MULTIMODAL CONTROL OF MUSIC AND FIRE PATTERNS

Todor Todoroff
Institut Numediart - UMons
Mons, BELGIUM
ARTeM - Brussels, BELGIUM
todor.todoroff@skynet.be

Radhwan Benmadhkour, Ricardo Chessini
Institut Numediart - UMons
Mons, BELGIUM
Radhwan.benmadhkour@umons.ac.be
Ricardo.chessini@umons.ac.be

ABSTRACT

We developed hardware and software tools to enable interactive performances whereby a dancer controls both music and fire. Pyro-technician Pierre D’haenens built at ShowFlamme [10] a patent pending ramp of 20 flames projectors. The height of each flame is individually controlled by software. We developed a light wireless wearable sensor system and video analysis tools for a stereoscopic camera in order to track reliably the gestures, the position and height of a dancer despite the presence of the flames. We combined these developments with previous ones related to real-time interactions with dancers and musicians. This allows the performer to control simultaneously flame patterns and the music, using position, height, gestures and sound analysis.

1. INTRODUCTION

There is an increasing need for techniques to track the gestures and movements of dancers on-stage, for various kinds of interactive performances. Techniques that demand only a light setup, both in terms of equipment and setup time, that are robust enough to take on tour and that don’t modify the dancer’s appearance. This excludes the use of some well established technologies used in motion pictures or in the gaming industry, like putting visible markers on the body and using a large array of cameras, or demanding the dancer to wear a special and cumbersome suit fitted with arrays of sensors. And the price should remain affordable for artistic projects.

In order to address those needs, we followed two complementary paths: we built a small and light wireless sensors system to be worn by the performer, providing kinetic and posture information thanks to the inclusion of 3-axes accelerometers, magnetometers and gyroscopes. And we developed tracking algorithms for a stereoscopic video camera to get the position of the performer on stage and his height. Those hardware and software tools can be used in many kinds of interactive performances, with one or more performer(s) on stage.

The closest existing device that would give us a visual idea of the fire ramp before it was build was the Rubens' flaming tube, where the pressure of gas at each flame hole is defined by an acoustical standing wave produced with a loudspeaker attached at one side of the tube. In our case, we benefit from individual control over each flame of the ramp. And we can achieve much higher flames. The aim was to generate complex flames patterns controlled by the movements of a dancer.

Figure 1. *Dominica Eyckmans dancing and playing viola.*

The same data would also control the music in combination with the tools we developed for the *Dancing Viola* project [12]: hit detection, DTW-based gesture recognition [1] and mapping by interpolation [13] allows us to blend three modalities: position tracking, gestures analysis and sound analysis. Figure 1 shows the performer playing the viola, controlling the flames, sound transformations of her acoustical instruments as well as triggering and modulating pre-recorded sounds.

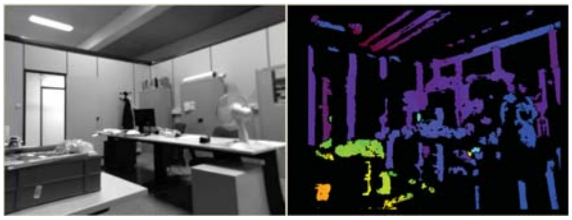

Figure 2. a. Left image of our stereoscopic camera. b. Depth map: orange shows a short distance to the camera.

2. POSITION TRACKING

We needed to track the position of the dancer in very difficult conditions. Indeed, the flames generate strong visible and infrared light, making the use of a classical IR camera unsuitable. But the variable flame patterns also constantly change the illumination of the scene, creating both strong light changes and moving shadows, rendering background subtraction also unusable. Using a stereoscopic camera, we could remove the lights and shadows on the ground thanks to the additional distance information.

2.1. Stereoscopic camera principles

A stereoscopic camera integrates two classical cameras. Figure 2 shows the left image and the depth map obtained with a stereoscopic camera. The depth map is not homogenous as is it not possible to estimate the depth when the corresponding point in the other image is not found. In those locations the displayed map is black. It happens when the surface is not textured or has no edges.

We use a Stereo-on-Chip Videre Design camera [5] that computes the depth map on board.

2.2. From image to world

The perspective projection gives the relation between an image point (u,v) and the corresponding coordinates of the scene point $(x_{cam}, y_{cam}, z_{cam})$.

$$\begin{cases} u = f\frac{x_{cam}}{z_{cam}} + u_0 \\ v = f\frac{y_{cam}}{z_{cam}} + v_0 \end{cases} \quad (1)$$

In those equations, f is the focal length, (u_0, v_0) are the coordinates of the principal point (point where the optical axis meets the image plane). Using this relation, we can retrieve the 3D point, but in the camera coordinate frame. If we apply a rotation and a translation to the coodinates to put the x and y axis on the ground, we get the position and the height in the new coordinates system (x_w, y_w, z_w).

The complete relation from pixel coordinates to ground positions is given in the expression 2 & 3 where R is a 3x3 rotation matrix and C is a vector containing the coordinates of the camera coordinate frame origin in the world coordinate frame [7].

$$\begin{bmatrix} u.z_{cam} \\ v.z_{cam} \\ z_{cam} \end{bmatrix} = \begin{bmatrix} f & 0 & u_0 & 0 \\ 0 & f & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_{cam} \\ y_{cam} \\ z_{cam} \\ 1 \end{bmatrix} \quad (2)$$

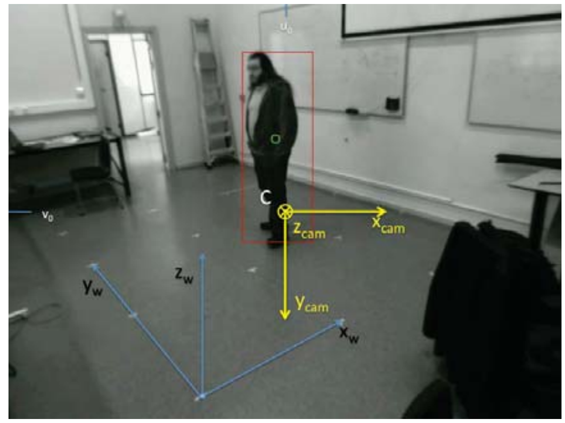

Figure 3. Position tracked in the world coordinates frame.

$$\begin{bmatrix} x_{cam} \\ y_{cam} \\ z_{cam} \\ 1 \end{bmatrix} = \begin{bmatrix} R & -RC \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \quad (3)$$

2.3. Walls, ground and fire elimination

As the camera reconstructs everything it sees, we also get the information about the fire and the walls. To get rid of this unwanted information, we define planes in front of the walls, the fire and over the ground (see fig. 4). Everything that is detected behind those planes is then suppressed and we only get the dancer's 3D reconstruction. Planes are defined given three non collinear points $P_1(x_1,y_1,z_1)$ $P_2(x_2,y_2,z_2)$ $P_3(x_3,y_3,z_3)$.

$$plane\ equation: ax + bx + cz + d = 0 \quad (4)$$

$$\begin{vmatrix} u_x & u_y & u_z \\ x_1 - x_2 & y_1 - y_2 & z_1 - z_2 \\ x_1 - x_3 & y_1 - y_3 & z_1 - z_3 \end{vmatrix} = (a,b,c) \quad (5)$$

$$d = -ax_1 - by_1 - cz_1 \quad (6)$$

The (a,b,c) coordinates of the normal vector to the plane are obtained by the cross-product of two vectors lying on it. Reintroducing a,b,c and $P_1(x_1,y_1,z_1)$ in the plane equation 4, we find d (eq 6).

2.4. Tracking

The tracking is done using classical blob tracking techniques [2] (implemented in cvblobslib). To create the blob, we binarize the depth map. The definition of a minimal blob size let us eliminate the noise. We then compute the barycenter of the blob in the three dimensions. The x and y values give the postion. We apply the transformation (2) (3) to the heighest pixel of the blob. The resulting z is the height of the dancer.

3. SENSORS

Commercial wireless sensor interfaces designed for artists (Eowave Eobody2 HF, Interface-Z Wiwi or Mini-HF, Infusion System I-cubeX, La Kitchen Kroonde, ...) have

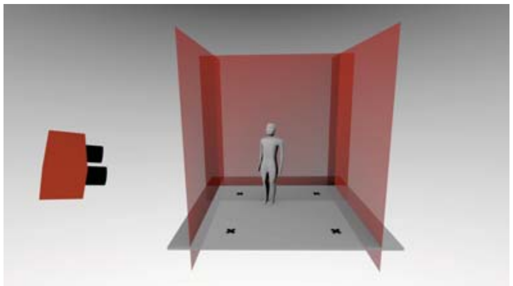

Figure 4. Fire, ground and walls elimination.

only up to 16 analog inputs, with 10 or 12 bits ADCs. While they allow users to connect various sensors without any additional programming, they provide only a limited number of channels and impose a heavy wire harness. In order to fit each sensor node with 3-axes accelerometer, magnetometer, gyroscope and temperature sensor for calibration, we would need 12 wires per node (10 channels plus ground and power). And we aimed at a system with 3 to 5 sensor nodes sampled at 100 Hz.

A system with sensor nodes of similar capacities had been described in [6], though with a different approach. There is an obvious trade-off between that system, truly wireless, even on the body, but with bigger nodes as each one carries its own emitter and battery, and our system with sensors connected through a digital bus on the body. We believe that our approach, with very flat sensors, invisible under the clothes, offers more freedom of movement to the dancer, particularly for movements on the ground, despite the need for cables on the body. Light-weight sensors have the additional advantage of having a small inertia which allows them to follow more closely the movements of the limbs of the dancer they are attached to.

Following the experience of the sensor system developed in 2006 at ARTeM [12] for the *Quartet Project* [9], and used for *De deux points de vue* [4] and *Dances with Viola* [3], we kept a master/sensor nodes architecture while reducing the form factor and adding sensing capabilities. They communicate through a 4 wire 400kHz I2C bus on the body: a bidirectional data (*SDA*) and clock link (*SCL*), a common ground (*GND*) and a power supply line (*VDD*).

3.1. Sensor chips choices

We made an extensive search in 2009 for our first prototype. There were obvious choices for 3-axes digital magnetometers (Honeywell HMC5843) and accelerometers (STMicroelectronics LIS302DLH or Analog Device ADXL345). We chose for the later both for its variable resolution, keeping 4mg/LSB up to ±16 g, and for its additional functionalities that could be used at a later stage (single and double click detection, FIFO). But 3-axes gyroscopes were not yet available and we had to choose a combination of an x/y-axes gyroscope and a z-axis one. Because of the amount of external components needed, we chose for the InvenSense IDG-650 and ISZ-650 rather than for the STMicroelectronics gyroscopes. We used a PIC18F2423 for its 12-bit DACs, adding four times oversampling for better precision. In our latest design, we use the Invensense ITG-3200 digital 3-axes Gyroscope and added 6 channels of ADC for optional additional sensors (pressure, flexion, light, ...), all on a 17x38mm PCB that fits into a tiny USB key box (Figure 5).

3.2. Wireless sensor system architecture

We tested several low power wireless technologies to see how far we could reduce the size and weight of the battery: ZigBee, SimpliciTI, Bluetooth. We found huge disparities between the announced data rates and the measured ones, as shown in table 1: ZigBee and SimpliciTI could not be used reliably with more than one or two sensor nodes. Though BlueTooth could handle three nodes at 100Hz, it only provided a decent latency at 50Hz, as packet sizes increased with data rate. A low power WiFi module could on the other handle up to 8 nodes at 100Hz, with the added benefit of a smaller and constant latency thanks to the use of a match character to send data in a single IP packet. Despite higher transmission power, the average WiFi power consumption was similar to BlueTooth, as the high throughput shortens the transmission time. It was chosen as shown in Figure 5.

	Wireless Personal Area Network		
Solution	Bluetooth	ZigBee	SimpliciTI
IEEE	820.15.1	802.15.4	based on 802.15.4
data sheet	720 kbps	250 kbps	250 kbps
functionality	216.9 kbps	69.5 kbps	41 kbps
with sensors	76.8 kbps	37.5 kbps	36 kbps

Table 1. Summary of throughputs.

Bi-directionnal communication allows the user to remotely and dynamically set up, directly from Max/MSP, the master sampling period and to define which on-board sensors and ADC channels need to be transmitted by each node, allowing the user to tailor his system and to optimize bandwidth. Unused on-board sensors can be put to sleep in order to economize power and many of their specific parameters can be adjusted in real-time: range, sampling and cut-off frequencies, self-test, degaussing, etc.

3.3. Attitude computation and skeleton

MARG (Magnetic, Angular Rate, and Gravity) sensors allow for a drift-free attitude computation using classical Kalman filtering. But we used a quaternion-based method [8] that gives good results even at low sampling rates while avoiding singularities associated to Euler angles. Depending on the connected number of sensors and the type of performance we can choose between 50, 100 or 200 Hz sampling rate. More detail on the sensors and on how the attitude can be used to reconstruct the skeleton, and hence find the relative positions of the limbs can be found in [11].

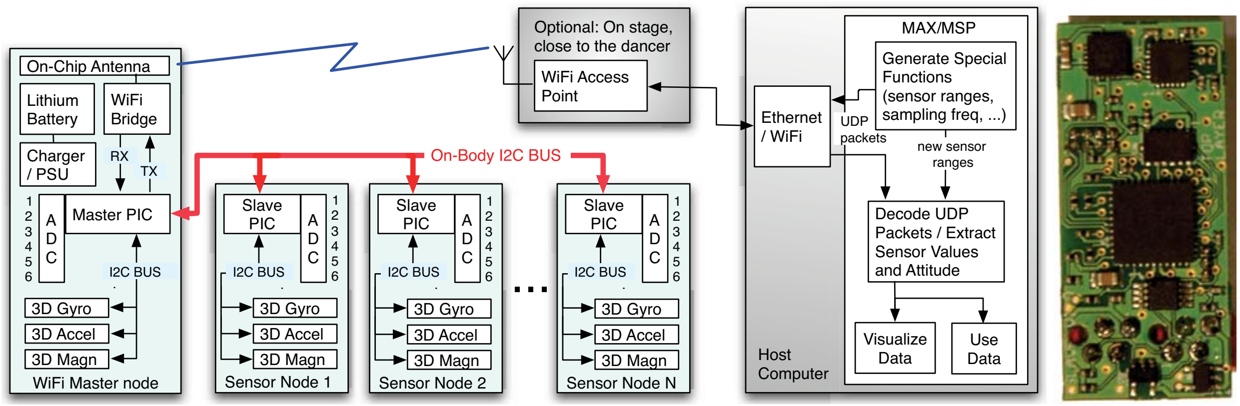

Figure 5. *Left: Wireless sensor system global architecture, with all transmission paths (network, wireless, serial, local and on-body I2C, analog). Right: 17x38mm sensor node (connectors are located on the back).*

4. MUSIC AND FLAME CONTROL

The whole system works within Max/MSP/Jitter, except for the video tracking, running on a separate Linux computer. The height of a flame or the intensity of a fire ball depends on the magnitude and duration of the valve opening. Real-time flame pattern generators respond to sound features, position and height tracking and gestures, using the mapping scheme described in [12]. The attitude extraction, the interpolation tools [13] and the DTW gesture recognition [1] can be combined to give increased control.

5. CONCLUSIONS

The position tracking proved very robust, reaching around 10 cm precision at all positions, despite the rough conditions. Plane elimination enables to define a precise interaction area. As the distance map is computed on chip, the position tracking is done at 25 frames per seconds.

We believe our sensors system is an improvement in size, capabilities and resolution over other systems in the same price range, specially when many sensors are needed. The combination of high resolution digital 3-axes accelerometers, magnetometers and gyroscopes allows for attitude computation and the choice of WiFi allows several performers to share the same wireless channel.

6. ACKNOWLEDGMENTS

We would like to thank D. Binon at SEMI-UMons as well as O. Schevens, D. Lekime and J.-Y. Parfait at Multitel for their help in designing the prototype and final version PCBs. Research supported by `Numediart`, a long-term research program centered on Digital Media Arts, funded by the Région Wallonne, Belgium (grant N°716631).

7. REFERENCES

[1] F. Bettens and T. Todoroff, "Real-time dtw-based gesture recognition external object for max/msp and puredata," in *Proc. SMC '09*, 2009, pp. 30–35.

[2] F. Chang, C.-J. Chen, and C.-J. Lu, "A linear-time component-labeling algorithm using contour tracing technique," *Computer Vision and Image Understanding*, vol. 93, no. 2, pp. 206–220, 2004.

[3] Dances with Viola. [Online]. Available: http://www.danceswithviola.org/

[4] De Deux Points de Vue. [Online]. Available: http://www.michele-noiret.be/index.php?page=de-deux-points-de-vue

[5] V. Design, "Stoc - videre design," April 2010. [Online]. Available: http://www.videredesign.com/

[6] M. Fernström, "Celeritas: Wearable wireless system," in *Proc. NIME '07*, 2007, pp. 205–208.

[7] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge University Press, 2004.

[8] S. O. H. Madgwick, "An efficient orientation filter for inertial and inertial / magnetic sensor arrays," 2010.

[9] Quartet. [Online]. Available: http://www.quartetproject.unsited.org/space/Sound+and+Gesture

[10] ShowFlamme. [Online]. Available: http://www.showflamme.be/

[11] T. Todoroff, "Wireles digital/analog sensors for music and dance performances," in *Proc. NIME '11*, Oslo, Norway, 2011.

[12] T. Todoroff, F. Bettens, W.-Y. Chu, and L. Reboursière, "Extension du corps sonore - dancing viola," in *Proc. NIME '09*, Pittsburgh, Pennsylvania, USA, 2009, pp. 141–146.

[13] T. Todoroff and L. Reboursière, "1-d, 2-d and 3-d interpolation tools for max/msp/jitter," in *Proc. ICMC '09*, Montreal, Quebec, Canada, 2009, pp. 447–450.