Gesture Recognition Technologies for Gestural Know-how Management

Preservation and Transmission of Expert Gestures in Wheel Throwing Pottery

Alina Glushkova[1,2,3] and Sotiris Manitsaris[2,3,4]

[1]*Multimedia Technologies and Computer Graphics Lab., University of Macedonia, Thessaloniki, Greece*

[2]*Rural Space Lab., University of Thessaly, School of Engineering, Volos, Greece*

[3]*Robotics Lab, MINES ParisTech, PSL Research University, Paris, France*

[4]*Equipe Interaction Son, Musique, Mouvement, Institut de Recherche en Coordination Acoustique/Musique, Paris, France*

Keywords: Motion Capture, Gesture Recognition, Know-how Management, Expert Gesture, Sensorimotor Feedback.

Abstract: The acquisition of gestural know-how in manual professions constitutes a real challenge since it passes from master to learner, through a many years long « in person » transmission. However this binding transmission is not always possible for practical reasons; the learner must train himself alone, by using traditional Knowledge Management tools such as e-documentation and multimedia contents. These tools present important limitations, only providing the learner expert knowledge in a descriptive way, with a low attractiveness and interaction level, without any sensorimotor feedback. It thus becomes crucial to find novel ways to preserve and transmit know-how. In this work we present the idea of a methodological framework for gestural know-how management in wheel throwing pottery, based on motion capture and gesture recognition technologies. In combination with machine learning techniques, they permit to model the practical, cinematic aspects of potter's expertise. These technologies can be used to compare experts' and learners' simulated performances and to provide real-time feedback to the learner, guiding him in the adjustment of his gestures. The final goal is to propose a novel and highly interactive embodied pedagogical application for gestural know-how transmission, supporting « self » trainings, and making them more efficient.

1 INTRODUCTION

In actual context of globalisation and knowledge-based economy it becomes more and more important to manage knowledge efficiently. Providing tools to deliver the right information to the right person at the right moment in the most appropriate way becomes the subject of Knowledge Management (KM) discipline. But what happens when we want to expand this idea not only to knowledge in general, but also to know-how, to precise practical tasks and gestures? In this case we can talk about Know-How Management (KHM).

Issues linked to KHM have been studied by different scientific fields such as anthropology, ethnology and sociology. Their main goal was to identify the components of know-how and to propose the most efficient way for their transmission. Methods and tools have been thus proposed delivering know-how to the learner mostly through documents and multimedia courses. In this work we present a methodological framework for gestural know-how management based on motion capture and gesture recognition technologies. The methodology has been applied in wheel throwing pottery.

2 STATE OF THE ART

2.1 Traditional KM Tools Used for KHM

Traditionally, gestural know-how is transmitted "in person" from master to learner, physically present in the same place at the same time. To better understand "in person" transmission and to propose pedagogical material it has been studied from an ethnological and anthropological point of view (Chevallier, 1991). Expert technical gestures have been analysed; their parameters such as trajectory

Glushkova A. and Manitsaris S..
Gesture Recognition Technologies for Gestural Know-how Management - Preservation and Transmission of Expert Gestures in Wheel Throwing Pottery.
DOI: 10.5220/0005475904050410
In *Proceedings of the 7th International Conference on Computer Supported Education* (CSEDU-2015), pages 405-410
ISBN: 978-989-758-107-6

and acceleration have been defined (Bril, 2011). Although "in person" transmission is not always possible for practical reasons such as geographical distance between the master and the learner, low expert's availability/accessibility or other factors.

Based on technological advances of the last decade, ethnographists have started to use traditional KM digital tools to create pedagogical content. E-documentation has been enriched with videos, images, audio recordings, to support teaching and "self" trainings without master intervention. In China a digital archive has been created presenting a traditional method of weaving with a Bamboo (Wang et al, 2011). In more industrial context, in a plant of electric energy production, a video camera has been placed on the helmet of the expert worker to record his gestures. These videos have been used to propose training material (Le Bellu, 2010).

However e-documentation presents two important limitations: a) it provides limited information about expert gesture's execution reducing it into two dimensions and b) it is based on passive multimedia content (listening and watching) and e-courses (speaking and writing) while know-how learning is achieved through doing (Dale, 1969) and through interaction with the master. When using e-documentation the learner only receives multimedia messages and cannot interact with the pedagogical tool.

2.2 Gesture Recognition Technologies for KH Capturing

Gesture recognition technologies (GRT) can be used to overcome some of the limitations mentioned above. They permit to capture biomechanical aspects of a gesture and not only a two dimensional image of it, providing a data that can be analysed and modelled.

For example in artistic applications, a marker-based approach has been used to capture and analyse violin player's performance (Rasamimanana et al., 2009). However this technology is expensive and not robust to occlusions that can easily occur in other applications. For joints tracking and dancing movements recognition, low cost technology has been used, such as a depth camera (Raptis, 2011). But this marker-less technology cannot provide precise information about hand gestures and is also self and scene occlusion dependent. Contrariwise, wireless inertial sensors are occlusion independent and well adapted to record continuously hand gestures. They have been used for capturing, modeling and recognition of expert gestures in wheel throwing pottery in our previous study (Manitsaris et al, 2014). However in research works mentioned above the use of GRT is limited to capturing expert gestures for know-how preservation, while in the methodology described in this paper we will enter the phase coming after and will propose a methodology for know-how transmission.

2.3 Sensorimotor Feedback Guidance for KH Transmission

The use of GRT can also permit to overcome the second disadvantage of traditional KHM tools, proposing an interaction between the learner and the pedagogical application. According to Piaget's theory (Piaget, 1976) embodied intelligence is acquired through this interaction with the environment, through senses and experiences.

Some studies in fields like sports or art have been inspired from this statement. Taking inputs from motion capture, sonic feedback is provided to a speed skater to make him correct a regular error observed in his performance (Godbout and Boyd, 2010). In i-maestro project, violin player's movements are analysed and instructive optical feedback is given to help him to improve his techniques (Ng et al., 2007). However in most of existing studies where feedback is used in a pedagogical perspective, reference gestures are characterised by simple trajectories, or periodicity and the feedback is provided based on a simple tracking of body joints. In our approach we aim to use machine learning techniques to model more complex expert gestures, such as wheel throwing pottery ones that will serve as reference gesture and will be compared in real time with learner's gestures.

3 RESEARCH QUESTIONS

The main goal of this research is to propose a novel and highly interactive embodied pedagogical application for gestural know-how transmission, supporting "self" trainings, and making them more efficient. To achieve this goal and to provide scientific evidence about our statement this research has been structured around 3 research questions.

- Can cinematic aspects of expert technical gestures be captured, modelled and recognized by the machine?

If machine is able to recognise different gestures executed multiple times by the same potter and the recognition accuracy is high then the hypothesis can be validated.

- Can GRT be used to evaluate pottery learner's performance during "self" trainings?

After having captured learner's gestures performed during "self" training we can compare them with expert's gestures. Machine's ability to recognise learner's gestures using expert's models as reference will be used as indicator of the efficiency of "self" training.

- Is "self" training with sensorimotor feedback more efficient than without?

To answer this question we capture learner's gestures performed using our application providing real-time sensorimotor feedback. Then we still train the system with expert models and use for recognition the gestures captured. If recognition accuracy is higher here than in the previous hypothesis, it will mean that pottery learner's gestures performed with feedback are closer to expert gestures and that "self" trainings with sensorimotor feedback are more efficient.

4 METHODOLOGY

4.1 Capture, Modelling and Recognition of Expert Gestural KH

The first step of our methodology consists on analysing and modelling expert's gestural know how with the use of GRT: a) knowledge is extracted through collaboration with the expert and a gesture vocabulary is created; b) then cinematic aspects of gestures from this vocabulary are captured with a suit containing 11 inertial sensors covering expert's upper body and recording joints' rotations; c) after the definition of the appropriate gestural descriptors, and data normalisation we proceed to d) stochastic modelling of expert's gestures using a hybrid machine learning approach based on Hidden Markov Models (HMM) and Dynamic Time Warping (DTW) (Bevilacqua, 2010). A single sample is used to define a gesture class. HMMs calculate in real-time computation measures between the models and the incoming data and define the likelihood that the hidden model generated the incoming observation sequence. Then, we use the Jackknife cross validation method, and the precision and recall metrics to evaluate the ability of our system to recognise different executions of different gestures performed by the same potter. During this phase we also use basic statistical analysis of expert's motion data to define the variance between the repetitions of his gestures.

This system (ArtOrasis) and methodological substeps are presented in details in the paper "Capture, modeling and recognition of expert technical gestures in wheel-throwing art of pottery" (Manitsaris, 2014).

4.2 Expert/Learner Gestures Comparison

In order to quantify and understand the limits of "self" trainings while practicing wheel throwing pottery it is necessary to capture the gestures from the vocabulary but this time performed by the learner. Then with ArtOrasis application we can train the machine with expert models and use learner's dataset for recognition. At this phase our statement is that more learner's execution of the gestures are close to master's more his data is close to the states inside the Hidden Markov Model and the system will be able to provide an accurate estimation of recognition probabilities. Additionally, DTW is used to align temporally the hidden model and the observation sequence. When two sequences are warped this permits to calculate the distance between them and to compare the set of master gestures with learner's performance. More learner data is close to expert's more recognition accuracy is high. Precision and recall can be thus used as a metric to evaluate learner's performance.

4.3 Sensorimotor Feedback Mechanism

Once the limit of "self" trainings defined we can proceed to the creation of the pedagogical application providing sensorimotor feedback. The goal of these real-time optical or sonic indications is to alert the learner about his errors (implicit feedback) and to guide him in the adjustment of the gestures (explicit), to provide him a constructive evaluation. Our statement at this phase is that this interaction established between the learner and the machine can contribute to efficient gestural know-how learning and it can make "self" trainings more efficient.

To verify this statement it is necessary to propose the feedback mechanism. This must be inspired from the types of feedback master gives during the "in person" transmission and it strongly depends on the case study. Generally, the transmission procedure starts by showing the gestures to the learner. A video presentation could correspond to this step. However in our methodology we desire to go a step further and we include to our application a video annotated with colocalizations, also called in literature direct manipulations, i.e. superimpositions of expert's gestures and visual indications pointing out the most important cinematic aspects of the gesture. Its' goal is to introduce the learning material and to attract learner's attention at the most difficult points.

Once the learner starts practicing, master observes him and provides with constructive comments that could be divided in 3 categories, as we can see in the table 1. We consider that sensorimotor feedback provided by the machine should be inspired from this structure.

Table 1: 3 types of comments provided by the expert.

	Preventive	Corrective	Evaluative
Function	Warning inform	Indicate corrections	Provide a score

At this stage we concentrate our work at the feedback intervening first, the preventive one.We propose an optical implicit feedback, warning the user that an error is identified in his gesture. For this we visualise the distance (in Euler angles rotations), between learner's and expert's performance calculated during the time warping. We also name this application embodied since the apprentice uses directly his body,without any intermediary devices such as the mouse or joysticks, to interact with the system. A high level interaction is thus achieved between the learner and the application that adapts the feedback provided depending on learner's gestural performance.

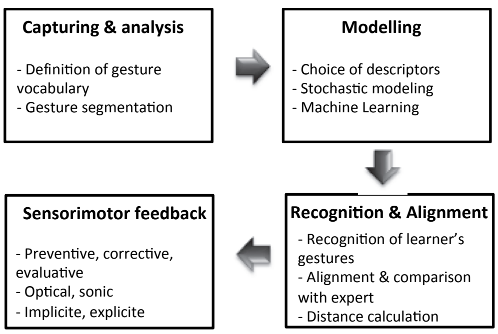

Figure 1: General overview of the methodology.

5 IMPLEMENTATION AND FIRST RESULTS

5.1 Potter's Gestural Kh Modelling

To answer to the first research question formulated in the section 3, we have conducted an experiment with the participation of 2 potters as described in the corresponding paper (Manitsaris et al., 2014). A gesture vocabulary with 4 or 6 gestures, used for the creation of a simple bowl (18/23cm diameter) has been created and 5 repetitions-subsets of each gesture have been captured. After that, raw data has been normalized and the appropriate descriptors have been selected.

When applying the jackknife method we use one of these repetitions of each gesture to train our GR system and the other repetitions for recognion sequence. All the data sets are once used for machine learning. In the table 2 we can see the very high real-time recognition accuracy for the 2 potters.

Table 2: Recognition accuracy rate of 2 expert potters.

	Precision	Recall
Potter A	100%	100%
Potter B	96%	97,5%

This machine's ability to recognise different expert's gestures constitutes a confirmation of the fact that cinematic aspects of gestures performed for bowl's creation, have been successfully modelled. It also means that a) the gestures (models) are different between them and it becomes evident from Levene's test results showing that the variances of experts' gestures are not equal ; b) the 5 repetitions of the same gesture are very similar and it can be concluded if we compare the angles distances on the 3 axis.

5.2 Comparison of Pottery Expert and Learner Gestures

In the second phase of the methodology, the goal is to evaluate learner's performance during « self » trainings, without receiving any feedback or guidance. To this end, the pottery learner of beginner level was asked to execute 5 times the same 4 gestures that the expert A showed him. It is important to notice that in wheel throwing pottery, when master teaches a beginner, the « in person » transmission often starts by virtual simulation of gestures. It helps the learner to memorise gestures trajectories before using the clay and the wheel.

Virtually performed gestures have been captured with the same 11 inertial sensors and a dataset of 20 gestures has been thus created. Then, we have selected one indicative expert data sequence and trained ArtOrasis system with it. However according to the statistical analysis done in (Volioti et al., 2014) all the 11 joints are not involved in pottery gestures in the same degree. Hands and head participate the most in the creation of a bowl. Based on this statement, we compare only wrists rotations data and use it for recognition. In the table 3 we present the jackknife results. At horizontal axis are indicated the 4 models and vertically the gestures used for recognition.

Table 3: Recognition accuracy rate of the learner.

	M1	M2	M3	M4	Recall
G1	**2**	0	1	2	40%
G2	1	**3**	1	0	60%
G3	0	0	**5**	0	100%
G4	2	2	1	**0**	0%
Precision	40%	60%	63%	0%	

From these results we can see that our system's ability to recognise pottery learner's gestures can be estimated at 50%, which is almost the half of expert's recognition accuracy. If we interpret these results from a semantic point of view they would mean that learner's performance declines from expert's by 50%. To compare more closely pottery learner's and expert's gestures we also use the DTW technique.

5.3 Learning Pottery with Colocalisations & Implicit Optical Feedback

To help the pottery learner reduce this distance we propose a pedagogical application, accompanying him in the learning process. Always inspired from "in person" transmission we start by providing the annotated video, reminding to the learner the most important cinematic of the gestures, such as body postures, gesture trajectories etc., as shown the figure 2.

Figure 2: Expert's video with colocalisations.

After the visualisation of the video the learner is invited to pass to the practical part with the use of sensors. For this, we have developed a simple user interface in MaxMSP environment, dynamically warning the user about his deviation, based on ArtOrasis system. More precisely, during the alignment of the model with the virtually performed, simulated gesture we calculate the normalized instant distance of rotation angles on 3 axes XYZ, between 2 sequences for 2 hands with the following equation.

$$k = \{x, y, z\} \qquad \delta = k_{learner} - k_{expert} \qquad (1)$$

Then, we visualise the absolute value of this distance in real-time. For this feedback we decided to ignore Z visualisation since wheel throwing pottery movements for bowl creation on this axis are limited. Learner's goal is to keep the distance lines as thin as possible. An interaction is thus installed between the user and the application.

This feedback is preventive and implicit since it's goal is only to warn about a deviation from expert's potter and not to give precise indications on how correct this deviation. At this stage we have opted for optical feedback because during virtual executions learner's vision can be used to receive information.

To active this implicit feedback we train the system with one indicative expert model of the first gesture, and we ask to the learner wearing the sensors to perform this gesture. To send the data flow from the sensors to our application in real-time we use the OSC protocol. At this stage HMMs are not mobilized since the system is trained only with the gesture the learner wants to practice. But DTW is aligning the 2 sequences and it permits us to calculate and to visualise the absolute value of learner's distance.

Figure 3: Implicit optical feedback - visualisation of angle deviations at X and Y axis for the left hand.

During this third experiment the learner in asked to perform each gesture with the use of our application 5 times and each repetition is captured. After that, we proceed to a jackknife where 4 expert models are used for the machine learning and 20 learner's repetitions are used for recognition.

Table 4: Recognition accuracy rate of the learner's gestures performed with feedback.

	M1	M2	M3	M4	Recall
G1	**5**	0	0	0	100%
G2	0	**5**	0	0	100%
G3	3	0	**2**	0	40%
G4	0	0	1	**1**	80%
Precision	63%	100%	67%	100%	

Then we perform Jackknife recognition tests, while still training the models with expert gestures and recognizing learner's performed with feedback. If we compare the results from the tables 3 and 4 we can see that the recognition accuracy and consequently machine's ability to recognise these pottery gestures have been improved, attending a precision and recall around 80%. We consider that it means that learner gestures performed with feedback are closer to expert gesture.

6 PERSPECTIVES

In this paper we present the idea of valorising GRT through an innovative KHM tool that could contribute to the efficient transmission of gestural know-how. The promising results presented in the section 5 constitute the first argument supporting the idea of this work. We can observe the tendency of improvement of pottery learner's gestures with the use of optical implicit feedback.

However to confirm the third hypothesis we need to conduct experiments with more than one user that will also subjectively evaluate the application through a questionnaire, and to test all the 3 types of sensorimotor feedback involving optical and sonic interaction. As underlined before, implicite optical feedback is effective to alert the learner about his errors but not to conduct him to their correction. Another important future research goal is to propose an efficient mechanism for corrective feedback activation based on a dynamically simulated statistical model.

ACKNOWLEDGEMENTS

The research project is implemented within the framework of the Action «Supporting Postdoctoral Researchers» of the Operational Program "Education and Lifelong Learning" (Action's Beneficiary: General Secretariat for Research and Technology), and is co-financed by the European Social Fund (ESF) and the Greek State.

REFERENCES

Bevilacqua, F., Zamborlin, B., Sypniewski, A., Schnell, N., Guédy, F. and Rasamimanana, N., 2010 'Continuous realtime gesture following and recognition', *LNAI 5934*, pp.73–84.

Bril, B., 2011. Description du geste technique: Quelles méthodes?. *Techniques & Culture*, (1), 243-244.

Chevallier, D., 1991, *Savoir faire et pouvoir transmettre. Transmission et apprentissage des savoir-faire et des techniques*. Les Editions de la MSH.

Dale E., *Audiovisual Methods in Teaching,* 1969, NY: Dryden Press.

Godbout, A., & Boyd, J. E., 2010, Corrective sonic feedback for speed skating: a case study. In Proceedings of the 16th international conference on auditory display, pp. 23-30.

Le Bellu S., Le Blanc B., 2010, How to Characterize Professional Gestures to Operate Tacit Know-How Transfer?, The Electronic Journal of Knowledge Management Volume 10 Issue 2, pp142-153.

Manitsaris, S., Glushkova, A., Bevilacqua, F., & Moutarde, F., 2014, Capture, modeling and recognition of expert technical gestures in wheel-throwing art of pottery. *ACM Journal on Computing and Cultural Heritage*.

Ng, K. C., Weyde, T., Larkin, O., Neubarth, K., Koerselman, T., & Ong, B., 2007, 3d augmented mirror: a multimodal interface for string instrument learning and teaching with gesture support. In *Proceedings of the 9th international conference on Multimodal interfaces*, pp. 339-345, ACM.

Piaget, J., 1976, *Piaget's theory*, pp. 11-23, Springer Berlin Heidelberg.

Raptis M., Kirovski D., Hoppe H., 2011, Real-Time Classification of Dance Gestures from Skeleton Animation. Eurographics/ In Proceedings of *ACM SIGGRAPH Symposium on Computer Animation*.

Rasamimanana N., Bevilacqua F., 2009, Effort-based analysis of bowing movements: evidence of anticipation effects. *The Journal of New Music Research*, 37(4): 339 – 351.

Volioti, C., Manitsaris, S., & Manitsaris, A., 2014, June,. Offline statistical analysis of gestural skills in pottery interaction. In Proceedings of the 2014 *International Workshop on Movement and Computing*, p. 172.

Wang, K. A., Liao, Y. C., Chu, W. W., Chiang, J. Y. W., Chen, Y. F., & Chan, P. C., 2011, Digitization and value-add application of bamboo weaving artifacts. In *Digital Libraries: For Cultural Heritage, Knowledge Dissemination, and Future Creation*, pp. 16-25. Springer Berlin Heidelberg.