A Sitting Posture Surveillance System Based on Kinect

Si He

College of Information Engineering, Wuhan University of Technology, Wuhan, China

Abstract. Modern life has made more and more people become sedentary, and long periods of unhealthy posture can lead to myopia, cervical spine and lumbar disease. Therefore, this paper presents a sitting posture surveillance system based on Microsoft Kinect. By getting joints’ three-dimensional position information from Kinect’s real-time skeleton tracking technology, and input the data into computer. After processing the joints’ location information with posture recognition algorithm, computer recognizes the user’s sitting posture, and warns the incorrect postures.

1. Introduction
In daily life, whether it is work, study or relaxation, sitting is the most commonly used posture, but the long-term incorrect sitting posture can cause great harm to the health of the human body. Unhealthy sitting posture can lead to the distance too close to the eye, and if things continue in this way, the ciliary of the eyes may lose their ability to regulate themselves and cause myopia[1]. And long periods of abnormal physiological curvature of the spine are likely to cause spinal deformity and various cervical diseases.

The occurrence and even getting into a habit of bad sitting posture is often involuntary, timely remind or correction from outside can effectively help keeping good sitting posture to habit. Traditional posture correction methods, besides human supervision and reminding, mostly are wearable physical orthotics or simple electronic orthotics based on acceleration sensors or level sensors. The effect is immediate but the precision is insufficient, and wearing for a long time can also hamper the body movement, bringing extra burden to users[2]. By contrast, using visual capture technology, through video or static images, judging posture by processing visual data is more accurate and flexible[3].

Gesture recognition is a hotspot in the field of machine vision and is widely used in human-computer interaction, behavior analysis, multimedia application and sports science[4].

2. Kinect
With the rapid development of camera technology in recent years, the cameras with 3D camera has gradually entered our horizon. TOF cameras emerged in the 1990s, the principle of which is to calculate the flight time of light and obtain the depth information of objects. Compared with the traditional 2D cameras, TOF cameras can easily separate foreground and background in the scene, thus the TOF cameras have the advantage in the aspect of undetermined target tracking and recognition. But the cost and relatively low resolution of the TOF cameras affected its development.

For the first time in November 2010, Microsoft launched a motion sensing device—Kinect. Kinect has functions of body and skeleton real-time tracking, motion capture and the microphone input, and can identify a series of actions. People can stand in front of the device and interact with computers...
through a variety of gestures. Kinect greatly reduce operating costs of human-computer interaction. Therefore, the research on the natural human-computer interaction based on Kinect has become a hotspot[5].

Different from normal cameras, Kinect senses the outside world with a CMOS infrared sensor. The sensor apperceives outside environment by black and white spectrum: pure white represents an infinite close, pure black represents infinity far, according to the physical distance to the sensor, objects are represented by corresponding grey level. It can sensitively perceive the space within the device's field of vision and generate a depth image represents the surrounding environment. The sensor generates a depth image stream at the frequency of 30 frames per second, making a real-time 3D reconstruction of the surrounding environment. Kinect detects 3D depth images and transfers them to skeletal tracking systems. The system can track the operator's 20 bone nodes, including the torso, limbs and fingers. In order to understand the user's action, Microsoft used machine learning technology and established a large image database, realized intelligent recognition that can understand the meaning of the user’s movement as far as possible[6].

![Figure 1. Physical and schematic diagram of Kinect.](image)

In June 2014, Microsoft launched the upgraded Kinect, Kinect2.0, with its physical and schematic diagram as shown in figure 1. Compared with the first-generation Kinect product, Kinect2.0 that we use in this paper enhanced the depth sensing capability, which can distinguish smaller objects through higher depth fidelity and dramatically improved noise base. Video resolution is upgraded from VGA level to 1080p level, providing full high definition video and smooth scene experience, the data processing amount can be up to 2GB/s; With USB3.0 interface, the delay reduced to 2 frames/s; Microsoft has also used the TO method to replace the existing "light measurement" method, which makes Kinect2.0 faster and more accurate than the first generation. The vision of Kinect2.0 increased 60%, support 6 operators at the same time, and can track each person's 25 joints information at the same time, optimize the tracking algorithm, and make the tracking posture more accurate and stable. Improved the detection performance of infrared sensor that can even work in the dark. In addition, Kinect2.0 also provides thumb tracking, fingertip tracking, gesture recognition, so that more detail elements can be added to future motion sensing applications.

3. Joint location

3.1 Posture recognition

Posture is a relationship of static position the between body parts, a posture can be uniquely determined by the relative position of joints, and the complexity of the postures directly affects the complexity of the recognition algorithm.
3.2 Skeleton capture
Because the amount of information of the depth image that the Kinect generates is limited, the Kinect will need extra information beyond depth data to develop applications with real value. That's why the skeleton tracking technology is presented. The characteristics of skeleton recognition are the integration of anthropometry and gait identification, identify skeleton by the physical characteristics and behavioral characteristics of the human body. Specifically, the skeleton recognition system accesses the human body static information from Kinect’s skeleton tracking function, then processes the depth image with methods such like matrix transformation and machine learning to get the coordinates of the joints[7]. Figure 2 shows the 25 joints that Kinect2.0 captures.

4. Posture identification

4.1 Judgement the credibility of joint’s information
Because of occlusion and other causes in the process of body movement, Kinect might lose some joints or recognize the wrong information of the joints, which can cause a serious impact on the recognition of the sitting posture. It is necessary to judge the credibility of the joints and recover the information.

4.1.1 Preliminarily judge according to the length of each part

![Figure 3. Schematic of arm joints.](image)

Take the arm as an example, figure 3 shows the schematic of the human arm joints, A is the shoulder joint, B is the elbow joint, and C is the wrist joint. Suppose the length of the upper arm is \( L_{AB} \), lower arm's length is \( L_{BC} \), joints’ information of the Kth frame are \( A(x_a, y_a, z_a) \), \( B(x_b, y_b, z_b) \), \( C(x_c, y_c, z_c) \), and assumes that the information of A is reliable, to determine the credibility of joint B. \( L_{ABK} \) is the length of the upper arm of the Kth frame, and set a threshold \( \tau \). When \( L_{ABK} \in (L_{AB} - \tau, L_{AB} + \tau) \), it is considered that the joint B is initially reliable, otherwise it is considered untrustworthy.

4.1.2 Further judge according to the continuity of movement
The movement of the joints is continuous, within a certain amount of time the velocity of the joints is limited, if the velocity beyond a certain range that the joint’s information is not credible.
First assumes that the joint B moves in uniform motion in each frame, set the time interval between the two frames as $t$, the coordinates of joint B at the (K-1)th frame and the Kth frame are $B(x_{b,K-1}, y_{b,K-1}, z_{b,K-1})$ and $B(x_{b,K}, y_{b,K}, z_{b,K})$. Therefore, the velocity of joint B at the Kth frame is:

$$
\begin{align*}
V_{Bx,K} &= \frac{x_{b,K} - x_{b,K-1}}{t} \\
V_{By,K} &= \frac{y_{b,K} - y_{b,K-1}}{t} \\
V_{Bz,K} &= \frac{z_{b,K} - z_{b,K-1}}{t}
\end{align*}
$$

(1)

Similarly, according to the formula can evaluate the velocity of joint B in each direction at the (K-1)th frame, set the threshold of velocity in each direction as $(v_x, v_y, v_z)$. According to a few frames before, estimate the velocity of B at the Kth frame is $(V_{Bx,K}, V_{By,K}, V_{Bz,K})$. And the real-time velocity of joint B at the Kth frame of should meet the formula 2, otherwise it is considered unreliable.

$$
\begin{align*}
V_{Bx,K} &\in (V_{Bx,K} - V_x, V_{Bx,K} + V_x) \\
V_{By,K} &\in (V_{By,K} - V_y, V_{By,K} + V_y) \\
V_{Bz,K} &\in (V_{Bz,K} - V_z, V_{Bz,K} + V_z)
\end{align*}
$$

(2)

4.2 Recovery of the joint's information
4.2.1 Recovery of intermediate joint
If joint A and C is known and reliable, B is not. The arm model shows that the joint B should be on the surface of the sphere $Q_A$ which centered at the joint with a radius of $L_{AB}$ and the sphere $Q_C$ which centered at the joint with a radius of $L_{BC}$ at the same time, and the two spheres intersect at the circle $S_B$, as shown in figure 4.

Figure 4. Two spheres intersect at $S_B$.

Suppose the coordinate of the joint B at the Kth frame is $P_{b,K}(x_{b,K}, y_{b,K}, z_{b,K})$. Therefore:

$$
\begin{align*}
L_{AB}^2 &= (x_{b,K} - x_{a,K})^2 + (y_{b,K} - y_{a,K})^2 + (z_{b,K} - z_{a,K})^2 \\
L_{BC}^2 &= (x_{b,K} - x_{c,K})^2 + (y_{b,K} - y_{c,K})^2 + (z_{b,K} - z_{c,K})^2 \\
(x_{b,K} - x_{b,K-1})^2 &+ (y_{b,K} - y_{b,K-1})^2 + (z_{b,K} - z_{b,K-1})^2 = V_{Bx,K-1}^2 + V_{By,K-1}^2 + V_{Bz,K-1}^2
\end{align*}
$$

(3)

There may be one intersection or two intersections situations. There is only one joint B, which can exclude two intersections situation. There might be no solution because of the errors in the length measurement and human tracking.

4.2.2 Recovery of terminal joint
For the recovery of terminal joint, can be performed according to the length of $L_{BC}$ and the velocity of the joint C. We know the average velocity of the joint C, and the speed keeps constant in a very short time, assuming that $V_{Cx} < V_{Cy} < V_{Cz}$ . Selects the two smaller velocity components of the joints, and then there is:
It is knowable from above that the angular velocity of joint C of i frames before is valued in the range of $(\alpha - \delta, \alpha + \delta)$. It is credible if the angular velocity of the recovered joint C valued within the range, if not, then select the other two velocity components $V_{Cy}$ and $V_{Cz}$, if is still not credible, select $V_{Cx}$ and $V_{Cz}$. As all three groups is beyond the range of threshold value, then take the group which is the most close to $\alpha$.

### 4.3 Coordinate transformation

![Figure 5. Coordinate systems of body and Kinect.](image)

Because of the need for work, users can’t always directly face the sitting posture surveillance system. When the user’s body rotate a small angle($\alpha$) around the axis that is perpendicular to the horizontal plane, as shown in figure 5, the coordinate system with subscript k represents the coordinate system of Kinect, and the coordinate system with subscript b represents the coordinate system of the user’s body. When a forward leaning motion occurs, the coordinates of shoulders and neck would lean to the negative direction of $X_k$ axis relative to the coordinates of spine or hip, the projection would not only on the negative direction of $Z_b$ axis as when user is facing the device, but also create a component in the direction, and would be judged as leaning to the right. Similarly, the left leaning motion would create a forward leaning component, and the right leaning motion would create a backward leaning component, so directly use coordinate system of Kinect can’t recognize the postures accurately. Therefore, coordinate transformation is required before posture recognition, transform the joints’ location data from Kinect coordinate system to the body coordinate system.

In most cases, the shoulders maintain a same altitude in the sitting posture. Due to the relative stable location relationship between the user’s shoulders. Calculate the angle $AG_\alpha$ between the connection line of the Shoulder_Left joint and the Shoulder_Right joint and the line that is parallel to the $X_k$ direction. As $AG_\alpha$ is known, the coordinate transformation is to transform the coordinates from Kinect coordinate system to the body coordinate system by anticlockwise rotating the coordinate system $X_kOY_k$ around the $Y_k$ axis by $AG_\alpha$ degrees, and the coordinate transformation formula is shown in formula 5.

$$
\begin{bmatrix}
x_b \\
y_b \\
z_b
\end{bmatrix} =
\begin{bmatrix}
\cos \alpha & 0 & -\sin \alpha \\
0 & 1 & 0 \\
\sin \alpha & 0 & \cos \alpha
\end{bmatrix}
\begin{bmatrix}
x_k \\
y_k \\
z_k
\end{bmatrix}
$$

### 4.4 Evaluation of tilt angle

In correct sitting posture, the cervical spine and lumbar vertebrae should be straight and perpendicular to the horizontal plane. When the cervical spine and lumbar vertebra are not in the same line or not perpendicular to the horizontal plane, the user is leaning to the left, right, forward or doing other bad
posture. Therefore, the core of judging the sitting posture is to evaluate the tilt angle of the cervical vertebra and lumbar vertebra. Since the system is generally placed in a relatively close area and is mainly used to monitor the posture and movement in sitting, it is generally only need to detect the 17 joints include Spine_Base joint and above. For easy detection, the nonlinearity of the curvature of the spinal is ignored.

Using the law of cosines, we use the coordinates of the Head joint, Neck joint and Spine_Shoulder joint to calculate the tilt angle of cervical vertebra ($AG_{CV}$), and calculate the tilt angle of lumbar vertebra ($AG_{LV}$) with the coordinates of the Spine_Shoulder joint, Spine_Mid joint and Spine_Base joint. Because we only need to evaluate the side tilt angle and the front tilt angle, the joints’ coordinates are projected onto the $X_bOY_b$ plane and the $Y_bOZ_b$ plane.

To evaluate the front tilt angle of cervical vertebra ($AG_{FCV}$) in the plane of $Y_bOZ_b$, we calculate the distance from Head joint to Neck joint, Neck joint to Spine_Shoulder joint, and Head joint to Spine_Shoulder joint first, as the formula 6. Then evaluate the $AG_{FCV}$.

\[
\begin{align*}
D_{FHN} & = \sqrt{(y_h - y_n)^2 + (z_h - z_n)^2} \\
D_{FNSpsh} & = \sqrt{(y_n - y_{spsh})^2 + (z_n - z_{spsh})^2} \\
D_{FHSpsh} & = \sqrt{(y_h - y_{spsh})^2 + (z_h - z_{spsh})^2} \\
AG_{FCV} & = \arccos\left(\frac{D_{FHN}^2 + D_{FNSpsh}^2 - D_{FHSpsh}^2}{2D_{FHN}D_{FNSpsh}}\right)
\end{align*}
\]

Similarly, the front tilt angle of lumbar vertebra ($AG_{FLV}$), the side tilt angle of cervical vertebra ($AG_{SCV}$) and the side tilt angle of lumbar vertebra ($AG_{SLV}$) can also be evaluated.

Because the side tilt is divided into left-leaning and right-leaning, and the cosine law used in this system is scalar calculation, it can’t get the direction of the side tilt angle directly. Therefore, it is necessary to add a direction judgment module, which adopts the most direct method of judgment here. If the cervical spine and spine are on the left side of the vertical line, it is left-leaning, whereas the other side is right-leaning.

5. Implement of the system

The sitting posture surveillance system is based on BodyBasics-D2D of Kinect for Windows SDK, using C++ language in the Visual Studio development environment. Acquisition and processing the 17 upper body joints data, identifying the user’s sitting posture in real time according to the above-mentioned evaluation method of the tilt angle of the cervical vertebra and lumbar vertebra, and report the bad posture in real time. The interface of the system as shown in figure 6, with the user’s skeleton information shown on the left side and the evaluation value of tilt angle of the cervical vertebra and lumbar vertebra be shown on the right in real time. When the user is sitting in the right posture, the sitting posture identification displays “Perfect sitting posture!” When the body is too forward, that is, any $AG_{CV}$ or $AG_{LV}$ is too large, use 15 degrees as the determining threshold here, sitting posture identification displays “You are leaning forward, please adjust sitting posture in time!” Similarly, when the body is too left or right, the sitting posture identification displays “You are leaning to the left, please adjust sitting posture in time!” and “You are leaning to the right, please adjust sitting posture in time!”

Due to contingency of the occurrence of bad sitting posture, such as the user is adjusting sitting position or get up to get items, or error occurs in joint data, the system would give the wrong posture remind and record wrong statistics. And it can seriously disturb the proper work of the user if the warning of incorrect sitting posture is too often, so on the basis of the above sitting posture surveillance system, add the time statistics module for each sitting posture. With the module, the system reports the error posture if and only if a certain posture’s duration exceeds a certain limit (here set to 5 seconds), if the posture is corrected during the time, the timer will be reset. When and only
when next error sitting posture occurs, the timer starts again. And in the meantime, there will be no any prompt information which greatly enhance the practicability of the system.

Figure 6. Interface of the system.

6. Conclusion
In this paper, the sitting posture surveillance system based on Microsoft Kinect motion-sensing development has successfully implement the real-time surveillance of the sitting posture of the user, with high accuracy of recognition. The system is only a combination of a Kinect device and a computer, with the advantages of low cost and high practicability. Although in most cases, Kinect has high recognition rate, and can get the joints’ coordinates accurately. For the joints covered by others parts of the body, the result of recognition is not satisfactory. The research in restoring the coordinates of the covered joints can further improve the accuracy rate of body recognition.

7. References
[1] Jones-Jordan L A, Mitchell G L, Cotter S A, etc. 2011 Visual activity before and after the onset of juvenile myopia Investigative Ophthalmology & Visual Science 52(3) 1841-1850.
[2] Jaimes A 2005 Sit straight (and tell me what I did today): a human posture alarm and activity summarization system In Proceedings of the 2nd ACM workshop on Continuous archival and retrieval of personal experiences 23-34.
[3] Wu L, Li K, Wu C 2010 A sitting posture surveillance system based on image and signal processing technology in 2010 2nd International Conference on Computer Engineering and Technology (ICCET), IEEE, 692-695.
[4] Kar A, Amitabha A, Guha M P 2010 Skeletal Tracking using Microsoft Kinect Methodology 1-11.
[5] Han J, Shao L, Xu D, et al. 2013 Enhanced computer vision with Microsoft Kinect sensor: a review IEEE Transactions on Cybernetics 43(5) 13-18.
[6] Boulos M N, Blanchard B J, Walker C, et al. 2011 Web GIS in practice X: a Microsoft Kinect natural user interface for Google Earth navigation International Journal of Health Geographics, 10(1) 45.
[7] Webb J, Ashley J 2012 Beginning Kinect Programming with the Microsoft Kinect SDK: Apress.