CHECKER: Real-time Scan Quality Checking and Associated User Experience

Matthew S. GILMER, Steven C. HAUSER, David BRUNER
Size Stream, LLC., Cary NC, USA

https://doi.org/10.15221/23.30

Abstract

In recent years there have been several substantial advances in mobile 3D body scanning technology and Size Stream has been, and will continue to be, engaged in this effort. In addition to improving measurement accuracy, Size Stream is improving the user experience of the scan process. We are providing accurate measurements and 3D body avatars that are suitable for fitness tracking, ready-to-wear/uniform sizing, and made-to-measure garment ordering with minimum friction. We will present our new machine-learning system designed to detect 'scan errors' in real time (CHECKER). We use the CHECKER results to guide the user in real time to take an error-free scan rather than alert the user after they have completed the scan process. The CHECKER system can detect obscured body parts and baggy clothing. Size Stream plans to extend the capabilities of CHECKER so that it can detect more 'scan errors' including, various pose errors, clutter, hair issues (down on the neck rather than tied up), and sub-optimal lighting.

Keywords: 3d body scanning, machine learning, apparel, user experience

1. Introduction

Size Stream has always been committed to being a leader in 3D object accuracy and the associated measurement quality. To this end we conduct several scan parties per year to test the performance of our technology and acquire data to use in improving it. This gives us a detailed understanding of how well our scanning systems perform in ideal conditions (one-on-one scanning instruction, clear scanning background, etc.).

In addition to our commitment to accuracy, we at Size Stream are also excited to share our technology with many more people and that is why we are working hard on ways to make the scan-at-home process as simple as possible without sacrificing accuracy. One of the ways to do this is to give real time feedback to the scanner when we detect issues that are known to affect accuracy. In this paper we first focus on the detection of these issues using machine learning and then we explore implementation options while judging them through a User Experience (UX) lens.

Good detectors require good training data and to this end we developed internal software to be used by an annotator which we will call the Annotation GUI (AGUI) which provides an intuitive interface for data annotation. The data accessible by the ScanGuard system comes from our customers who gave consent to have their photos saved and is extremely valuable because it allows us to see how our technology is used in a less supervised setting than our scan parties and to train scan-error-detection models on these settings. Our database currently contains over 10,000 scans with RGB images and over 1,000,000 scans with silhouettes. The latter can be used to train models that take only a silhouette input, but these models are outside the scope of this paper.

For context in the following sections, we provide a short description of our current scan process. First, the user is played a short video or slide show communicating the Pre-Scan Instructions (summarized in Table 1. If the user decides to proceed and initiate a scan our pose guidance system (using Google's MLKit Pose) coaxes the user into an A-pose facing the device, takes a picture, and saves the final pose information from MLKit. After the 'front view', the user is instructed to turn 90 degrees to the side (either side is okay) with their hands straight down by their sides. Once in position the user is prompted to look straight ahead (not at the device) and takes a picture saving the final pose information for the 'side view'. An example of captured images from a successful scan is provided in Figure 1. Segmentation is then performed on device and resultant silhouettes are uploaded for cloud processing along with two pose files (from the front and side views) and the subject's user-inputted height, weight, age, and gender. These are the inputs into our avatar generation and measurement extraction system, the details of which are outside the scope of this paper.

Table 1. Pre-Scan Instructions

Prep	Skin-tight clothing, tie up long hair, remove shoes and accessories
Space	Blank wall in a clear area about 4 x 7 feet. Your whole body needs to be in view
Scan	Place your phone at knee to waist height.

Figure 1. Example of a successful scan's front (left) and side (right) view images

2. Methods

2.1. Scan Error Identification and Characterization

To identify the important scan errors, we look for any failure to comply with our scan instructions. We assess the frequency and impact to guide prioritization of detection efforts. The frequency determination is subject to some uncertainty from human error which we minimize by writing clearly defined scan error definitions to be referenced by any reviewer (see Table 2). Reviewers also undergo training sessions with our head reviewer. The impact estimation is also currently subject to some uncertainty since it is dependent on the concept of 'poor fit'. A return is processed for any customer who claims poor fit and thus depends strongly on fit preference. The best data that we currently have for predicting poor fit comes from comparing scan results from scans where a scan error was present to reference scans (either scans without the scan error or scans taken with our booth scanner, the SS20). We look at the magnitude of the deviation in key measurements and leverage our fit experts to determine the impact. Here we focus on the five scan errors for which we have developed detection methods. They are the five listed in Table 2 and shown in Figures 2 and 3.

Table 2. Scan Error Definitions

Scan Error Type	View(s)	Definition
Feet Cut Off	Front and Side	Either Ankle in the front view or the front ankle in the side view is obscured or off screen
Baggy Shirt	Front	Excess cloth over waist by (1/2" by 4")
Baggy Pants	Front	Excess cloth under waist by (1/2" by 4")
Hand On Thigh	Side	Arm covering front of body by at least a palm (1/2" x 4")
Arm Blocking Back	Side	Arm covering back of body by at least a palm (1/2" x 4")

Figure 2. Examples of Scan Errors in the Front View (from left to right: 'Feet Cut Off', 'Baggy Shirt' and 'Baggy Pants')

Figure 3. Examples of Scan Errors in the Side View (from left to right: 'Feet Cut Off', 'Hand On Thigh', and 'Arm Blocking Back')

2.1.1. Feet Cut Off

An example of 'Feet Cut Off' is shown in the left-most images of Figures 2 and 3. The issue is that the surface the phone is resting on (a mattress in this case) is obscuring the feet and MLKit Pose is designed to estimate position even when certain landmarks are not visible (we require full visibility of the subject for our avatar generation engine). To solve this issue, we built our own detector that is designed to predict 'visibility' as well as position for each landmark. Our method works for any landmark being obscured but our training data did not have enough instances of scans with other landmarks obscured. However, since the frequency of other landmarks being obscured is so small it is relatively unimportant.

2.1.2. Baggy Shirt and Baggy Pants

An example scan exhibiting both ‘Baggy Shirt’ and ‘Baggy Pants’ is shown in the right-most image of Figure 2. Baggy clothes add pixels to the body in segmentation which get processed into an avatar that is too large. We found in our internal tests that measurements resulting from scans with baggy clothes are biased large especially for torso circumferences in the case of ‘Baggy Shirt’, and hip and thigh circumferences in the case of ‘Baggy Pants’. We trained our detector only on front view images. As will be discussed in Section 4.2, we plan to give real-time feedback in the event of detection and since the side view is further along in the scan process the friction caused by notifying the user at this later stage would be too great.

2.1.3. Hand On Thigh

An example of ‘Hand On Thigh’ is shown in the central image of Figure 3. Correction of this scan error is not currently part of CHECKER but the possibility of its inclusion is discussed in Section 4.3. This scan error adds pixels to the front in the segmentation and thus increases key torso and hip circumferences.

2.1.4. Arm Blocking Back

An example of ‘Arm Blocking Back is shown in the right-most image of Figure 3. Like, ‘Hand On Thigh’, this scan error is not currently part of CHECKER but the possibility of its inclusion is discussed in Section 4.3. This scan error adds pixels to the back in the segmentation and thus increases key torso circumferences.

2.2. Scan Errors and Annotation

Using AGUI, we annotated over 10,000 scans. We developed a pose detector for which one purpose was the detection of ‘Feet Cut Off’. In general, the detector’s purpose was to take either a front or side view RGB image as input and turn it into coordinates and visibilities for a list of landmarks. To facilitate annotation for this detector in AGUI, keyboard shortcuts were used to mark the coordinates as ‘in view’ or not and coordinates were modifiable via clicking or dragging and dropping. For the rest of the scan errors, namely ‘Baggy Shirt’, ‘Baggy Pants’, ‘Hand On Thigh’, and ‘Arm Blocking Back’, we marked either positive or negative via checkboxes on a per-scan-per-view basis.

2.3. Thresholding

After a model is trained, we evaluate it on a validation set (never seen by the detector during training). We calculate precision, recall, and F1-score (the harmonic mean of the precision and recall) for thresholds across the range (0, 1). We choose the threshold that gives the maximum F1-score.

3. Results

3.1. Feet Cut Off

For this scan error we found the threshold did not depend strongly on gender. The precision and recall for the best F1-score turned out to be 99% and 82%, respectively.

3.2. Baggy Shirt and Baggy Pants

Using separate thresholds depending on the scan error (‘Baggy Pants’ or ‘Baggy Shirt’) and gender (male or female) we can chart the results in Table 3.

Table 3. ‘Baggy Shirt’ and ‘Baggy Pants’ Precision and Recall

Precision [%] | Recall [%]	Male	Female
‘Baggy Shirt’	92 | 92	85 | 82
‘Baggy Pants’	92 | 94	85 | 84

3.3. Hand On Thigh and Arm Blocking Back

Using separate thresholds depending on the scan error (‘Baggy Pants’ or ‘Baggy Shirt’) and gender (male or female) we can chart the results in Table 4.

Table 4. Hand On Thigh and Arm Blocking Back Precision and Recall

Precision [%] | Recall [%]	Male	Female
‘Hand On Thigh’	94 | 86	96 | 85
‘Arm Blocking Back’	95 | 82	93 | 77

4. User Experience (UX)

4.1. Introduction

Here we will present some ideas for putting our detectors to use during scanning in real-time. This describes CHECKER in its current incomplete state. Some revision is necessary before deploying CHECKER in any of our scanning apps and continued revision may be prudent. For all scan errors detected we will monitor the success rate of scanners after they have received feedback from CHECKER. If these success rates show room for improvement, we may deploy app versions with revised UX, perhaps employing AB testing.

4.1. Feet Cut Off

We can detect this condition in either the front or side view (see Table 2). In rare cases the ankles will be visible in the front view but not in the side view, however for most cases we just need to detect it in the front view and give the user a message. This is the first stage of CHECKER. Figure 4 describes a draft of such a message. The subject shown has baggy clothes and this will be corrected to show a subject in proper scan wear before it is deployed. Additionally, a similar message will be designed for the rare case of detection in the subsequent side view.

Figure 4. Draft UX for a positive 'Feet Cut Off' detection is a video showing just a transition from the left image to the right image with the accompanying voice prompt: "Your feet are out of view. Please reposition so that they can be seen. The scan will restart."

4.2. Baggy Shirt and Baggy Pants

We can detect this condition in only the front view (see Table 2). This is the second stage of CHECKER and takes place directly after the front view image is captured. Figure 5 describes a draft of the message the subject would receive after a positive detection of either 'Baggy Shirt' or 'Baggy Pants'. The subject shown in the right panel still has baggy clothes and this will be corrected to show a subject in proper scan wear before it is deployed.

Figure 5. Draft UX for a positive 'Baggy Shirt' and/or 'Baggy Pants' detection is a video showing just a transition from the left image to the right image with the accompanying voice prompt: "Please check that your clothes are not baggy or loose. We will start the scan again."

4.3. Hand On Thigh and Arm Blocking Back

At present we have no plans to include the detection of these conditions as part of CHECKER, but it may be included in future iterations. Before doing so we would need to develop a pose guidance algorithm using the detector during the side view. This development would likely have a few testing and tweaking iterations. The voice prompt accompanying detection would need to short but also descriptive enough to explain the issue. Since this scan error is more complex than the previous ones, this would be no easy task.

5. Summary

Here we have described our CHECKER system for real-time scan error correction in its current pre-deployment state. We have summarized our normal scanning process for context and examined some scan errors ('Feet Cut Off', 'Baggy Shirt', 'Baggy Pants', 'Hand On Thigh', and 'Arms Blocking Back') that sometimes occur. For each of these scan errors we have provided what we can of our detector development process including data collection, annotation, and thresholding. We presented our results from our detector validation for each considered scan error. Finally, we presented our plans for the UX that will accompany CHECKER. We have plans to extend the number of scan errors handled by CHECKER in the future to include pose errors (probably beginning with 'Hand On Thigh' and 'Arms Behind Back'), clutter, hair issues (down on the neck rather than tied up), and sub-optimal lighting.