SingleF

PSNR 49.35 0.9919

PSNR 49.35 0.9919

stills cameras (> 8k) to capture the person in a small volume (< 8m³). The 3D model reconstructed with these static HR images captures fine details in both shape and appearance. In this paper we propose a practical solution that enhances the appearance of a low-resolution (LR) dynamic video performance capture acquired with a sparse set of cameras through super-resolution (SR) appearance transfer from the same human subject acquired with DSLR cameras used for static 3D reconstruction.

2 Method

We propose FSTD, ‘From Static to Dynamic’, a pipeline (shown in Figure 1) that performs local super-resolution and global colour correction of a 4D dynamic performance reconstruction from multi-view video using HR static capture from multiple DSLR cameras. The input of our pipeline is a 4D video sequence of a person as well as a set of K HR images \( (I_{HR})^K_{i=1} \) of the same person acquired with static cameras. The 4D video sequence consists of M LR frames \( (I_f)_{i=1}^M \), reconstructed meshes and texture maps. No geometric information of the HR capture is used.

LR-HR image couple identification: After foreground-background separation, pairs of reference images \( (I_{HR}) \) and target frames \( (I_f) \) with similar content are identified for the colour transfer. To identify the couples, we propose a new automatic method that computes the similarity between them in the texture map domain instead of using the image domain due to the different acquisition systems. We first apply Densepose [1] to reconstruct partial texture maps of the frames \( T_p(I_f) \) and of the HR images \( T_p(I_{HR}) \). These partial texture maps are invariant to the camera orientation and position. Similarity between partial texture maps \( T_p(I_f) \) and \( T_p(I_{HR}) \) is evaluated using the SSIM metric. A LR frame \( I_f \) is coupled with the HR image \( I_{HR} \) whose partial texture map is the most similar defined by the SSIM metric:

\[
I^p_f \leftrightarrow I^p_{HR} \text{ where } \arg \max \{ \text{SSIM}(T_p(I^p_f), T_p(I^p_{HR})) \} \quad (1)
\]

where \( I^p_f \) is the i-th frame of the video camera j and \( I^p_{HR} \) is the HR image of the camera q.

Colour transfer: Once the couples are identified, we select those whose \( I_f \) corresponds to the first frame of LR cameras where the models are in a T Pose. The image colour transfer approach [2] is extended to multi-view images as input to learn a colour transfer function \( \phi(x) \) from HR to LR images. The function \( \phi(x) \) is modelled as a Thin Plate Spline that depends on a set of parameters \( \Theta \). This set is computed by minimizing the following energy function with a gradient descent algorithm:

\[
\theta = \frac{1}{N} \sum_{j=1}^{N} \arg \min_{\theta_l} \{ ||p_f|^{2} - 2 < p_f/p_l > \} \quad (2)
\]

where \( N \) is the number of input couples, \( p_f \) is the Gaussian distribution of the colours of \( I_f \) with parameterised mean \( \phi_{

3 Results

The qualitative results obtained applying FSTD show improvements in the quality of the appearance of dynamic performances. In particular, Figure 2 shows that the colour transfer stage enhances the global appearance of the reconstructed model by colour correction to match the HR image colour distribution. Figure 3 illustrates how the SR enhances the fine detail appearance of the low-resolution 4D model. Table 1 shows that our training approach achieves higher values of PSNR and SSIM for two performers compared to the original RCAN.


