Reframing Haute Couture Handcraftship: How to Preserve Artisans' Abilities with Gesture Recognition

Gustavo Marfia, Marco Roccetti, Andrea Marcomini, Cristian Bertuccioli, and Giovanni Matteucci

Alma Mater Studiorum - University of Bologna
Mura Anteo Zamboni 7, 40127 - Bologna, Italy
{gustavo.marfia,marco.roccetti,andrea.marcomini, cristian.bertuccioli,giovanni.matteucci}@unibo.it

Abstract. Computer gaming has often represented a fertile ground for the implementation and testing of novel and engaging human-computer interactions systems. Such phenomenon has occurred first with mice and joysticks and keeps on going, increasing in complexity and realism, with body-based interfaces (e.g., Wii, Kinect). Now, many fields and applications could benefit from these advances, starting with those where interactions, rather than physical objects, play a key role. Relevant exemplars can be found within many specimen of intangible cultural heritage (e.g., music, drama, skills, craft, etc.), whose preservation is possible only thanks to those tradition bearers that patiently bestow their knowledge upon new generations. Italian luxury crafts, which range from sports cars to high-end clothing, for example, often obtain their high quality and consequent reputation from a mix of intangible artistic and technological skills. The preservation of such skills and the persistent creation of such handcrafts has been possible, in time, thanks to those "master-apprentice" relations that have retained the quality standards that stand behind them. Nowadays such type of relations remain no longer easy to implement, as creation and production paradigms have undergone radical changes in the past two decades (i.e., globalization of production processes), making the transfer and preservation of skills challenging. Inspired by the advances made in human-computer interaction schemes for gaming, in this work we propose a non-invasive encoding of artisans manual skills, which, based on a set of vision algorithms, is able to capture and recognize the gestures performed by one or both hands, without needing the use of any specific hardware but a simple video camera. Our system has been tested on a real-world scenario: we here present the preliminary results obtained when encoding the gestures performed by an artisan while working at the creation of haute couture shoes.

Keywords: Algorithms, Design, Performance, Experimentation

1 Introduction

Craftsman skills are the result of experience and technique learned over time through consolidated careers in designing and manufacturing a specific product, where tradition is often mixed with art and the use of new technologies, just as envisioned long ago by the Bauhaus scholars [1]. However, the "master-apprentice" teaching model, which has been at the basis of the transfer of such skills, is no longer as common as in the past. Hence it is not easy to teach and pass on knowledge and skills learned by specialized craftsmen who create and develop sophisticated and unique craft products, in a perfect balance between tradition and modernity. For these reasons, it is clear that, within a crafting enterprise, it is very important to find ways of archiving the most relevant collections of intangible assets and knowledge. This is a process where technologies that have been devised in the realm of computer gaming can make a difference and lead to the preservation, through digital encodings of gestures and movements, of the knowledge and experience required in the craft industry [2].

Such benefits, in addition, can be extended beyond the preservation of knowledge, and also be utilized to analyze and assess the processes and gestures that are employed to design and create particular products, for example. Also for this reason, the use systems that can track and recognize human movements within the context of handcrafting opens up to a new range of opportunities and applications that would have been unimaginable before now.

Now, the contribution of this work is the design and implementation of a system capable of tracking and encoding the actions performed by an artisan at work. It operates in two steps. The first is that of recognizing and tracking the features of a human body, while performing coarse movements. The second is the recognition and tracking of the actions performed by one or both hands, while making fine, but also fast, gestures within a restricted area. With these two steps we are able to follow the complete set of movements and gestures, which an artisan can perform while working.

Before ours, a few works have been proposing the idea of preserving cultural heritage with the use of modern technologies, for example in the context of knitting [3]. However, the cited stream of work principally leveraged on all those technologies that can support and emphasize the social aspects that intervene while crafting (e.g., exchange of information with social networks) rather than on obtaining, through the tracking and recognition of gestures, a digital coding of experiences and skills. An interesting technique, in this latter sense, has been proposed in [4], [5], where the authors demonstrated the feasibility of tracking and recognizing handmade gestures with the sole use of a video camera and a pair of special gloves, specifically customized (i.e., colored with a give pattern) for such purpose. However, when applied to haute couture crafting enterprises, the requirement of wearing special gloves can represent a disturbing factor for an artisan, thus jeopardizing the performance of the natural gestures executed while creating.

Following the recent achievements that have been got in the field of gestural gaming interfaces, in the remainder we will show that it is possible to encode and maintain the knowledge of craft skills through a gesture recognition system based on non-invasive technology. Starting from the idea that any interaction with our system must be realized without the use of any tangible device or equipment, i.e., without using dedicated and specialized hardware ([2], [6]), we realized a system that can encode craft gestures, with the sole use of a single video camera and a set of specialized software algorithms. As we shall shortly see, such system has been successfully applied to the tracking and recognition of the gestures involved in the creation of haute couture shoes.

2 Our Approach

The actions performed by any artisan can be roughly of two types, depending on the gesture that it performs: fine and coarse. Intuitively, actions that involve fine gestures are performed in all those cases where a high degree of precision is required, as in the cases where a shoe is stitched or when a mechanical wristwatch is tweaked, for example. We consider coarse all the remaining actions where, instead, the same precision is not mandatory (e.g., moving an object from one end to another of a lab). In the following we will first focus on how fine-grained gestures can be tracked and recognized, and then move on to how coarse ones can be followed as well, in the context of the production of haute couture shoes.

The production of shoes involves a long set of steps which range from, choosing the leather of which they will be made of to stitching together the midsole, the sole, the vamp and the heel sections that compose them. Among all the possible actions, we selected two significant gestures in the production of shoes: carving and stitching a sole (Figs. 1.a and 2.a). These two actions are typically performed in a sequence, one after the other: at first, a sole is carved in order to create a groove; then the obtained groove is used as a track to sew together the remaining leather parts of the shoe to the given sole. The coding of these gestures lies in the determination of: (a) the number of times that a particular gesture is repeated and (b) the path that is followed while performing that gesture. Our system supports this type of coding through the tracking of the hand movements of a craftsman at work (Figs. 1.c and 2.c).

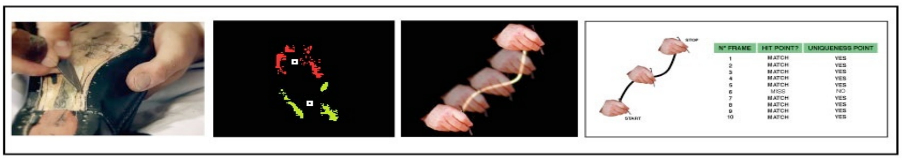

Fig. 1. Left to right: (a) carving the sole, (b) distance of the center of clusters, (c) tracking one hand, (d) system performance while carving.

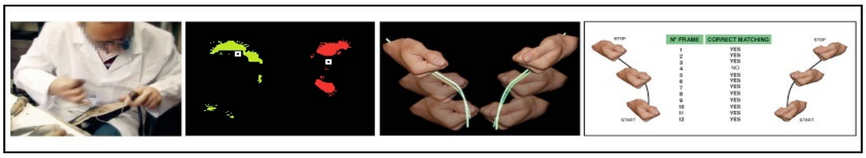

Fig. 2. Left to right: (a) sewing leather to a sole, (b) distance of the cluster centers, (c) tracking two hands, (d) system performance while sewing.

Technically, our system operates as follows. A video camera is placed in front of the areas where an artisan works, at a close distance from where gestures and movements are typically performed (we tested our system with this distance set to 50 cm and to 120 cm). After capturing two consecutive video frames, the system first of all applies a Gaussian filter to remove any high-frequency noise. Every second frame is then subtracted from the first one in order to obtain the difference frame between the two. Such frame, in fact, contains the information pertaining any change that occurred in time, between the moment when the first frame was captured and that when the second was acquired. Within the difference, hence, our system searches for any macro area where the two frames effectively differ, as this may be indicating that, single movements, or also multiple movements, have been performed. This step is performed dividing the difference frame into a grid where a square becomes active when it contains an active macro area. However, an active square may not represent a solid indication of motion, if it lies in an isolated position (the motion of one hand also involves the motion of the connected arm, for example). For this reason we also apply a filter that eliminates all those active squares that fall into isolated positions, since they cannot be representing significant movements.

Now, once the areas where motion has occurred have been individuated, an interesting problem is that of determining whether an artisan is working, utilizing one or both of its hands. In order to detect whether both hands are used, we adopted the following approach. In particular, we first utilize a k-means clustering algorithm (with k equal to 2) that takes as an input all active squares in order to separate them into two clusters. In fact, with $k = 2$ we are considering, a priori, that any movement may have been performed by two hands. Obviously this is not always the case: in fact, there may be actions, which require a single hand and others, which require both. Hence, in order to discriminate whether one or both hands accomplished a given movement, we adopted the following heuristics based on the distance between the two cluster centers that have been individuated at the preceding step (Figs. 1.b and 2.b): a distance that exceeds a given threshold value indicates that both hands accomplished a movement (Fig. 3.a, frames 1 through 7), while if that distance is less than that threshold value the movement is probably performed by a single hand (Fig. 3.a, frame 8, and Fig. 3.b). Finally, after having determined whether one or two hands are moving in front of the camera, their position is computed as follows. If only one moving hand is detected, its position will be identified as the cluster center positioned higher in the grid (Fig. 3.b); otherwise, if two hands are in motion, their positions will be estimated as the center of the two separated clusters. Clearly, with such approach discontinuities may appear, yielding situations where, due to a number of reasons (e.g., illumination level), first one and then two hands are detected, and vice versa; to overcome such type of problem, we integrated a Kalman Filter in our system, whose role is that of tracking consistently and gracefully the trajectories along which a particular movement is carried out.

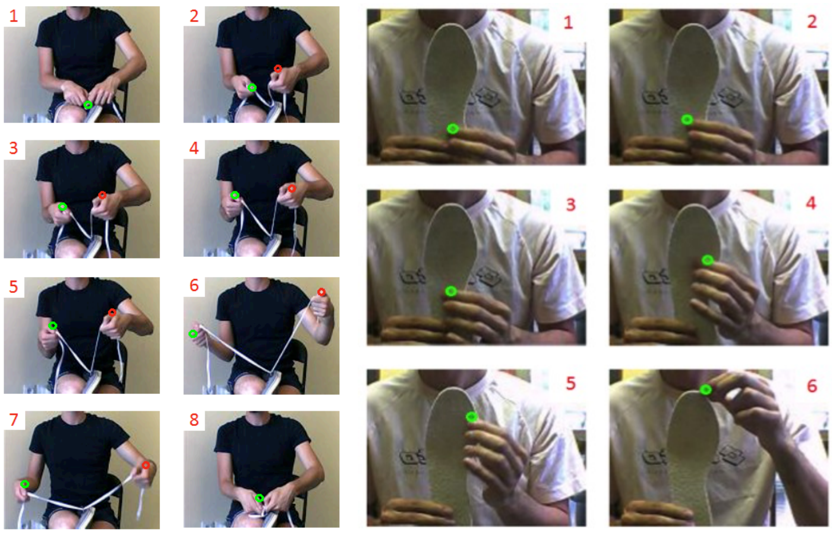

Fig. 3. Left to right: (a) tracking two hands, (b) tracking one hand.

Now, the tracking and recognition of any coarse gesture can be performed utilizing a video camera that is placed at some distance from an artisan, hence capturing its whole figure. The algorithm that we implemented to carry out this task lies along the lines of the one that has been presented in [7] and works as follows. The basic idea that is implemented is that of determining the position of five crucial points of the human body: the head, the hands and the feet of a person (Fig. 4.a). In brief, this is performed by first determining the positions of the two feet, which may be detected as the two minimum points of a human blob which has been separated from the static background. After this first step, and utilizing anatomical proportions, the same human blob is divided into four sectors (Fig. 4.b), two that include its bottom part (i.e., below the hip line) and two that, instead, incorporate its top (i.e., above the hip line). Within the top sector, our algorithm proceeds searching for three maxima: if these are successfully found our algorithm has detected also the hands and the head of the moving person. If not, our algorithm assumes that the global maximum represents the head of the person and adopts additional strategies to find the remaining positions (i.e., those of one or both hands), again leveraging on the idea of searching for the maximum of the blob in a given direction. This said, this methodology can be adopted to track and recognize a wide set of actions that are performed when crafting a shoe, ranging from those where an artisan transports any leather or other material to its workstation, to those that are executed while interacting with different machinery and tools (e.g., automatic stitchers, leather polishing machines, etc.).

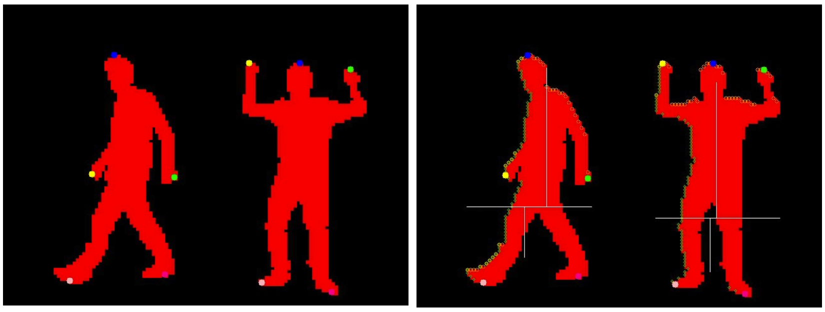

Fig. 4. Detecting coarse actions. Left to right: (a) detected points, and, (b) human body sectors.

Obviously, once the data has been obtained, the goal can be twofold: (a) use this as a benchmark and compare it against individuals for performance feedback, and, (b) use the acquired information as a training set that can be adopted by unskilled apprentices. All this leaves an open set of questions unanswered, which require further research and investigation. In fact, how could it be possible to efficiently compare the gestures performed by two different artisans working at the same task? These gestures comprise actions that are not only based on simple repetitions, but also synthetize creativity aspects that could give different, but similarly remarkable, results. To this date, our approach to this problem has only relied on a geometrical analysis and comparison of the trajectories that are returned by our tracking activities. In particular, two gestures are considered to be similar, if their corresponding trajectories begin (and end) in nearby areas. More precisely, our algorithm recognizes as correct all those trajectories that flow within a stripe of a given size (the idea is that a certain degree of tolerance is admitted to capture any difference produced by creativity). Finally, each given trajectory must terminate within a given period of time, after which that gesture has been performed too slowly to be considered correct. This mechanism was devised to make our recognizer able to consider as correct a wider set of movements with slightly different trajectories that differentiate only on the basis of a few geometric differences, nonetheless there is awareness on our side that this methodology is rough and needs to be refined, for example inspired by relevant intuitions provided in [3].

3 Empirical Results

The performance of our system has been assessed in reality, thanks to the collaboration with a renowned Italian haute couture shoe brand. In particular, we were able to assess how our system behaved while performing fine gestures when two different actions were being carried out: (i) carving the sole of a shoe, which required the movement of a single hand, and, (ii) sewing the leather to a sole, which, instead, required the use of two hands.

Table 1. Single hand.

Euclidean distance (pixel)	Camera distance (cm)	Match (%)	Uniqueness point (%)
35	50	96.65	98.08
35	120	96.8	100
45	50	96.15	100
45	120	93.57	100

All the results discussed in this Section were obtained utilizing a camera whose frame rate was set to 30 fps. The performance figures pertaining our algorithm while the first movement was executed are shown in Table 1. In particular, when dealing with a single hand, we notice that it is possible to encounter situations where, erroneously, two hands are detected. However, when choosing a threshold value for the distance between the two centers that fell in the 35 to 45 pixels range, we obtained a very close match and a good estimate in determining the movement of a single hand (Fig. 1.d).

Table 2. Two hands.

Euclidean distance (pixel)	Camera distance (cm)	Match (%)
35	50	94.54
35	120	84.78
45	50	83.82
45	120	90.9

The results pertaining movement (ii) are, instead, shown in Table 2. Also in this case the threshold value between the cluster centers (i.e., distance between the two hands) plays an important role. In particular, the closer are the hands to a camera, the larger is the size of the motion-sensible areas. As a consequence, the distance between the cluster centers will be small/large when the hands are close/distant from the camera. This means that the threshold value needs to be small in all those cases where the camera is kept close to an artisan, and large in all other cases. We hence can better determine if we have one or two hands involved during the performance of a given movement (Fig. 2.d).

4 Conclusion

We implemented a prototype, which, with sole use of a webcam, is able to track and encode the trajectory of a movement made by a craftsman at work. Our first results indicate that there is a good accuracy in following fine gestures, as well as coarse ones, when a video camera is both close or far from a craftsman at work. What is relevant is that all this has been made possible thanks to techniques and experiences drawn from the computer gaming field and allied technologies [8], [9], [10], [11].

5 Acknowledgements

The authors wish to thank the ALTER-NET Project for the financial support.

References

1. Maulsby, L.M.: Bauhaus Modern and Bauhaus Culture: From Weimer to the Cold War. Wiley Journal of Architectural Education 63, 1 (2009)
2. Roccetti, M., Marfia, G. and Zanichelli, M.: The Art and Craft of Making the Tortellino: Playing with a Digital Gesture Recognizer for Preparing Pasta Culinary Recipes. Computers in Entertainment, ACM, 8 (4) (2010)
3. Rosner, D.K. and Ryokay, K.: Reflections on Craft: Probing the Creative Process of Everyday Knitters. In: 7th ACM Conf. on Creativity and Cognition, Berkeley, pp. 195-204 (2009)
4. Wang, R.Y. and Popovic, J.: Real-time Hand-Tracking with a Color Glove. ACM Trans. Graph. 28 (3) (2009)
5. Palazzi, C.E., Roccetti, M., Marfia, G.: Realizing the Unexploited Potential of Games on Serious Challenges. Computers in Entertainment, ACM, 8 (4) (2010)
6. Marfia, G., Roccetti, M., Matteucci, G. and Marcomini, A.: Technoculture of Handcraft: Fine Gesture Recognition for Haute Couture Skills Preservation and Transfer in Italy. In: 39th ACM International Conference and Exhibition on Computer Graphics and Interactive Techniques Posters, Los Angeles (2012)
7. Roccetti, M., Marfia, G., Bertuccioli, C.: Day and Night at the Museum: Intangible Computer Interfaces for Public Exhibitions. Submitted for publication (2012)
8. Palazzi, C.E., Ferretti, S., Roccetti, M., Pau, G., Gerla, M.: How Do You Quickly Choreograph Inter-Vehicular Communications? A Fast Vehicle-to-Vehicle Multi-Hop Broadcast Algorithm, Explained. In: 4th Annual IEEE Consumer Communications and Networking Conference, pp. 960-964 (2007)
9. Aldini, A., Gorrieri, R., Roccetti, M., Bernardo, M.: Comparing the QoS of Internet Audio Mechanisms via Formal Methods. ACM Transactions on Modeling and Computer Simulation 11 (1), pp. 1-42 (2001)
10. Leontiadis, I., Marfia, G., MacK, D., Pau, G., Mascolo, C., Gerla, M.: On the effectiveness of an opportunistic traffic management system for vehicular networks. IEEE Transactions on Intelligent Transportation Systems 12 (4), pp. 1537-1548 (2011)
11. Marfia, G., Roccetti, M.: TCP at last: Reconsidering TCP's role for wireless entertainment centers at home. IEEE Transactions on Consumer Electronics 56 (4), pp. 2233-2240 (201)