HEAD TRACKING FOR 3D AUDIO USING THE NINTENDO WII REMOTE

Mauricio Ubilla
School of Enginering
Department of Computer Science
Pontificia Universidad Catolica de Chile

Domingo Mery
School of Enginering
Department of Computer Science
Pontificia Universidad Catolica de Chile

Rodrigo F. Cadiz
Centro de Investigacion en Tecnologias de Audio, Music Institute, and
Department of Computer Science
Pontificia Universidad Catolica de Chile

ABSTRACT

Head Tracking systems with head pose estimation have several applications in human-computer interaction and constitute a key component of 3D audio environments. Most developed systems to date are restricted to a range of poses in which the user has to be facing the camera, are sensible to the lightning conditions and are not identity-invariant. These constrains make these approaches not very appropriate for 3D audio. In this work, we propose a head tracking system with accurate head-pose estimation in the infrared spectrum that uses the Nintendo Wii Remote to track a circlet with infrared diodes. The system doesn't restrict the user to be facing the camera, is less sensitive to the lighting conditions and is completelly identity-invariant. Additionally, the system eliminates the camera inclination using the accelerometer data of the device and the head coordinates are transmitted over the network using the OSC protocol, allowing communication with most of the 3D audio engines available today.

1. INTRODUCTION

Head Tracking (HT) refers to the process of detecting a human head in a video sequence, ideally determining its position and orientation in space. In this way, there are different levels of HT, that range from the simple detection of the user's head (usually achieved with *Face Detection* [1]) to the estimation of its orientation or pose (topic also known as *Head Pose Estimation* [2]).

HT has been well studied in the last 15 years [2], mainly because of:

- The great interest in its applications in Human-Computer Interaction (HCI). During the last 20 years humans have interacted with computers mostly through a keyboard and a mouse. HCI is a research area that looks to develop new ways of interaction. HT is part of a set of these new ways of interaction, known as Vision-Based Interfaces (VBIs) [3].
- The massification of web cams, which makes this technology more accessible. Nowadays, these systems can be used by any person.

Most of the devoloped HT techniques estimate the user's head orientation from a single view. They can be classified in two categories [3]:

- Appeareance-based methods: obtain the information directly from the acquired images without trying to build a 3D representation of the user's head [4], [5].
- Model-based methods: use a 3D model to represent the user's head, typically cylindrical [6], [7] or ellipsoidal [8]. The images are used to estimate the 6 parameters that define the position and orientation of the model in space.

Other techniques that use stereo-vision have also been developed to obtain better results [9], [10]. However, all these systems share the limitation of requiring that the user has to be facing the camera to detect his face.

A very important HT application is to obtain virtual spatial audio through headphones [11], where the user's head orientation is used to spatialize sounds in real time. In this kind of application, it is of especial interest that the HT system doesn't restrict the user's head orientation. In [11], the authors used a HT system in the infrared (IR) spectrum, which uses an Object Pose algorithm [12], to reconstruct the position of 4 IR emitters mounted over the headphones. This system provides more freedom to the user's head orientation, but still has problems differentiating some poses.

Most of the HT systems developed to date [4,5,6,7,8,9,10] share several limitations:

- They are restricted to poses in which the user is facing the camera in order to capture facial characteristics.
- They are sensible to the ambient's lightning conditions and the user's physical characteristics.
- They ignore the effects that small camera inclinations could have over the estimation of the user's head inclination.

In this paper we propose a HT system in the IR spectrum that utilizes the Nintendo Wii Remote [13] to detect a circlet with IR emitting diodes. The system determines the head position and orientation rapidly and robustly for a wide variety of poses, thanks to the circlet which has diodes in its whole perimeter. Additionally, the system

detects and eliminates the camera inclination using the information provided by the Wii Remote accelerometers.

We chose this approach in the IR spectrum using the Wii Remote for the following reasons:

1. The characteristics of the Wii Remote's camera: It provides 1024x768 pixels of resolution, and a refresh rate of 100 Hz, outperforming similarly priced web cams, which typically provide 640x480 pixels at 30 Hz. The camera resolution directly affects the precision of the HT system.
2. Infrared: Working in the IR spectrum simplifies a lot the implementation of the system by practically not requiring preprocessing of the images, it makes the system less sensible to the ambient lightning conditions, and is is completely independent of the user's physical characteristics.
3. The Wii Remote accelerometers: They allow to determine the device inclination in rest position, in order to estimate the absolute inclination of the user's head relative to the Earth plane.

The HT system was implemented in C# using the Managed Library for Nintendo's Wii Remote developed by Brian Peek [16]. Additionally, the system can send the position and orientation of the user's head to other applications using the Open Sound Control (OSC) protocol [14]. OSC was created to send data between computers, sound synthesizers and other multimedia devices. For the OSC connection we used the Bespoke OSC library developed by Bespoke Software [17]

The rest of the paper is structured as follows. In section 2 we describe the proposed HT system. In section 3 we comment the results of the system, in terms of precision, speed, and robustness. Finally, in section 4 we discuss the strengths and weaknesses of the system, proposing possible improvements.

2. PROPOSED SYSTEM

In this section we describe the proposed HT system. Section 2.1 shows how to reconstruct the position of three points in space, from their projections in an image. Section 2.2 shows how to use this method to perform HT in the IR spectrum using a circlet with just three IR diodes. Section 2.3 presents the proposed system, which extends the system of section 2.2, using a circlet with diodes in its whole perimeter. Finally, section 2.4 shows how to use the Wii Remote's accelerometers to detect the camera inclination and eliminate it from the HT system's reconstructions.

2.1 3D reconstruction from one view and three points

In this section we describe the method used to reconstruct the position in space of the circlet's diodes from their projections in the image. As it will be shown further on, it is enough to detect 3 points whose relative distances are known, to be able to reconstruct their original positions in space.

2.1.1 Projective geometry

We consider a classical Pinhole camera model as the one shown in figure 2.1.1. In this model, the 3D reference system or *camera system*, is centered at the camera's optical center O. The image plane is placed at a distance *f* from O, known as *focal distance*. The Z axis of the camera system coincides with the optical axis of the camera, and its X and Y axis are parallel with the x and y axis of the image system.

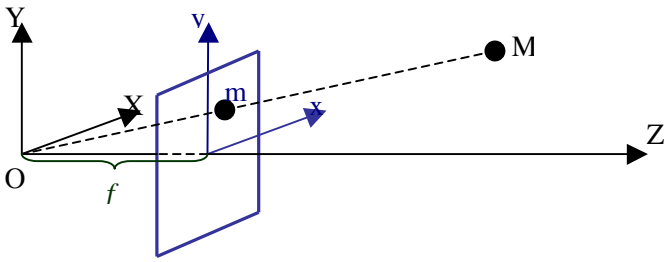

Figure 2.1.1: Projection of a point in the Pinhole camera model.

A point **M** of coordinates (*X*, *Y*, *Z*) in space is projected to a point **m** of coordinates (*x*,*y*) in the image. The coordinates of **m** are determined by the coordinates of **M** and by the focal distance *f*:

$$x = Xf / Z \qquad y = Yf / Z \qquad (2.1.1)$$

The inverse problem, of obtaining the coordinates of **M** from **m**, is under determined. On clearing *X*, *Y* and *Z* from (2.1.1), *Z* stays as a free parameter:

$$X = Zx / f \qquad Y = Zy / f \qquad Z = Z \qquad (2.1.2)$$

2.1.2 3D reconstruction from three points

Let $\mathbf{M_1}$, $\mathbf{M_2}$ and $\mathbf{M_3}$ be tree points in space of unknown coordinates (X_i, Y_i, Z_i), i=1..3, but whose relative distances are known: $\left\|\overline{\mathbf{M}_1\mathbf{M}_2}\right\| = a$, $\left\|\overline{\mathbf{M}_2\mathbf{M}_3}\right\| = b$, and $\left\|\overline{\mathbf{M}_1\mathbf{M}_3}\right\| = c$.

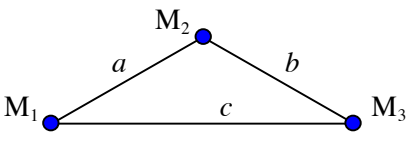

Figure 2.1.2: Distances between the points $\mathbf{M_1}$, $\mathbf{M_2}$ y $\mathbf{M_3}$.

Writing the equations for the distance between 2 points for $\mathbf{M_1}$, $\mathbf{M_2}$ and $\mathbf{M_3}$, we obtain:

$$\begin{aligned}(X_2 - X_1)^2 + (Y_2 - Y_1)^2 + (Z_2 - Z_1)^2 &= a^2 \\ (X_3 - X_2)^2 + (Y_3 - Y_2)^2 + (Z_3 - Z_2)^2 &= b^2 \\ (X_3 - X_1)^2 + (Y_3 - Y_1)^2 + (Z_3 - Z_1)^2 &= c^2\end{aligned} \quad (2.1.3)$$

Let $\mathbf{m_1}$, $\mathbf{m_2}$ and $\mathbf{m_3}$ be their projections in the image, of known coordinates (x_i,y_i), i =1..3. Thee coordinates of $\mathbf{M_1}$, $\mathbf{M_2}$ and $\mathbf{M_3}$ are connected with the coordinates of $\mathbf{m_1}$, $\mathbf{m_2}$ and $\mathbf{m_3}$, by the equations (2.1.2). Replacing them in (2.1.3), and expanding the binomials and factorizing by Z_1, Z_2 and Z_3 the system takes the form:

$$\begin{aligned}S_1 Z_1^{\,2} + 2W_{12} Z_1 Z_2 + S_2 Z_2^{\,2} &= a^2 \\ S_2 Z_2^{\,2} + 2W_{23} Z_2 Z_3 + S_3 Z_3^{\,2} &= b^2 \\ S_1 Z_1^{\,2} + 2W_{13} Z_1 Z_3 + S_3 Z_3^{\,2} &= c^2\end{aligned} \quad (2.1.4)$$

Where S_i are the constants accompanying the quadratic terms Z_i^2 and W_{ij} are the constants accompanying the mixed terms Z_iZ_j.

$$\begin{aligned}S_i &= \left(\frac{x_i}{f}\right)^2 + \left(\frac{y_i}{f}\right)^2 + 1 \\ W_{ij} &= \frac{x_i x_j}{f^2} + \frac{y_i y_j}{f^2} + 1\end{aligned} \quad (2.1.5)$$

By solving the system (2.1.4) the Z coordinates of $\mathbf{M_1}$, $\mathbf{M_2}$ and $\mathbf{M_3}$ can be obtained, and then the other two coordinates can be determined using the equations (2.1.2), completing the 3D reconstruction of these points. Nevertheless, this system is polynomial and is difficult to solve directly, for which we used a numerical method to solve it.

Let g_1, g_2 y g_3 be $\Re^3 \rightarrow \Re$ functions of Z_1, Z_2 y Z_3 such that:

$$\begin{aligned}g_1(Z_1,Z_2,Z_3) &= S_1 Z_1^{\,2} + 2M_{12} Z_1 Z_2 + S_2 Z_2^{\,2} - a^2 \\ g_2(Z_1,Z_2,Z_3) &= S_2 Z_2^{\,2} + 2M_{23} Z_2 Z_3 + S_3 Z_3^{\,2} - b^2 \\ g_3(Z_1,Z_2,Z_3) &= S_1 Z_1^{\,2} + 2M_{13} Z_1 Z_3 + S_3 Z_3^{\,2} - c^2\end{aligned} \quad (2.1.6)$$

Let $\mathbf{G} : \Re^3 \rightarrow \Re^3$ be a vectorial function such that:

$$\mathbf{G}(Z_1,Z_2,Z_3) = (g_1(Z_1,Z_2,Z_3), g_2(Z_1,Z_2,Z_3), g_3(Z_1,Z_2,Z_3)) \quad (2.1.7)$$

Solving the system (2.1.4) for Z_1, Z_2 and Z_3, is equivalent to finding a zero (vector) of function $\mathbf{G}$. In other words, a (Z_1, Z_2, Z_3) trio is solution of the system (2.1.4) if and only if it is root of the function $\mathbf{G}$. To find a zero of this function and solve the system, we used the Newton-Raphson method [15] which is a numerical algorithm to find zeros or roots of real functions.

Using this method we can reconstruct the position in space of three points whose relative distances are known, from their projections in a 2D image. In the next section we describe how to apply this method to perform HT in the IR spectrum using the Nintendo Wii Remote.

2.2 Basic Head Tracking System

In this section we explain how to use the reconstruction method described in the previous section to implement a simple HT system in the IR spectrum, using the Nintendo Wii Remote to detect a circlet with three IR diodes placed in the user's head. Figure 2.2.1 shows a scheme of the system.

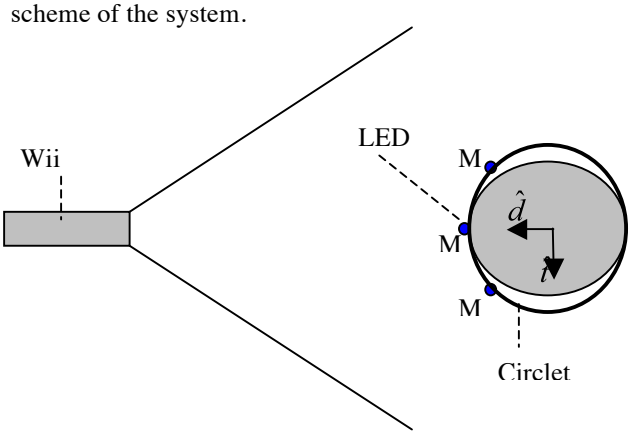

Figure 2.2.1: Scheme of the basic HT system (top view).

Vector $\hat{d}$ shows the direction where the user's nose is pointing and vector $\hat{t}$ shows the direction where the left ear is pointing. The three IR diodes are placed over the user's forehead in a circular circlet of known radius r. The central diode (M_2) is placed at the center of the forehead. The other two (M_1 and M_3) are placed at both sides of the central diode, forming an angle of 45° with it.

The Nintendo Wii Remote incorporates an IR camera [13], of 1024x768 pixels of resolution. The FoV (Field of View) of the camera in the horizontal plane is of 45°, which determines a focal distance of 1326 pixels (units). Using the reconstruction method described in section 2.1, we can obtain the 3D position of the three diodes of the circlet, and subsequently obtain a good estimation of the position and orientation of the user's head. The only needed data to perform the reconstruction are the relative distances between the diodes and their position in the image.

2.3 Extended Head Tracking System

The HT system presented in the previous section requires the user to be facing the camera, because the

program needs to see the three IR diodes placed in his forehead to perform the 3D reconstruction (see figure 2.2.1). This restriction may not be a problem for many HT applications, but can be a limitation for others, such as the spatial audio system mentioned in the introduction.

In this section we propose a way to extend the HT system presented in the previous section, to obtain a solution that would not require the user to remain facing the camera, and that could be used for a wider set of applications.

Now we use a similar circlet to the one presented in the previous section, but that incorporates five additional IR diodes, completing a regular octagon. The new diodes allow the Wii Remote to see the circlet from any point, no matter where the user is looking. Figure 2.3.1 shows a scheme of the system.

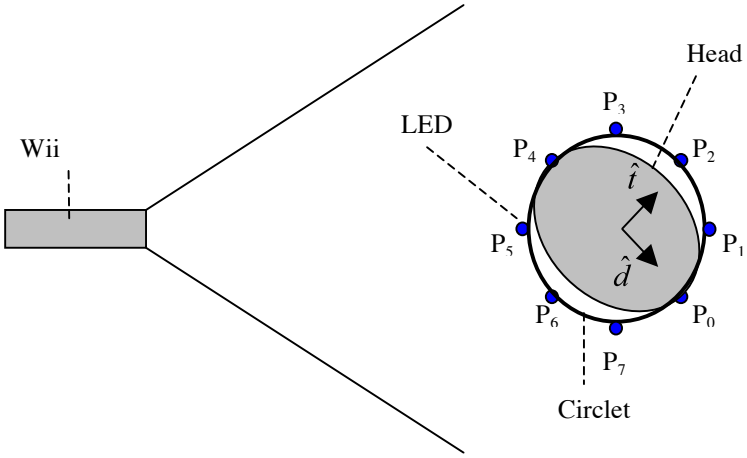

Figure 2.3.1: Scheme of the extended HT system (top view)

Vector $\hat{d}$ shows the direction where the user's nose is pointing and vector $\hat{t}$ shows the direction where the left ear is pointing. The eight diodes of the octagon are enumerated from the frontal diode P_0 to the right, as shown in figure 2.3.1.

The problem of this approach lies in identifying to which part of the circlet belong the diodes being seen by the camera, and therefore, where the user is looking. To help with this task, we added a couple of additional diodes (R_f y R_r), placed over the frontal diode (P_0) and the opposite to this one (P_4) to mark the front and the rear of the circlet. The diodes placed in circle (P_i) will be called from now on, "principal diodes", while the two additional diodes (R_f y R_r) will be called "reference diodes".

3D reconstruction of the circlet

Once the positions of 3 principal diodes of the circlet were reconstructed, we can easily determine the position of the circlet's center and the remaining principal diodes.

Let M_1, M_2 and M_3 be the positions of the 3 visible principal diodes reconstructed by the algorithom of section 2.1 (any trio). Let P be the position of the center, and M_i, i = 4..8, the position of the remaining principal diodes.

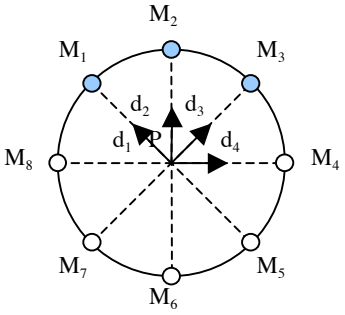

Figure 2.3.6: Reconstruction of the circlet

P can be reconstructed from M_1, M_2 and M_3, using the equation 2.2.3. To reconstruct the positions M_i of the remaining diodes, it's required to determine the 4 directions that go from P to M_1, M_2, M_3 and M_4. Let d_i, i = 1..4, be such directions. These are given by:

$$\begin{aligned}\mathbf{d}_1 &= (\mathbf{M}_1 - \mathbf{P})/\|\mathbf{M}_1 - \mathbf{P}\| \\ \mathbf{d}_2 &= (\mathbf{M}_2 - \mathbf{P})/\|\mathbf{M}_2 - \mathbf{P}\| \\ \mathbf{d}_3 &= (\mathbf{M}_3 - \mathbf{P})/\|\mathbf{M}_3 - \mathbf{P}\| \\ \mathbf{d}_4 &= (\mathbf{M}_3 - \mathbf{M}_1)/\|\mathbf{M}_3 - \mathbf{M}_1\|\end{aligned} \qquad (2.3.1)$$

The positions of the remaining principal diodes are given by:

$$\begin{aligned}\mathbf{M}_4 &= \mathbf{P} + r\mathbf{d}_4 \\ \mathbf{M}_5 &= \mathbf{P} - r\mathbf{d}_1 \\ \mathbf{M}_6 &= \mathbf{P} - r\mathbf{d}_2 \\ \mathbf{M}_7 &= \mathbf{P} - r\mathbf{d}_3 \\ \mathbf{M}_8 &= \mathbf{P} - r\mathbf{d}_4\end{aligned} \qquad (2.3.2)$$

To determine the rotation matrix R of the circlet, we use the equation 2.2.5, but for this, first we must redefine the vectors d and t for the circlet used in this section. Let P_0 be the position of the circlet's frontal diode and P_2 the position of the circlet's left diode (see figure 2.3.1). The direction d of the user's nose can be defined as:

$$\mathbf{d} = (\mathbf{P}_0 - \mathbf{P})/\|\mathbf{P}_0 - \mathbf{P}\| \qquad (2.3.3)$$

Finally, the direction t of the ears can be defined as:

$$\mathbf{t} = (\mathbf{P}_2 - \mathbf{P})/\|\mathbf{P}_2 - \mathbf{P}\| \qquad (2.3.4)$$

2.4 Elimination of the camera's inclination

In this section we describe how to use the Wii Remote's accelerometers to detect the inclination of the device and correct the reconstruction of the circlet, recalculating the position of its diodes relative to a new system of reference with no inclination with respect to the Earth plane.

Let S be the camera's system of reference, of axes X, Y and Z, represented in gray in figure (2.4.1), which point in the directions defined in section 2.1. This system is inclined with respect to the Earth plane, represented as π in the same figure.

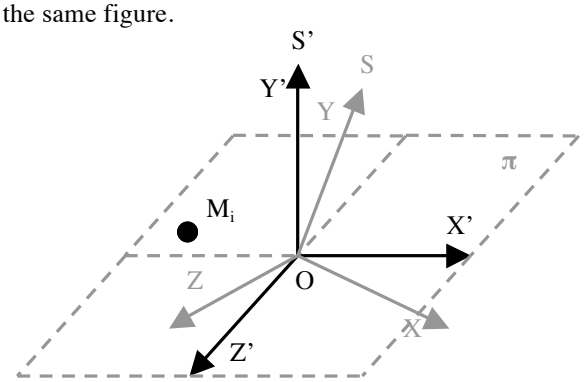

Figure 2.4.1: Correction of the camera's inclination.

Let M_i be the position of one of the circlet's diodes relative to the system S of the camera. The goal is to find the position of this diode relative to a new system of reference S', of axes X', Y' and Z' (represented in black), centered in the same origin as S, but:

- whose Y' axis is normal to the Earth plane π and,
- whose Z' axis is the orthogonal projection of the Z axis over π.

The X' and Z' axes of the system S', are contained in the Earth plane π.

To obtain the position M_i' of the diode relative to the system S', we must determine the inclination of the camera relative to the earth plane, which is determined by two angles:

- A pitch angle, corresponding to the angle formed by the Z and Z' axes.
- A roll angle, corresponding to the angle formed by the X and X' axes.

The yaw angle cannot be determined in this case.

To determine this inclination, we can use the information of the accelerometers built into the Wii Remote [13]. The device counts with three accelerometers that measure accelerations in each axis. When it is a rest, the only present acceleration is the acceleration of gravity, so we can determine its inclination with respect to the Earth plane, watching how the acceleration vector is projected onto each of the three axes.

Figure (2.4.2) shows the disposition of the accelerometers in the Nintendo Wii Remote. Note that these axes do not coincide with the axes of the camera's system.

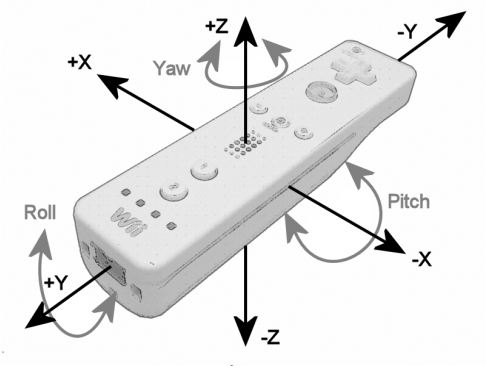

Figure 2.4.2: Disposition of the Wii Remote's accelerometers.

The Wii Remote gives acceleration values normalized by the acceleration of gravity in the range [-3g,+3g] [13]. This means that if the device rests on one of its sides over a perfectly horizontal surface, the registered accelerations should be 1g in the axis normal to the surface and 0 in the other two. If it were not the case, a calibration process must be made to determine the error of each accelerometer and subtract it from the values.

2.4.1 Determination of the pitch and roll of the camera

Let a_x, a_y and a_z be the magnitude of the accelerations registered by the Wii Remote in each of the three axes. The vector $\vec{a} = (a_x, a_y, a_z)$, shows the direction of the raw acceleration of the device. When it is at rest, the accelerometers only register the acceleration of gravity, so $\|\vec{a}\| = 1g$, but most importantly, $\vec{a}$ is normal to the Earth plane. Figure (2.4.3) shows a particular instance of vector $\vec{a}$ in black:

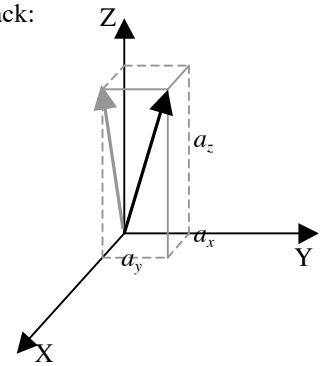

Figure 2.4.3: Watching the acceleration of gravity

The pitch angle (from now on, φ) of the camera corresponds to the one formed between the $\vec{a}$ vector and its projection over the XZ plane (gray vector), which is given by:

$$\varphi = \text{atan}\left(\frac{a_y}{\sqrt{a_x^{\,2} + a_z^{\,2}}}\right) \quad (2.4.1)$$

For its part, the roll angle (from now on, ρ) corresponds to the one formed between the projection of $\vec{a}$ over the XZ plane (gray vector) and the Z axis, which is given by:

$$\rho = \text{atan}\left(\frac{a_x}{a_z}\right) \quad (2.4.2)$$

Both rotation angles are independent. It can be verified looking at figure (2.4.2) that if $a_y = 0$, there's no pitch rotation. In the same way, if $a_x = 0$, there's no roll rotation.

2.4.2 Correction of the 3D reconstruction of the circlet

Once the angles of inclination of the camera are known, we can determine the 3D position of the diodes with respect to the new system of reference S', whose Y axis is normal to the plane of the Earth.

Let M_i be the position of one of the circlet's diodes in the system of reference S of the camera. To get the position M_i' with respect to the new system S', a linear transformation of rotation **R** must be done to M_i to eliminate the rotations of roll and pitch of the camera.

$$\mathbf{M}_i' = \mathbf{R}\mathbf{M}_i \quad (2.4.3)$$

The rotation matrix R is the product of two rotation matrixes of pitch and roll:

$$\mathbf{R} = \mathbf{R}_{pitch}\mathbf{R}_{roll} \quad (2.4.4)$$

These rotation matrixes are given by:

$$\mathbf{R}_{roll} = \begin{bmatrix} \cos(\rho) & -\sin(\rho) & 0 \\ \sin(\rho) & \cos(\rho) & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2.4.5)$$

$$\mathbf{R}_{pitch} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\varphi) & -\sin(\varphi) \\ 0 & \sin(\varphi) & \cos(\varphi) \end{bmatrix} \quad (2.4.6)$$

The matrix $\mathbf{R}_{roll}$ rotates the system S of the camera around its Z axis, making its X axis coincide with the X' axis of the system S' (see figure 2.4.1). Finally, the matrix $\mathbf{R}_{pitch}$ rotates the system S around its X axis, making its remaining axes coincide with their respective axes of the system S'. Both rotations are not commutative because of how the rotation angles were defined. The roll rotation must be done in the first place.

Multiplying the rotation matrix R with the position M_i of each diode of the circlet, we can correct its reconstruction, obtaining the inclination of the user's head with respect to the plane of the earth, no mattering the inclination of the camera.

3. EXPERIMENTS AND RESULTS

3.1 Precision

To measure empirically the error of the circlet's reconstructions done by the system, we would have to know the exact original positions of the diodes with respect to the camera's system, to be able to compare them with the reconstructed positions. Since this is nearly impossible in practice, we made a simulation in Matlab to measure this error theoretically.

The simulation considers different positions and rotations for the circlet in space, computing the 3D positions of the diodes, and projecting them to an image according to the Pinhole camera model (equations 2.1.1). The (x,y) position of the diodes in the image is discreetized, which means a loss of information. From these projections it reconstructs the 3D position of the diodes and determines the position and rotation of the circlet. Finally, it compares the reconstructed position and rotation angles with the original ones.

We considered a circlet with three diodes as the one of section 2.2, of 10.5 cm of radius. The circlet was placed in 11 rectangles along the Z axis, from Z = 100 cm to Z = 200 cm every 10 cm. Each rectangle goes from X = -30 cm to X = 30 cm and from Y = -20 cm to Y = 20 cm and contains 35 different positions, every 10 cm (7x5). Finally, each position considers 5 posible rotations in each angle from -20° to +20° every 10°, giving a total of 125 posible composed rotations. The total number of simulated instances was 48,125.

The error of the reconstructions increases with the distance as expected. Nevertheless, the mean error of the position stays under 1 cm, and of the angles under 1° when the user stands at less than 2 m of the camera.

The error of the reconstructions in the simulation is mostly due to the loss of information produced by projecting a 3D point of continuous coordinates to a discreet image. In this lies the importance of the camera's resolution. By eliminating the discretization in the simulation's projections, the error of the reconstructions converges to zero.

In practice, we have to add to the previous error the one caused by imperfections in the circlet's construction. The circlet considered in the simulations is an "ideal" circlet without imperfections, for which it only gives us a lower bound of the real error.

3.2 Speed

The C# implementation of the extended HT system with the elimination of the camera's inclination turned on, runs without slowdowns at 100 Hz, making maximum use of the Wii Remote's IR camera's refresh rate.

The 3D reconstructions of the system converge in 5 to 6 iterations of the Newton-Raphson when the circlet is placed from 100 to 200 cm of the camera, using a constant starting point $\mathbf{Z_0} = (100,102,100)$ and $\varepsilon = 0.1$ (in centimeters). Reusing the previous solutions as new starting points, Newton-Raphson converges in 4 to 5 iterations for the same ε.

3.3 Robustness

The HT system works robustly in absence of IR noise. We consider IR noise any source of IR light detected by the Wii Remote that doesn't correspond to a diode of the circlet. The main sources of IR noise identified are:

- The light of the sun: contains IR radiation that can reach the Wii Remote directly through windows, or reflected over mirrors or other surfaces.
- Reflections of the circlet's diodes: mirrors or other plain surfaces placed too close to the circlet may reflect the IR light of the diodes towards the Wii Remote.
- Dirt from the lens: can distort the light received by the Wii Remote, splitting the light of one diode in many, or merging several lights in one.

This IR noise confuses the HT system, making it identify noise as diodes of the circlet, which produces wrong and unstable reconstructions.

Finally, the robustness of the extended HT system is limited by the robustness of the method used by the Tracking Module to find the reference diodes. This one fails when the circlet presents pronounced pitch rotations with respect to the camera's system, leading to wrong detections of the reference diode, and therefore, to wrong circlet's reconstructions. The system works robustly only for moderated pitch rotations (± 30°).

4. CONCLUSIONS

We have proposed a HT system in the IR spectrum that uses the Nintendo Wii Remote to detect a circlet with IR emitting diodes, placed over the user's head. The system accurately detects the position and orientation of the head in space, additionally presenting two new features with respect to previous works:

1. It works without requiring the user to be facing the camera, thanks to the circlet, which incorporates IR diodes in its whole perimeter.
2. It detects and eliminates the camera inclination using the Wii Remote accelerometers, obtaining the absolute inclination of the user's head relative to the Earth plane.

Other positive features of the system are:

- Low cost: The Wii Remote can be purchased for a price comparable to a conventional web cam's[1], and the used circlet can be built for less than 10 USD.
- Performance: with its 1024x768 pixels of resolution and 100 Hz of refresh rate, the Wii Remote widely outperforms the conventional web cams which typically provide 640x480 pixels at 30 Hz. The camera resolution directly affects the precision of the reconstructions. On the other hand, the HT system runs at 100 Hz maximizing the capabilities of the Wii Remote.
- Simplicity: The chosen approach in the IR spectrum using the multi-object tracking (MOT) engine of the Wii Remote, highly simplifies the implementation by not requiring major image pre-processing. Additionally, the simplicity of the circlet's geometry makes the geometric analysis easier.
- Modularity: The proposed system attacks the problem decomposing it in two modules: a tracking module entrusted of identifying the circlet's diodes, and a reconstruction module entrusted of reconstructing its geometry in space. This scheme is valid for any system with a similar approach and allows proposing improvements easily.
- Usability: The HT system output can be sent to other programs through the Open Sound Control (OSC) protocol, allowing its usage in widely used audio applications.

Regarding the negative aspects of the system, we consider that it counts with two main weaknesses:

1. The circlet's symmetry: The frontal and rear parts of the circlet look exactly equal for the system, because it has no way to differentiate the frontal and rear reference diodes. To solve this issue, we determined that the user has to be facing the camera at the beginning of the tracking, to count with an initial known state. This requirement is disadvantageous for two reasons:
    - Makes the system not autonomous, for depending of an initial user set-up.
    - Introduces the risk that a permanent error may occur in the long term. For example, if the circlet momentarily leaves the FoV of the camera, it may

[1] Amazon.com offers it at 34.96 USD in August 2009.

confuse the front with the rear of the circlet.

A good way to solve this problem is to break with the circlet's symmetry, placing the rear reference diode under the plane of the principal diodes. The method to detect the reference diodes would be the same, but they would be differentiated for being over or under the plane of the principal diodes.

2. The reference diodes detection: The tracking module finds the reference diodes looking for a trio of diodes forming a straight enough angle. This criteria works well anytime the circlet doesn't present very pronounced pitch rotations with respect to the camera's system. Otherwise, the most straight angle may be formed in a wrong trio of diodes, leading to an incorrect detection of the reference diode.

   The criteria of the straight angles to detect the reference diodes, restricts the possible pitch rotations that the circlet can adopt to be correctly detected, for which it is necessary to look for a better alternative.

REFERENCES

[1] E. Hjelmas and B.K. Low, "Face detection: A survey," *Computer Vision and Image Understanding*, vol. 83, no. 3, pp. 236-274, 2001.

[2] E. Murphy-Chutorian E and M.M. Trivedi, "Head Pose Estimation in Computer Vision: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 4, pp. 607-626, 2009.

[3] M. Porta, "Vision-based user interfaces: methods and applications," *International Journal of Human-Computer Studies*, vol. 57, no. 1, pp. 27-73, 2002.

[4] F. Bérard, "The perceptual window: head motion as a new input stream," *Proceedings of the 7th IFIP conference on Human–Computer Interaction (INTERACT)*, Edinburgh, Scotland, pp. 238–244, 1999

[5] K. Toyama, ""look, ma—no hands!" Hands-Free Cursor Control with Real-Time 3D Face Tracking," Proc. Workshop Perceptual User Interfaces, pp. 49-54, 1998.

[6] M. La Cascia, S. Sclaroff, and V. Athitso, "Fast, Reliable Head Tracking under Varying Illumination: An Approach Based on Registration of Texture-Mapped 3D Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*. vol. 22, no. 4, pp. 322-336, 2000.

[7] W. Ryu, and D. Kim, "Real-time 3D Head Tracking and Head Gesture Recognition," *16th IEEE International Conference on Robot & Human Interactive Communication*. August 26-29, 2007, Jeju, Korea. pp. 169-172, 2007.

[8] K.H. An, and M.J. Chung, "3D Head Tracking and Pose-Robust 2D Texture Map-Based Face Recognition using a Simple Ellipsoid Model," *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*. Acropolis Convention Center. Nice, France, Sept. 22-26, 2008, pp. 307-312, 2008.

[9] L.-P. Morency, A. Rahimi, N. Checka, and T. Darrell, "Fast Stereo-Based Head Tracking for Interactive Environments," *Proc. IEEE Int'l Conf. Automatic Face and Gesture Recognition*, pp. 375-380, 2002.

[10] R. Yang and Z. Zhang, "Model-Based Head Pose Tracking with Stereovision," *Proc. IEEE Int'l Conf. Automatic Face and Gesture Recognition*, pp. 242-247, 2002.

[11] A. Mohan, R. Duraiswami, D.N. Zotkin, D. DeMenthon, L.S. Davis, "Using Computer Vision to generate Customized Spatial Audio," *Proceedings of the 2003 International Conference on Multimedia and Expo*, ICME 2003. vol. 3, pp. 57-60, 2003.

[12] D. DeMenthon, and L.S. Davis, "Model-Based Object Pose in 25 Lines of Code," *International Journal of Computer Vision*. vol. 15, no. 1-2, pp. 123-141, 1995.

[13] J.C. Lee, "Hacking the Nintendo Wii Remote," *IEEE Pervasive Computing*, vol. 7, no. 3, pp. 39-45, 2008.

[14] M. Wright and A. Freed, "Open Sound Control: A New Protocol for Communicating with Sound Synthesizers," *International Computer Music Conference*, Thessaloniki, Greece, 1997, pp. 101-104, 1997.

[15] R. Hartley and A. Ziserman, "Multiple View Geometry in Computer Vision", Cambridge University Press, 2000.

[16] WiimoteLib - .NET Managed Library for the Nintendo Wii Remote. http://www.brianpeek.com/blog/pages/wiimotelib.aspx

[17] Bespoke - .NET OSC Library. http://www.bespokesoftware.org/wordpress/?page_id=69