

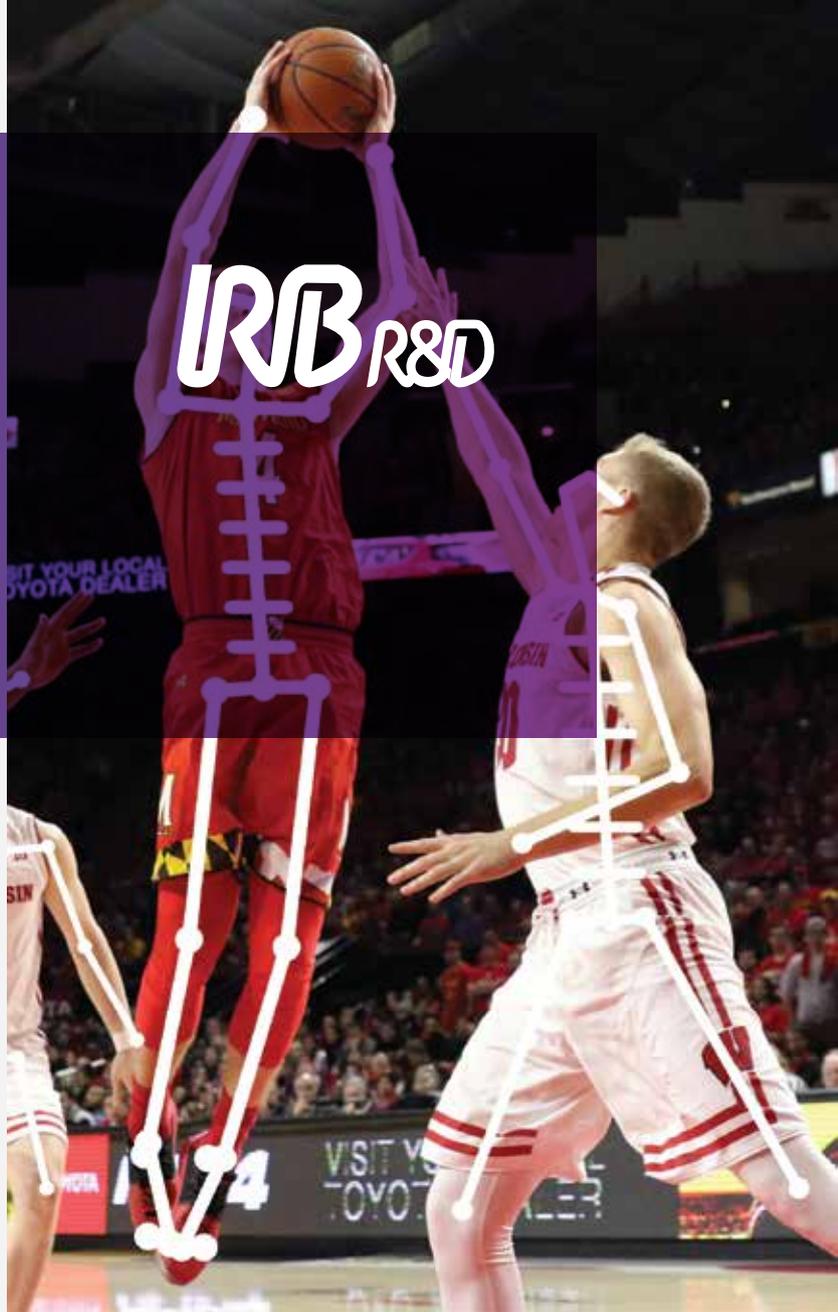
Deep MOCAP

Human Motion Capture Using Deep Convolutional Neural Networks

Human Motion Capture

Motion capture is the process of recording human movements. It has many uses in medical, military, sports, computer vision, and cinema industry.

Current motion capture systems fall into two main categories: optical and non-optical systems. The main disadvantage of current motion capture systems is the need for special hardware and corresponding software. The cost of such systems could be prohibitive for small productions. In addition, the system may need some requirements for the space it operates in, for example Chroma key (green screen) or magnetic distortion limitations.



IRIB R&D implemented a method for human motion capture based on deep convolutional neural networks (CNN). The method consists of two main phases:
1) k2D human pose estimation.
2) k2D data aggregation to capture 3D data.





Islamic Republic of Iran Broadcasting (IRIB) Research and Development

IRIB R&D has made major advances in applying basic research and innovation in media technology. It develops innovative new products and technologies in broadcast/broadband industry that leads the development of future technologies in IRIB. Our R&D projects often involve collaborations with public or private entities, including universities, government laboratories, technology start-ups and incubators, research institutes and partner companies.

✉ rd@irib.ir

🌐 www.rd.irib.ir/en

☎ +98-21-22164060

📞 +98-21-22164181

🌐 www.linkedin.com/company/iribrd

📍 IRIB R&D Center, IRIB, Valiasr Ave.,
Tehran, Iran,

P.O: 193953895

More Info

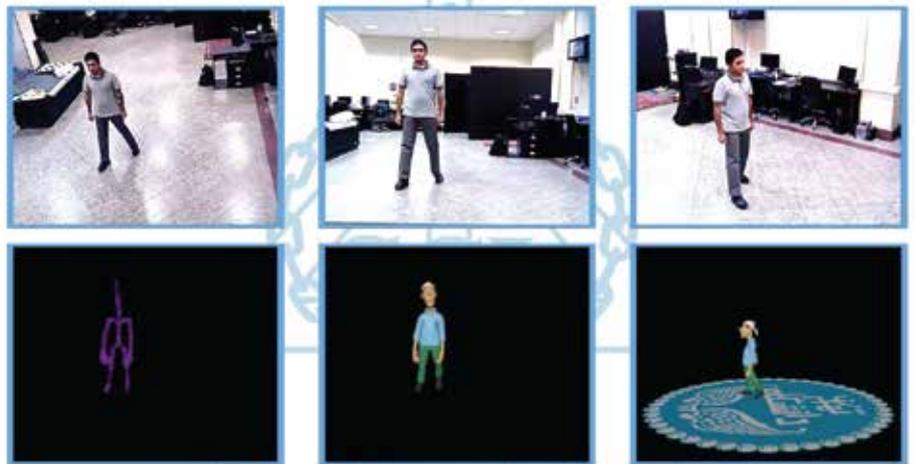


Website: rd.irib.ir/en

Advantages

First, the captured image by each camera (cameras are assumed to be calibrated) is processed by the designed CNN. The CNN is trained to output human joints in the 2D input image. The input to the CNN is a raw image in RGB format and the output is the location of detected human joints in the image.

The system detects all humans in the image and is independent of the number of people who are present in the input image. It is also capable of capturing fingers, if the input image has enough resolution. Then, the 2D positions in each image are aggregated to construct a single 3D map of human structure and motion. To do so, we need intrinsic and extrinsic matrices of all cameras. These matrices are formed using calibration methods. Finally, the 3D points are mapped to a virtual character to follow the motions.



Secure Communication Lab

Alongside human movement, the system could also capture human facial points. The procedure is the same as the body joints. Facial points could be used in both 2D and 3D formats, because of the symmetric features and limitations of face.

The main advantage of the proposed method is that it does not need any special hardware, since our method works with low cost ordinary cameras and no sensors. The load of the system has been transferred from equipment and hardware (which is the case in traditional systems) to the calculations take place in computer. This is because of the fact that the main part of a motion capture system, which is the detection of special positions on subject, is done through complex algorithm using our designed CNN. We need no addition hardware. However, because of the CNN model in the core of the MOCAP system, a GPU is recommended.

