

Natural Interaction and Computer Graphics Applications

Samuel A. Iacolina¹, Alessandro Lai¹, Alessandro Soro^{1,2}, Riccardo Scateni¹

¹University of Cagliari, Dept. of Mathematics and Computer Science - Italy

²CRS4 - Center for Advanced Studies, Research and Development in Sardinia - Italy

Abstract

Natural Interaction with computers has been a challenging topic of research since the very beginning of the digital era and refers to the possibility, on the user's part, of exploiting natural abilities to control the machine and interpret its outputs. If in the infancy of computer graphics this meant using visual representation and pen pointing, nowadays more refined techniques are needed to fit the wide range of applications, from home entertainment to virtual and augmented reality. This paper describes some advances in gesture, tangible and surface computing, showing how such interaction models, if treated as a continuum, improve the usability, accessibility and overall experience of computer graphics applications.

Categories and Subject Descriptors (according to ACM CCS): I.3.6 [Computer Graphics]: Methodology and Techniques —Interaction techniques

1. Introduction

The origin of modern computer graphics is traditionally traced back to 1963, with the creation first interactive CAD system, the SketchPad by Ivan Sutherland [SE63]. This date represent a milestone also for human-computer interaction research, as it was the dawning of a *new area of man-machine communication*. Only a few years later the invention of the mouse by Douglas Engelbart [Eng70] would settle the appearance of our desks for the next 40 years. Grudin discusses the evolution of computers and interaction [Gru90], pointing out five ages, each of which is characterized by a different definition of 'interface': beginning with the hardware (such as mouse and keyboard) to current tangible and multi-touch interfaces, both based on cognitive peculiarities of the human mind. An analogous perspective is expressed by Dourish [Dou04], that remarks how the milestones of HCI correspond to the adoption of paradigms that allow people to exploit more *natural* skills in the interaction, such as, from earliest to more recent, linguistic abilities (programming languages), visual memory and spatial organization (GUIs), epistemic action (tangible user interfaces).

Parallel to such evolutions, since the mid-70s and up to the present time the focus of HCI research shifted from Interface design to interaction design, i.e. from a software perspective to a social/organizational one, driven by (or driving) the revolution in ICT spread, use and goals. This paper

presents the rationale and the technical issues arising from the design of an enhanced activity space meant to foster and study human-human interaction in a technology enriched environment. Natural interaction *in* the physical space and *with* the virtual space (documents) happens by means of gestures, manipulation and tangible artifacts. We discuss new insights on such interaction model(s) and trace a research roadmap in this field.

2. Natural Interaction

The concept of natural interaction has been variously defined and misused. In general it refers to the exploitation of natural (i.e. intuitive, familiar, innate, universal, cross-cultural, etc.) skills or abilities for controlling, either implicitly or explicitly, a computer system. Actually such a definition includes in practice any human activity and behavior, e.g. gestures, manipulation of physical objects, manipulation of virtual objects, facial expressions, head movements, body movements, body postures, natural (spoken) language, sign languages, use of real world metaphors. For the scope of this paper, natural interaction refers to the direct manipulation either of physical or virtual objects (such as with tangibles or multi-touch displays) accompanied by a narrow class of gestures for disambiguation and negotiation of the interaction space. The following sections will better define these topics under a broader perspective.

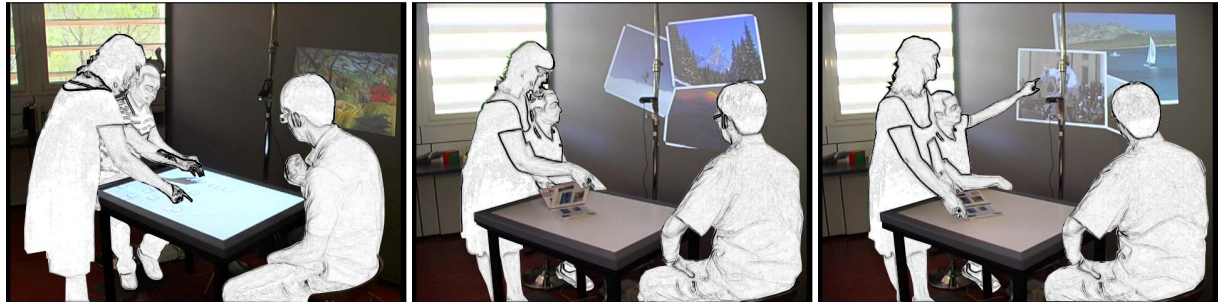


Figure 1: Envisioning collaborative activities in the interactive space. From left to right: a multitouch table game; tangible exploration of multimedia contents, multi-touch interactive videowall.

2.1. Gestures

The broad topic of gesture recognition and gestural interaction has been variously addressed by computer scientists (see [WH99] and [PSH97] for surveys), initially as part of multi-modal user interfaces (following seminal work by Bolt [Bol80, BH92]), and more recently, as a dominant aspect of, to name a few themes, tangible interaction [IU97], kinetic interaction [BH09], emotions recognition [CKC08]. However, there is still the need for a comprehensive understanding of what meaning and function gestures have in human communication, reasoning and behavior. The very problem of defining what a gesture is has been often tailored to the specific needs of applications (e.g. pen computing) or technologies (e.g. multi-touch displays). Additionally, gesture has been mostly regarded to as an alternative to input devices, that allows a more 'natural' form of interaction. This position is arguable by itself (as Wexelblat points out in [Wex98]), but, first and foremost, it is fitted on an interaction model (the personal computer, or the interaction as a private matter involving one person and one computer) that doesn't reflect the way people work, play, communicate today. Taking into account the cognitive or social role of gestures throws a new light on the 'natural' interaction: exploiting more and more natural abilities [Dou04], and fitting gently (unobtrusively) into human activities.

2.2. Manipulations

Though information technologies have faced an impressive evolution over the last decades, the way we access and explore multimedia contents (and digital information in general) has not changed much after the widespread adoption of graphical user interfaces. As Dourish points out [Dou04], not only the interaction paradigms, but the very appearance of computers seem to be *firmly stuck in an age of beige boxes*. Ironically, the typical metaphors of desktop computing, such as windows, icons etc. are making their way also to more personal and informal scenarios, such as home life, TV sets, and mobile phones. However new models of interaction based on manipulations are emerging. A key aspect

of manipulative interaction has been exposed by Kirsh and Maglio [MK96, Kir94]: they observed that skilled Tetris players tend to execute lots of fast and apparently useless moves on the bricks while playing. Their hypothesis (confirmed by many experiments) is that these moves are executed by the player in order to reach a more convenient cognitive status, rather than to directly achieve one's goal.

The advantages of such behavior are:

- the complexity of the task is moved from the head of the user to the world, available strategies and possible solutions to a given problem appear at a glance
- the (limited) resources of attention and memory are not wasted to concentrate on the strategy and can be used to explore alternative solutions
- such exploration performed by means of manipulations on the world (or tools) are easier (less cognitive effort) and faster (less time) than it is to do so mentally

Kirsh and Maglio also pointed out that epistemic action increases with skill: skilled computer users don't rely on the shorter (but mind consuming) strategy to solve a problem, even if the strategy is known, but rather execute more commands to evaluate their exact effect on the overall task, and eventually undo the last command when the result is not satisfactory. Execute a command, evaluate its result, and then in case rollback the command (epistemic action) is easier and faster for skilled computer users than representing in the mind the results and then issue the commands (pragmatic action).

2.3. Tangible Interaction

Tangible User Interfaces (TUIs) exploit physical objects to control the state of digital data: the physical control is coupled to its digital model to grant easy access to the model's functionalities. TUIs represent a growing and increasingly popular research area that encompass ergonomics, psychology and cognitive science, robotics and try to fill the gap among physical world and digital objects by letting the user

manipulate directly the latter. The peculiarities of TUIs are that:

- the interaction takes place in the physical space: instead of manipulating graphical entities that represent digital objects, the user manipulates objects themselves;
- the interaction and its effects on an artifact happen at the same time and in the same place;
- the interface encompasses the state of the model, i.e. the user interface is not meant to represent the state of the system, but rather the interface is the state of the system.

Early work of Fitzmaurice [FIB95, Fit96] on Graspable User Interface and then the work of Ishii on TUIs informed most research on Tangible Interaction. A key aspect of tangible interaction is that it allows people to actively explore and make sense of the world (either physical or digital).

2.4. multi-touch interaction

Pioneer work on multi-touch sensing devices can be tracked back to the mid-eighties, see for example [LBS85] [KGH85] [KN84]. An overview of the evolution of multi-touch technologies is maintained in [Bux07]. Given that multi-user interaction is a straightforward extension of multi-touch sensing, the obvious playground in this field consists in displays capable of accommodating a number of users, such as tabletop and wall-size displays. [DML02] and [Wil05] are examples of the former, [Wil04] and [DH05] of the latter. Several techniques have been exploited to implement multi-touch sensing devices, each one with strengths and constraints: [DML02] consists of an array of antennas whose signals get transmitted, through the body of the user, to a receiver that elaborates touch events. Among optical techniques [Wil04] exploits stereo cameras to compute hands position, but the cameras are located behind the semi-transparent screen, thus the system is bounded to front/rear projected display. The same holds in [Han05], which relies on an infrared camera that captures the light that escapes the display surface when finger contact occurs. In [OSK02] the optical sensor is located above the display surface, and thus the hands of the user(s) stay between the camera and the screen.

However, the recent explosion of (research) interest on multi-touch interaction follow the well-known work of Jeff Han on FTIR systems [Han05, Han06]. multi-touch interaction can in some respects be considered to sit at the borderline of gestures and manipulations: the user operates basic manipulations on virtual object, that can be transformed and organized just like common GUIs would allow, but at the cost of loosing the haptic feedback that would come from real tools (such as with tangibles). More expressive manipulations (e.g. selection of multiple objects) require conventional (and arbitrary) gestures, a simple sign language, to be executed touching the screen.

Anyway, it's worth remarking that more than its manipulative-oriented interaction paradigm, the key aspect

of multi-touch displays (or, more precisely, multi-touch tables and walls) is that it allows multi-user co-located interaction, i.e. more than one person can operate the device at the same time. This is a first step towards the *interface at the work setting* anticipated by Grudin [Gru90] back in 1990, though here we prefer to talk about *interaction* and not to restrict the application domain to collaborative *work* alone.

3. Collaborative Activity Space

As introduced in the previous sections, the new and rich field of computer supported collaborative activities is tightly bounded to the understanding (and technological support) of human gestures and manipulations. Though from a technological point of view such topics can be (and mostly *have*, so far) addressed independently, doing so from the interaction design perspective would be non-effective.

Our research is focused in designing and testing Collaborative Activity Spaces, i.e. areas in which either work, learning, playing or social activities are carried on with the support of technological artifacts. Our goal is to asses design practices and evaluation methods to understand:

- the benefits of the technological support to specific activities, for example, do games become more exciting/enjoyable/engaging if played with natural interaction in a technology rich space?
- the negotiation of the activity space among users: does people perform/feel better in collaborative/competitive tasks supported by the system described? Related works have shown that this is an open issue, and experiments with the use of such systems leads often to unexpected results (see for example [PKS*08])
- the appeal of the system, i.e. How people can be encouraged to engage in the use of the system at a first visit;
- accessibility, in the broad sense of universal access [ES05]: is it possible to design a system based on gestures/manipulations that overcome cultural/age/cognitive barriers? And if so, what gesture and manipulations are more suited to such purpose?

To this purpose a prototype activity space has been created, that comprises:

- a FTIR multi-touch table** whit several improvements to allow pre-contact feedback and robustness to changing lighting conditions [ISS10];
- interactive wall** , with multi-touch sensing based on bevel cameras [SPL08], the display is composed of a tile of commodity projectors [LSS10];
- surveillance cameras** to monitor the activity in the experimental area and to sense motions/gestures and use of fiducials for tangible interaction;

video-cameras, multi-touch table and the interactive wall are handled by means of a custom software framework, designed almost from scratch [MD10], that provides software



Figure 2: On the left the multi-projector video-wall (detail of the blending process). On the right the FTIR interactive table.

abstractions over physical sensors, abstract and concrete multi-touch widgets (supporting manipulation of images, web pages, movies, etc.), fiducials recognition and tracking, and a distributed asynchronous event subscription/delivery architecture.

The following sections describe in more depth some technical and research issues in the design and implementation of the single components.

3.1. Interactive Video Wall

Interactive video walls require almost invariably the adoption of display tiles, made either of LCD screens or projectors. Choosing the first solution, although possible in principle, one has to face high costs, logistic hassles, high power consumption and heat emission, all factors discouraging to adopt it. It's true that, apart from this, arranging an array of multi-touch displays is barely a problem of setting up a scaffold holding it, and connecting the array to an appropriate hardware that supports a display of that size. In the second case, when using projectors, the cost per surface unit is reduced, and the final result can be absolutely seamless, due to the absence of any type of frame inside or around the display; on the other hand, this seamlessness is obtained at the cost of facing and solving the problem of blending, in terms of geometry, colors and lightness, the images coming from each different projector.

The blending problem is already theoretically solved by many previous works as it is well summarized in [MB07] and practically implemented in many ways; these solutions mainly relies on hardware, using expensive projectors with in-hardware blending capabilities, or in software, typically bounded to the video architecture and to the specific application to be displayed, thus restraining the portability of the system. A typical example of this is the Chromium framework, targeted to OpenGL based applications [HHN*02].

This issue appears even more important in the development of multi-touch video-wall applications. As an example, coordinate transformation (from sensor space to GUI space) is affected by the blending functionality, and is better addressed if the blending is realized at the application level, rather than at the device level.

Projectors get calibrated one at a time: a black and white checkerboard sample is captured by a camera positioned just in front of the projection. The camera itself need to be calibrated to avoid lens distortion (this task is easily done using OpenCV [Int00]). Tilt and orientation with respect to the display surface must be known. In absence of any distortion from camera and projector lenses the image projected on the screen and the one captured by the camera would have identical proportions. OpenCV allows to precisely determine the position of the internal corners of a chessboard pattern, and we use it to compute the deformation matrix for the projector. The resulting (inverted) transformation is then applied just before the rendering phase. Different (partially overlapping) areas of the model are then rendered separately, and each one is deformed according to the appropriate matrix before being rendered to the screen. In this way we achieve geometrical consistency between projectors, using this alignment to compensate the space wasted by overlapping projection regions. Finally, a darkening mask is applied to obtain a luminance consistence.

3.2. Interactive Table

As known, a key technology for the design of multi-touch systems is Frustrated Total Internal Reflection (FTIR). Common FTIR setups have a transparent acrylic pane with a frame of LEDs around the side injecting infrared light. When the user touches the acrylic, the light escapes and is reflected at the finger's point of contact.

The infrared sensitive camera at the back of the pane can

clearly see these reflections. As the acrylic is transparent a projector can be located behind the surface (near to the camera) yielding a back-projected touch sensitive display. The software part consists in a basic set of computer vision algorithms applied to the camera image to determine the location of the contact point. An advantage of FTIR based sensors over competing solutions (such as DI, DSI [SBD*08]) is that this technology suffers less from ambient IR noise, and is thus more robust to changing lighting conditions.

On the other hand, it is well known that FTIR has some disadvantages:

- it does not sense finger proximity, the user must touch the surface;
- it is difficult to track the fingers during movements;
- though more robust to changes in ambient light, it still relies on a control over lighting conditions.

To partly address such issues we propose to take advantage of the shadows that the hands of the user project on the interaction surface. Our experiments show that such solution allows to effectively sense user interaction in an uncontrolled environment, and without the need of screening the sides of the multi-touch table.

Tracking infrared shadows to improve the quality of multi-touch interaction has been studied before. Echtler and co-workers [EHK08] describe a system to sense hovering on the surface, and thus provide pre-contact feedback in order to improve the precision of touch on the user's part. However the system they describe is based on a controlled IR lighting source above the table. In this sense their system exploits an additional artificial lighting source, increasing the dependence on the lighting conditions.

Our solution, as further described below, exploits natural uncontrolled light to improve the tracking algorithm. We take advantage of the natural IR noise to aid tracking, thus turning one of the main issues of MT sensors into a useful quality, making it possible to enhance tracking precision and implement pre-contact feedback. The hands of the user project a shadow on the surface (that will appear as a dark area in the noisy background). Such dark area is easily tracked because it is almost completely free of noise. Furthermore, fingertips correspond to the darker parts of the shadow, and can be recognized with good accuracy. Note that tracking the shadow is more and more effective as the ambient light increases (as opposite from IR blobs tracking), thus IR tracking and shadow tracking tend to complement each other, the former working better in full darkness, the latter in full daylight. Such complementarities are key aspects of our work: it allows the system to work in less controlled environments, and to be more robust to changing lighting condition, as may easily happen in real world, off-lab installations. This latter is, as known, one of the major issues for computer vision based interactive systems.

4. Conclusion and Further Work

We have shown the design and rationale of an interactive space for mixed gesture/manipulative interaction. Our plan for future work is to progress on two fronts:

1. identify the preferred combination of gesture and manipulations that people would use when describing an animation on a graphical model (e.g. a bird that flies) and identify gaps (missing functionalities) that prevent such expression from being used in HCI;
2. implement such functionalities in our system, by means of gesture sensing (e.g. HMM bases recognition) or geometrical algorithms (e.g. automatic skeletonization and segmentation).

Our goal thus is not to define or recognize an ad-hoc sign language (i.e. a given set of standardized gestures) but rather to automate a mapping between a generic pantomime or manipulation and a corresponding transformation on a 3D model, to allow people to use their natural gestural skills in describing a task to the system just like they would to a human.

References

- [BH92] BOLT R. A., HERRANZ E.: Two-handed gesture in multi-modal natural dialog. In *UIST '92: Proceedings of the 5th annual ACM symposium on User interface software and technology* (New York, NY, USA, 1992), ACM, pp. 7–14.
- [BH09] BRUEGGER P., HIRSBRUNNER B.: Kinetic user interface: Interaction through motion for pervasive computing systems. In *UAHCI '09: Proceedings of the 5th International on Conference Universal Access in Human-Computer Interaction. Part II* (Berlin, Heidelberg, 2009), Springer-Verlag, pp. 297–306.
- [Bol80] BOLT R. A.: "put-that-there": Voice and gesture at the graphics interface. In *SIGGRAPH '80: Proceedings of the 7th annual conference on Computer graphics and interactive techniques* (New York, NY, USA, 1980), ACM, pp. 262–270.
- [Bux07] BUXTON B.: Multi-touch systems that i have known and loved (overview). 2007.
- [CKC08] CASTELLANO G., KESSOUS L., CARIDAKIS G.: Emotion recognition through multiple modalities: Face, body gesture, speech. 92–103.
- [DH05] DEMPSKI K. L., HARVEY B.: Supporting collaborative touch interaction with high resolution wall displays. In *Proceedings of the 2nd Workshop on Multi-User and Ubiquitous User Interfaces (MU3I)* (2005).
- [DML02] DIETZ P. H., MA H., LEIGH D. L.: Multi-user touch surface, 2001/05/24 2002.
- [Dou04] DOURISH P.: *Where the Action Is: The Foundations of Embodied Interaction*, new edition ed. The MIT Press, September 2004.
- [EHK08] ECHTLER F., HUBER M., KLINKER G.: Shadow tracking on multi-touch tables. In *AVI '08: Proceedings of the working conference on Advanced visual interfaces* (New York, NY, USA, 2008), ACM, pp. 388–391.
- [Eng70] ENGELBART D.: X-y position indicator for a display system. Stanford Research Inst., Patent and Trademark Office - US patent 3,541,541, 1970.

- [ES05] EMILIANI P. L., STEPHANIDIS C.: Universal access to ambient intelligence environments: opportunities and challenges for people with disabilities. *IBM Syst. J.* 44, 3 (2005), 605–619.
- [FIB95] FITZMAURICE G. W., ISHII H., BUXTON W. A. S.: Bricks: laying the foundations for graspable user interfaces. In *CHI '95: Proceedings of the SIGCHI conference on Human factors in computing systems* (New York, NY, USA, 1995), ACM Press/Addison-Wesley Publishing Co., pp. 442–449.
- [Fit96] FITZMAURICE G. W.: *Graspable user interfaces*. PhD thesis, Toronto, Ont., Canada, Canada, 1996. Adviser-Buxton, William.
- [Gru90] GRUDIN J.: The computer reaches out: the historical continuity of interface design. In *CHI '90: Proceedings of the SIGCHI conference on Human factors in computing systems* (New York, NY, USA, 1990), ACM, pp. 261–268.
- [Han05] HAN J. Y.: Low-cost multi-touch sensing through frustrated total internal reflection. In *UIST '05: Proceedings of the 18th annual ACM symposium on User interface software and technology* (New York, NY, USA, 2005), ACM, pp. 115–118.
- [Han06] HAN J. Y.: Multi-touch interaction wall. In *SIGGRAPH '06: ACM SIGGRAPH 2006 Emerging technologies* (New York, NY, USA, 2006), ACM, p. 25.
- [HHN*02] HUMPHREYS G., HOUSTON M., NG R., FRANK R., AHERN S., KIRCHNER P. D., KLOSOWSKI J. T.: Chromium: a stream-processing framework for interactive rendering on clusters. *ACM Trans. Graph.* 21, 3 (2002), 693–702.
- [Int00] INTEL CORPORATION: *Open Source Computer Vision Library - Reference Manual*, 2000.
- [ISS10] IACOLINA S. A., SORO A., SCATENI R.: *Improving FTIR Based Multi-touch Sensors with IR Shadow Tracking*. Tech. rep., CRS4, 2010.
- [IU97] ISHII H., ULLMER B.: Tangible bits: towards seamless interfaces between people, bits and atoms. In *CHI '97: Proceedings of the SIGCHI conference on Human factors in computing systems* (New York, NY, USA, 1997), ACM, pp. 234–241.
- [KGH85] KRUEGER M. W., GIONFRIDDO T., HINRICHSEN K.: Videoplacement—an artificial reality. *SIGCHI Bull.* 16, 4 (1985), 35–40.
- [Kir94] KIRSH D.: On distinguishing epistemic from pragmatic action. *Cognitive Science* 18, 4 (December 1994), 513–549.
- [KN84] KASDAY L. R., NJ P.: Touch position sensitive surface, 1981/12/23 1984.
- [LBS85] LEE S., BUXTON W., SMITH K. C.: A multi-touch three dimensional touch-sensitive tablet. *SIGCHI Bull.* 16, 4 (1985), 21–25.
- [LSS10] LAI A., SORO A., SCATENI R.: Interactive calibration of a multi-projectors system in a video-wall multi-touch environment. In *UIST 2010 Adjunct Proceedings* (2010), pp. 437–438.
- [MB07] MAJUMDER A., BROWN M. S.: *Practical Multi-projector Display Design*. A. K. Peters, Ltd., Na-tick, MA, USA, 2007.
- [MD10] M DERIU A. SORO G. P.: *XPlaces: An Open Framework to Support the Digital Living at Home*. Tech. rep., CRS4 – (accepted at IEEE/ACM IOTS 2010), 2010.
- [MK96] MAGLIO P. P., KIRSH D.: Epistemic action increases with skill. In *In Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society* (1996), Erlbaum, pp. 391–396.
- [OSK02] OKA K., SATO Y., KOIKE H.: Real-time tracking of multiple fingertips and gesture recognition for augmented desk interface systems. In *FGR '02: Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition* (Washington, DC, USA, 2002), IEEE Computer Society, p. 429.
- [PKS*08] PELTONEN P., KURVINEN E., SALOVAARA A., JACUCCI G., ILMONEN T., EVANS J., OULASVIRTA A., SAARIKKO P.: It's mine, don't touch!: interactions at a large multi-touch display in a city centre. In *CHI '08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems* (New York, NY, USA, 2008), ACM, pp. 1285–1294.
- [PSH97] PAVLOVIC V. I., SHARMA R., HUANG T. S.: Visual interpretation of hand gestures for human-computer interaction: A review. *IEEE Trans. Pattern Anal. Mach. Intell.* 19, 7 (1997), 677–695.
- [SBD*08] SCHÖNING J., BRANDL P., DAIBER F., ECHTLER F., HILLIGES O., HOOK J., LÖCHTEFELD M., MOTAMEDI N., MULLER L., OLIVIER P., ROTH T., VON ZADOW U.: *Multi-Touch Surfaces: A Technical Guide*. Tech. Rep. TUM-I0833, University of Münster, 2008.
- [SE63] SUTHERLAND I. E., EDWARD I.: Sketchpad: A man-machine graphical communication system. *AFIPS Spring Joint Computer Conference* (May 1963).
- [SPL08] SORO A., PADDEU G., LOBINA M.: *Multitouch Sensing for Collaborative Interactive Walls*, vol. 272. Springer Boston, 2008, pp. 207–212.
- [Wex98] WEXELBLAT A.: Research challenges in gesture: Open issues and unsolved problems. In *Proceedings of the International Gesture Workshop on Gesture and Sign Language in Human-Computer Interaction* (London, UK, 1998), Springer-Verlag, pp. 1–11.
- [WH99] WU Y., HUANG T. S.: Vision-based gesture recognition: A review. In *GW '99: Proceedings of the International Gesture Workshop on Gesture-Based Communication in Human-Computer Interaction* (London, UK, 1999), Springer-Verlag, pp. 103–115.
- [Wil04] WILSON A. D.: Touchlight: an imaging touch screen and display for gesture-based interaction. In *ICMI '04: Proceedings of the 6th international conference on Multimodal interfaces* (New York, NY, USA, 2004), ACM, pp. 69–76.
- [Wil05] WILSON A. D.: Playanywhere: a compact interactive tabletop projection-vision system. In *UIST '05: Proceedings of the 18th annual ACM symposium on User interface software and technology* (New York, NY, USA, 2005), ACM, pp. 83–92.