122nd ASEE Annual Conference & Exposition
June 14 - 17, 2015
Seattle, WA
Making Value for Society
Paper ID #12413

Gait-Based Gender Classification Using Kinect Sensor

Mr. Mohammed Eltaher, Dept. of Computer Science and Engineering, University of Bridgeport, Bridgeport, CT

Mohammed is a full-time Ph.D. student of Computer Science and Engineering at the University of Bridgeport. He received his B.S degree in Computer Science from Sebha University, Libya in 2000 and the M.S. degree in Intelligent System from University Utara Malaysia in 2005. He worked as assistant lecturer at Department of Computer Science, Sebha University from November 2005 to January 2008. Mohammed has research interests in the areas of data mining, information retrieval and multimedia database.

Mr. Yawei Yang, University of Bridgeport

Yawei is a full-time Master student of Computer Science and Engineering at the University of Bridgeport. He received his B.S degree in Software Engineering from Tianjin University of Technology, China in 2012. He worked as assistant at Department of Computer Science, University of Bridgeport from August 2013 to December 2014.

Dr. Jeongkyu Lee, University of Bridgeport

Dr. Jeongkyu Lee is currently associate professor in Department of Computer Science and Engineering at University of Bridgeport. He received Ph.D. in Computer Science from the University of Texas at Arlington in 2006. Before he pursued his doctorate, he worked as a database administrator for seven years with companies including Hana Bank and IBM Korea. His primary research area is in the multimedia database management system and analysis. Research interests include graph-based multimedia data modeling, indexing structure, big data system/analysis, social media data mining, and user profiling.

Gait-Based Gender Classification Using Kinect Sensor

Abstract. Gender classification plays an important role in many applications such as surveillance systems and medical applications. Most of approaches for gender classification are based on the features of human face, voice and gait. Recently, gait-based approaches have been focused on more and more since the way they collect information about human gait is non-contact and non-invasive. Therefore, in this paper we propose a novel method to classify human gender from video stream, which is based on gait features. In order to classify gender, we collect silhouettes of human walking pattern from a Microsoft Kinect sensor, and extract two gait features, i.e., Gait Energy Image (GEI) and Denoised Energy Image (DEI) from a sequence images. GEI is an appearance-based gait representation, while DEI is used to remove the noises from GEI. For the gait features, we use feature vector with low dimension. We employ Support Vector Machine (SVM) for the classifier with the gait-based feature vector. The extracted feature dataset are divided into two parts, i.e., training and testing datasets. The training data set are used for training a SVM classifier while the testing dataset are used for the evaluation. According to the experimental results, we know that GEI is an applicable feature for human gait representation. Despite of the limitation of the dataset, e.g., different races and thickness of clothes which weaken the distinct differences between males and females, the average accuracy of the proposed approach reaches up to 87% under 10 times holdout validation.

1. Introduction

Gender classification plays an important role in many applications such as surveillance systems and medical applications. Most of approaches for gender classification are based on the features of human face, voice and gait. Recently, gait-based approaches have been focused on more and more since the way they collect information about human gait is non-contact and non-invasive. Gait is an information source that can differentiate the difference of gender appropriately, and it can be acquired in a non-intrusive way. In addition, gait features are accessible even at video stream with low resolutions [5].

Gait, referencing to the pattern of walking or locomotion, has been used as an efficient biometric feature in human identification [7]. The goal of this paper is to classify the people's gender whose walking silhouette are captured by Kinect as a sequence of images.

In earlier gait-based gender classification, trackers are attached to the main joints of the body, and sometimes the target subjects, i.e., people, are even asked to wear special suits [4]. These methods are inconvenience and not user-friendly at all. It is difficult to apply such solutions in real applications. With the help of Kinect sensor [9], which provides RGB stream, depth map image stream as well as user skeleton stream simultaneously, it becomes much easier to capture human silhouette images with inexpensive devices.

In this paper, we employ Microsoft Kinect sensor as our input device to capture gait information from peoples. This is stable and reliable approach since we only use the depth stream and skeleton stream, which are not sensitive to illuminating conditions. We utilize Microsoft Kinect to capture human silhouette sequence, and extract the feature vector from GEI, i.e., Gait Energy Image, proposed in [1]. Then, we apply Support Vector Machine (SVM) as a classifier to classify human gender.

In the remainder of this paper, Section 2 introduces related work on gait based gender recognition. Section 3 describes the overview of proposed method. Section 4 specializes the detailed procedure of creating GEI and DEI, and the method to extract feature values. Section 5 elaborates the experimental result, and Section 6 concludes the paper.

2. Related Works

The existing works for gender recognition published in the literatures is based on face or voice. Gait features can also be used for gender classification, and some prominent researchers have worked on it [3]. Some biometric characteristics are very indicative for human gender recognition such as the body sway, shoulder-hip ratio as well as waist-hip ratio. For example, males tend to swing their shoulders more than hips while females swing their hips more than shoulders. Males normally have wider shoulders than females and females normally have thin waists and wider hips [4]. However, it is challenging to extract these model-based features from video stream. There is another category of feature for gait representation, which uses the appearance-based features. Although these features are typically used in gait recognition, they could be used for gender classification.

Gait Energy image (GEI) is one of appearance-based gait representation proposed by Han *et al.* [1]. GEI is used as feature vector for gender classification by Li *et al*. [2]. Yu *et al*. [3] divide GEI into five components and apply different weights for each component. Hu *et al*. [7] use the same method to divide GEI into five components. However, they applied PCA and FLDA to each component to reduce the dimension of feature vectors. Original GEI works well in gender classification. However, the amount of feature values usually is very large and complicated to work out them.

3. Framework of Gender Classification

There are two major components in the proposed system, i.e., training section and testing section. Training section takes samples from the training dataset to find the support vector for classifier. Afterwards, testing section classifies the gender of input silhouettes sequence using the result from training section. Figure 1 shows the overview of the proposed system.

4. GEI and DEI of Gait

This section describes the detailed steps of preprocessing the original data stream captured by Kinect device. The goal of preprocessing is to get the Gait Energy Image (GEI) and Denoised Energy Image (DEI) of a user who walks through the active area of Kinect sensor.

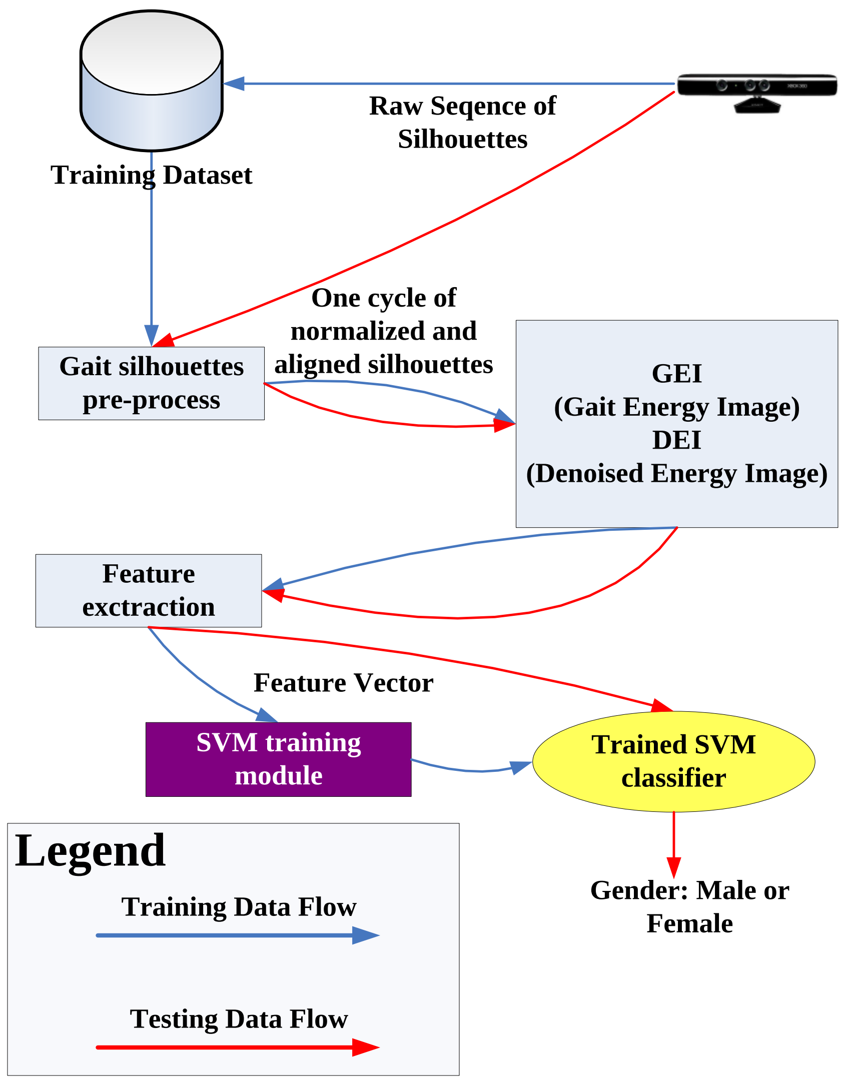

Fig. 1 Overview of Gender Classification System using Microsoft Kinect

Silhouettes Segmentation

Microsoft Kinect is a motion sensing input device for Microsoft XBox 360 and Windows PCs. In addition to the regular RGB camera, Kinect also has a depth sensor which consists of an infrared laser projector together with a CMOS sensor. In the software aspect, there are three data streams for the Kinect sensor including RGB stream, depth map stream and skeleton stream. Specifically, the depth images are particular bitmaps in which each pixel represents the distance between the target object to the sensor in the corresponding position. The skeleton stream includes the spatial information of user torso and limbs. Major joints such as shoulder joint and ankles are labeled in skeleton images.

In this paper, we utilize the depth stream and skeleton stream to capture human silhouettes. The Kinect SDK provides APIs to get human silhouettes from the depth and skeleton stream. Figure 2 shows a sample silhouette from Kinect user segmentation API.

Gait Period Detection

In the real-time video, there are several cycles of gait within a sequence of silhouettes. Therefore, we regard the gait as a sequence of repeating walking patterns. To address the problem, we use only one cycle in the gait to extract the features. Thus, we can identify the boundary between two gait cycles.

Fig. 2. Sample silhouette from Kinect

With the help of skeleton stream in Kinect sensor, we are able to measure the distance variation between two ankles as shown in Figure 3, i.e., blue circled line. We consider the silhouettes between two full stride stances as one cycle. In other words, we pick silhouettes as a sequence of frames from one peak of ankle to another peak, which is regarded as ankles' distance, which is called as a gait cycle.

Note that there might be noises in the ankles' distance sequence due to the instability of the device. We apply 3-points linear smooth algorithm on the original ankles' distance to overcome the noises. Then, we can easily find all peaks within the sequence. Following is the algorithm for smoothing the ankles' distance sequence:

$$dist_i = \frac{dist_{i-1}+dist_i+dist_{i-1}}{3} \; ; i = 1,2,\dots,n-2 \qquad \textbf{(1)}$$

, where $dist_i$ represents the ankles' distance in the i^{th} silhouette, n is the amount of all silhouette frames. We could apply this filter into the original sequence several times to get rid of most noises. In figure 3, we plot the smoothed ankles' distance in red stared line.

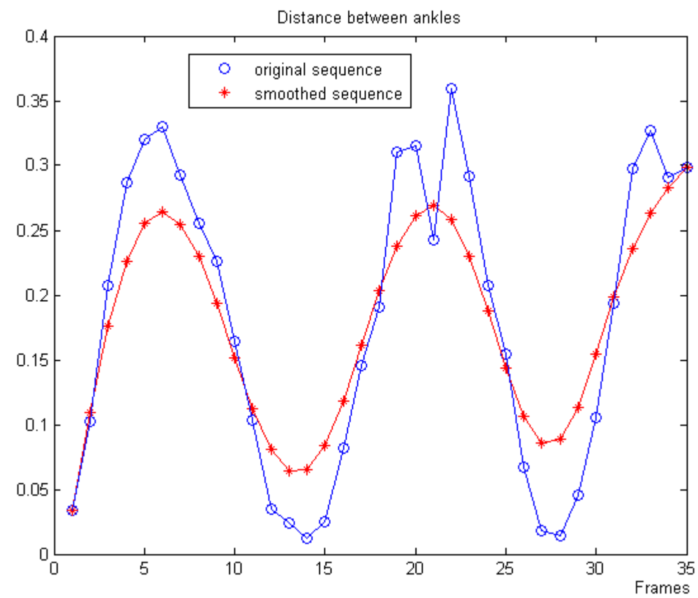

Fig. 3. Blue circled line indicates original ankles' distance sequence while red stared line indicates smoothed ankles' distance sequence.

Silhouettes Normalization

The captured silhouettes may be different in size due to the variation of distance between user and Kinect camera, thus it is necessary to apply normalization to these silhouettes images. In this step, we first trim the gait frames into rectangles that fit the human silhouettes, and then resize them to ensure that all silhouettes have the same height. In our experiment, we take 180 pixels height for all silhouette images. Figure 4 shows a silhouette after normalization applied.

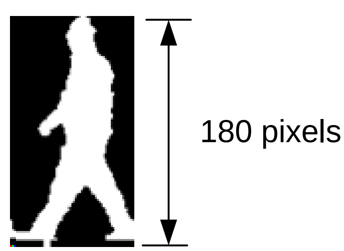

Fig. 4. A normalized silhouettes

Gait Energy Image (GEI)

Gait energy image is proposed by Han *et al* in [1], which represents both spatial and temporal information of gait silhouettes in one image. This energy image has dramatically reduced the storage space of gait comparing with the original binary a sequence of silhouettes. In addition, it will take much less time to process gait than to analyze the original sequence of silhouettes.

In order to create GEI, first we center the upper half of the silhouette with respect to its horizontal centroid to ensure that the torso is aligned in the sequence. Afterwards, we create the GEI by computing the average of a sequence of silhouettes in a gait cycle (Composed of T frames) [5]. Equation (2) is the formal definition of GEI:

$$F_{i,j} = \frac{1}{T}\sum_{t=1}^{T} I_{i,j,t} \quad \textbf{(2)}$$

, where *i* and *j* are the image coordinates, and $I_{i,j,t}$ is the binary silhouette image obtained from the t^{th} frame [5]. Therefore, GEI is an average image of a sequence of silhouette in a gait cycle. Figure 5 shows an example of GEI for a sequence of silhouette images. Note that we take a square whose side width equals to the silhouettes' height as background. Therefore, all the GEIs will have the same dimension. GEI is employed due to the following reasons:

- GEI is time-normalized and it accumulates human walking silhouettes within a complete cycle.
- GEI has weakened the noise in each single silhouette image.
- A pixel with higher intensity value in GEI means that human walking occurs more frequently in this position [1].

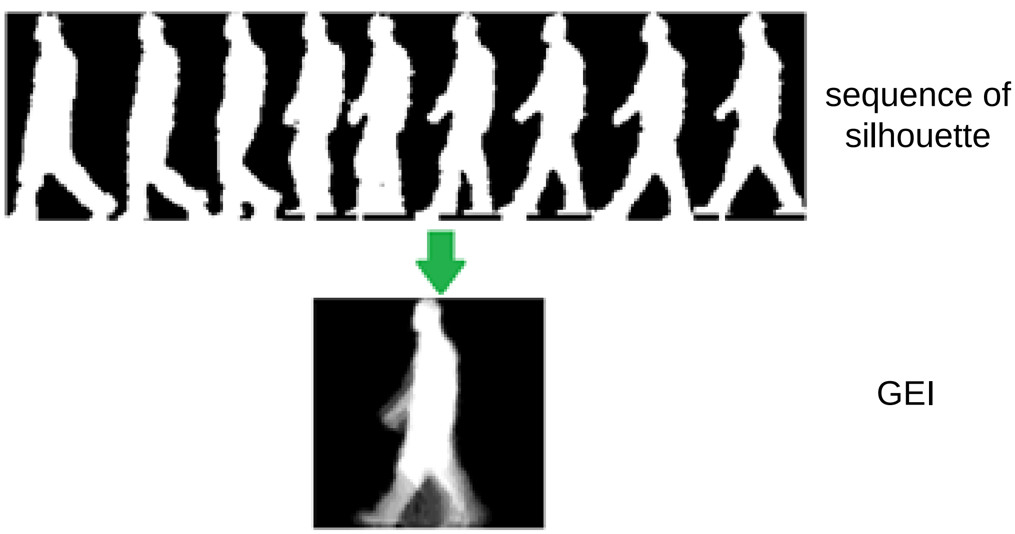

Fig. 5. Example of GEI

Denoised Energy Image (DEI)

After we have GEI from a sequence of silhouettes, DEI, i.e., Denoised Energy Image, is computed. DEI can be created by removing noises from GEI [6]. Using DEI, the giant variance between male and female can be acquired. DEI can be obtained as follows:

$$D_{i,j} = \begin{Bmatrix} 1, if\ G_{i,j} \geq T \\ 0, otherwise \end{Bmatrix} \quad \textbf{(3)}$$

, where $D_{i,j}$ is the pixel after noise removal, and $G_{i.j}$ is the pixel values of the GEI image. T is a threshold value. Figure 6 is an example of DEI.

Fig. 6. DEI

Feature Extraction and SVM Classifier

Both [2] and [3] take entire GEI image as feature vector, which has a very high dimensional data. When the size of GEI is 180 pixel × 180 pixels, the feature vector would have 32,400 elements. In order to address the problem, we introduce a simple method to extract lower-dimension feature vector from DEI. The proposed feature vector *F* is defined as follow:

$$F = \{X, Y\} \quad \textbf{(4)}$$

, where *F* represents the feature vector, *X* is a sub-vector of horizontal features and *Y* is sub-vector of vertical features. Following is the definition of sub-vector *X* and *Y*.

$$X_i = \sum_{j=0}^{DW-1} D_{i,j} \quad \textbf{(5)}$$

$$Y_i = \sum_{j=0}^{DH-1} D_{i,j} \quad \textbf{(6)}$$

, where X_i is the i^{th} horizontal feature value, Y_j is the j^{th} vertical feature value, *DW* and *DH* represent the width and height of DEI, respectively. *i* and *j* are the coordinates of DEI. According to the definition, *X* and *Y* show the distribution of white pixels in DEI from both horizontal and vertical direction. Moreover, this feature vector has only (*DW* + *DH*) in size. In case of taking 180 × 180 pixels DEI as an example, the feature vector size will be only 360 elements.

SVM (Support Vector Machine) is a supervised learning model with associated learning algorithms that analyze data and recognize patterns. We employed SVM technique in our system to train and test the extracted feature vector for gait-based gender classification. SVM can simultaneously minimize the empirical regression error and maximize the geometric margin between different classes. A hyper-plane will be found in multi dimensional space to separate different classes [3]. In our system, we employ SVM algorithm provided by EmguCV [7] library. We take the polynomial kernel function and auto training option to initialize the SVM object.

5. Experimental Results

In order to evaluate the proposed algorithm, we perform the experiment with Kincet sensor. As shown in Figure 7, we place Kinect on a table of 1.3 meter height, i.e., 4.3 feet. The users walk through in the direction vertical to the central sight line of Kinect.

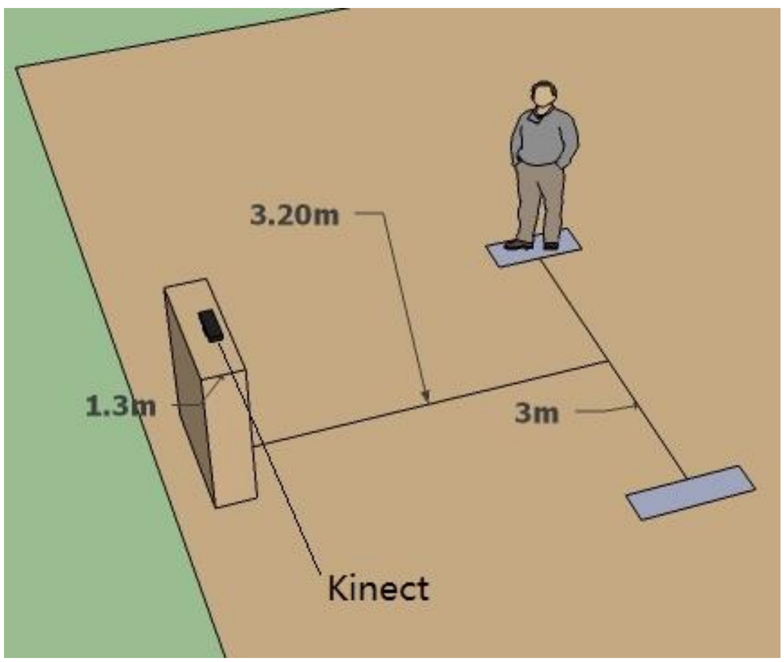

Fig. 7. Experimental setup using Kincet sensor

Regarding to the implementation, we program on a computer with Intel Core i5-2430 2.0 GHz CPU, 4 GB DDR main memory. The Kinect sensor is connected to this computer through USB port. We program under 64-bit Windows 7 operating system with Kinect SDK 1.7, XNAS40 as well as EmguCV version 2.9.0. The programming IDE for this project is Microsoft Visual Studio 2010 with .NET framework 4.0.

We have 72 volunteers including 38 female and 34 male, and ask them to walk in front of the sensor back and forth. The program records gait silhouettes as well as gender of the tested person.

The following are the steps for training and testing:

Step1*:* Pick one cycle from each individual's silhouettes sequence.

Step2: Compute GEI/DEI and extract feature vectors for all of them.

Step3: Shuffle all these feature vectors and randomly divide them into two groups (52/20). We labeled the larger data set with known genders and use them to train the classifier. We use the remaining dataset for testing.

Step4: Repeat Step3 for 10 times.

By executing the steps above, i.e., 10 times holdout validation, we get the accuracy of our methods is between 80% and 95%, and the average accuracy of 10-folds validation is 87% with 5.9% standard deviation as shown in Table 1. The main reason of 5.9% standard deviation is because of relatively small testing data set, i.e., missing 1 person means losing 5% (i.e., 1 out of 20) in terms of accuracy. Table 1 shows the summary of experimental results.

According to the experimental results, we know that GEI is an applicable feature for human gait representation. However, the proposed approach in this paper have several important limitations including (1) *thickness of clothes* which weaken the distinct differences between males and females, (2) *different cultures* which affect the characteristics of gaits, and (3) *Kinect setup* which is important to get the proper feature values. In order to overcome such limitations, our future works will focus on developing more reliable feature vector from gaits of human. Specifically, 3-D based gait features will help us to improve the accuracy.

Table 1. Experimental results

	Training Samples	**Testing Samples**	**Male/Female (training)**	**Male/Female (testing)**	**Accuracy (%)**
1	52	20	26 / 26	8 / 12	90
2	52	20	28 / 24	6 / 14	80
3	52	20	27 / 25	7 / 13	95
4	52	20	23 / 29	11 / 9	90
5	52	20	24 / 28	10 / 10	80
6	52	20	22 / 30	12 / 8	85
7	52	20	25 / 27	9 / 11	90
8	52	20	26 / 26	8 / 12	95
9	52	20	25 / 27	9 / 11	80
10	52	20	25 / 27	9 / 11	85
Average					**87%**

6. Concluding Remarks

In this paper, we propose a method to classify human gender based on their gaits using Kinect sensor. To address the problem, our proposed approach extracts two gait features, i.e., Gait Energy Image (GEI) and Denoised Energy Image (DEI) from a sequence. For the gait feature, we use a feature vector with a low dimension in this paper. The extracted feature dataset are divided into two groups, i.e., training and testing datasets, for SVM classifier. The experimental results show that the proposed method reaches 87% on average under 10 times holdout validation.

References

1. Ju Han, and Bir Bhanu "Individual Recognition Using Gait Energy Image" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, No.2, pp.416-322, Feb 2006

2. Xuelong Li, Stephen J. Maybank, Shuicheng Yan, Dacheng Tao, and Dong Xu "Gait Components and Their Application to Gender Recognition" *IEEE Transactions on systems, MAN and Cybernetics-Part C: Applications and Reviews*, Vol.38, No.2, pp.145-155, Mar 2008

3. Shiqi Yu, Tieniu Tan, Kaiqi Huang, Kui Jia, and Xinyu Wu "A Study on Gait-based gender classification" *IEEE Transactions on Image Processing*, Vol 18, No.8, Aug 2009

4. N.F.Troje "Decomposing biological motion: A framework for analysis and synthesis for gender classification of human gait patterns" J.Vis., Vol.2 No.5, pp.371-387, 2002

5. Laura lgual, Agata Lapedriza, and Richard Borras "Robust gait-based gender classification using depth cameras" E*URASIP Journal on Image and video processing*, 2013

6. Li-Hong Juang, Shin-An Lin, and Ming-Ni Wu "Gender Recognition Studying by Gait Energy Image Classification" International Symposium on Computer, Consumer and Control, 2012

7. Maodi Hu, Yunhong Wang "A New Approach for Gender Classification Based on Gait Analysis" *Image and Graphics, 2009. ICIG '09. Fifth International Conference*, pp. 869-874

8. EmguCV, http://www.emgu.com/wiki/index.php/Main_Page

9. Kinect, Microsoft Co., http://www.xbox.com/en-US/kinect