 DOI: 10.3217/978-3-85125-524-9-25

On Quality Assurance of 3D Bust Reconstructions

Gernot Stuebl, Christoph Heindl, Harald Bauer, and Andreas Pichler[1]

***Abstract*— In this paper a non-reference method for quality assurance in 3D bust reconstruction is presented. The proposed approach is part of an automatic parametrization concept for 3D reconstruction applications with no ground-truth data available. It is based on a novel concept of pair-wise view comparisons, which is new in this field. Evaluation on a dataset of human bust scans shows perfect prediction of human votes.**

I. INTRODUCTION

Exact reconstruction of the human body especially the bust is an application field which got boosted by the raise of low-cost 3D printers and online 3D printing services. Nevertheless creating a high fidelity 3D reconstruction often involves manual post processing.

Recent publications present systems which are able to do reconstructions on a quality level which makes post processing unnecessary, see Heindl et al. [1]. However for these the quality strongly relies on a correct parametrization of the system. Unfortunately parametrization is dependent on the scan data. So no golden standard for a parameter setting exists and the parameter values have to be adopted for each reconstruction individually. In principle human interaction has been shifted from direct manipulation/correction of 3D data to the selection of correct parameter values. Having this in mind, an (semi-)automatic configuration of the parameter values is desirable.

The paper is outlined as followed: first Section II gives an overview of traditional quality assurance methods for 2D and follows with related work in the field of 3D quality assurance. The main approach is described in Section III, whereas Section IV presents the results on a dataset of 3D bust reconstructions. This is followed by a discussion on the applicability of the approach in Section V as well as a conclusion and outlook to future research in the last section.

II. RELATED WORK

A vital part of an automatic parametrization system is a component for assessing the reconstruction quality. The following subsections covers related work in this domain with an introduction of traditional 2D measures and the main emphasis on 3D quality assurance.

A. 2D Quality Assurance

In 2D there are traditional (dis-)similarity measures which are used for quality assurance. Some of these can also be adopted to 3D. A simple one is the Root-Mean-Squared Error (RMSE) [5] of two images I, K which is defined as

$$\mathrm{RMSE}(I,K) := \sqrt{\frac{1}{mn}\sum_{p=0}^{m-1}\sum_{q=0}^{n-1}(I(p,q)-K(p,q))^2} \quad (1)$$

and measures the deviation in each pixel. Based on this the Peak Signal to Noise Ratio (PSNR) [5] is defined as

$$\mathrm{PSNR}(I,K) := 20\cdot\log\frac{I_{\max}}{\mathrm{RMSE}(I,K)} \quad (2)$$

with $I_{\max}$ the maximum possible value in the image (e.g. 255 for monochromatic 8 bit images). PSNR measures the signal fidelity between an original and a disturbed image. A more complex measure is Structural Similarity index (SSIM) [2] which is designed to judge signal fidelity in the way the human vision system does. It is sensitive to structural distortions such as noise contamination, blurring, and insensitive to non-structural distortions such as luminance and contrast change. The mathematical definition is

$$\mathrm{SSIM}(\vec{x},\vec{y}) = \frac{(2\mu_x\mu_y+c_1)(2\sigma_{xy}+c_2)}{(\mu_x^2+\mu_y^2+c_1)(\sigma_x^2+\sigma_y^2+c_2)} \quad (3)$$

with $c_1 = (k_1L)^2$, $c_2 = (k_2L)^2$ as stabilization constants for the division with weak denominators, where $L = 2^b - 1$ denotes the dynamic range of pixel-values with b as the number of bits per pixel and $k_1 = 0.01$ and $k_2 = 0.03$.

B. 3D Quality Assurance

Generally, quality assurance algorithms are divided into full-reference (FR), reduced-reference (RR) and no-reference (NR) algorithms. This distinction is based on the amount of information that is available.

Full-reference algorithms rely on a ground-truth data, e.g. early attempts to judge quality through texture and geometric resolutions belong to this category, see Pan et al.[3]. Also a broad range of algorithms which measure the quality of 3D codecs or stereoscopic 3D are full-reference based, see Mekuria et al. [4]. You et al. [5] give a good overview on how traditional 2D measures can be used for FR 3D quality assurance.

For reduced-reference algorithms the ground-truth is not fully available. Instead of this, selected features are calculated from the ground-truth and used as input of the quality assurance system, see Wang et al. [6] or Rehman and Wang [7].

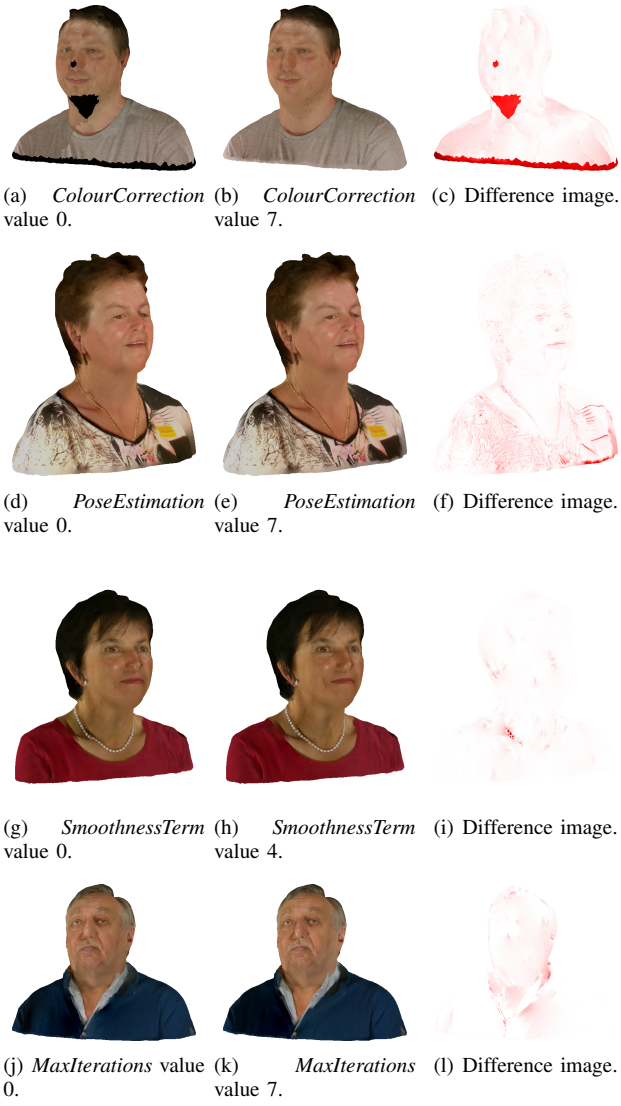
A recent example for a no-reference algorithm is presented by Alexiadis et al. [8]. In this work the 2D key frames which are needed to build the 3D reconstruction are compared to

[1]PROFACTOR GmbH, 4407 Steyr-Gleink, Im Stadtgut A2, Austria {Forename.Surname}@profactor.at

synthesized versions of it. The authors utilize a SSIM based measure to adjust the reconstruction settings. This is close to the proposed approach in this paper. The main difference is that we do not need to process the available key frames but instead work only on synthetic views.

III. AUTOMATIC PARAMETRIZATION

The aim of automatic parametrization is to determine optimal values for different reconstruction parameter-types. In this case optimality means that the parameter value is near or equal the value a human operator would have chosen for the given data. Figure 1 depicts examples for the influences of different parameter-types.



(a) *ColourCorrection* value 0. (b) *ColourCorrection* value 7. (c) Difference image.

(d) *PoseEstimation* value 0. (e) *PoseEstimation* value 7. (f) Difference image.

(g) *SmoothnessTerm* value 0. (h) *SmoothnessTerm* value 4. (i) Difference image.

(j) *MaxIterations* value 0. (k) *MaxIterations* value 7. (l) Difference image.

Fig. 1. Reconstruction effects of different parameter-types shown on Model 0001 in Subfigures (a) to (c), Model 0019 in Subfigures (d) to (f), Model 0010 in Subfigures (g) to (i), and Model 0033 in Subfigures (j) to (l). The images in the last column highlight the differences. The parameter values are set to the extremes, to better demonstrate the effects.

The following procedure illustrates how a non-professional human operator could select a good parameter setting:

1) The operator sets or alters a parameter value.
2) The operator lets the reconstruction run.
3) The operator inspects the result from different views if it is better or worse than before.
4) The operator repeats the steps until some level of reconstruction quality is reached.

Based on this we propose an approach using pair-wise view comparison of different reconstructions. For a parameter-type α the accumulator matrix M_α is a symmetric matrix defined as

$$M_\alpha(k,l) := \sum_{i:=0}^{n} \mathrm{SIM}(V_i(R_{\alpha,k}), V_i(R_{\alpha,l})) \tag{4}$$

where k,l are elements of the ordered parameter value set P_α and $R_{\alpha,k}$ is the reconstruction. Elements in P_α are chosen such there is an increasing influence of the parameter to the observed visual effect. n is the number of equally spaced views around $R_{\alpha,k}$, whereas each view V_i is a 2D projection of the 3D object. For comparison of two images as similarity SIM the Peak Signal to Noise Ratio (PSNR) is used.

We assume that for humans the skin area is very important for quality judgement. Therefore the images are converted into the Hue, Saturation, Value (HSV) colour space before comparison. This should make the comparison more sensitive to skin parts, see Sedlacek [9]. A detailed evaluation and discussion of this step follows in Subsection IV-C and Section V.

Pixel-wise comparison is performed only on the bust itself, since the background is masked out during comparison. Given this framework we propose the optimal parameter value $o_\alpha \in P_\alpha$ to be defined as

$$o_\alpha := \arg\max_{k \in P_\alpha} \sum_{l \in P_\alpha} M_\alpha(k,l). \tag{5}$$

Literally speaking the parameter value o_α creates 2D views which are most similar to the views created with all other values. The hypothesis is that this is also a good parameter value which a human would choose.

IV. EVALUATION

Due to the lack of free datasets for bust reconstruction, an own dataset has been built up during an open house presentation in the company.

A. Dataset

The dataset contains 32 3D human bust scans showing different people, further called models. The data is acquired with a turntable and an off-the-shelf RGB-D sensor. Each individual is scanned in eight key poses. For the detailed set-up of the scan process see Heindl et al. [1].

For the reconstruction four different parameter-types are inspected: colourcorrection level, number of steps for pose estimation, surface smoothness term and maximal iteration of the bundle adjustment. These types form the parameter set $S = \{$ *ColourCorrection, PoseEstimation, SmoothnessTerm, MaxIterations* $\}$. For a detailed explanation of the reconstruction software and the parameter semantics see again Heindl et al. [1].

Eight models are assigned to each parameter-type $\alpha \in S$. A model is reconstructed with the full range of parameter values, which are 8 values for *ColourCorrection, PoseEstimation, MaxIterations*, and 5 values for *SmoothnessTerm*. The parameter space is discrete and the values form the individual parameter value sets P_α.

In a questionnaire 32 people (16 male, 16 female) between 19 and 55 years old, were asked to choose the most aesthetic reconstruction for each model. Since every reconstruction is mapped to a parameter value, they implicitly chose the parameter value which led to the best reconstruction quality.

The best parameter choices according to the human votes have been counter-checked to produce reasonable reconstructions. During this result preparation, one model which was assigned to *SmoothnessTerm*, had to be omitted because of inconsistencies in the data. In detail the parameter value with the most human votes for this model leads to a failed reconstruction similar to the bottom right picture of Figure 4. Therefore the final dataset consists of 31 human judged model reconstructions.[1]

B. Evaluation Criteria

Figure 2(a) and Figure 2(b) show example distributions of human decisions for specific models. One can see the variances in the votes. To cover these variances we define the following correctness criterion:

Definition 1. *A parameter value estimation is correct if it is inside $\mu \pm \sigma$ of the human decisions.*

To test this criterion, human judgements have been simulated with random values. In detail for each decision distribution (e.g. Subfigure 2(b)) a uniformly distributed random value in the same discrete parameter range was generated. If the random value fulfilled the correctness criterion for the decision distribution, it was counted as correct, otherwise as incorrect. With 1000 trials this lead to a mean accuracy of 0.5095 and $\sigma = 0.0841$ which can be seen as baseline for the following tests.

C. Results

In an evaluation which is run on each decision distribution in the dataset, the best parameter value for the reconstruction of a model is estimated using Equation 5 with PSNR as similarity measure. After that the parameter value is checked against the decision distribution with Definition 1. Therefore if the parameter value is inside $\mu \pm \sigma$ of the human decisions, the parameter value estimation is counted as correct and false otherwise. This procedure lead to an estimation accuracy of 1 on the dataset of 31 judged reconstructions.

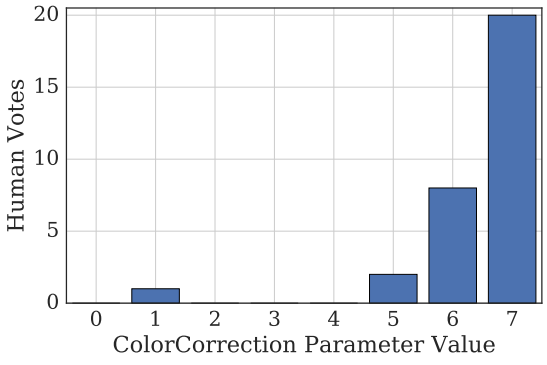
The evaluation has also been run with two other (dis-) similarity measures: Root-Mean-Squared Error (RMSE) and Structural Similarity index (SSIM), see Table I.

The first is a standard measure for deviations. Applying it the accuracy drops to 0.9032. This is further interesting since the RMSE is also the denominator in Equation 2. One can see that the logarithm in the equation is important in this context.

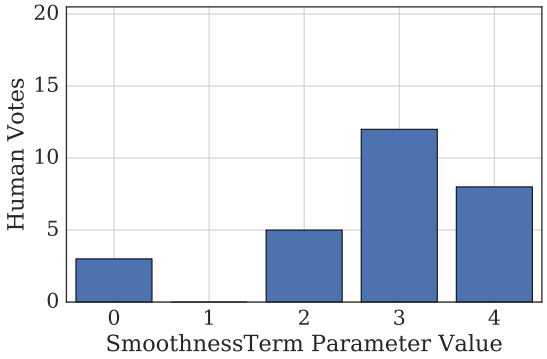
When applying SSIM, which should reflect human perception, the accuracy drops to 0.9032. A detailed look on the results reveals that RMSE as well as SSIM fail on the models assigned to *ColourCorrection*.

A similar comparison has been performed with different colour spaces, see Table II. Beside the HSV colour space Red, Green, Blue (RGB), YCbCr, Grayscale and CIE-Lab colour spaces have been evaluated. RGB is a standard in image representation. When using it the accuracy drops to 0.7742. Recent publications indicate that YCbCr colour space shows advantages in skin detection, see Shaik et

[1]The full dataset can be requested by emailing the main author.



(a) Parametertype *ColourCorrection* on Model 0001.



(b) Parametertype *SmoothnessTerm* on Model 0093.

Fig. 2. Examples of human decision distributions for parameter-types *ColourCorrection* on Model 0001 in Subfigure (a) and *SmoothnessTerm* on Model 0093 in Subfigure (b). One can see the variance in the data.

(Dis-)similarity	Accuracy
PSNR	1
RMSE	0.9032
SSIM	0.9032

TABLE I

ACCURACY ON THE DATASET EVALUATED WITH DIFFERENT IMAGE (DIS-)SIMILARITIES FOR FORMULA 4. EVALUATED MEASURES ARE PEAK SIGNAL TO NOISE RATIO (PSNR), STRUCTURAL SIMILARITY INDEX (SSIM) AND ROOT-MEAN-SQUARED ERROR (RMSE). PSNR PERFORMS BEST.

Colour space	Accuracy
HSV	1
RGB	0.7742
YCbCr	0.7419
Grayscale	0.7419
CIE-Lab	0.7188

TABLE II

ACCURACY ON THE DATASET EVALUATED USING DIFFERENT IMAGE COLOUR SPACES. PEAK SIGNAL TO NOISE RATIO (PSNR) IS USED AS SIMILARITY MEASURE. EVALUATED COLOUR SPACES ARE HUE, SATURATION, VALUE (HSV), RED, GREEN, BLUE (RGB), YCBCR, GRAYSCALE AND CIE-LAB. USING HSV SHOWS THE HIGHEST ACCURACY.

al. [10]. Nevertheless by using this colourspace the accuracy drops to 0.7419. On the other hand with Grayscale colourspace the accuracy drops also to 0.7419. This is of further interest since the applied grayscale conversion algorithm simply takes the Y component of YCbCr and omits the colour channels. This procedure is common usage in photo editing software like Photoshop [2] or GIMP [3]. A further look on the results uncovers that YCbCr and Grayscale have their wrong estimations on the same models. Therefore CbCr colour encoding adds no benefit to using the Y channel alone in this application. CIE-Lab colour space was also evaluated since it approximates human vision, unfortunately in this application the accuracy dropped to 0.7188.

D. Comparison with state-of-the-art

A comparison with state-of-the-art is difficult, since the algorithms are usually embedded into a certain application scenario which is not always exchangeable.

Nevertheless the *Evaluation of the appearance quality* part in the publication of Alexiadis et al. [8] has been adopted to our set-up: The parameter value of which the reconstructed views are most similar to the ground-truth key-frames is chosen as best value. Like in Alexiadis et al. the similarity measure is SSIM and the colour space HSV.

When run on the dataset the accuracy is at 0.4062. This is not a fair comparison since the appearance evaluation is only a part of the whole framework of Alexiadis et al. and only confirms that the set-ups of both approaches cannot be intermixed.

V. DISCUSSION

This section contains a discussion about the applicability of the approach as well as considerations on the runtime.

A. Applicability

The proposed approach relies on an interesting property of the reconstruction principle: changes in the parameter value lead to mainly distinct local deviations in the model.

PSNR as the chosen similarity measure has to be sensitive to this deviations. To visualize this a metric Multi Dimensional Scaling (MDS) [11] algorithm is utilized. An MDS algorithm tries to position each object in multi-dimensional space such that the between-object distances are kept as well as possible. This gives more insight into the working principle of the proposed approach since it illustrates which images are similar from the view of PSNR.

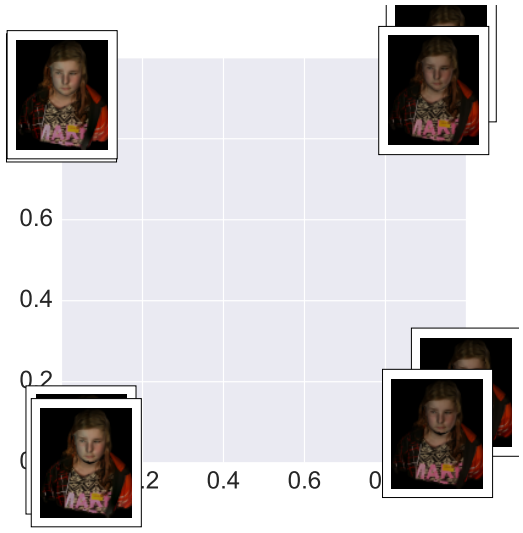


Fig. 3. Multidimensional scaling layout of all frontal views for parameter-type *ColourCorrection* on Model 0093 using Peak Signal to Noise Ratio (PSNR) as similarity measure and Hue, Saturation, Value colour space. Top right is the best choice, bottom left and right show deviations on the pine, top left on the right cheek. The farther the images are away from each other the more they are different in the meaning of PSNR. The images form clusters according to local deviations in the reconstruction.

In Figure 3 all frontal view reconstructions in the whole parameter value range for *ColourCorrection* of a specific model (0093 in the dataset) are laid out with an MDS algorithm. To create the necessary distance matrix for the algorithm, the similarities in M_α were converted to distances. On the top right is the optimal reconstruction. Bottom left and bottom right show deviations on the pine, whereas top left deviates on the left cheek. It can be seen that images with similar deviations are clustered together.

However the increasing visual effect of the parameter values, mentioned in Section III, is not visible in the layout, on the one side because MDS is a form of non-linear dimensionality reduction and on the other side PSNR as underlying measure does not fully reflect the human visual perception.

Figure 4 depicts also a MDS layout for a whole parameter value range (parameter-type SmoothnessTerm on Model 0098 in the dataset). On bottom right is the rare case of a complete failed reconstruction, which has a high distance to the other images. One can see that the case of a global deviation is treated well, as long as it is not in the majority of the images.

The dependency on distinct local deviations can be a loss of generality of the approach. However especially in the area of human 3D reconstruction there should be a wide range of possible applications. Furthermore our approach is not dependent on a certain reconstruction principle.

[2]http://www.adobe.com/at/products/photoshop.html
[3]https://www.gimp.org

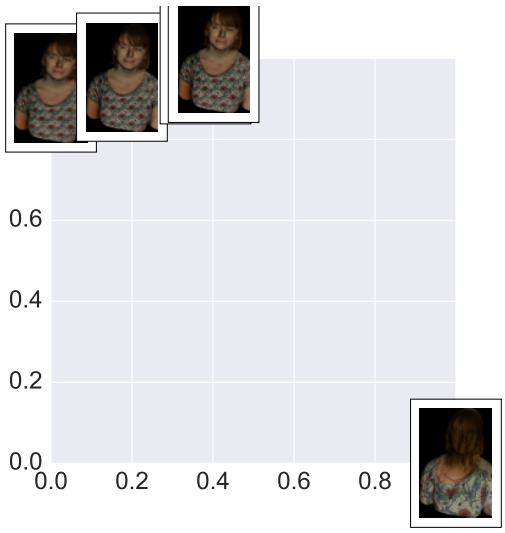


Fig. 4. Multidimensional scaling layout of all frontal views for parameter-type *SmoothnessTerm* on Model 0098 using Peak Signal to Noise Ratio (PSNR) as similarity measure and Hue, Saturation, Value colour space. The farther the images are away from each other the more they are different in the meaning of PSNR. Therefore the failed reconstruction on bottom right has a high distance to the other images.

A further eventual loss of generality is the coupling to a specific colour space (HSV) together with the assumption that human decisions are dependent on skin deviations. All models in the dataset are Central Europeans with white skincolour. It is not sure that the proposed approach in this configuration works also with models having other skincolours. Nevertheless the approach is a good starting point for future work, see Section VI.

A final point regarding applicability is that the proposed approach inspects all parameter-types isolated, see Section VI on future work to this issue.

B. Runtime Considerations

The proposed method utilizes a brute force evaluation of all parameter values. While the final comparison of the views is computationally cheap, the reconstruction itself is time consuming: On an Intel Core i5-200 CPU with a NVIDIA Geforece GTX 560 and 16GB RAM it takes in the mean 145 s to do a reconstruction. To overcome this issue, the reconstruction has been implemented as web service in the Amazon Cloud.

Since the reconstructions are independent of each other, they could be run fully in parallel, benefiting from the virtually infinite computational power in the cloud. However in practice we run the parallelization in a way such that one parameter-type can be fully evaluated at once.

VI. Conclusion and Future Work

In this work an approach utilizing pairwise comparison of 2D views from different 3D model reconstructions has been demonstrated, which simulates human quality choices. The approach shows perfect prediction on the given dataset.

The essential part of the approach is to select the reconstruction which is most similar to all others. The effect is that reconstructions with local deviations are sorted out. This idea is new and might inspire other scientific work.

From the technical side there are two main possibilities of improvement, which are caused by the nature of the used dataset. First the dataset only covers white-skinned Central Europeans and the approach is coupled to a specific colour space. So there could be a loss in generality when inspecting models with other skincolours. To overcome this a future work could use a face detector as pre-step and parametrize the comparison to the actual skincolour. For this new models have to be added to the dataset.

Another future work may approach the issue of isolated parameter-type evaluation. Unfortunately with the available questionnaire, combinations of parameters cannot be evaluated since they are not in the data. However for future work this would be very interesting, since it could provide further insights to the generality of the approach. In case that there will be significant dependencies between parameter-types a future version may include some kind of genetic algorithm to find the best combination.

Acknowledgment

This research is carried out within the ”FTI-Project ProTechLab” project funded by the State of Upper Austria through the Strategic Economic and Research Program ”Innovatives OÖ 2020”.

References

[1] C. Heindl, S.C. Akkaladevi, and H. Bauer. *Capturing Photorealistic and Printable 3D Models Using Low-Cost Hardware*, pages 507–518. Springer International Publishing, Cham, 2016.

[2] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *Image Processing, IEEE Transactions on*, 13(4):600–612, April 2004.

[3] Y. Pan, I. Cheng, and A. Basu. Quality metric for approximating subjective evaluation of 3D objects. *IEEE Transactions on Multimedia*, 7(2):269–279, April 2005.

[4] R. Mekuria, P. Cesar, I. Doumanis, and A. Frisiello. Objective and subjective quality assessment of geometry compression of reconstructed 3D humans in a 3D virtual room, 2015.

[5] J. You, G. Jiang, L. Xing, and A. Perkis. *Quality of Visual Experience for 3D Presentation - Stereoscopic Image*, pages 51–77. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.

[6] X. Wang, Q. Liu, R. Wang, and Z. Chen. Natural image statistics based 3D reduced reference image quality assessment in contourlet domain. *Neurocomputing*, 151, Part 2:683 – 691, 2015.

[7] A. Rehman and Z. Wang. Reduced-reference image quality assessment by structural similarity estimation. *IEEE Transactions on Image Processing*, 21(8):3378–3389, Aug 2012.

[8] D.S. Alexiadis, A. Chatzitofis, N. Zioulis, O. Zoidi, G. Louizis, D. Zarpalas, and P. Daras. An integrated platform for live 3D human reconstruction and motion capturing. *IEEE Transactions on Circuits and Systems for Video Technology*, PP(99):1–1, 2016.

[9] M. Sedlacek. Evaluation of RGB and HSV models in human faces detection. Central European seminar on computer graphics, Budmerice. In *IIIA.1-5 - Conference on Computer Systems and Technologies - CompSysTech2004*, page 125131, 2004.

[10] K.B. Shaik, P. Ganesan, V. Kalist, B.S. Sathish, and J.M.M. Jenitha. Comparative study of skin color detection and segmentation in HSV and YCbCr color space. *Procedia Computer Science*, 57:41 – 48, 2015. 3rd International Conference on Recent Trends in Computing 2015 (ICRTC-2015).

[11] I. Borg and P.J.F. Groenen. *Modern Multidimensional Scaling: Theory and Applications (Springer Series in Statistics)*. Springer, 2nd edition, August 2005.