Improved Interactive Reshaping of Humans in Images

Nidhi Arora
IIT Delhi
India
nidhi@cse.iitd.ac.in

Harsh Kumar
IIT Delhi
India
harsh23031989@gmail.com

Jagjeet S. Dhaliwal
IIT Delhi
India
cs5080212@cse.iitd.ac.in

Prem Kalra
IIT Delhi
India
pkalra@cse.iitd.ac.in

Parag Chaudhuri
IIT Bombay
India
paragc@cse.iitb.ac.in

ABSTRACT

In this paper, we present an interactive and flexible approach for realistic reshaping of human bodies in a single image. For reshaping, a user specifies a set of semantic attributes like weight and height. Then we use a 3D-morphable model based image retouching technique for global reshaping of the human bodies in the image such that they satisfy the semantic constraints specified by the user. We address the problem of deformation of the environment surrounding the human body being reshaped, which produces visible artifacts, especially noticeable at regions with structural features, in prior work. We are able to separate the human figure from the background. This allows us to reshape the figure, while preserving the background. Missing regions in the background are inpainted in a manner that maintains structural details. We also provide a quantitative measure for distortion and compare our results with the prior work.

Keywords

image retouching, image manipulation, warping, deformation, reshaping, inpainting.

1 INTRODUCTION

Professional image editing packages like Adobe Photoshop often limit themselves to local modification of the image whereas retouching tasks such as increasing or decreasing the height or weight of the human bodies in the images require global consistency to be maintained across the human body. This requires professional skills and novice users find it hard to achieve such consistency. Hence, one comes across many such images where the use of Photoshop is quite evident.

Zhou et al. [ZH10] propose an interactive approach based on a 3D-morphable model to deform individual parts of the body to achieve global editing consistency and desired spatially-varying deformation within and across individual body parts. The approach doesn't rely on 3D-reconstruction of the human body from the image but applies a body-aware warping to the image for reshaping human bodies in a single image. The method comprises of the following steps:

- A morphable 3D-model is roughly matched to the human body in the image by user interaction.
- The 3D-model is deformed by changing the semantic attributes like height and weight.
- The image is then warped such that the human body in the image reflects the changes in the matched 3D-model by following the changes in the 2D-contours of the body with respect to the pose of the human in the image.

The above method is effective for human bodies with a variety of poses, shapes and in presence of loose clothing. It, however, introduces artifacts in the background of the image, noticeable especially in the structural features like walls, floor and other features. Consider, for example Figure 1, where the height of the person is decreased. We can observe that artifacts are introduced in the horizontal structure and the shadow of the person as highlighted in Figure 1(b). We address this problem of deformation of the background of the image by separating the warping of the human body from the background of the image (refer Figure 1(c)). Our solution relies on inpainting of the background image after delineating the human shape that needs a change. We use optical flow as a quantitative measure to obtain the distortion in the background and show that our approach has a lower distortion. The separation of the image into foreground with the human shape and background offers us flexibility to even change the pose of the human body in a limited way.

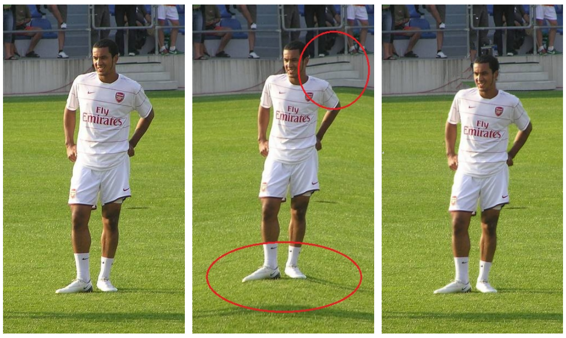

(a) (b) (c)

Figure 1: Structural artifacts introduced in the image using [ZH10]; (a) Input image, (b) Resulting image with decreased height and noticeable deformation of the background, (c) Image obtained using our approach.

The paper is organized as follows. In Section 2, some previous research carried out in this area and its allied fields is discussed. Section 3 provides a brief overview of the approach followed by Zhou et al. [ZH10] and highlights its shortcomings. Section 4 describes our proposed approach. Section 5 presents some of the results followed by Section 6 that concludes our work.

2 RELATED WORK

This work is related to a variety of fields. We briefly describe the state-of-the-art techniques for the relevant domains. Since our work is based on the method proposed by Zhou et al. [ZH10], some of the related work cited in that paper is revisited in this section. In addition, we also look at other relevant literature.

2.1 Image Retouching, Warping and Resizing

Retouching images incorporates the use of several image editing tools, for e.g., tone adjustment, recoloring, image composition, image repairing, image denoising, image warping etc. Most existing retouching tools operate at the pixel level and are effective for low level editing tasks [EP08]. However, they are not suitable for high-level editing tasks because they involve user interaction for maintaining the coherence of editing operations [AC02]. Image warping and resizing methods like Moving Least Squares (MLS) [SM06] and Radial Basis Function (RBF) [AR95] propagate the changes from the control handle to the rest of the image. Zhou et al. [ZH10] propose a warping approach for parametric reshaping of humans in images to adhere the human shape to the specified semantic attributes by resizing the human body along and orthogonal to skeletal bones to achieve a global consistency.

2.2 3D-Morphable Models with Pose Fitting and Shape Selection

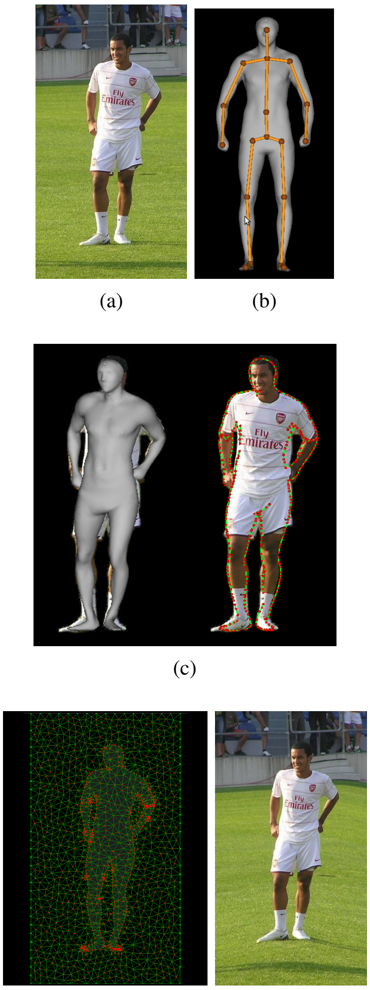
SCAPE [AS05] encodes the pose and shape of the human body separately. The pose is stored with the help of an underlying skeleton while the shape of the person is encoded using variational methods or envelope skinning. The outcome of the two steps is combined for the generation of new shapes. Hasler et al. [HS09] introduce a model which encodes both the shape and pose of the model using a translation and rotation invariant encoding. Many automatic pose estimation methods from a single image have been proposed (see [HW09] and references therein). However, it requires a certain amount of user assistance to obtain more reasonable poses ([DA03], [HK07], [CT12], [TM11]). Richter et al. [RV12] present a system for real-time deformation of the shape and appearance of people who are standing in front of a depth+RGB camera, such as the Microsoft Kinect.

In the next section, we give a brief overview of the method proposed by Zhou et al. [ZH10] and outline the limitations of this method.

3 PARAMETRIC RESHAPING OF HUMAN BODIES

Changing the semantic attributes of the human body in the image requires global consistency to be maintained in and across the individual body parts to produce visually pleasing results. A morphable 3D-model of the human body is used to guide the global reshaping of the human body in the image.

The method (refer to Figure 2) requires matching a 3D-morphable model, by adjusting its pose and shape parameters to the human body in the image. The pose and shape parameters of the morphable model are taken as user input for defining the pose and the semantic attributes for the desired change of shape. After matching the 3D-morphable model to the human body in the image, the model is morphed to change the semantic attributes according to the user input. The changes in the fitted model with respect to its 3D-skeleton are used to guide the image warping especially at the projected contours of the fitted model. A body-aware image warping coherently resizes the body parts along the direction parallel and perpendicular to the bone axes of the 3D-skeleton to incorporate the length changes in the corresponding directions. Thus, the human body is parametrically reshaped with the help of user interaction.


(a) (b)

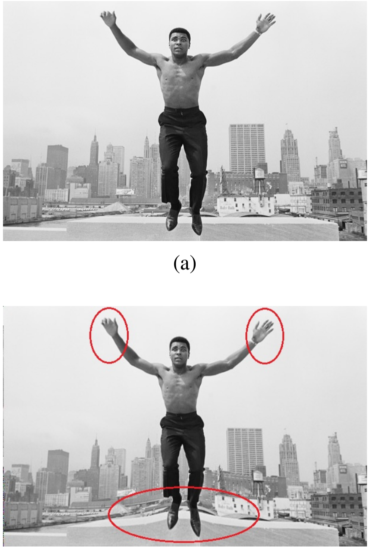
(c)

(d) (e)

Figure 2: Overview of the parametric reshaping using 3D-morphable model [ZH10]: (a) Original image, (b) mesh rigged with skeleton, (c) pose and shape deformation to roughly match the human body in the image and the image contour, (d) image embedded into 2D-triangular mesh using image subject contour, (e) model is deformed to adhere to new semantic values, changes in the 3D-morphable model are used to drive image warping.

However, we observe that the above approach causes artifacts to the surrounding environment of the human body that is reshaped. For example, in Figure 1, deformation of the background (especially the structural artifacts like stairs and the railings of the stadium and the shadow of the human) is quite noticeable and hence, the resultant image shows visible artifacts when the height of the person is decreased.

Consider Figure 3, the height and weight of Muhammad Ali has been decreased in the resultant image. The unnatural elongation of the hands and the feet of Muhammad Ali as highlighted in the figure occurs because of the proximity of the hands and feet to the border of the image which remains unchanged.


(a)

(b)

Figure 3: (a) Original image, (b) Unnatural elongation of hands and feet of the person is visible in the output image.

In the next section, we address the above mentioned shortcomings and improve this technique.

4 PROPOSED APPROACH

As discussed in the previous section, the above mentioned technique introduces artifacts in the resultant image and hence, the results are not visually pleasing under moderate to severe deformations even though they achieve the required change in the semantic attributes of the human. It is observed that pinning of the image boundaries results in unnatural elongation of the body parts close to the boundaries. Pinning leads to artifacts due to the deformation of triangles near the image boundaries when the human body undergoes deformation according to the changes in the 3D-morphable model. One can pad the image while keeping the vertices on the boundary of the padded image fixed or pad the image and relax the constraints at the boundary of the padded image. This, however, gives only limited improvement.

We observe that the body-aware warping of the whole image results in structural artifacts as the changes in the human body are propagated to the background. These structural artifacts can be removed by separating the warping of the human body from the background of the image by using the proposed technique consisting of the following steps:

- The human bodies, to be warped, are segmented from the image as foreground using GrabCut [RK04].

- The hard segmentation, thus, obtained is used as input for extracting the foreground color, background color and alpha value for each pixel of the image.
- Body-aware image warping, based on the parametric deformation of the 3D-morphable model, is then applied to the foreground image and the alpha matte.
- The hole generated in the background is inpainted using a suitable technique.
- The resulting image is then, generated by combining the warped foreground image and the inpainted background using the warped alpha values for each pixel.

4.1 Segmentation of the Human Body

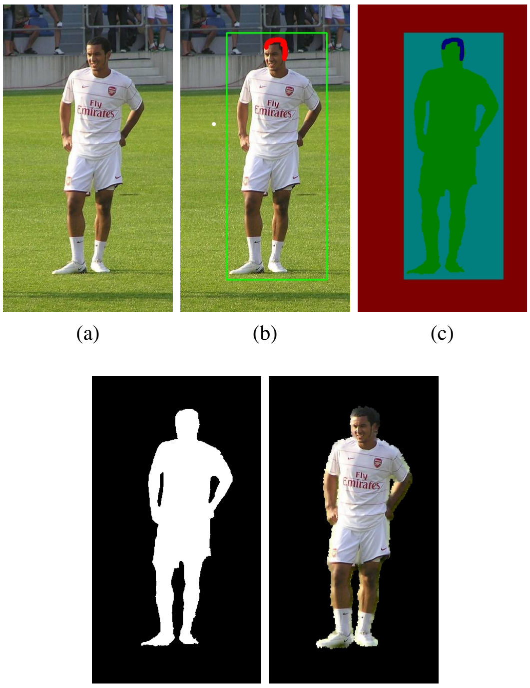

(a) (b) (c)
(d) (e)

Figure 4: Segmentation of the human body from the image using GrabCut: (a) Input image, (b) Scribbles provided to drive the segmentation, (c) Regions formed by GrabCut: red, blue, green and dark blue represent background, probable background, probable foreground and foreground respectively, (d) Hard segmentation mask obtained by combining foreground and probable foreground regions, and (e) Human body segmented from the image using hard segmentation mask.

The human body to be warped is segmented from the image using GrabCut [RK04] where the user interaction is limited to drawing a rectangle to mark the background and probable foreground area. GrabCut [RK04] uses an iterative minimization graph-cut algorithm using Gaussian Mixture Models to segment the foreground and reduces the amount of the user interaction. If the results are not approximately correct, more information is provided using scribbles, marking the areas of the image as background and foreground and running the iterative mechanism once again till the segmentation of the human body from the image is approximately correct. The foreground scribbles and the probable foreground are then used to form the hard segmentation of the human body that is to be warped (refer Figure 4). Many other methods exist for segmenting the desired foreground from the image. Magic Wand tool uses the user input to form color statistics from the specified region and computes a connecting region of pixels such that the selected pixels fall within some adjustable tolerance of the color statistics. We have found GrabCut [RK04] suitable for our purpose.

4.2 Image Matting


(a) (b) (c)

Figure 5: (a) Input image, (b) Alpha matte, (c) Foreground color image.

In our approach, hard segmentation obtained from segmenting the foreground object from the complex background is transformed into a trimap by using erosion on the user input (scribbles) to drive the calculation for the three components in a narrow strip around the hard boundary (refer Figure 5). Thus, we separate the foreground color, background color and alpha value for each pixel of the image. There exist a number of methods for matting [CS01], [SJ04], [LW06] , we have adopted a simple approach and found it adequate for our purpose.

4.3 Image Warping

To warp the image, we follow the same method as proposed by Zhou et al. [ZH10] as described in the previous section. For pose matching, we have found that the data of 3D-morphable models need axis alignment to the coordinate system induced by the image. The mesh needs to be transformed so that it is upright and facing the user. To achieve the desired orientation of the mesh, it is rotated to align the plane of symmetry to the global YZ-plane. To compute the plane of symmetry of the human mesh placed arbitrarily in the space, the original mesh is registered to the triangular mesh obtained by reflecting the original mesh (termed as mirrored mesh) about any arbitrary plane in the space. The original mesh and the mirrored mesh are roughly aligned to each other by performing a coarse registration using the optimal set of Persistent Feature Histograms [AR95]. The axes of the minimum area bounding box of the vertices projected on the plane of symmetry are aligned to the world coordinate axes. For fine alignment of the mesh obtained so far, it is registered against the template mesh using a corresponding point set registration. Now, the mesh obtained from the database is upright and facing the user.

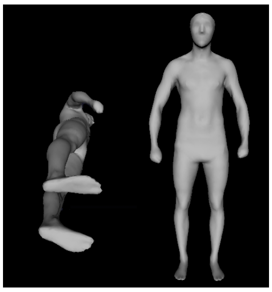
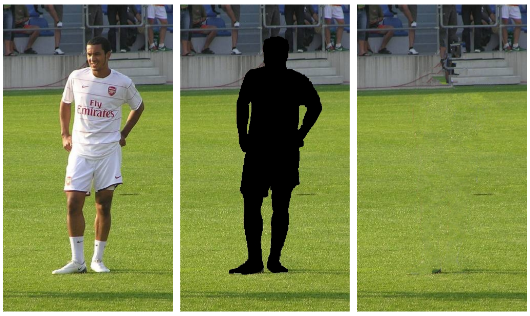


Figure 6: Sample mesh generated from the database using given semantic attributes and the required orientation of the mesh.

4.4 Background Filling

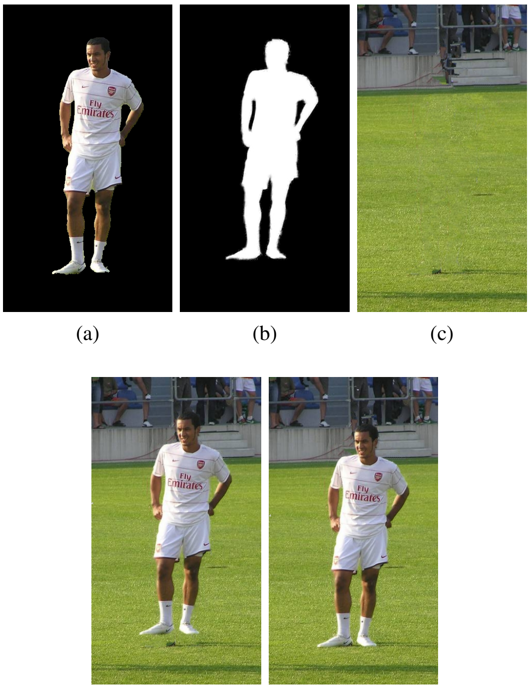


(a) (b) (c)

Figure 7: (a) Input image, (b) Target region for inpainting, (c) Inpainted background image.

Background completion is performed by using an image inpainting technique [CP04]. The target region is specified by creating a mask using alpha values obtained in the previous step. A threshold is chosen on the alpha values such that all the pixels with alpha value more than the threshold are taken to be foreground and hence are removed from the image and form the target region to be inpainted as shown in Figure 7(b). Arora et al. [AK12] propose a method to improve exemplar based inpainting [CP04] by tweaking the values of various parameters like patch size, shape and size of the mask. They use the technique of inpainting for filling in the cracks and restoring old paintings. We have used their approach [AK12] for background completion.

4.5 Remaining Steps



(a) (b) (c)

(d) (e)

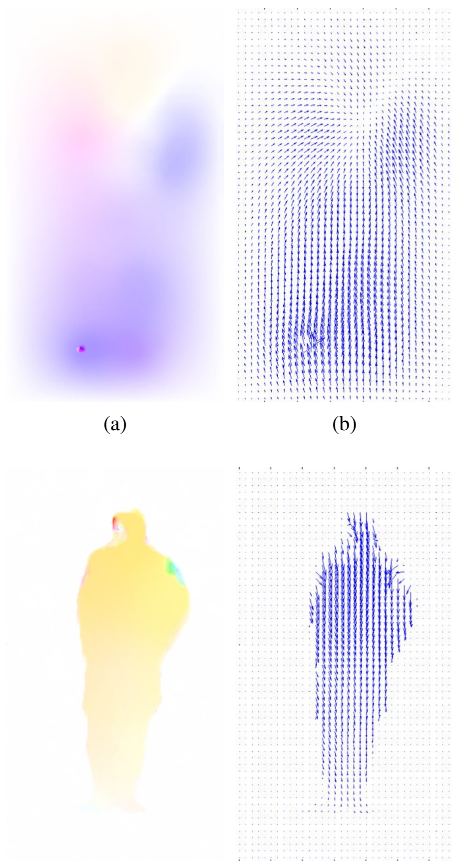
Figure 8: Combining the warped foreground color image and the inpainted background to form the resultant image; (a) Warped foreground color image, (b) Warped alpha matte, (c) Inpainted background color image, (d) Combining the above without any translation of foreground with respect to the background, and (e) Combining the above after translating the foreground image for a more visually-pleasing result.

Thus, we obtain the alpha matte, the foreground image and the inpainted background image by following the above mentioned steps. The 3D-morphable model is roughly matched with the human body in the blended foreground image (foreground image combined with the alpha matte). The padded blended foreground image is embedded into a 2D-triangular mesh using the image contour and the boundaries of the padded image and the changes in the 3D-morhpable model (conforming to the new semantic attribute values) are then used to warp the 2D-triangular mesh to obtain the new 2D-triangular mesh. The new 2D-triangular mesh is used to warp the foreground image and the alpha matte of the segmented human body. The foreground image and the inpainted background image are then blended together to form the resultant image using the alpha matte. The foreground image and the alpha matte sometimes need to be translated to produce more-visually pleasing effects. For example, in Figure 8 the foreground image and the alpha matte need to be translated so that the shadow and the human are in contact with each other and hence, produce a more-visually pleasing and plausible result.

5 RESULTS

Our proposed method is applied on a variety of images of human subjects with various poses and shapes. The human subjects in the image are morphed to reflect the changes in the semantic attributes as desired by the user. Figure 10, Figure 11 and Figure 12 shows the comparison of the results from [ZH10] and our approach. Artifacts are clearly visible in the regions with structural features as highlighted in the images. The environment surrounding the human body being reshaped is noticeably deformed in Zhou et al.[ZH10] results. Our results do not suffer from such problems because of the use of inpainting in our pipeline. In Figure 12, we have also elongated the shadow of the girl along with increasing her height.

We also compare the results in terms of optical flow field, a quantitative measure of distortion. Another measure of distortion is Gradient Vector Flow (GVF) [SB12] but we have used optical flow field for evaluation. In Figure 9, we compare the optical flow field obtained from Zhou et al. result and our result for Figure 1. Figure 9(a) and Figure 9(c) are the color coded output for optical flow. Color coding represents the magnitude of flow field. The range of the colors is from yellow to violet where violet represents large magnitude of the flow and yellow color means a lower value of flow field. Figure 9(b) and Figure 9(d) are the corresponding pictorial representation in the form of arrow diagrams. Table 1 denotes the quantitative measure of the magnitude of optical flow for both the results. Total flow value is the sum total of the magnitude of the optical flow of all the pixels of the image, then we divide it by the total number of pixels in the image to obtain the per-pixel optical flow value. To calculate the flow values for the background pixels in the image, mask is used to separate foreground from background. Same mask is also used to compute the background in Zhou result to have uniformity of comparison. In our result, both the flow value per pixel and the flow value for background is lower than the Zhou et al. result which means that our result has lower distortion and specifically, lower distortion of the background. Similarly, in Figure 10 and Figure 11, comparison in terms of optical flow fields is also shown. Table 2 represents the quantitative comparison of the per-pixel flow values for Fig 10 and Fig 11. Flow values for our results are much lower than Zhou et al. results.


(a) (b)

(c) (d)

Figure 9: Optical flow field obtained for Fig 1: (a), (b) Flow field obtained from Zhou et al. result and (c), (d) Flow field obtained from our result.

Results	Flow value per pixel	Flow value for background(per-pixel)
Zhou et al. result	12.3911	12.8948
Our result	4.1860	0.7214

Table 1: Optical flow values for Fig 9 (Size: 600×331).

5.1 Human Subject with Loose Clothing

We have handled the case of loose clothing in the same way as Zhou et al. [ZH10] have handled. We create a larger mask covering the entire clothing of the person for inpainting the background. Pose fitting and shape selection are performed in the same way as explained in our pipeline. In Figure 13, waist girth of the girl has been decreased.

5.2 Multiple Human Subjects

We address another application in which an image comprise of multiple human subjects. Our objective is to change the semantic attributes of both the subjects. This is done in two passes. We create three masks, two for the individual humans and a combined mask covering both the subjects. We change the semantic attributes of

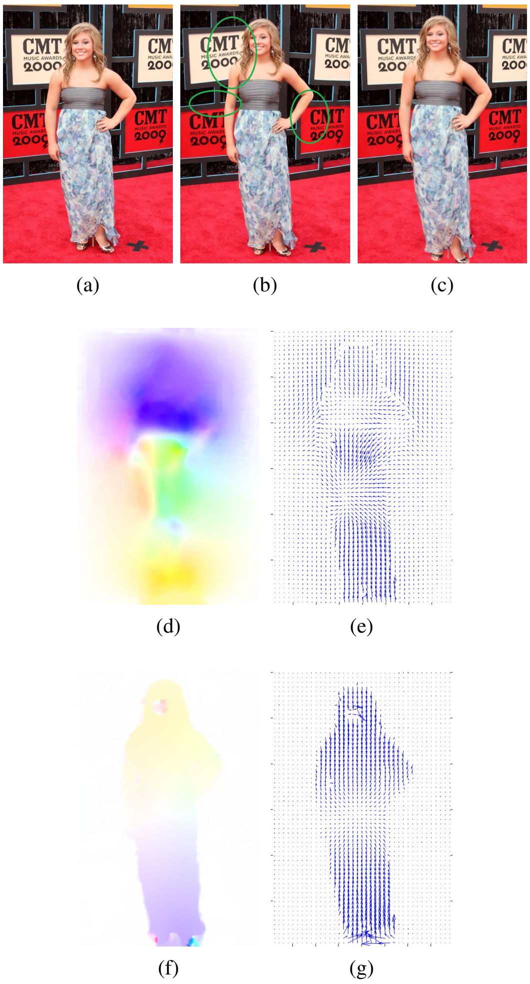

(a) (b) (c) (d) (e) (f) (g)

Figure 10: Comparison of results from Zhou et al. [ZH10] and our approach: (a) Input image, (b) Result using [ZH10] with height increased, (c) Result from our approach, (d), (e) Optical flow field obtained from Zhou et al. result and (f), (g) Optical flow field obtained from our result.

both the subjects individually following the above mentioned pipeline. The mask covering both the humans is used for inpainting the background. We then, paste the reshaped subjects onto this inpainted background. For instance, in Figure 14, height of both the subjects is decreased and in Figure 15, weight is decreased and height is increased of both the subjects.

5.3 Change of Background Setting

We can also change the semantic attributes of a person and place it in a different setting altogether. According to our approach, we segment out the foreground from the background and parametrically reshape the foreground, which basically contains the human body. We then, composite this reshaped subject onto a different background. In Figure 16, we change the height of the character in Figure 16(a) and place her in a different background setting in Figure 16(b). Shadow is cast in the resultant image assuming the position of the light source and is then composited with the background.

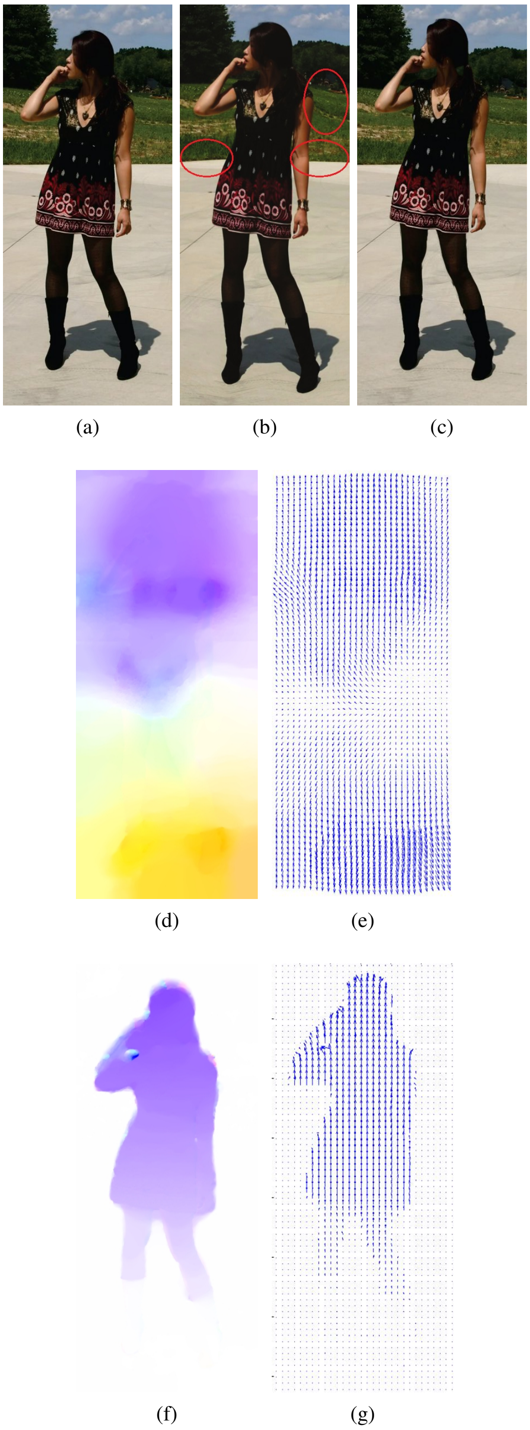

(a) (b) (c) (d) (e) (f) (g)

Figure 11: (a) Original image, (b) Result using [ZH10] with increased height of the human subject and (c) Resulting image with increased height from our approach, (d), (e) Optical flow field obtained from Zhou et al. result and (f), (g) Optical flow field obtained from our result.

Figures	Zhou et al. results(Flow value per pixel)	Our results(Flow value per pixel)
Fig.10(600×393)	7.6880	4.5066
Fig.11(726×307)	10.8101	3.8052

Table 2: Optical flow values for Zhou et al. results and our results for Fig 10 and Fig 11.


(a) (b) (c)

Figure 12: (a) Original image, (b) Result using [ZH10] with increased height of the human subject and (c) Resulting image with increased height from our approach.


(a) (b)

Figure 13: (a) Original image and (b) Resulting image with decreased waist girth of the human subject.


(a) (b)

Figure 14: (a) Original image and (b) Resulting image with decreased height of both the human subjects.


(a) (b)

Figure 15: (a) Original image, (b) Resulting image with decreased weight and increased height of both the human subjects.

Since the method proposed by Zhou et al. [ZH10] does not separate the foreground and background, so, it cannot be used to change the background setting of a human subject.

(a) (b)

Figure 16: (a) Original image, (b) Human subject placed in a different background setting.

6 CONCLUSION AND FUTURE WORK

This paper presents an interactive and flexible approach for realistic reshaping of human bodies in an image. We perform parametric reshaping using a 3D-morphable model to achieve globally consistent manipulation effects. A user specifies a set of semantic attributes like weight, height and others as proposed earlier in 3D-morphable model based image retouching technique [ZH10] for global reshaping of human bodies in an image. We address the problem of deformation of the background of the image because of the propagation of the retouching effects to the background. We follow the approach in which we separate the foreground and background. Foreground is reshaped and background is inpainted maintaining the necessary structural details.

The main contribution of this paper is to combine set of techniques to obtain improved results. The improvement is shown both in terms of visual results and a quantitative measure.

As a future work, we are interested in extending our approach for reshaping subjects in videos.

7 ACKNOWLEDGEMENTS

We would like to thank Nils Hasler [HS09] for providing the database of scanned human meshes. We would like to thank Shizhe Zhou [ZH10] and Ligang Liu [ZH10] for the use of some of the images that we use as input to test our methods. We have taken these images from Zhou et al. [ZH10] paper and their supplementary materials available online.

8 REFERENCES

[AC02] W. A. Barrett and A. S. Cheney, “Object-based image editing,” *ACM Transaction on Graphics, Proceedings of SIGGRAPH 2002*, vol. 21, issue. 3, pp. 777-784, July 2002.

[AS05] D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers and J. Davis, “SCAPE: shape completion and animation of people,” *ACM Transactions on Graphics, Proceedings of SIGGRAPH 2005*, vol. 24, issue. 3, pp. 408-416, July 2005.

[AR95] N. Arad and D. Reisfeld, “Image warping using few anchor points and radial functions,” *Computer Graphics Forum*, vol. 14, pp. 35-46, 1995.

[AK12] N. Arora, A. Kumar and P. Kalra, “Digital restoration of old paintings,” *WSCG International Conference on Computer Graphics, Visualization and Computer Vision 2012*, pp. 347-356, June 2012.

[CT12] T. Chen, P. Tan, L. Q. Ma, M. M. Cheng, A. Shamir and S. M. Hu, “Poseshop: Human image database construction and personalized content synthesis,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, issue. 5, pp. 824-837, May 2013.

[CS01] Y. Chuang, B. Curless, D. Salesin, and R. Szeliski, “A bayesian approach to digital matting,” *in proceedings of IEEE CVPR 2001*, vol. 2, pp. 264-271, Dec. 2001.

[CP04] A. Criminisi, P. Pérez and K. Toyama, “Region filling and object removal by exemplar-based image inpainting,” *IEEE Transactions on Image Processing*, vol. 13, issue. 9, pp. 1200-1212, Sept. 2004.

[DA03] J. Davis, M. Agrawala, E. Chuang, Z. Popovic and D. Salesin, “A sketching interface for articulated figure animation,” *SCA ’03*, pp. 320-328, 2003.

[EP08] K. Eismann and W. Palmer, “Adobe photoshop restoration and retouching,” *New Riders Press, 3rd edition.*

[HS09] N. Hasler, C. Stoll, M. Sunkel, B. Rosenhahn and H. P. Seidel, “A statistical model of human pose and body shape,” *Computer Graphics Forum (Proceedings of Eurographics)*, vol. 28, issue. 2, pp. 337-346, March 2009.

[HK07] A. Hornung, E. Dekkers and L. Kobbelt, “Character animation from 2D pictures and 3D motion data,” *ACM Transactions on Graphics*, vol. 26, issue. 1, Jan. 2007.

[HW09] Z. Hu, G. Wang, X. Lin and H. Yan, “Recovery of upper body poses in static images based on joints detection,” *Pattern Recognition Letters*, vol. 30, issue. 5, pp. 503-512, Apr. 2009.

[LW06] A. Levin, D. Lischinski, and Y. Weiss, “A closed form solution to natural image matting,” *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR ’06*, vol. 1, pp. 61-68, June 2006.

[RV12] M. Richter, K. Varanasi, N. Hasler and C. Theobalt, “Real-time reshaping of humans,” *International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission*, pages 340-347, 2012.

[RK04] C. Rother, V. Kolmogorov and A. Blake, “GrabCut: interactive foreground extraction using iterated graph cuts,” *ACM Transactions on Graphics, Proceedings of SIGGRAPH 2004*, vol. 23, issue. 3, pp. 309-314, Aug. 2004.

[SM06] S. Schaefer, T. McPhail and J. Warren, “Image deformation using moving least squares,” *ACM Transactions on Graphics, Proceedings of SIGGRAPH 2006*, vol. 25, issue. 3, pp. 533-540, July 2006.

[SJ04] J. Sun, J. Jia, C. K. Tang and H. Y. Shum, “Poisson matting,” *ACM Transaction on Graphics, Proceedings of SIGGRAPH 2004*, vol. 23, issue. 3, pp. 315-321, Aug. 2004.

[TM11] M. T. Islam, K. M. Nahiduzzaman, Y. P. Why and G. Ashraf, “Informed Character pose and proportion design,” *The Visual Computer: International Journal of Computer Graphics - Special Issue on CYBERWORLDS 2010*, vol. 27, issue. 4, pp. 251-261, Apr. 2011.

[ZH10] S. Zhou, H. Fu, L. Liu, D. Cohen-Or and X. Han, “Parametric reshaping of human bodies in images,” *ACM Transactions on Graphics, Proceedings of SIGGRAPH 2010*, vol. 29, issue. 4, July 2010.

[SB12] S. Battiato, G. M. Farinella, G. Puglisi, D. Raví, “Content-aware image resizing with seam selection based on gradient vector flow,” *in proceedings of IEEE International Conference on Image Processing ICIP 2012*, pp. 2117-2120, Sept. 2012.