A Discriminative Parts Based Model Approach for Fiducial Points Free and Shape Constrained Head Pose Normalisation In The Wild

Abhinav Dhall[1] Karan Sikka[2] Gwen Littlewort[2] Roland Goecke[3,1] Marian Bartlett[2]
[1]iHCC, Australian National University, Australia
[2]Machine Perception Laboratory, University of California San Diego
[3]Vision & Sensing Group, HCC Lab, University of Canberra, Australia
abhinav.dhall@anu.edu.au, gwen@mplab.ucsd.edu, roland.goecke@ieee.org, {ksikka, mbartlett}@ucsd.edu

Continuous Confidence Map Based Normalisation: While continuous head pose normalisation is not the goal of this paper, we demonstrate as a proof of concept that it is possible to extend the current method for continuous head pose normalisation. For dealing with faces in videos [1], continuous head pose normalisation is required. [2] argue that the appearance of a part does not changes with a subtle pose change, therefore a detector for part i in pose angle p can be shared for the same part i for a pose angle $p + \delta$. Further experiments in [2] showed that sharing based models and independent model have comparable performance. However, sharing based models are faster upto ten times as compared to the independent models [2]. The confidence maps based methods (CM-HPN$_{PS}$ and CM-HPN$_{PI}$) can be extended from discrete to continuous by sharing part-specific regression models R, which are shared among neighboring pose angles.

For a video V with n frames, the first frame's head pose p and frontal reconstructed points are computed using Algorithm 2 (Same Algorithm 2 in main paper). For the consecutive frames, the part-wise regression models specific to the pose computed for first frame are used for normalisation (Algorithm 3). If the current frame's head pose $p + \delta$ has a large difference from the current p, it may lead to a poor facial landmark reconstruction (based on low $Score^*_f$, which if less than a threshold will discard the reconstruction). Based on this the current frame's head pose p and normalised frontal points are re-computed using 2 and the new value of head pose p is used for upcoming frames. Algorithm 1 defines the process.

Algorithm 1: Continuous head pose normalisation

Require: Video V.
  $p, L^*_f = Algorithm2(V_1, p)$ {For the first frame}
  for $i \leftarrow 2$ to $lenght(V)$ **do**
    $Score_p, L^*_f = Algorithm3(V_i, p)$
    if $L^*_f == null$ **then**
      $p, L^*_f = Algorithm2(V_i, p)$ {Update head pose p}
    end if
  end for

Algorithm 2: Pose Invariant Confidence Map Regression

Input: Frame I
Output: Pose p and frontal parts configurations L^*_f
for *pose* $p \in P$ **do**
  $Score_p, L^p_f = Algorithm3(I, p)$
end
$L^*_f = L^p_f$ for the largest $Score_p$ return p with the largest $Score_p$

References

[1] A. Dhall, R. Goecke, S. Lucey, and T. Gedeon. A semi-automatic method for collecting richly labelled large facial expression databases from movies. *IEEE Multimedia*, 2012.

[2] X. Zhu and D. Ramanan. Face detection, pose estimation, and landmark localization in the wild. In *CVPR*, pages 2879–2886, 2012.

Algorithm 3: Frontal Virtual Points Reconstruction using Pose-Specific Confidence Map Regression (Same as Algorithm 1 in the main paper)

Input: Image I and pose p
Output: $Score'$ and L_f^*
for *part* $i \in V$ **do**
  Compute part wise confidence maps,
  $C_i^p = \theta(I, i, p)$ (Eq. 5)
  Divide C_i^p into k blocks B
  $B = \{B_1^p B_2^p ....B_k^p\}$
  for *a = 1 : k* **do**
    Reconstruct $B_a^f \leftarrow B_a^p$ using corresponding model from $\mathcal{R}_i$
  end
  Rejoin reconstructed blocks
  $C_i^f \leftarrow \{B_1^f, B_2^f ...B_k^f\}$
end
$FC \in \sum_{i \in V} C_i^f$
Compute frontal $Shape_f(L)$ and maximise $Score'$ (Eq. 6)
$L_f^* = \max_L(Score'(I, L, p)$