

POLITECNICO DI MILANO

DEPARTIMENTO DI ELETTRONICA, INFORMAZIONE E BIOINGEGNERIA

IMAGE ANALYSIS

**3D VISUAL BODY SCAN**

Méndez Méndez, Jorge Armando: 835814

Rodríguez Sánchez, Rafael Alberto:

Milano, July 2015.

**Contents**

[Problem Formulation 3](#_Toc423029011)

[Solution Approach 3](#_Toc423029012)

[Obtention of World Information 3](#_Toc423029013)

[Three Images Approach: General Description 3](#_Toc423029014)

[Computation of Reference World Information 4](#_Toc423029015)

[Computation of World Information of Specific Frame 4](#_Toc423029016)

[Obtention of Rotation Information 5](#_Toc423029017)

[Obtention of Image Points 5](#_Toc423029018)

[Obtention of Convex Hull 5](#_Toc423029019)

[Results 5](#_Toc423029020)

[Obtention of World Information 5](#_Toc423029021)

[Computation of Reference World Information 5](#_Toc423029022)

[Computation of World Information of Specific Frame 5](#_Toc423029023)

[Obtention of Rotation Information 5](#_Toc423029024)

[Obtention of Image Points 5](#_Toc423029025)

[Obtention of Convex Hull 5](#_Toc423029026)

[Analysis 6](#_Toc423029027)

[Obtention of World Information 6](#_Toc423029028)

[Computation of Reference World Information 6](#_Toc423029029)

[Computation of World Information of Specific Frame 6](#_Toc423029030)

[Obtention of Rotation Information 6](#_Toc423029031)

[Obtention of Image Points 6](#_Toc423029032)

[Obtention of Convex Hull 6](#_Toc423029033)

[Conclusions and Recommendations 6](#_Toc423029034)

[References 7](#_Toc423029035)

# Problem Formulation

The general goal of this project is to obtain three dimensional measures of a person’s body, in order to be able to customize clothing items for them. This way, a person’s body measures could be obtained in the distance, thus eliminating the need for tailors to meet directly with their customers.

In order to achieve this, a video of a person rotating around their own axis is obtained. The height of such person should be specified, so that real world information can be acquired. The area of the person corresponding to their torso will be studied over all the frames, obtaining measures that can be of interest at the time of computing a 3D reconstruction of the abovementioned section of the body.

# State of the Art

asdasd

# Solution Approach

The solution sought starts from an already internally calibrated camera. The height of the person must be introduced. The inputs of the algorithm are:

* Three photos of the person:
  + One at a certain distance from the camera, close to the center of the image.
  + One at a distance , shifted to the left of the center of the image.
  + One at a distance , shifted to the right of the center of the image.
* A video of a person rotating around its own axis:
  + The person must hold a OBJECT over its head at all times.
  + The camera must be held perpendicular to the floor at all times.

The background of the video and images must be uniform and there should be enough contrast to be able to detect edges. The OBJECT must be of a distinct color, so that color segmentation successfully detects it.

From this point on, the solution is separated into four different modules, each of them responsible, respectively, of: obtaining world information, obtaining rotation information, obtaining image points and obtaining the 3D reconstruction of the body.

## Obtention of World Information

In order to obtain real world information, a specific module is implemented. This whole module is based on an approach which requires three images of a single person at different positions to be acquired.

### Three Images Approach: General Description

The idea of using three images of the person is the following. A person stands at a certain distance from the camera, aiming to be close to the center of the image, and a picture of this image is acquired. Then, the person takes a few steps forward and to the left to obtain a second image, and then by taking a few steps to the right, the final image is taken. Assuming the camera is completely static during this procedure, the three images can be overlapped obtaining a single image in which an object with the same measures stands at different positions.

For obtaining this overlapped image, a Canny edge detector is used. The thresholds for the edge detector are manually selected by the user via a simple user interface with a track bar. Once all three images have been converted to Boolean values, they are all logically summed (via the *or* logical operator). Ideally, the obtained overlapped image should look like the one observed in Figure 1.

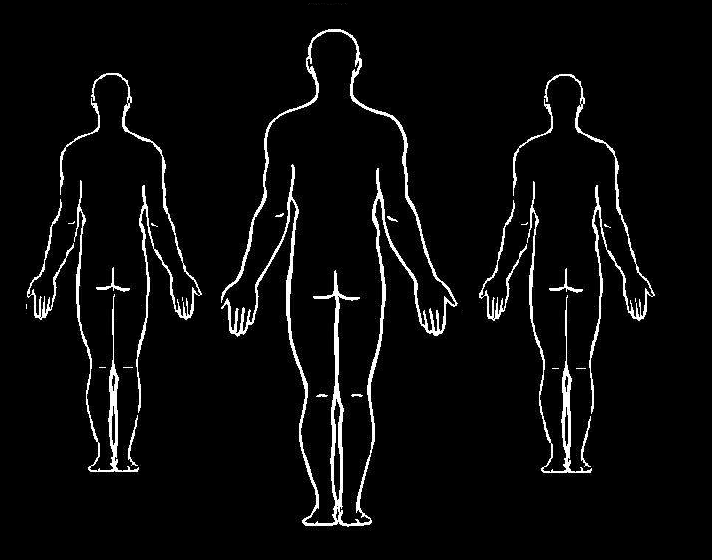


Figure 1. Ideal overlapped image.

The obtained overlapped image is displayed to the user, who is asked to select an area for each of the heads and each of the feet, in a specific order. Once this is done, it is possible to compute the real plane that is parallel to the floor plane and goes through the head of the person, in 3D coordinates with respect to a reference frame fixed in the camera center. Figure 2 illustrates the plane on which the top of the head lies, whichever the person’s location might be.

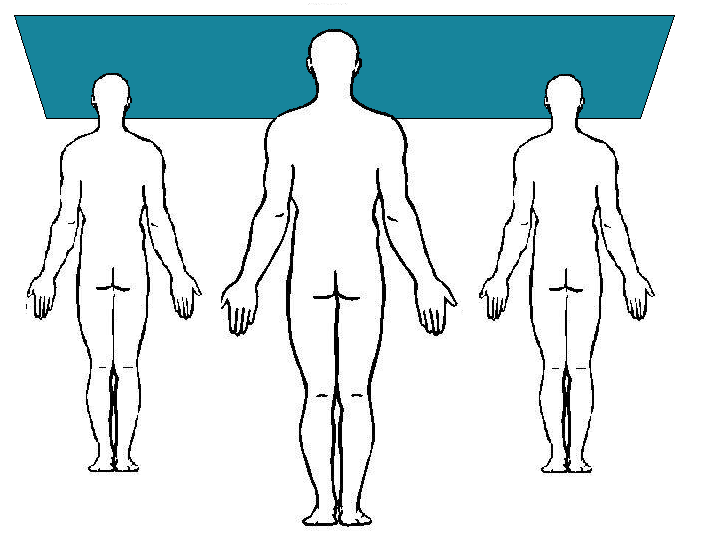


Figure . Plane that goes through three heads on the overlapped image.

### Computation of Reference World Information

In order to obtain the plane described above, some simple geometrical computations are required. The first step is to compute the real 3D points in world coordinates, corresponding to the heads and feet of the person (since the and coordinates are not relevant, either of the feet can be chosen). In order to do this, the known height of the person is used to compute the distance from each of the persons to the camera center. Observing the triangles formed by the person’s real height, the person’s height in the image, the distance to the person and the focal distance in Figure 2, it is easy to obtain that:

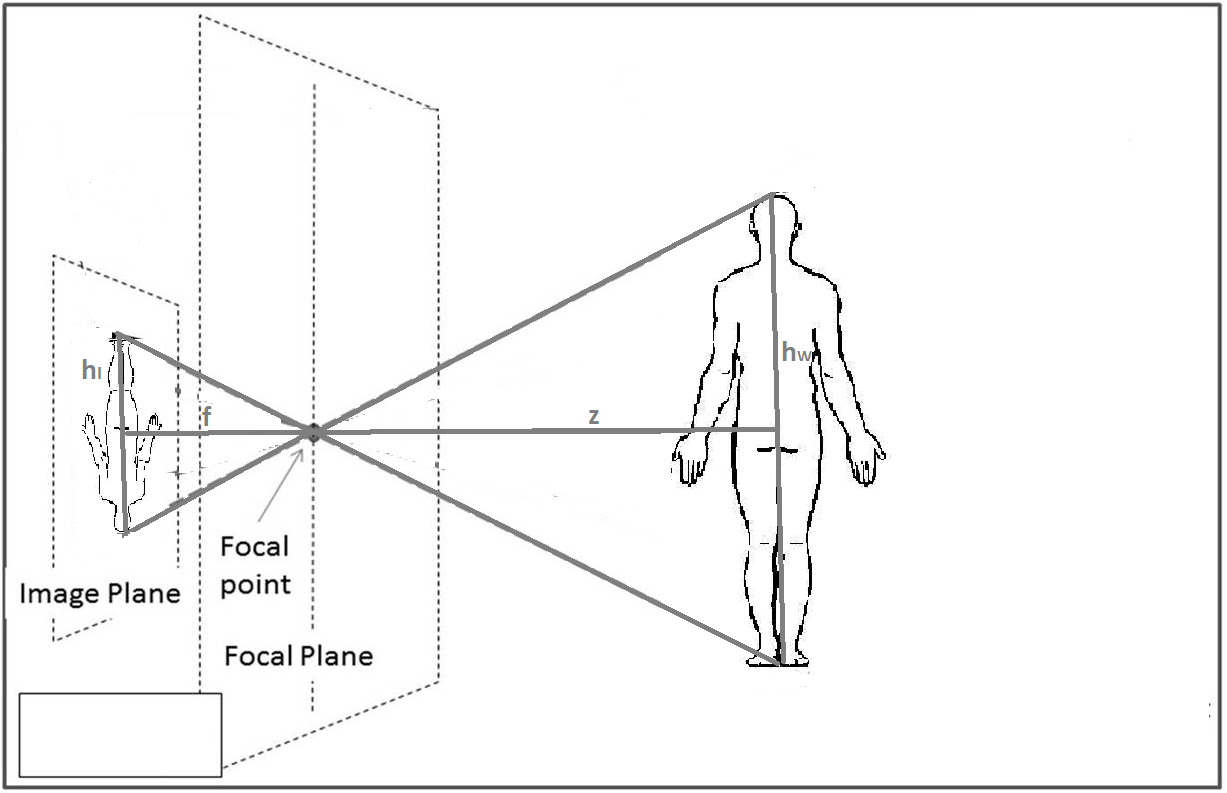


Figure 3. Transformation of the person's height onto the image.

On the equation above, and are the height of the person in world and image coordinates, respectively, is the focal distance of the camera, and is the obtained coordinate of the person. is obtained by subtracting the image coordinates of the points corresponding to the head and to the feet of the person.

After this point, an plane is computed with constant value . The line corresponding to the image ray coming from each of the points is also computed, and then intersected with the plane, thus obtaining the remaining coordinates. Maybe another figure.

The image rays are computed assuming that the center of the camera is the origin of the coordinate system, and that there is no rotation. Hence, the formula applied is the following, where is the matrix of intrinsic parameters of the camera, and is the image point in homogeneous coordinates:

On the other hand, the intersections with the planes are

Once all these points are obtained, the desired plane can be obtained, described as its normal vector and a point on the plane . Given the three points that correspond to the head of the person (, and ), and the normal vector can be obtained as:

This plane is very useful since, even though the person might move back and forth, left and right, during the video, the highest point on their head will always lie on this plane.

### Computation of World Information of Specific Frame

Based on the information obtained up to this point, three functionalities are added to obtain real world information from each of the frames in the video. The first of them consists con calculating the *distance to the person*, and their displacement with respect to the center, in each image. In order to obtain this information, the image line coming from the image ray corresponding to the highest point in the head of the person is intersected with the plane parallel to the floor at the height of the head, obtaining the point . The coordinate of the obtained point is returned as the distance to person. The coordinate is returned as the displacement of the person with respect to the center. Figure

The second implemented functionality is to compute the *height of a point with respect to the head*. To do this, the point is computed in the same way as before. Then, the line coming from the ray corresponding to the point is intersected with the XY plane at constant value , obtaining the point . The distance in , , is returned as the height of the point with respect to the head. Figure

The last implemented functionality is to obtain the *image coordinate at a given height from the head* in the real world. The first step for this is, again, obtaining the point . Then, a point at the desired height from the head is computed as . This point is then converted to image coordinates assuming, again, that the world reference is in the same as the camera reference. Thus, the computed point is , and the value is returned as the desired image coordinate. This, together with the previous functionality, is very useful at the moment of obtaining points at the same height level of the person on every frame. Figure

## Obtention of Rotation Information

### Detection of Rotation Information Object

Detection in the image: color, corners of color segment.

### Computation of Rotations

The length of the object in the image is computed for each of the studied frames, storing the largest value. This maximum length corresponds to the image with 0° rotation. Then, it can easily be seen that, on every frame, the cosine of the rotation angle is the ratio of the current length to the maximum length observed for the object. This comes from considering that the object observed in the image is a projection of the real object. Thus, we can obtain the rotation from the triangle observed in Figure 1, by considering that the maximum length is the real length of the object projected onto the image without any rotation.

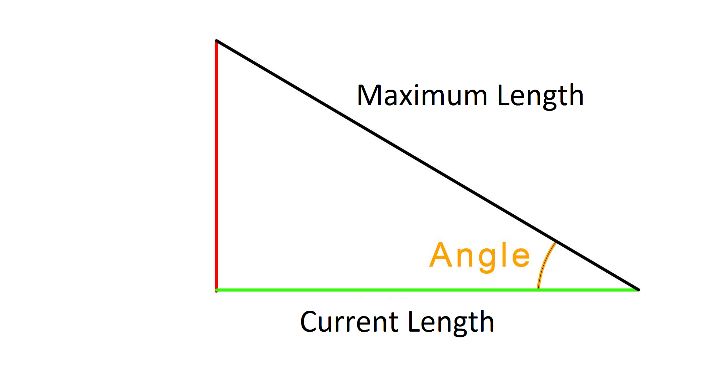


Figure 4. Projection of rotation information object onto image plane.

In the case in which the object is not correctly detected, the rotation is obtained by assuming uniform rotation between the last and the next computed frames.

## Obtention of Image Points

The first step toward obtaining the interest points in the image is computing the Region Of Interest (ROI). For this, the four corners of a rectangle are defined in the following way. The leftmost coordinate () is of the image width, and the rightmost one () is of the width. This gets rid of possible edges on the borders of the images, and focuses only on the central area, where the person is expected to be.

At this point, an uppermost coordinate is defined at of the image height to feed this information to the head finding procedure, which will be explained later. Then the head is found, and the uppermost coordinate () is recomputed as the one at from the head, and the lowermost one () is computed as the one at from the head.

The defined ROI is used to find two types of interest points. The first one is the head of the person, which is found by horizontally scanning the edges of the image from to starting at the coordinate at of the image, to ignore the upper border. The algorithm is stopped either when the first white pixel is found, which corresponds to the highest one and thus the top of the head of the person, or no pixel is found before arriving to the coordinate . This is the stopping condition because is at the beginning of the ROI, which is by definition below the head, since it is on the body.

The second type of interest points found is the silhouette points. The first step for doing this is updating the ROI to match the original heights used in the first frame. For this, the height from the head in world coordinates is computed for both and from the first frame, obtaining and . Then, the new and are obtained by finding the image coordinate on the new image at and from the head, respectively.

Once the ROI is updated, its height is subdivided in as many levels as specified in code. Then, for every level, the edges of the image are scanned horizontally from left to right, from to the first white pixel, and then from right to left, from to the first found white pixel. These two points represent the outermost points of the person in a given height, thus by finding them at every height level, the silhouette of the person can be approximated. If as much as one point in one level is not found, then the image is discarded.

## Obtention of 3D Reconstruction

May also be simple enough for no subtitles. Computation of camera centers, rotation of the rays, projection of the lines, intersections…some sort of point selection.

In order to compute the 3D reconstruction of the body, a convex hull approach is sought. The idea consists in assuming the person is static and the camera is rotating about the axis that goes vertically through the center of the person. Then, for each level, the lines corresponding to the image rays of the two points coming from each of the rototranslated cameras, are projected onto the plane that transversely cuts the person at the given height.

As seen on Figure, the computed projected lines correspond to the lines tangent to the person. Hence, by obtaining enough lines and intersecting them appropriately, the silhouette of the person at a given height can be well approximated. Two intersection points are computed for each pair of consecutive images, on each level. The first one, between the leftmost ray of the first image and the leftmost ray of the second image; the second one, between the rightmost ones. As can be seen in figure, these points are the vertexes of the polygon that most fitly encloses the silhouette of the person, given the amount of frames used.

In order to compute the rototranslation of the camera centers, information of the distance from the camera to the person, the displacement of the person with respect to the center of the camera, and the rotation angle of the person is used. RIGID TRANSFORMATION, P MATRIX, IDENTITY…

Thus, the translation vector and rotation matrix obtained are the following:

The rototranslation obtained from them is applied to the center of the camera described by removing the effect of the movement of the person front or back, and left or right:

On the above equations, is the distance from the camera to the person on the current frame, is the rotation angle with respect to the (vertical) axis, is the displacement of the person with respect to the camera center in the axis, and is the difference between and the distance from the person to the camera center in the first frame.

# Results

## Obtention of World Information

Figure 2 shows the overlapped image obtained for the dataset used.

a

Figure 5. Obtained overlapped image.

### Computation of Reference World Information

The normal vector and point on plane that describe the obtained result for the plane that always contains the head are as follows:

Computing the inclination of this plane with respect to a plane horizontal in the reference frame used, that is, a plane with normal vector , the obtained result is:

### Computation of World Information of Specific Frame

To test results of the computation of the distance from the camera to the person, said distance was measured on four images: the three corresponding to the overlapped image, and the first one of the rotating video. In every case, the real distance was also measured, obtaining the results computed in Table 1.

|  |  |  |
| --- | --- | --- |
|  | Measured Distance [cm] | Computed Distance [cm] |
| Center Image |  |  |
| Left Image |  |  |
| Right Image |  |  |
| First Frame |  |  |

Table 1. Distance to Person Results.

Similarly, to test the result of the computation of the height with respect to the head, the height from head to feet was computed on the three images corresponding to the overlapped image, and another one from the video with rotation close to 180°. The results, along with the real height of the person, are shown in Table 2.

|  |  |
| --- | --- |
|  | Computed Height [cm] |
| Center Image |  |
| Left Image |  |
| Right Image |  |
| 180° Image |  |
| Real Value |  |

Table 2. Height from Head to Point Results.

Later on, the computation of the image coordinate at a given height from head was tested on the same images as the previous case, finding the point at full height of the person from their head. Table 3 contains the values of the coordinates obtained for manually selected feet points and the corresponding computed coordinate for the person’s full height.

|  |  |  |
| --- | --- | --- |
|  | Original image  coordinate | Computed image  coordinate |
| Center Image |  |  |
| Left Image |  |  |
| Right Image |  |  |
| 180° Image |  |  |

Table 3. Image Y Coordinate at given Height Results.

## Obtention of Rotation Information

The images shown in Figure 3, Figure 4, Figure 5, Figure 6 and Figure 7 show different frames of the video, and the rotation angles obtained for each of them.

a

Figure 6. Obtained rotation angle (1).

a

Figure 7. Obtained rotation angle (2).

A

Figure 8. Obtained rotation angle (3).

A

Figure 9. Obtained rotation angle (4).

A

Figure 10. Obtained rotation angle (5).

## Obtention of Image Points

The images shown in Figure 8, Figure 9, Figure 10, Figure 11 and Figure 12 display small dots on every computed point, both for the silhouette points and for the head point, for a set of images.

A

Figure 11. Obtained image points (1).

A

Figure 12. Obtained image points (2).

A

Figure 13. Obtained image points (3).

A

Figure 14. Obtained image points (4).

A

Figure 15. Obtained image points (5).

## Obtention of 3D Reconstruction

Figure 13 shows the projected image rays for one height level of the person, to show how they intersect with one another. Then, Figure 14 shows the obtained points from the computation of the convex hull for that same height level.

A

Figure 16. Projected image rays for one level.

A

Figure 17. Camera centers projected on plain in red, points computed for convex hull approximation in blue.

Finally, Table 4 holds the results for the perimeter measures obtained for several height levels; that is, the final body measures obtained from the video for the 3D reconstruction of the body. These would be the values to use in the design of clothing for a given person. The real values, measured directly on the person, are also shown for comparison purposes.

Table 4. Perimeter measures results.

|  |  |  |
| --- | --- | --- |
| Height Level | Measured Perimeter [cm] | Computed Perimeter [cm] |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

# Analysis

## Obtention of World Information

### Computation of Reference World Information

Just include a bunch of percentage errors on every subtitle.

### Computation of World Information of Specific Frame

.

## Obtention of Rotation Information

Observing the obtained results for the rotations, it can be seen that they hold some important errors. On the first frames, the angles seem quite accurate and show results that, at least to the eye, seem to work. However, as the person begins to rotate and the stick used to compute rotations begins to move away from the camera, results begin to deviate from what was expected.

## Obtention of Image Points

Observing the results obtained for the finding of interesting points on images, it could be seen that for almost every frame used, the image points were found without problems. The point that was found with the least accuracy was the one corresponding to the head, which was sometimes obtained with some displacement from the real center of the head.

It is also important to note that on some frames it was impossible to find the desired points, since there were no edges obtained for the interest sections of the body.

Another point to be noticed is that, among all different frames, the image points computed for the silhouette of the body all correspond to roughly the same positions on the body. That is, they all start just below the arms and go down up to the hips.

## Obtention of 3D Reconstruction

.

# Conclusions and Recommendations

General overall conclusions about results, about positive and negative aspects, about results worth saving for future. Recommendations about paths to follow, different methods to study that might throw better results, about further computations that could have been made, about simplifications of use (such as automations) that should be applied, etc.