\texttt{id}}] \leftarrow \texttt{median}\{\mathbf{F}_{\texttt{dict}}[F_{\texttt{id}}]\}$
  end for
end function

function STEP_III
  $\mathcal{T}_{i \to j}^{s}$, d_pix, $O_{xy} \leftarrow$ step_II()
  for $(u_x, u_y) \in O_{xy}$ **do**
    label $\leftarrow$ face_to_label$[F_t^i[u_x, u_y]]$
    if label $\notin$ d_pix.keys() **then**
      label $\leftarrow$ sym_parts[label]
      if label $\notin$ d_pix.keys() **then**
        $\mathcal{T}_{i \to j}^{s}[u_x, u_y] \leftarrow [\#, \#, \#]$
      end if
    end if
    $v_q \leftarrow [u_x, u_y]$
    $v_{pix}, v_{xy} \leftarrow$ d_pix[label]
    $d_L \leftarrow L_2(v_q, v_{xy})$
    $\mathcal{T}_{i \to j}^{s}[u_x, u_y] \leftarrow v_{pix}[\texttt{argmin}(d_L)]$
  end for
end function

function STEP_II
  $\mathcal{T}_{i \to j} \leftarrow 0_{w \times h \times 3}$
  $\mathbf{F}_{\texttt{dict}} \leftarrow$ step_I$(x^i; F^i)$
  d_pix $\leftarrow \{\}$ ▷ label to pixel
  $O_{xy} \leftarrow []$
  for $F_{\texttt{id}}, (u_x, u_y) \in F_t^j$ **do**
    if $F_{\texttt{id}} \in \mathbf{F}_{\texttt{dict}}$.keys() **then**
      $\mathcal{T}_{i \to j}[u_x, u_y] \leftarrow \mathbf{F}_{\texttt{dict}}[F_{\texttt{id}}]$
      label $\leftarrow$ face_to_label$[F_{\texttt{id}}]$
      $F_n \leftarrow [\mathbf{F}_{\texttt{dict}}[F_{\texttt{id}}]; (u_x, u_y)]$
      d_pix[label] $\leftarrow$ d_pix[label] $\cup F_n$
    else
      occluded $\leftarrow$ True
      for $l \in \{1, \ldots, n\}$ **do**
        $F_l \leftarrow \mathbf{F}[l_{F_{\texttt{id}}}][l]$
        if $F_l \in \mathbf{F}_{\texttt{dict}}$.keys() **then**
          $\mathcal{T}_{i \to j}[u_x, u_y] \leftarrow \mathbf{F}_{\texttt{dict}}[F_l]$
          occluded $\leftarrow$ False
          label $\leftarrow$ face_to_label$[F_l]$
          $F_n \leftarrow [\mathbf{F}_{\texttt{dict}}[F_l]; (u_x, u_y)]$
          d_pix[label] $\leftarrow$ d_pix[label] $\cup F_n$
        end if
      end for
      if occluded **then**
        $O_{xy} \leftarrow O_{xy} \cup (u_x, u_y)$
      end if
    end if
  end for
end function

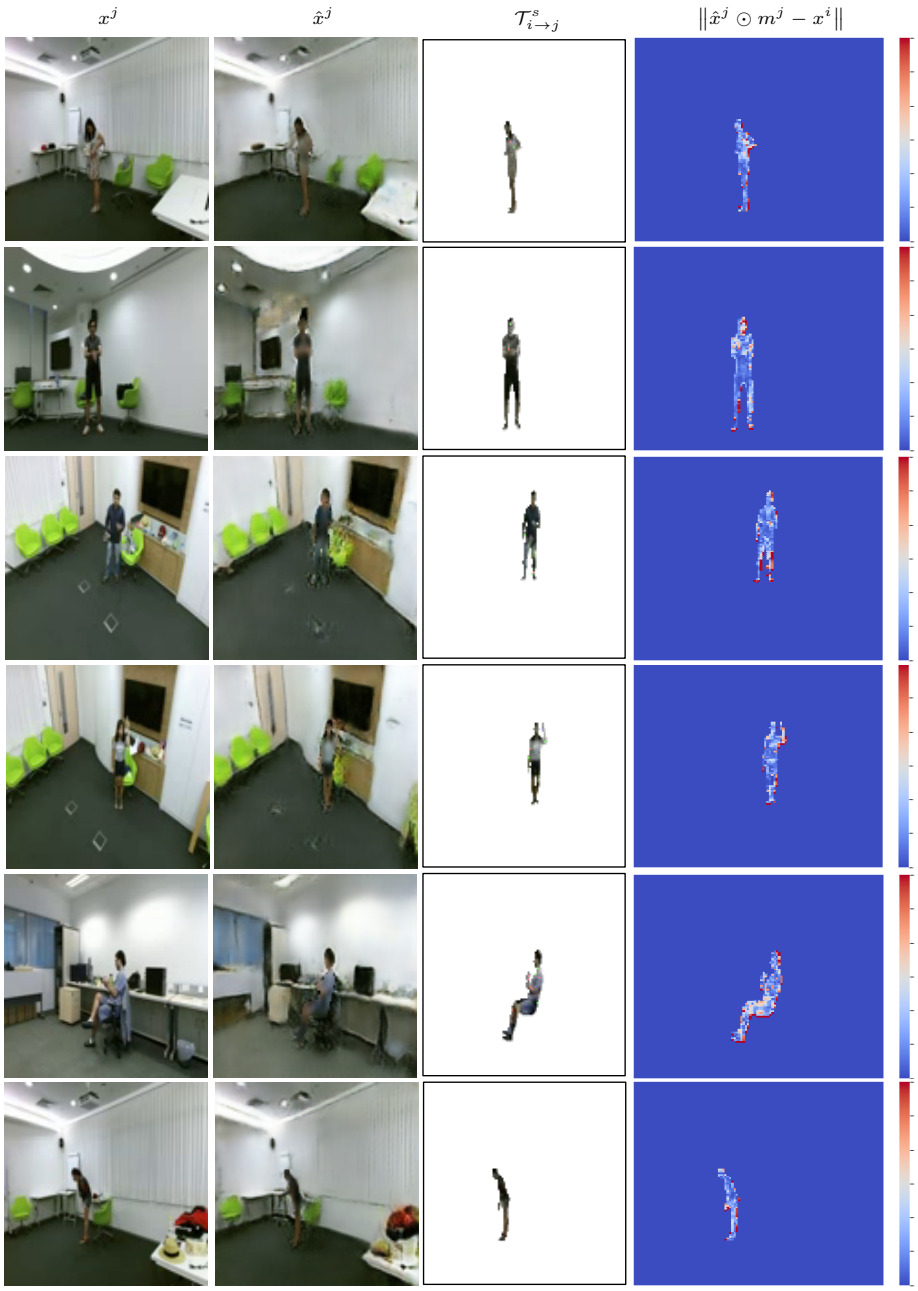

Fig. 2: Additional results of the learned residual of $\mathcal{T}^s_{i\to j}$ using GTNet.