http://dx.doi.org/10.14400/JDC.2016.14.1.189

# Kinect Sensor를 이용한 실시간 3D 인체 전신 융합 모션 캡처

김성호
상지대학교 컴퓨터정보공학부

# Realtime 3D Human Full-Body Convergence Motion Capture using a Kinect Sensor

**Sung-Ho Kim**
**School of Computer, Information and Communication Engineering, Sangji Univ.**

**요 약** 최근 카메라, 캠코더 및 CCTV 등의 사용이 활발해지면서 영상 처리 기술의 수요가 급증하고 있다. 특히 키넥트 센서와 같은 깊이(Depth) 카메라를 사용한 3D 영상 기술에 대한 연구개발이 더욱더 활성화되고 있다. 키넥트 센서는 RGB, 골격(Skeleton) 및 깊이(Depth) 영상을 통해 인체의 3D 골격 구조를 실시간 프레임 단위로 획득할 수 있는 고성능 카메라이다. 본 논문에서는 키넥트 센서를 사용하여 인체의 3D 골격 구조를 모션 캡처하고 범용으로 사용되고 있는 모션 파일 포맷(*.trc 및 *.bvh)으로 선택하여 저장할 수 있는 시스템을 개발한다. 또한 본 시스템은 광학식 모션 캡처 파일 포맷(*.trc)을 자기식 모션 캡처 파일 포맷(*.bvh)으로 변환할 수 있도록 하는 기능을 가진다. 마지막으로 본 논문에서는 키넥트 센서를 사용하여 캡처한 모션 데이터가 제대로 캡처되어졌는지 모션 캡처 데이터 뷰어를 통하여 확인한다.

**주제어** : 키넥트 센서, 모션 캡처, 모션 파일 포맷, 3D 골격, 모션 포맷 변환기

**Abstract** Recently, there is increasing demand for image processing technology while activated the use of equipments such as camera, camcorder and CCTV. In particular, research and development related to 3D image technology using the depth camera such as Kinect sensor has been more activated. Kinect sensor is a high-performance camera that can acquire a 3D human skeleton structure via a RGB, skeleton and depth image in real-time frame-by-frame. In this paper, we develop a system. This system captures the motion of a 3D human skeleton structure using the Kinect sensor. And this system can be stored by selecting the motion file format as trc and bvh that is used for general purposes. The system also has a function that converts TRC motion captured format file into BVH format. Finally, this paper confirms visually through the motion capture data viewer that motion data captured using the Kinect sensor is captured correctly.

**Key Words :** Kinect sensor, Motion capture, Motion file format, 3D skeleton, Motion format converter

* 본 논문은 2014년도 상지대학교 교내연구비 지원에 의한 것임.

Received 23 November 2015, Revised 25 December 2015
Accepted 20 January 2016
Corresponding Author : Sung-Ho Kim
(The Society of Digital Policy)
Email: kimsh1204@sangji.ac.kr

ISSN: 1738-1916

© The Society of Digital Policy & Management. All rights reserved. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

## 1. 서론

모션 캡처 시스템은 광학식, 자기식 및 기계식 등 다양한 종류의 장비들로 구성[16]되어져 있으며 컴퓨터 게임, 애니메이션, 영화 등 캐릭터의 동작을 필요로 하는 거의 모든 분야에서 필수적으로 사용되고 있는 최첨단 3D 동작 데이터 획득 방법이다. 그러나 기존의 모션 캡처 시스템들은 대부분 고가여서 중소기업들이 직접 구매하여 사용하기에는 경제적인 부담이 매우 컸다. 물론 정부 및 공공기관 등에서 고가의 모션 캡처 장비를 구매하고 저가로 임대하여 사용할 수 있도록 하는 제도가 마련되어 있으며 많은 중소기업체들이 사용하고는 있다. 그러나 시간적, 경제적 문제들이 많아 여러 가지 불편을 초래하는 경우도 여전히 존재하고 있다. 그런 이유로 저가의 모션 캡처 장비를 개발하기 위한 연구들이 수 년 동안 진행되어져 온 것도 사실이다. 그리고 나름대로의 저가 장비들이 많이 출시되어졌으며 일부에서는 아직도 사용 중인 곳도 있다. 그러나 저가 장비들은 대부분 응용 및 복합적인 동작들을 위한 모션 캡처에서는 제대로 기능을 발휘하지 못하는 경우가 대부분이라서 가장 기본적인 동작으로 제한되어지는 단점을 가지고 있다.

본 논문은 마이크로소프트사의 키넥트 센서를 사용하여 매우 쉽게 3D 모션 캡처를 수행하고, 범용 3D 모션 데이터를 획득하는 것을 목적으로 한다. 키넥트 센서는 RGB 및 깊이(Depth) 이미지 인식 센서를 갖춘 카메라이며, 가장 자연스러운 상태에서 모션 캡처를 할 수 있는 NUI(Natural User Interface) 장비들 중의 하나[17]이다. 키넥트 센서는 기존의 모션 캡처 장비에 비하면 매우 저가이면서 그 기능이 매우 강력하여 교육, 의료, 게임 등 매우 다양한 분야에서 사용되고 있으며 매우 활발하게 응용 프로그램들이 연구 개발되고 있다. 그러므로 본 논문에서는 키넥트 센서를 사용하여 실시간으로 인체의 전신 모션 캡처를 시행하고, 3D 모션 캡처 데이터를 생성하여 기존의 모션 캡처 장비에서 캡처한 모션 데이터와 동일한 포맷을 가지는 3D 모션 캡처 데이터를 생성하기 위한 시스템을 개발하고자 한다. 그리고 키넥트 센서로 캡처한 모션 캡처 데이터를 모션 데이터 뷰어를 통하여 제대로 캡처가 되었는지 가시적으로 확인할 수 있도록 한다.

## 2. 관련 연구

본 논문의 주제와 관련된 국내외 연구동향을 살펴보면 키넥트 센서를 사용한 인체의 동작 인식[15], 동작 추적 등에 대한 주제가 대부분이다. 특히 모션 캡처와 관련해서는 대부분 기존의 3D 그래픽스 및 애니메이션 툴과 키넥트 센서를 연동하게 함으로서 단일 응용프로그램이 아니라 기존의 3D 그래픽스 및 애니메이션 툴에서만 실행시킬 수 있는 콘텐츠들로 구성되어져 있다. 여기서 문제는 기존의 3D 그래픽스 및 애니메이션 툴을 보유하고 있지 않은 사용자들에게는 키넥트 센서를 사용하기 위한 접근성이 매우 약한 문제점이 있다. 또한 기존의 3D 그래픽스 및 애니메이션 툴(Tool)을 보유하고 있다고 하여도 키넥트 센서를 연동하기 위한 플러그인 프로그램이 유료이기 때문에 꼭 사용해야하는 경우가 아니라면 플러그인 프로그램 구매를 꺼리게 된다. 물론 본 연구의 목적과 매우 밀접한 관계가 있는 연구결과물로서 키넥트 센서를 사용한 모션 캡처를 주제로 하는 연구 논문들[3,4,5,6,7,8,9,10]도 최근에 출판되었다. 한 연구 결과는 마커(Marker)를 사용하는 기존의 모션 캡처 방식과 달리 한 개에서 네 개까지의 키넥트 센서를 사용한 모션 캡처 기법을 소개[1]하고 있다. 또 다른 연구 결과에서는 하나의 키넥트 센서를 사용하여 캡처한 모션 데이터의 동작에 있어서의 정확도가 낮은 문제점을 해결하기 위하여 모션 데이터의 정확도를 높이기 위한 연구[2]가 진행되어졌다.

## 3. 모션 캡처 시스템 및 파일 포맷

### 3.1 모션 캡처 시스템

모션 캡처 시스템은 광학식, 자기식 및 기계식 등 여러 가지 다양한 기술력을 기반으로 연구 개발되어 지금까지 사용되어지고 있다. 광학식은 사람이나 사물에 마커(Marker)를 부착하고 적외선 카메라를 비추어 반사되는 마커의 위치를 추적함으로서 3차원 공간의 위치를 캡처하는 방식이다. 자기식은 전기적으로 자기장을 형성해주는 장치와 11개 내외의 센서로 구성되며, 센서의 회전과 이동 값을 실시간으로 캡처할 수 있다. 기계식은 인체에 부착된 회전 센서로부터 데이터를 받는 형태로서 각 관

절에 3개축(X축, Y축 및 Z축)에 대응하도록 센서가 부착되며 회전 값은 자이로(Gyro)를 사용하여 처리 및 인식된다. 그러므로 모션 캡처 시스템의 종류에 따라 캡처된 모션 데이터의 파일 포맷은 모두 다른 형식을 가지고 있다.

### 3.2 모션 캡처 파일 포맷

모션 캡처 시스템의 종류에 따라 모션 데이터 파일 포맷은 모두 달라지는데, 그 이유는 모션 캡처되는 원리 및 기술이 모두 다르기 때문이다. 지금까지 주로 많이 사용되는 모션 데이터 파일 포맷의 종류[11]로는 TRC, BVH, C3D 등이 있었다. 그러나 본 논문에서는 범용으로 사용되고 있는 TRC 및 BVH 파일 포맷에 한하여 제안하고 사용하기로 한다. TRC 파일 포맷은 Motion Analysis사에서 제작한 모션 캡처 파일 포맷이다. TRC 파일 포맷은 [12] 및 <Table 1>과 같이 얼굴의 각 근육 및 전신의 각 관절 부분에 부착한 마커에 대한 3D 공간 좌표를 획득하고 저장한다. 그래서 TRC 파일 포맷은 다른 파일 포맷과 달리 골격을 기반으로 하고 있지 않다.

〈Table 1〉 Structure of TRC file format

| | | | | |
|---|---|---|---|---|
| PathFileType 4 (X/Y/Z) MotionAnalysis | | | | |
| DataRate CameraRate NumFrames NumMarkers Units | | | | |
| OrigDataRate | | | | |
| OrigDataStartFrame OrigNumFrames | | | | |
| 30 60 55 24 mm 60 0.0 600 | | | | |
| Frame# Time Markers HeadTop LeftHead RightHead RightShoulder RightElbow RightWrist ..... | | | | |
| | X1 | Y1 | Z1 | X2 ... |
| 1 | 0 0.00000 | 0.00000 | 0.00000 | 18.1347 |
| | 649.971 | 689.404 | -14.6755 | 626.967 |
| ..... | | | | |

BVH(BioVision Hierarchy)는 BioVision사에 의해서 개발된 계층형 모션 데이터 파일 포맷으로서 애니메이션 동작을 표현하기 위해 사용되어진다. 그래서 최근 가장 인기 있는 범용 모션 데이터 파일 포맷 중의 하나이다. BVH 파일 포맷의 인체 골격 구조[11]를 잘 나타내고 있는 파일 포맷의 구조는 <Table 2>와 같다.

〈Table 2〉 Structure of BVH file format

```
HIERARCHY
ROOT Hips
{
  OFFSET   0.00          0.00          0.00
  CHANNELS 6 Xposition Yposition Zposition Zrotation
Xrotation Yrotation
  JOINT Chest
  {
      OFFSET              0.00          5.21          0.00
      CHANNELS 3 Zrotation Xrotation Yrotation
      JOINT Neck
      {
         OFFSET           0.00          18.65         0.00
         CHANNELS 3 Zrotation Xrotation Yrotation
         JOINT Head
         {
            OFFSET        0.00          5.45          0.00
            CHANNELS 3 Zrotation Xrotation Yrotation
            End Site
            {
               OFFSET     0.00          3.87          0.00
            }
         }
      }
.....
}
MOTION
Frames:    2
Frame Time: 0.33333
 8.03 35.01   88.36 -3.41 14.78 -164.35 13.09 40.30 -24.60 .....
```

### 3.3 모션 데이터의 대응하는 골격 및 관절 위치

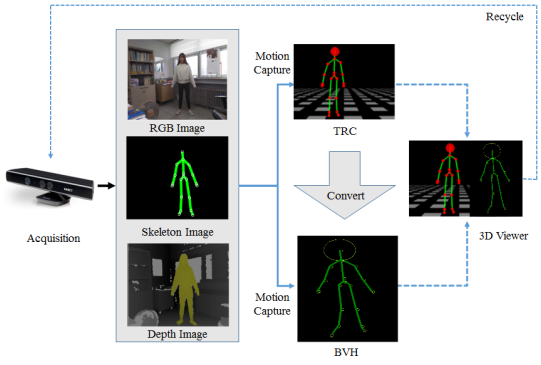
본 논문에서는 키넥트 센서를 사용하여 TRC 및 BVH 파일 포맷 등으로 사용자의 필요에 의해 선택하여 캡처하는 것이 가능하도록 한다. 이것은 기존의 서로 다른 종류의 고가 모션 캡처 장비를 사용하여 서로 다른 모션 캡처 파일 포맷의 모션 데이터를 획득하는 것보다 경제적, 시간적으로 훨씬 저렴하며 쉽고 빠르게 사용자가 필요로 하는 모션 캡처 데이터 파일 포맷을 생성해 낼 수 있다는 점에서 매우 유용하기 때문이다. 키넥트 센서에서 인식하는 인체의 각 부위별 관절의 위치는 [11]을 참조한다. <Table 3>은 본 논문에서 사용하기 위한 TRC 파일 포맷의 마커, BVH 파일 포맷의 센서 및 키넥트 센서의 관절 위치에 대한 명칭을 비교하여 나타낸 것이다.

〈Table 3〉 Skeleton and Joint position corresponding of the motion data

| | TRC | BVH | Kinect |
|---|---|---|---|
| 1 | Root | Hips | HipCenter |
| 2 | MidBack | Chest | Spine |
| 3 | Sternum | | ShoulderCenter |
| 4 | NeckRear | Neck | |
| 5 | HeadTop | Head | Head |
| 6 | LeftHead | | |
| 7 | RightHead | | |
| 8 | LeftShoulder | LeftCollar | LeftShoulder |
| 9 | LeftElbow | LeftUpArm | LeftElbow |
| 10 | LeftWrist | LeftLowArm | LeftWrist |
| 11 | LeftHand | LeftHand | LeftHand |
| 12 | RightShoulder | RightCollar | RightShoulder |
| 13 | RightElbow | RightUpArm | RightElbow |
| 14 | RightWrist | RightLowArm | RightWrist |
| 15 | RightHand | RightHand | RightHand |
| 16 | LeftHip | LeftUpLeg | LeftHip |
| 17 | LeftKnee | LeftLowLeg | LeftKnee |
| 18 | LeftAnkle | | LeftAnkle |
| 19 | LeftFoot | LeftFoot | LeftFoot |
| 20 | RightHip | RightUpLeg | RightHip |
| 21 | RightKnee | RightLowLeg | RightKnee |
| 22 | RightAnkle | | RightAnkle |
| 23 | RightFoot | RightFoot | RightFoot |

## 4. 시스템 설계 및 사용자 인터페이스

[Fig. 1]은 본 논문에서 제안하는 시스템의 흐름도이다. 즉, 키넥트 센서를 사용한 모션 캡처를 위해서 키넥트 센서의 원리[13]를 이해하고, 키넥트 센서로부터 RGB, 골격(Skeletal) 및 깊이(Depth) 이미지를 획득한다.



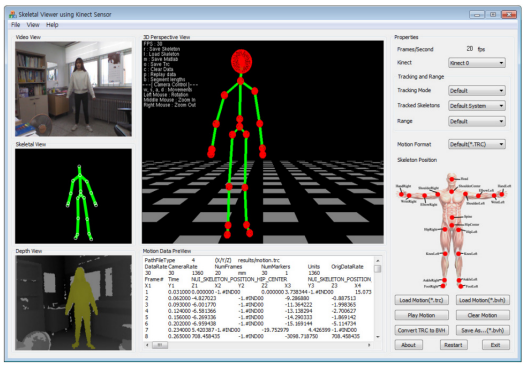

[Fig. 1] Flowchart of system

그리고 골격(Skeletal) 및 깊이(Depth) 이미지로부터 BVH 및 TRC 파일 포맷의 모션 데이터를 사용자의 선택에 의하여 획득하도록 설계하였다. 또한 캡처된 TRC 파일 포맷을 BVH 파일 포맷으로 변환(Converter)이 가능하게 함으로서 모션 데이터 사용의 편리성을 향상시켰다. 본 시스템은 3D 뷰어에서 TRC 및 BVH 파일 포맷을 각각 읽어 재생시켜 봄으로서 모션 캡처 데이터가 제대로 캡처되어졌는지를 가시적으로 확인할 수 있도록 하였다. 그러므로 본 시스템의 사용자 인터페이스는 RGB, 골격(Skeletal), 깊이(Depth), 원근 뷰(Perspective View)와 실시간 모션 데이터 뷰어, 3D 뷰어 및 속성(Properties) 등으로 구성되어진다. 여기서 원근 뷰는 키넥트 센서로 캡처되어지는 모션 데이터를 3D 공간에서 실시간으로 보여주며, 실시간 모션 데이터 뷰어는 캡처되어지는 모션 데이터를 출력시켜준다. 속성에는 모션 캡처 파일 포맷(*.trc, *.bvh) 선택, 재생, 저장 등 여러 가지 기능들로 구성되어진다.

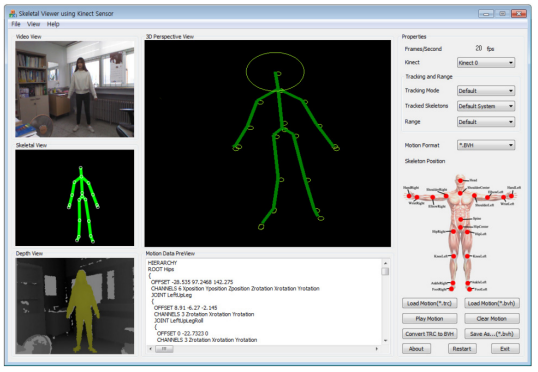
본 시스템은 키넥트 센서, 4GB RAM, i3-3220 3.30GHz CPU 등과 같은 H/W 및 Windows 7(x86), Kinect SDK v1.8, Kinect Runtime v1.8, Kinect Developer Toolkit v1.8, OpenNI v2.2, NiTE v2.2, OpenGL API, Visual Studio 2013 C++ 등과 같은 S/W 환경에서 구현하였다.

## 5. 모션 캡처 및 결과 확인

[Fig. 2]는 본 논문에서 키넥트 센서를 사용하여 TRC 및 BVH 파일 포맷으로 모션 캡처를 실행하기 위해 개발한 시스템의 실행과정이다. (a)는 광학식 모션 캡처 파일(*.trc) 포맷으로 캡처하는 과정이고, (b)는 자기식 모션 캡처 파일(*.bvh) 포맷으로 캡처하는 과정이다. 본 시스템에서는 실시간 모션 캡처를 수행한 후 모션 캡처 파일로 저장하고, 시스템에 내장된 모션 데이터 뷰어를 통하여 모션 캡처가 제대로 수행되었는지 재확인할 수 있도록 하였다. 모션 데이터 뷰어는 OpenGL로 구현한 3D 환경이기 때문에 카메라의 이동, 회전 및 확대/축소 등의 기능들을 기본적으로 사용하여 다양하게 입체적으로 확인할 수 있다.



[a] For TRC file format



[b] For BVH file format

[Fig. 2] Motion Capture Scene using Kinect sensor

## 6. 결론 및 향후 연구

일반적으로 3D 모션 캡처 데이터는 3D 캐릭터 애니메이션이 적용되는 PC 및 스마트폰 게임, 애니메이션, 광고 및 영화 등 매우 다양한 분야에서 광범위하게 사용되어지고 있는데, 본 논문에서 키넥트 센서를 사용하여 캡처한 모션 데이터도 교육, 연구 및 각종 산업 분야에서 충분히 사용되어질 수 있다. 그 이유는 모션 캡처 데이터의 파일 포맷이 3D 캐릭터 애니메이션을 위해서 범용으로 사용되어지고 있기 때문이다. 그러므로 본 논문에서는 키넥트 센서를 사용하여 TRC 및 BVH 파일 포맷 등으로 실시간 모션 캡처하고 그 결과를 모션 캡처 파일로 저장할 수 있는 시스템을 개발하였다. 또한 저장된 TRC 파일을 BVH로 변경할 수 있는 기능과 함께 모션 데이터 3D 뷰어를 통하여 재확인할 수 있도록 하였다. 기존 고가의 모션 캡처 장비들을 사용하여 모션 캡처를 하게 되면 서로 다른 파일 포맷을 위해 경제적, 시간적으로 매우 불편하고 어려운 과정을 거쳐야만 가능했다. 그러나 본 연구의 결과물은 키넥트 센서와 같은 매우 저가의 장비를 사용하여 매우 쉽고 빠르게 각 분야에서 필요로 하는 모션 캡처 데이터를 획득할 수 있다는 점에서 매우 유용하다고 판단된다. 그러나 키넥트 센서 1개를 사용하여 캡처한 모션 데이터는 제대로 된 3D 모션 데이터라고 보기에는 아직 무리가 있다. 그러므로 향후 연구에서는 3개 이상의 키넥트 센서를 정면, 측면, 후면 등과 같이 서로 다른 방향에 설치하고 하나의 데이터로 보정 및 후처리함으로서 간단한 동작뿐만 아니라 복잡한 응용동작까지 제대로 된 3D 모션 데이터를 획득할 수 있도록 해야 할 것이다.

## ACKNOWLEDGMENTS

This research was supported by Sangji University Research Fund, 2014

## REFERENCES

[1] K. Berger, K. Ruhl, C. Brümmer, Y. Schröder, A. Scholz, M. Magnor, “Markerless motion capture using multiple color-depth sensors,” In Proc. Vision, Modeling and Visualization (VMV), 2011.

[2] Xiaolin Wei, Peizhao Zhang, Jinxiang Chai, “Accurate realtime full-body motion capture using a single depth camera,” ACM Transactions on Graphics (TOG) - Proceedings of ACM SIGGRAPH Asia 2012.

[3] A. Hadjakos. “Pianist Motion Capture with the Kinect Depth Camera,” In Proceedings of the International Conference on Sound and Music Computing, Copenhagen, Denmark, 2012.

[4] Jeff K. T. Tang, Jacky C. P. Chan, Howard Leung, “Interactive dancing game with real-time recognition of continuous dance moves from 3D human motion capture,” Proceedings of the 5th International Conference on Ubiquitous Information

Management and Communication Article No. 50, 2011.

[5] Zheng Xiao, “3D Human Postures Recognition using Kinect,” 2012 4th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC), pp. 26-27, PP.344-347, 2012.

[6] M. bin Mohd Sidik, M. bin Sunar, I. bin Ismail, M. bin Mokhtar, and N. Jusoh, "A study on natural interaction for human body motion using depth image data," in Proc. IEEE International Workshop on Digital Media and Digital Content Management (DMDCM, 2011), pp. 97-102, 2011.

[7] Andrea Sanna1, Fabrizio Lamberti1, Gianluca Paravati1, Felipe Domingues Rocha1, “A kinect-based interface to animate virtual characters,” Journal on Multimodal User Interfaces, © OpenInterface Association 2012, 10.1007/s12193-012-0113-9, 2012.

[8] Myung-hwan Choi, Byoung-keon Park, Bon-yeol Koo, Je-wook Chae, Jay-jung Kim, “Human motion analysis based on a markerless motion capture using Kinect,” The 8th Conference on National Defense Technology, pp. 619-623, 2012.

[9] Mankyu Sung, “3D Character Motion Synthesis and Control Method for Navigating Virtual Environment Using Depth Sensor,” Journal of Korea Multimedia Society Vol.15 No.6, pp. 827-836, 2012.

[10] Hyun-Cheol Yoon, Jong-Seung Park, “Gesture-based Game Character Animation Using Motion Data Captured From the Kinect Sensor,“ Korean Society for Computer Game, Vol.26 No.2, pp. 147-153, 2013.

[11] List of motion and gesture file formats, http://en.wikipedia.org/wiki/List_of_motion_and_gesture_file_formats

[12]http://blogs.msdn.com/b/eternalcoding/archive/2011/06/13/unleash-the-power-of-kinect-for-windows-sdk.aspx

[12] http://cta-diderot.brucity.be/?page_id=11

[13]http://www.laserfocusworld.com/articles/2011/01/lasers-bring-gesture-recognition-to-the-home.html

[13] http://tech.integrate.biz/kinect_mocap.htm

[14]http://channel9.msdn.com/coding4fun/kinect/Saving-Kinect-Streams-Skeleton-Stream-to-BVH-Converter-and-Stream-Saver-Application

[15] Han Suk Choi, “Kinect-based Motion Recognition Model for the 3D Contents Control,” Journal of the Korea Contents Association, Vol.14, No.1, pp.24-29, 2014.

[16] Yeong-Hyeon Byeon, Myung-Won Lee, and Keun-Chang Kwak, “A Trend Analysis of Motion Capture Systems for Sports Motion Analysis,” Journal of KIIT. Vol.11, No.5, pp.191-201, 2013

[17] Zhang, Zhengyou. "Microsoft kinect sensor and its effect." MultiMedia, IEEE 19.2 (2012): 4-10.

**김 성 호(Kim, Sung Ho)**

- 1998년 8월 : 숭실대학교 컴퓨터학과(공학석사)
- 2005년 2월 : 숭실대학교 컴퓨터학과(공학박사)
- 2006년 3월 ~ 현재 : 상지대학교 컴퓨터정보공학부 교수
- 관심분야 : 컴퓨터 그래픽스, 컴퓨터 애니메이션, 모션 캡처, 가상현실/증강현실, 컴퓨터게임, Web3D, 멀티미디어, 지리정보시스템
- E-Mail : kimsh1204@sangji.ac.kr