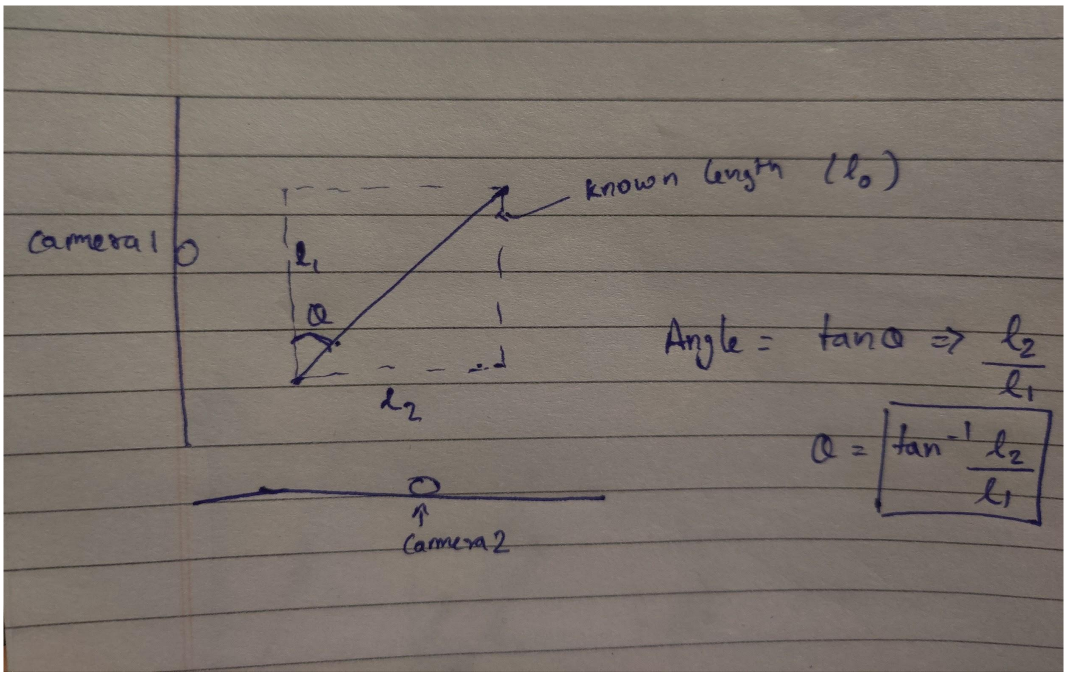
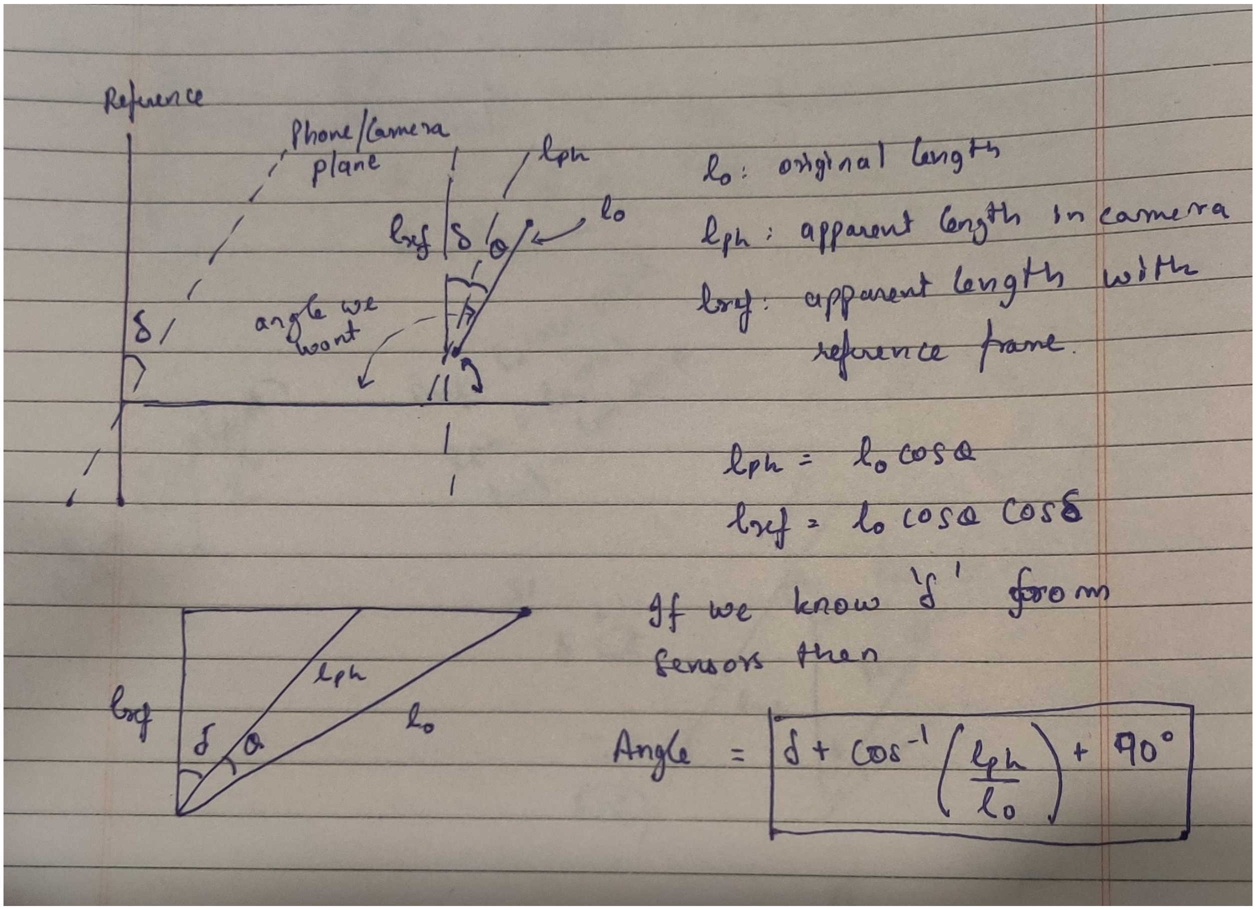
After researching the topic for some time, I have identified three potential ways to address the problem of estimating the angle of a skeleton using two cameras:

1. <u>Using Two Camera configuration</u>: In the figure shown, two cameras are used to estimate the angle of a skeleton. One camera views the skeleton from one angle, while the other camera views it from a different angle. The angle estimate is calculated by taking the arctan of the ratio of the apparent lengths of the skeleton as seen by each camera. However, this method relies on the assumption that the two cameras are perfectly perpendicular to each other and are well-aligned with a reference frame. This may not be practical in real-world situations, as it would be difficult for a user to set up two cameras in this manner and ensure their proper alignment.



2. <u>A single camera configuration</u>: Instead of a two camera configuration, a single camera can also work, using some sensor data. First, we need to find the true length of the skeleton(L0). Next, I choose a reference object such as a wall, let the orientation angle of the device with respect to the wall be 'δ'. Using the technique mentioned above, we can measure the length of the object with respect to the device, let the angle be 'θ'. The total angle that should be estimated will be therefore '$\theta + \delta$'. Combining data from gyroscope and accelerometer[1] it is feasible to compute 'δ' with some error due to the sensor but using this we can trigonometrically determine the angle of the skeleton.



3. Machine Learning based angle estimator: Aberman et al [2] makes a component that encodes the view-angle.In this work, it is assumed the component is static. Once the latent components are extracted they are utilized to yield new motions, allowing a loss to be computed and optimized. They implicitly learn to cluster motions in the dynamic latent space, where each cluster consists of similar motions performed by different individuals. Another interesting paper by Xu et al[3] found that the human visual system processes the angle feature in two stages: encoding the orientation of the bounding lines and combining them into an angle feature and by estimating the angle in an orthogonal internal reference frame. At last, Wayfair tech blog[4] had an interesting architectural design: quasi-VGG network that predicts a numerical value of the angle that takes 3 inputs: image along with sine and cosine values of shot angle.

References:

[1] Carlson A. Tracking Pose Using Common Mobile Phone Sensors (https://digitalcommons.bard.edu/cgi/viewcontent.cgi?article=1208&context=senproj_s2018)
[2] Aberman et. al. Learning Character-Agnostic Motion for Motion Retargeting in 2D (https://arxiv.org/pdf/1905.01680.pdf)
[3] Xu et. al The human visual system estimates angle features in an internal reference frame: A computational and psychophysical study. [https://jov.arvojournals.org/article.aspx?articleid=2718450]
[4] Wayfair: (https://www.aboutwayfair.com/tech-innovation/shot-angle-prediction-estimating-pose-angle-with-deep-learning-for-furniture-items-using-images-generated-from-3d-models)