Markerless Motion Capture Integrated with Human Modeling for Virtual Ergonomics

Giorgio Colombo[1], Daniele Regazzoni[2], and Caterina Rizzi[2]

[1] Department of Mechanical Engineering, Polytechnic of Milan, Milan, Italy
giorgio.colombo@polimi.it
[2] Department of Engineering, University of Bergamo (BG), Dalmine, Italy
{daniele.regazzoni,caterina.rizzi}@unibg.it

Abstract. This paper refers to the context of virtual ergonomics and specifically addresses a case study of the commercial refrigeration industry. The aim is to develop a computer-aided platform to analyse end-users' postures and movements and ergonomically validate the design of device a man or woman may deal with. This paper describes the integrated use of human modeling and motion capture (Mocap) systems to perform ergonomic analysis relying exactly on real movements. Two optical Mocap systems, both low cost and markerless, have been considered: one based on six Sony Eye webcams and another one on two Microsoft Kinect sensors. Analogously, two human modeling tools have been adopted: Jack, specifically targeted for ergonomics and integrated with Microsoft Kinect, and LifeMod, a biomechanical simulation package. The proposed virtual ergonomics solutions have been experimented considering the case study of vertical refrigerator display units.

Keywords: Virtual ergonomics, Mocap, Digital human modeling, commercial refrigeration.

1 Introduction

Human factors are involved in several steps of product life (design, manufacture, maintenance, use, etc.) and keeping them into account since the early stages of the development process can be a key issue for a successful product. Several aspects, such as handling capability, physical condition and risk prevention, have to be considered as well as that human beings with different needs will interact with the product. This requires the development of products centered on human beings and suitable to the widest range of population characterized by different sizes, genders, ages, preferences and abilities [1] and the adoption of strategies and tools that permit to consider ergonomics aspects since the conceptual design stage.

In such a context, virtual ergonomics is the discipline that permits engineers to create and manipulate virtual humans to investigate the interactions between the users and the product. For example human factors, such as positioning, visibility, reaching, grasping and lifting of weights can all be evaluated by using virtual humans, providing a feedback to designers without the need for physical prototyping.

V.G. Duffy (Ed.): DHM/HCII 2013, Part II, LNCS 8026, pp. 314–323, 2013.

This paper provides an insight on this topic with particular attention to commercial refrigeration industry specialized in display units. It describes the use of virtual ergonomics techniques in the design process of display units used in supermarkets.

In this work we consider the integrated use of human modeling [1, 2] and motion capture (Mocap) [3] systems to perform ergonomic analysis relying exactly on real movements executed by operators in everyday activities and provide designer with guidelines to improve workers' environment and users' interaction. Actually, such an approach should avoid any potential approximation or mistake due to standardization of movements. The aim of the research is to develop a computer-aided platform to analyze workers' postures and movements and ergonomically validate the design of device a man or woman may deal with.

Regarding Mocap systems, we decided to use low cost techniques, developed for video games and entertainment, to verify their usability and performance in industrial context. Such technologies benefit from a huge investment on research that leads to a rapid evolution, but on the same time they keep affordable prices because of the target market they refer to.

The paper, after a description of the case study, presents the technical solutions adopted. The application to the case study and preliminary results of the experimentation are finally presented and discussed.

2 The Case Study and Users' Categories

As mentioned, we consider the design process of display units. The case study concerns the family of vertical refrigerator display units with or without doors, which are installed in groceries or supermarkets. Such machinery should accommodate the full range of users during its life cycle. At least we can distinguish three main groups of users that interact with the refrigerated display unit: customers picking up goods, workers in charge of checking exposed products and filling out shelves with new ones, and maintenance technicians who need to access to some specific components (Fig. 1). For each category some ergonomic aspects are more relevant than others: i.e., visibility and reachability of goods for customers, reachability of some display components for technicians and, the most important, posture and stress for operators who repeat the same task for hours and may occur in musculoskeletal disorders.

We mainly focus the attention on costumers and workers with the main goal of determining goods reachability for the first ones and suitability of working conditions for the latter. In fact, the task of loading the shelves may generate health disorders due to repeated actions, holding and lifting loads, uncomfortable postures, among the others. A preliminary study, based on interviews and direct observations, has been conducted to determine how the operators really behave and main operations to be reproduced during the tests in the lab to verify the feasibility of our approach. We also identified some occasional postures that permit to improve reachability but unacceptable because of dangerous postures or hygienic reasons (e.g., stepping on the first shelf).

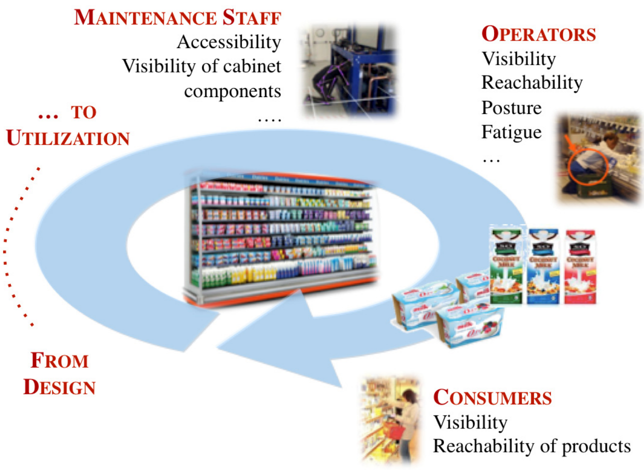


Fig. 1. Vertical display unit and main users' groups

3 Virtual Ergonomics Solutions

Our main goal has been to experiment the integration of low-cost Mocap systems with human modeling tools for virtual ergonomics of refrigerated display unit.

Figure 2 shows the virtual ergonomics solutions considered.

Regarding Mocap system, several solutions are available for different contexts such as military, entertainment, sports, and medical applications. According to the working principle four main categories can be identified [3]: optical, mechanical, inertial, and magnetic that can be used for testing and validation also in a combined way [4]. We considered two systems, both optical and markerless: the former based on Sony Eye webcams and the latter on Microsoft Kinect [5]. Initially, the Microsoft Kinect has been developed for game environment but it is attracting more and more attention from different contexts [6,7]. It is a markerless system bases on structured light that can track body motion with a depth sensor.

We have been using them to acquire the real movements and postures of operators so that the following simulation can be based on real data. They are not expensive and can be easily moved and, with some precautions, used also outside the lab in potentially any work environment we want to acquire.

Both systems foresee the adoption of iPi Motion Capture™ software [8], a non-real time markerless system developed to work with Sony Eyes webcams and recently with Microsoft Kinect. Its main features are:

- Possibility to use from 3 to 6 webcams and 1 or 2 Kinect sensors.
- A maximum acquisition area of 7 m x 7 m.
- Non real-time tracking.
- Input file format: MPEG.

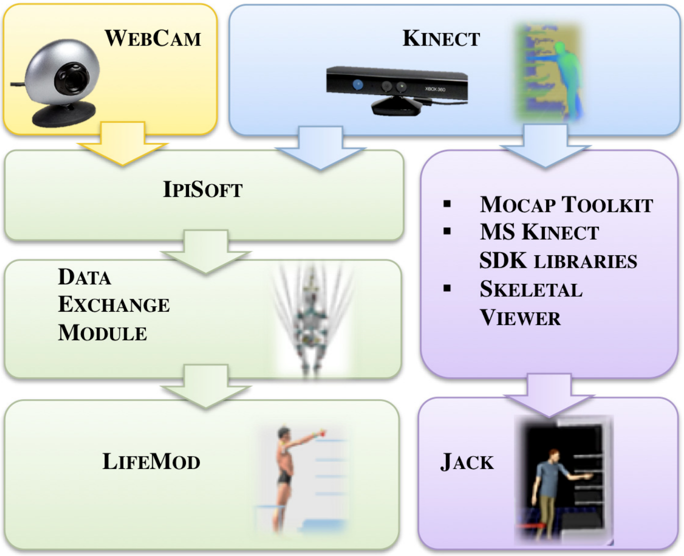


Fig. 2. Virtual ergonomics solutions

Concerning human modeling, we can find various tools of different complexity depending on the target application. They can be grouped in [1,9]: virtual human/actors for entertainment, mannequins for clothing [10], virtual manikin for ergonomic analysis [11, 12], and, finally, detailed biomechanical models [13].

We decided to use Siemens Jack [14, 15] and LifeMod [16]. Jack, a well-known human modeling system, has been chosen since it specifically targeted for ergonomics and is integrated with Microsoft Kinect. It permits to define complex scenes with virtual manikins and objects, simulate many tasks and evaluate posture and ergonomics factors also using analysis tools such as RULA (Rapid Upper Limb Analysis) to investigate work-related upper limb disorders, NIOSH (U.S. National Institute for Occupational Safety and Health) lifting equations to evaluate lifting and carrying tasks, and OWAS (Owako Working Posture Analysis System) to analyze postures during work. The second system, LifeMod, is a biomechanical simulation software that permits to generate a complete biomechanical model of the human body. It is a plug-in of ADAMS software, a multi-body analysis system. The creation of a model normally starts with the generation of a basic set of connected human segments based on the dimension contained in an anthropometric database; then, the joints, the muscles and the tendons are created and contact force with objects are defined.

3.1 Webcam-Based Solution

This solution comprises:
Hardware

- Six/four Sony Eye webcams with a resolution of 640x480 pixels at 60 Hz mounted on photographic tripods with semicircle disposition.
- Portable workstation Dell Precision M6500.

Software

- iPisoft Recorder (calibration and capture).
- iPisoft Desktop Motion Capture (tracking).
- Data Exchange module.
- LifeMod.

The system acquires synchronized video sequences obtained with the camera without having to apply physical markers on the operator's skin. It automatically recognizes the different body segments and, then for each time step, calculates joints position and orientation. Precisely, iPisoft Recorder synchronizes images recorded from the webcams while iPisoft Desktop Motion Capture recognizes and applies the segmentation of body and tracking of movement. iPisoft output contains the recorded movement in BVH (Biovision Hierarchical Data) format. iPisoft adopts a skeleton made of 27 joints hierarchically organized, each characterized by proper d.o.f. and constraints.

To reproduce the movements within the human modeling system, it has been necessary to develop a module for data exchange. Two are the reasons: firstly LifeMod uses another data format (SLF) and secondly acquired data are related to the real position of the human being's joints while the biomechanical software uses external markers placed on the skin surface. Therefore, we developed an ad-hoc algorithm in Matlab, which translates the information relative to the joint hierarchy and to the motion contained in the BVH file to a SLF one and a CMD script that relocates the markers accordingly to LifeMod representation.

Once the model is defined, simulation phase can begin. To obtain accurate simulations with the muscles and the articulations it's necessary to execute a first inverse dynamic simulation to drive the body with motion agents describing the movements to execute. Once that the movements are stored a direct dynamic simulation is run to calculate the forces created by the muscles and the stresses the body is subjected to. The outcome provided by the system consists of forces and momenta acting on each joint in each time step of the analysis.

3.2 Kinect-Based Solution

Kinect has been integrated both with LifeMod and Jack. Therefore, this solution comprises:

Hardware

- Two Microsoft Kinect sensors with a resolution of 640x480 pixels at 30 fps mounted on photographic tripod and connected via USB cable.
- Portable workstation Dell XPS.

Software

- iPi soft Recorder (calibration and capture).
- iPisoft Desktop Motion Capture (tracking).
- LifeMod and the Data Exchange module.

or

- Siemens Jack with Mocap Toolkit, a module specifically developed by Siemens for Kinect sensor, MS Kinect SDK v 1.0 libraries and SkeletalViewer to transfer data streaming acquired with Kinect sensor to Jack.

If we use LifeMod, as in the previous case iPi Recorder manages the recording of images and depth videos coming from Kinect, while the iPi Studio performs environment calibration and video analysis.

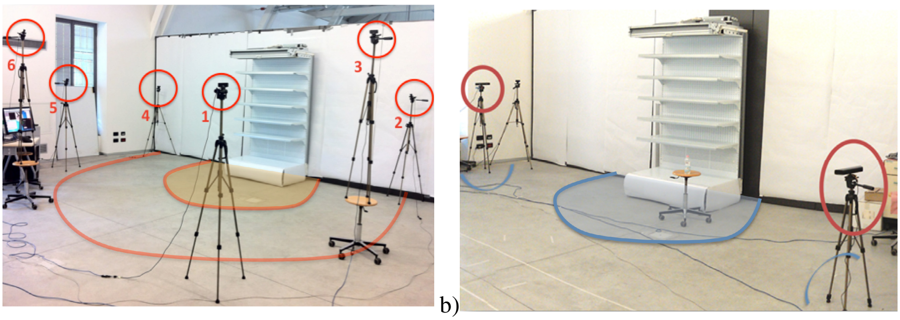
While using Jack, iPiSoft is not necessary. First, using Jack the virtual scene composed by the 3D model of the refrigerated unit and the operator avatar is created; then, the Kinect plug-in and the SkeletalViewer software are launched and the acquisition session can start. The skeleton used in the transition from Mocap to Jack is made of 21 joints whose positions and movements are tracked and there is not a hierarchy among them. This skeleton is less complex than Jack's one and some details cannot be taken into account (i.e., head rotation, fingers).

4 Experimentation

The solutions have been experimented with two vertical units: one with six shelves and doors for frozen goods and another one with 5 shelves and without doors for fresh goods. We used a simplified version of the real display units because some elements of the complete units (e.g., lateral walls) may interfere with the operator during motion capture.

We involved testers of both genders and different heights to evaluate if the motion capture system is affected by any problem. The pick and place operation of a bottle on each shelf has been considered and tasks performed by the testers have been acquired simultaneously with proposed solutions.

First we performed the calibration of the webcam-based solution that initializes the system and permits to correctly locate each camera in space. A semicircle disposition of the webcams at different heights around the operator is the best choice. Instead, the Kinect sensors were placed laterally to the actor. Figure 3 shows the lab setup/layout for the webcam and Kinect solution.



a) b)

Fig. 3. Acquisition scenarios: a) Webcam-based; b) Kinect-based

Then, we validated the data exchange module as well as different ways of automatically changing the position of joints depending on height and structure of subjects. Once, preliminarily tested the technical solutions we started the real campaign.

Ten actors were asked to perform as much as possible as if they were in the real environment and to follow a precise routine to produce comparable results with operators characterized by different anthropometric measures. The routine defines the initial and final position of each movement to be performed.

Figures 4, 5 and 6 portray some examples of acquisition related to loading task for the vertical units. Figure 4 shows an example of acquisition carried out for the vertical unit with doors. One can see the representation of data related to the loading of the fourth shelf during the three main steps with the webcam-based solution. The first one refers to the environment where the webcam images are captured and elaborated for each time step to gather joints positions shown in the second image. The third representation comes from LifeMod where data have been converted, corrected and integrated with anthropometric databases.

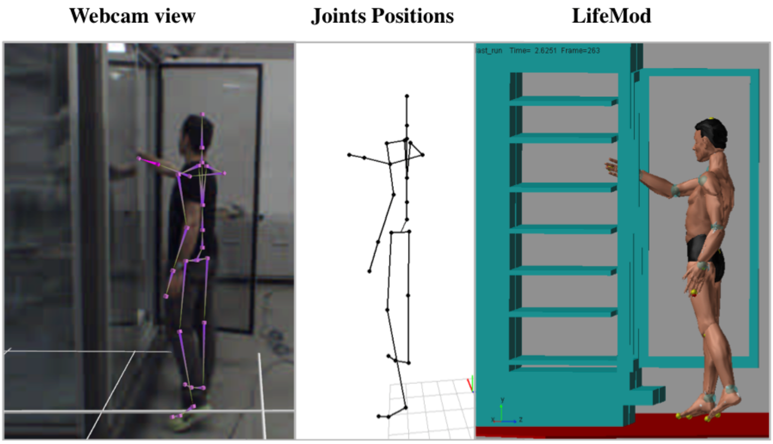


Fig. 4. Vertical unit with doors, Webcam+LifeMod: Loading the fourth shelf

Figure 5 shows the loading task of the highest shelf and of the middle one of the vertical unit without door using the Kinect-based solution and LifeMod. The images are similar to those of Figure 4; only the first one differs being the depth map of the Kinect sensor. Finally, Figure 6 shows the acquisition steps related to the loading of the second shelf using Jack. Precisely, Figure 6a shows the image acquired, Figure 6b the corresponding depth map, Figure 6c the corresponding reconstructed skeleton and Figure 6d the virtual avatar of the tester.

The Mocap systems have been also tested with more complex machines and tasks, such as maintenance of a compression machine feeding the refrigerated display units. Figure 7 shows two screen shots with the worker in two different postures taken from of the tracking sequence obtained by using the webcams as acquisition means. The image on the left shows the iPisoft avatar overlapped to the silhouette of the real worker, while the image on the right shows the skeleton obtained calculating the position of the key joints.

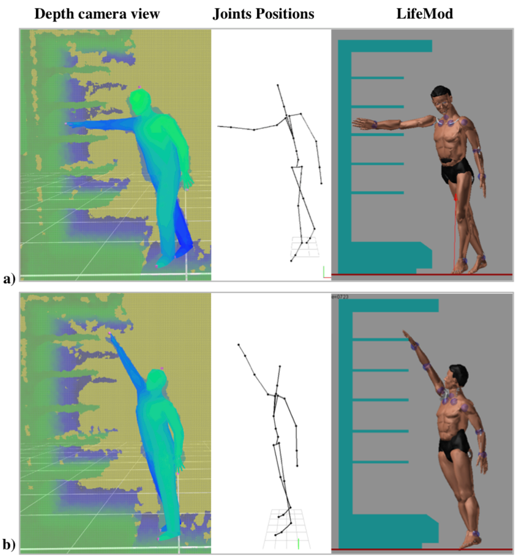


Fig. 5. Vertical unit without doors, Kinect+LifeMod: a) Loading highest shelf; b) loading the middle shelf

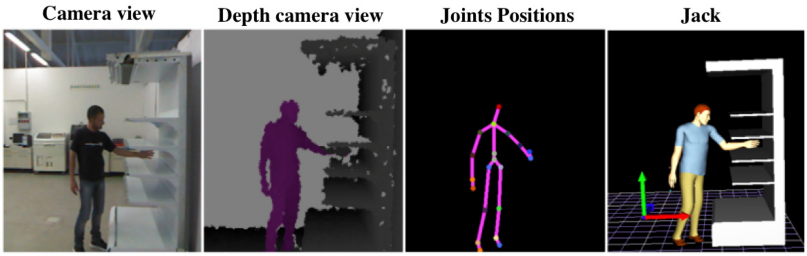


Fig. 6. Vertical unit without doors, Kinect+Jack: a) Loading the second shelf

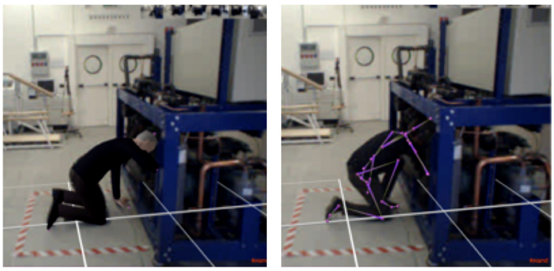


Fig. 7. Frames of Mocap process with worker in two different postures

Results have been considered satisfying for both markerless optical systems and performances are adequate for work related analysis and interesting for research activity. Complex scenes benefit from redundant points of view (webcam solution), while simple environments allow easier depth camera acquisition (Kinect solution).

Webcam-based solution requires a calibration routine and is sensitive to environment light and contrast with actor. Besides, Kinect-based solution does not require complex routine to calibrate sensors and, thus, the overall setup is much shorter and it is less sensible to light conditions. This contributes to a better portability in almost any working environment where the scene is small enough to be seen by both sensors. In fact, only two (iPisoft) or one (Jack) sensors are supported and the area of acquisition is smaller.

Regarding the human modeling system, the main advantages of Jack are the direct integration with Kinect sensor and real time determination of ergonomics parameters; however it supports only one Kinect and there are some limitations affecting its performance when used with Jack. The skeleton does not always perfectly match the subject posture, this is particularly true when body's area overlap but improvements are under development and a new version will be available soon. LifeMod permits to create more detailed human model and an easy extrapolation of data for joint and segments but on the other hand there is the need for data exchange module and marker relocation and ergonomics parameters are not automatically calculated. Known problems about Kinect sensors mutual interference are significant as long as we adopted only two sensors. Anyway further investigations will take into account also scenarios with more sensors and, thus, solutions do depth camera interference.

5 Conclusions

This work aimed at exploiting low-cost motion capture solutions (optical-markerless) outside the entertainment domain together with a detailed biomechanical DHM solution and verifying their applicability in industrial domain.

The results reached so far are promising and valuable for several industrial applications since the accuracy of the Mocap system is good enough to assess ergonomics of potentially any workplace. In case of a product being directly used by customers, as the display unit we tested, the ability to simulate the interaction of real people brings important and unpredictable feedbacks to designers. The high portability of the Mocap systems allows using it also outside the lab so that real workers' motion can be captured and further developments comprehend an acquisition campaign in a supermarket. Actually, starting from real operators performing in the real workplace allows not only to be more precise but also to consider unknown postures or movements, either correct or wrong, performed by operators in their everyday activities.

References

1. Sundin, A., Ortengren, R.: Digital Human modelling for CAE applications. In: Salvendy, G. (ed.) Handbook of Human Factors and Ergonomics, 3rd edn., John (2006)
2. Duffy, V.G. (ed.): HCII 2007 and DHM 2007. LNCS, vol. 4561. Springer, Heidelberg (2007)
3. Furniss, M.: Motion capture. MIT Communications Forum, http://web.mit.edu/commforum/papers/furniss.html (accessed March 2013)
4. Schepers, H.M.: Ambulatory Assessment of Human Body Kinematics and Kinetics. PhD Thesis, Universiteit Twente, The Netherlands (2009), http://www.xsens.com/images/stories/PDF/ThesisSchepers.pdf
5. Microsft Kinect, http://www.xbox.com/en-US/kinect (accessed February 2013)
6. Stowers, J., Hay, M., Bainbridge-Smith, A.: Altitude control of a quadrotor helicopter using depth map from Microsoft Kinect sensor. In: Proceedings of the IEEE International Conference on Mechatronics, Istanbul, Turkey, April 13-15, pp. 358–362 (2011)
7. Raptis, M., Kirovski, D., Hoppe, H.: Real-time classification of dance gesture from skeleton animation. In: Proceedings of 10th Annual ACM SIGGRAPH/Eurographics Symposium on Computer Animation, Vancouver, BC, Canada, August 5-7, pp. 147–156 (2011)
8. http://www.ipisoft.com (accessed March 2013)
9. Rizzi, C.: Digital Human Models within product development process. In: Innovation in Product Design, pp. 143–166. Springer (2011)
10. Volino, P., Luible, C., Magnenat-Thalmann, N.: Virtual clothing. John Wiley and Son Ltd. (2008)
11. Green, R.F., Hudson, J.A.: A Method for Positioning Digital Human Models in Airplane Passenger Seats. In: Advances in Applied Digital Human Modeling. CRC Press (2011)
12. Colombo, G., Regazzoni, D., Rizzi, C.: Ergonomic design through virtual Humans. Computer-Aided Design and Applications 10(5), 745-755 (2013)
13. Colombo, G., Facoetti, G., Rizzi, C.: Virtual testing laboratory for lower limb prosthesis design. Computer Aided Design and Applications 10(4), 671–683 (2013)
14. Badler, N., Phillips, C.B., Webber, B.L.: Simulating Humans: computer graphics animation and control. Oxford University Press, Oxford (1993)
15. http://www.plm.automation.siemens.com/en_us/products/tecnomatix/assembly_planning/jack/ (accessed February 2013)
16. http://www.lifemodeler.com (accessed March 2013)