A practical computer based vision system for posture and movement sensing in occupational medicine

Gustavo Marfia and Marco Roccetti

University of Bologna, Italy

Abstract

Back pain and upper extremities injuries due to overexertion account for over twenty percent of leave days from work in the US. This explains why a vast amount of initiatives have been, to this date, carried out aiming at reducing the occurrence of such type of injuries. However, although such type of lesions are among the most studied in occupational medicine, no automatic detection and prevention technologies are pervasively available, to this date, at workplaces. Such deficiency is ascribable to the absence of any flexible and cost-effective tectaphnology that may play such role. This work aims at filling such gap: the contribution of this paper is the design and implementation of a movement-posture computer-vision based system that, performing as a sensor, can detect overexertion movements, helping avoid the most common injuries that these cause. Such tasks are carried out with the use of a simple webcam, thus not requiring any expensive or specialized (e.g., Microsoft Kinect) hardware device. The proposed technology is, hence, easily affordable by any type of company and production plant throughout the world and easy adaptable to recognize and detect a wide set of movements and postures. The validity of such approach is demonstrated in realistic settings through a wide set of experiments.

Keywords: *Multimedia, computer system, computer vision, occupational injury, injury risk detection technology, well being, kinect.*

---

1. Introduction

The December 2014 news release from the US Bureau of Labor Statistics reports that in 2013 strains, sprains and tears accounted for one-third of all the days-away-from-work cases [1]. Among these, injuries that involved the back, shoulder and upper extremities amounted to over 50% (the back representing the most commonly injured body part) with the leading event being overexertion [2]. Overexertion (i.e., an excessive expenditure of energy by skeletal muscles) is

the third leading cause of injuries in the US, accounting for approximately 3.3 million emergency room visits in 2008. A variety of studies demonstrate that overexertion could be avoided: (a) keeping a correct posture, (b) limiting the amount of carried weight, and, (c) carefully dosing the duration of repetitive movements while working, as these are a well known cause of muscular tension [3], [4], [5].

Although musculoskeletal and, more in general, all those injuries that may occur at workplaces during working activities have been studied for long (occupational medicine sees its birth in 1671 when the book *Disease of workers*, by Bernardino Ramazzini, is first published), their prevention has been mainly addressed following two approaches. The first approach amounts to imagine in advance, during the design phases of production plants, all the movements and actions that may be required to complete a given task. The advantage of such approach is that it helps: (a) individuating harmful postures or gestures, in order to (b) modify those working procedures that may exhibit any type of risk. The disadvantage, instead, depends on the fact that even very well designed production processes could lead trained personnel to perform sequences of dangerous movements (not always the most comfortable or rapid way of doing something is also the safest one). In fact, physical fatigue and unexpected or out of control events may always cause wide varieties of dangerous situations during production. The second approach aims at understanding the magnitude of tissue stress that could be tolerated while performing a particular manual task. This type of studies generally relies on the construction of complex biomechanical models, also involving the use of controlled laboratory settings.

Summarizing, the two main trends that have been followed, to this date, in occupational medicine research may be simply summarized as the study of injuries, ex post, in order to ban ex ante all those movements that could be dangerous for workers. These approaches, however, assume that laborers are always in control and follow at all times an approved *movement code*, in order to avoid any harmful actions.

Recently, a number of human body tracking systems have seen birth. A review of such works reveals a widespread use of marker-based motion capture and electromyographic devices, together with sensors (e.g., gyroscopes, magnetometers, accelerometers, pedometers, goniometers, pressure and inertial sensors) adopted in advanced rehabilitation programs [2], [3], [4], [5]. Compared to such type of proposals, which rely on expensive wearable sensors, this work embraces a different perspective: verify the feasibility of tracking important body features, i.e., features that may reveal injury risks, only resorting to webcams, avoiding any on body sensors (which may be expensive, but also disturbing for a person busy at work). In essence, this work presents the design, implementation and deployment of a movement-posture tracking application that, based on vision technologies, could decide whether a given person were at risk of injuries (or not) in real-time. This work does not, instead, present a novel activity recognition scheme, the activities that are performed at specific locations are assumed to be known, as it normally happens in assembly lines. A first advantage of such approach is that the use of webcams and surveillance cameras could dramatically increase the penetration ratio of injury prevention technologies, as they are cheap and in many cases they may be found, deployed for security reasons, in many public and private facilities. Under such perspective, simple webcams

could be turned into wellbeing sensors capable of detecting when harmful movements are performed: all this could be possible with the integration of webcams and video cameras with posture and movement recognition software algorithms.

A second advantage is simply given by the fact that the use of webcams leads to the development of a low cost system, which is capable of improving the safety conditions of also those production plants, which operate on a low budget. In fact, while a cost difference of a few hundred dollars, for monitoring technology, in a developed country may be irrelevant, as the labor and worker replacement cost may justify such expense (e.g., $100 sum up to the cost of one or two hours of specialized labor, which may not be easy to replace), that same cost may not be justifiable in a low budget plant, where tasks are typically simpler and workers are more easy to replace. Hence, the use of simple webcams could really amount to a turning point fostering the spread of injury prevention applications in a great variety of environments and countries.

In order to assess the feasibility of such approach, we focused on two particular situations, two specific tasks that have been recognized as being at high risk of overexertion. The first amounts to hammering an object, which typically involves a movement pattern where a worker, a craftsman for example, holds a tool (i.e., a hammer) and swings rapidly, up and down, his/her arm. Depending on the weight of the utensil and on the speed at which the movement is performed, both short-term and long-term musculoskeletal damage could occur. The second task, instead, amounts to lifting an object from the ground. Posture here plays a key role, as such action, when performed in an incorrect way, could lead to back injuries. We have chosen these two postures/movements because: (a) both are risky, according to occupational injuries

assessment procedures, as explained in Section 3.1, and, (b) both threaten health, although in different ways, as the main risk of a hammering movement is its repetitiveness (it can gradually wear out a worker's arm and shoulder joints), whereas, even if not performed repeatedly, a single lifting movement, performed in an erroneous manner, could lead to lower-back pain.

This said, to the best of the authors' knowledge, this is the first work that discusses the opportunity (and the problems) of assessing physical stress with a technology that does not include the use of any complex sensors, but a simple webcam. An irrefutable advantage of such approach is that low cost off-the-shelf webcams can be introduced in manufacturing plants with very limited budgets. In addition, this work indicates a path that may be extended beyond the workplace scenario, resulting hence generally beneficial for the improvement of the wellbeing of a person.

The remainder of this paper is organized as follows. In Section 2 we review the most relevant related work that has been produced in the addressed field. We then move on to describe the rationale and the methodologies underlying our approach, while in Section 4 we validate our approach through a series of experimental scenarios, providing a further discussion in Section 5. We finally conclude with Section 6.

2. Related Work

A wealth of research has been carried out so far in the area of occupational medicine with a particular emphasis on musculoskeletal disorders [2]-[27]. Whereas all of such work has certainly aimed at improving the well being of employees and workers, the majority has been carried out evaluating the amount of stress that given movements induce in body tissues.

Along such lines, the authors of [5] studied how repetitive, constrained, varied and mobile work could damage the muscular tissues of the neck and of the upper limbs. Such analysis was carried out over a heterogeneous set of workplaces, including offices, industrial plants and other types of settings, comparing injury risks between females and males. This scientific contribution analyzed workers building upon: (a) their answers to a detailed set of questions, and, (b) their direct physical examination. The authors were able to conclude that in all settings (i.e., office, industrial) repetitive and constrained tasks showed elevated risks when compared with varied and mobile ones, without finding any risk elevation dissimilarities between sexes. Such stream of research has the merit of providing important insights on which are the prevalent causes of injuries at work. A missing step this contribution leaves open to future work amounts to devising practical real-time injury prevention systems.

Although a limited amount of such work has been carried out so far, a few examples exist where ergonomic risks are assessed not ex post or ex ante, but in real time [20]-[35]. In [20], for example, the authors present a system that permits a real-time ergonomic assessment of manual tasks in an industrial environment. Such system relies on an on-body sensor network composed of seven wireless Colibri inertial measurement units, where each lightweight sensor (weighting approximately 50 g) is equipped with a tri-axial accelerometer, a tri-axial gyroscope and a tri-axial magneto-inductive magnetic sensor. Such type of technology is also being marketed on the consumer market, as a mobile app, the Lumo App, which monitors the posture of a given person in real-time [34], [35]. In essence, when a user maintains an incorrect posture, the sensor network triggers the mobile phone to alert such situation. Although body sensor network

approaches appear very interesting and promising for the future, their costs are still considerable (e.g., the Lumo app price is set to approximately $150), as they require important investments. Many companies and factories are instead equipped with webcams and surveillance cameras, low cost devices (e.g., web cams may cost as low as $10), which can be conveniently turned into injury prevention systems. For this reason, in this paper, we exploit such opportunity as a low cost alternative that may reach many different working environments in very different geographic settings.

3. Challenges and Methods

Building upon the observation that the use of on body sensors should be minimized in practical systems, as they may be poorly tolerated, we focused on understanding the challenges posed by those postures and movements that more frequently cause overexertion injuries for vision-based sensing and detection procedures. We, hence, proceed: (a) identifying the postures and movements that pose greatest risks of overexertion, (b) selecting the vision technologies that may provide the best performance at the lowest possible cost, and, (c) individuating the algorithms that can efficiently alert the injury risks of interest.

3.1. Risky movements and postures

The identification of the movements and postures that present highest risks is at the center of occupational health and ergonomics studies. During the years, many different methods and tools have been created for the risk assessment of manual tasks. In this work, we resort to the Rapid

Upper Limb Assessment (RULA) index to identify high-risk movements and postures, as this represents one of the de facto standards in occupational health literature [11].

In particular, the RULA index amounts to a global risk score that ranges from one to seven (higher values correspond to higher risks). To compute such score, a user answers two groups of questions posed on the RULA's Employee Assessment Sheet [36]: (a) questions concerning the use of the arm and the wrist, and, (b) questions concerning the use of the neck, trunk and leg. A user's score increases as his/her answers reveal incorrect postures, increased muscle use, weight of loads, task duration, and repetitiveness. The movement patterns that, however, most frequently emerge as high risk ones are those involving extensive joint movements, i.e., extensive rotations and torsions of the arms, wrists, neck, trunk and legs of a person. When an extensive movement is performed, higher muscular tensions are activated, as well as higher injury risks (e.g., bending the body trunk by a 60 degrees angle increases the global score by four points in RULA). To promptly detect the occurrence of such situations, a real-time tracking of the positions of given body parts is key. In fact, performing such operation it is possible to assess the extension and the speed of the movements that are performed while completing a given task.

We here concentrate on two specific movements, both at a high risk of developing overexertion injuries, according to RULA scores. Although both present overexertion risks, they do in different ways, on different time scales. The first, repetitively hammering an object, creates long-term muscle tension, which could lead to injuries when sufficient recovery time between strikes were not available. The second, lifting a weight from the ground, is often carried out in incorrect ways, as for example in all those cases where awkward postures are mistakenly adopted (i.e.,

without bending the knees), placing an excessive amount of stress on a specific part (i.e., the back) of a person's body, i.e., even one single lifting movement could be fatal for the lower back.

3.2. The technological problem

We previously discussed why sophisticated and expensive technologies might not represent the best options in workplace environments (e.g., offices, industrial production plants etc.). These sites exhibit also another opportunity: workplaces are usually equipped with camera-based surveillance mechanisms, deployed for purposes that range from security and safety to production control ones. Video camera based systems hence appear interesting technological enablers when designing injury prevention technologies targeted for workplace scenarios, as they could be into systems capable of tracking body parts. In addition, the use of such type of technology would greatly limit deployment costs.

Nonetheless, using computer vision techniques could not be as straightforward as expected, as working locations, as public ones, can be insidious operating locations [37], [38]. Vision algorithms, in fact, are sensitive to a number of factors, including the distance and the angulation of a camera's point of view with respect to the position of a scene of interest. Illumination interferences, not rare in certain working environments (e.g., the sparks created by a blacksmith at work), could also contribute to the degradation of tracking operations.

Working locations, however, could also offer favorable conditions. In industrial plants, for example, workers often wear reflective suits to emphasize their presence for safety reasons. Such suits not only signal the presence of a worker, but also place distinguishing tags on given body

parts. Such tags clearly represent a valuable piece of information that could be put to good use to improve tracking operations.

Summarizing, working locations are peculiar sites that deserve the creation of systems tailored to follow the movements and postures of busy workers in specific environments. Each of the previously individuated challenges (e.g., webcam location, illumination levels) could be approached increasing the number of the hardware components (e.g., multiple video cameras, depth sensing cameras) and the complexity of the algorithms that process video feeds. For the purpose of this work, we avoid addressing the problem of finding the optimal number and placement of webcams as a function of workplace specific factors (e.g., working position, illumination, suit, etc.). We will focus, instead, on developing vision algorithms capable of detecting overexertion risks during the two postures/movements of interest.

3.3. Algorithms

We here concentrate on describing the algorithms we developed to track the two postures/movements of interest, treating each in isolation. For the hammering movement, injury risks rise as a worker repeatedly performs the same movement again and again. For this reason we concentrated on developing an algorithm that could follow the body parts of interest and provide all the information required to compute risk values in real-time. For the lifting movement, instead, we focused on devising a scheme capable of differentiating between correct and incorrect ways of lifting objects from the ground. In this case we will show that it is possible to concentrate on two specific body parts to detect when the movement is being performed and distinguish a wrong action from a right one.

3.3.1. Hammering an object

The hammering movement, according to the RULA score sheet, exhibits increasing injury risks as the: (a) the weight of the utensil increases (e.g., hammer), (b) the speed of the swinging movement is higher, and, (c) the amplitude of the angular movement widens.

A feasible way of identifying the weight of a given utensil in a video stream simply amounts to assign different colors to different weights (if such approach were adopted, it could also be put to good use to easily determine (b), the position and the speed of the hand while performing the hammering task). Such procedure is not sufficient to determine (c), which, instead, requires the implementation of a more elaborate procedure, as the one presented in this work.

The approach adopted to determine (a), (b) and (c) builds upon a few observations of anatomical nature. We first describe a simple model used to represent the hammering movement, i.e., a triangle composed by the arm, forearm and the shoulder-hand vector (Fig. 1). The variables of interest are:

- T, the time taken to complete a swing;
- $\gamma(T)$, the angular movement of the shoulder joint;
- $\beta(T)- \beta(0)$, the angular movement of the elbow joint;
- α, the angular displacement of the shoulder-hand vector.

Leveraging on these variables, we observe that tracking the shoulder-hand vector at all times (dotted vector in Fig. 1) could be put to good use to determine T and α, but such information

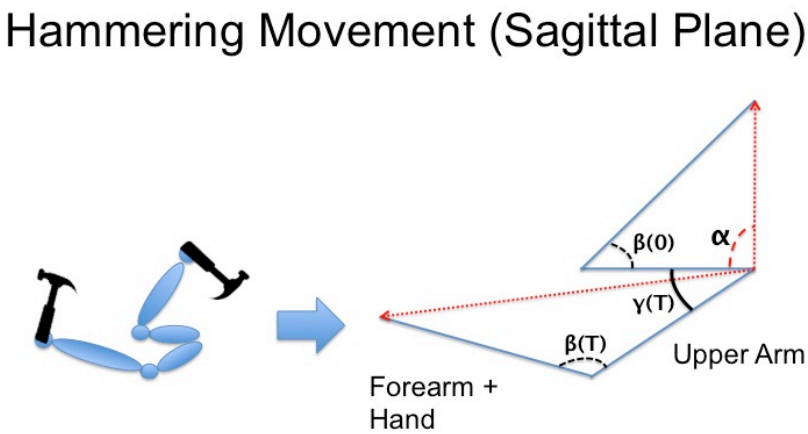


Fig. 1. Hammering movement: isolating the interested joints.

results insufficient to compute the other variables of interest. Combining anatomical and geometrical observations, however, it is possible to minimize the amount of information (and effort) required to also compute γ(T), β(T) and β(0). The value of L_2 (the length of the upper arm) can be assumed equal to the value of L_1 (the sum of the lengths of the forearm and the hand of a person), as indicated in Fig. 2, as this represents a good anatomical approximation utilized in literature [25]. However, even with the $L_1 = L_2 = L$ assumption, we still miss the length of the arm to estimate the elbow joint angle β(t) at all times and the maximum shoulder joint angle γ(T). This can be accomplished in different ways. The most complicated and time-consuming approach would require the implementation of an initial calibration phase: at the beginning of his/her shift, a worker stand with his/her arm straight and not attached to the body, in this way it is possible to simply estimate the total distance that separates his/her shoulder from the hand during the full extension of the arm (Fig. 3). Such approach is unrealistic, though, and in most cases unnecessary. Many production plants now require workers to *login*, for example swiping their badge or with RFID tag authentication, when working at specific tasks. It is hence possible

to also imagine a realistic scenario where, when a worker authenticates, profile information (including anatomical information, such as arm length) is retrieved and utilized for safety purposes. With any of the two aforementioned simple operations, the length of the upper arm plus that of the forearm may be obtained. An estimate of the values of β(t) and γ(T) can be found, hence, resorting to the simple trigonometric relations defined in Equations 1 and 2:

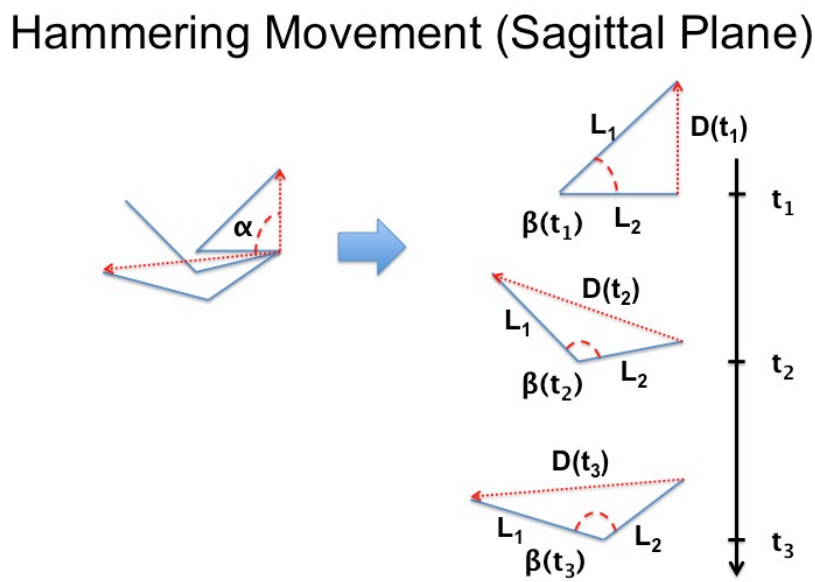


Fig. 2. Hammering movement: isolating the interested joints.

$$\beta(t) = \arccos\left(1 - \frac{D^2(t)}{2 \times L^2(t)}\right), \qquad (1)$$

$$\gamma(T) = \alpha + \frac{\pi - \beta(T)}{2}, \qquad (2)$$

where D(t) amounts to the length of the shoulder-hand vector at time t. It is now clear that, considering a generic hammering movement, the variables required to estimate all quantities of interest are the positions of: (a) the joint that is supporting the given movement (i.e., the shoulder), and, (b) the hand [43], [44].

Now, a way of tracking the shoulder-hand vector amounts to search for the maxima and minima of the human silhouette, obtained as the contour of the blob (obtained with a gradient operator)

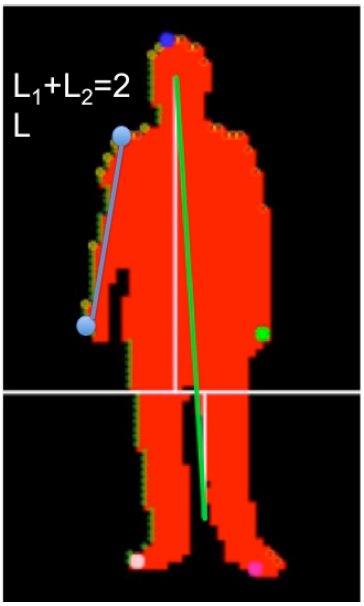


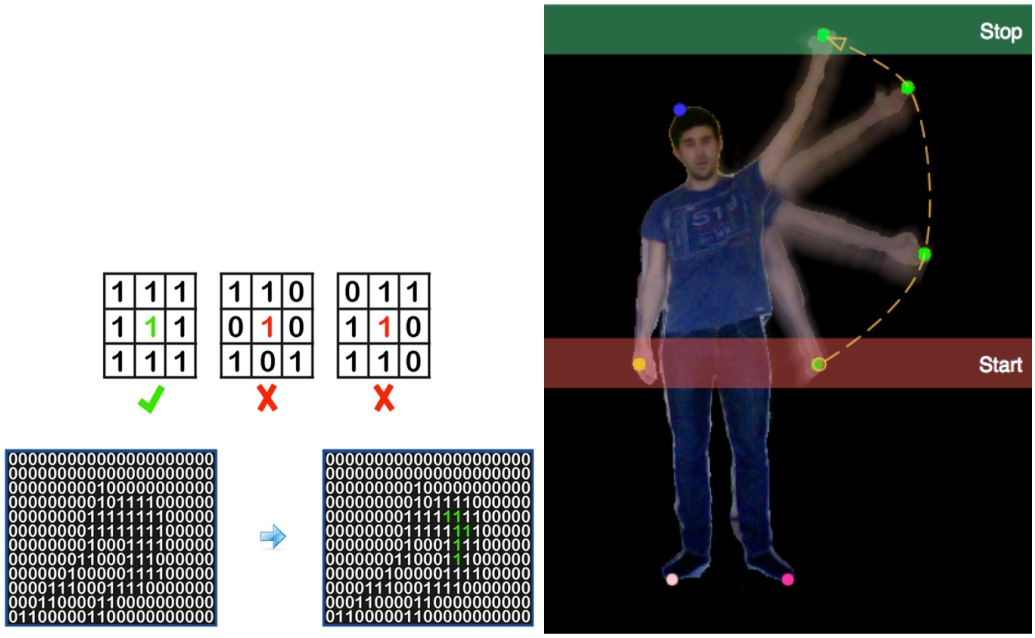
Fig. 3. Estimating the length of the entire arm (shoulder-hand vector).

individuated subtracting the frame where a worker is present from the static background image [42]. The obtained area can successively be divided into squares and the current position of the hand will be given by the square that, surrounded by other squares that also fall within the area, is located at the topmost position (Fig. 4.a and 4.b). As such methodology may suffer from discontinuities (the position of the hand may be missed in some difference frames), a Kalman filter may be utilized to estimate the current location of the square of interest in a consistent and continuous way. This procedure is feasible the moment a person's hand is located at an elevation that is higher than his/her head. Such condition is not always fulfilled, though, when hitting an object with a hammer. A solution to this problem can be found resorting to two video cameras, one pointing at the frontal and one pointing at the, lateral, sagittal plane. The video stream received from the frontal camera is used to track the positions of the upper extremities of interest

of the human body (i.e., head, hands and shoulders), as shown in Fig. 5. Such result is obtained determining the maxima of the blob in the *y* positive direction (i.e., upper direction), and, then proceeds recognizing whether those maxima are significant body parts (e.g., head, hands, shoulder). A procedure that detects whether a maximum is a significant body part, or not, may rely on the preventive knowledge of how many maxima are searched for (e.g., one to three in the positive *y* direction, when seeking for a person's head and hands) and on the depth of the gaps (the height of the local minimum between two maxima) separating them, i.e., a higher difference between minimum and maxima values indicating the presence of two significant body parts (e.g., head and hand in Fig. 5). How this can be practically implemented is presented by the pseudo code shown in Table I. In particular, the first operation that is performed (line 1), is that of saving the position of the absolute maximum in variable (i_M, $f(i_M)$). After individuating the global maximum, the algorithm then cycles (lines 2-7) seeking and saving (if eventually found) the positions of a second maximum value separated from the initial one by a local minimum (lines 3-4) and for a local minimum (lines 5-6).

Table I. Individuating hand and shoulder elevation from frontal camera blob.

```
1.  y_M = f(i_M), y_m = f(i_M), store(i_M, f(i_M));
2.  for (i = i_M; i < MAX; i++) {
3.          if (f(i) > y_m + GAP)
4.              y_M = f(i); store(i, f(i));
5.          if (f(i) < y_M − GAP OR f(i) < y_m)
6.              y_m = f(i); store(i, f(i));
7.          else  y_m = y_M; }
```



(a) **(b)**

Fig. 4. Following the extremities of a person with a frontal camera: (a) algorithmic solution, (b) practical result.

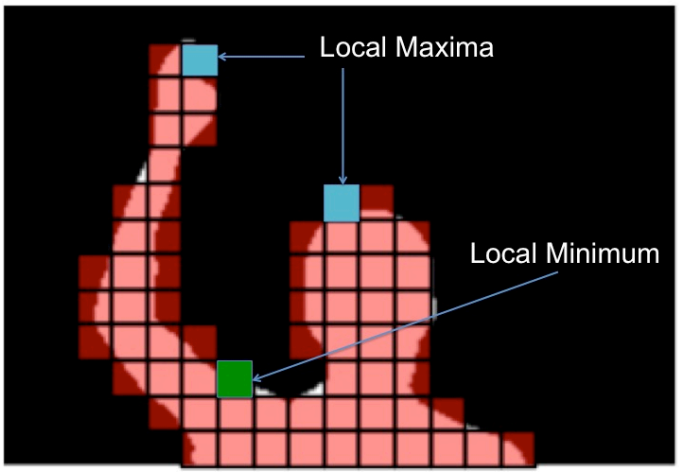


Fig. 5. Finding the shoulder as a local minimum.

Applying such methodology to our specific case of interest, it is hence possible to identify the elevations of the hand and shoulder of a person, first individuating the position of the head and then retaining the positions of the local minima and the local maxima that may be found in the

positive and negative *x* directions (e.g., one minimum and one maximum are individuated in Fig. 5). The second webcam, the lateral one pointing at the sagittal plane, can instead be exploited to assess the lateral positions of the shoulder, which may be approximated with the lateral position of the body, and the hand. Such information is sufficient to individuate the space-time coordinates of the shoulder-hand vector. A diagram summarizing the relevant phases of the hammering movement risk assessment is shown in Fig. 6.

3.3.2. Lifting an object from the ground

The attention now moves to the second task of interest: the posture adopted by workers while lifting a weight from the ground. The problem here amounts to develop a methodology capable

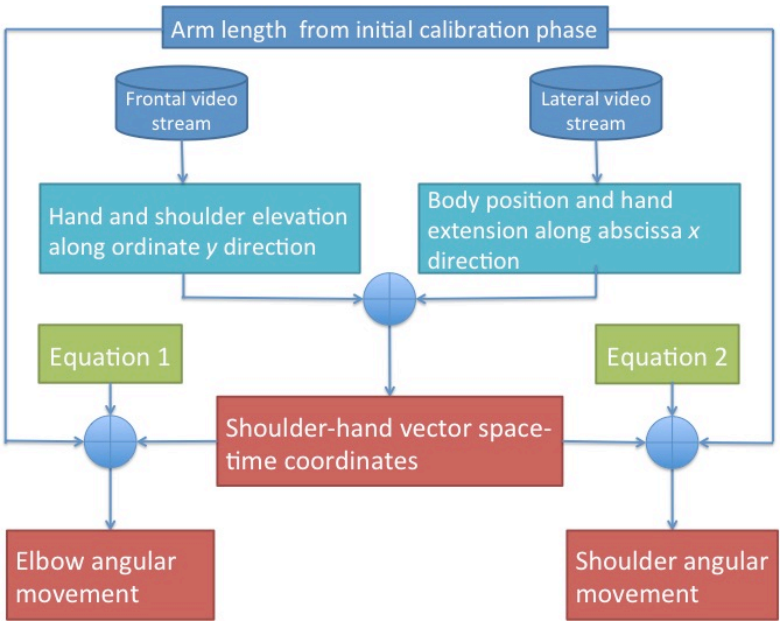


Fig. 6. Finding the shoulder as a local minimum.

of distinguishing incorrect from correct ways of performing such action: (a) one where a worker

bends his/her back (incorrect, as all stress concentrates on lower-back muscles), keeping the legs straight, and, (b) another where a worker lowers bending his/her knees (correct, all muscles contribute to the lifting action). Such problem has been approached, at first, investigating whether an incorrect posture (case (a)) could be detected observing the movement of a person's head. This would be, in fact, very convenient, as we have shown in Section 3.3.1 that determining the position of a worker's head with a very good precision is possible at all times. Such conjecture has been exploited performing a large set of experiments where the height of a person has been measured while lifting an object from the ground in the correct (incorrect) way. The experiments, unfortunately, revealed that simply following the elevation from the ground of a worker's head is not sufficient to determine the correctness (or not) of the utilized posture/movement. This can be explicitly observed examining the examples shown in Figs. 7 and 8: no pattern clearly distinguishes a situation where a correct movement is performed from one where an incorrect one is carried out.

Additional experiments have, hence, been carried out. Such experiments entailed processing the video stream obtained from a camera pointing at the sagittal plane, as represented in Fig. 9. In particular, we concentrated on the dynamics and on the relative positions of four different body features: head, lower back, hands, and feet. Such points have been obtained resorting to an algorithm that follows the philosophy adopted in Section 3.3.1: the algorithm computes the maxima and minima in the four directions of the plane (i.e., up, down, left and right) and employs methodologies that enable the continuous tracking of such points (i.e., Kalman filter).

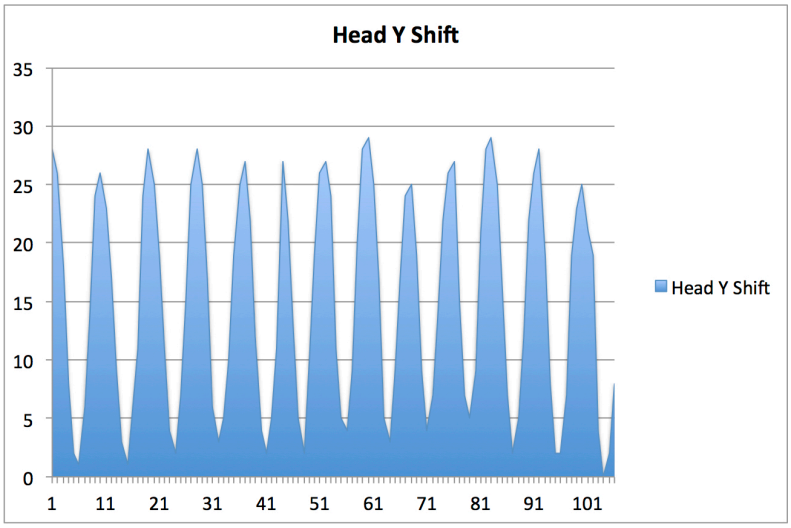


Fig. 7. Displacement of the head of a person while grabbing an object from the ground while doing the right movement.

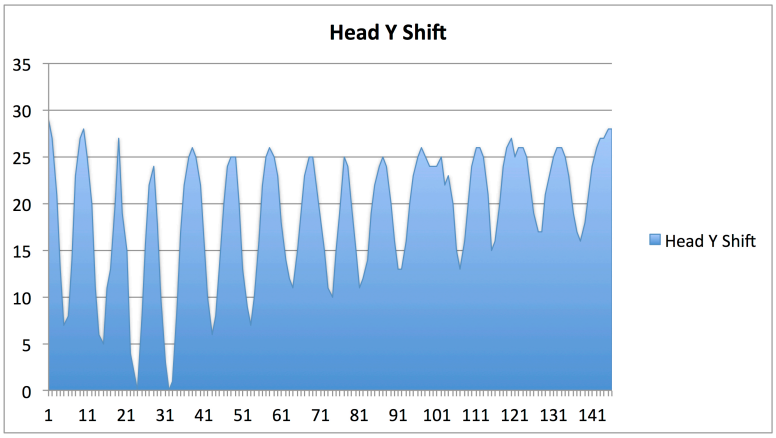


Fig. 8. Displacement of the head of a person while grabbing an object from the ground while doing the wrong movement.

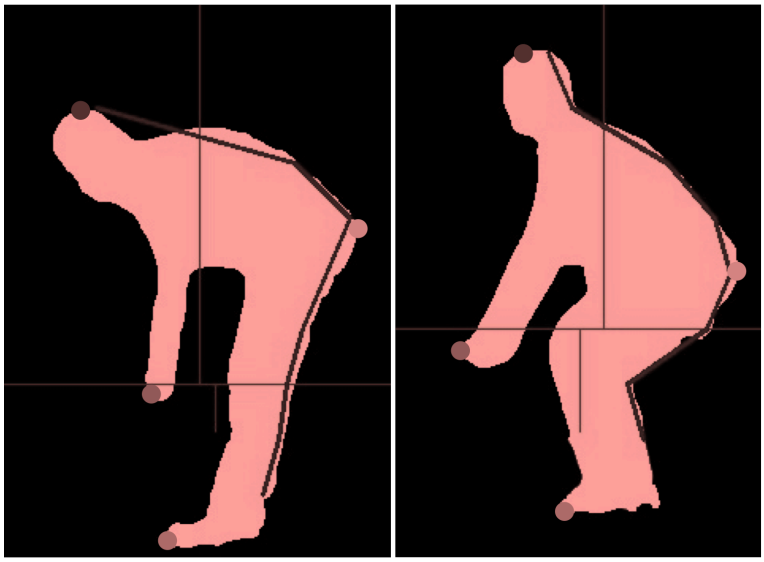


Fig. 9. Grabbing an object from the ground: incorrect (left) and correct (right) movements

We anticipate here what the experiments reveal in Section 4: a clear way of finding whether an (in)correct posture is being adopted depends not so much on an exact estimation of the position of each point, but on the extension of the movement of the lower back of a person. This is intuitively easy to picture, when a person bends his/her legs, the lower back moves more extensively than in a situation where the legs stand straight. This observation, however, is not

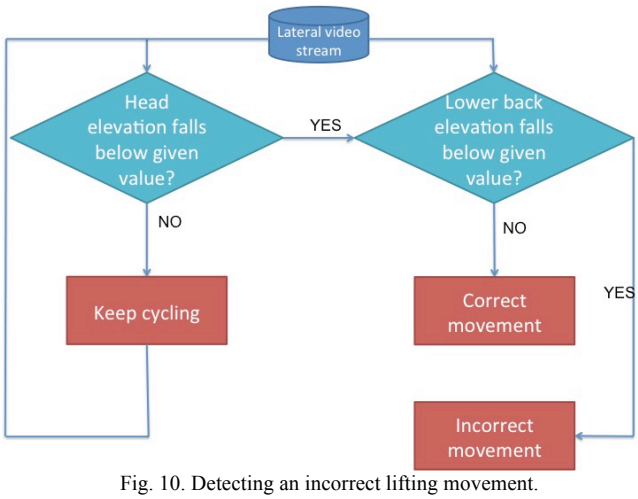


Fig. 10. Detecting an incorrect lifting movement.

sufficient to build a computer system capable of deciding when right or wrong movements are performed. Distinguishing a correct lifting movement from an incorrect one, in fact, requires three successive steps:

1. Detect whether an object is being lifted from the ground;
2. If an object is being lifted from the ground, follow the position of the lower back;
3. If the lower back does not move beyond a given elevation while performing the movement, the movement has been performed in an incorrect way.

While we revealed that the second and third steps could be developed following the lower back, no approach has so far been suggested for the first step. Figs. 7 and 8 provide us sufficient information to implement this step: following the elevation of the head of a person it is possible to establish whether that a person is bending to grab something from the floor. The flowchart represented in Fig. 10 summarizes how such system should work.

4. Results

We now proceed describing the results obtained when tracking a person while (Subsection 4.1):

- Hitting an object with a hammer;
- Lifting a weight from the ground;

Finally, for the sake of completeness, briefly contrasting the low cost solution that is here presented with a hardware based one which utilizes a depth-sensing camera (Subsection 4.2).

4.1. Webcam based results

For the first case of interest, our experiments prove the feasibility of tracking the positions of the hand and the shoulder of a worker while holding a hammer, frame by frame. In particular, it was possible to track both of the given body features with an almost perfect precision (97 to 99% of times) during experiment sequences with three different workers who performed fifty hammering repetitions (Table II). Figure 11 provides a sequence of three different frames taken from the camera pointing at a person while performing the hammering movements (each repeated action lasted approximately one second, yielding from 9 to 15 frames). The circles indicate the identified position of both the shoulder and the hand.

TABLE II

FOLLOWING THE HAND AND SHOULDER OF A WORKER WHILE HAMMERING: PERCENTAGE OF FRAMES WITH CORRECT TRACKING.

TOTAL NUMBER OF REPEATED ACTIONS: 50; FRAMES PER ACTION: FROM 9 TO 15.

Webcam distance\Body part	Hand	Shoulder
50 cm	**97%**	**99%**
120 cm	99%	99%

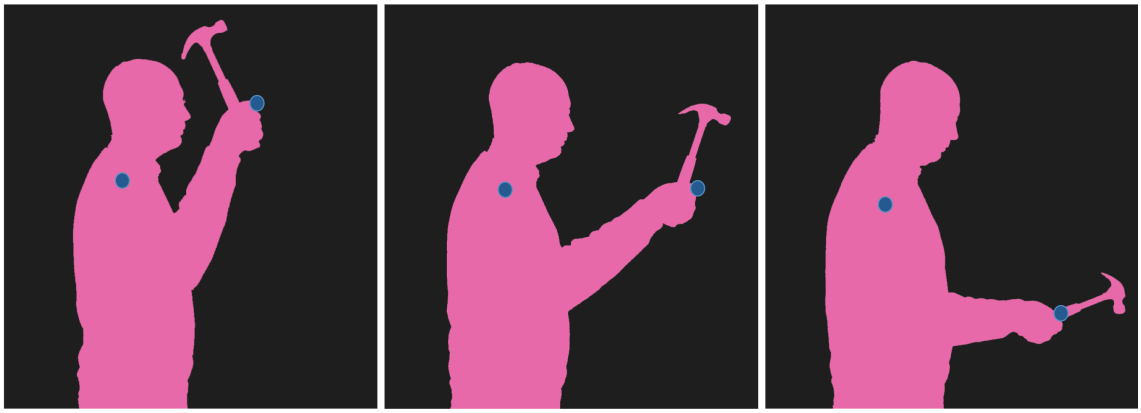


Fig. 11. Tracking the shoulder and the hand while hammering, image taken at a distance of 120 cm.

In addition, such approach could be flexibly adapted to follow the given body features while performing other different actions, finer than the hammering one. Table III shows the performance of a modified version of the algorithm presented in Subsection 3.3.1, while a worker carves the sole of a shoe, hence while performing a movement that is more confined in space, but slower, when compared to the hammering movement. As a further test, such methodology has also been extended to detect a worker's hand, while creating a shoe, i.e., sewing a piece of leather to the sole of a shoe (Table IV). As highlighted by the two Tables, the system tracking the carving and sewing movements performs as well as in the situation where a hammering movement was followed, with a single exception which will be clarified later.

TABLE III

FOLLOWING THE HAND AND SHOULDER OF A WORKER WHILE CARVING A SHOE: PERCENTAGE OF FRAMES WITH CORRECT TRACKING.

TOTAL NUMBER OF REPEATED ACTIONS: 50; FRAMES PER ACTION: FROM 9 TO 15.

Webcam distance\Body part	Hand	Shoulder
50 cm	**97%**	**99%**
120 cm	97%	99%

TABLE IV

FOLLOWING THE HAND AND SHOULDER OF A WORKER WHILE SEWING A SOLE: PERCENTAGE OF FRAMES WITH CORRECT TRACKING.

TOTAL NUMBER OF REPEATED ACTIONS: 50; FRAMES PER ACTION: FROM 9 TO 15.

Webcam distance\Body part	Hand	Shoulder
50 cm	**95%**	**99%**
120 cm	85%	99%

Compared to the situation where a hammering movement was tracked, sewing and carving are more complex actions, as they utilize both hands, which both need to be recognized as they move. The hands of a person who is sewing or carving a sole are, in fact, not easily isolated as one of the extreme parts of the human body, regardless of the plane. Nevertheless, the algorithm presented in Subsection 3.3.1 can be utilized to perform: (a) background subtraction along with blob identification, and, (b) Kalman filtering for tracking. It should here be noted, though, that (a) is accomplished in a slightly different way than before. In fact, for the two aforementioned movements, the body of the worker is also acquired and utilized as part of the background when aiming at identifying the position of the hands, as it stands in a static position while performing these jobs. Subtracting the background, hence, the areas that are individuated are those that move, in essence the hands, and the tools that the hands are holding.

Sewing and carving, however, typically require two hands, posing an additional technical problem, as hands should be distinguished. To this aim, the following extensions have been implemented. The system detected whether both hands were used, resorting to a k-means clustering algorithm (with k equal to 2) that took as an input all active squares in order to separate them into two clusters. In fact, setting $k = 2$ indicates any movement may have been performed by two hands. To discriminate whether one or both hands accomplished that movement, the following heuristics based on the distance between the two cluster centers that have been individuated at the preceding step were implemented: a distance that exceeded a given threshold value indicated that both hands accomplished two separate movements (Fig. 12.a, frames 1 through 7; as per the two different markers associated to the two hands), while if that

distance was less than that threshold value only one relevant movement was assumed to be performed by a single hand (Fig. 12.a, frame 8, and Fig. 12.b). The dynamics of the abovementioned approach is further illustrated in Figs. 13 (carving) and 14 (sewing).

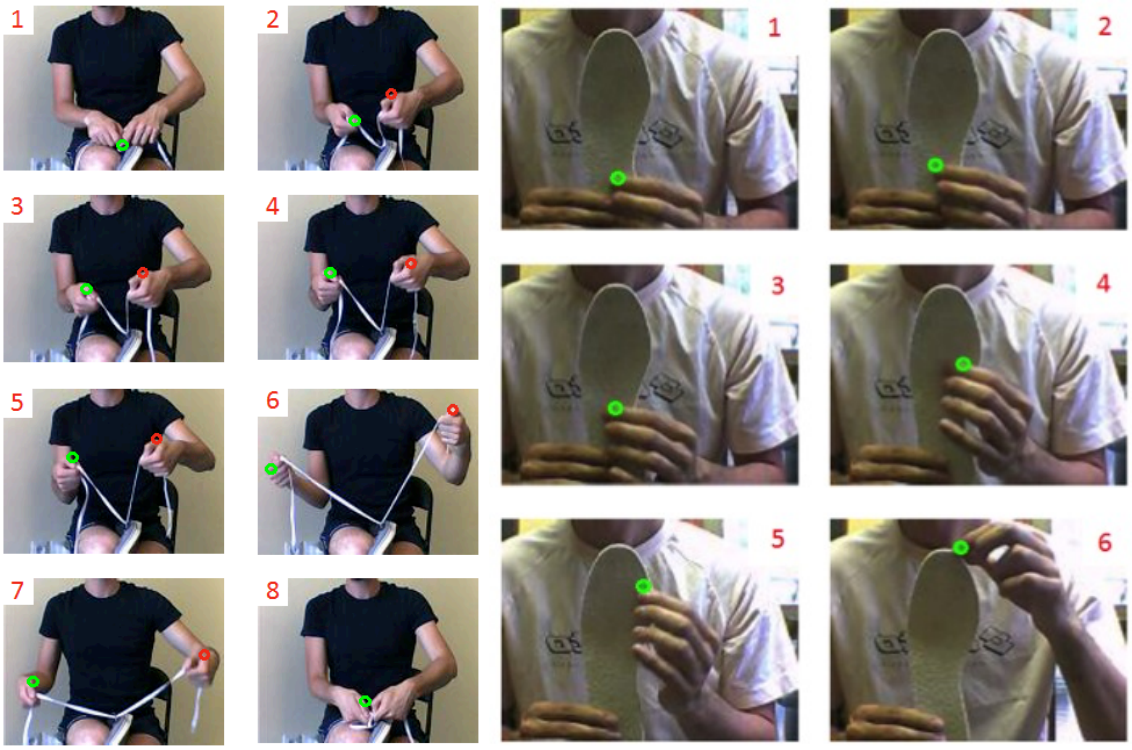


Fig. 12. Left to right: (a) tracking two hands (distance 120 cm), (b) tracking one hand (distance 50 cm).

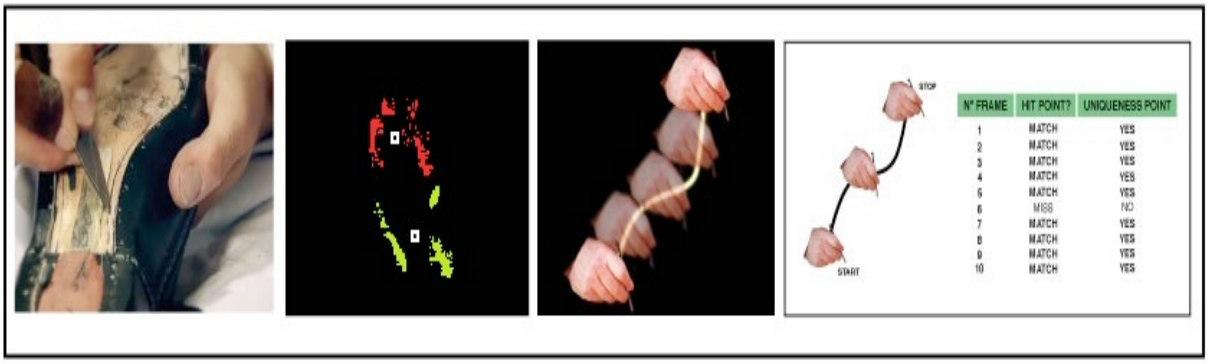


Fig. 13. Left to right: (a) carving the sole (distance 50 cm), (b) cluster center distances, (c) tracking one hand, (d) system performance while carving.

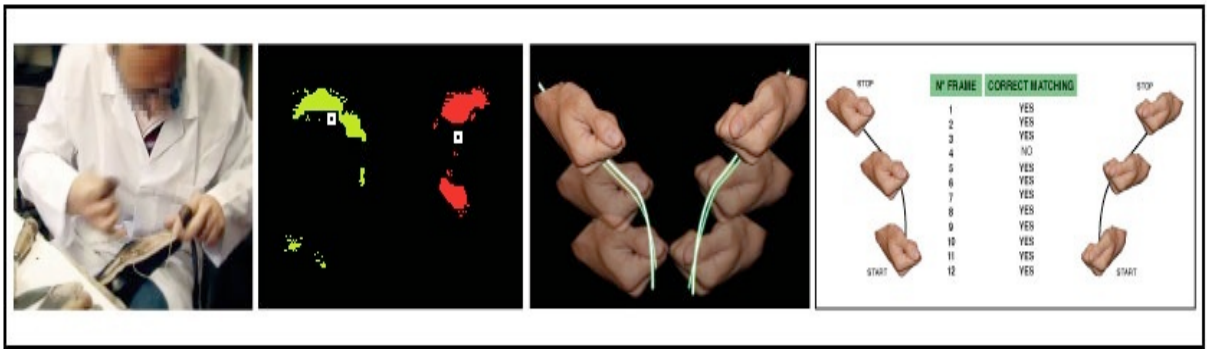


Fig. 14. Left to right: (a) sewing leather to a sole (distance 120 cm), (b) cluster center distances, (c) tracking two hands, (d) system performance while sewing.

As a final comment concerning the actions of sewing and carving, it is interesting to note that: (a) the shoulder is detected almost perfectly at all times following the same procedure utilized during the hammering movement (this is clearly due to the fact that it moves very slowly during the analyzed postures/gestures), while, (b) the precision with which the hand which sews the leather is tracked decreases by 10% when the webcam distance increases from 50 cm to 120 cm. The reason for such performance decrease was due to the interference that the second hand produced when moving close (in the same area) to the first one (the hand that holds the shoe): erroneously, in a few frames the second hand was followed instead of the first one.

After reviewing the results obtained for hammering (and also carving and sewing movements), some final considerations on the lifting action are in order. Precisely, Figs. 15 and 16 show the displacement along the Y-axis of the lower back position of a person at work while lifting a box from the ground, performing the incorrect and correct movement, respectively. Fig. 15 highlights what happens during a wrong movement: the extension of the lower back is limited as, in such

situation, it is the upper part of the body figure (the trunk and the arms) that moves most, erroneously. When a correct movement is performed, the lower back of a person moves extensively along the Y-axis, as shown in Fig. 16. The gap between the two body parts provides valuable information: the distance between the highest and lowest positions reached by the lower back indicates whether someone is adopting a correct posture (or not), while lifting an object from the ground. Related to this, it is interesting to notice that the distance covered (in terms of squares utilized along the Y axis) while moving in the correct way is always at least twice as long the distance covered while performing the wrong movement.

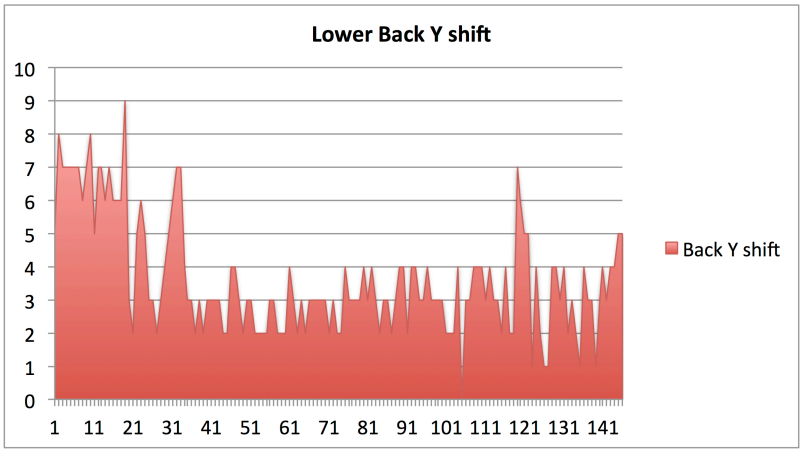


Fig. 15. Displacement of the lower back of a person while grabbing an object from the ground doing the wrong movement.

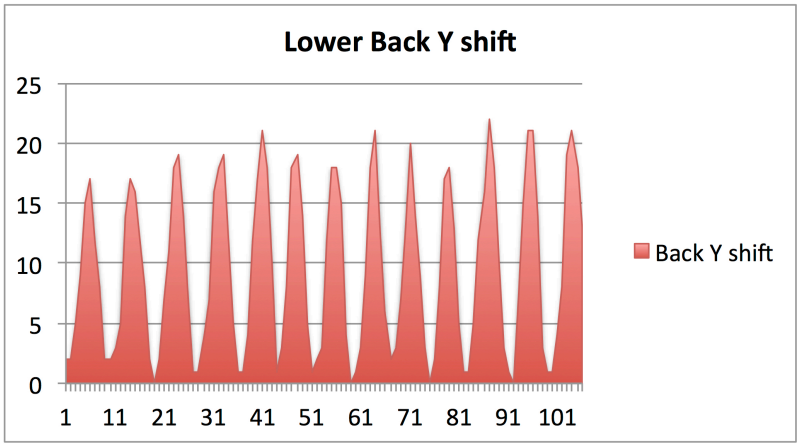


Fig. 16. Displacement of the lower back of a person while grabbing an object from the ground doing the right movement.

This information could be leveraged to automatically reveal when a given movement may lead to an injury or not. We here, instead, will limit to provide the results concerning the ability of our algorithms to precisely follow given parts of a human body (i.e., the lower back) with a simple camera pointing at the sagittal plane. These results are obtained comparing at given time instants the actual position of the lower back of a human body with the corresponding position revealed by the algorithms. On a set of fifty repetitions (i.e., lifting an object from the ground fifty times), we found that the correct position of the lower back was individuated in approximately 96% of frames.

TABLE V
ASSESSING FURTHER POSTURES/MOVEMENTS: PERCENTAGE OF FRAMES WITH CORRECT TRACKING.
TOTAL NUMBER OF REPEATED ACTIONS: 50; FRAMES PER ACTION: FROM 9 TO 15.

Movement	Webcam distance/Body part	Hand	Shoulder
Reaching frontal far away point.	**120 cm**	**95%**	**99%**
Grabbing an object placed above a person's head	**120 cm**	**94%**	**99%**

As a further and final proof of the flexibility of the proposed approach utilized in the case of grabbing an object from the ground, the system was tested while performing two special movements: (a) reaching a frontal far away point with both hands, and, (b) grabbing an object from a box placed above a person's head. Utilizing a webcam placed on the sagittal plane of a person that repeated such movement fifty times, the correct position of the person's hands was determined in the 95% and 94% of times for (a) and (b), respectively, finally proving that such approach can be easily extended to multiple types of postures and movements (Table V).

4.2 A depth sensing solution

While the use of webcams in a working setting may result challenging, and thus lead to prefer specialized hardware solutions such as depth sensing cameras (e.g., Kinect) or wearable devices, their low cost and widespread availability imposes a reflection on the effective opportunities and difficulties they set. With this in mind, we proceed comparing the characteristics of our original proposal, solely based on webcams, and a possible development of such proposal, which resorts to the use of the Microsoft Kinect sensor. Such task has been carried out re-implementing a few of the operations that have been described so far. Compared to an approach based on a webcam, the Microsoft Kinect SDK, offers developer functions, which directly return the position of 18 pre-defined joints of a person. In addition, the Microsoft Kinect SDK also provides direct information regarding the fact a user is leaning, or not, from the vertical position through the Lean Tracking APIs. Hence, from an algorithmic, but also a software development point of view, compared to the use of one or two webcams, the step of segmenting the position of a worker's body and the step of individuating the position of a given body feature of interest are not needed.

Hence, using the Kinect simplifies the tracking of the shoulder, elbow and hand joints, during the hammering movement, as well as the tracking of the head and the lower back during the lifting weight movement [41], [42], [43].

This said, a depth sensing camera as the Kinect may appear as a solution to many of the problems that have been approached in the previous Sections. Nonetheless, it is well known that a Kinect based solution is not the panacea for the scenarios of our interest. Recent studies

confirm, in fact, that the joints of the upper limbs of a person may be estimated incorrectly in variety of situations [44], [45], [46].

In addition, the approaches that have been so far presented in literature have been more deeply concerned with the problem of recognizing a particular activity performed by a person in a generic setting, while we focus on the characteristics of an activity which is already known [47], [48]. For example, in a working environment certain activities are performed only at given locations (this is how assembly lines are designed), and hence the activity recognition step offered by Kinect based solutions is not necessary. Instead, it is important to concentrate on the characteristics of those given activities performed at precise locations.

With the above discussion in mind, additional experiments have been performed to assess the difference, in terms of delay, between the use of a webcam and the use of a Kinect sensor. Table VI provides our results, together with other figures of merit pertaining this succinct comparison between the use of a depth sensing camera and the use of a webcam. In essence, no particular difference has been found in terms of speed (the particular application scenario can easily deal with a frame processing delay of a few hundreds of milliseconds) and in terms of background subtraction complexity (both a webcam approach and a Kinect approach suffer the presence of very dynamic backgrounds). As anticipated, the main differences fall in their costs and in the amount of complexity required on the software systems that they would respectively use.

TABLE VI
DEPTH SENSING-WEBCAM COMPARISON.

Technology\Characteristic	Cost	Tracking and detection delay	Software operations	Background dynamics
Webcam	$10	~50 ms	Body segmentation body part recognition and incorrect movement detection	static
Kinect	$150	50-200 ms	Joint tracking	slow, static

5. Conclusion

This paper introduces a novel injury risk detection system, solely based on the use of webcams. The convenience of such approach is witnessed by two facts: (a) it enables a real-time estimation of injury risks, and, (b) it is cost effective, as it relies on the use of simple off-the-shelf webcams. With such ideas in mind, it has been proven that it is possible to estimate in real-time the magnitude of those variables that can be put to good use to quantify the risk of remaining injured while working. Concluding, with this work it has been shown that complex tasks, such as injury prediction and detection, could be pervasively deployed with the use of off-the-shelf consumer technologies.

Acknowledgements

The authors feel indebted to Cristian Bertuccioli and Andrea Marcomini for the aid provided during the experiments phases of this work.

References

[1] Available, online, accessed on May 25th 2015: http://www.bls.gov/iif/oshcdnew.htm

[2] G.D. Herrin, M. Jaraiedi, C.K. Anderson (1986). Prediction of overexertion injuries using biomechanical and psychophysical models. *The American Industrial Hygiene Association Journal*, *47*(6), 322-330.

[3] Rainville, J., Smeets, R. J., Bendix, T., Tveito, T. H., Poiraudeau, S., & Indahl, A. J. (2011). Fear-avoidance beliefs and pain avoidance in low back pain—translating research into clinical practice. *The Spine Journal*, *11*(9), 895-903.

[4] Como, J. J., Smith, C. E., & Grabinsky, A. (2012). Trauma epidemiology, mechanisms of injury. *Essentials of Trauma Anesthesia*, 1.

[5] Yoon, J., Shiekhzadeh, A., & Nordin, M. (2012). The effect of load weight vs. pace on muscle recruitment during lifting. *Applied ergonomics*, *43*(6), 1044-1050.

[6] A.J. van den Bogert, T. Geijtenbeek, O. Even-Zohar, F. Steenbrink, E.C. Hardin (2013). A real-time system for biomechanical analysis of human movement and muscle function. *Medical & biological engineering & computing*, *51*(10), 1069-1077.

[7] S. Machado, R. Osório, N. Silva, M. Magini (2010). Biomechanical analysis of the muscular power of martial arts athletes. *Medical & biological engineering & computing*, *48*(6), 573-577.

[8] S. Zhang, H. Hu, H. Zhou (2008). An interactive Internet-based system for tracking upper limb motion in home-based rehabilitation. *Medical & biological engineering & computing*, *46*(3), 241-249.

[9] H. Zheng, N.D. Black, N.D. Harris (2005). Position-sensing technologies for movement analysis in stroke rehabilitation. *Medical and biological engineering and computing*, *43*(4), 413-420.

[10] G. Andreoni, M. Mazzola, O. Ciani, M. Zambetti, M. Romero, F. Costa, E. Preatoni (2009). Method for movement and gesture assessment (MMGA) in ergonomics. *International Conference on Digital Human Modeling*, San Diego, CA, USA, pp. 591-598.

[11] L. McAtamney, E.N. Corlett (1993). RULA e a survey method for the investigation of work-related upper limb disorders. *Applied Ergonomics,* 24, 91-99.

[12] M. T. Knauf, I. Z. Schultz, A. M. Stewart, R. J. Gatchel, (2014). Models of Return to Work for Musculoskeletal Disorders: Advances in Conceptualization and Research. In *Handbook of Musculoskeletal Pain and Disability Disorders in the Workplace* (pp. 431-452). Springer New York.

[13] F. Gerr, N. Fethke, L. Merlino, D. Anton, J. Rosecrance, M. P. Jones, M Marcus, and A Meyers. A Prospective Study of Musculoskeletal Outcomes Among Manufacturing Workers I. Effects of Physical Risk Factors. *Human Factors: The Journal of the Human Factors and Ergonomics Society* (2013): 0018720813491114.

[14] R. J. Gatchel, I. Z. Schultz, (2014). Future Research Directions for Preventing and Treating Occupational Musculoskeletal Disorders. In *Handbook of Musculoskeletal Pain and Disability Disorders in the Workplace* (pp. 485-498). Springer New York.

[15] M. S. Fimland, O. Vasseljen, S. Gismervik, M. B. Rise, V. Halsteinli, H. B. Jacobsen, P. C. Borchgrevink, H. Tenggren, and R. Johnsen. Occupational rehabilitation programs for musculoskeletal pain and common mental health disorders: study protocol of a randomized controlled trial. *BMC public health* 14, no. 1 (2014): 368.

[16] H. S. Miedema, H. F. Molen, P. P. F. M. Kuijer, B. W. Koes, A. Burdorf, (2014). Incidence of low back pain related occupational diseases in the Netherlands. *European Journal of Pain*, *18*(6), 873-882.

[17] R. J. Gatchel, N. D. Kishino, A. M. Strizak, (2014). Occupational Musculoskeletal Pain and Disability Disorders: An Overview. In *Handbook of Musculoskeletal Pain and Disability Disorders in the Workplace* (pp. 3-17). Springer New York.

[18] C. A. Mustard, A. Chambers, S. Ibrahim, J. Etches, P. Smith, (2014). Time trends in musculoskeletal disorders attributed to work exposures in Ontario using three independent data sources, 2004–2011. *Occupational and environmental medicine*, oemed-2014.

[19] R. Raja, S. R. Kingsbury, E. Wise, P. G. Conaghan, (2014). Primary care approaches to musculoskeletal multiple-site joint pain pharmacological therapy: a survey of general practitioners. *Prim Health Care Res Dev*, 1-6.

[20] N. Vignais, M. Miezal, G. Bleser, K. Mura, D. Gorecky, F. Marin (2013). Innovative system for real-time ergonomic feedback in industrial manufacturing. *Applied Ergonomics*, 44, 566-574.

[21] J.M. Hollerbach, T. Flash (1982). Dynamic interactions between limb segments during planar arm movement. *Biological cybernetics*, *44*(1), 67-77.

[22] C. Nordander, K. Ohlsson, I. Åkesson, I. Arvidsson, I. Balogh, G. Hansson, U. Stromberg, R. Rittner, S. Skerfving (2009). Risk of musculoskeletal disorders among females and males in repetitive/constrained work. *Ergonomics*, *52*(10), 1226-1239.

[23] B.P. Bernard, V. Putz-Anderson (1997). Musculoskeletal disorders and workplace factors: a critical review of epidemiologic evidence for work-related disorders of the neck, upper extremities, and low back. DHHS (NIOSH) Publication No. 97B141.

[24] B.R. da Costa, E.R. Vieira (2010). Risk factors for work‐related musculoskeletal disorders: a systematic review of recent longitudinal studies. *American journal of industrial medicine*, *53*(3), 285-323.

[25] A. Osborne, C. Blake, B.M. Fullen, D. Meredith, J. Phelan, J. McNamara, C. Cunningham (2012). Prevalence of musculoskeletal disorders among farmers: a systematic review. *American journal of industrial medicine*, *55*(2), 143-158.

[26] W. Yu, I. Yu, Z. Li, X. Wang, T. Sun, H. Lin, S. Xie (2012). Work-related injuries and musculoskeletal disorders among factory workers in a major city of China. *Accident Analysis & Prevention*, 48, 457-463.

[27] T. Massey, G. Marfia, A. Stoelting, R. Tomasi, M. Spirito, M. Sarrafzadeh, G. Pau (2011). Leveraging social system networks in ubiquitous high-data-rate health systems. *Information Technology in Biomedicine, IEEE Transactions on*, *15*(3), 491-498.

[28] R.G. Radwin, M.L. Lin (1993). An analytical method for characterizing repetitive motion and postural stress using spectral-analysis. *Ergonomics* 36, 379-389.

[29] A. Freivalds, Y. Kong, H. You, S. Park (2000). A comprehensive risk assessment model for work-related musculoskeletal disorders of the upper extremities. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting.* 44(31), 5-728.

[30] E. Bernmark, C.C.P.S. Wiktorin (2002). A triaxial accelerometer for measuring arm movements. *Applied Ergonomics*, 33, 541-547.

[31] D. R. Mullineaux, S. M. Underwood, R. Shapiro, J. W. Hall, 2012. Real-time biomechanical biofeedback effects on top-level rifle shooters. *Applied Ergonomics*. 43, 109-114.

[32] G.C. David (2005). Ergonomic methods for assessing exposure to risk factors for work-related musculoskeletal disorders. *Occupational Medicine.* 55, 190-199.

[33] H.C. Agustin, Posture position sensor. U.S. Patent 5,168,264, issued December 1, 1992.

[34] Available, online, accessed on May 5th 2014: http://www.lumobodytech.com/lumoback/

[35] Available, online, accessed on May 5th 2014: http://www.fastcoexist.com/3025324/fund-this/does-this-posture-sensing-device-really-keep-you-sitting-up-straight

[36] Available, online: http://ergo.human.cornell.edu/pub/AHquest/CURULA.pdf

[37] M. Roccetti, G. Marfia, A. Semeraro (2012). Playing into the wild: A gesture-based interface for gaming in public spaces. *Journal of Visual Communication and Image Representation*, *23*(3), 426-440.

[38] G. Marfia, M. Roccetti, G. Matteucci, A. Marcomini, (2012, August). Technoculture of handcraft: fine gesture recognition for haute couture skills preservation and transfer in Italy. *ACM SIGGRAPH 2012 Posters* (p. 81). ACM.

[39] E. Spyropoulos, E. Chroni, P. Katsakiori, G. Athanassiou (2013). A quantitative approach to assess upper limb fatigue in the work field. *Occupational Ergonomics*, *11*(1), 45-57.

[40] J.N. Côté, D. Raymond, P.A. Mathieu, A.G. Feldman, M.F. Levin (2005). Differences in multi-joint kinematic patterns of repetitive hammering in healthy, fatigued and shoulder-injured individuals. *Clinical Biomechanics*, *20*(6), 581-590.

[41] D.S. Alexiadis, P. Kelly, P. Daras, N.E. O'Connor, T. Boubekeur, M.B. Moussa, (2011, November). Evaluating a dancer's performance using kinect-based skeleton tracking. In *Proceedings of the 19th ACM international conference on Multimedia* (pp. 659-662). ACM.

[42] D. González-Ortega, F.J. Díaz-Pernas, M., Martínez-Zarzuela, M. Antón-Rodríguez (2014). A Kinect-based system for cognitive rehabilitation exercises monitoring. *Computer methods and programs in biomedicine*, 113(2), 620-631.

[43] A. Da Gama, T. Chaves, L. Figueiredo, V. Teichrieb (2012, March). Poster: improving motor rehabilitation process through a natural interaction based system using kinect sensor. In *3D User Interfaces (3DUI), 2012 IEEE Symposium on* (pp. 145-146). IEEE.

[44] A. Kar (2010). Skeletal tracking using microsoft kinect. *Methodology*, *1*, 1-11.

[45] S. J. Ray, J. Teizer, (2012). Real-time construction worker posture analysis for ergonomics training. *Advanced Engineering Informatics*, *26*(2), 439-455.

[46] Available, online, accessed on May 25th 2015: http://structure.io/openni

[47] K. Wang, X. Wang, L. Lin, M. Wang, W. Zuo, (2014). 3D Human Activity Recognition with Reconfigurable Convolutional Neural Networks. In *Proceedings of the ACM International Conference on Multimedia* (pp. 97-106). ACM.

[48] B. Ni, Y. Pei, Z. Liang, L. Lin, P. Moulin, (2013). Integrating multi-stage depth-induced contextual information for human action recognition and localization. In *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on* (pp. 1-8). IEEE.