LISTENING TO UNANIMATED OBJECTS' STORIES FOR TREATMENT AND REPAIR: A COMPUTER VISION APPROACH

Marco Roccetti, Alessandro Amoroso, Cristian Bertuccioli, Andrea Marcomini
Department of Computer Science and Engineering
Mura Anteo Zamboni 7, 40127 Bologna
{roccetti, amoroso}@cs.unibo.it
+39 051 2094503

Gustavo Marfia, Giovanni Matteucci
Department of Life Quality Studies
Corso D'Augusto 237, 47921 Rimini
{gustavo.marfia, giovanni.matteucci}@unibo.it
+39 0541 434044

ABSTRACT

In any medical practice an important part of the treatment strategy includes listening to what a patient has to say. Considering a broken object as a subject of treatment, we claim that listening to an object's story provide an aid to its repair. A prominent question emerges hence: how can objects tell their stories? Based on the consideration that an object's story corresponds to its peculiar lifecycle, we are confident that it can be acquired by "observing" how that object is crafted, manipulated and operated along its life. Two sweaters, for example, may be exactly the same at the beginning, but depending on how they have been worn, and on the decay process to which they have been subjected, their correspondent stories may be very different. With this in mind, we have tried to exploit fine gesture recognition algorithms that have been put to good use to capture the stories of unanimated objects that are crafted and then manipulated by human beings. This approach appears very promising when searching for mechanisms capable of automatically understand what objects have to tell to us.

***Index Terms*—** Repair, storytelling, talking objects, gesture recognition, computer vision, experiments

1. INTRODUCTION

The typical process that leads to the diagnosis of a disease walks through an initial phase where a doctor listens to its patients' stories, then searches for any visible sign of the reported symptoms, finally evaluating the results of a number of relevant lab tests. Hopping back and forth between these information sources, a doctor typically ends formulating a theory regarding which are the possible causes of the observed symptoms and proposing a treatment plan for their definite eradication. All in all, we can summarize this process in two basic steps: (a) listening (i.e., to a patient, to visible symptoms, to lab test results, to medical know how) and (b) taking action (i.e., formulating a diagnosis and a treatment plan). Now, it is important to emphasize the importance of the words that a patient says. Without those words (e.g., treatment of infants) the clinical evaluation of a subject lays additional difficulties in front of a medical doctor. Hence, listening, in its widest meaning, that embraces both any general (e.g., medical science) and specific knowledge (e.g., any chronic disease) of the subject under study, represents the initial step of any treatment, or repair, process.

Interestingly, we can find similar situations also in the realm of unanimated physical objects. For example, when electronic components were not as pervasive as today, mechanics typically diagnosed the *diseases* of motor vehicles by simply listening to the sounds that they emitted. Today such type of approach has been augmented, mechanics can still listen to the *symptoms* that their patients report, but can also perform lab tests connecting computers to the sensors that modern cars mount.

Clearly, not all objects are capable of telling their stories and not everyone is capable of interpreting them, when available. When a mechanical wristwatch stops, no particular sound provides a hint regarding the cause that made it stop. But also when an object emits sounds that indicate how well it works, like in the motor vehicle example, a casual person might not be able to distinguish those that are made when broken (e.g., a seized engine) from normal ones.

Now, in a contemporary scenario where the process of repairing, just as those of reusing, re-scoping and recycling, represents a fundamental component of a sustainable society, we here aim at discussing the possible approaches that could be taken when answering to the following two questions: (a) how can objects report the nature of their symptoms, and, (b) how can those reports aid anyone (i.e., also the unskilled) in their repair process?

In order to make a step in the direction of answering to both questions, let us now jump a long way forward and imagine a scenario where an object *contains* all the information required to accomplish its repair. We hence imagine that when an object breaks, a system is capable of indicating the type of repair that it needs, along with the tools and the gestures (i.e., practices) that the specific repair requires. Remaining open to imagination, this entails that a person could view a video showing how the repair could be accomplished, or wear a pair of gloves that physically guide through the performance of a set of mending gestures, or listen to a succession of detailed instructions that explain what actions should be taken. However, to permit to a system to provide instructions on how a given object can be repaired/treated the peculiar lifecycle of that object should be known, along with all the manipulations to which that objects was subjected, To this aim, in the remainder of this paper we try to demonstrate that fine gesture recognition algorithms can be used to capture the stories of unanimated objects that are crafted and then manipulated by human beings.

2. ANIMATING OBJECTS

Moving on to turning imagination into reality, the first technical problem that emerges amounts to how an object can contain any information at all. Interestingly, the authors of [1] recently addressed this problem proposing a system capable of integrating digital information into objects. In particular, they designed and implemented Spyn, a system based on mobile platforms that is capable of supporting the pinning of digital data on fabric. In brief, while knitting, a person can mark positions on fabric with special ink. This ink is used to create patterns that can be easily recognized by any simple visual recognition algorithm. These patterns are then utilized as virtual links between locations on fabric and digital data of any kind. A knitter can then *insert* a piece of information (e.g., an audio recording) directly from its smartphone by: (a) marking a position, and, (b) linking that information to the given position. Hence, simply pointing that smartphone's camera at that position it is then later possible to retrieve the information that was there stored.

Hence, taken for granted that some mechanism for pinning digital information on objects exists, the problem moves on to understanding what type of information may be useful in a repair process that an object should carry on. While a general answer to this question is far from being available, we here propose two non-alternative directions of work.

The first draws its inspiration from the Tortellini X-Perience system [2]. Tortellini X-Perience teaches how Tortellini pasta can be prepared implementing a set of steps where it: (a) shows on a video the actions that are required during each preparation phase, (b) requires that a person repeats those actions, and, (c) checks (utilizing a set of gesture-recognition algorithms) whether the person has repeated the given actions correctly (see Figure 1 and 2 above). In case the actions that a person performs are wrong, Tortellini X-Perience requires their repetition in order to move forward to the next step of the preparation recipe.

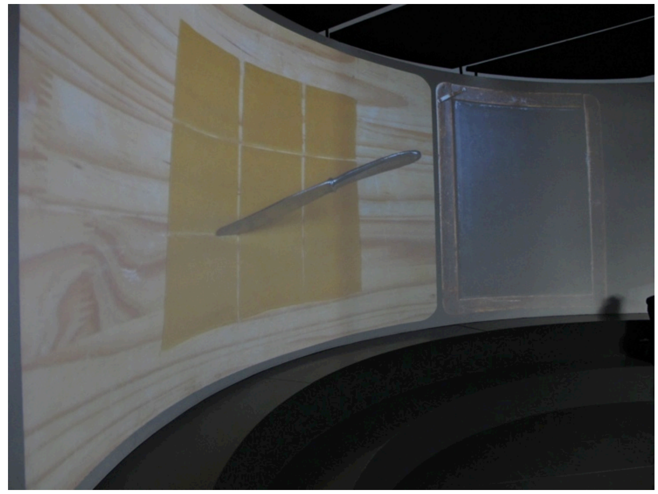

Figure 1. Tortellini X-Perience: recognizing when you cut the dough

Now, moving back to the repair context, we can envision a system (or a context) that digitalizes and records (in some given form, may it be a video or the codification of their trajectories or whatever) the actions that are performed while creating an object. But why should this information be useful? We here rely on conventional wisdom, considering that often a clear knowledge of how an object is created can also give an insight on how that object can be (or not) repaired.

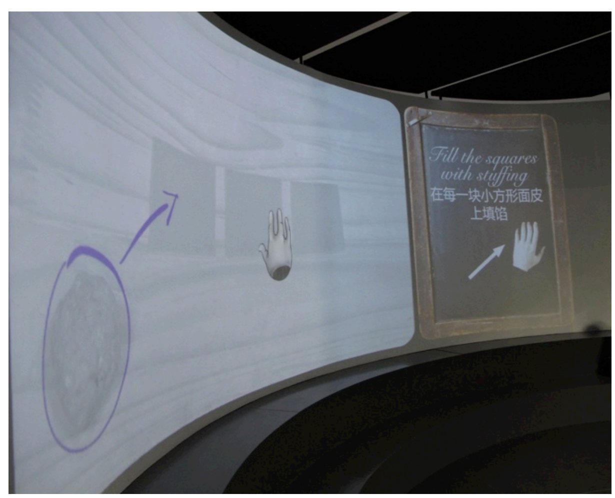

Figure 2. Tortellini X-Perience: explaining how you can fill in the pasta.

However, recognizing the fact that the knowledge of how an object has been created cannot always be useful or beneficial to its repair process, we also propose the following approach. Such approach origins from another piece of conventional wisdom: most of the objects that we use break in a very limited set of ways in very specific positions. Tears on clothing, for example, will typically occur along seams. Dishwashers, as well as other appliances, have common problems and common solutions [3]. Many producers of objects and appliances already acknowledge such fact when providing tutorials and frequently asked question sections written on manuals and sketched on parts. How could such existing initiatives be boosted with the aid of advanced computer-based technologies? Again, such initiatives can be supported, just to cite the simplest example, taking videos and codifying the actions required to accomplish given common repairs.

However, an additional step ahead can be taken in such direction. In fact, although breakages typically fall in given sets, each has its peculiarities and can require more or less care when performing certain repair actions. Some degree of customization might be very useful, hence, also when a common type of deterioration occurs. For this reason we envision repair support systems, which will rely on: (a) where an object broke, (b) how it was created, (c) how it is used or naturally decays, (d) how it is typically fixed at that point, and, (e) how a typical fix should be adapted to the peculiarities of a given deterioration.

Generalizing the approach that we have just described, we can hence classify any information related to an object in two groups: (a) preexisting and hence generic (e.g., the design of a sweater), and, (b) related to its own production and lifecycle and hence specific (e.g., how worn out is a given sweater). While any *generic* information can be easily coded according to the standard construction process of an object (i.e., some form of a storyboard that codifies the steps) and does not require any additional technological intervention to be acquired, the same cannot be said when thinking of the *specific* characteristics of an object.

Such type of information, in fact, is related to the peculiar lifecycle of an object and can only be acquired as the object is crafted, manipulated and operated along its life. Two sweaters, for example, may be exactly the same, but depending on how they have been worn, and on the decay process to which they have been subjected, they could present very different problems (i.e., tears and holes).

Modern vision technologies and fine gesture recognition and interpretation algorithms, like in [4, 5], clearly appear very promising when searching for mechanisms capable of fulfilling such type of task, but many open questions remain. The prominent one amounts to understand how such two pieces of knowledge can be stored, mixed together and put to good use to support repair operations. In conclusion, following this approach, the problem of repairing an object resembles to humans listening to very different stories told by different objects. Computer vision algorithms can be exploited for this task as discussed in the following Section.

3. COMPUTER VISION ALGORITHMS FOR UNDERSTANDING OBJECTS' STORIES: A PRELIMINARY ASSESSMENT

Fine gesture recognition and interpretation algorithms can be used to capture all the manipulations to which objects are subjected, starting from the time instant when they were crafted, along their entire lifecycle. The idea is that of understanding those actions that are performed on some given unanimated objects to capture their stories that can be used for repair.

This Section reports on a set of experimental results as drawn from a preliminary assessment activity conducted with some gesture recognition and interpretation algorithms we developed that only use a simple camera. It is not our intention to illustrate here such algorithms whose description can be found in [5]. Rather we report here the main results we achieved.

In Figure 3 the various phases are illustrated with a human that brings screws from a shelf to a workbench.

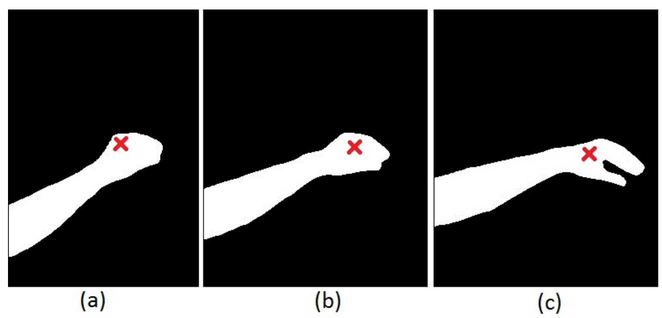

Figure 3. Taking screws from a shelf. (a) initial gesture, (b) intermediate gesture, (c) gesture during the intake

Results got with our algorithms to recognize those gestures as performed by a real person in action are depicted in Table 1.

Table 1. Bringing screws from a shelf to a workbench using one hand.

	50 cm	150 cm
Gesture following	96,2 %	94,5 %
Gesture recognition	94 %	92 %

The results related to "gesture following" represent the ability of our algorithms to precisely follow a moving hand, with a camera that is set at a distance of respectively 50 cm and 150 cm apart. The percentage of positive hits is shown in the Table with respect to the total amount of captured frames (24 fps).

With "gesture recognition" instead we intend here the fact that the trajectory that a moving hand draws is recognized as exactly that of a hand that takes a screw from a shelf and brings it to the workbench. This is obtained by comparing at given time instants the real position of a hand with the one where that hand should be if the hand were performing a given gesture. The total number of checkpoints are 24 per second.

Again the results are given in terms of positive hits with respect to the total amount of captured frames. The same action was repeated 20 times.

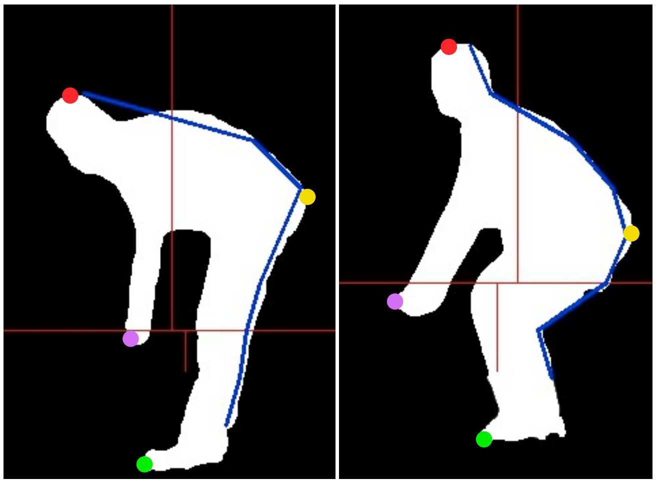

Figure 4. Lifting a box: recognizing the most important parts of a human body.

In Figure 4 the main positions are shown of a human body with a person that lifts a box from the floor to the workbench using both hands.

Table 2. Lifting a box from the ground to the workbench.

	Head	Hands	Feet	Waist
Accuracy	100 %	94,7 %	100 %	97,5 %
Recognition	100 %	85 %	100 %	87 %

Results got with our algorithms to recognize those postures as taken by a real person in action are depicted in Table 2.

The results related to "accuracy" represent the ability of our algorithms to precisely follow all the parts of a human body with a simple camera. The percentage of positive hits is shown in the Table with respect to the total amount of captured frames (24 fps).

With "recognition" we intend again the fact that the trajectory that a given part of a human body draws is recognized as the correct one. This is obtained by comparing at given time instants the real position of a part of a human body with the one where that part of a human body should be if a given gesture were performed. The number of checkpoints were 24 per second.

Again the results are given in terms of positive hits with respect to the total amount of captured frames. The same action was repeated 30 times.

In Figure 5 the main positions are illustrated of a human body with a person that tries to stretch a fabric.

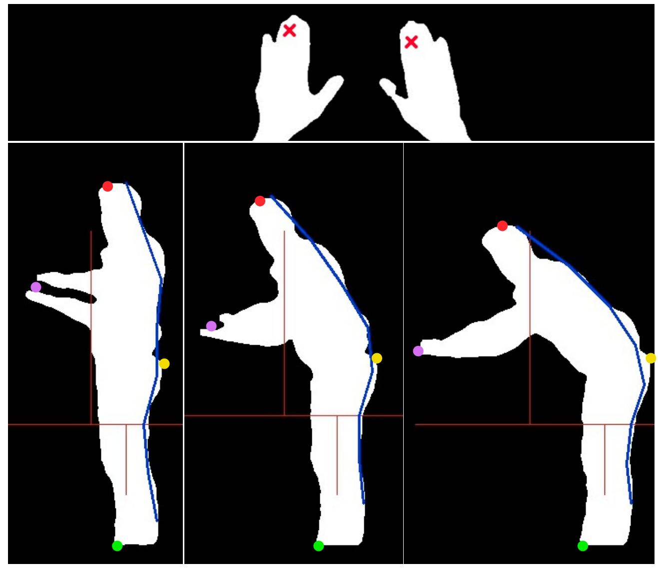

Figure 5 Stretching fabric. Top: hand detection. Bottom: body's postures.

Results got with our algorithms to recognize the gestures performed by the hands are shown in Table 3.

Table 3. Stretching fabric on the workbench using both hands.

	50 cm	150 cm
Gesture following	98,0 %	97,7 %
Gesture recognition	95 %	94 %

"Gesture following" means the ability of our algorithms to precisely follow both hands, with a camera that is set at a distance of respectively 50 cm and 150 cm apart. The percentage of positive hits is shown in the Table with respect to the total amount of captured frames (24 fps).

With "gesture recognition" we intend the fact that the trajectory that the hands draw is recognized as correct. This is obtained by comparing at given time instants the real position of the hands with those where the hands should be with a given gesture. The total number of checkpoints are 24 per second.

Again the results are given in terms of positive hits with respect to the total amount of captured frames. The same action was repeated 15 times.

The body's postures are analyzed in Table 4. The results related to "accuracy" represent the ability of our algorithms to precisely follow the postures of a human body with a camera. The percentage of positive hits is shown in the Table with respect to the total amount of captured frames (24 fps).

With "recognition" we intend again the fact that the trajectory that a given part of a human body draws is recognized as the correct one. This is obtained by comparing at given time instants the real position of a part of a human body with the one where that part of a human body should be if a given gesture were performed. The number of checkpoints were 24 per second.

Again the results are given in terms of positive hits with respect to the total amount of captured frames. The same action was repeated 15 times.

Table 4. Stretching fabric on the workbench: body's postures recognition.

	Head	Hands	Feet	Waist
Accuracy	100 %	100 %	100 %	91,9 %
Recognition	100 %	100 %	100 %	90 %

4. CONCLUSIONS

Just as a doctor listens to the symptoms reported by its patients, a person could listen to the deterioration symptoms reported by its objects, along with possible ways of repair. Listening and learning, in fact, are both part of do-it-yourself procedures, where one aims at fixing, rather than at replacing a malfunctioning appliance or tool. Under such perspective, we have here proposed an approach to two fundamental problems in the repair process: (a) how can objects be augmented with digital technologies so as to tell their stories to humans, and, (b) what an object should exactly say to humans to be repaired. As part of our proposal, we have also individuated the existing approaches that fall closest to the implementation of such ideas, for the development of future repair systems [6-12]. In this paper, fine gesture recognition algorithms have been used to capture all those manipulations to which objects were subjected, the idea being that of understanding the actions that are performed on them to capture those stories that can be then used for repair. Several preliminary experimental results have been discussed that confirm the efficacy of this approach.

5. REFERENCES

[1] Rosner, D.K., Ryokay, K., "Reflections on Craft: Probing the Creative Process of Everyday Knitters." Proceedings of the 7th ACM Conf. on Creativity and Cognition, Berkeley, pp. 195-204, 2009

[2] Roccetti, M., Marfia, G., Zanichelli, M., "The Art and Craft of Making the Tortellino: Playing with a Digital Gesture Recognize for Preparing Pasta Culinary Recipes." Computers in Entertainment, ACM, 8 (4), art. no. 28, 2010

[3] Available, online: http://apwagner.wordpress.com/2008/05/20/5-common-dishwasher-problems/, accessed on December 21 2012

[4] Marfia, G., Roccetti, M., Marcomini, A., Bertuccioli, C., Matteucci, G.,``Reframing Haute Couture Handcraftship: How to Preserve Artisans' Abilities with Gesture Recognition", Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 7624 LNCS, 437-444

[5] Roccetti, M., Marfia G., Semeraro, A, "Playing into the Wild: A Gesture-based Interface for Gaming in Public Spaces", *Journal of Visual Communication and Image Representation 23 (3*), 426-440

[6] Jackson S.J., "Rethinking Repair", in *Media Meets Technology*, eds. Pablo Boczkowski, Kirsten Foot, and Tarleton Gillespie (MIT Press: forthcoming, 2013).

[7] Ferretti S., Roccetti, M.,, "Fast Delivery of Game Events with an Optimistic Synchronization Mechanism in Massive Multiplayer Online Games", ACM International Conference Proceeding Series 265 , 405-412, 2005 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology, Valencia, 2005.

[8] Ferretti, S., Furini, M., Palazzi, C.E., Roccetti, M., Salomoni, P., "WWW recycling for a Better World", *Communications of the ACM 53 (4*), 139-143.

[9] Ferretti, S., Mirri, S., Roccetti, M., Salomoni, P., "Notes for a collaboration: On the design of a wiki-type educational video lecture annotation system", ICSC 2007 International Conference on Semantic Computing, art. no. 4338406, 651-656, Irvine, September 2007.

[10] Palazzi, C.E., Teodori, L., Roccetti, M., "Path 2.0: A participatory system for the generation of accessible routes", 2010 IEEE International Conference on Multimedia and Expo, ICME 2010 , art. no. 5583240, 1707-1711.

[11] Rosner, D. K., "The Material Practices of Collaboration", Proceedings CSCW 2012.

[12] Rosner, D. K., Taylor, A.S.,. "Binding and Aging", *Journal of Material Culture 17 (4*), 405-424.