DRINK INTAKE MONITORING USING A DEPTH CAMERA

Gasparrini S., Cippitelli E., Spinsante S., Gambi E. - Università Politecnica delle Marche
{s.gasparrini, e.cippitelli, s.spinsante, e.gambi}@univpm.it

ICVSS 2014
Sicily ~ 13-19 July
International Computer Vision Summer School

Abstract

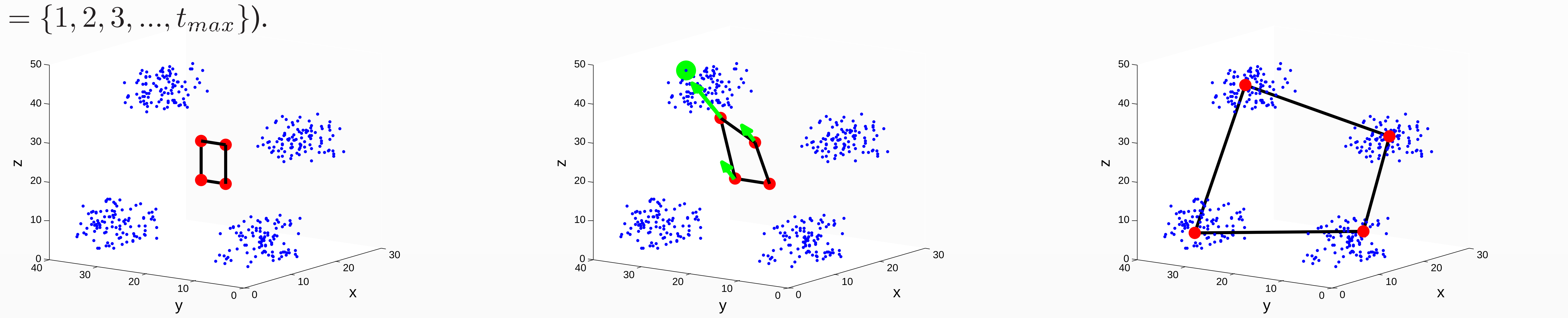
Health of elderly people is directly influenced by food and drink intake behaviors. This monitoring is usually done by the patient himself or by the healthcare staff but unfortunately their judgment is subjective and this aspect represents a limitation to the assessment. The information provided by the Microsoft Kinect can be exploited to automatize the process. This sensor provides RGB and Depth (D) streams to realize a data fusion algorithm that recognizes the patient gestures typically performed when he is drinking. In this work, the device is placed in ceil configuration and a skeleton, characterized by eleven joints, is automatically fitted to the person shape. At each frame, the system monitors the distances between head and hands to identify a drink intake action.

Self Organizing Maps(SOM) algorithm applied to Point Cloud

A preprocessing step, constituted by background subtraction and filtering, is applied to the raw D frame to obtain a point cloud (PC) of D data that represents the human shape. At this point the SOM algorithm is exploited to place the joints of the skeleton in the PC. The SOM is an unsupervised neural network (clustering) [1], already used in similar contexts [2]. The left figure shows the first step of the algorithm, when an initial model is placed in the PC. At each step a specific element of the PC is evaluated and it modifies the position of the nearest joint(Euclidean distance) and its neighbors. When all the elements are being taken into account, the result is a model that covers all the PC. The equations to displace nearest joint are:

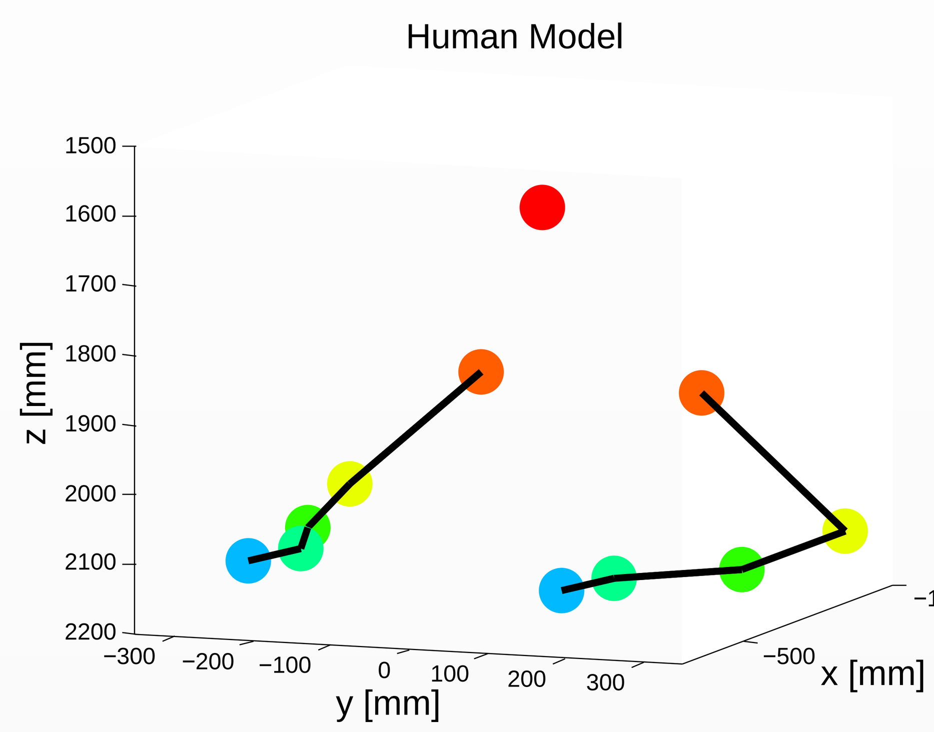
$$\widehat{v}^{t+1} = \widehat{v}^t + \widehat{\epsilon}^t (x - \widehat{v}^t) \qquad \widehat{\epsilon}^t = \epsilon_i \left(\frac{\epsilon_f}{\epsilon_i}\right)^{\frac{t}{t_{max}}} \qquad \widetilde{v}^{t+1} = \widetilde{v}^t + \widetilde{\epsilon}^t (x - \widetilde{v}^t) \qquad \widetilde{\epsilon}^t = 0.5\widehat{\epsilon}^t$$

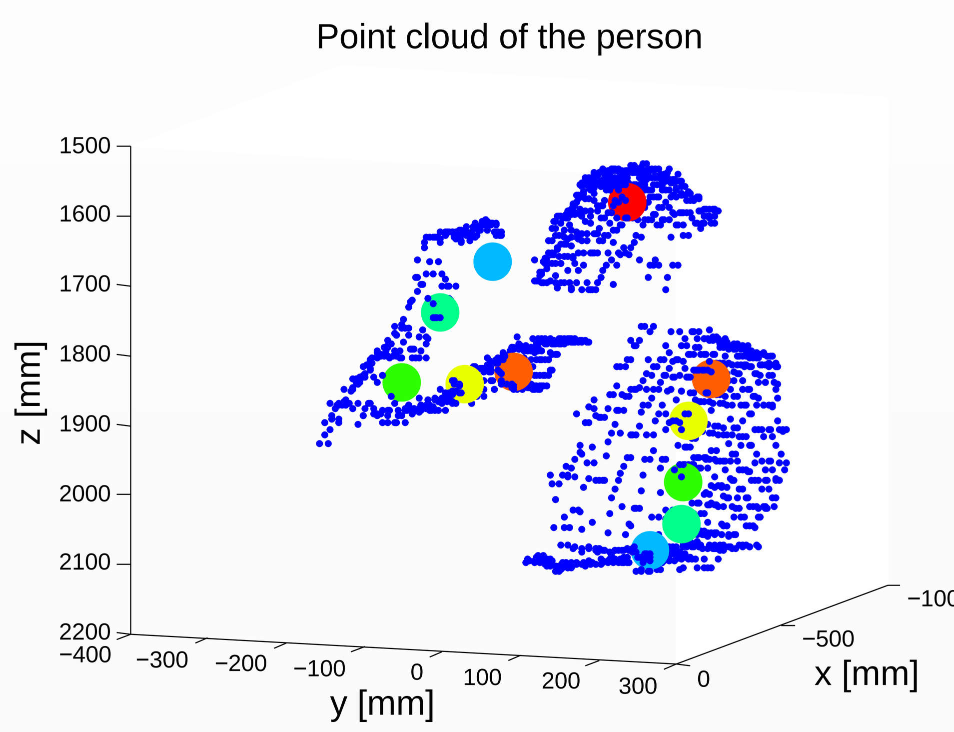
where x is an element of PC, $\widehat{v}$ is the nearest joint to x and $\widetilde{v}$ are the neighbor joints of $\widehat{v}$. ϵ_f and ϵ_i are the coefficients that must be tuned to allow the convergence of the process, t_{max} is the number of repetition of the algorithm through all the PC($t = \{1, 2, 3, ..., t_{max}\}$).

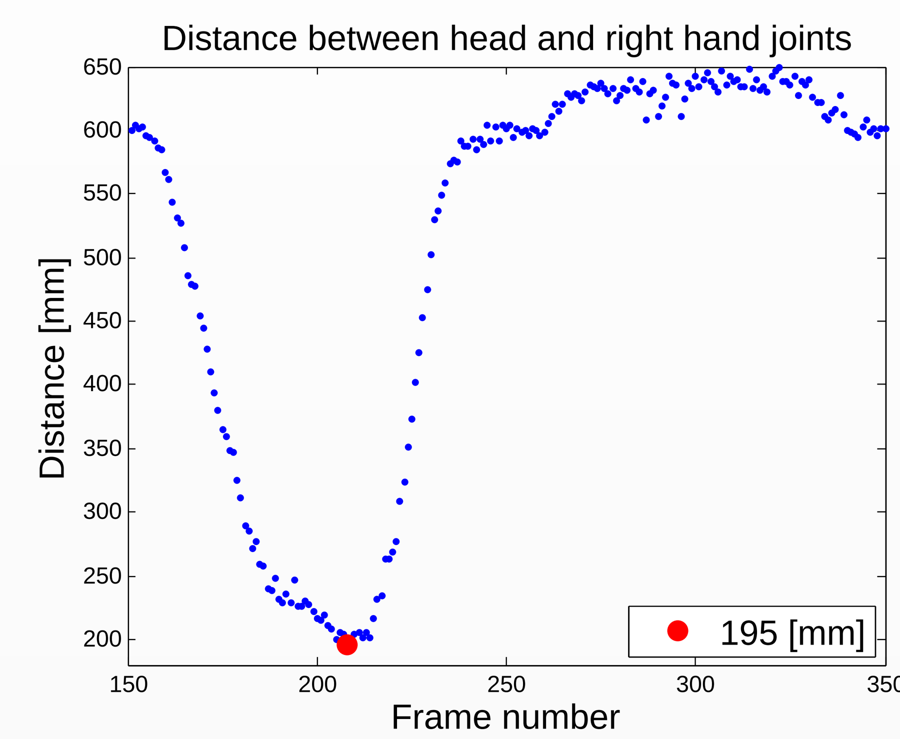


Preliminary results

The left image represents the initial model applied to the PC, all joints are connected with their neighbors except the head joint. The algorithm, after the exploiting of the SOM in the PC, fits the model to estimate the position of the hands joints. When the person is grasping the glass, the system monitors the distance between hand and head joints. The right figure depicts these distance for each frame. Approximately at 210^{th} frame a minimum is reached, and the associated PC and skeleton model is showed in the middle figure.







References

[1] T., Kohonen, The self-organizing map, in *Proceedings of the IEEE*, pp. 1464-1480, Sep 1990

[2] M., Martin, B., Martinetz, E., Barth, Haker, Self-Organizing Maps for Pose Estimation with a Time-of-Flight Camera, in *Springer Berlin Heidelberg*, pp. 142-153, Sep 2009