Monocular 3d pose estimation for humans

Viktor Pantushev

29. Februar 2024

Contents

1 Summary

This paper presents a reconstruction generated from 2D data for monocular 3D motion analysis. The clever application of trigonometric formulas eliminates bottlenecks and inefficiencies in existing algorithms and data structures.

The results are particularly promising for the field of medicine (in developing countries). This also includes physiotherapy services and the correction of movement exercises in the medical and sports sectors.

2 Introduction

Motion analysis software plays a central role in various areas and facilitates tasks ranging from performance evaluation in sports to the animation of virtual figures. Despite significant advances in the field, optimizing the 3D performance of such software remains a critical challenge. In this paper, a strategy for improving efficiency while maintaining the same hardware is investigated.

2.1 Background and Motivation

The demand for sophisticated tools for motionanalysis has risen sharply with technological advances and increasing applications in all areas. Despite the increasing complexity of software, optimization without compromising accuracy is a major challenge. This is a reason for a closer examination of methods to improve the 3D performance of motion analysis software.

3 Methodology

The latest development in 3D motion analysis works with the help of heat maps. This method combines 2D motion analysis with an estimation of how large a body part appears on the image.

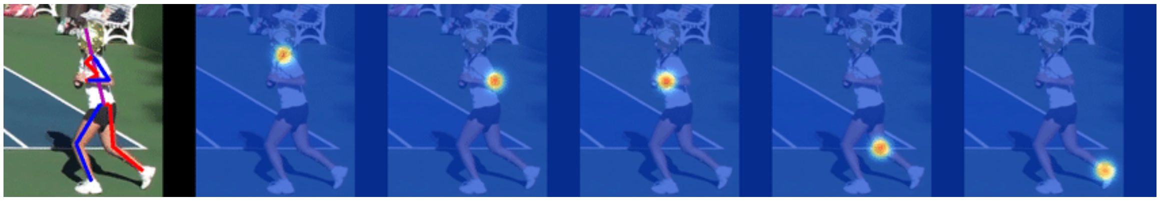


Figure 1: Heatmap Visualization
Source: `http://doi.org/10.1016/j.jvcir.2021.103055`

4 Development and implementation of mathematical optimizations

4.1 Alpha angle calculation

It is assumed in the work that each part of a body is connected to another and is fixed at a certain point, i.e. a circle is the maximum movement when the movement is parallel to the camera. This is shown graphically below:

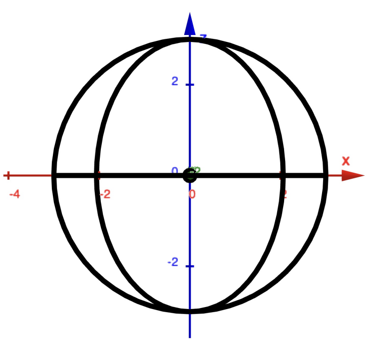


Figure 2: Sideview

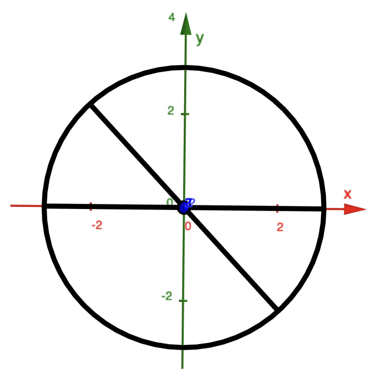


Figure 3: Birdview

Figure 1 shows that a circle becomes an ellipse when it is rotated towards the camera. Viewed from above, it can now be calculated using the cosine.

$$cos(\alpha) = Ankathete/Hypothenuse$$
$$\alpha = arccos(Ankathete/Hypothenuse)$$

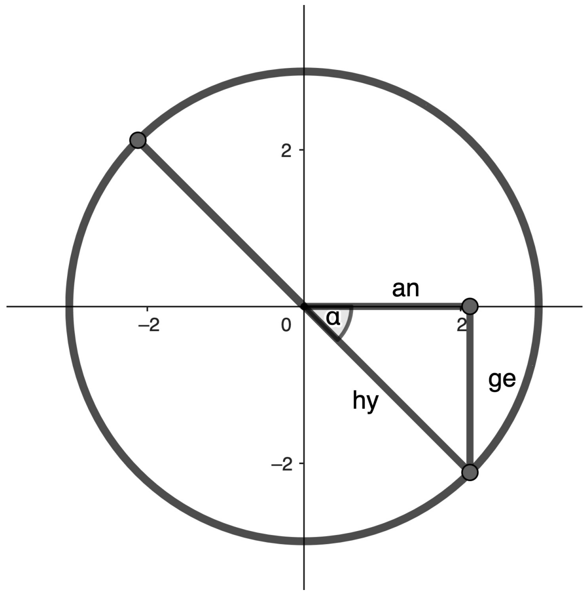


Figure 4: Bird's eye view (and cosine display)

Now the following problem is encountered: the shortest distance from the center of the ellipse cannot be determined in Figure 1. This means that the arc cosine function cannot yet be formed. Instead, however, we know the X and Y values noted in the camera image (distance EF and FM, where M is the center [0,0,0]):

You can also see the distance EF and GH, which is the same distance from the center point M when viewed from above. An important observation is that G and E are equidistant from the center point in the bird's eye view and that |EF| = |GH| applies.

$$|\vec{EF}| = |\vec{GH}|$$

Since EF = Y-value of the camera image

$$y = \vec{EF}$$

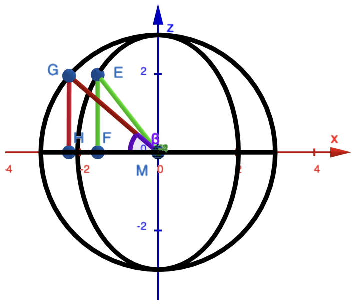


Figure 5: Sideview

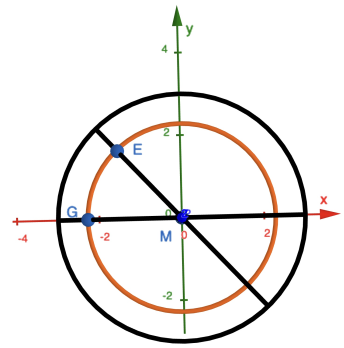


Figure 6: Birdview

Another finding is that the X value of the camera image can be read directly. This is FM.

$$x = F\vec{M}$$

Ebenfalls wissen wir den Radius r des originalen Kreises mit Mittelpunkt M:

$$r = |G\vec{M}|$$

Now HM can be calculated using the Pythagorean theorem:

$$r * r = |G\vec{H}|^2 + |H\vec{M}|^2$$

$$r * r = |\vec{Y}|^2 + |H\vec{M}|^2 \quad \text{Da |GH| = |EF| = Y-Wert}$$

$$|H\vec{M}| = \sqrt{r^2 - |\vec{Y}|^2} \quad \text{Since r and Y are known, now also known to |HM|}$$

Now the arc cosine can be calculated:

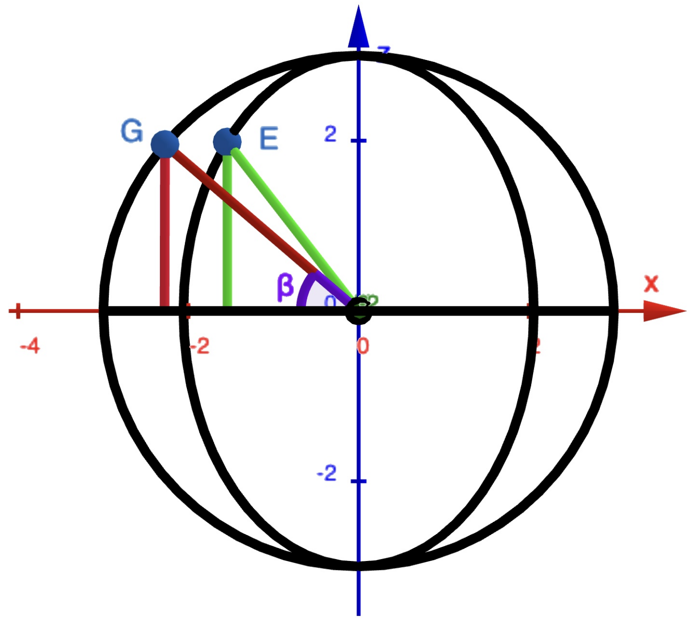
$$\alpha = arccos(Ankathete/Hypothenuse)$$

$$\alpha = arccos(\frac{|E\vec{F}|}{|H\vec{M}|})$$

$$\alpha = arccos(\frac{X}{|H\vec{M}|}) \quad \text{da: } X = |E\vec{F}|$$

$$\alpha = arccos(\frac{X}{\sqrt{r^2 - Y^2}}) \quad \text{da: } |H\vec{M}| = \sqrt{r^2 - |\vec{Y}|^2}$$

4.2 Beta angle calculation



$$tan(\beta) = \frac{|\vec{HM}|}{|\vec{GH}|}$$

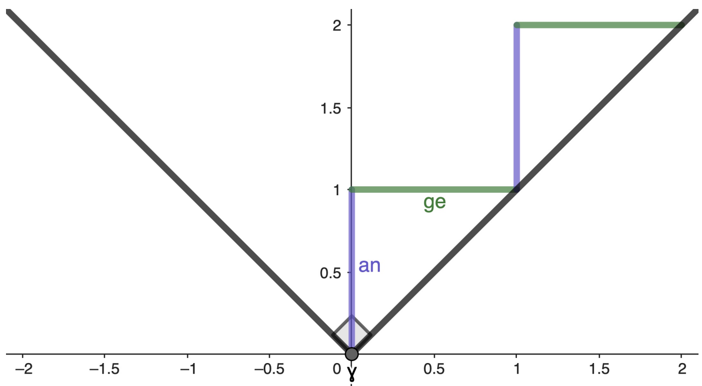
$$\beta = arctan(\frac{|\vec{HM}|}{|\vec{GH}|})$$

Bzw.:

$$\beta = arctan(\frac{\sqrt{r^2 - X^2}}{Y})$$

4.3 Calculation of body parts that are closer to the camera

It is easy to calculate the shoulder and hip angles as these are connected to the back and the distance to the camera can be measured using the length of the back. The parallel value of a body part is determined when the person is parallel to the camera. For all body parts indirectly connected to the back, this parallel value must be reduced or increased, depending on whether the body part is closer or further away. To do this, the distance must be measured, adjusted and offset against the camera angle in order to multiply the magnification. The image size on the camera sensor is indirectly proportional to the distance from the sensor. The imaging angle of the camera (gamma) is an essential part of this. The proportion of the body part imaged by the camera in relation to the camera is indirectly proportional to the overall image, as this becomes smaller at equal distances from the camera.



Actually:

$$Gegenkathete = Ankathete * tan(\frac{\gamma}{2})$$

If anathema is defined as 1:

$$Gegenkathete = tan(\frac{\gamma}{2})$$

Since there is an indirect proportionality:

$$Entfernung = \frac{k}{AnteilAmBild}$$

Distance difference w:

$$w = cos(\alpha) * \frac{cos(\beta)}{LaengeDesKoerpers}$$

Share a of w in the total image:

$$a = \frac{k}{AnteilAMBild - w}$$

Factor f between w and distance:

$$f = \frac{Entfernung}{a}$$