The Kinect sensor as human-machine-interface in audio-visual art projects

Matthias Kronlachner
Student
mail@matthiaskronlachner.com

IOhannes m zmölnig
zmoelnig@iem.at

Institute of Electronic Music and Acoustics

University of Music and Performing Arts Graz

Introduction

For several years now, the entertainment and gaming industry has been providing multifunctional and cheap human interface devices which can be used for artistic applications.

Since November 2010 a sensor called Kinect™ for Microsoft's Xbox 360 is available. This input device is used as color camera, 4 channel microphone array, and provides - as an industry-first - a depth image camera at an affordable price. Soon after the release of the Xbox-only device, programmers from all over the world provided solutions to access the data from a multitude of operating systems running on ordinary computers.

Operating system independent ways are presented to access and interpret Kinect data streams within Pure Data/Gem[a]. Overviews of case studies for interactive audio and visual art projects are given.

All developed software is Open Source and working under Linux, Mac OS X and Windows.

[a]Pure Data is a visual programming environment used for computermusic and interactive multimedia applications. Gem stands for *Graphics Environment for Multimedia* and extends Pure Data to do realtime OpenGL based visualizations.

Kinect streams in Pure Data/Gem

using `pix_freenect`, `pix_depth2rgba`, `pix_threshold_depth`

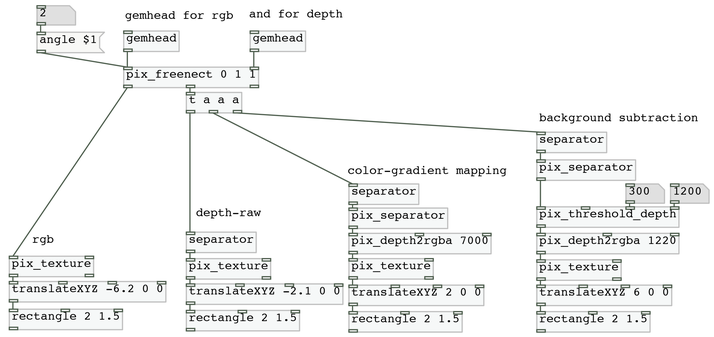

Fig. 1: RGB stream, raw depth stream, color gradient mapping, background subtraction

Representation of depth data

A solution for handling depth data is proposed using RGBA or YUV colorspace.
For RGBA output the 16 bit depth data is divided into the upper and lower eight significant bits and stored in the red (R) and green (G) channel. The blue channel (B) is used for additional information about the pixel. For example, if a user is present in that specific pixel, the specific user-id is set.

R	G	B	A
3/8 msb	8 lsb	0 or userid (OpenNI)	255

YUV422 (2 bytes per pixel)
11 bit/16bit depth values

For development and visualization purposes `pix_depth2rgba` is used to map distance values onto a color gradient (Fig. 2).

Fig. 2: Gradient for displaying depth data - near to far

Various methods using CPU or GPU processing power can be used to filter out certain regions of the depth image and use them for tracking or as projection stencil.

Skeleton tracking with `pix_openni`

`pix_openni` gives access to higher level functions as multiple user, hand and skeleton tracking (Fig. 3). The output rate depends on the framerate of the depth sensor, appr. 30 Hz.

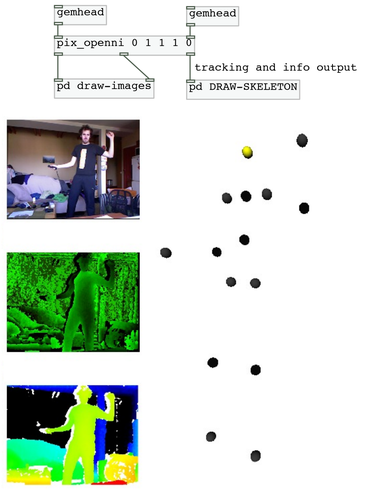

Fig. 3: Displaying skeleton retrieved by pix_openni

vertimas - übersetzen

The piece *übersetzen - vertimas* for dancer, sound and projection features the Kinect sensor to translate body movements on stage into sound and turns the dancers body into a hyperinstrument. Additionally, the depth video stream is used to gather the outline of the dancer and project back onto her body in realtime (Fig. 4).
Therefore an data-flow filtering and analysis library has been developed, providing quickly adjustable methods for translating tracking data into sound or visual control data (Fig. 5).

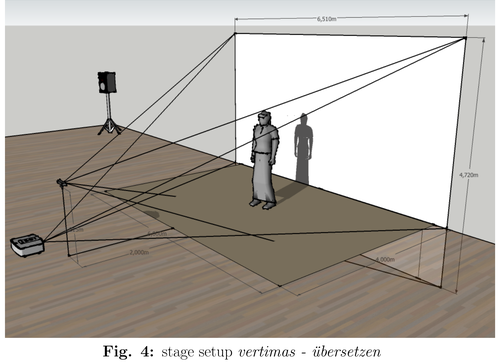

Fig. 4: stage setup *vertimas - übersetzen*

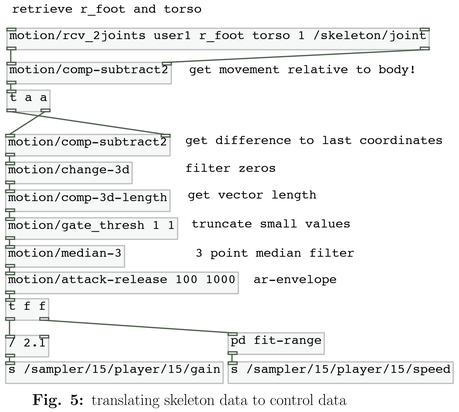

Fig. 5: translating skeleton data to control data

IEM Computermusic Ensemble

ICE is a group of electronic musicians, each playing with a notebook and individual controllers. The target is to play contemporary music, adapted or written for computermusic ensembles.
In March 2012 a network concert between Graz and Vilnius took place. One Kinect sensor was used to track the skeleton of two musicians. The tracking data allowed each musician to play his virtual instrument without handheld controllers. Additionally the Kinect video stream showing the stage in Vilnius was sent to Graz and projected on a canvas for the remote audience. (Fig. 6)

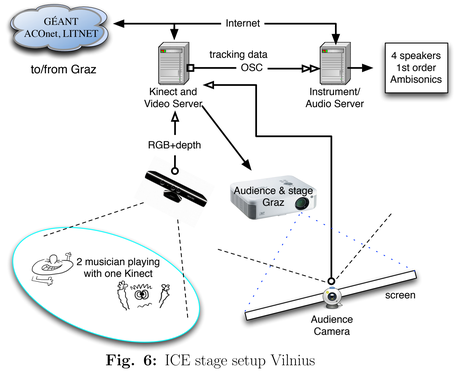

Fig. 6: ICE stage setup Vilnius

Head pose estimation

Based on a paper and software by Gabriele Fanelli[2], the external `pix_head_pose_estimation` has been developed which takes the depth map of the Kinect as input and estimates the Euler angles and position of multiple detected heads in the depth map. The head tracking can be used to rotate an Ambisonics soundfield for monitoring 3D sound environments via headphones (Fig. 7).

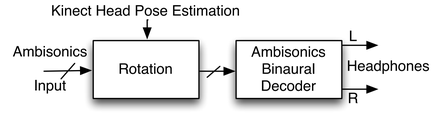

Fig. 7: rotating sound field according to head movements

References

[1] M. Kronlachner, *The Kinect distance sensor as human-machine-interface in audio-visual art projects, project report*, Institute of Electronic Music and Acoustics, Graz, 2012.

[2] G. Fanelli & T. Weise & J. Gall & L. Van Gool, *Real Time Head Pose Estimation from Consumer Depth Cameras*, DAGM'11.

Examples and source code available from the homepage!
www.matthiaskronlachner.com