|  |  |
| --- | --- |
| Project | **Human Factor for Immersive Content Working Group**  <<http://sites.ieee.org/sagroups-3079/> **>** |
| Title | **Human Body Keypoint Data Format for Mixed Reality Motion Recognition** |
| DCN | **3079-21-0008-01-0002** |
| Date Submitted | **July 15, 2021** |
| Source(s) | **Yoon, Kyoungro**  yoonk@konkuk.ac.kr **(Konkuk Univ.)**  **Bae, Hyo Chul**  th1q@nate.com **(Konkuk Univ.)**  **Jeong, Sangkwon Peter** [ceo@joyfun.kr](mailto:ceo@joyfun.kr) **(JoyFun Inc.,)** |
| Re: |  |
| Abstract | This standard defines an interface that the basis for the transfer of human body keypoint information between applications in virtual reality and users in the physical world. |
| Purpose | The purpose of this standard is a body keypoint data format for transferring user body keypoint information from the physical world to the virtual world. |
| Notice | This document is offered as a basis for discussion and is not binding on the contributing individual(s) or organization(s). The material in this document is subject to change in form and content after further study. The contributor(s) reserve(s) the right to add, amend or withdraw material contained herein. |
| Release | The contributor grants a free, irrevocable license to the IEEE to incorporate material contained in this contribution, and any modifications thereof, in the creation of an IEEE Standards publication; to copyright in the IEEE’s name any IEEE Standards publication even though it may include portions of this contribution; and at the IEEE’s sole discretion to permit others to reproduce in whole or in part the resulting IEEE Standards publication. The contributor also acknowledges and accepts that IEEE 3079 may make this contribution public. |
| Patent Policy | The contributor is familiar with IEEE patent policy, as stated in [Section 6 of the IEEE-SA Standards Board bylaws](http://standards.ieee.org/guides/opman/sect6.html#6.3) <[http://standards.ieee.org/guides/bylaws/sect6-7.html#6](http://127.0.0.1:4664/cache?event_id=757737&schema_id=1&s=5X0vID10lu_E6yrIkWkNd4Wz2H8&q=hancock)> and in *Understanding Patent Issues During IEEE Standards Development* <http://standards.ieee.org/board/pat/faq.pdf> |

Human body key point data format for Mixed Reality Motion Recognition

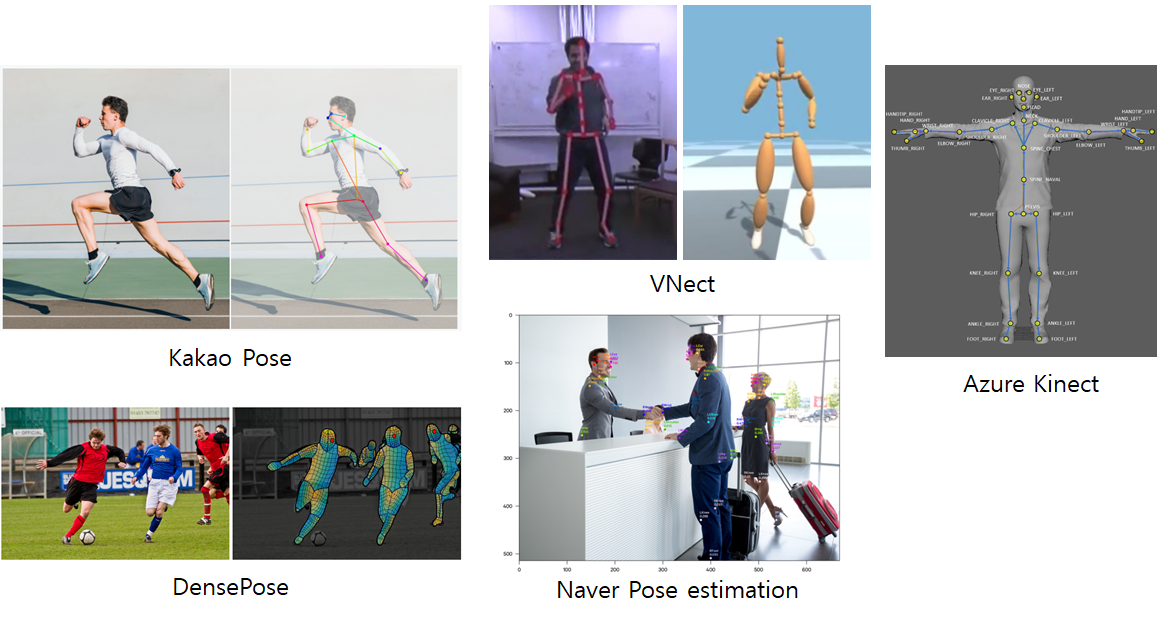
1. **Coverage**

This standard is applied to a system using human body key point. Human body key point format is composed of environmental information' and 'key point information'.

1. Human body key point data format
   1. Introduction



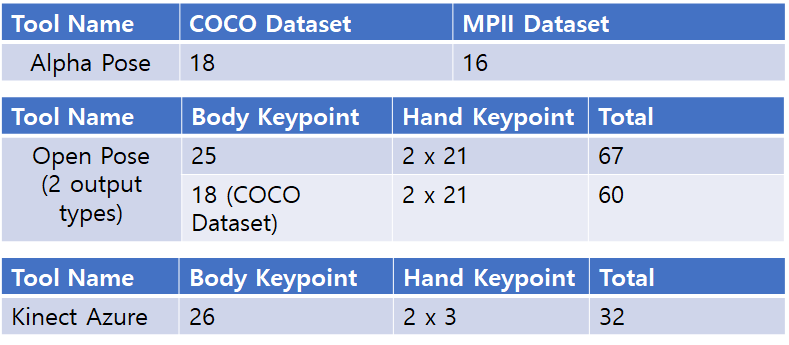
(Figure 2-1) Various systems using human body key point



(Figure 2-2) Human pose estimation

Human body key point is used in various systems that recognize human motion, such as sign language recognition, home training, and abnormal behavior detection, as shown in (Figure 2-1). Human body key point generally extracts human body key point information through various methods through RGB camera or depth camera sensor, as shown in (Figure 2-2). However, each of these pose estimation methods differ in the image size, coordinate system, and number of body key points. (Table 2-1) shows the number of body key points according to each human pose estimation method.

(Table 2-1) Number of body key points in human pose estimation method



If a standard for the format of human body key point data extracted through various human pose estimation techniques is established, human body key point data required in mixed reality applications can be utilized in various application software regardless of the extraction tool.

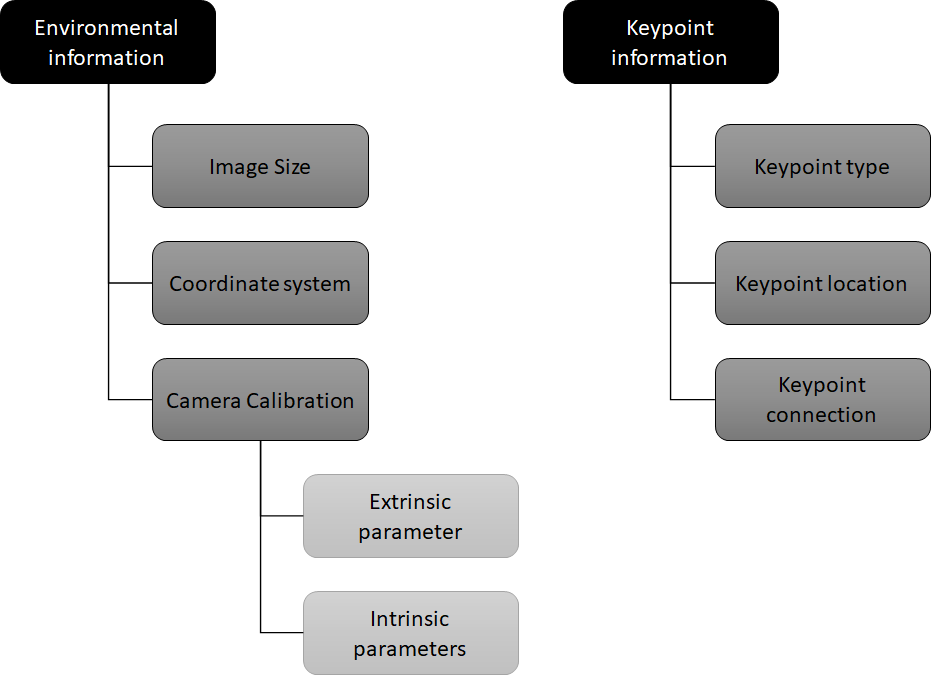
This proposal intends to establish a standard for the representation of human body key point information extracted through various posture estimation techniques for systems using human body key points.

* 1. Structure of Human body key point data format

The human body key point data format consists of an 'environmental information' and a 'key point information' as shown in (Figure 2-3).

The 'environmental information' consists of image size, coordinate system (world coordinate system, camera coordinate system, pixel coordinate system, normalized coordinate system) and camera calibration information such as extrinsic parameter and intrinsic parameters as follow <Table 2-1>.

The 'key point information' consists of the location of the user's key points, connection information between key points, and the type of key points as follow <Table 2-2>. Among them, the location of the key point may be composed of all or part of the lists in <Table 2-3>. The position of each keypoint consists of integer values of x, y, and z.



(Figure 2-3) Human body key point data format structure

* + 1. Environmental information

<Table 2-1> Struct of environmental information

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Definition | Unit | Value Type |
| Image Size | Describe the width and height of the image size used for human pose estimation. | Pixel | Integer |
| Coordinate system | Describe the four coordinate systems.  Select one of the world coordinate system, camera coordinate system, normal coordinate system, and pixel coordinate system. | - | String |
| Camera Calibration | Describes the transformation relationship between 3D spatial coordinates and 2D image coordinates, or a parameter describing the transformation relationship | - | - |
| Extrinsic parameter | Describes the transformation relationship between the camera coordinate system and the world coordinate system. | - | - |
| Rotation | Describes the rotation between the camera coordinate system and the world coordinate system. |  | Matrix (float) |
| Translation | Describes the origin of the world coordinate system expressed in the coordinates of the camera-centered coordinate system. |  | Matrix (float) |
| Intrinsic parameters | Parameter for obtaining the projected position on the image or restoring the 3D spatial coordinate from the image coordinate | - | - |
| Focal length (fx, fy) | Describe the distance between the center of the lens and the image sensor (CCD, CMOS, etc.). | Pixel | Float |
| Principal point (cx, cy) | Describe the position of the center(pinhole) of the camera lens. | Pixel | Float |
| Skew coefficient | Describe the distortion coefficient between the x-axis and y-axis of the image sensor. | Degree | Float |

* + 1. Key point information

<Table 2-2> Struct of key point information

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Definition | Unit | Value Type |
| Key point type | Describe a list of names and numbers of available key points. |  | Array (String) |
| Key point location | Describe the location of key points.  For 2D key points, X, Y are described, and for 3D key points, X, Y, Z are described. | pixel | Array (Integer) |
| Key point connection | Describes connection information between key points. |  | Array (Integer) |

<Table 2-3> List of key point location

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Definition | Unit | Value Type |
| CervicalVertebra1 | 1st cervical vertebra that supports the skull | pixel | Integer |
| CervicalVertebra2 | 2nd cervical vertebra (neck) | pixel | Integer |
| CervicalVertebra3 | 3rd cervical vertebra (neck) | pixel | Integer |
| CervicalVertebra4 | 4th cervical vertebra (neck) | pixel | Integer |
| CervicalVertebra5 | 5th cervical vertebra (neck) | pixel | Integer |
| CervicalVertebra6 | 6th cervical vertebra (neck) | pixel | Integer |
| CervicalVertebra7 | 7th cervical vertebra, lowest cervical vertebra | pixel | Integer |
| LeftEye | Left eye (center of pupil) | pixel | Integer |
| RightEye | Right eye (center of pupil) | pixel | Integer |
| LeftEyeBrow | the rightmost tip of the left eyebrow | pixel | Integer |
| RightEyeBrow | the leftmost tip of the right eyebrow | pixel | Integer |
| LeftEar | left ear canal | pixel | Integer |
| RightEar | right ear canal | pixel | Integer |
| Nose | the highest part of the nose | pixel | Integer |
| LeftClavicle | The part of the left clavicle that connects | pixel | Integer |
| LeftShoulder | The part where the left shoulder blade connects to the arm bone | pixel | Integer |
| LeftElbow | upper and lower joints of the left arm | pixel | Integer |
| LeftWrist | The joint where the left arm and hand connect | pixel | Integer |
| LeftHand | center of left hand | pixel | Integer |
| RightClavicle | The part of the right clavicle that connects to the sternum | pixel | Integer |
| RightShoulder | The part where the right shoulder blade connects to the arm bone | pixel | Integer |
| RightElbow | The upper and lower joints of the right arm | pixel | Integer |
| RightWrist | The joint where the right arm and hand connect | pixel | Integer |
| RightHand | right palm center | pixel | Integer |
| ThoracicVertebra | spine, roughly in the middle of the stomach | pixel | Integer |
| ThoracicVertebra1 | 1st thoracic vertebrae, connecting to the cervical vertebrae | pixel | Integer |
| ThoracicVertebra2 | 2nd thoracic vertebrae (spine) | pixel | Integer |
| ThoracicVertebra3 | 3rd thoracic vertebrae (spine) | pixel | Integer |
| ThoracicVertebra4 | 4th thoracic vertebra (spine) | pixel | Integer |
| ThoracicVertebra5 | 5th thoracic vertebrae (spine) | pixel | Integer |
| ThoracicVertebra6 | 6th thoracic vertebrae (spine) | pixel | Integer |
| ThoracicVertebra7 | 7th thoracic vertebrae (spine) | pixel | Integer |
| ThoracicVertebra8 | 8th thoracic vertebrae (spine) | pixel | Integer |
| ThoracicVertebra9 | 9th thoracic vertebrae (spine) | pixel | Integer |
| ThoracicVertebra10 | 10th thoracic vertebra (spine) | pixel | Integer |
| ThoracicVertebra11 | 11th thoracic vertebra (spine) | pixel | Integer |
| ThoracicVertebra12 | 12th thoracic vertebrae (spine), connected to the lumbar vertebrae | pixel | Integer |
| LumbarBertebra | The lumbar spine, roughly in the area of ​​the navel | pixel | Integer |
| LumbarBertebra1 | 1st lumbar vertebrae, connected to the thoracic vertebrae | pixel | Integer |
| LumbarBertebra2 | 2nd lumbar vertebra | pixel | Integer |
| LumbarBertebra3 | 3rd lumbar vertebra | pixel | Integer |
| LumbarBertebra4 | 4th lumbar vertebra | pixel | Integer |
| LumbarBertebra5 | 5th lumbar vertebrae, connected to the lumbar vertebrae | pixel | Integer |
| Sacrum | sacrum, tailbone | pixel | Integer |
| Pelvis | part of the hip and tailbone that connects to the thigh | pixel | Integer |
| LeftPelvis | left middle part of the hipbone | pixel | Integer |
| RightPelvis | right middle part of the hipbone | pixel | Integer |
| LeftHip | The joint between the left thigh and hip bone | pixel | Integer |
| RightHip | The joint between the right thigh and hip | pixel | Integer |
| LeftKnee | left leg knee | pixel | Integer |
| LeftAnkle | left leg ankle | pixel | Integer |
| LeftHeel | heel of left foot | pixel | Integer |
| LeftFoot | center of left foot | pixel | Integer |
| RightKnee | right leg knee | pixel | Integer |
| RightAnkle | right leg ankle | pixel | Integer |
| RightHeel | heel of the right foot | pixel | Integer |
| RightFoot | center of right foot | pixel | Integer |

1. summary

Human body key point data format is used in various systems that recognize human motion, such as sign language recognition, home training, and abnormal behavior detection. Human body key point generally extracts human body key point information through various methods through RGB camera or depth camera sensor. However, each of these pose estimation methods differ in the image size, coordinate system, and number of body key points. This proposal intends to establish a standard for the representation of human body key point information extracted through various posture estimation techniques for systems using human body key points.