A Robotized Projective Interface for Human-Robot Learning Scenarios

Davide De Tommaso, Sylvain Calinon, Darwin Caldwell

Department of Advanced Robotics, Istituto Italiano di Tecnologia, via Morego 30, 16163 Genova
Emails: davide.detommaso@iit.it   sylvain.calion@iit.it   darwin.caldwell@iit.it

Abstract: *In this work we discuss a novel robotics interface with perception and projection capabilities for facilitating the skill transfer process. The interface aims at allowing humans and robots to interact with each other in the same environment, with respect to visual feedback. During the learning process, the real workspace can be used as a graphical interface for helping the user to better understand what the robot has learned up to then, to display information about the task or to get feedback and guidance. Thus, the user can incrementally visualize and assess the learner's state and, at the same time, focus on the skill transfer without disrupting the continuity of the teaching interaction. We also propose a proof-of-concept, as a core element of the architecture, based on an experimental setting where a pico-projector and an rgb-depth sensor are mounted onto the end-effector of a 7-DOF robotic arm.*

Keywords: *Human-Robot Interaction, Learning from Demonstration, Augmented Reality.*

1. Introduction

While accuracy and speed have, for a long time, been at the top of the agenda for robot design and control, the development of new actuators and control architectures is now bringing a new focus on passive and active compliance, energy optimization, human-robot collaboration, easy-to-use interfaces and safety. The considerable growth of the number of service robots has brought machines closer to the human, involving aspects of daily life. The cooperation between robots and people without technical skills is becoming even more common in different fields

and applications. Therefore, the classic methods for interfacing with the robot do not satisfy the new requirements of the modern world in which the final user should not need to be an expert-programmer to use the interface. Instead of stand-alone programming, dynamic bidirectional models of interaction are required, in which the robot (learner) actively acquires the task demonstrated by the user (teacher).

Our recent studies have specifically addressed the issue of finding new user-friendly physical interfaces in order to reduce the complexity gap between humans and machines and to speed-up the skill transfer [16]. The aim of this paper is to discuss the novelty of the interface proposed in [16] and to present new developments in the experimental part.
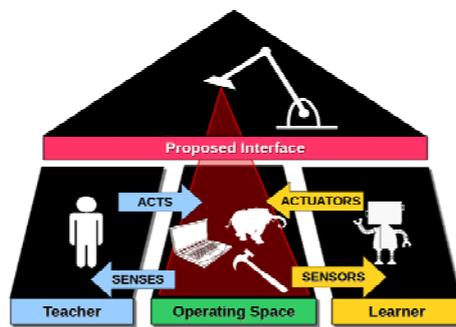

Fig. 1. Architecture overview of a typical learning scenario, in which the Robot/Learner and the User/Teacher interact with each other in the same Operating Space

The proposed interface is designed to provide a visual augmented operating space shared between the learner and the teacher, as shown in the scheme in Fig. 1. The aim is to share a common understanding of the task needed to be transferred, by using the operating space as a graphical interface where the task features are graphically superimposed in an augmented reality fashion. Consequently, the teacher can understand what the robot is learning, by observing the surrounding environment, and eventually refine or rectify on-line the skill whenever a robot mistake occurs. This adaptive learning process enables the user to always be aware of the learner's state and to continue the task without interrupting the teaching phase. In order to highlight the role of the proposed interface in human-robot learning tasks, three different illustrative scenarios have been presented in [16].

Our experimental setup jointly adopts an rgb-depth sensor and a pico-projector, both mounted onto the end-effector of a robot arm. Adopting such a mobile configuration, instead of a fixed setup, leads to a number of key advantages, such as: a) an extended field of view due to the different angle of views reachable by the robotic arm; b) the possibility to actively handling occlusions and facilitating tracking of task-relevant features and also c) the adaptive multi-resolution for perception and projection features.

In the paper we propose an experiment without focusing on a specific learning task. Here the aim is just to show the implementation of the core framework of the architecture. The experiment consists of a first phase in which the user physically interacts with the robotics interface for choosing the place where the projection will

appear. The robot is gravity compensated, providing a user friendly interface that can be easily moved by hand. In this phase the system adapts the perspective of a still image on the basis of the projector's viewpoint with respect to the destination surface. After, the robotic arm is able to autonomously superimpose visual information in the selected area and actively adapt to perturbations. The perspective and the size of the projection are kept constant, although the joints configuration of the robot can vary during the task.

2. Related work

2.1. Mixed reality in Human-Robot Interaction

Augment Reality (AR) enables users to see virtual graphical elements superimposed on real objects. The use of Augmented Reality using projectors to superimpose graphical elements directly on the task space has been introduced from Bimber and Raskar in [5] as Spatial Augmented Reality. The recent technology of hand-held projectors promises a rapid growth of applications, for enabling the user to interface with computers or robots.

In [6, 7] the authors use pico-projectors to visualize augmented digital information over real objects as HCI interfaces. A lightweight mobile camera projector unit is used in [6] to augment a paper map with additional information. This virtual map, the *Map Torchlight*, is tracked over the paper map and can precisely highlight points of interest, streets, and areas to give directions or other guidance information. In [7] a digital pen embedded with a spatially-aware miniature projector is used to explore the interaction design space of a paper document, providing the user with immediate access to additional information and computational tools. HRI interfaces with the help of hand-held projectors have also been studied, such as a robotic control interface for visualizing manipulation tasks [8], as an interface for controlling the robot without the direct manipulation by the user [9], or an alternative to the anthropomorphic interface using a projected display to interact with the user [10]. We also took inspiration from *LuminAR* [11], a project redefining the original concept of a desk lamp. Combining the technology of robotics and computer science, the authors use the light from a pico-projector, mounted on an articulated robotic arm, to show digital information to the user directly over the desk or any other surface. The joint use of a camera allows the user to interact with this virtual interface through hands motion, such as reading emails or navigating through a website.

Recently, V o g e l et al. [12] tackled the safety issue in Human-Robot collaboration task by using a projector-based solution. The authors propose a spatial augmented reality interface able to establish a physical safety area in a shared workspace between users and robots, by using a camera projector pair. The projective device gives feedback to the user about the safe working area, by projecting virtual barriers directly aligned with the real portion of space. The perception device helps the system to actively monitor the physical state of the user and the robot within the safety area, by changing position, shape and orientation of the projected image dynamically.

Another recent work about projective interfaces in the field of wearable computing is *OmniTouch*, in which H a r r i s o n et al. [13] suggest an innovative way to access digital information everywhere. *OmniTouch* is a wearable device that enables the user to interact with a GUI projected on any physical surface by using gestures. By exploiting the perception capabilities of the rgb-depth sensor, the system is able to detect suitable surfaces in which to project the GUI, by using a pico-projector, and to interact with fingers like a mouse pointer.

3. Motivation

For learning processes that require natural human interaction for transferring skills to robots, the design and development of interfaces between the teacher and the learner play a key role. In LfD strategies, expertise in robotics should not be expected of the final user. This makes it necessary to develop a shared communication protocol for transferring skills from humans to machines. Although several studies have investigated the social and technological aspects of the Human-Robot Interaction, many issues remain largely unexplored.

We emphasize in [15] the importance of providing an active role to the human-teacher in the learning process. The effectiveness and the generalization of the acquired skill do not depend only on the number of demonstrations but mostly on the pedagogical quality of these. The way to transfer a skill may be affected by the different nature of the learner and the teacher involved in the interaction and by several psychological factors related to the user during the teaching process. In the Human-Robot Interaction learning process, the way of giving several demonstrations of the task and the way of refining the learned skill by observing new reproduction attempts are often considered separately. We propose in [15] a learning paradigm to allow the user-teacher to incrementally see the results of demonstrations. This attempts to establish a two-ways interaction during the teaching process and to allow the user to feel involved in the task acquisition process.

3.1. A projective interface for learning scenarios

A critical issue in LfD is to visualize the skill that the robot has learned in the robot's environment, prior to executing it on the real robot for safety reasons. Virtual reality techniques and robot simulators have the drawback that the whole robot's environment, namely the objects involved in the interactions and the robot itself, need to be modelled. Accuracy errors of the synthesized model might introduce discrepancy between real and simulated movements. Physics and dynamics of the system also have to be taken into account to develop the simulator, which is sometimes difficult to develop. Even if powerful frameworks and tools are nowadays available to simulate robot environments, physical GUI user interactions are still required. Indeed, several operations such as: zoom-in/out, view point changes, removing occluding objects, are performed by the user through input devices (mouse, keyboard, touch-screen) causing interruptions during the teaching process.

Our work in LfD takes the perspective that the development of compliant actuators will bring gradual changes in the way skills and motions are represented by the algorithms in LfD. The machine learning tools that have been developed for precise reproduction of reference trajectories need to be re-thought and adapted to these new challenges. For planning, storing, controlling, predicting or re-using motion data, the encoding of a robot skill goes beyond its representation as a single reference trajectory that needs to be tracked or set of points that needs to be reached. Instead, other sources of information need to be considered, such as the local variation and correlation in the movement.

The system proposed could be used in this context to project trajectories or flow fields by using planar surfaces in the environment. In this way, the user can select the surface of interest and move around the robot to see different views while keeping his/her gaze towards the robot's workspace.

To make the process transparent, learning information about the task should be presented in an area of the workspace that is convenient for the user.
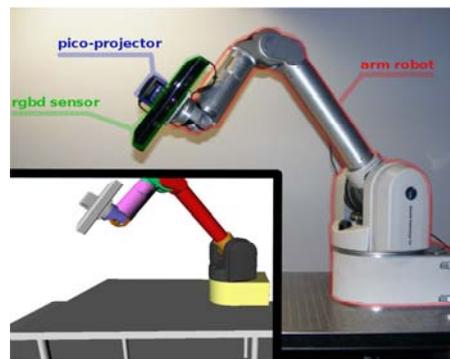

Fig. 2. The projection system using the compliant Barrett WAM 7-DOF
robot endowed with a Kinect device and a pico-projector

4. Experimental setup

Our experimental setup consists of a compliant Barrett WAM 7-DOF arm robot endowed with a plastic support mounted at the end-effector holding rigidly a Microsoft Kinect and an AXAA laser pico-projector (see Fig. 2). We consider such manipulator as an interface that can move, perceive and project in its environment. These features and its light weight well fit with the requirements of human-robot learning scenarios, in which the physical contact between the user and the manipulator represents an important modality of interaction. The Kinect has been extensively exploited in different fields of research as depth sensor, introducing an affordable option for point cloud tracking and detection [14]. For the projection capability, we selected a pico-projector because it is small enough to be mounted on top of the Barrett WAM and its laser technology allows us to project at any distance

without requiring to adjust focus. In the experiment the robot and the user share the same working space.
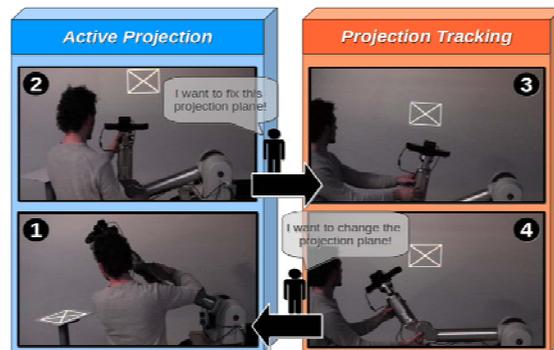
Fig. 3. Operating Modes of the system. The Active Projection enables the user to freely move the robot for finding suitable planar surfaces of projection, while the system projects undistorted images with fixed size and perspective. In the Projection Tracking mode the system acts for maintaining the projection in the selected plane of projection. Although the joints configuration of the robot can be modified by the user, the system continuously adapts the projection by maintaining fixed the geometric constraints previously selected

For selecting appropriate projecting surfaces, we decided to exploit both the control capabilities of the robot and the perception capability of a depth sensor. Instead of using structured-environments or tag-based surfaces, the system actively projects distortion-free images on the basis of the geometry of the planar surface of projection, detected by the Kinect. Accordingly, whenever the user wants to select the position and orientation of the projected display, she/he just needs to manually move the robotic arm in an appropriate position, while the robot compensates for the weight of its arm and friction in its joints. Once the projection surface is selected, its geometric features can be fixed in the robot's frame. Thus, for continuously tracking the projection, the system autonomously reacts to the changes of the robot arm configuration by 1) changing the orientation of the end-effector and 2) re-computing the perspective of the projected image.

5. Developed prototype

The system involves two mutually exclusive operating modes: 1) *Active Projection* and 2) *Projection Tracking* (see Fig. 3).

5.1. Active Projection

For allowing the user to select the projection plane, the system starts with the *Active Projection* mode. In this phase the process of warping the projection is carried out by using jointly the Kinect and the pico-projector, while the robot is only controlled by compensating for the gravity. This enables the user to easily change the joints configuration of the robot by looking for the pose of the end-effector allowing the projection in the desired plane. According to the end-effector pose, the source

image I_s, in Fig. 4-D, is warped in the image to project I_p in Fig. 4-C by using the perspective transformation

(1)
$$\underbrace{\begin{bmatrix} w\,p_x \\ w\,p_y \\ w \end{bmatrix}}_{I_p(x,y)} = \underbrace{\begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}}_{H_\pi} \underbrace{\begin{bmatrix} s_x \\ s_y \\ 1 \end{bmatrix}}_{I_s(x,y)}.$$

Since the 3D points of the projection surface lie on the same plane, the views of the Kinect and the pico-projector are related by an homography. For estimating the homography matrix two sets of four 2D points are required, four points in the source image I_s (e.g., the matrix elements (0,320), (240,640), (320,480), (320,0) see Fig. 4-D and four points $p_1^p, \cdots, p_4^p$ in the destination image I_p (see Fig. 4-C). Such points in I_p correspond to the 3D points $P_1^k, \cdots, P_4^k$ in the Kinect's frame, and can be found by changing the coordinate system from the Kinect to the pico-projector.
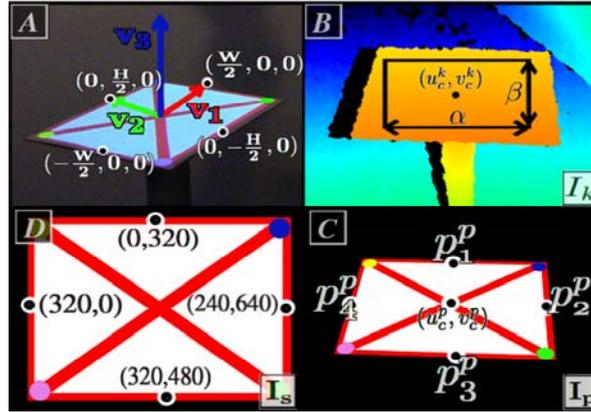

Fig. 4. (A) The frame of reference of the projection is defined by the orthogonal vectors v_1, v_2, v_3 whose origins correspond to the center of the projected image. (B) The Kinect's depthmap is used for estimating the equation of the plane chosen for the projection. Principal Component Analysis is applied to points samples from a rectangular area of I_k. (C) The input image of the pico-projector is the 640x480 RGB matrix I_p representing the result of the perspective transformation for fitting the image in the projection plane. (D) The 640×480 RGB matrix I_s represents the source image to project

The four corners of the projection are automatically detected by following the geometry of the planar surface and can be found by

(2)
$$\begin{bmatrix} P_1^k \\ P_2^k \\ P_3^k \\ P_4^k \end{bmatrix} = T \begin{bmatrix} 0 & -\frac{W}{2} & 0 & \frac{W}{2} \\ \frac{H}{2} & 0 & -\frac{H}{2} & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

(see Fig. 4-A). The 4×4 matrix T is the transformation between the Kinect and the projection frames (see Fig. 6), which can be written as

102

$$(3) \qquad T = \begin{bmatrix} R & C^k \\ 0 & 1 \end{bmatrix}, \quad R = [v_1\ v_2\ v_3]$$

where the equation of the projecting plane is

$$(4) \qquad \pi. \ r = r_0 + sv_1 + tv_2.$$

Fig. 5. Active Projection. The snapshots show how the user physically interacts with the robot for finding a suitable projection plane. The system aims to project a fixed size undistorted image following the geometric perspective of the projected surface

The geometry of the plane, namely the vectors v_1 and v_2 (see Fig. 4-A), can be estimated by *Principal Component Analysis* (PCA) applied to a set of 3D points in the Kinect's frame. As shown in Fig. 4-B, such point cloud is extracted from a rectangular area of the depthmap I_k, defined by the center $(u_c^{\ k}, v_c^{\ k})$, the height β and the width α. The 2D point $(u_c^{\ k}, v_c^{\ k})$ corresponds to the pico-projector's principal point $(u_c^{\ p}, v_c^{\ p})$ in I_p while C^k represents the projection of $(u_c^{\ k}, v_c^{\ k})$ in the Kinect's frame. The Active Projection mode is summarized in Algorithm 1, in which the procedure $AP()$ takes as input the source image I_s and the depthmap of the Kinect I_k and provides as output the image to project I_p.

Algorithm 1 Active Projection

```
1: function AP(I_k, I_s)
2:     T = paramsPlaneEstimation(u_c^k, v_c^k, α, β)
3:     P_1^k,..P_4^k = selectingFour3DPoints(T, W, H)
4:     p_1^P,..p_4^P = fromKinect2Projector(P_1^k,..P_4^k)
5:     H_π = homography(p_1^P,..p_4^P, (0, 320),..(320, 0))
6:     I_P = warpImage(I_s, H_π)
7:     return I_P
8: end function
```

5.2. Projection tracking

Once the user has selected a suitable projecting plane, the system switches from the *Active Projection* to the *Projection Tracking* mode. In such process, the robot is still maintained in gravity compensation. The *Projection Tracking* mode consists of an iterative process involving two main functionalities. The first computes the rotational forces to apply to the end-effector for pointing towards the center of the projection plane, which is represented as a 3D point fixed in the robot's frame. The robot actively reacts to perturbation of its joints configuration by keeping constant the size and the perspective of the projected image. The second functionality computes an updated perspective transformation, based on the actual end-effector pose, for enabling the resulting image to be undistorted despite the changes of the geometric parameters of the plane in the pico-projector's frame.

103

According to these two features, the system iteratively computes 1) the end-effector orientation to send to the robot $\boldsymbol{R}_{end}$ and 2) the warped projected image I_p. Four 3D points $P_1, \cdots, P_4$ in the robot's frame are obtained from points $P_1^k, \cdots, P_4^k$ in the Kinect's frame which defines the selected projection plane. In line 1 of Algorithm 2 the geometric transformation

$$\begin{bmatrix} P_1 \\ P_2 \\ P_3 \\ P_4 \end{bmatrix} = T_k \begin{bmatrix} P_1^k \\ P_2^k \\ P_3^k \\ P_4^k \end{bmatrix}$$

is computed by using the transformation matrix T_k between the Kinect and the robot frames (see Fig. 6).
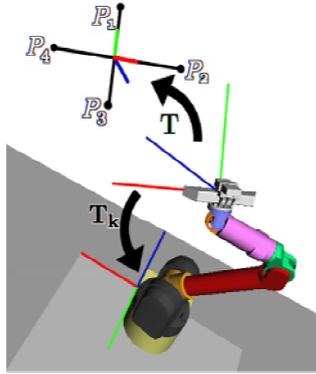

Fig. 6. A preliminary calibration process between the frames of reference of the robot, the pico-projector and the Kinect enables the system to find the corresponding transformation matrices

Therefore, the target towards which the end-effector has to point, namely the projection's center, is computed in line 2. Then, in the main loop the end-effector orientation and the warped image are continuously computed based on the actual robot's configuration. The *LookAt*() function, in line 3, provides the end-effector orientation matrix $\boldsymbol{R}_{end}$ for looking towards the point **C**, while the homography/warp operations on lines 7 and 8 enables the projected image to appear undistorted

---

Algorithm 2 Projection Tracking

1: $\mathbf{P_1}, ..\mathbf{P_4} = fromKinect2Robot(\mathbf{P_1^k}, ..\mathbf{P_4^k})$
2: $\mathbf{C} = findProjectionCenter(\mathbf{P_1}, ..\mathbf{P_4})$
3: **while** true **do**
4: $\quad \mathbf{R_{end}} = LookAt(\mathbf{C})$
5: $\quad sendOrientationToRobot(\mathbf{R_{end}})$
6: $\quad \mathbf{p_1^P}, ..\mathbf{p_4^P} = fromRobot2ProjectorImage(\mathbf{P_1}, ..\mathbf{P_4})$
7: $\quad \mathbf{H_\pi} = homography(\mathbf{p_1^P}, ..\mathbf{p_4^P}, (0, 320), ..(320, 0))$
8: $\quad \mathbf{I_p} = warpImage(\mathbf{I_s}, \mathbf{H\pi})$
9: **end while**

104
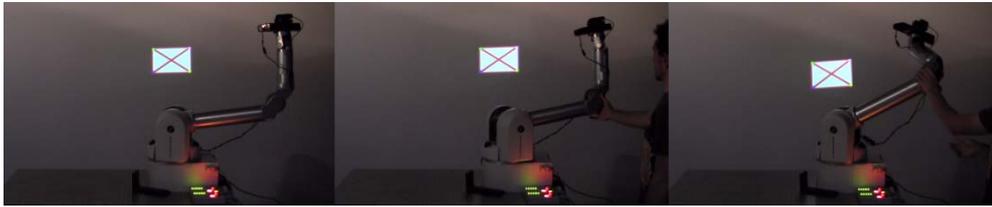
Fig. 7. Tracking Projection. The snapshots show how the system aims to track the size, the position and the perspective of the projection despite the changes of the joints configuration of the robot

6. Conclusion

We discussed the novelty of a robotic interface, originally designed in [16], for the assessment of the robot skills acquisition through active sensing and interactive data visualization. We focused on the importance of designing interfaces for non-expert users, who should not need to be skilled in robotics and computer programming to socially teach skills to robots. We also implemented a prototype to demonstrate the technical feasibility of the proposed interface, by combining a perceptive and a projection device mounted on an actively compliant robot manipulator. A human-robot collaboration scenario involving the task of finding suitable projecting surfaces and managing perturbations has been presented. We conducted an experiment showing that the user can manually interact with the gravity-compensated robot for finding a suitable end-effector pose which enables the projection to be superimposed on a desired surface. Therefore, once the position of the projection has been selected, we showed how the robot arm can actively adapt the projection when faced with changes of its joints conjuration. In the context of social bidirectional teaching interactions we believe that the proposed system can help the instructor for giving or receiving feedbacks in a natural and intuitive manner, helping him keep his focus of attention to the task without disrupting the teaching activity. The proposed experiment opens new research perspectives that will be part of our future work. Firstly, the proposed architecture needs to be tested in a realistic context, by identifying and collecting the parameters to measure the quality of the teaching process and analysing the resulting data in user studies. Such results can be compared to the performances obtained in the same scenario without the help of the proposed active interface. By taking insight from effective computing studies, psychological factors may be similarly taken into account to study the role of the device as a social actor in the teaching interaction.

R e f e r e n c e s

1. I s h i i, H., B. U l l m e r. Tangible Bits: Towards Seamless Interfaces Between People, Bits And Atoms. – In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI, ACM, New York, USA, 1997, 234-241.
2. D o u r i s h, P. Where the Action Is: The Foundations of Embodied Interaction. Cambridge, MA, USA, MIT Press, 2001.
3. K l e m m e r, S. R., B. H a r t m a n n, L. T a k a y a m a. How Bodies Matter: Five Themes for Interaction Design. – In: Proceedings of the 6th Conference on Designing Interactive Systems, DISACM, New York, USA, 2006, 140-149.

4. G i l l e t, A., M. S a n n e r, D. S t o¤e r, A. O l s o n. Tangible Augmented Interfaces for Structural Molecular Biology. – IEEE Comput. Graph. Appl., Vol. **25**, 2005, 13-17.

5. B i m b e r, O., R. R a s k a r. Spatial Augmented Reality: A Modern Approach to Augmented Reality. – In: Proceedings of Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH. ACM, New York, USA, 2005.

6. S c h ö n i n g, J., M. R o h s, S. K r a t z, M. L ö c h t e f e l d, A. K r ö u g e r. Map Torchlight: a Mobile Augmented Reality Camera Projector Unit. – In: Adjunct Proceedings of the 27th International Conference on Human Factors in Computing Systems, CHI EA, ACM, 2009, 3841-3846.

7. S o n g, H., T. G r o s s m a n, G. F i t z m a u r i c e, F. G u i m b r e t i e r e, A. K h a n, R. A t t a r, G. K u r t e n b a c h. Penlight: Combining a Mobile Projector and a Digital Pen for Dynamic Visual Overlay. – In: Proceedings of the 27th International Conference on Human Factors in Computing Systems, CHI, ACM, New York, USA, 2009, 143-152.

8. H o s o i, K., V. N. D a o, A. M o r i, M. S u g i m o t o. Visicon: A Robot Control Interface for Visualizing Manipulation Using a Handheld Projector. – In: Proceedings of the International Conference on Advances in Computer Entertainment Technology, ACM, New York, USA, 2007, 99-106.

9. K a z u k i, K., Y. S e i j i. Extending Commands Embedded in Actions for Human-Robot Cooperative Tasks. – International Journal of Social Robotics, Vol. **2**, 2010, No 2, 159-173.

10. P a r k, J., G. J. K i m. Robots with Projectors: An Alternative to Anthropomorphic Hri. – In: Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction, HRI, ACM, New York, USA, 2009, 221-222.

11. L i n d e r, N., P. M a e s. LuminAR: Portable Robotic Augmented Reality Interface Design and Prototype. – In: Adjunct Proceedings of the 23nd Annual ACM Symposium on User Interface Software and Technology, UIST, ACM, New York, USA, 2010, 395-396.

12. V o g e l, C., M. P o g g e n d o r f, C. W a l t e r, N. E l k m a n n. Towards Safe Physical Human-Robot Collaboration: A Projection-Based Safety System. – In: Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on, 25-30 September 2011, 3355-3360.

13. H a r r i s o n, C., H. B e n k o, A. D. W i l s o n. Omnitouch: Wearable Multitouch Interaction Everywhere. – In: Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology, UIST, ACM, New York, USA, 2011, 441-450.

14. R u s u, R. B., S. C o u s i n s. 3D is Here: Point Cloud Library PCL. – In: Proceedings of IEEE International Conference on Robotics and Automation (ICRA). Shanghai, China, 2011.

15. C a l i n o n, S., A. B i l l a r d. What is the Teacher's Role in Robot Programming by Demonstration? – Toward Benchmarks for Improved Learning. Interaction Studies. – Special Issue on Psychological Benchmarks in Human-Robot Interaction, Vol. **8**, 2007, No 3, 441-464.

16. D e T o m m a s o, D., S. C a l i n o n, D. C a l d w e l l. A Tangible Interface for Transferring Skills – In: International Journal of Social Robotics, Springer Netherlands, 2012, 1-12.