Hybrid Approach for Orientation-Estimation of Rotating Humans in Video Frames Acquired by Stationary Monocular Camera

David Baumgartner
Christoph Praschl
RG Advanced Information Systems and Technology, Research and Development GmbH, University of Applied Sciences Upper Austria
Softwarepark 11
4232 Hagenberg, Austria
first.last@fh-hagenberg.at

Tobias Zucali
amb-technology.ai, AMB GmbH
Hafenstraße 47-51
4020 Linz, Austria
t.zucali@amb-technology.ai

Gerald A. Zwettler
RG Advanced Information Systems and Technology, Department of Software Engineering, University of Applied Sciences Upper Austria
Softwarepark 11
4232 Hagenberg, Austria
gerald.zwettler@fh-hagenberg.at

ABSTRACT

The precise orientation-estimation of humans relative to the pose of a monocular camera system is a challenging task due to the general aspects of camera calibration and the deformable nature of a human body in motion. Thus, novel approaches of Deep Learning for precise object pose-estimation in robotics are hard to adapt to human body analysis. In this work, a hybrid approach for the accurate estimation of a human body rotation relative to a camera system is presented, thereby significantly improving results derived from poseNet by applying analysis of optical flow in a frame to frame comparison. The human body in-place rotating in T-pose is thereby aligned in the center, applying object tracking methods to compensate for translations of the body movement. After 2D skeleton extraction, the optical flow is calculated for a region of interest (ROI) area aligned relative to the vertical skeleton joint representing the spine and compared frame by frame. To evaluate the eligibility of the clothing as a fundament for good feature, the local pixel homogeneity is taken into consideration to restrict the optical flow to heterogeneous regions with distinctive features like imprint patterns, buttons or buckles besides local illumination changes. Based on the mean optical flow with a coarse approximation of the axial body shape as ellipsis, an accuracy between 0.1° and 2.0° by a target rotation of 10° for orientation-estimation is achieved on a frame-to-frame comparison evaluated and validated on both, Computer Generated Imagery (CGI) renderings and real-world videos of people wearing clothing of varying feature appropriateness.

Keywords
Object Tracking, Orientation-Estimation, Optical Flow, Image Alignment, Human Skeleton Extraction, Human Pose Estimation, Pixel Homogeneity

1. INTRODUCTION

Accurate knowledge of the orientation of a human from a monocular video is besides skeleton analysis of highest importance to allow for 3D reconstruction of the body, e.g., for subsequent analysis of the size of a garment. Furthermore, in the field of collaborative human-computer interaction with respect to industry 4.0 or general person tracking in the surveillance field, the analysis of human pose and orientation is of high interest. Using only one static monocular camera system without any knowledge of extrinsic camera parameters, the human as the target object needs to be recorded from several views, e.g., by in-place rotating in T-pose. With a static camera system and rotating jet centric aligned objects, it is inevitable to gain accurate knowledge of the exact orientation relative to the camera system to allow for subsequent analysis and post-processing such as the 3D reconstruction of the individual body shape. The deformable and non-rigid nature of a human body in motion with all intrinsic and decoupled movements of hips, head and ankles harden the task of defining exact orientation relative to a camera system. Furthermore, a human body with its rotational-symmetric and homogeneous shape will project to very similar silhouettes in the 2D video recordings if only marginally varying the rotation

https://doi.org/10.24132/CSRN.2020.3001.5

angle making it hard to derive specific, unambiguous estimations.

2. Related Work

2.1 Action Recognition

The field of action recognition and pose classification can be seen as a predecessor of nowadays joint-based human pose estimation. For most of these approaches, it is generally assumed that human detection, i.e., removing the background, already has been achieved somehow [Ger10a], and thus the silhouette contour is available for processing. While simple local features such as the shape modeled as the histogram of oriented gradients (HOG) can be exploited to deduce the human activity, incorporation of motion as the histogram of optical flow (HOF) [Lap08a] for spatio-temporal context introduces additional robustness in video processing. Besides shape, also spatio-temporal texture features can be utilized by applying local binary patterns (LBP) [Kell08a]. Body shape as silhouette derived after background removal can be directly used as a feature vector for clustering in the vector space [Singh08a]. State of the art feature detectors such as Speed-Up Robust Features (SURF) allows deriving robust markers as input for action recognition [Ben14a]. Nowadays, action recognition is achieved using neural networks or even Deep Learning architectures [Ron16a].

2.2 Deep Learning for Human Pose Estimation

With the evolution of Deep Learning frameworks and methodologies, accurate human pose estimation is now feasible with 2D joint skeletons derived from input video frames, e.g., with DeepPose introduced by Toshev [Tos14a]. To derive the joint skeleton in 3D, it can either be derived from a single planar RGB image utilizing Deep Learning [Li14a] or from predicting the 3D skeleton by projecting assumed 3D positions back to 2D for evaluating the best match [Che16a]. A 3D skeleton can easily be derived if incorporating multiple views with particular 3D approximation to combine the results with a simple neural network [Rho18a]. If the relative orientation of the human within the frames is known, the topic of this paper, plain 2D skeletons and epipolar geometry are sufficient for an approximation of 3D skeleton joint positions [Yan98a].

2.3 Human Orientation Estimation

For the domain of human orientation-estimation, specific deep learning models can be trained [Cho16a] at validation accuracy slightly above 80%. Given the 3D skeleton joint positions, the human orientation can further be directly estimated by e.g. calculating the orthogonal vector from a plane between neck and left/right leg [Cho16a] or by assessing the hip-rotation with a plane between chest and left/right hip [Wei19a]. Likewise, the human face orientation can be approximated from relative landmark positions such as nose, mouth, eyes from 2D [Sug05a] [Gou04a] or from 3D marker positions for full pose recovery [Der17a].

While the estimation of the human body as an elastic and deformable multi-joint kinetic skeleton is hard to determine and unambiguous, for solid bodies in case of a priori known 3D shape Deep Learning has recently led to significant improvements in 3D orientation and pose estimation [Xia18a].

A similar approach using pixel motions to evaluate the rotational changes is presented in [Pra20a]. There we describe a method to determine the rotation in the context of head-mounted augmented reality (AR) devices. In contrast to the presented paper, we are using the method for inside-out orientation determination for AR devices, while this paper introduces an outside-in approach.

3. Hybrid Approach for Human Orientation-Estimation

3.1 Overview of our approach

Image data utilized in this research work is acquired with a particular camera setup. One monocular static camera system is used to record a person rotating in T-pose on a spot facilitating simplified camera calibration. The body skeleton is tracked in 2D by utilizing OpenPose pose estimation [Cao17a]. To compensate for body movement, the 2D frames are scaled and aligned according to the spline-joint of the tracked skeletons.

A rough orientation estimation is directly derived from the body skeleton evaluating the shoulder to hip plane, while significant refinement is achieved by evaluating the optical flow. Preconditioned nearly homogeneous object position and distance from the camera, the person elliptic axial cross-section is estimated from 0° and 90° views. Utilizing the modeled person cross-section and the quantitative phase shift of optical flow when comparing two frames, the relative rotation can be calculated based on the derived translation. The pixel homogeneity, via a co-occurrence matrix, allows excluding of homogeneous areas from optical flow analysis, resulting in a more stable and precise result.

3.2 MATERIAL

With a very broad field of application in terms of the used camera system, this paper differentiates between tests using Computer Generated Imagery (CGI) data and those using real-world data to address a wide range of scenarios. The first of the two used materials is shown in Figure 1 and uses the game engine Unity to render a virtual environment containing a model of a person, which rotates around its spine.

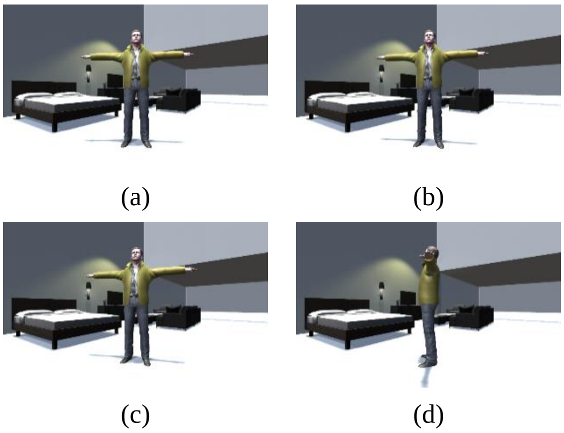

(a) (b)
(c) (d)

Figure 1: Sample images of a virtual person model with a rotation of (a) 0°, (b) 10°, (c) 20° and (d) 90°.

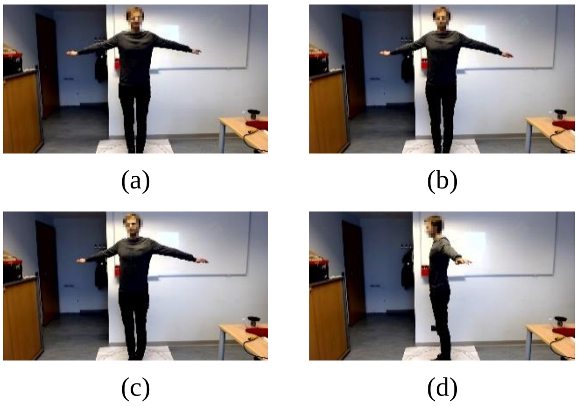

(a) (b)
(c) (d)

Figure 2: Sample images of a real-world person standing on the degree circle shown in Figure 3 with a rotation of (a) 0°, (b) 10°, (c) 20° and (d) 90°.

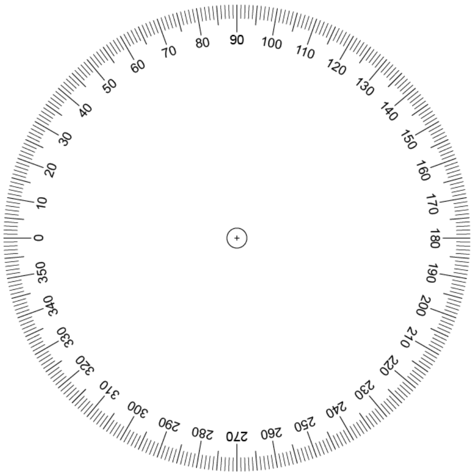


Figure 3: Degree circle underlay used as a ground truth for the real-world tests.

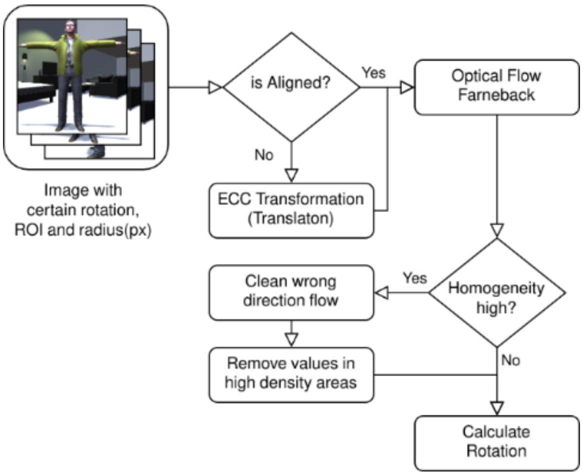


Figure 4: Shows the required overall workflow steps of the rotation calculation.

That allows us to verify the approach in a perfect setting with a precisely known rotation as the ground truth and without additional distortions related to the used camera or, e.g., movements of the camera. However, it also allows us to evaluate the effect of such influences, which can be configured/simulated in such a virtual environment.

In the second scenario, videos are captured form a person who is rotating in place around their spine, shown in Figure 2. The usage of real camera footage is associated with, among other things, the problems mentioned above, but allows us to evaluate our approach in a real-world scenario, in which it finally should find application. In this test environment, we also distinguish between two camera settings – a positional static and a dynamic one. In the first setup, we are using a fixed camera on a tripod, and in the second one, the camera is held by a second person. This additional differentiation allows comparing the synthetic scenario tests with the static camera. First, to evaluate the influence of different camera-related impacts, and secondly, the influence of real-world conditions that contain additional sources of errors, such as (minimal) movements of the camera. As the ground truth in the real-world scenarios, we use a motorized turntable that can be rotated degree-wise and is mounted on a printed degree circle, shown in Figure 3.

In both, the synthetic scenario as well as the real-world scenario, we evaluate the rotation of the person at 0°, 10°, 20°, and 90° each based on a single image frame with 24-bit color depth. The approach is tested with images of size 1920x1080px. For this, the synthetic scenario scene is exported with this resolution at the corresponding rotations. The real-world images are taken with a Logitech C920 that is mounted on a tripod in the static scenario and in the dynamic one with an iPhoneXR.

3.3 METHODOLOGY

Identifying the rotation of an object with a monocular camera requires various aspects to be checked and corrected. First of all, three frames are necessary, as described in section 3.2, one as origin, one as the target, and one with a 90° view. Handheld cameras usually introduce movement between consecutive images. The target image therefore has to be aligned to prevent such an error first. This alignment has to be a translation only and there must not be any Euclidian, affine, or homography transformation. This transformation allows us to align the background and to preserve the motion between the frames. Figure 4 shows the overall abstract workflow. It highlights the required steps for this approach to calculate the rotation of an object within images.

Following the alignment correction, the optical flow gets calculated for the entire image but evaluated only within a small region of the rotating object. Since the flow in the direction of the vertical axis is not relevant, the subsequent process steps ignore it. Within the resulting features (a two-dimensional array), the direction of the flow is essential. The main rotation direction is determined by checking for one direction that has more than 75% of all flow values in its direction. If this is not guaranteed, it is impossible to calculate the rotation from the given samples. That leads to dropping all values in the wrong direction and are therefore classified as noise by the optical flow calculation. After those steps, another check further cleans the remaining values. Within the small region of interest, the homogeneity is vital to decide for further cleaning tasks or to calculate the rotation directly. Via the co-occurrence matrix based on the grayscale values in the horizontal direction, reliable homogeneity information can be acquired. Combined with the statistical variance within the horizontal flow, it is possible to decide for further cleaning or not.

To determine the homogeneity of the texture in the ROI, a gray value co-occurrence matrix M is calculated for $range = 64$ as number of bins for 8bit unsigned input image I leading to range factor $rF = \frac{range}{256}$ with

$$M(\langle T_i \rangle, k, l) = \frac{1}{width * height * |\langle T_i \rangle|}$$

$$\sum_i^{|\langle T_i \rangle|} \sum_x^{width} \sum_y^{height} \begin{cases} 1 & if\ floor(\frac{I(x+\Delta x_i)}{rF}) = k \\ & \wedge floor(\frac{I(y+\Delta y_i)}{rF}) = l \\ 0 & else \end{cases} \quad (1)$$

applying the translation vector $T_1(1,0)$ and $T_2(0,1)$ provided as a parameter set $\langle T_i \rangle$ with translation T_i as pair $(\Delta x_i, \Delta y_i)$. Based on co-occurrence matrix M, homogeneity is calculated as

$$homogeneity(M) = \sum_k^{range} \sum_l^{range} \frac{M(k,l)}{1+|k+l|}. \quad (2)$$

The following cleaning steps intend to remove unbalance within the remaining optical flow values. If the image has many homogenous areas, then the optical flow contains peaks within the histogram of the residual values. In order to determine if such peaks exist, the histogram is split into ten equal-width slices. Values in a slice are dropped if the slice contains more than the 85 quantiles of the number of values. That cleaning results in a more uniform distribution of all flow values. This step is valid under the assumption that the image contains many homogenous areas were no real optical flow can be calculated and would bias the further calculation steps. This allows to boost/reinforce the remaining values in their contribution to the rotation calculation.

Finally, after pre-processing, the core rotation calculation requires some additional information besides the optical flow data. It involves the width of the cross-section of the rotating object. Therefore, it is mandatory to know a priori where the rotating object and its outline resides. If the object is a can, then the cross-section is always equal, but in the case of a person, a 90° rotated view is required to extract the cross-section width. This information is used as the diameter in pixels. Reducing this value by a factor of two results in a first approximation of the radius and is further used as an adjacent leg. Flowing that is the mean value of the remaining optical flow values used as the opposite side value as shown in Figure 5 as *x Flow*. The optical flow values are further corrected by calculating a scale factor with the arccosine, dividing

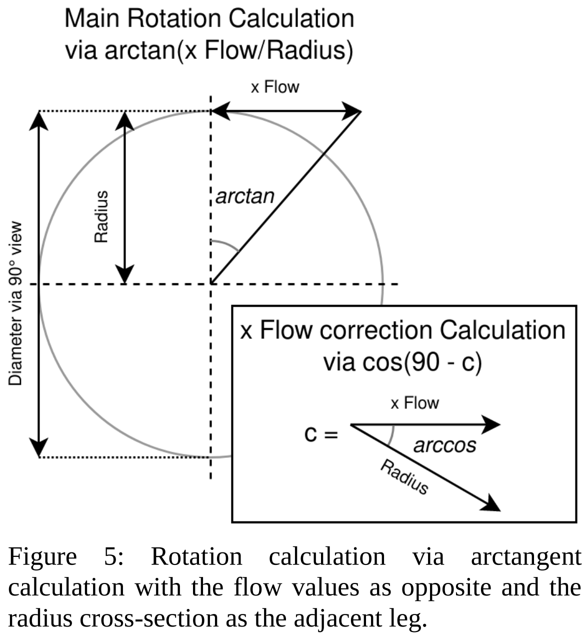
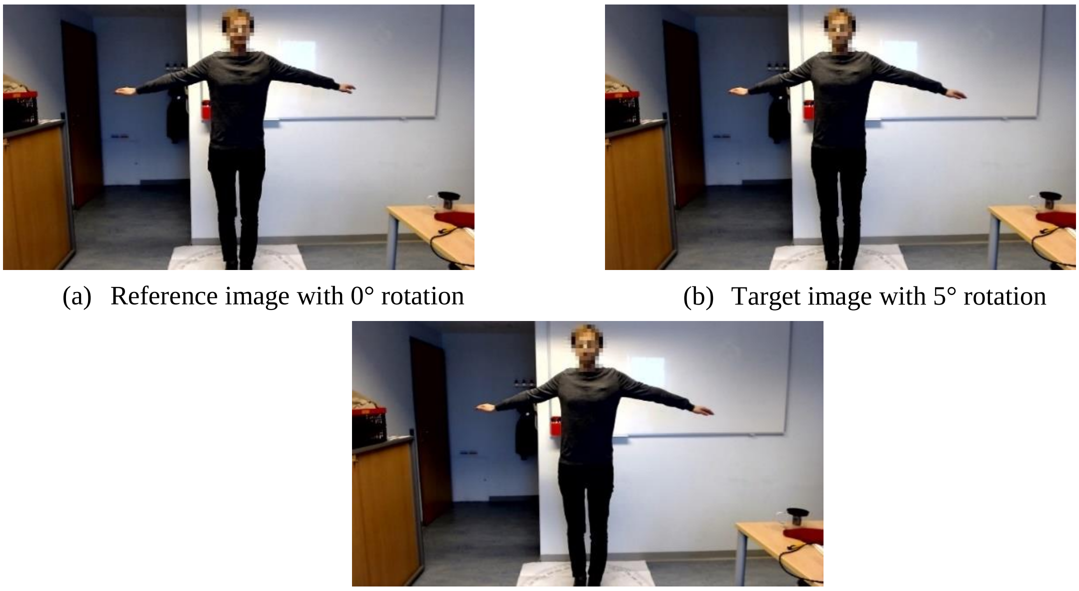

Main Rotation Calculation
via arctan(x Flow/Radius)
x Flow
Radius
arctan
Diameter via 90° view
x Flow correction Calculation
via cos(90 - c)
x Flow
c =
arccos
Radius

Figure 5: Rotation calculation via arctangent calculation with the flow values as opposite and the radius cross-section as the adjacent leg.

the values by that factor and recalculating the opposite side value. The arctangent of the fixed values represents the final rotation value.

4. IMPLEMENTATION

Our rotation estimation requires image processing steps and statistical calculations. Respective steps are described in section 3. For those tasks, we chose Python (v.3.7) as our environment to test our approach. For image processing, we rely on OpenCV (v.4.1.2.30) with the Python-Wrapper available via PyPI. Since there is no co-occurrence matrix calculation implemented in OpenCV, we additionally use Scikit-image (v.0.16.2) for calculating the matrix and deriving the homogeneity within the image from it. OpenCV additionally requires the NumPy library when used within the Python environment. NumPy (v.1.17.4) further provides the possibility for calculating quantiles and histograms. For the geometric calculation, we rely on the math implementation of the Python environment. We use the following OpenCV functions throughout the implementation: *findTransformECC, warpAffine, calcOpticalFlowFarneback.*

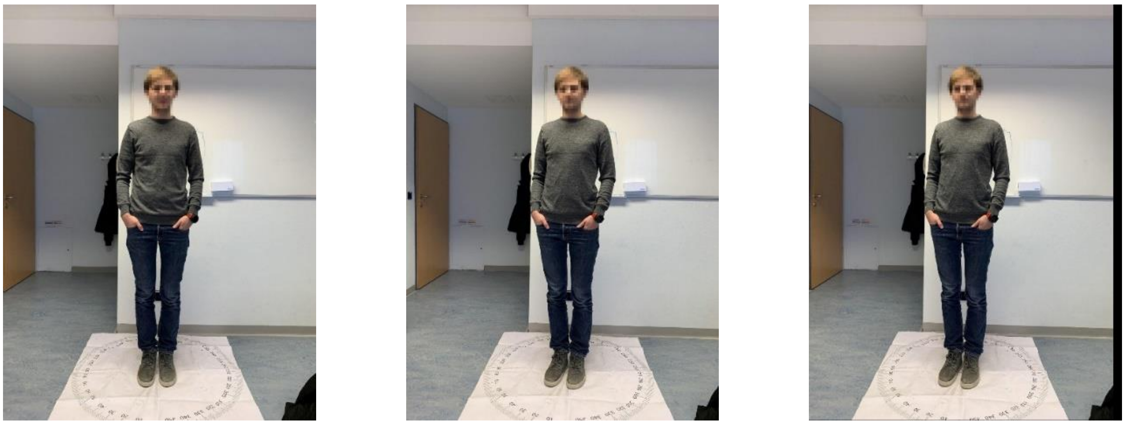


(a) Reference image with 0° rotation

(b) Target image with 5° rotation

(c) Target image with 5° rotation and alignment to the reference image, where no alignment is required due to the static mounted camera

Figure 6: Test sample with statically mounted monocular camera and no alignment required



(a) Reference image with 0° rotation

(b) Target image with 10° rotation

(c) Target image with 10° rotation and alignment to the reference image

Figure 7: Test sample with handheld monocular camera and alignment required, visible as black rectangles (c) due to the alignment

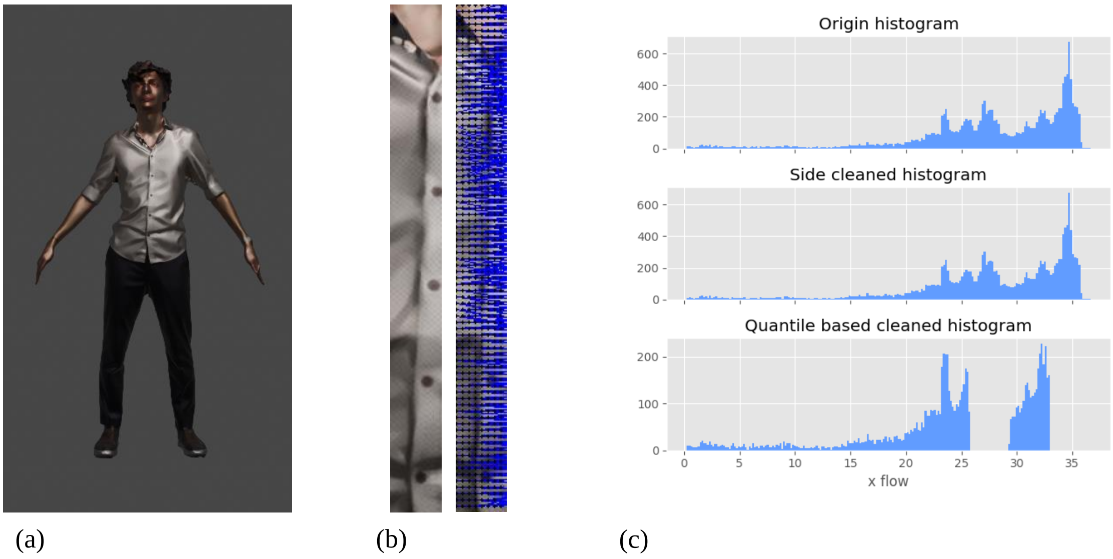


Figure 8: (a) CGI rendering sample, (b) ROI within the rendering and the corresponding optical flow values with a rotation to the right from camera view point, (c) density of all optical flow values and the remaining densities after the cleaning

5. RESULTS

Although the person is expected to rotate in T-pose, the follow test cases cover different kinds of poses to prove the applicability of the optical-flow based orientation estimation.

5.1 Alignment Texture Normalization

Figure 6 contains a test sample with alignment. This case shows that alignment is not required because the recording happened with a statically mounted monocular camera.

The scene in Figure 7 shows, that the alignment is required in the case of cameras handheld by a second person. The aligned target image is therefore moved to the left and a bit up, visible as black rectangles in the Figure 7 (c) The movement of the hand has been corrected to calculate the flow information from the same region of interest (ROI). Otherwise, two non-aligned areas could be compared with no significant overlap of information in the ROI.

Alignment via Enhanced Correlation Coefficient (ECC) estimates the required transformation based on maximizing the correlation coefficients between the reference and target image. The best fitting alignment in our case is with translation transformation because it does not distort the content of the target image.

5.2 Tests on CGI Renderings

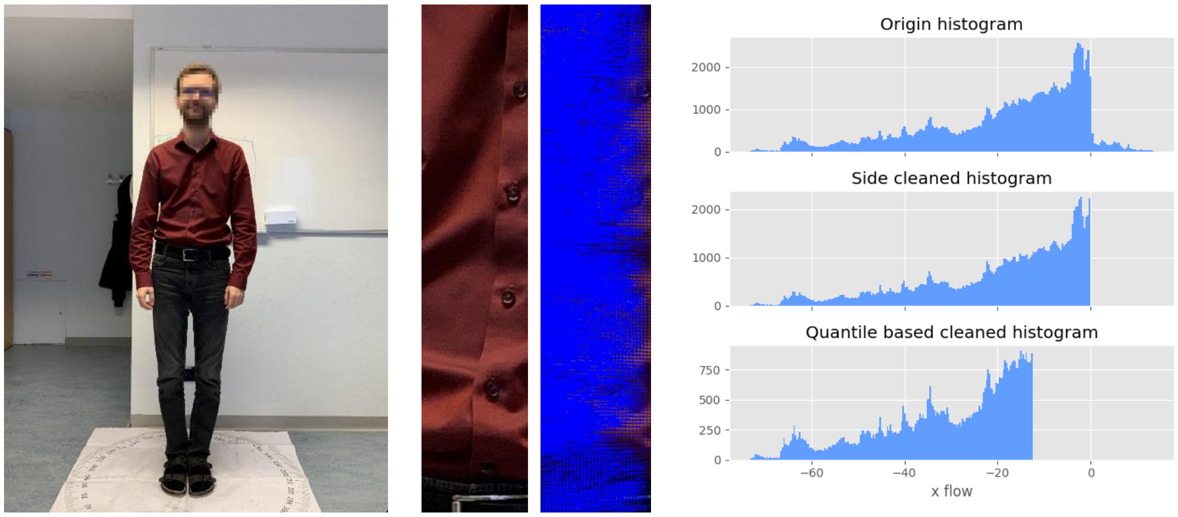
Figure 8 (a) shows the initial CGI scene. The ROI in the scene contains several features and has, therefore, a low homogeneity value of 0.267. The Figure 8 (b) contains the ROI clip out and a visualization of the optical flow. Figure 8 (c) contains the histogram of the optical flow values. It highlights a small density of noise in the range of 0 to 15. The mean is 27.97, and the median at 28.79. The calculated radius for this scene is 110. The calculation result is 14.73° with mean and 15.14° as the median value, after the correction. The expected target rotation for this sample was 15° and is consequently below the ±2° bound of the target rotation. The quantile-based cleaning, as the histogram shows in Figure 8 (c) worsens the result to a rotation value of 12.72°.

All synthetic scenario scenes results are shown in Figure 10. There are eight cases for 5°, four cases for 10° and one case for 15°. The 5° rotation are between (0°, 5°) and (5°, 10°). The 10° rotation are always between (0°, 10°). Further tests with less rotation in the range of 0° to 10° were tested with equal accurate results as the 5° target.

5.3 Tests on Real-World Recordings

The real-world recordings achieved similar results as with CGI renderings. Figure 9 shows an example from a real recording done with a handheld camera. The homogeneity in the ROI states a value of 0.487, which concludes to apply all preprocessing cleaning steps.

Figure 9 (c) highlights the steps of the cleaning process. First, the side cleaning and finally the quantile cleaning. The target rotation in this sample is 10°, starting at 0°. Without cleaning, a rotation of 6° can directly be calculated. After both cleaning steps and correction, the final rotation results in 10.05°, which represents a very accurate result. The median flow value is, therefore, 30.71, the median value 26.92, and the radius 176. The rotation based on the median value only achieves 8.8°.



(a) (b) (c)

Figure 9: (a) real-world recording with printed degrees below, (b) ROI within the recoding and the corresponding optical flow values with a rotation to the left from camera viewpoint, (c) density of all optical flow values and the remaining density

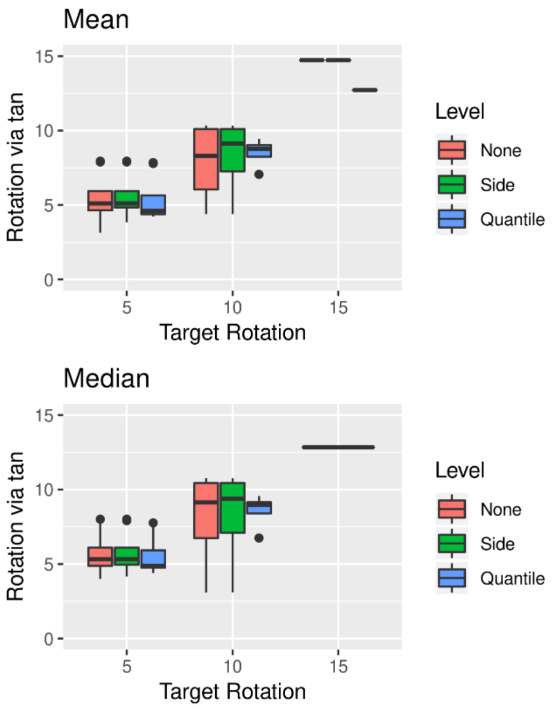


Figure 10: All synthetic scenario results with the cleaning steps and the resulting rotation by mean and median value from the optical flow.

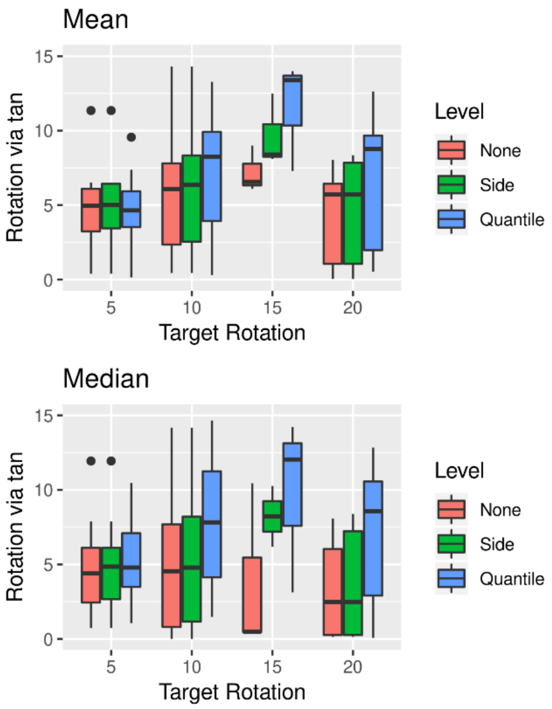


Figure 11: All real-world results with the cleaning steps and the resulting rotation by mean and median value from the optical flow.

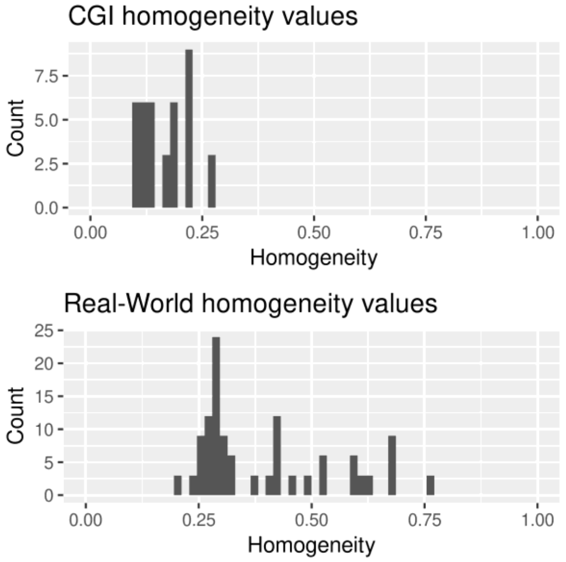


Figure 12: Homogeneity values between CGI rendering samples and real-world samples showing a significantly higher value.

All real-world results are summarized in Figure 11. The target rotation cases range from 5° to 20° with nine for 5°, 21 for 10°, three for 15° and seven for 20°. It is clear to see that with a rotation > 10° as target, the results get more incorrect. The samples itself inflict that because no feature information cannot be compensated like a too homogeneous ROI.

5.3 Homogeneity between CGI Rendering and Real-World Recordings

As expected, the homogeneity is utterly different between the CGI renderings and the real-world recordings, as visible in Figure 12. The CGI scenes usually have more features based on the perfect texture. The homogeneity values are significantly lower and centered at 0.167 and the real-world values at 0.396. In real-world recordings, it depends on the sharpness of the image, such as the focus, for example.

6. DISCUSSION AND CONCLUSION

The results show that the approach is working from controlled scenarios such as CGI rendering but also for real-world recordings. In all cases, it is visible that the rotation cannot be calculated as perfect as wanted in every scenario. Since the approach is relying on optical flow information, it is not possible to calculate a rotation higher than 45°. On the other side, the rotation calculation between contiguous frames from a video allows tracking of the rotation in a range lower than 10°. In this range, as stated, it is possible to extract the information at a high confidence.

We also tested scenarios with people wearing T-shirts with only horizontal textures. For those cases, the homogeneity is high, and by ignoring this information, the resulting rotation is far away from the desired target. That further concludes that reliable optical flow information is required. Without that, all pre and in-between cleaning steps do not automatically repair the calculated features. Therefore, it works not in every case, but with the described checks, it is possible to decide if the result is reliable or not.

Our simplified model of a person with a circle has some limitations. Naturally a person does not have a shape like this, but a more ellipse-like one instead, which has to be improved. Further, the rotation is typically not around the center of such a model. It is more likely around an asymmetric point between the center of our model and the backside. The first tests conclude that the resulting rotation can be more accurate by moving the center nearer to this desired arbitrary point.

7. OUTLOOK

Future tests will focus on the applicability of the presented orientation estimation for human shape reconstruction with real-world data. Due to the skeleton-based alignment, the presented accuracy should be sufficient with ±2° to allow for 3D human body reconstruction from a single static monocular video feed.

As the static monocular video acquires frames from one entire rotation of 360°, some potential for improvement as post-processing similar to VSLAM loop closure exists. If, e.g., the orientation evaluated from the neighboring frame pairs $(n_0, n_1), (n_1, n_2)$, …, (n_m, n_0) sums up to $360 + err$, then all frame-to-frame estimations can get scaled with $s = \frac{360}{360+err}$. Furthermore, the input video can be processed at various sampling rates, e.g., ten frames, 20 frames, and so on combining the particular results by applying maximum likelihood estimation for higher robustness.

8. ACKNOWLEDGMENTS

Our thanks to the Austrian Research Promotion Agency FFG for facilitating project TrueSize with basic funding program number 872105, fed by research budget of the Federal Republic of Austria.

9. REFERENCES

[Ben14a] Ben-Musa, A.S., Singh, S.K., and Agrawal, P. Suspicious Human Activity Recognition for Video Surveillance System. Proc. of the Int. Conf. on Control, Instrumentation, Comm. & Comp. Technologies (ICCICCT-2014), 2014.

[Cao17a] Cao, Z., Simon, T., Wei, S. E., & Sheikh, Y. Realtime multi-person 2d pose estimation using part affinity fields. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017.

[Che16a] Chen, A.-H., and Ramanan, D. 3D Human Pose Estimation = 2D Pose Estimation + Matching. CoRR, 2016.

[Cho16a] Choi, J.C., Lee, B-J., and Zhang, B.-T. Human Body Orientation Estimation using Convolutional Neural Network. CoRR, 2016.

[Der17a] Derkach, D., Ruiz, A., and Sukno, F.M. Head Pose Estimation Based on 3-D Facial Landmarks Localization and Regression. 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), 2017.

[Far03a] Farneback, G. Two-Frame Motion Estimation Based on Polynomial Expansion. Proce. of the 13th Scandinavian Conf. on Image Analysis, 2003.

[Ger10a] Geronimo, D., Lopez, A.M., Sappa, A.D., and Graf, T. Survey of pedestrian detection for advanced driver assistance systems. IEEE Trans Pattern Anal Mach Intell., vol 32(7), 2010.

[Gou04a] Gourier, N., Hall, D., and Crowley, J.L. Estimating Face orientation from Robust Detection of Salient Facial Structures. FG Net Workshop on Visual Observation of Deictic Gestures, 2004.

[Kell08a] Kellokumpu, V., Zhao, G., and Pietikäinen, M. Human Activity Recognition Using a Dynamic Texture Based Method. Proc. of the British Machine Conf, 2008.

[Lap08a] Laptev., I., Marszalek, M., Schmid, C., and Rozenfeld, B. Learning realistic human actions from movies. Proc. of the IEEE Comp. Soc. Conf. on Comp. Vision and Pattern Recog, 2008.

[Li14a] Li, S., and Chan, A.B. 3D Human Pose Estimation from Monocular Images with Deep Convolutional Neural Network. Asian Conf. on Comp. Vision (ACCV), 2014.

[Pra20a] Praschl, C., Krauss, O., Zwettler, G. Enabling Outdoor MR Capabilities For Head Mounted Displays: A Case Study. International Journal of Simulation and Process Modelling 2020.

[Rho18a] Rhodin, H., Salzmann, M., and Fua, P. Unsupervised Geometry-Aware Representation for 3D Human Pose Estimation. CoRR, 2018.

[Ron16a] Ronao, C.A., and Cho, S.-B. Human activity recognition with smartphone sensors using deep learning neural networks. Expert Systems with Applications, vol 59, 2016.

[Singh08a] Singh, M., Basu, A., and Mandal, M. Human Activity Recognition Based on Silhouette Directionality. IEEE Trans. on Circuits and Systems for Video Technology, 2008.

[Sug05a] Sugimoto, A., Kimura, M., and Matsuyama, T. Detecting human heads with their orientations. Electronic Letters on Computer Vision and Image Analysis 5(3), 2005.

[Tos14a] Toshev, A., and Szegedy, C. DeepPose: Human Pose Estimation via Deep Neural Networks. IEEE Conference on Computer Vision and Pattern Recognition, 2014.

[Wei19a] Wei, G., Lan, C., Zeng, W., and Chen, Z. View Invariant 3D Human Pose Estimation. CoRR, 2019.

[Xia18a] Xiang, Y., Schmidt, T., Narayanan, V., and Fox, D. PoseCNN: A Convolutional Neural Network for 6D Object Pose Estimation in Cluttered Scenes. CoRR, 2018.

[Yan98a] Yaniz, C., Rocha, J., and Perales, F. 3D Part Recognition Method for Human Motion Analysis. Proc. of the Int. Workshop on Modelling and Motion Capture Techniques for Virtual Environments, 1998.