Łódź, 09.06.2013

**PBL project – Kinect-based human movement analysis.**

Supervised by Wojciech Sankowski PhD.

Delivered by:

Telecommunication and Computer Science Group:

Piotr Detmer, 178627

Grzegorz Izydorczyk, 178628

Bartosz Janel, 178629

Krzysztof Karolczak, 178631

Michał Kunikowski, 178632

Przemysław Piestrzeniewicz, 178633

Adrian Titienko, 178634

# Problem statement

Hypothesis formulation:

*Can skeleton data from Microsoft Kinect be used for creating unique   
gait signatures that will be a reliable base for biometric human identification.*

Gait - the coordinated, cyclic combination of movements that result in human locomotion [1].

Gait recognition is the process of identifying an individual by the manner in which he or she walks. It is an unobtrusive biometric marker, which offers the possibility to identify people at a distance, without any interaction or co-operation from the subject. This is the property which makes it so attractive as a method of identification.

This project aims to develop a software base for semi-automatic gait-based recognition system.

# Introduction

The aim of given task was to develop a set of applications allowing to record, process and classify motion captured sequences obtained from Microsoft Kinect sensor.

Microsoft library supporting skeleton detection for Kinect device was used in order to track the movement of separate joints. Acquired data samples were serialized, thus enabling offline analysis, and processed automatically to obtain set of features according to human gait recognition based on Bezier curves algorithm [3]. Computed features were used for classification using k-NN method preceded by feature extraction using principal component analysis (PCA).

## Kinect

Kinect is a motion sensing input device presented in 2010 by Microsoft as a part of Xbox 360 console. In 2012 new version of the device was presented being compliant with Window operating systems.

Kinect sensor case contains:

* 1280x960 resolution RGB camera allowing to capture colour images.
* Infrared emitter and sensor for casting and reading infrared beams that are converted to depth information measuring distances between the sensor and objects.
* A 3-axis accelerometer with maximum 2G range, which can be used to determine current orientation of the device.
* Microphone array consisting of four separate microphones placed in different locations allowing to record audio, as well as, find the location of the sound and the direction of audio wave.



Fig 1. Kinect device sensors.

Some of specifications of the device are presented below:

|  |  |
| --- | --- |
| Kinect | Array Specifications |
| Viewing angle | 43° vertical by 57° horizontal field of view |
| Vertical tilt range | ±27° |
| Frame rate (depth and colour stream) | 30 frames per second (FPS) |
| Audio format | 16-kHz, 24-bit mono pulse code modulation (PCM) |
| Audio input characteristics | A four-microphone array with 24-bit analogue-to-digital converter (ADC) and Kinect-resident signal processing including acoustic echo cancellation and noise suppression |
| Accelerometer characteristics | A 2G/4G/8G accelerometer configured for the 2G range, with a 1° accuracy upper limit. |

## Interaction space

Interaction space is defined by a field of view by Kinect cameras. This is the area in front of the sensor where infrared and colour sensors are able to track the objects in front. It is a vital part of presented solution, since skeleton tracking is supported only in the boundaries of interaction space. Sensor itself, has a tilt extension enabling to increase the area of interaction.

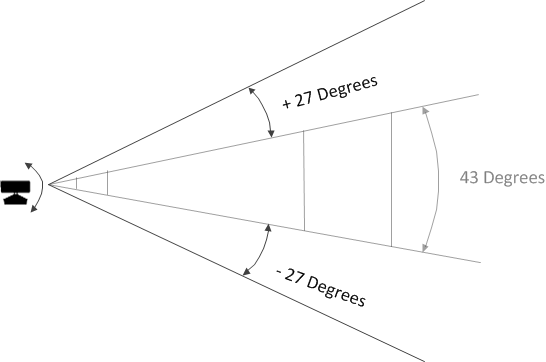


Fig 2. Tilt extension

In default range mode, Kinect can see people standing between 0.8 meters (2.6 feet) and 4.0 meters (13.1 feet) away; users will have to be able to use their arms at that distance, suggesting a practical range of 1.2 to 3.5 meters

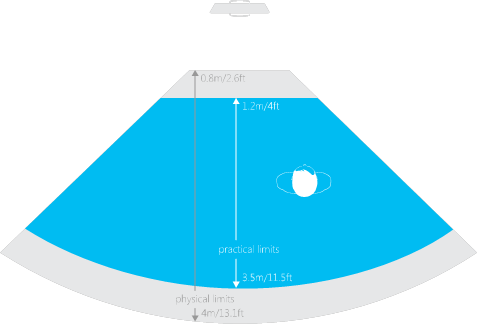


Fig 3. Horizontal field of view in default range.

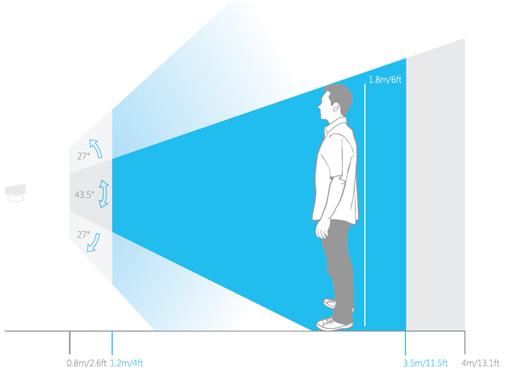


Fig 4. Vertical field of view in default range.

In near range mode, Kinect can see people standing between 0.4 meters (1.3 feet) and 3.0 meters (9.8 feet); it has a practical range of 0.8 to 2.5 meters.

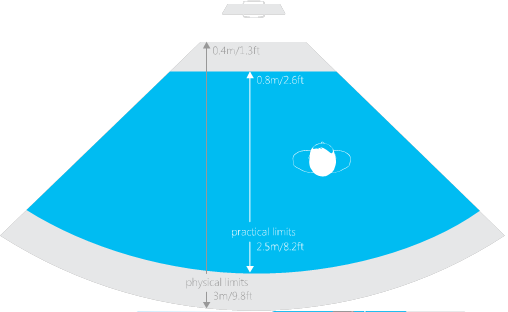


Fig 5. Horizontal field of view in near range mode.

## Skeletal Tracking

With IR camera Kinect is able to recognize 6 people in its field of view and track two of them. Application can locate the joints of users and track their movements over time. This is the main feature of the device used for this project for gait recognition.

Skeletal recognition has two modes of detection: for users standing and sitting, in both cases facing the sensor. The requirement of person tracking with his/her face visible caused some problems during motion capturing and led to some additional sequence processing.

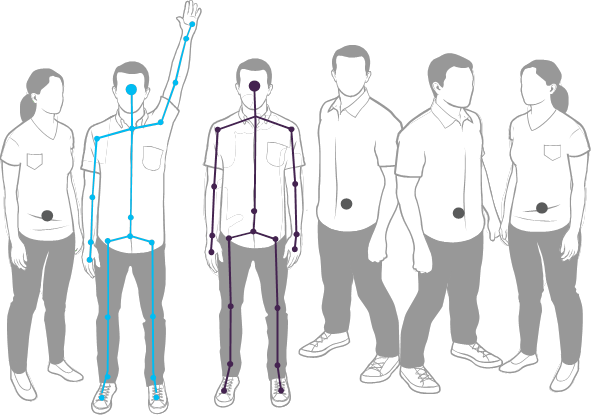


Fig 6. Example of tracking 2 users, with 6 people recognized

In order to be recognized by the device user has to stand in front of the sensor, no additional calibration is required as long as all the body parts are visible for the device.

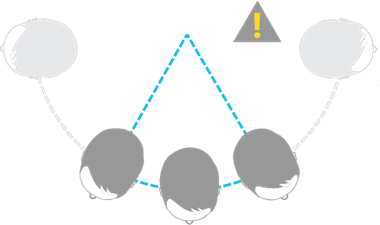


Fig 7. Skeleton tracking works only for users facing the sensor.

Usage of more than one sensor for skeleton tracking may lead to the possibility of interference. No other infrared sources may point at the skeleton at the time when detection is performed.

Behaviour of lost joints for skeleton is unpredictable. Generally lost point will stay at the last detected place or will start jittering together with fluctuation. All the unpredictable joint movements are being filtered later on.

# Kinect-based gait recognition algorithm

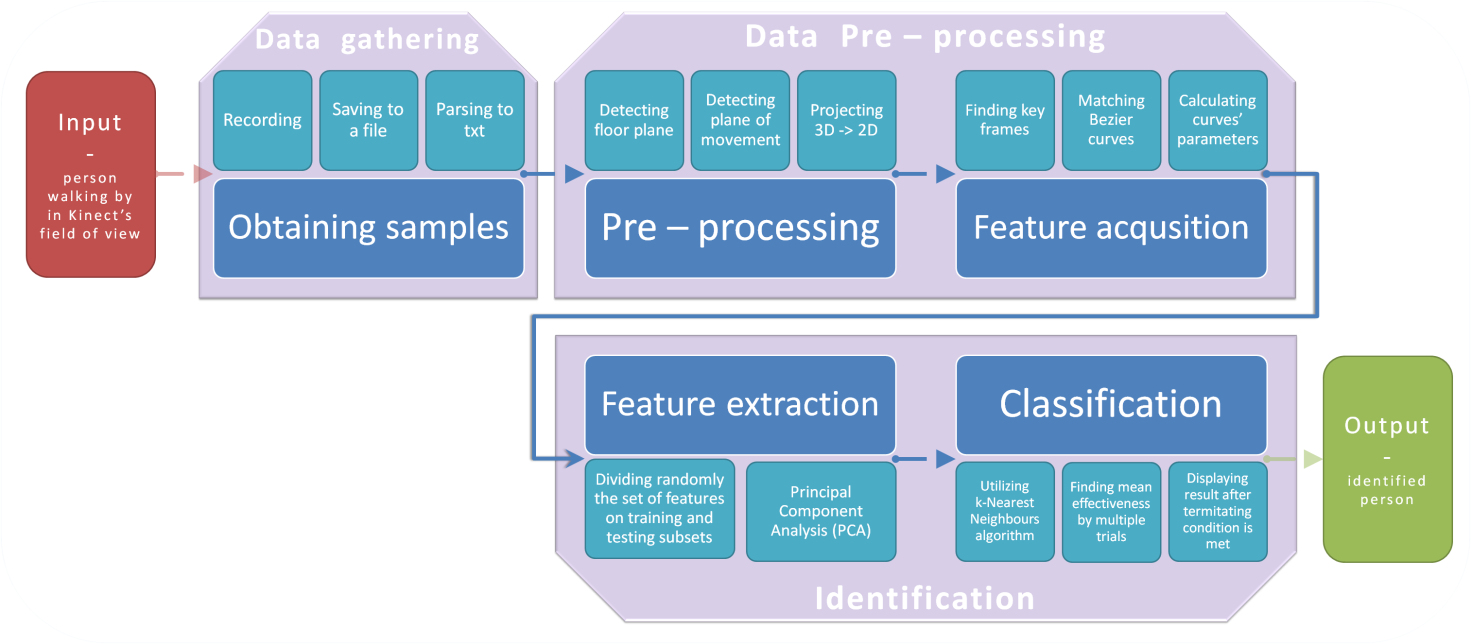


Fig 8. Full description of the data-flow in the created recognition algorithm

## Data pre-processing

The aim of data pre-processing was to create a set of points that could be characterized with the following features:

* Outcome set of points would retain the key features of the input signal
* Outcome set of points would be independent of
  + the direction of a person movement
  + the slope of the surface on which a person is moving
  + the global position of a person body but the relative position of the limbs would be preserved
* Noise of the measurement samples would be minimized
* Random behaviour due to lost points, or jitter of skeleton points would be decreased

Therefore, it would decided to subject the input sample a series of operations resulting in the set of points that would have all the above characteristics. Moreover, to make the further analysis simpler (and hence quicker), dimensionality of input samples was also decreased by projecting on a 2D plane.

Fig 9. Consecutive steps in data pre-processing

All the numeric values of used parameters and most of the joint types to be used were chosen empirically. At first, input samples was subjected to zero-phase 3rd order Butterworth filter of normalized cut-off frequency equal to 0.06. Then all points were translated such that crotch point would be moved to the origin of coordinate system. Based on all samples from the recording the floor plane was computed by minimization the sum of all point-to-plane distances for ankle samples. Having done this, the movement plane was found as one perpendicular to both the floor and back planes (the latter one was determined based on both hips’ positions). Finally, each 3D joint was projected on the movement plane and new coordinates were found according to the main axis defined by floor and back planes’ vectors.

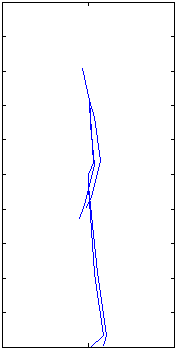
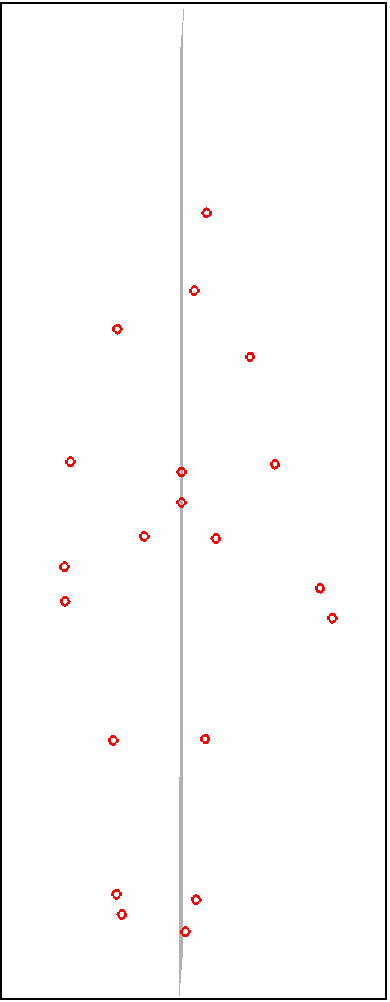
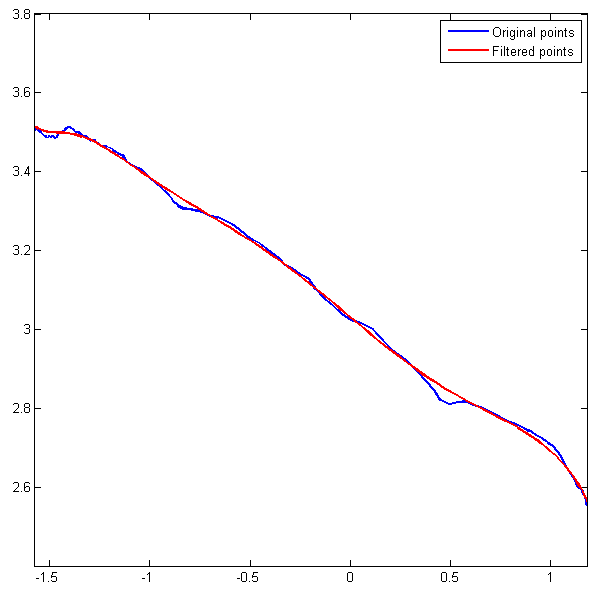


Fig 10. Sample after pre-processing

## Feature selection and classification

An important issue in gait recognition is proper feature selection, that will effectively represent all the gait characteristics. Feature has to be easy to calculate and process and is highly dependent on the hardware used and form of captured frames.

Proposed algorithm starts with **key frame generation.** Key frames are recognized by observing different phases of human movement cycle. From these we obtain four following frames.

* Frame A – right leg straight behind, left leg straight in front, full stride
* Frame B – left leg standing straight, right leg above the ground with knee bent
* Frame C – left leg straight behind, right leg straight in front, full stride
* Frame D – right leg standing straight, left leg above the ground with knee bent

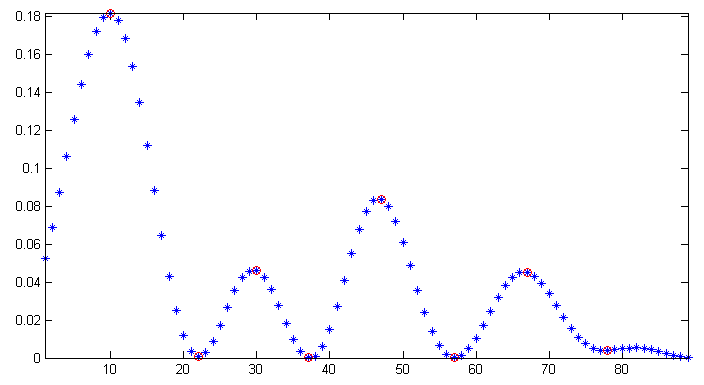


Fig 11. Feet distance as a function of time

For the purposes of formulation of a more robust phase detection algorithm, the above presented conceptual phases had to be slightly redefined. In order to detect the key frames, the algorithm examines the distance *d* between feet along x-axis (horizontal axis). On this basis four phases of gait have been distinguished and defined:

* Phase 1 - the maximum distance between left foot and right foot – Frame A
* Phase 2 – feet intersection in x-axis, between phase 1 and phase 3 – Frame B
* Phase 3 – the maximum distance between right foot and left foot – Frame C
* Phase 4 – feet intersection in x-axis,, between phase 3 and phase 1 – Frame D

Since the algorithm for proper operation needs full gait sequences i.e. containing at least four consecutive phases (1, 2, 3, 4) / (2, 3, 4, 1) / (3, 4, 1, 2) / (4, 1, 2, 3), it verifies if the recorded sequence satisfies this condition. In case the condition is not satisfied the algorithm rejects whole sequence.

Next step is **Bezier curve computation** based on control points chosen for each frame. Initially, five control point described in [2] were selected. However, after several tests, the set of control points was changed to make the Bezier curves for each user as distinguishable as possible.

Therefore, selected control points are: left feet, left ankle, left hip, left shoulder, left hand, neck (since user was approaching the sensor from his/her left side). Bezier curves are generated by 6th order Bernstein polynomials. Exemplary Bezier curves for two different people are shown below.

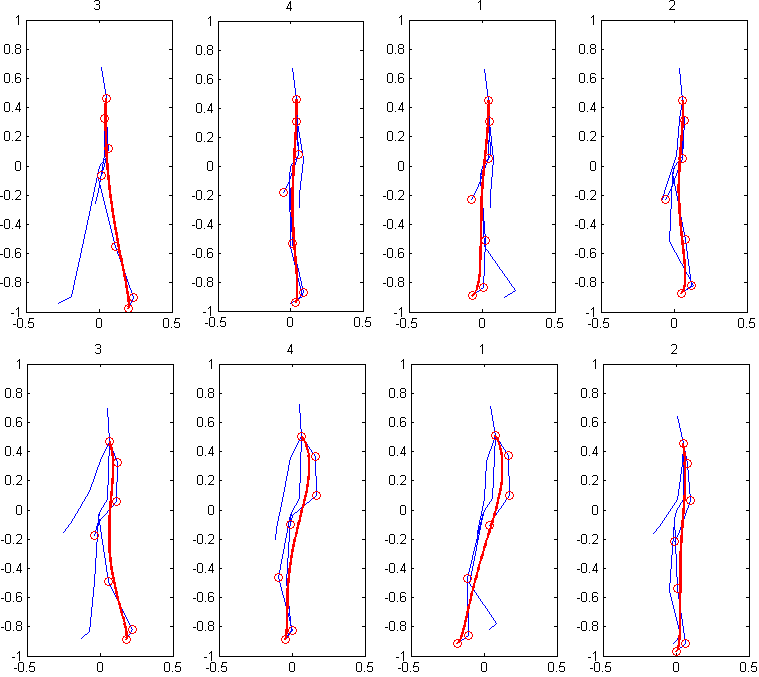


Fig 12. Bezier curves calculated for two subject and four key frames

**Feature generation** is based on previously computed curves and starts with mean value generation for all curve points in all control frames together with variance calculation. All calculations are performed separately for X and Y axis. This results in 3 features for each detected key frame.

Having three full steps recorded the algorithm would detect 12 key frames, find control points, compute curves and calculate features giving us 48 features all together (12 for each step recorded). This was new approach, as in already existing algorithms [2] features were not calculated for each key frame but for whole two steps.

# Results

After completion of the whole algorithm cycle the feature set was evaluated using PCA (Principle Component Analysis). Firstly, features were assessed for its contribution to identification of people. It turned out that selecting a single feature was not enough to properly recognize more than one/two people (Fig. 13).

Hence, the extended analysis was performed, this time PCA was used to extract three best features. Such operation led to significant improvement in classification (Fig. 13). One could conclude that obtaining large set of features that was subjected to PCA algorithm significantly improved the chance of proper detection.

|  |  |
| --- | --- |
| Fig 13. Single feature. | Fig 14. Visualisation of 3D feature space |

# Summary

# References

1. [1] **Biometric Gait Recognition -** Jeffrey E. Boyd, James J. Little / Department of Computer Science University of Calgary, 2011.

[2] **Human Gait Recognition Using Bezier Curves -** Pratibha Mishra et al. / International Journal on Computer Science and Engineering (IJCSE), 2011.

[3] **Microsoft MSDN** - http://msdn.microsoft.com