THE INHERENT CONTEXT AWARENESS OF NATURAL USER INTERFACES: A CASE STUDY ON MULTITOUCH DISPLAYS

Bojan Blažica

Investing in your future
OPERATION PART FINANCED BY THE EUROPEAN UNION
European Social Fund

Doctoral Dissertation
Jožef Stefan International Postgraduate School
Ljubljana, Slovenia, May 2013

Evaluation Board:
Prof. Dr. Franc Novak, Chairman
Jožef Stefan International Postgraduate School, Ljubljana, Slovenia
Prof. Dr. Roman Trobec, Member
Jožef Stefan International Postgraduate School, Ljubljana, Slovenia
Asst. Prof. Dr. Boštjan Murovec, Member
University of Ljubljana Faculty of Electrical Engineering, Ljubljana, Slovenia

MEDNARODNA PODIPLOMSKA ŠOLA JOŽEFA STEFANA
JOŽEF STEFAN INTERNATIONAL POSTGRADUATE SCHOOL

Bojan Blažica

THE INHERENT CONTEXT AWARENESS OF NATURAL USER INTERFACES: A CASE STUDY ON MULTITOUCH DISPLAYS

Doctoral Dissertation

INHERENTNA KONTEKSTNA OZAVEŠČENOST NARAVNIH UPORABNIŠKIH VMESNIKOV: ŠTUDIJA PRIMERA VEČDOTIČNIH ZASLONOV

Doktorska disertacija

Supervisor: Prof. Dr. Dunja Mladenić
Jožef Stefan International Postgraduate School, Ljubljana, Slovenia
Co-supervisor: Dr. Daniel Vladušič
XLAB Research, Ljubljana, Slovenia

Ljubljana, Slovenia, May 2013

To all those who, intentionally or by chance,
taught me what I know and shaped who I am

Contents

Abstract

In computer science, context-awareness refers to the capability of a computing device to sense, understand and react to contextual information, i.e. information that is not at the centre of an activity but is still relevant for that activity. A computing device does not necessarily interact with humans at a given moment, but when it does, its Context-Awareness has many implications for human-computer interaction. In this thesis we look at Natural User Interfaces from a Context-Awareness perspective. On one hand, we show that considering natural user interfaces as context-aware systems further increases the expressive power of these interfaces and, on the other hand, we show that natural user interfaces can also represent essential building blocks for context-aware systems and are therefore a viable way towards context-awareness. Research prospects arising from this perspective, which this thesis explores, are: to what extent are natural user interfaces already inherently context-aware, how to increase the expressiveness of natural user interfaces through context-awareness, do natural user interfaces provide enough information to perform biometric user identification, and how to take advantage of information implicitly conveyed by the user during interaction with natural user interfaces. The specific natural user interfaces used in this thesis are multitouch displays.

First, the fields of natural user interfaces and context-awareness are reviewed. Regarding natural user interfaces, as this is an emerging research field, special care is taken to survey all currently available definitions of the term. Similarly, multitouch displays and multitouch interaction are described in more detail as they are considered in the case study for this thesis. Other related fields such as ubiquitous/pervasive computing, ambient intelligence etc. are also briefly explained. The presented overview does not merely introduce the topic of the thesis, but also shows how interconnected these fields are and how natural user interfaces are indeed inherently context-aware.

We have shown how the increased amount and variety of data from natural user interfaces can be exploited to acquire contextual information by developing a biometric user identification method and a clustering algorithm for hand detection, both for multitouch displays. The method for user identification, named MTi, is based on features obtained only from the coordinates of the 5 touchpoints of one of the user's hands. This makes it applicable to (almost) all multitouch displays without requiring additional hardware and regardless of the display's underlying sensing technology. The method was tested on a dataset of 34 users and reported 94.69 % identification accuracy. The method also proved to scale well and has an above-average usability. Next, we address the problem of hand detection, i.e. detecting how many hands are currently on the surface and associating each touch point to its corresponding hand. The presented solution – a clustering algorithm with simple heuristics based on the anatomy of the human hand – is software-based and thus again applicable to all multitouch surfaces regardless of their construction. Along with these two, other related methods that increase the expressiveness of multitouch displays are surveyed. Finally, this thesis explores the possibility to use implicit human-computer interaction to aid personal photo collection management. The idea is that the way we interact with natural user interfaces can implicitly disclose additional (contextual) information, which helps a context-aware system to better understand the user. More specifically, we take into account the user's personal relationship with a single photo; whether the photo is of particular importance to the user. We call this personal relationship the user's *affinity* for a photo. Experiments revealed that affinity is correlated with the time a user spends viewing a picture. Furthermore, by looking at viewing times, it is also possible to distinguish the task a user is currently performing.

The positive examples of context acquisition on multitouch displays presented confirm that natural user interfaces are inherently context-aware and show how their expressive power can be further increased by viewing them from a context-aware perspective.

Povzetek

V računalništvu se pojem kontekstna ozaveščenost nanaša na sposobnost računalniškega sistema, da zazna, razume in se odzove na informacije, ki izvirajo iz konteksta, v katerem se nahaja in deluje. Imenujemo jih kontekstne informacije in jih definiramo kot tiste informacije, ki sicer niso v centru neke aktivnosti, a so zanjo še vedno pomembne. V primeru, da računalniški sistem interagira s človekom, ima lahko kontekstna ozaveščenost sistema velik vpliv na samo komunikacijo človek-računalnik. Ta disertacija z vidika kontekstne ozaveščenosti obravnava naravne uporabniške vmesnike. V njej pokažemo, da z obravnavanjem naravnih uporabniških vmesnikov kot kontekstno ozaveščenih sistemov po eni strani povečamo njihovo že tako veliko izrazno moč, po drugi strani pa lahko naravni uporabniški vmesniki predstavljajo osnovne gradnike kompleksnejših kontekstno ozaveščenih sistemov. Naravni uporabniški vmesniki so torej prava pot za doseganje kontekstne ozaveščenosti računalniških sistemov. Iz tega pogleda izhajajo naslednja raziskovalna vprašanja, obravnavana v tej disertaciji: v kolikšni meri so naravni uporabniški vmesniki že sami po sebi kontekstno ozaveščeni, kako dodatno povečati izrazno moč naravnih uporabniških vmesnikov s pomočjo kontekstne ozaveščenosti, ali naravni uporabniški vmesniki nudijo dovolj informacij za biometrično razpoznavanje uporabnikov in kako izkoristiti formacije, ki jih uporabnik med interakcijo implicitno poda računalniškemu sistemu. Naravni uporabniški vmesniki, ki jih bomo uporabljali v tej disertaciji, so večdotični zasloni.

Disertacija najprej poda pregled področij naravnih uporabniških vmesnikov in kontekstne ozaveščenosti. Ker je predvsem prvo področje še v nastajanju, damo poseben poudarek pregledu različnih definicij pojma naravnih uporabniških vmesnikov. Podrobneje so opisani tudi večdotični zasloni in večdotična interakcija, ki jih uporabimo pri praktičnem prikazu. Poleg tega disertacija na kratko povzema naravnim uporabniškim vmesnikom in kontekstni ozaveščenosti sorodna področja, kot so vseprisotno računalništvo in ambientalna inteligenca. Pregled področij, poleg predstavitve teme disertacije, prikaže tudi tesno povezanost teh področij in potrjuje trditev, da so naravni uporabniški vmesniki že sami po sebi kontekstno ozaveščeni.

Z razvojem metode za biometrično razpoznavanje uporabnikov in algoritmom za zaznavanje rok, (oboje na večdotičnih zaslonih) smo pokazali, kako lahko količino in raznolikost podatkov, ki nam jih naravni uporabniški vmesniki ponujajo, izkoristimo za dodatno povečanje izrazne moči teh vmesnikov. Metoda za razpoznavanje uporabnikov, imenovana MTi, temelji na značilkah, ki jih izračunamo zgolj na podlagi koordinat dotikov 5 prstov ene izmed uporabnikovih rok. To pomeni, da je metoda splošno uporabna na (skoraj) vseh večdotičnih zaslonih, ne glede na njihovo konstrukcijo in brez dodatne strojne opreme. Natančnost metode pri bazi s 34 uporabniki je 94.69 % in je sorazmerno neobčutljiva na večanje baze. Metoda se je ob tem izkazala tudi kot nadpovprečno prijazna do uporabnikov.

Kot naslednjega raziščemo problem zaznavanja rok na večdotičnih zaslonih. Ta se ukvarja z ugotavljanjem, koliko rok je na zaslonu in kateri roki pripadajo posamezni prsti. Predstavljena rešitev – algoritem za rojenje na osnovi hevristik, ki izvirajo iz anatomije človeške roke – je povsem programskega značaja in, ker ne zahteva dodatnih strojnih delov, uporabna na vseh večdotičnih zaslonih.

Poleg omenjenih metod v disertaciji raziščemo še druge načine, ki prav tako povečajo izrazno moč večdotičnih zaslonov. Nazadnje obravnavamo še možnost uporabe implicitne komunikacije človek-računalnik za pomoč pri urejanju osebnih zbirk slik. Osnovna ideja je, da lahko način, kako interagiramo z računalniškim sistemom, implicitno razkrije dodatne (kontekstne) informacije, s pomočjo katerih lahko sistem lažje razume uporabnika in njegove namene. Glede na to stališče smo obravnavali uporabnikov oseben odnos do posamezne slike: ali ima slika za uporabnika poseben pomen ali ne? Ta oseben odnos smo imenovali uporabnikova *afiniteta* do slike. Eksperimenti so pokazali, da je afiniteta korelirana s časom, ki ga uporabnik porabi za ogled slike. Poleg tega je iz časa ogleda mogoče določiti uporabnikovo trenutno opravilo (iskanje določene slike, ogledovanje zbirke slik, ali priprava izbora slik).

Predstavljeni uspešni primeri pridobivanja kontekstnih informacij na večdotičnih zaslonih potrjujejo, da so naravni uporabniški vmesniki sami po sebi, inherentno, kontekstno ozaveščeni in da se izrazna moč teh vmesnikov še poveča, če nanje gledamo s stališča kontekstno ozaveščenih sistemov.

Abbreviations

ANN	=	artificial neural network
CLI	=	command line interface
CNUI	=	continuous natural user interface
CVS	=	comma separated value
GUI	=	graphical user interface
HCI	=	human-computer interaction
IUI	=	intelligent user interface
MT	=	multitouch
NUI	=	natural user interface
OUI	=	organic user interface
PUI	=	perceptual user interface
SUS	=	system usability scale
SVM	=	support vector machine
TUI	=	tangible user interface
UBICOMP	=	ubiquitous computing
WIMP	=	windows icon menu pointer

1 Introduction

This chapter introduces the main concepts of the thesis - context-awareness and natural user interfaces. Multitouch displays and use cases for multitouch interaction are discussed in more detail. Related topics, such as intelligent user interfaces, ubiquitous computing, pervasive computing and others are also briefly explained and the relationship between them discussed. At the end of the chapter, the hypotheses and goals of the thesis are presented and the main scientific contributions listed.

1.1 Background

"Computing machines can do readily, well, and rapidly many things that are difficult or impossible for men and men can do readily and well, though not rapidly, many things that are difficult or impossible for computers. This suggests that a symbiotic cooperation, if successful in integrating the positive characteristics of men and computers, would be of great value" (Licklider, 1960). In 1960, Licklider identified the problems for realization of such 'man-computer symbiosis' in speed mismatch between men and computer, memory hardware requirements, memory organization requirements, differences between human and computer language and input and output equipment, the latest being the most problematic. Fifty years later this still holds true.

Today Human-Computer Interaction (HCI) is concerned with the above-mentioned problems i.e. the design, evaluation and implementation of interactive computing systems for human use and the study of major phenomena surrounding them (Hewett et al., 2009). Figure 1.1 shows a schematic overview of the topics covered in HCI. It is clear that HCI is an interdisciplinary area deriving its knowledge from computer science (application design and engineering of human interfaces), psychology (the application of theories of cognitive processes and the empirical analysis of user behaviour), sociology and anthropology (interactions between technology, work, and organization), and industrial design (interactive products).

Interaction between human and computers takes place at the user interface. This is why research and development of user interfaces lies at the very core of HCI. We will first briefly present the development that led to today's user interfaces and then focus on multitouch displays, their construction and use cases.

1.1.1 From Punch Cards to Natural User Interfaces

Early digital computers used batch interfaces. They consist of punch cards for programs and data input and prints as output. Batch interfaces are non-interactive, which means that the user has to specify everything prior to processing and is never prompted for additional input until the job is done. With the connection of teletype machines to computers and the advent of dedicated text-based CRT terminals in the late 50s, batch interfaces gave space to command line interfaces (CLI). CLIs brought speed and interactivity to human computer interaction. In the early 60s, Douglas Engelbart invented the first mouse prototype used for

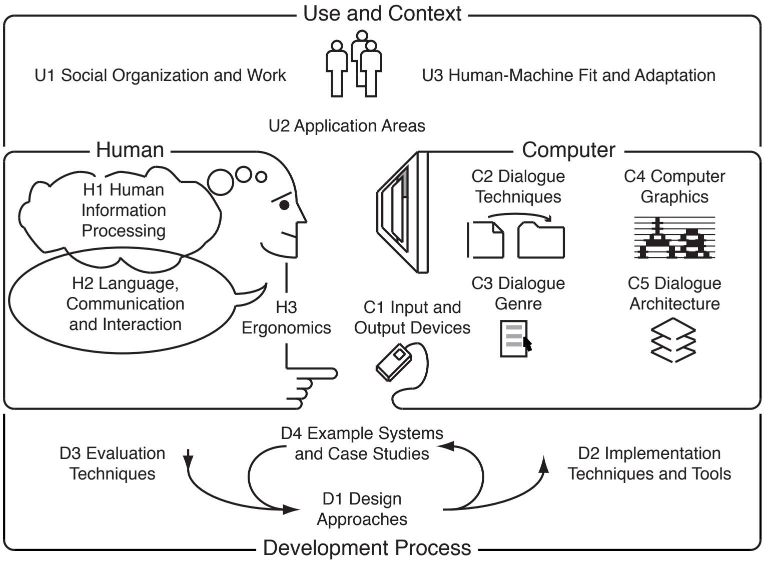

Use and Context
U1 Social Organization and Work
U3 Human-Machine Fit and Adaptation
U2 Application Areas
Human
H1 Human Information Processing
H2 Language, Communication and Interaction
H3 Ergonomics
Computer
C1 Input and Output Devices
C2 Dialogue Techniques
C3 Dialogue Genre
C4 Computer Graphics
C5 Dialogue Architecture
D4 Example Systems and Case Studies
D3 Evaluation Techniques
D1 Design Approaches
D2 Implementation Techniques and Tools
Development Process

Figure 1.1: A schematic overview of Human-Computer Interaction (Hewett et al., 2009).

manipulating text-based hyper-links as a part of his 'augmenting human intellect' project (Engelbart, 1962). Later, researchers at Xerox PARC extended the concept of hyper-links to graphics and by doing so created the first graphical user interface (GUI). In respect to the CLI, GUI facilitates a more intuitive interaction because there is no need to learn commands by heart as available commands are presented on the screen in the form of windows, icons, menus and a pointer. This interaction paradigm can be emphasized by the acronym WIMP. Further efforts focused on the development of interfaces that interact with humans in a seamless way by understanding natural expressions of the user's intent instead of forcing him/her to learn new rules for interaction. The results of these efforts are, among others, various speech and gesture recognition techniques known as Natural User Interfaces or NUIs.

1.1.2 Natural User Interfaces

NUI Group, a global research community focused on the open discovery of natural user interfaces, defines natural user interfaces as an emerging computer interaction methodology which focuses on human abilities such as touch, vision, voice, motion and higher cognitive functions such as expression, perception and recall. A natural user interface seeks to harness the power of a much wider breadth of communication modalities which leverage skills people gain through traditional physical interaction (NUI Group Community, 2009). Some representative examples of such interfaces are:

- Multitouch displays (Han, 2005): tracking and recognizing multiple fingers (and objects) on a display, which leads to new interaction techniques. These displays differ from touchscreens capable of detecting one touch in that they allow for a different interaction paradigm, while touchscreens are just a replacement for the input device, usually the mouse, in a GUI.

- Speech recognition (Rosenfeld et al., 2000): interfaces based on speech recognition, speech synthesis and natural language processing.

- Space gesture recognition (Hay et al., 2008): 3D tracking or motion capture using Nintendo Wii or similar controllers as stereo vision systems. The user interacts with the system by performing, usually predefined, spatial gestures.

- Interfaces based on electrophysiological signals (e.g. Brain-Computer Interfaces, Wolpaw (2002)): these interfaces determine the intent of the user from electrophysiological signals caused by brain, muscle, and cerebral cortex activity. For example, classifying finger gestures by interpreting forearm electromyography (EMG) signals caused by muscle movements (Saponas et al., 2009). These interfaces are also known as direct neural interfaces, mind-machine interfaces or brain-machine interfaces.

The main property of these interfaces is their intuitiveness. Ideally, a natural user interface does not require the user to undergo any training in order to interact with it. In other words, NUIs excel in terms of learnability and discoverability. NUIs build on the knowledge users get in their everyday life and exploit interaction metaphors with real-world objects to interact with digital objects in the digital world. The actual definition of NUIs is subject of ongoing debate and there are many definitions available. Some of the more concise definitions are:

- TechTarget[1]: "A natural user interface (NUI) is a system for human-computer interaction that the user operates through intuitive actions related to natural, everyday human behaviour ... A NUI may be operated in a number of different ways, depending on the purpose and user requirements. Some NUIs rely on intermediary devices for interaction but more advanced NUIs are either invisible to the user or so unobtrusive that they quickly seem invisible."
- TechTerms[2]: "A NUI is a type of user interface that is designed to feel as natural as possible to the user. The goal of a NUI is to create seamless interaction between the human and machine, making the interface itself seem to disappear."
- Wikipedia[3], Ron George: "...a user interface that is effectively invisible, or becomes invisible with successive learned interactions, to its users."

Some authors offer a more lengthy definition of natural user interfaces or give a short definition accompanied by a lengthier explanation, like Joshua Blake's definition of NUIs:

> "A natural user interface is a user interface designed to reuse existing skills for interacting directly with content (Blake, 2013)."

He explains his definition as follows:

> **"NUIs are designed.** First, this definition tells us that natural user interfaces are designed, which means they require forethought and specific planning efforts in advance. Special care is required to make sure NUI interactions are appropriate for the user, the content, and the context. Nothing about NUIs should be thrown together or assembled haphazardly. We should acknowledge the role that designers have to play in creating NUI style interactions and make sure that the design process is given just as much priority as development;
>
> **NUIs reuse existing skills.** Second, the phrase 'reuse existing skills' helps us focus on how to create interfaces that are natural. Your users are experts in many skills that they have gained just because they are human. They have been practicing for years skills for human-human communication, both verbal and non-verbal, and human-environmental interaction. Computing power and

[1] `http://whatis.techtarget.com/definition/natural-user-interface-NUI`

[2] `http://www.techterms.com/definition/nui`

[3] `http://en.wikipedia.org/wiki/Natural_user_interface`

> input technology has progressed to a point where we can take advantage of these existing non-computing skills. NUIs do this by letting users interact with computers using intuitive actions such as touching, gesturing, and talking and presenting interfaces that users can understand primarily through metaphors that draw from real-world experiences. This is in contrast to GUI, which uses artificial interface elements such as windows, menus, and icons for output and pointing device such as a mouse for input, or the CLI, which is described as having text output and text input using a keyboard. At first glance, the primary difference between these definitions is the input modality – keyboard versus mouse versus touch. There is another subtle yet important difference: CLI and GUI are defined explicitly in terms of the input device, while NUI is defined in terms of the interaction style. Any type of interface technology can be used with NUI as long as the style of interaction focuses on reusing existing skills;
>
> **NUIs have direct interaction with content.** Finally, think again about GUI, which by definition uses windows, menus, and icons as the primary interface elements. In contrast, the phrase 'interacting directly with content' tells us that the focus of the interactions is on the content and directly interacting with it. This doesn't mean that the interface cannot have controls such as buttons or checkboxes when necessary. It only means that the controls should be secondary to the content, and direct manipulation of the content should be the primary interaction method."

In his keynote speech at Interact 2011, Antão Almada[1] emphasizes Blake's 'interaction style over input modality' point of view: "People feel naturally what they are supposed to do. Natural means ease of use, non-invasive sensors, to simplify as much as possible, to give intelligence to the interface so that the users get only the useful information." The company he works for uses NUIs mainly in marketing applications, where people do not have time to learn how to interact with the applications they build and the need for easy, learnable interaction based on skills people already have is particularly strong.

Richard Monson-Haefel's presents yet another variant of NUI definition: "A Natural User Interface is a human-computer interface that models interactions between people and the natural environment." He continues specifying a couple of aspects of this definition that are especially important:

> "NUI is a form of HCI. It is important that we make that explicit;
>
> NUI models natural interactions. That means it leverages and uses as a template the interactions people have with each other (e.g. speech and gestures);
>
> NUI also models interactions between people and the natural environment (e.g. water and rocks) as opposed to their artificial environment (e.g. computers and cars.)."

Dennis Wixon talks about NUIs and interfaces in general in terms of principals and guidelines (Wixon, 2008). He believes that from a historical analysis of user interfaces we can extract a set of principles and guidelines for each interface type. "The principles are what drives the design. The guidelines are simply derivations of the principles for an individual context. Principles are what's important. Principles and data drive successful design." In this sense he analyses three different interface types that primarily build upon text—command line interfaces, graphics—graphical user interfaces and objects—natural user interfaces. He notes that interaction in CLIs builds upon the psychological function of

[1] `http://interact2011.moodle.uab.pt/mod/forum/discuss.php?d=634`

recall as the user is disconnected from the static system he interacts with and has to direct the system by learning/recalling a high number of commands. He continues by describing GUIs as an exploratory and responsive type of interaction, where the user scans through menus and recognizes the commands he had to learn and recall within CLIs. Wixon argues that interaction with GUIs is indirect as the user actually controls the mouse or the keyboard and not the GUI itself. NUIs, on the other hand, provide an unmediated interaction that is evocative and thus relies on user intuition. The commands in a NUI are few and the interaction is fast. Wixon states that NUIs are also contextual and that they understand the environment they are in and react to it naturally. He concludes this analysis of interfaces by extracting a set of principles that should apply to NUI design: the principal of performance aesthetics (the interaction should be enjoyable), the principle of direct manipulation, the principle of scaffolding (the system should support actions as you move forward and should reveal itself in those actions), the principle of contextual environments and the principal or the super real (to take real things and extend them in a logical yet unrealistic way in the digital world). These 5 principles build upon three core principles: social, seamless, spatial. These are explained on an example - the Microsoft Surface multitouch tabletop computer. Surface is *social* as it encourages social interaction around a shared information space and brings people together instead of isolating them, the actions of a Surface user and the interaction with the Surface are the same, which makes the Surface *seamless* and *spatial* as the objects in the interface have an implied physicality and the interaction with the Surface leverages spatial memory.

Wixon co-authored a book about NUIs with Daniel Wigdor, called ‘Brave NUI world’ (Wigdor and Wixon, 2011). The book presents guidelines stemming from the above-mentioned principles and tries to define NUIs from a user centred perspective: “The word ‘natural’ in natural user interfaces describes the user’s feelings during interaction and not the interface itself. Interfaces often cited as NUIs, e.g. multitouch displays with their high-bandwidth input modality, are not natural per se - they instead provide a higher potential for developing a natural user interface, if and only if an all-new interface is designed with new input actions, new affordances - in short, a new paradigm” (Wigdor and Wixon (2011), p. 10). In their view, creating a natural user interface is a design goal. It can be achieved through “a clear viewpoint, hard work, careful design, rigorous testing and some luck.” The *clear viewpoint* here is their vision of natural user interfaces: “Our vision is that a natural user interface is one that provides a clear and enjoyable path to unreflective expertise in its use. It makes skilled behaviour seem natural in both learning and expert practice. It makes learning enjoyable and eliminates the drudgery that distracts from skilled practice. It can make you a skilled practitioner who enjoys what you are doing. Natural in this sense does not mean raw, primitive, or dumbed down.” Finally, they summarize their definition of natural user interfaces as follows:

> “A NUI is not a natural user interface, but rather an interface that makes your user act and feel like a natural. An easy way of remembering this is to change the way you say ‘natural user interface’ - it’s not a natural *user interface*, but rather a *natural user* interface.”

Bill Buxton supports this *natural user* interface over natural *user interface* point of view: “It’s not about speech, it’s not about gesture, it’s not even about the phone, and it’s not about human-to-human communication,” he says. “How these things work together in a natural and seamless way that reduces complexity for the users — that’s what we’re about. Getting these things right opens up another dimension in how we have technology integrated into our lives” (Buxton, 2010). His understanding of what natural in natural user interfaces means is: “designing to take advantage of the skills we acquired in a lifetime of living in the real world. These skills are motor-sensor skills, cognitive skills and social skills.” In this sense, a designer of natural user interfaces must take care not to waste people’s skills. This design paradigm also leads to interfaces that are natural to some users and not natural to others. For example, someone that has invested the time and effort to learn touch typing may find editing text in ‘vi’[1] natural. On the other hand, despite touch being usually regarded as a natural means for interaction, Foehrenbach et al. (2008) surprisingly report how tactile feedback added to gestural interaction with high resolution interactive walls increases error rates by 10 %. Kurfess (2013) shares this hardware-agnostic point of view on NUIs. In his opinion, natural user interfaces are not about input modalities, but about interaction style. A user interface becomes natural with careful interaction design and planning; the word natural here means that interaction is appropriate for the user, the content, and the context in which the interaction takes place. This can be achieved by reusing skills that the user already has and building upon experience and expertise often unrelated to computer use. Kurfesses key guidelines for developing natural user interfaces are the use of direct manipulation where possible, enabling instant expertise for the user, reducing cognitive load and inducing progressive learning.

The roots of this debate about the essence of *natural* interaction can be traced back to the late 70s, and early 80’s. Although at the time the expression natural user interfaces was not yet born, Ben Shneidermann’s work on *direct manipulation* belongs in this context. In (Frohlich, 1993) Frohlich et al. summarized *direct manipulation* as “a style of interaction characterized by the following three properties:

1. Continuous representation of the object of interest,

2. Physical actions or labelled button presses instead of complex syntax, and

3. Rapid incremental reversible operations whose impact on the object of interest is immediately visible.”

The benefits of direct manipulation are “ease of learning, ease of use, retention of learning, reduction and ease of error correction, reduction of anxiety and greater system comprehension” (Shneiderman (1982), p. 253). Due to the overlapping of these benefits and the various definitions of natural user interfaces and their goals we can say that findings related to direct manipulation can be applied to NUIs as well. In 1985, Hutchins et al. (1985) deconstruct the term *directness* in direct manipulation in two parts: the psychological *distance* between user goals and the action a specific interface requires to achieve these goals, and to psychological *engagement* of feeling oneself to be controlling the computer directly rather than through some intermediary. “Essentially distance refers to the mismatch between the way a user normally thinks about a problem domain and the way it is represented by a computer. Systems which reduce distance reduce this mismatch and the associated

[1] “vi is a modal editor: it operates in either insert mode (where typed text becomes part of the document) or normal mode (where keystrokes are interpreted as commands that control the edit session). For example, typing i while in normal mode switches the editor to insert mode, but typing i again at this point places an ‘i’ character in the document. From insert mode, pressing the escape key switches the editor back to normal mode. A perceived advantage of vi’s separation of text entry and command modes is that both text editing and command operations can be performed without requiring the removal of the user’s hands from the home row. As non-modal editors usually have to reserve all keys with letters and symbols for the printing of characters, any special commands for actions other than adding text to the buffer must be assigned to keys which do not produce characters, such as function keys, or combinations of modifier keys such as Ctrl, and Alt with regular keys. Vi has the advantage that most ordinary keys are connected to some kind of command for positioning, altering text, searching and so forth, either singly or in key combinations. Many commands can be touch typed without the use of Shift,Ctrl or Alt. Other types of editors generally require the user to move their hands from the home row when touch typing” (Wikipedia, 2013b).

mental effort of working out what can be done in the system (semantic distance) and how to do it (articulatory distance). Engagement, on the other hand, refers to a particular style of representation based on a model world metaphor rather than on a conversational metaphor for interaction. Systems can encourage a feeling of engagement by depicting objects of interest graphically and allowing users to manipulate them physically rather than by instruction."(Frohlich (1993), p. 3) NUIs' strategy to reduce the above-mentioned psychological distance between user goals and action is to exploit skills users already possess. It is this 'recycling/remixing/reusing of skills' that makes NUIs intuitive. However, a lower *distance* between goals and actions is not always good - to accomplish more abstract tasks, this distance is actually beneficial, for example: repetitive tasks are easier to accomplish within a command line interface with a 'for loop' than manually in a more direct user interface. In this sense, Hutchins et al. conclude that direct manipulation systems benefit from the simplified mapping between goals and actions needed in an interface to achieve those goals, but at the same time lose some expressive power due to the loss in abstraction corresponding to the simplified mapping. As Bill Buxton often puts it: "Everything is best for something and worst for something else." Therefore in 1993 Frolich (Frohlich, 1993) proposed a shift in interaction from *directness* to *gracefulness*: "The puzzle of desirable indirectness in interaction is solved if we shift to a more social definition of directness as interaction in which there is least collaborative effort expended to achieve a users' goals. Activities which are cognitively indirect can then be seen as socially direct in that they have the effect of minimizing the joint work carried out by system and user entailed in achieving task success." Frolich's notion of graceful interaction builds upon Hayes' and Reddy's (Hayes and Reddy, 1983), who in 1983 proposed a decomposition of the term *graceful interaction* into a set of skills: "skills involved in parsing elliptical, fragmented, and otherwise ungrammatical input; in ensuring robust communication; in explaining abilities and limitations, actions and the motives behind them; in keeping track of the focus of attention of a dialogue; in identifying things from descriptions, even if ambiguous or unsatisfiable; and in describing things in terms appropriate for the context." We can say that *graceful* interaction addresses some of the generally acknowledged problems of NUIs. Don Norman in (Norman, 2010) argues that a pure gestural system makes it difficult to discover the set of possibilities and the precise dynamics of execution that an interface requires: "It is also unlikely that complex systems could be controlled solely by body gestures because the subtleties of action are too complex to be handled by actions – it is as if our spoken language consisted solely of verbs. We need ways of specifying scope, range, temporal order, and conditional dependencies. As a result, most complex systems for gesture also provide switches, hand-held devices, gloves, spoken command languages, or even good old-fashioned keyboards to add more specificity and precision to the commands. Gestural systems are no different from any other form of interaction. They need to follow the basic rules of interaction design, which means well-defined modes of expression, a clear conceptual model of the way they interact with the system, their consequences, and means of navigating unintended consequences." He continues: "Whether it is speech, gesture, or the tapping of the body's electrical signals for 'thought control,' all have great potential for enhancing our interactions, especially where the traditional methods are inappropriate or inconvenient. But they are not a panacea."

If direct and graceful manipulation have been terms used to discuss NUI-like interaction in the past, some authors suggest new terms for future discussion. Among them are Oviatt and Cohen's Perceptual User Interfaces (PUI) (Oviatt and Cohen, 2000), Wixon's Organic User Interfaces (OUI) (Wixon, 2008) and Petersen and Stricker's Continuous Natural User Interfaces (CNUI) (Petersen and Stricker, 2009). These are often restatements or combinations of other established and NUI related research fields we will discuss in section 1.1.5. For example, perceptual user interfaces are a combination of natural user interfaces and intelligent user interfaces: “PUIs are characterized by interaction techniques that combine an understanding of natural human capabilities (particularly communication, motor, cognitive, and perceptual skills) with computer I/O devices and machine perception and reasoning. They seek to make the user interface more natural and compelling by taking advantage of the ways in which people naturally interact with each other and with the world — both verbally and non-verbally” (Oviatt and Cohen, 2000).

To sum up, we can say that all the presented definitions generally agree that natural user interfaces should not revolve around the interface itself but should be focused on how a user perceives the interface. A concise and in this sense correct definition would be:

> “Natural User Interfaces are interfaces that are intuitive to use.”

The word cloud presented in Figure 1.2 concurs with this definition. We created the word cloud with the text of this section, which surveys various definitions of NUIs. After stop words removal[1], the emphasis of each word is proportional to its normalized frequency in the text. We can see that the word, which strikes out the most, is *‘skills’*; for an interface to feel natural it must be intuitive and therefore it must rely on skills that the user has already obtained in life. For example, due to their widespread use, the keyboard and mouse may also be considered as natural user interfaces. In this sense we can say, that once the skills to operate an interface are acquired, any interface can be regarded as natural. To some people, the gesture of waving on a sidewalk (to stop a taxi) is an example of a natural user interface, while to some it is not. It depends on the user, but also on the context in which the interaction takes place. Exploiting contextual information to the benefit of human-computer interaction broadens the communication channel between user and computer even further which in turn opens up new possibilities for designing interaction that exploits skills that users already possess. On one hand, natural user interfaces ease the acquisition, understanding and exploitation of context and on the other hand, context makes achieving natural interaction easier.

1.1.3 Context-Awareness

The fields of natural user interfaces and context-awareness, or context-aware computing, share the same goal: making devices and systems *easy to use*. Another similarity is that the definition of the term *context-awareness* also comes in different flavours. It has first been mentioned by Schilit et al. in 1994: “Such context-aware software adapts according to the location of use, the collection of nearby people, hosts, and accessible devices, as well as to changes to such things over time. A system with these capabilities can examine the computing environment and react to changes to the environment” (Schilit et al., 1994). This definition and the birth of context-awareness was due to the developments in other fields of computer science. Smaller form factors, networking, the advent of mobile computing and the increase of processing power all contributed to the shift from designing systems for a single anticipated context of use to thinking about the different contexts in which a computing device is likely to be used and how it could adopt to these contexts.

The Oxford English dictionary defines context as “the circumstances relevant to something under consideration”. In this sense we can say that in context-aware computing the *thing* under consideration is the user’s interaction with, or use of, a certain computing device and the relevant circumstances include location, time, identity, etc. Abowd et al. identified a “minimal set of necessary” context: “where, who, when, what, and why” (Abowd et al.,

[1] We removed words that bear no specific information such as common short function words (the, is, at etc.) as well as words with no specific meaning for this particular text (natural, user, interface, NUI and NUIs)

Figure 1.2: Word cloud based on words from Section 1.1.2 about NUIs. The most emphasized word, ‘skills’ highlights the importance of making NUIs feel intuitive to use, which can be achieved by designing the interface so that it relies on skills that users already possess.

2002). Similarly, Dey defines context as: “any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves” (Dey, 2000). He further identifies certain types of context that are more important than others: location, identity, activity and time. Brown et al. (Brown et al., 1997) describe context more generally: “the environment, situation, state, surroundings, task and so on.” Along those lines Dourish understands context as “information of middling relevance,” where this information is not something so central to an activity that it defines it, neither is it formed of details which have no bearing on the activity Dourish (2004). Finally, Chalmers in his book offers a brief and concise definition: “Context is the circumstances relevant to the interaction between a user and their computing environment” Chalmers (2011). To make this definition more explicit, he additionally lists some aspects of context: location, co-location, and related locations, identity of the user and of co-located people, current activity, time, sound and light levels, motion, both macro-level speed and location traces as well as micro-level patterns of acceleration, vibration and orientation, available network bandwidth and delay, available computing power, memory and storage, availability of particular interfaces, such as screens, speakers, microphones, and screen size and colour depth. Table 1.1 shows how these aspects overlap with aspects from definitions provided by other authors.

Chalmers further lists a set of categories of context or how this contextual information can be used (Chalmers et al., 2004):

- Context display: the contextual information gathered can be presented to the user, for example the current geographical location or exact orientation of a hand-held device;
- Contextual augmentation: annotating data with the context of its generation, for example a GPS-enabled photo camera;
- Context-aware configuration: for example printing a document on the nearest printer;
- Context triggered actions: dimming the lights of a GPS navigation device after dark;
- Contextual mediation: the use of context to modify services provided or the data requested to best meet the needs and limits arising from the context of the interaction;
- Context-aware presentation: the adaptation of a user interface or the presentation of data based on contextual data, for example responsive websites fall in this category.

The field of context-aware computing tightly relates to ubiquitous computing. In 1991, Weiser (Weiser, 1991) wrote: “We are therefore trying to conceive a new way of thinking about computers, one that takes into account the human world and allows the computers themselves to vanish in the background.” The *background* he is talking about is the *context* that context-awareness computing tries to capture, understand and exploit. Weiser also coined the term *calm technology* to describe an approach to ubiquitous computing, where computing moves back and forth between the centre and periphery of the user’s attention (Weiser and Brown, 1997). In achieving this vision of calm technology context-aware computing plays an important role as it strives to collect contextual information through automated means and make it easily available to an application. It is then up to the designer of the application to decide what information is relevant and how to deal with it. This frees the user from the need to explicitly interact with the application and helps the computing device running the application to disappear in the background as envisioned by Weiser.

Elementary building blocks of context-awareness are: context acquisition, context modelling and representation, and reasoning based on contextual information.

Context acquisition is the first step in each context-aware application pipeline, the step that defines not only what contextual data is and in what form this data is available but, to some extent, it also defines the architectural style of the system (Chen, 2004). There are three basic approaches to context acquisition: direct sensor access, middleware infrastructure, and context server. Direct sensor access is suitable when the device providing a context-aware service is capable of directly communicating with the sensor, when the sensor is integrated in the device and there is no need for additional data processing. Such approach is straightforward to implement but lacks the ability to handle more complex situations that require managing multiple concurrent sensors. Adopting the middleware infrastructure approach makes it easier to separate context acquisition from context managing and/or context use. Hiding low-level sensing details also eases system extensibility and code reusability. Finally, the context server approach extends the middleware infrastructure approach by allowing access to remote data sources. Delegating context data acquisition and any needed processing to an external source also reduces the resource intensive burden on,

ASPECTS OF CONTEXT		
Dey (2000)	Abowd et al. (2002)	Chalmers (2011)
Any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves.	WHERE	Location Co-location Related locations
	WHO	Identity of the user Identity of co-located people
	WHAT	Activity
	WHEN	Time
	WHY	
		Sound and light levels Macro-level speed and location traces Micro-level patterns of acceleration, vibration and orientation Available network bandwidth and delay Available computing power, memory and storage Availability of particular interfaces, such as screens, speakers, microphones Screen size and colour depth

Table 1.1: Aspects of context according to different authors.

usually mobile, resources-scarce, devices running context-aware applications. Independently of the approach taken to acquire contextual data, the sensors used in the process can be categories as physical, virtual and logical (Indulska and Sutton, 2003). The most frequently used are physical sensors which measure physical properties like light, speed, acceleration, audio, temperature, presence and location of touch, blood pressure etc. For virtual sensors the source of data are different applications or services, such as e-mails, electronic calendars and keyboard or mouse input dynamics. Finally, logical sensors are those that combine physical or virtual data from physical or virtual sensors with additional knowledge or data from databases or similar sources in order to extract higher level information. The identification technique presented in Chapter 2 can be regarded as a logical sensor for user identity.

Once contextual data is acquired, it must be stored or represented in a model that will later facilitate the use of this data (by direct exploitation or by additional reasoning based on contextual information). A recent survey of context modelling and reasoning techniques (Bettini et al., 2010) listed the following seven requirements for context models and context management systems: ability to cope with *heterogeneity and mobility* of different contextual information sources, possibility to represent *relationships and dependencies* between sources and data, *timelessness* delivered by a system for handling context histories, robustness to *imperfection* in contextual data, computationally efficient *reasoning* to determine whether there has been a change in context and if the change requires the system to react or adapt, *usability of modelling formalisms* that eases context use and context modelling for application designers, and *efficient context provisioning* which is not trivial to achieve in the presence of large models and numerous data objects. Similar requirements were also identified by other authors (Strang and Linnhoff-Popien, 2004; Korpipää and Mäntyjärvi, 2003) and various solutions on how to meet them were proposed. Briefly, these solutions are: key-value models, markup scheme models, graphical models, object-oriented models, logic-based models and ontology-based models. Furthermore, Bettini et al. (2010) explore the possibilities to combine different models and reasoning techniques into hybrid models that can truly address all the above mentioned requirements.

In (Makkonen et al., 2009), the authors present a short survey of context-awareness use cases. The first is a general case named *technology enhancing HCI*, where context-awareness is used to enable devices to act smartly and make life easier for the user. Examples range from smart home appliances like gesture controlled DVD players to enhanced desktop applications like a messaging system that minimizes interruptions to the user or adopts the possible input space based on the context. Ambulatory monitoring, remote assisted rehabilitation, abnormality detection and activity monitoring are examples of the second use case for context-awareness, i.e. *healthcare.* The next use case consists of context-aware systems aimed at *diaries and memory support by tagging*; an example of such an application for photo collection management will be presented in Chapter 4. Context-awareness is also beneficial for *mobile guides*: on-site tourist guides, recommendation systems that exploit information about the user, current location and perhaps information about what other users liked. These same information can also be exploited by context-aware systems for *advertising.* Finally, the two most useful use cases are *work assistance* and *learning.* Here context-aware systems can remind the worker of the correct order of working phases, perform automatic quality control, enable learning at any time and any place in situations that are the most appropriate for the skill being learned etc.

The possibilities where context-awareness can be applied are endless and can be very different from each other. This diversity poses a lot of challenges and concerns that need to be addressed before context-awareness can reach maturity and wide adoption. In (Schmidt, 2003) Schmidt identified the central research challenges in context-awareness as:

- **Understanding the concept of context:** how is context connected to situations in the real world, how can context be represented and stored in a universal way;

- **How to make use of context:** once context information is available, what is it useful for, what type of applications can be enhanced, what about ambiguity and reliability, the joint interpretation of ‘standard’ and contextual input;

- **How to acquire context information:** the process of capturing a real world situation, assessing its relevant features and storing it in an abstract representation, a prerequisite for any context-aware system;

- **Connecting context acquisition to context use:** in situations where context acquisition and context use is distributed, mechanisms for communication, common understanding and representation of contextual information are required;
- **Understanding the influence on human computer interaction:** the user's understanding and control of the system;
- **Support for building context-aware ubiquitous computing systems:** providing support for context acquisition, context provision, and context use in order to make the process of implementing context-aware applications much simpler;
- **Evaluation of context-aware system:** evaluation must be done in context, which may in turn influence the evaluation itself.

1.1.4 Multitouch Displays

This thesis explores natural user interfaces from a context-awareness perspective and uses multitouch displays as an exemplary NUI to do so. To better understand the use cases in Chapters 2, 3 and 4, some background on multitouch displays is needed.

> Multitouch displays are displays capable of detecting multiple points of contact.

In contrast to multitouch trackpads, which are also capable of detecting multiple touchpoints and serve only as an input device, multitouch displays combine input and output in the same device and on the same location. This leads to a more direct type of interaction as the user interacts with objects on the display by directly touching them. The ability of detecting multiple touchpoints also opens up numerous possibilities for designing multitouch interaction such as gestural interaction and multi-user interaction.

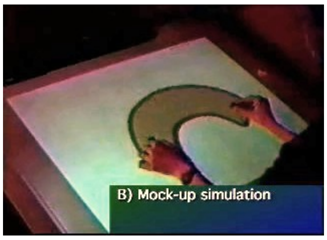
Although MT displays gained broader attention only in the last few years, the first true multitouch display was invented by Bob Boie at Bell labs in 1984 (Buxton, 2009). It consisted of a transparent capacitive foil overlaying a CRT monitor. Even before, in 1983, Myron Krueger explored the possibilities of unencumbered (i.e., no gloves, mice, styli, etc.) rich gestural interaction by developing a vision based system for tracking hands and enabling multiple fingers, hands and people to interact using a rich set of gestures' (Buxton, 2009). Table 1.2 lists some of the most influential work connected with MT displays and gives a brief historical overview of the field.

BRIEF HISTORY OF MULTITOUCH DEVELOPMENT	
The sensor frame (1985) (McAvinney, 1986): optical sensors placed in the corners of a CRT monitor capable of detecting the point and angle of touch of up to three fingers.	

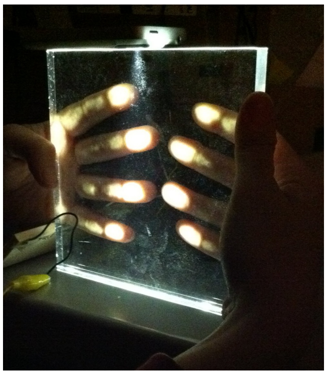
Bi-Manual Input (1986) (Buxton and Myers, 1986): a study on bi-manual input which showed that continuous bi-manual control was easy for the user and that it increased productivity.

Simon (1992) (Buxton, 2009): IBM and Bell South released Simon, a phone that used a touch interface without any buttons. Though Simon was a single touch device, it forecasted features that we see today in multitouch mobile phones.

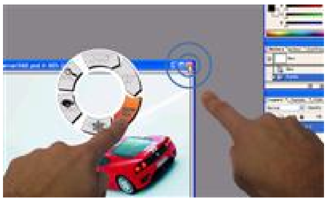
Graspable/tangible interfaces (1995) (Fitzmaurice et al., 1995): a system capable of sensing the identity and location of objects on a tabletop display. This work introduced the notion of tangible interfaces.



MT sensing using frustrated total internal reflection (2005) (Han, 2005): a cheap, yet robust multitouch optical sensing technique scalable to large installations that popularized multitouch displays and multitouch interaction. Image taken from Jain and Low (2011).



Precise selection techniques for multi-touch screens (2006) (Benko et al., 2006): a paper dealing with precise pointing and selection on MT displays.



Apple iPhone(2007): the first and most popular mobile phone with a multitouch interface. It supports various MT gestures, like for example the 'pinching' gesture for zooming introduced by Krueger in 1983.

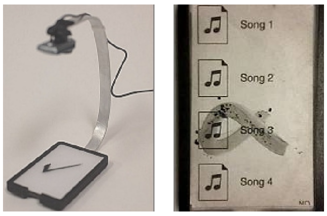
Daniel Wigdor (Wigdor et al., 2009, 2007): various studies exploring the possibilities of using a multitouch display for collaborative work, in mobile devices and other use cases.

Multi-touch technologies (2009) (Ğetin G., 2009): NUI Group released a book about multitouch technologies. Besides hardware details, it also addresses the problem of tracking and identifying fingers on the screen.

NUI Group

Windows 7 (2009): first operating system with native support for multitouch interaction.

Gesture research (Mauney et al., 2010): a global study aimed at identifying the most common user-generated gestures for control of a touchscreen within a given country and across cultures.



Microsoft Surface II (2011): a 40 inch screen with optical multitouch sensing capabilities only 4 inch thick.	

Table 1.2: Some important milestones in the development of multitouch displays and multitouch interaction.

These examples show the direction of research in the field of multitouch interaction. One research direction aims at improving multitouch interaction from a hardware point of view; questions like scalability to small (e.g. mobile phones) and large (e.g. multimedia walls) displays or touch detection and tracking accuracy are addressed here. Another research direction tries to define multitouch as an interaction paradigm by searching for appropriate gestures, understanding what users expect from multitouch interaction and how they naturally react while interacting with a MT display. Finally, great effort is undertaken by some researches to identify the most appropriate use cases for multitouch interaction. Working along one of these research directions is rarely the case, because of the influence between them, for example, a particular use case may require particular hardware specifications.

Enabling Factors

A user interface is defined by the hardware (physical part; input and output devices) it uses and the software (logical part) that enables the user to interact with the computer by means of the interface's hardware. As they interact with each other, both physical and logical aspects of an interface must be taken in consideration in the phase of interface development and in identifying potential use cases. It is therefore important to understand the hardware and software mechanisms underlying an interface.

Hardware: Although a user interface is much more than just an input/output device, it is crucial to understand how this device works, its limitations, physical dimensions and the data it provides - the expressive language it offers. On a hardware level, we roughly divide MT sensing methods in two categories: optical and non-optical. The first are mostly used in small devices, while the second in larger installations.

Non-optical sensing methods: non optical methods are represented by capacitive, resistance and surface acoustic wave displays (Schöning et al., 2008). They were first developed for displays capable of detecting one touch point and later adapted to multitouch sensing by the addition of special controllers, multiple sensing layers etc. This means that not all of them are fully multitouch capable. For example, most surface wave acoustic displays are only capable of detecting two touch points. Other features that vary depending on the method are size, display opacity, robustness, energy consumption and the capability to detect stylus touch.

Optical sensing methods: simple construction, low cost and scalability are the key properties of optical MT sensing methods, such as frustrated total internal reflection (FTIR), diffused illumination (DI), laser/led light plane (LLP) and others. Each of these methods consists of an optical sensor (typically a camera), infrared light source, and visual feedback in the form of projection or an LCD monitor (Ğetin G., 2009). As Figure 1.3 shows, the

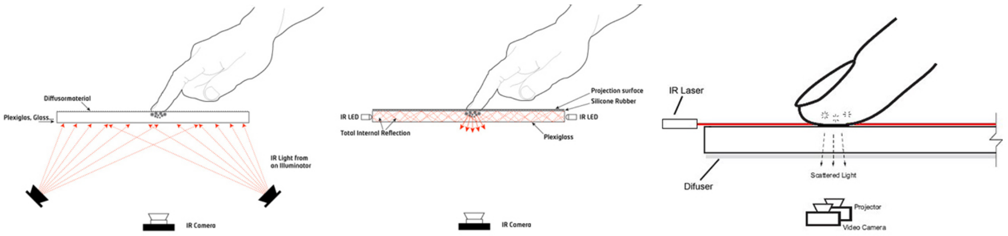


Figure 1.3: Optical multitouch sensing methods techniques (Ğetin G., 2009): diffused illumination (DI, left), frustrated total internal reflection (FTIR, middle), and laser/led light plane (LLP, right).

idea is to capture infrared light that reflects from the fingers, when the surface is touched. At the same time visual feedback is provided by a projector (not shown in the picture).

The difference between optical methods is mainly in the source of IR light. As with non-optical displays, different construction methods bring different properties. DI displays are capable of sensing objects (Jordà et al., 2007), FTIR offers superior finger tracking (Han, 2005) and LLP, when combined with an LCD for visual feedback, has the greatest picture quality (Motamedi, 2008).

Software: Figure 1.4 shows a schematic overview of multitouch software using optical sensing methods. The main components are the tracker and the gesture recognizer.

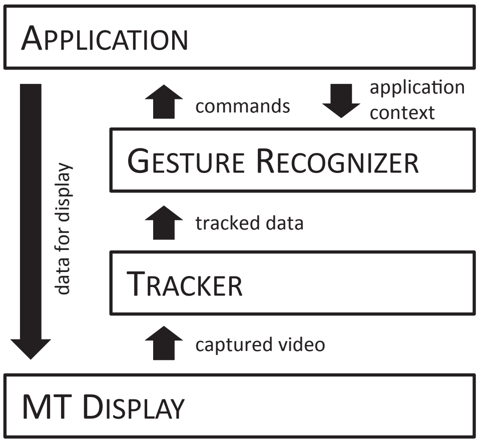


Figure 1.4: Schematic overview of multitouch software (Blažica, 2009).

The tracker: (omitted if using non-optical sensing methods) processes raw captured video from the camera and identifies the same finger on successive frames with a K-Nearest Neighbours approach (Ğetin G., 2009). This is achieved through static or dynamic background subtraction and with threshold and high-pass filtering. The tracking process is simplified by hardware construction; using IR light sources and bandpass camera filters of the same wavelength removes background noise (Blažica, 2009).

The gesture recognizer analyses tracked data and the current state of the application. Upon this, it identifies the presence of gestures and their targets in the application. Gesture recognition is usually implemented using Hidden Markov Models, artificial neural networks (Ğetin G., 2009) or rules (Blažica, 2009). The difficulty of MT gesture recognition arises from their variety as gestures can be progressive or non-progressive (drawing a circle vs zooming), simple or compounded (one finger flick vs pinching) and dynamically or statically modelled (depending on the consideration of gesture speed and acceleration) (De Nardi, 2008).

Strengths and weaknesses: first of all, MT displays offer a natural style of interacting with computers as they leverage the knowledge and skills we learn in the real world (e.g. rotating a picture with two fingers). The expressiveness of the language MT displays provide in comparison with GUIs is enormous. Instead of a single cursor that moves in two dimensions and a few buttons, MT displays offer an almost infinite number of cursors and gestures (limited only by the displays size and the number of contemporary users). Another advantage is the possibility of replacing different specialized input devices with a lot of buttons (e.g. a sound mixer) with one MT display and customized software[1]. On the other side, except for rare examples like Tactapad[2], MT displays don't give any physical feedback to the user as the keyboard does. This also means, that MT displays can't be used by blind people or in situations where visibility is affected, like outdoors in extreme sunlight conditions (Buxton, 2009). Fatigue caused by prolonged interaction with vertically mounted displays known as 'gorilla arm' is another problem concerning MT displays. Typically a user sits in front of his/her computer and it is impractical for him/her to hold his/her hand raised reaching for the MT display for a long time.

Use Cases

Today, the use of multitouch displays has been explored in various fields. An in-depth overview can be found in (Shaer and Hornecker, 2010), where Shaer and Hornecker identified the key areas where tangible interaction has been applied. Tangible interaction or Tangible User Interface (TUI), is an umbrella term introduced in 1997 by (Ishii and Ullmer, 1997) and encompasses interactive surfaces such as multitouch displays, graspable objects, such as cards and books associated with digital information, and the use of ambient media such as sound, light and airflow for interaction in the background (at the periphery of human perception). Areas of TUIs application are: learning, problem solving and planning, information visualization, tangible programming, entertainment, play and edutainment, music and performance, social communication, tangible reminders and tags. What follows is a brief summary of the cited survey and the categories of MT use cases.

The use of TUIs (and MT displays) as computer supported *learning* tools has two main reasons. The first is that learning researchers and toy designers have always followed the strategy of augmenting toys to increase their functionality and attractiveness, while the second reason is that physical learning environments engage all senses and thereby support the overall development of the child (the theory on learning poses great importance on physical movement, multimodal interaction and also suggests that gestural interaction supports thinking and learning). TUIs for learning can be further split in the following categories: *digital manipulatives* - "computationally enhanced versions of physical objects that allow children to explore concepts, which involve temporal processes and computation", *computationally enhanced instruction kits* - "make concepts accessible on a practical level that are normally considered to be beyond the learner's abilities and age-related level of abstract thinking," such as exploring concepts of volume and area with blocks, *storytelling* - supporting early literacy education with applications that augment traditional toys and play environments or books, *learning for children with special needs* - "physical interaction here has benefits of slowing down interaction, training perceptual-motor skills, providing sensorial experience", and *diagnostic tools* - "the kinds of mistakes and steps taken in building

[1] `http://www.jazzmutant.com/`

[2] `http://tactiva.com`

a spatial structure after a given model can indicate the level of cognitive spatial abilities a child has developed." A strong use case for multitouch displays related to learning are interactive whiteboards. Several studies demonstrate the positive effect interactive whiteboards have on the learning process. A case study on creative teaching and learning in literacy and mathematics concludes that special features such as interactivity, speed, capacity and range enhance the delivery and pace of the learning session. The research also indicates that it is the skill and the professional knowledge of the teacher who mediates the interaction, which is critical to the enhancement of the whole-class teaching and learning processes (Wood and Ashfield, 2008).

Support for epistemic actions, physical constraints and tangible representations of problems are three aspects of TUIs responsible for their use in *problem solving and planning.* An *epistemic action* is an action performed by the user that changes the physical world with the intention to aid the user's mental process, task or action. A typical example of an epistemic action supported by MT displays is rotating or aligning physical objects on the display. In terms of problem solving, *physical constraints* can communicate interaction syntax and limit the solution space. By doing so, physical constraints decrease the need for learning explicit rules and thus ease the use of a computational system for the task at hand. Finally, *tangible representations of problems* enabled by TUIs have been successfully exploited in spatial or geometric applications, such as urban planning and architecture, where the physical arrangement and manipulation of objects has a direct mapping to the represented problem. This increases users' spatial cognition, reduces cognitive load, and enables more creative immersion in the problem.

Rich multimodal representation as well as the possibility of two-handed input are two features of TUIs, which hold great potential for interacting with *information visualizations.* Some examples are the use of TUIs for interactive visualizations in the fields of neurosurgery, geophysics, and structural molecular biology. More possibilities are open and need to be explored and proposed solutions further validated, however, these first studies already report several advantages of such TUI-based visualizations. Most notably, increased efficiency and ease of learning. A well-known and warmly accepted adaptation of a multitouch display for data visualization is Perceptivepixel's commercial product, the 'magic wall', first used during CNN's coverage of the 2008 US presidential election. It was credited to make numbers and data more accessible to viewers.

Music and performance applications are one of the oldest and most popular areas for TUIs. Properties that make them so adequate for musical performance are support of collaboration and sharing of control, continuous, real-time interaction with multidimensional data, and support of complex, skilled, expressive, and explorative interaction. On a high-level, TUI applications for music and performance can be divided in four categories: fully controllable sound generators or *synthesizers*, *sequencers* for mixing and playing audio samples, *sound toys* which offer only limited user control, and *controllers* that remotely control an arbitrary synthesizer. An exhaustive list of around 90 example applications is presented in Kaltenbrunner (2013). Music-related applications of TUIs are also systems that support VJ-ing — the creation or manipulation of imagery in real-time through technological mediation and for an audience, in synchronization to music (Amerika, 2009).

Another strong use case for TUIs is *tangible programming* - the use of tangible interaction techniques for constructing computer programs. Tangible programming systems have mostly been applied to teach children to program and help them learn while at the same time offering some level of entertainment. Applications of tangible programming spread also in other fields such as database querying and industrial work in plants. It has to be noted that most tangible programming examples are built around physical objects and do not necessarily involve a multitouch display. The same is true for TUIs used in *entertainment, play, and* *edutainment*, where the principles of physical input, tangible representation, and digital augmentation are exploited. Examples include museum interactives that combine hands-on interaction with digital displays, augmented traditional board games and interactive playgrounds. Care should be taken while designing these systems for a wide population; a case study of using a MT tabletop in a museum (Hornecker, 2008) showed that the multitouch display actually distracted visitors from the content and that visitors perceived it more as a toy for children.

Social communication is another category of successful TUI application. Examples range from figurines that represent users in video-conferencing systems to prototypes for remote intimacy. In terms of multitouch displays, it has been noticed that their use stimulates more equal participation in case of co-located groupware, where a group of people gathers around a tabletop to collaboratively solve a problem Morris et al. (2006). Furthermore, MT walls, tabletops and laptops in systems that provide functionalities to layout and manipulate multiple live desktops (select and pull-out any user chosen applications from their own laptops onto the wall or the table, enable visualization, overlay and mark up of live visual renderings from any of the user's own applications) give all group members equal access to touch manipulation around a tabletop. These functionalities allow easier data sharing, spontaneous walk-up collaboration, larger display areas and multi-touch input models (Wigdor et al., 2009).

The final category of TUI use Shaer and Hornecker list in (Shaer and Hornecker, 2010) are tangible *reminders and tags*. An example of this category connected to multitouch displays are tangible reminders in form of vacation souvenirs that, when placed on an interactive surface, open an associated photo collection.

A category not explicitly mentioned in (Shaer and Hornecker, 2010) is the use of multitouch displays in *mobile devices* - small devices such as mobile phones, portable music players, navigation systems and others. These devices have contradictory needs: the need to be small, have a large display and a lot of buttons. A MT display can meet all this needs without compromising. In addition, these devices can also take advantage of gestures. Another category left out by the above mentioned survey is *desktop use*. Despite several commercially available monitors capable of MT and native MT support already in Windows 7, MT has not been successfully implemented for desktop use yet. This is due to the already mentioned 'gorilla arm' effect and the fact that most application still don't take full advantage of MT interaction as they were developed for the GUI interaction paradigm. Perhaps projects like 10/GUI[1] (a concept aimed at overcoming these problems by rethinking the desktop as we know it today) or Windows 8 will make it more clear how to approach multitouch interaction for desktop use.

This overview of multitouch use cases shows how they can be successfully exploited in extremely varied, and sometimes overlapping, scenarios; for example, a museum installation that supports learning but at the same time offers entertainment. It is plausible to expect that further research efforts will both, improve already existing use cases and make new use cases possible.

1.1.5 Related Fields

This section briefly summarizes definitions of research fields closely-related to natural user interfaces and context-awareness. The aim is to highlight their common ground and to show how all strive towards the same goal: a general improvement in Human-Computer Interaction by bringing the E (as environment or context) in the HCI picture. Figure 1.5

[1] `http://10gui.com/`

illustrates the communication triangle composed of users, context, and computing devices. Fields concerned with some aspect of this interaction triangle are:

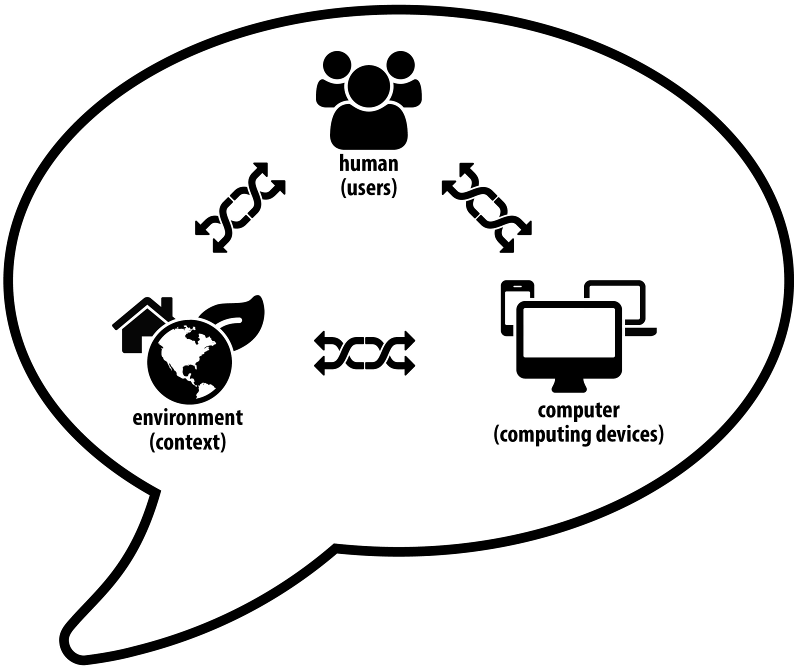


Figure 1.5: The communication triangle of Human-Computer Interaction; communication is possible on all levels, between each pair of the involved entities: human, computer and environment. Two types of arrows represent two types of communication: explicit and implicit.

- Intelligent User Interfaces (IUI): “Intelligent user interfaces (IUIs) are human-machine interfaces that aim to improve the efficiency, effectiveness, and naturalness of human-machine interaction by representing, reasoning, and acting on models of the user, domain, task, discourse, and media (e.g., graphics, natural language, gesture). As a consequence, this interdisciplinary area draws upon research in and lies at the intersection of human-computer interaction, ergonomics, cognitive science, and artificial intelligence and its subareas (e.g., vision, speech and language processing, knowledge representation and reasoning, machine learning/knowledge discovery, planning and agent modelling, user and discourse modelling)” (Maybury, 1998).

- Ubiquitous Computing (UBICOMP): “Ubiquitous computing is the method of enhancing computer use by making many computers available throughout the physical environment, but making them effectively invisible to the user” (Weiser, 1993).

- Pervasive Computing: “A device can be a portal into an application-data space, not a repository of custom software that a user must manage. An application is a means by which a user performs a task, not software written to exploit a device’s capabilities. And a computing environment is an information-enhanced physical space, not a virtual environment that exists to store and run software” (Saha and Mukherjee, 2003).

- Physical Computing: “Physical computing, in the broadest sense, means building interactive physical systems by the use of software and hardware that can sense and respond to the analog world” (Wikipedia, 2013a).

- Ambient Intelligence (AmI): “Ambient Intelligence (AmI) is about sensitive, adaptive electronic environments that respond to the actions of persons and objects and cater for their needs. This approach includes the entire environment – including each single physical object – and associates it with human interaction,” (Aarts and Wichert, 2009) or from a more philosophical point of view: “Ambient Intelligence is the way for us to re-immerse ourselves in life, and not in technology” (Epstein, 1998).

- Everyware: “In everywhere, all information we now look to our phones or Web browsers to provide becomes accessible from just about anywhere, at any time and this is delivered in a manner appropriate to our location and context” (Greenfield, 2006).

- Internet of things (IoT): “The basic idea of this concept is the pervasive presence around us of a variety of things or objects – such as Radio-Frequency Identification (RFID) tags, sensors, actuators, mobile phones, etc. – which, through unique addressing schemes, are able to interact with each other and cooperate with their neighbours to reach common goals” (Atzori et al., 2010).

Among these research fields UBICOMP is the oldest and perhaps the most influential one. It started considering the environment in HCI and predicted a future where computing devices disappear in the background and users interact with them through the physical environment. The shift from past-HCI to future-UBICOMP-HCI is illustrated in Figure 1.6.

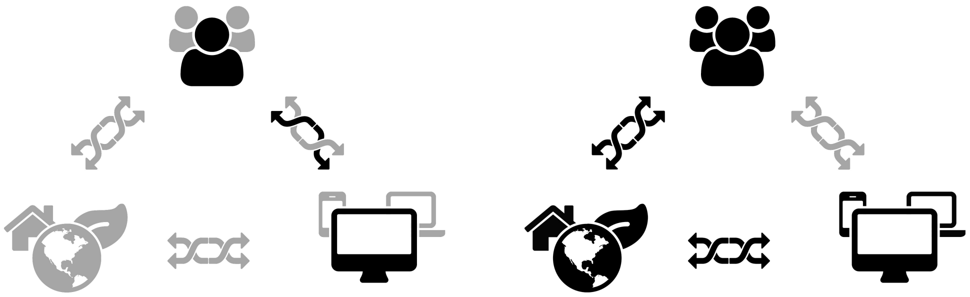


Figure 1.6: Illustration of HCI’s past (a single user explicitly interacts with a single computer; left) and HCI’s future as envisioned by UBICOMP (many seemingly invisible computing devices available to users, right).

1.2 Hypotheses and Goals

The purpose of this thesis is to show that natural user interfaces are a viable way towards context-aware systems. Previous sections introduced the fields of natural user interfaces and context-awareness and highlighted the open research questions in these fields. Arguably the most significant are the questions of user identity and the question of understanding implicit interaction clues. Both questions fall under the ‘context acquisition’ challenge, which we consider to be the most important as it is a prerequisite for further research in context-awareness; we need to be able to extract context if we want to use it in applications or to study how it affects human-computer interaction.

This thesis addresses the problem of context extraction and understanding in terms of natural user interfaces. The presented solutions are showcased on multitouch displays. The thesis’ hypotheses are:

1. Our first hypothesis is that Natural User Interfaces (NUIs) and Multitouch (MT) displays are, to some extent, inherently context-aware; that the information they carry is sufficient to build context-aware systems.

2. The second hypothesis is that understanding and exploiting context-awareness further extends the expressiveness of NUIs.

3. The third hypothesis is that MT displays provide enough information to perform user identification.

4. The fourth hypothesis is that the way we interact with NUIs can implicitly disclose additional (contextual) information that can be exploited by a context-aware system to better understand the user.

The goals and expected contributions of the dissertation are the following:

- Provide a thorough overview of the literature of the fields of context-awareness and natural user interfaces/multitouch displays;

- Define a clustering algorithm for hand detection on multitouch displays;

- Define a biometric method for user identification on multitouch displays;

- Define a model for implicit extraction of information from user interaction;

- Implement and evaluate the proposed algorithms on artificial and/or real-world data;

- Set up a database for further research on user identification on multitouch displays.

1.3 Scientific Contribution

Broadly speaking, the thesis provides an overview of the intersection of the fields of context-awareness and natural user interfaces as well as some closely related, but often too distant, fields such as intelligent user interfaces, pervasive computing, ubiquitous computing, ambient intelligence etc. The research covered by the thesis increments our understanding of context-awareness, how it can be achieved through natural user interfaces and how it augments interaction with these interfaces.

More specifically, the contributions of this thesis are:

- A novel solution to user identification on multitouch displays; where related work uses additional hardware or restricts the solution to only a subset of multitouch displays, the method proposed only considers data common to all multitouch displays and thus provides a universally applicable solution. The description of the method, its implementation and evaluation are published in the International Journal of Human Computer Studies (Blažica et al., 2013a);

- A clustering algorithm for hand detection is presented that opens up design possibilities to this date merely theoretically envisioned in literature (Partridge and Irani, 2009). The algorithms's description and evaluation are published in the Lecture Notes in Computer Science series (Blažica et al., 2013).

- The development of an implicit human-computer interaction method for photo collection management. An implementation of the method in a tablet application was used to successfully validate the method. The method's explanation and preliminary results are published in the Lecture Notes on Computer Science series (Blažica et al., 2011), while the implementation and validation in a tablet application is published in the Personal and Ubiquitous Computing journal (Blažica et al., 2013b).

- User identification, hand detection and extraction of implicit information for human-computer interaction represent three concrete contributions in the context-awareness open challenge of acquiring contextual information.

At the same time, these contributions can be directly mapped to features that increase the expressive power of natural user interfaces. Finally, this thesis' relevance spreads also in the realm of personal information management by showing how implicitly conveyed contextual information can be exploited to facilitate organization of photo collections.

1.4 Thesis Structure

This thesis is structured as follows. Chapter 1 introduces the broader field of the thesis, namely context-awareness and natural user interfaces, as well as some closely related and partly overlapping fields. This chapter also reviews the more narrow field of tangible interaction and multitouch displays and states the hypotheses, goals and contributions of the thesis. These are in turn more precisely presented in the main three chapters, which are based on original research papers published in internationally recognized journals of the human-computer interaction field. Chapter 5 discusses the implications of the presented research in relation to the research hypotheses and goals stated in Chapter 1. Chapter 5 also concludes the thesis and provides some guidelines for future work.

2 MTi: a Method for User Identification on Multitouch Displays

In this chapter, the paper (Blažica et al., 2013a) titled "MTi: a method for user identification on multitouch displays" by Bojan Blažica, Daniel Vladušič and Dunja Mladenić is presented. The paper is published in the International Journal of Human Computer Studies[1].

The paper first provides an exhaustive overview of literature concerned with user identification and user distinction methods on multitouch multi-user displays. State-of-the-art methods are presented by considering three key aspects: user identification, user distinction and user tracking. Next, the paper proposes a novel method for user identification on multitouch display. The method, called MTi, is capable of user identification based solely on the coordinates of touch points, which makes the method universally applicable to all multitouch displays, regardless of their construction. The method's accuracy was tested on two different datasets; data for the smaller dataset composed of 34 users was gathered on a home-made led light plane (LLP) multitouch display, while data for the larger database composed of 100 users was extracted from images in hand geometry database. Additionally, a usability study was performed to see how users react to the proposed identification method, to pinpoint the method's strengths and weaknesses and to frame its scope.

In terms of context-awareness, the MTi method is an example of how a natural user interface and its increased amount and variety of data can fruitfully be exploited to acquire additional contextual information. The method falls under the context acquisition challenge and captures one of the key aspects of context, namely the user identity (as described in Section 1.1.2).

The first author of the paper conceived the idea of user identification based on touch coordinates and developed the features at the core of the method. He also implemented the method, conducted the experiments needed to assess the method's identification accuracy, carried out the usability evaluation and wrote the paper. Co-authors contributed with advice at several steps in the process.

[1] IF 2011 = 1.171; JI - ergonomics: 2 quartile; VJ - psychology, multidisciplinary: 2 quartile

Available online at www.sciencedirect.com

SciVerse ScienceDirect

Int. J. Human-Computer Studies 71 (2013) 691–702

International Journal of Human-Computer Studies

www.elsevier.com/locate/ijhcs

MTi: A method for user identification for multitouch displays

Bojan Blažica[a,b,*], Daniel Vladušič[a], Dunja Mladenić[b,c]

[a]XLAB Research, Slovenia
[b]Jožef Stefan International Postgraduate School, Slovenia
[c]Artificial Intelligence Laboratory, Jožef Stefan Institute, Slovenia

Received 16 May 2012; received in revised form 28 February 2013; accepted 11 March 2013
Communicated by J. Laviola
Available online 19 March 2013

Abstract

This paper describes MTi, a biometric method for user identification on multitouch displays. The method is based on features obtained only from the coordinates of the 5 touchpoints of one of the user's hands. This makes MTi applicable to all multitouch displays large enough to accommodate a human hand and detect 5 or more touchpoints without requiring additional hardware and regardless of the display's underlying sensing technology. MTi only requests that the user places his hand on the display with the fingers comfortably stretched apart. A dataset of 34 users was created on which our method reported 94.69% identification accuracy. The method's scalability was tested on a subset of the Bosphorus hand database (100 users, 94.33% identification accuracy) and a usability study was performed.

Keywords: Multitouch; User identification; Biometrics; Identity-enabled surfaces

1. Introduction

The origins of multitouch interaction can be traced back to the early 80s, to Mehta's (1982) flexible machine interface, Nakatani and Rorlich's (1983) soft machines as well as other pioneering work in the field. Since then, multitouch interaction evolved significantly, mostly in terms of the hardware supporting it, i.e. multitouch displays. There are now various types of sensing technologies which are roughly divided into optical and non-optical technologies. Optical technologies (e.g. frustrated total internal reflection and diffused illumination, see Çetin et al., 2009 for a comprehensive list) are suitable for larger installations and capable of detecting a virtually unlimited number of touches, while non-optical (e.g. capacitive sensing) technologies are used mostly for smaller displays. Other properties that differ depending on the underlying sensing technology of a display are the ability to detect objects placed on the display, the provided information regarding the acquired touchpoints (coordinates of touch, orientation of touch), the capability to identify users, etc. In short, what all technologies have in common is the ability to provide information concerning the coordinates of multiple touch points.

Multitouch interaction is more than just the ability to sense multiple touchpoints — it also incorporates the way we interpret these touchpoints and the information they contain. A common way of exploiting this informational potential is the implementation of gestural interaction, for example the pinch gesture for zooming/scaling on smartphones (the gesture was first proposed by Krueger et al. (1985)). Other examples of scenarios that were made possible by (or gained benefit from) the intuitiveness, directness and expressive power of multitouch interaction were interactive tables in education (Higgins et al., 2011), public multitouch installations (Jacucci et al., 2010), music production (Jordà et al., 2006), data visualization (Arnell et al., 2008) and computer supported collaborative work (Ardaiz et al., 2010). In all of these (and other) scenarios, interaction and the resulting user experience can be further augmented by the introduction of user aware multitouch displays; 'role taking', 'interface adaptation' and 'access restrictions' are just a few features that user

*Corresponding author. Tel.: +386 41 502 959.
E-mail addresses: bojan.blazica@xlab.si, bojan.blazica@gmail.com (B. Blažica).

1071-5819/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ijhcs.2013.03.002

identification brings to the interaction table. Currently, only a small subset of multitouch displays is capable of identifying users, while the majority still lack this technology (see Section 3).

The contribution of our work is the development and evaluation of MTi, a biometric identification method for multitouch displays, that is based only on the touch coordinates of one of the user′s hands. This means that the method is independent of the display′s touch detection technology and can be implemented on all displays capable of accommodating a human hand and detecting 5 or more touchpoints.

This paper is structured as follows: in Section 2, we describe the motivation for our work by reviewing literature concerned with identity aware multitouch interaction. This is followed by an overview of user identification methods and related work described in Section 3. We continue in Section 4 with a description of our approach to user identification — the MTi method. The method's accuracy and scalability are evaluated in Section 5, while the method′s usability is evaluated in Section 6. Following the discussion in Section 7, we conclude with a concise definition of the contribution of our work and list some possibilities for further research.

2. Motivation

The motivation for our research can be divided into two aspects. The first focuses on research that deals with user identification on multitouch displays from a more theoretical standpoint and is concerned with *concepts and frameworks* while the second deals with studies of particular *scenarios* that could take advantage of user-awareness.

A conceptual approach to user identification on interactive surfaces is presented by Ryall et al. (2005) and their iDwidgets — specialized identity differentiating widgets. iDwidgets are elementary GUI components for user-aware environments with user or role specific customizations of functionality, appearance, content, etc. A platform that could host iDwidgets is presented by Partridge and Irani (2009) along with a discussion of the advantages of identity-enabled (IE) surfaces over non-IE surfaces and problems concerning the implementation of IE surfaces. Our research addresses one of these problems, i.e. the problem of providing identity recognition for systems which have no native support for identity recognition. The same problem is also addressed by Kim et al. (2009); they argue that mechanisms such as additional video input, capacitance or social and software protocols are needed for user identification on multitouch displays. Moreover, they also propose a ‘general model to describe the user identification process for camera-based multi-touch tabletop displays’ (Kim et al., 2009, p. 4).

One of the many scenarios that involve user identification on multitouch displays is the use of collaborative gestures for co-located groupware. Morris et al. (2006) formalized the notion of cooperative gestures as ‘interactions where the system interprets the gestures of multiple group members collectively in order to invoke a single command’ (Morris et al., 2006, p. 1209). Their conclusion was that the ‘use of cooperative gestures can add value to applications as a means of increasing participation, drawing attention to important commands, enforcing implicit access control, facilitating reach on large surfaces, and/or enhancing social aspects of an interactive experience’ (Morris et al. 2006, p. 1209). Similarly, Buisine et al. (2011) report that in a creative problem-solving domain, interactive tabletop systems have a positive impact on group collaboration. They suggest that this may be due to an increase in social comparison in the configuration of users around a tabletop and that this effect could be further emphasized by an explicit real-time feedback of the user's performance — a feature that can be made possible via user identification. Along the same lines, Rogers et al. (2009) report that tangible interfaces stimulate participation from those who find it hard to talk or are incapable of verbal communication (e.g. non-native speakers, shy people, children with learning disabilities). Furthermore, the study showed that the tabletop and physical-digital conditions resulted in more equitable participation, which led the authors to conclude that ‘where creativity and democracy are valued, then having tangible and easily accessible entry points within information and physical spaces can be an effective way of facilitating collaboration’ (Rogers et al., 2009, p. 106). On the other hand, collaborative tasks involving command and control systems require constraints so that not everything is accessible to everyone, thus facilitating division of work and assumption of roles. The use of a tangible tabletop interface in such a scenario, collaborative decision-making in maritime operations, is investigated by Scott et al. (2010). The need of military personnel (decision makers) to quickly respond in complex situations demands access to up-to-date information from different electronic data sources and the ability to share this information with other decision makers involved in the process. This need is addressed with the use of a multitouch tabletop and its rich interaction metaphors and direct and intuitive manipulation. Concordant with Szymanski et al. (2008), they found that multitouch interfaces in general lack the possibility to uniquely identify users — a feature that would support different operator roles along with corresponding security, interface personalization or tailoring, logging on a per-user basis, etc. In contrast to scenarios where people join together to collaborate on a task, the dynamics of interaction with a public multitouch display installation are expected to be different. Indeed, this is what Peltonen et al. (2008) found after examining data (e.g. log files, video footage) gathered from a large multitouch display installation in the center of Helsinki. They conclude: ‘In particular, we are thinking of established “norms” of conduct that apply to other “older” publicly available objects. We find it important to think about design separately for (small or large) groups of users versus individual users. Design should support performative acts and facilitate asymmetric and ad hoc role-taking, thus letting users learn the opportunities for interaction from their peers’ (Peltonen et al., 2008, p. 1294). The authors argue that this implies some kind of user awareness from the display itself. Finally, the possibilities of multitouch multi-user interaction are also being explored in typically single-user desktop-based business applications. Besacier (2011) motivates his multitouch tabletop adaptation of Microsoft Excel, the Tablexcel, with the need for co-located cooperative work on spreadsheets, often manifested in business environments. A straightforward application of user identification in such a scenario is granting read and write permissions based on the role/identity of the user.

In this section, we presented related research that motivated our work and explained *why* identity can be of paradigm importance in certain applications, while in the next section we review related work concerned with *how* user identification can be implemented on multitouch displays.

3. Related work

Identification can be based on an item the user has — a key or an ID card, on what he knows — a password or PIN and, finally, on who he is. The latter is represented by biometric identification methods which exploit the uniqueness of a person's particular characteristic. Some of the characteristics explored so far and used for identification include fingertips, voice, iris patterns, palm print, hand geometry and ECG signal. Along with physical characteristics, biometric methods also exploit behavioral patterns like gait or keystroke dynamics.

Our approach to user identification on multitouch displays can be viewed as a hand geometry based identification method that uses features extracted from a very limited set of input information. According to de-Santos-Sierra et al. (2011) and Dutağaci et al. (2008), other hand geometry based identification methods use a varied range of features including finger lengths, measurements along different axes (Euclidean distance), alignment of finger shapes and shape distance measurements, width, height and angle measurements, distorted patterns of the back of the hand, coefficients of FFT and DCT (Euclidean distance), elliptical model and fingertip/valley information, fusion of 3D and 2D hand geometry features, wavelet features, as well as others. Most of these features are extracted from a picture/scan of the user's hand, which is the input information required for methods that are based on these features. To the best of our knowledge, no hand geometry based identification method has ever been proposed with such a small set of required input information (coordinates of five touchpoints/fingertips) as the one proposed in this paper.

Some of the widely accepted criteria to evaluate the performance of biometric identification methods are False Acceptance Rate (FAR), False Rejection Rate (FRR) and classification accuracy. FAR is the percentage of erroneously identified users (e.g. an impostor is identified as one of the legitimate users), FRR represents the percentage of rejected identification attempts (e.g. a legitimate user is rejected by the system), while classification accuracy reports the percentage of correctly classified users. According to data from Bhattacharyya et al. (2009) presented in Table 1, we can see that the method presented in this paper (MTi), with its accuracy of 94.69%, FAR 1.4% and FRR 4.9%, offers a performance that is comparable to current state-of-the-art biometric identification methods. The last column in Table 1 shows the number of enrolled users in the database during testing; good performance combined with high scalability make fingerprint based identification one of the most widely adopted biometric methods.

Table 1
Performance of common biometric identification methods (data from Bhattacharyya et al. (2009)).

Method	FAR [%]	FRR [%]	Enrolled users
Fingerprint	2	2	25,000
Voice	2	10	30
Iris	0.94	0.99	1224
Keystroke	7	0.1	15
Hand geometry	2	2	129
MTi (this paper)	1.4	4.9	34

To enable a truly multi-user environment on a multitouch display, three capabilities are needed: user identification, user distinction and user tracking. User *identification* is the ability of the system to uniquely recognize a user, while user *distinction* is the ability to distinguish between different users interacting with the display, without knowing their exact identity. Once the users are identified or distinguished, a system capable of user *tracking* is always able to tell to whom each touchpoint belongs. Table 2 summarizes related work reviewed in this section and shows which of these capabilities are provided.

Currently available solutions to the problem of user identification for multitouch displays rely mostly on hardware — either these solutions work only on displays built with specific multitouch sensing technology or they impose the use of additional hardware.

DiamondTouch (Dietz and Leigh, 2001) can distinguish between four users by exploiting an array of antennas (each antenna transmits a unique signal embedded in the touch surface) and special seats that work as receivers. When a user touches the surface, a small signal is coupled from the antennas near the touch through his/her body to the receiver. This technology supports two-handed interaction and distinguishes between users. Schmidt et al. (2010) discuss the benefits of user identification (and hand detection) for multitouch interaction. They also present a prototype display augmented with an overhead camera. The camera tracks hands and identifies users based on the hand's contours. Instead of hand contours, Dohse et al. (2008) use skin color segmentation to distinguish and identify users with an overhead camera. Here the term identification is intended more loosely – in the case where two users are interacting, each on his/her side of a tabletop display, the system will be able to distinguish which user the touch belongs to. However, if the users swap positions, the system will fail to notice it. Ouellet et al. (2012) combine biometric identification (face recognition) and the Kinect motion sensor for identification and tracking of users interacting with a tabletop display. Similarly, Ackad et al. (2012) use the Kinect sensor positioned above the tabletop to track users that identify themselves with their mobile phones or tablets, a method designed not only to enable user identification but also to provide a content sharing mechanism.

Besides relying on additional hardware for user identification, some methods also require that the user wears the needed

Table 2
Overview of currently available identification methods for multitouch displays and their capabilities in terms of user identification (I), user distinction (D) and user tracking (T).

Reference	Hardware requirements/identification method.	I	D	T
Dietz and Leigh (2001)	DiamondTouch table/based on sensor signals.		*	*
Schmidt et al. (2010)	Overhead camera/hand contour recognition.	*	*	*
Dohse et al. (2008)	Overhead camera/skin color segmentation		*	*
Ouellet et al. (2012)	Kinect sensor/face recognition.	*	*	*
Ackad et al. (2012)	Kinect sensor and mobile phone/association with mobile phone.	*	*	*
Meyer and Schmidt (2010)	IdWristbands (IR emitting wristbands)/association with emitter.	*[a]	*	*
Hodges et al. (2007)	Microsoft Sur40 display and objects with fiducials/association with objects.	*[a]		
Marquardt et al. (2010)	Special gloves/association with gloves.	*[a]	*	*
Scott et al. (2010)	Anoto pen/association with pen.	*[a]	*	*
Siddalinga (2010)	Ultrasonic emitters/association with emitter.	*[a]	*	*
Schöning et al. (2008)	Mobile phone/association with phone.	*		
Kim et al. (2010)	/, password based.	*		
Dang et al. (2009)	Display capable of detecting finger orientation/based on finger orientation.		*	
Zhang et al. (2012)	Display capable of detecting finger orientation/based on finger orientation.	*[b]	*	*
MTi (this paper)	/, hand geometry based biometric.	*	*[c]	*[c]

[a] If the object (e.g. Anoto pen) needed is associated to and owned by a single person.
[b] The identity of the user is associated to his/her position around the table, which means that the user is not allowed to move freely around the table.
[c] In combination with the 'See Me, See You' method (Zhang et al., 2012).

hardware; IdWristbands (Meyer and Schmidt, 2010) identify users that wear a wristband with two IR diodes emitting a coded signal. Another hardware based hand detection solution are the fiduciary-tagged gloves presented by Marquardt et al. (2010). The gloves are equipped with fiducials that enable recognition of various parts of the hand like fingertips, palms, sides, etc. Hand and user detection is achieved in a similar fashion. Microsoft′s Sur40[1] (Hodges et al., 2007) table provides an analog means for identification of objects and users (if users are connected to objects). In a similar way, Scott et al. (2010) associated the unique id of an Anoto pen with a user and/or role. Another hardware based approach is the use of ultrasonic signals in combination with simple triangulation methods to localize the source (an emitter worn by the user) of the signal (Siddalinga, 2010). Schöning et al. (2008) use a mobile phone to authenticate users that interact with a large-scale multitouch wall in the scenario of a team operating the surface with different roles (and the accompanying access rights, authority levels, functionality, etc.).

Kim et al. (2010) discuss problems connected to authentication on multitouch surfaces like shoulder surfing (the problem of entering a PIN in full view of one or more observers) and propose some input mechanisms (e.g. Pressure-Grid, Shield-PIN, Color-Rings) that enable the entering of PINs and passwords on multitouch displays to occur more privately and thus with improved security. Related to the method presented in this paper is also Dang et al. (2009) method for hand distinction; it is intended for single user scenarios and does not perform user identification, but rather distinguishes if the user is touching the display with the left or the right hand. Similarly, Ewerling et al. (2012) proposed an approach for hand detection and hand distinction aimed at optical displays; hand detection is the ability to group touchpoints that are caused by fingers of the same hand. Sometimes knowing which of the current users is interacting with the display, without knowing his/her exact identity can be enough; 'See Me, See You' (Zhang et al., 2012) is a method for user distinction based on finger orientation.

In this section we reviewed work specifically related to our research, i.e. other identification methods for multitouch displays, as well as more generally related work such as common biometric identification methods and evaluation criteria. Next, we present and evaluate MTi — our method for identification on multitouch displays.

4. The MTi identification method

Our main goal was to determine whether it is possible to perform user identification on multitouch displays without additional hardware and without relying on a particular sensing technology. This posed a strict constraint on the information we were allowed to use, as the description of touchpoints with coordinates is the only information that all multitouch displays have in common. The main novelty of our approach is, therefore, the definition of features, based solely on touchpoint coordinates, which enable user identification (with an 'off-the-shelf' classifier). We named the identification method based on these features MTi (MultiTouch identification).

In this section, we will introduce MTi. First, we will describe the data used to develop and evaluate the method. Next, we will present the core of the method — the features used and how they were chosen. Finally, the choice of the classifier used in the method will be explained. Although the classifier will already be mentioned when selecting the features, we decided to present the MTi method in the above-mentioned order, as it follows the natural flow of events when MTi is used; when the user places his/her hand on the screen *data* is generated, this data is then transformed

[1] http://www.microsoft.com/en-us/pixelsense/default.aspx.

Table 3
Properties of the two datasets used for evaluation and development.

Dataset	MTiDB	Bosphorus
Number of users	34	100
Samples per user	6–12, average 9.4	3
Min class variance	875.17	212.66
Max class variance	26120.26	15634.33
Average class variance	7337.24	2122.26
Samples source	LLP multitouch display.	Flatbed scanner+fingertips extraction.
Filtering (manual)	deletion and correction of erroneously recorded samples.	choice of samples conveying to our method's hand pose restriction.

into MTi *features* and, based on these features, the *classifier* identifies the user.

4.1. Data

In our evaluation, two different datasets were used: the multitouch identification dataset (MTiDB) with fewer users and more samples per user and the Bosphorus hand database (Dutağaci et al., 2008) with more users and fewer samples per user.

Table 3 summarizes the properties of both databases. In this table, class variance is defined as the sum of variances across the features used (presented in Fig. 2) for a particular user.

We created the MTiDB dataset using our multitouch display (LED laser plane construction, see Çetin et al., 2009 for details) in the following manner. Each user was asked to comfortably stretch the fingers of his/her right hand apart as far as possible and then place the fingers on the display in the sequence thumb, index finger, middle finger, ring finger, little finger. The coordinates of the touchpoints detected were recorded in the database along with the ID of the user. Samples were collected from 34 users, 30 male and 4 female, between the ages of 18 and 39. Each user repeated the above-mentioned procedure 10 or more times, erroneously collected samples were either corrected or filtered out. For example, if the user accidentally touched the screen with the palm of the hand and four fingers, this sample was removed from the dataset, while the incorrect identification of fingers (e.g. if the user did not touch the display in the above explained sequence) was corrected. The fingers of such a sample were relabeled and the corrected sample added to the dataset. In total, 320 samples were collected. On average, each user in the dataset is represented by 9.4 samples with the least represented user having 6 samples and the most represented user having 12 samples.

The Bosphorus hand database consists of hand images obtained with a flatbed scanner. When collecting data, 6 images were obtained for each user; 3 images were scanned for the left and 3 images for the right hand. With 642 users, this database is, to the best of our knowledge, the largest database for hand geometry based identification. From these images, we extracted the required information for our identification approach — the positions of the fingertips. In the Bosphorus hand database, the users were not restricted to a specific hand pose, as was required of the users in our dataset, meaning that of the three images from a single user, the finger spacing could be far apart in one image and significantly closer in the remaining two images. Therefore, we manually selected 100 users, where all three samples had the fingers comfortably stretched apart and were thus appropriate for our identification method.

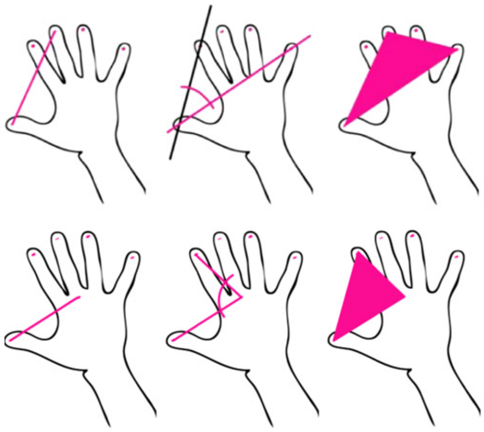


Fig. 1. Examples of three types of features: distances (left), angles (middle) and areas (right).

4.2. Features

In a 'latent semantics' fashion, we can say that our idea is to infer the geometry of a user's hand by observing the relative positions of the fingertips. The only condition is that the user touches the display with his fingers spread as far apart as possible (yet still comfortably).

We used three different types of features for identification: distances, angles and areas; we avoided the direct use of coordinates to prevent dependency on rotation and translation of the user's hand. In the top row of Fig. 1, we see how different types of features are extracted: 'distances' are the Euclidean distances between all possible combinations of two fingers (10 distances), 'angles' are measured between the line that is defined by the thumb and the index finger and the lines defined by all other pairs of fingers (9 angles) and 'areas' are the areas of possible geometric shapes defined by the fingers (10 triangles, 5 quadrilaterals and a pentagon). There are 35 features altogether. We also considered using features presented in the bottom row of Fig. 1 and used by Micire et al. (2011) for left/right hand identification and finger detection. Again, there are three types of features: distances between the centroid of the touchpoints and the touchpoints (5 distances), angles defined by two adjacent touchpoints with the vertex in

Table 4
Evaluation of identification accuracy for different combinations of feature types. Symbols used for representing different feature types are D for distances, ϕ for angles, △ for areas of triangles, ⏢ for areas of quadrilaterals and ⬠ for the area of the pentagram.

	Step	D	ϕ	Areas			Accuracy [%] (MTiDB)	Accuracy [%] (Bosphorus DB)
				△	⏢	⬠		
Features used	1	*					88.75	56.00
			*				80.63	77.33
				*			70.31	40.67
					*		64.69	13.67
						*	23.13	0.67
	2	*	*				92.81	91.33
		*		*			94.69	76.33
		*			*		92.50	71.33
		*				*	89.69	63.67
	3	*	*	*			**94.69**	**94.33**
		*		*	*		92.50	76.33
		*		*		*	93.75	76.67
	4	*	*	*	*		94.06	91.33
		*	*	*		*	94.69	94.33
	5	*	*	*	*	*	94.69	91.33

the centroid and the outer angle defined by the thumb and the little finger with the vertex in the centroid (5 angles) and areas of triangles defined by two adjacent touchpoints and the centroid (4 areas). These features resulted in lower classification accuracy and were omitted from further evaluation.

We used the following greedy procedure to evaluate combinations of different types of features. In step 1, we evaluated the accuracy for each feature type alone. In step 2, we evaluated the combination of two feature types — the best feature type from step 1 and a second type. In step 3, we considered three feature types — the best combination from step 2 and an additional feature type. Using the same principle, step 4 evaluates combinations of 4 feature types, while in step 5 identification accuracy is evaluated using all 5 feature types. The measure used was classification accuracy (10 fold cross validation; see Section 4.3 for details on the classifier used). Table 4 reports the results of using this procedure on our dataset; for comparison, we repeated the same steps on the Bosphorus hand database. Since both cases yielded the best accuracy when using a combination of distances, angles and areas of triangles, these features were used for further evaluation. The selected 29 features are presented in Fig. 2.

4.3. Classifier

To see whether it is possible to identify a user based on the features described above, we tried out different classifiers. Preliminary results showed that neural nets (95.31%), Bayesian networks (92.81%), random forests (94.69%) and support vector machines (94.69%) were all suitable for the task. Due to their resistance to over-fitting, their good results in terms of classification accuracy and their relatively short required training time, SVMs with a linear kernel (other kernels yielded a lower accuracy) were chosen for further evaluations of MTi. The choice of classifier was of secondary importance and finding the optimal classifier is outside the scope of this paper.

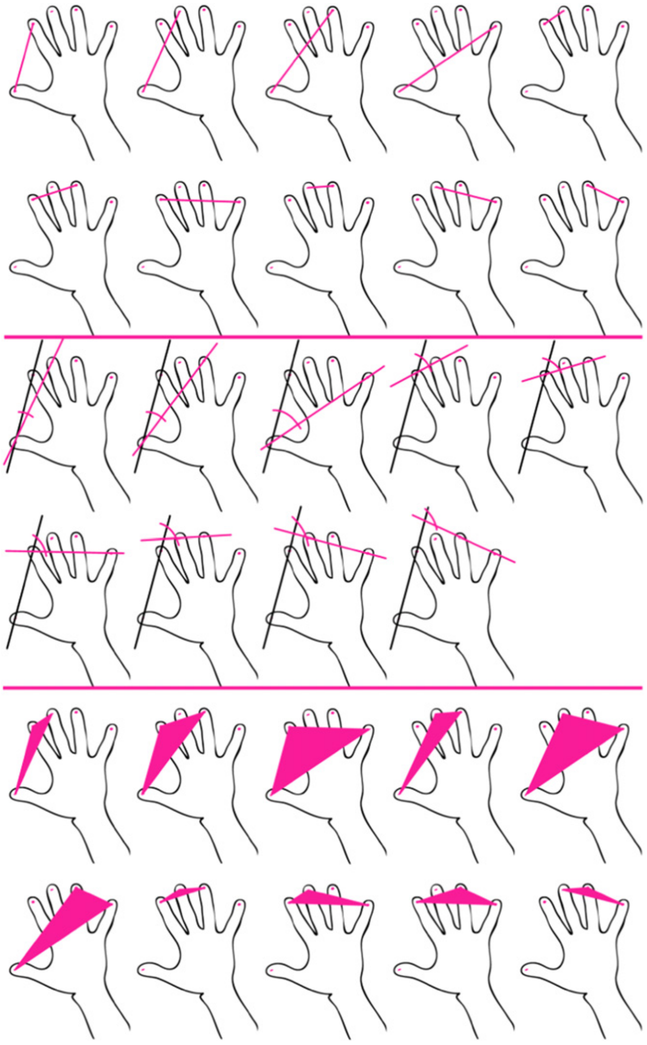


Fig. 2. Illustration of all of the 29 features used for identification: distances between pairs of fingers (top), angles between the line defined by the thumb and the index finger and lines defined by other fingers (middle) and areas of triangles defined by all possible combinations of three fingers (bottom).

Experiments were conducted in Weka (Hall et al., 2009), version 3.7, with the LibSVM (Chang and Lin, 2011) implementation of support vector machines. A one-versus-one (or all-versus-all) scheme was used for multi-class classification.

5. Evaluation of identification performance

To evaluate the identification performance of MTi, we used identification accuracy as the main measure of performance. Because our dataset was obtained on an actual multitouch display, we used it to evaluate our approach′s real world performance, while the larger Bosphorus hand database was used to get an approximation of how well our approach scales. Additional evaluations on subsets of our dataset were performed to understand the effect of data quantity and quality on classification accuracy. To compare our method to other identification methods, FAR and FRR were also calculated and are reported in Table 1.

5.1. Accuracy

All results were obtained with 10 fold cross validation[2] and are presented in Table 5. Because users in the Bosphorus hand database are represented with 3 samples, we also made an evaluation of our identification method on a subset of our dataset using only 3 samples per user. These samples were chosen so that the variance of each class (defined in Section 4.1) was minimized. Prior to classification, data was standardized to have the mean value 0 and a standard deviation of 1.

In Tables 6 and 7, we further investigate the effects of the quality and quantity of data on classification accuracy by choosing a subset (3, 4, 5 and 6 samples) for each user. These samples were chosen in two different ways: randomly and so that the class variance for each user was minimized. The reported accuracies represent the average accuracy of 10 runs.

When evaluating the effect of the quantity of data in the training set on classification accuracy, we used the selected subset to train the classifier and the remaining samples to evaluate the classifier. The results are presented in Table 6. We can see that classification accuracy increases with the increased amount of samples in the training set. Another thing to notice is the lower accuracy when the training set is composed of samples that minimize class variance. This is due to the fact that, by selecting the best samples, we create a training set that leads to a classifier less capable of generalization and, at the same time, leave the worst samples for testing.

When evaluating the effect of the quality of data in the training set on classification accuracy, we performed 10 fold cross validation on the selected subset. The results are presented in Table 7 and indicate that it is possible to obtain higher classification accuracy by imposing a variance threshold on the training data (during the process of user enrollment).

5.2. Accuracy vs database size

The scalability of MTi was tested on a subset of the Bosphorus hand database (described in Section 4.1). For each database size n, from the interval [2, 100] in steps of 2, identification accuracy was calculated as the average accuracy from 10 runs, where in each run identification accuracy was obtained from a 3 fold cross validation of n randomly chosen users. In this case 10 fold cross validation (used elsewhere in the paper) was not possible due to the small number of samples at small database sizes. For example, a database with 2 users consists of only 6 samples. For two users in the database, the accuracy was calculated as 98.33%. It then slowly drops to 97.67% for 10 users, 95.73% for 50 users and finally to 94.33% for a database with 100 users.

Table 5
Identification accuracy (10 fold cross validation) of our identification method on our dataset (MTiDB) and on the Bosphorus hand database (Bosphorus).

Dataset	DB size [users]	Number of samples	Average number of samples per user	Accuracy [%]
MTiDB	34	320	9.4	94.69
Bosphorus	100	300	3	94.33

Table 6
The effect of the *quantity* of data on classification accuracy (evaluated on MTiDB data).

Train on selected, test on removed

	Identification accuracy [%]			
Number of instances selected	3	4	5	6
Random	93.39	95.33	96.87	97.33
Min class variance	84.86	83.15	84.00	86.21

Table 7
The effect of the *quality* of data on classification accuracy (evaluated on MTiDB data).

Cross validation on selected

	Identification accuracy [%]			
Number of instances selected	3	4	5	6
Random	81.86	86.47	89.70	93.88
Min class variance	100	100	97.65	96.08

5.3. Usability evaluation

To see how MTi behaves in 'real life' and how users respond to it, we performed a usability study. We implemented the method and used it in an application that consists of four parts: enrollment, playground, balloon game and quiz game. The application was developed in the MT4j[3] framework (Laufs et al., 2010) and ran on a 40 inch Samsung SUR40 multitouch tabletop. MTi was implemented as described in the previous sections. Due to its size, the multitouch tabletop can accommodate four users simultaneously. Therefore, we performed the usability test with groups of four users. To assess the usability of MTi, we used post- and pre-test questionnaires along with data collected by the application.

[2] 'The standard way of predicting the error rate of a learning technique given a single, fixed sample of data is to use stratified tenfold cross-validation. The data is divided randomly into 10 parts in which the class is represented in approximately the same proportions as in the full dataset. Each part is held out in turn and the learning scheme trained on the remaining nine-tenths; then its error rate is calculated on the holdout set. Thus, the learning procedure is executed a total of 10 times on different training sets (each set has a lot in common with the others). Finally, the 10 error estimates are averaged to yield an overall error estimate.' (Hall et al., 2009, p. 153).

[3] http://www.mt4j.org/.

5.3.1. Enrollment

During enrollment, each user was presented with a component (seen in Fig. 3, left) with which he/she entered his/her username, password and MTi 'hand' samples. The component also associated each user with a color, a unique ID, which allowed the users to see all the entered samples and reported the variance of the samples. While gathering data for the development of the MTi method (described in Subsection 4.1), we noticed that many samples were erroneously stored and had to be manually corrected; the problem occurred due to the fact that users had to place their fingers on the table in the sequence thumb, index finger, middle finger, ring finger, little finger and this procedure was prone to error. For the usability study, we developed an algorithm that automatically detects which touchpoint represents which finger. The algorithm works on the premise that the fingers are placed on the display as requested by the method while comfortably stretched apart. First, the thumb is identified as the finger most distant to the other four. It then looks for the touchpoint closest to the thumb and marks it as the index finger, then marks the touchpoint closest to the index as the middle finger and so on. In case an error still occurred, or the user accidentally placed his/her fingers on the component, the user was able to enter a new sample without storing the incorrect one. When the user entered all of the required data (username, password and samples) he/she pressed finish and the enrollment was over. When all four users completed the enrollment, their samples were gathered and the application built an SVM identification model used in the other three parts of the application. The application also stored the parameters (mean value and variance for each feature) used to standardize the samples so that new samples (entered during identification) could be standardized in the same way.

5.3.2. Playground

After enrolling, the participants were given some time to get accustomed to identification using the MTi method. Four identification components appeared on the table, one for each participant. The identification component can be seen in action in Fig. 3, right; when idle, the component is a gray rectangle, but when a user places his/her fingers on it the component identifies the user and signals his/her identity by displaying the user's name and coloring the border of the component using the color associated with the identified user. With its visuals, this component enforces a social protocol between users, which prevents errors arising if two users place their fingers close together.

5.3.3. Balloon game

This part of the application was designed to simulate a scenario in which the speed of identification is crucial and to compare MTi identification to a baseline identification method. We chose a password based identification on an on-screen keyboard as the baseline as it is the only identification method currently available on all multitouch displays. The balloon game is a single player game where the goal is to capture as many balloons as possible before they fly off the screen. There are 8 balloons in each level and at least 2 must be captured to advance to the next level. With each level, the speed of the balloons increases. The game ends when the participant fails to capture the minimum amount of balloons in that level. To capture the balloon, the participant has to tap on it. Half of the balloons in each level require authentication to be captured. The time from when the user taps on the balloon (and the balloon requests authentication) and when the authentication is finished, is recorded. Of the four participants in each group, two used password identification while the other two used MTi identification.

5.3.4. Quiz game

The quiz game is the fourth and last part of the application developed in the usability test of the MTi method. It is intended as a use case of the method and a test of the method in a collaborative scenario. The group of participants was prompted with a question and three answers were displayed on the screen. To answer the question, a participant had to place his/her hand on the answer he/she thought was correct. The answer would then turn either green or red depending on if it was correct or incorrect and the color of the border of the answer gave feedback on which user was identified by the system (Fig. 3, middle). The identified user was assigned two points for a correct answer and one negative point for an incorrect answer. Each question could only be answered once (by the fastest participant). There were a total of 20 questions in the quiz.

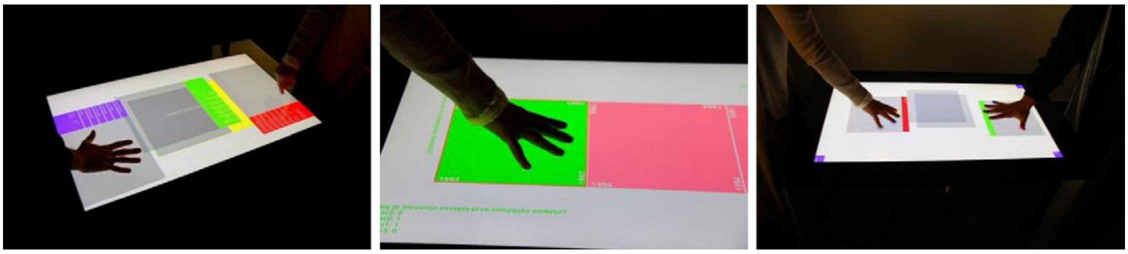


Fig. 3. Parts of the application used to test MTi: enrollment (left), quiz game (middle) and playground (right). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

5.3.5. *Usability assessment*

The usability of the MTi identification method was evaluated with the System Usability Scale (SUS) score (Brooke, 1996). It is a widely adopted and studied (Lewis and Sauro, 2009; Sauro, 2011) usability measure and consists of 10 Likert type items (see Table 9). The questionnaire was answered after the users finished all four parts of the test. Prior to the test, participants filled out a short background survey (age, gender, educational background and experience with multitouch displays). Those participants who used password based identification in the balloon game filled out two SUS questionnaires, one for each identification method (in the quiz game all participants used the MTi method). Along with the SUS score, methods were also compared using the time it takes to authenticate. At the end of the questionnaire, the users were given the possibility to leave a comment. In total, 18 participants, 15 male and 3 female, between the ages of 23 and 37 took part in the usability study. Most of them reported previous experience with multitouch devices such as smartphones and tablets, but only one had experience interacting with a bigger multitouch display. All of them have a background in engineering or natural sciences. Participants were randomly assigned to four groups of four and two participants repeated the experiment to accommodate the remaining two participants (but filled out the questionnaire only once, following their first trial).

6. Results

The results of the usability evaluation are summarized in Table 8. We can notice that password based identification received a below average[4] SUS score (53.75), while the MTi method received an above average SUS score (79.03). Furthermore, as expected, MTi is also considerably faster than password based identification; it takes, on average, 1.5 s to authenticate using MTi and 7.7 s to enter a password. The latter is also dependent on the length of the password and would be significantly reduced if the participants were more experienced in interacting with big multitouch displays. This type of experience would bring the average level achieved in the Balloon game under the password condition closer to the one achieved under the MTi condition.

A more detailed look at the items in the SUS questionnaire (Table 9) reveals that, in general, the participants liked the MTi method and found it easy to use (item 1, 3 and 9) as well as easy to learn (items 4, 10 and 7). Some of the items in the questionnaire do not have relevant meaning; for example item 5 ‘I found the various functions in this system were well integrated’ has no meaning as the system is only composed of one function, i.e. identification. The items where MTi and password identification differ the most are items 3, 1 and 8. This means that participants found MTi to be easier to use and would like to use it frequently, while they found password identification to be cumbersome.

Table 8
Average SUS score, average identification time and average level achieved for both identification methods involved in the usability study.

Identification method	MTi	Password
Number of participants	18	8
Average SUS score	79.03	53.75
Average time spent during authentication	1526 ms	7771 ms
Average level achieved in the Balloon game	9	4

Comments left by the participants offer additional insight into the usability of the method and can be divided into 3 groups:

- general appreciation of the MTi method (‘I prefer MTi identification as it is quicker and you do not need to remember anything.’, ‘I found the MTi method quicker and simpler, while the password method is more secure.’),
- problems or lack of experience with multitouch hardware (‘If the hardware would be 100% reliable and responsive, MTi would be really great.’, ‘A technician would not help, but a physical keyboard would.’, ‘Bad votes due to hardware non-responsiveness.’).
- and suggestions for improvements or use cases (‘It would be interesting to see if using the whole palm would be easier and more accurate.’, ‘Identification with the whole hand would be more appropriate, more robust and would yield more consistent results.’, ‘The system is great for games!’, ‘MTi would be useful for applications where a middle level of security is required — access to a certain type of file or GUI personalization or access to bookmarks. For modification of these resources, a password would still be required. In this ‘soft security’ mode, all interaction would be logged and in case of a security breach, the owner could trace back the impostor's activity.’).

7. Discussion

First and foremost, the results of the experiments presented show that all multitouch displays that are big enough to accommodate a human hand and can detect 5 or more touchpoints are capable of user identification. The identification method we presented, the MTi, works on coordinates of touchpoints — data common to all multitouch displays. The method presented was also warmly accepted by participants during the usability study.

It is interesting to notice, that between the two datasets that were evaluated, the class/user variance is lower when using the Bosphorus database, which was recorded with no restrictions regarding hand placement. It is therefore reasonable to conclude that our restriction to place the hand on the display with the fingers comfortably apart is not too stringent. This assumption was implicitly confirmed by the usability study, as no participant placed any complaint regarding the

[4] The average SUS score is 68 (http://www.measuringusability.com/sus.php).

Table 9
Detailed overview of the SUS scores for both identification methods. For each item participants answered on a scale from 1 to 5, where 5 means they strongly agree and 1 that they strongly disagree with the statement.

Item	Question	Average score	
		MTi	Password
1	I think that I would like to use this system frequently.	3.89	1.88
2	I found the system unnecessarily complex.	1.39	3.00
3	I thought the system was easy to use.	4.72	2.29
4	I think that I would need the support of a technical person to be able to use this system.	1.72	1.13
5	I found the various functions in this system were well integrated.	3.50	2.75
6	I thought there was too much inconsistency in this system.	2.33	3.00
7	I would imagine that most people would learn to use this system very quickly.	4.83	3.75
8	I found the system very cumbersome to use.	2.00	3.88
9	I felt very confident using the system.	3.39	3.38
10	I needed to learn a lot of things before I could get going with this system	1.28	1.25

restriction. The lower variance in the Bosphorus database also offers an explanation for the similar identification accuracy reported for both datasets despite the difference in dataset size. Furthermore, this leads to possibilities for applying variance thresholds to acquired data in order to further improve accuracy. Finally, a possible and reasonable explanation as to why the variance of our dataset is higher is the inaccuracy of the LLP sensing technology of our multitouch display.

The success of MTi is a result of the features used (although an in-depth analysis of the feature space is beyond the paper's scope as well as the effects of feature selection on accuracy). We see the first two types of features, distances and angles, to be a sort of transformation from the coordinate system of the screen where the touchpoint coordinates are recorded, to a polar hand centric coordinate system. This allows the method to be invariant to the position and rotation of the user's hand. The other feature type, namely ‘areas’, brings hand geometry inference to the method. The fact that these features are extracted from multiple fingers, makes them inherently susceptible to the geometry of the whole hand.

From Tables 6 and 7, we can see that MTi is also susceptible to the quality of both training and test data and the size of the training set. Choosing only the best few samples for each user dramatically improves identification accuracy, while a smaller, but still significant, increase in identification accuracy can be noticed with an increase of samples in the training set. These facts provide two interesting directions for future improvements of the method.

The evaluations conducted on the Bosphorus database showed that by increasing the number of users enrolled in the database, identification accuracy drops slightly to 94.33% for 100 users. Considering that in most of the possible use cases for MTi the number of users expected is below 100, we can say that the method scales well. The only use case in which MTi's scalability might become an issue is in a public interaction display scenario. Another application area where MTi's scalability might seem problematic is education where a single display might be shared by the whole educational institution. However, the database could be divided into smaller parts, where each part would represent a meaningful subgroup of users that always interact with the display simultaneously, e.g. a class.

In Section 3, we mentioned three basic capabilities that make a multitouch display a multi-user device: user identification, distinction and tracking. These capabilities are also present in the design criteria for multitouch identification methods proposed by Siddalinga (2010):

- support for a wide variety of multitouch hardware (underlying sensing technology as well as configuration — horizontal vs vertical),
- support for user mobility (the ability to maintain the same user association even if the user switches sides or moves to a different part of the display),
- support for user authentication (to protect data and territories from unauthorized access) and
- continuous user tracking (the ability to detect the user for each touch he/she makes).

Due to the restrictions imposed by hardware, some of the currently available multitouch identification methods are only applicable to horizontal/tabletop displays as they require an overhead camera, or, in the case of DiamondTouch, a seat. Other methods, like the fiduciary gloves, have no problems with display orientation or user mobility but are lacking in terms of authentication capabilities and/or require the use of special, sometimes cumbersome, hardware. To the best of our knowledge, there is no multitouch user identification method that completely meets the above-mentioned design criteria. Neither does MTi. It supports most multitouch hardware, user mobility and authentication, but provides only very limited continuous user tracking. For example, after user A places his/her hand on the display for identification, the fingers from the hand used are tracked and associated with user A only until they leave the display. Another way to partly overcome the problem of continuous user tracking is to conclude actions or handling of items that require identification with an identification gesture (placing five fingers on the display); for example, after dragging an item on the display to a new position, the user would be prompted to identify himself/herself by placing his/her fingers on the display. A more sophisticated solution would be to combine MTi with the 'See Me, See You' technique (Zhang et al., 2012) in the following manner: the method 'See Me, See You' distinguishes between users based on the orientation of the touch. The method was tested for the condition that all the users touch the tabletop with the right index finger, although the authors report that the method could be generalized to all fingers. Another, more strict condition, is that the users cannot move freely around the table; the method enables a particular touch to be associated with the position of the user around the table not the user himself/herself. If two users swap positions, the system fails to notice it. In the cited paper, the authors address this problem using an interface component called the Position avatar. The Position avatar links a user with a position; if a user changes position, he/she must move his/her avatar to the new position. MTi could be used as a replacement or implementation of the Position avatar, enhancing it with identification capabilities. We can see that MTi and 'See Me, See You' complement each other; MTi adds identification capabilities (role taking, personalization, security…) to 'See Me, See You', while 'See Me, See You' adds continuous user tracking to the MTi method.

Regarding the authentication capabilities of MTi, it has to be noted that this is not a method to be utilized for high security demands. Due to the fact that the method operates on the coordinates of the touchpoints, it is possible for a user to position his/her fingers (or other objects) on the display in a way that would make the method recognize him/her as someone else. But not all use cases are of a high security nature; for example, when using identification for interface personalization, it is unlikely that a user would be interested in stealing someone else's personalization settings. Games are another field were high security is not demanded and where the possibility of 'cheating' could also be intentionally exploited by players as well as game designers. Eventually, as some participants in the usability study suggested, the method could be improved by allowing the user to place the entire palm on the display. This would make the method even more reliable and secure.

8. Conclusion: contribution and future work

Our work showed that it is possible to provide user identification capabilities on every multitouch display that is large enough to accommodate a human hand and can detect 5 or more touchpoints, independently from the display's underlying sensing technology, as our identification method (MTi) relies only on the coordinates of the touchpoints. With regard to data gathered from 34 users on an LLP multitouch display, MTi reported an accuracy of 94.69%. Using data extracted from a selection of 100 users from the Bosphorus hand database (3 samples per user), the scalability of our method (94.33% accuracy on 100 users) was also shown. Finally, an evaluation and trial run with users showed that the usability of the MTi identification method is above average (SUS score 79.03).

Possibilities for future work include, but are not limited to: finding the optimal classifier and fine-tuning its parameters for further improvements in identification performance, exploring feature selection algorithms to find the optimal set of features, considering a scheme that takes into account a certain threshold based on class (user) variance for data acquisition during both enrollment and verification, a study on how our method performs on multitouch displays with different sensing technology and adaptation of the method so that users can place the entire hand on the display during identification.

Acknowledgments

This research was funded in part by the European Union, European Social Fund, Operational Program for Human Resources, Development for the Period 2007–2013 and by the Slovenian Research Agency.

The authors would like to thank their colleagues who, by 'lending a hand', contributed to the multitouch identification database MTiDB, the participants in the usability study and Bulent Sankur at the Boğaziçi University for kindly granting us access to their database. Bojan would specially like to thank Jolanda Modic, for help with clean-up of the datasets and motivation.

References

Arnell, O., Björk, A., Dahlbäck, N., Pennerup, J., Prytz, E., Wikman, J., 2008. Infotouch: an explorative multi-touch visualization interface for tagged photo collections. In: Proceedings of the 5th Nordic Conference on Human-Computer Interaction: Building Bridges, pp. 491–494, http://dx.doi.org/10.1145/1463160.1463227.

Ardaiz, O., Arroyo, E., Righi, V., Galimany, O., Blat, J., 2010. Virtual collaborative environments with distributed multitouch support. In: Proceedings of the 2nd ACM Sigchi Symposium on Engineering Interactive Computing Systems, pp. 235–240, http://dx.doi.org/10.1145/1822018.1822055.

Ackad, C., Clayphan, A., Maldonado, R.M., Kay, J., 2012. Seamless and continuous user identification for interactive tabletops using personal device handshaking and body tracking. In: Proceedings of Extended Abstracts of CHI 2012: ACM Annual Conference on Human Factors in Computing Systems. Austin, TX, May 5–10, 2012, pp. 1775–1780.

Buisine, S., Besacier, G., Aoussat, A., Vernier, F., 2011. How do interactive tabletop systems influence collaboration? Computers in Human Behavior 2011, 49–59. http://dx.doi.org/10.1016/J.CHB.2011.08.010.

Besacier, G., 2011. Tablexcel: a multi-user, multi-touch interactive tabletop interface for microsoft excel spreadsheets. In: Proceedings of the 13th IFIP TC 13 International Conference on Human-Computer Interaction — volume part IV (interact'11), pp. 366–369.

Bhattacharyya, D., Ranjan, R., Alisherov, A.F., Choi, M., 2009. Biometric authentication: a review. Biometric Technology Today 2 (3), 13–28 Retrieved from:⟨http://www.sersc.org/journals/ijunesst/vol2_no3/2.pdf⟩.

Brooke, J., 1996. SUS: a "quick and dirty" usability scale. In: JorDan, P.W., Thomas, B., Weerdmeester, B.A., Mcclelland (Eds.), Usability Evaluation In Industry. Taylor & Francis, London, UK, pp. 189–194.

Çetin, G., Bedi, R., Sandler, S., 2009. Multi-Touch Technologies, 1st ed. Retrieved from http://www.nuigroup.com.

Chang, C.-C., Lin, C.-J., 2011. Libsvm: a library for support vector machines. ACM Transactions on Intelligent Systems And Technology 2 (27), 1–27. http://dx.doi.org/10.1145/1961189.1961199 Software available at: ⟨http://www.csie.ntu.edu.tw/~cjlin/libsvm⟩.

de-Santos-Sierra, A., Sánchez-Ávila, C., Bailador Del Pozo, G., Guerra-Casanova, J., 2011. Unconstrained and Contactless Hand Geometry Biometricssensors 11 (11), 10143–10164. http://dx.doi.org/10.3390/S111110143.

Dutağaci, H., Sankur, B., Yörük, E., 2008. Comparative analysis of global hand appearance-based person recognition. Journal of Electronic Imaging 17, 011018.

Dietz, P.H., Leigh, D.L., 2001. Diamondtouch: a multi-user touch technology. In: Proceedings of ACM Symposium on User Interface Software and Technology, UIST, pp. 219–226.

Dohse, K.C., Dohse, T., Still, J.D., ParkHurst, D.J., 2008. Enhancing multi-user interaction with multi-touch tabletop displays using hand tracking. In: Proceedings of the First International Conference on Advances in Computer-Human Interaction (ACHI'08), pp. 297–302, http://dx.doi.org/10.1109/achi.2008.11.

Dang, C.T., Straub, M., Andre, E., 2009. Hand distinction for multi-touch tabletop interaction. In: Proceedings of the 2009: ACM International Conference on Interactive Tabletops and Surfaces 2009, Banff, Ab, November 23–25 2009, pp. 101–108.

Ewerling, P., Kulik, A., Froehlich, B., 2012. Finger and hand detection for multi-touch interfaces based on maximally stable extremal regions. In: Proceedings of the 2012 ACM International Conference on Interactive Tabletops and Surfaces (ITS'12). ACM, New York, NY, USA, pp. 173–182. http://dx.doi.org/%2010.1145/2396636.2396663 http://doi.acm.org/10.1145/2396636.2396663.

Higgins, S., Mercier, E., Burd, E., Joyce-Gibbons, A., 2011. Mulit-touch tables and collaborative learning. British Journal of Educational Technologyhttp://dxdoi.org/10.1111/J.1467-8535.2011.01259.X.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H., 2009. The WEKA Data Mining Software: An Update. SIGKDD Explorations, 11; .

Hodges, S., Izadi, S., Butler, A., Rrustemi, A., Buxton, B., 2007. Thinsight: versatile multi-touch sensing for thin form-factor displays. In: Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology, UIST'07, Newport, Rhode Island, USA, October 07–10, 2007. ACM, NY, pp. 259–268.

Jacucci, G., Morrison, A., Richard, G.T., Kleimola, J., Peltonen, P., Parisi, L., Laitinen, T., 2010. Worlds of information: designing for engagement at a public multi-touch display. In: CHI'10 Proceedings of the 28th International Conference on Human Factors in Computing Systems, pp. 2267–2276, http://dx.doi.org/10.1145/1753326.1753669.

Jordà, S., Geiger, G., Alonso, M., Kaltenbrunner, M., 2006. The reactable: exploring the synergy between live music performance and tabletop tangible interfaces. In: Proceedings of the 1st International Conference on Tangible and Embedded Interaction (2007), pp. 139–146, http://dx.doi.org/10.1145/1226969.1226998.

Krueger, M.W., Gionfriddo, T., Hinrichsen, K., 1985. VIdeoplace — an artificial reality. In: Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI'85), pp. 35–40.

Kim, D., Dunphy, P., Briggs, P., Hook, J., Nicholson, John, Nicholson, James, Olivier, P., 2010. Multi-touch authentication on tabletops. In: Proceedings of the 28th International Conference on Human Factors in Computing Systems, Chi 10.

Kim K., Kulkarni T., Elmqvist N., 2009. Interaction workspaces: identity tracking for multi-user collaboration on camera-based multi-touch tabletops. In: Proceedings of the Workshop on Collaborative Visualization on Interactive Surfaces, Retrieved from: 〈https://www.engineering.purdue.edu/~elm/projects/iwspaces/iwspaces.pdf〉.

Laufs, U., Ruff, C., Zibuschka, J., 2010. Mt4j — A Cross-Platform Multi-Touch Development Framework. ARXIV Preprint Arxiv10120467 Abs/1012.0, pp. 52–57.

Lewis, J.R., Sauro, J., 2009. The factor structure of the system usability scale. In: Proceedings of the Human Computer Interaction International Conference (HCII 2009), San Diego, CA, USA.

Mehta, N., 1982. A Flexible Machine Interface. M.A.Sc. Thesis. Supervised by Professor K.C. Smith. Department of Electrical Engineering, University of Toronto.

Morris, M.R., Huang, A., Paepcke, A., Winograd, T., 2006. Cooperative gestures: multi-user gestural interactions for co-located groupware. In: Proceedings of the Sigchi Conference on Human Factors in Computing Systems, pp. 1201–1210, http://dx.doi.org/10.1145/1124772.1124952.

Meyer, T., Schmidt, D., 2010. Idwristbands IR-based user identification on multi-touch surfaces. In: Proceedings of ACM International Conference on Interactive Tabletops and Surfaces, pp. 277–278.

Marquardt, N., Kiemer, J. Greenberg, S., 2010. What caused that touch?: Expressive interaction with a surface through fiduciary-tagged gloves. In: Proceedings of ACM international Conference on Interactive Tabletops and Surfaces, ITS 10, pp. 139–142.

Micire, M., Mccann, E., Desai, M., Tsui, K.M., Norton, A., Yanco, H.A., 2011. Hand and finger registration for multi-touch joysticks on software-based operator control units. In: Proceedings of IEEE Conference on Technologies for Practical Robot Applications TEPRA 2011, pp. 88–93.

Nakatani, L.H., Rohrlich, J.A., 1983. Soft machines: a philosophy of user-computer interface design. In: Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 83), pp. 12–15.

Ouellet, J.-N., Harvey, E.R., Echevarria, J., Franck, G., Scott, S.D., 2012. Computer vision application using the kinect sensor for the identification and tracking of users interacting with a surface computing platform. In: Presentation at the 80th Congress of the Association Francophone Pour La Savour, (ACFAS), Montreal, PQ, May 7–11, 2012.

Partridge, G.A., Irani, P.P., 2009. Identtop: a flexible platform for exploring identity-enabled surfaces. In: Proceedings of CHI 2009 Extended Abstracts, pp. 4411–4416, http://dx.doi.org/10.1145/1520340.1520675.

Peltonen, P., Kurvinen, E., Salovaara, A., Jacucci, G., Ilmonen, T., Evans, J., Oulasvirta, A., Saarikko P., 2008. It's mine, don't touch!: interactions at a large multi-touch display in a city centre. In: Proceeding of the Twenty-Sixth Annual Sigchi Conference on Human Factors in Computing Systems, pp. 1285–1294, http://dx.doi.org/10.1145/1357054.1357255.

Ryall, K., Esenther, A., Everitt, K., Forlines, C., Morris, M.R., Shen, C., Shipman, S., 2005. Idwidgets: parameterizing widgets by user identity. In: Proceedings of Interact, pp. 1124–1128, http://dx.doi.org/10.1007/11555261_120.

Rogers, Y., Lim, Y.-K., Hazlewood, W.R., Marshall, P., 2009. Equal opportunities: do shareable interfaces promote more group participation than single users displays? Human-Computer Interaction 24 (1), 79–116.

Schmidt, D., Ki Chong, M., Gellersen, H., 2010. Handsdown: hand-contour-based user identification for interactive surfaces. In: Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries (Nordichi 10), pp. 432–441, http://dx.doi.org/10.1145/1868914.1868964.

Siddalinga, P.R., 2010. Sensor augmented large interactive surfaces. M.A.Sc. Thesis. Iowa State University. Retrieved from: 〈http://archives.ece.iastate.edu/archive/00000599/01/thesis.pdf〉.

Schöning, J., Rohs, M., Kruger, A., 2008. Using mobile phones to spontaneously authenticate and interact with multi-touch surfaces. In: Proceedings of Workshop on Designing Multitouch Interaction Techniques for Coupled Public and Private Displays, pp. 41–45.

Scott, S.D., Allavena, A., Cerar, K., Franck, G., Hazen, M., Shuter, T., Colliver, C., 2010. Investigating tabletop interfaces to support collaborative decision-making in maritime operations. In: Proceedings of ICCRTS 2010: International Command and Control Research and Technology Symposium, Santa Monica, CA, USA, June 22–24, 2010.

Szymanski, R., Goldin, M., Palmer, N., Beckinger, R., Gilday, J., Chase, T., 2008. Command and control in a multitouch environment. In: Paper Presented at the 26th Army Science Conference, Orlando, Flori da.

Sauro, J., 2011. A Practical Guide to the System Usability Scale. Denver, Co. Createspace.

Zhang, H., Yang, X.-D., Ens, B., Liang, H.-N., Boulanger, P., Irani, P., 2012. See me, see you: a lightweight method for discriminating user touches on tabletop displays. In: Proceedings of CHI 2012: ACM Annual Conference on Human Factors in Computing Systems. Austin, TX, MAY 5–10, 2012, pp. 2327–2333.

3 HDCMD: a Clustering Algorithm to Support Hand Detection on Multitouch Displays

In this chapter, the manuscript titled "HDCMD: a Clustering Algorithm to Support Hand Detection on Multitouch Displays" by Bojan Blažica, Daniel Vladušič and Dunja Mladenić is presented. A shortened version of this manuscript (Blažica et al., 2013) appears in the Lecture Notes in Computer Science (LNCS) series in the proceedings of the SouthCHI 2013 conference (due to space restrictions).

The paper addresses the problem of hand detection on multitouch displays. Currently, most multitouch displays are able to detect multiple touchpoints, but lack the ability to group separate touchpoints into hands. This also means that the display cannot detect how many hands are present. In the literature several use cases for hand detection have been identified and some solutions proposed. Mostly these solutions rely on additional hardware or are restricted to multitouch displays of a certain type, for example optical displays. The paper provides an overview of the existing body of work concerning hand detection and proposes a simplistic, yet in some cases efficient enough approach based on clustering.

In terms of context-awareness we are again, as with the MTi method presented in Chapter 2, exploiting the rich data provided by multitouch displays to extract additional contextual information from interaction. The aspect of context being extracted is the co-location of different people interacting with the display. Some use cases for this information are presented in the reviewed literature in the paper.

The authors together conceived the idea for the algorithm, while the first author carried out the needed experiments and wrote the majority of the paper.

HDCMD: a Clustering Algorithm to Support Hand Detection on Multitouch Displays

Bojan Blažica,[1,2] Daniel Vladušič,[1] and Dunja Mladenić[2]

[1] XLAB Research, Pot za Brdom 100, SI-1000 Ljubljna, Slovenia
{bojan.blazica, daniel.vladusic}@xlab.si
[2] Jožef Stefan International Postgraduate School, Jamova 39, SI-1000 Ljubljana, Slovenia
dunja.mladenic@ijs.si

Abstract. This paper describes our approach to hand detection on a multitouch surface i.e. detecting how many hands are currently on the surface and associating each touch point to its corresponding hand. Our goal was to find a general software-based solution to this problem applicable to all multitouch surfaces regardless of their construction. We therefore approached hand detection with a limited amount of information: the position of each touch point. We propose HDCMD (Hand Detection with Clustering on Multitouch Displays), a simple clustering algorithm based on heuristics that exploit the knowledge of the anatomy of the human hand. The proposed hand detection algorithm's accuracy evaluated on synthetic data (97%) significantly outperformed XMeans (21%) and DBScan (67%). We also elaborate on the dependencies between display size, hand detection accuracy and the number of hands on the display.

Keywords. Hand Detection; Multitouch; Clustering; Co-located Groupware.

1 Introduction

Intuitiveness and directness - properties associated with multitouch interaction since the advent of the first multitouch displays in the early 80s [1]. Intuitive, because they allow us to interact with digital objects similarly to the way we interact with physical objects in our everyday life. Direct, because the display represents the input and the output of the computer thus enabling us to manipulate directly with what we see. Furthermore, multitouch displays greatly increase the possibilities of interaction between humans and computers. Whereas the ubiquitous WIMP paradigm (windows, icons, menus, pointer) only offers to the user one cursor that moves in two dimensions, multitouch extends this to up to ten cursors per user. Moreover, by detecting an unlimited number of fingers, support for multiple users working simultaneously on the same surface is implied. A recent survey [2] conducted among multitouch interaction researchers showed that multiuser support is one of the top features of multitouch displays. According to their survey, Benko et al. defined an interactive tabletop as "a large surface that affords direct, multi-touch, multi-user interaction". However, despite the ability of sensing multiple touchpoints, a multitouch display cannot be regarded as a multiuser device per se, as it does not distinguish between different users, nor is it capable of assuming how many users are operating with it or even how many hands are currently on the screen. We can thus say that multitouch displays are only finger-aware but not hand-aware or user-aware. To overcome this, we need a way to group touches (fingers) into hands and to associate hands with users. This higher-level information broadens the possibilities of interaction with a multitouch display and transforms it into a multiuser device. Even when the display is used by a single user, hand detection could improve interaction by complying to design guidelines that stem from the nature of bimanual interaction [3, 4].

In this paper we present an incremental clustering algorithm, which makes any multitouch display hand-aware. In Section 2, we motivate our work with a review of literature related to hand-aware interaction on multitouch displays followed by an overview of hand detection methods and related work in Section 3. We continue in Section 4 with a description of general characteristics of clustering for hand detection and the definition of our clustering algorithm. Section 5 presents the methods and results of the evaluation on synthetic data, which is then discussed in Section 5.2. We conclude with a concise definition of the contribution of our work.

2 Motivation

The need for and the positive effects of multiuser support in multitouch displays have been examined by various studies. Researchers have been concerned with the impact of multitouch interaction on group behavior. On the other hand, investigations about the nature of tabletop interaction in the physical world have been conducted to gain valuable insight on how to design digital tabletop interaction with a multitouch display. In this section we present some of these studies that motivated us in our work.

Ringel Morris et. al. [5] explored the potential of collaborative gestures for co-located groupware. They formalized the notion of cooperative gestures as “interactions where the system interprets the gestures of multiple group members collectively in order to invoke a single command”. Their conclusion was that the “use of cooperative gestures can add value to applications as a means of increasing participation, drawing attention to important commands, enforcing implicit access control, facilitating reach on large surfaces, and/or enhancing social aspects of an interactive experience.” In such a scenario, hand detection is a crucial requirement for the development of groupware interaction.

When designing natural user interfaces (e.g. multitouch interaction, freehand gestures) the underlying paradigm is that interaction with digital objects should resemble interaction with physical objects as much as possible. Terrenghi et al. [6] observed that given a task, participants divided it in the same subtask when performing it in the physical, as well as in the digital domain, although they eventually performed these subtasks differently - most of the bimanual interaction present in the physical domain was lost. The authors conclude that instead of blindly copying real world interaction, we should always think about how a certain task can optimally be performed in the digital domain. Furthermore, this suggests that in the digital world there may be another kind of bimanual interaction, which must not only be supported but also stimulated by an application. Hand detection could thus represent a valuable tool to achieve this.

Rogers et al. [7] performed an experiment where groups of three people were asked to complete a garden planning task under three different conditions: using a laptop, using a tabletop or in a physical-digital setup (by manipulating physical objects on a tabletop). One key finding was that tangibility and accessibility stimulate more participation from those who find it hard to talk or are incapable of verbal communication (e.g. non-native speakers, shy people, and children with learning difficulties). Furthermore, the study showed that the tabletop and physical-digital conditions resulted in a more equitable participation which led the authors to the conclusion that "where creativity and democracy are valued, then having tangible and easily accessible entry points within information and physical spaces can be an effective way of facilitating collaboration." On the other hand, collaborative tasks involving command and control systems require constraints, so that not everything is accessible to everyone, thus facilitating division of work and assumption of roles. In the first case hand and user detection can be applied to stimulate non-verbal communication between users, while in the second it is required to allow the adoption of roles.

If the previously reviewed scenarios, in which hand detection is beneficial, are of a more general nature, the interaction techniques described in [8, 9] are more specific. The first presents a vision-based hand tracking system showcased by a set of one and two-handed gestures in a picture manipulation application, while the second explores multitouch interactions within a room planning scenario. Another specific scenario that implicitly requires hand detection is presented in [10], where Peltonen et al. explore the dynamics of interactions around a public multitouch display installation. They conclude that "design should support performative acts and facilitate asymmetric and ad hoc role-taking, thus letting users learn the opportunities for interaction from their peers."

In this section we explained what motivated our work and why hand awareness is a sought-after feature in multitouch interaction; in the next sections we will overview existing methods for hand detection and show how it can be achieved on every multitouch display.

3 Related Work

The problem of hand detection for multitouch displays can be tackled from a hardware and/or software point of view. Existing solutions mostly rely on additional hardware, while ours is completely software-based. Here we briefly review the existing solutions.

DiamondTouch [11] can distinguish between four users by exploiting an array of antennas, where each antenna transmits a unique signal, embedded in the touch surface and special seats that work as receivers. When a user touches the surface, a small signal is coupled from the antennas near the touch through his body to the receiver. This technology supports two-handed interaction and distinguishes between users. In

[12], Schmidt discusses the benefits of hand detection and user identification for multitouch interaction. He also presents a prototype display augmented with an overhead camera. The camera tracks hands and identifies users based on the hand's contours [13]. Instead of hand contours, Dohse et al. [14] use skin color segmentation to distinguish and identify users with an overhead camera. Another similar approach was adopted by Echtler et al. [15]; with an additional light source placed on the ceiling above the display and a dedicated circuit to control the lights and the camera, they were able to detect shadows cast by the users hand's with the camera already present in the display. Besides hand detection, this enabled the authors to implement mouse-like 'hover' functionality. Another hardware based hand detection solution are the fiduciary-tagged gloves presented by Marquardt et al. in [16]. The gloves are equipped with fiducials that enable recognition of various parts of the hand like fingertips, palms, sides etc. Hand and user detection is achieved in a similar fashion.

The methods described are all capable of distinguishing between hands and also between users. Their common drawback is the need for additional hardware, which makes these methods cumbersome and inapplicable to existing multitouch displays.

To the best of our knowledge, the only software-based method for hand detection is the one presented by Dang et. al. in [17]. They adopt "a simple heuristic for mapping fingers to hands that makes use of constraints applied to the touch position combined with the finger orientation." For two touchpoints, this technique first checks if they are within a certain distance. If so, their intersection is checked next. The intersection of two touchpoints is the intersection of the lines described with the fingers' positions and orientations. If this intersection is behind the touchpoints, the touchpoints can be associated to the same hand. This decision is not yet final as other conditions are applied for further disambiguation; we wish to point out this condition because it relies on the definition of what is 'behind' and what 'in front' on the display. This implies that all the users approach the display from the same side, which is plausible in a single user environment as envisioned in the paper, but restricts the possible multiuser expansion of the method to horizontally mounted displays. In other words, the method cannot be applied in a multiuser tabletop scenario. On the other hand, the method's strengths are the fact that it can be implemented on every display that provides information about finger orientation and its reported 97.5% ($\sigma = 0.48$) overall accuracy in distinguishing a single user's left and right hand.

4 Clustering for Hand Detection

This section is divided as follows: in subsection 4.1 we overview the general characteristics of clustering for hand detection on multitouch displays, subsection 4.2 explains why we chose DBScan and XMeans for baseline comparison and finally subsection 0 describes our proposed clustering algorithm.

4.1 General Characteristics of Clustering for Hand Detection on Multitouch Displays

Generally speaking, a multitouch display is a touchscreen capable of detecting an unlimited number of touches. Depending on the underlying sensing technology, some constraints may apply. Technology also determines what data we get from the screen. Optical, computer-vision based technologies (e.g. FTIR [18], diffused illumination; see [19] for a full list) are unrestricted in terms of the number of detected touchpoints and provide a rich description of touches. For example, a possible set of data provided by an optical multitouch display is described by the TUIO protocol [20]; each touchpoint is described by a session ID, class ID, position, angle, dimension, area, velocity vector, rotation velocity vector, motion acceleration, rotation acceleration and a free parameter. Besides fingers, some optical displays are also capable of detecting objects placed on the display. On the other hand, non-optical sensing technologies (e.g. capacitive, resistive[19]) can detect only a limited number of touchpoints (e.g. PQ Labs G3 Basic, up to 6 touchpoints[1]) and provide a limited set of information, usually only the coordinates of the touches.

The lowest common denominator of all multitouch displays is the description of touchpoints with x, y coordinates. Therefore, if a hand detection technique is to be generally applicable to all multitouch displays, it must rely only on these data. This is what shaped our goal as ‘the development of a method/technique for hand-detection based on the coordinates of the touchpoints.’ The goal, as we put it, is similar to the definition of clustering: to determine the intrinsic grouping in a set of unlabeled data.

Furthermore, clustering for hand detection on a multitouch display is characterized by the following properties: a small and highly variable number of instances (touchpoints), the human hand’s anatomy, unknown number of clusters and the continuous nature of interaction. These properties must be taken into account when choosing an appropriate clustering algorithm or when developing one. On the one hand, the small number of instances causes problems to most clustering algorithms; on the other hand, heuristics derived from the hand’s anatomy can fruitfully be exploited when developing an algorithm from scratch as we will show later.

4.2 Suitable Clustering Algorithms for Hand Detection: DBScan and XMeans

First, we wanted to evaluate the performance of existing clustering algorithms. According to [21], they can be divided as follows: partitioning, hierarchical, density-based, grid-based, model-based, and ensembles of different algorithms. Our choice was dictated by the nature of our problem: assessing the correct number of hands on the screen means that we needed an algorithm capable of automatically detecting the number of clusters in the data. Furthermore, the algorithm should be adept to work in an incremental fashion as fingers come and leave the screen. This led us to the choice of two algorithms: XMeans and DBScan. The first is an extension of the k-means

[1] http://multi-touch-screen.com/store.html

algorithm capable of determining the number of clusters (k) automatically [22], while the second is a clustering algorithm “relying on a density-based notion of clusters which is designed to discover clusters of arbitrary shape [23].” The notion of density of clusters used by DBScan is defined with two parameters: the neighborhood size Eps and minimal number of points minPts. Basically, for a point to be part of a cluster it must satisfy the condition that at least minP points are present in its Eps neighborhood. This is only partly true as points on the border of a cluster are an exception to this condition (see [23] for details). Our domain of hand detection determines the choice for both parameters; minPts must be set to 1, so that hands with only one finger can be detected and a hand span seems a sound choice for Eps - 15.9 cm resulted in the highest accuracy (parameters were fine-tuned for performance, data not shown).

Algorithm 1. HDCMD: hand detection with clustering on multitouch displays

```
Input: fingers - a list of all touchpoints present on the screen.
Output: hands - a list of detected hands on the screen and the associated fingers.
for each finger f1 in fingers do
  if f1 is unclustered
  then
    create hand hand
    add hand to hands
    add f1 to hand
  else
    continue
  end
  for finger f2 in fingers do
    if (f2 is unclustered) and (distance(hand, f2) < maxDistance)
    then
      add f2 to hand
    end
    if size(hand) > =5
    then
      break
    end
  end
end
```

4.3 HDCDM: Hand Detection With Clustering on Multitouch Displays

We propose an algorithm HDCMD (Hand Detection with Clustering on Multitouch Displays) that builds upon two premises: the size of the hand span and the fact that a human hand has five fingers. HDCMD maintains a list of hands and each hand in this list maintains a list of fingers associated with it. The algorithm works as follows: when a finger touches the screen, if there is no other already identified hand within maxDistance (approximately half a hand span distance), a new hand is added to the hands list and the finger is associated with it. If one or more hands are within maxDistance, the finger is added to the nearest hand that has less than 5 fingers associated with it. If only hands with five fingers are near, a new hand is created. We achieved the best results with maxDistance set to 10 cm (the parameter was experimentally fine-tuned, data not shown). For evaluation we used the algorithm as presented in Algorithm 1.

5 Accuracy of Hand Detection

The main goal of our evaluation was to determine the accuracy of hand detection for XMeans, DBScan and HDCMD. Additionally, we also performed two tests. One aimed at confirming the intuitive assumption that HDCMD's accuracy is affected by the multitouch display's size and the other aimed at quantifying the maximum number of hands that HDCMD can detect with respect to the display's size and the desired detection accuracy.

5.1 Accuracy of hand detection

To evaluate the performance of clustering for hand detection, a substantial amount of data is needed, therefore we implemented a 'touchpoints data generator' (TDG)[2] to create a suitable database. Besides the amount of data needed, another argument in favor of our simulation approach is the fact that real interaction is highly dependable on the application used while gathering the data and can therefore influence test results. TDG is an algorithm that, given the number of hands, the number of fingers on them and the size of the display, returns randomly generated pairs of coordinates for each touchpoint. It works as follows: for each hand TDG generates a center point that represents the center of the palm. Then it generates touchpoints that are no more than 8 cm away (we empirically determined 8 cm as approximately half of an average hand span) around the palm's center. This center point must be at least a hand span away from all previously generated center points so that hands do not overlap completely thus leading to a better resemblance of real multitouch interaction. In contrast to the dynamic nature of multitouch interaction, the data generated in this way is static and represents a still frame or a snapshot of what is on the screen at a given point in time. With TDG we created a dataset of 26946 snapshots, 499 for each combination of hands and fingers, where the number of hands ranged from 1 to 8 and the number of fingers from 1 to 5. We also created snapshots with a random number of hands and

[2] A similar approach was adopted in the development of Microsoft Kinect sensor's skeletal tracking. (C. Bishop, Microsoft Research Cambridge, online - last accessed 25.1.2013, http://techtalks.tv/talks/54443)

fingers from the intervals [1,8] and [1,5] for hands and fingers respectively. Snapshots with a fixed number of hands and fingers are useful for analysis, while those with a random number of hands and fingers give a better approximation of real interaction. The size of our virtual table was 200 cm x 200 cm. The combinations of fingers and hands and the size of the table was chosen so that the dataset can represent up to four users interacting with the screen with both hands or 8 users interacting with only one hand.

Table 1. Accuracy [% out of 499 runs] of hand detection with XMeans on synthetic data. Overall accuracy is 21% (5709/26946).

Hands\Fingers	Random 1-5	1	2	3	4	5
random 1-8	22	13	29	27	23	25
1	99	100	100	96	98	99
2	67	0	85	84	87	88
3	0	0	0	0	0	0
4	0	0	0	0	0	0
5	0	0	0	0	0	0
5	0	0	0	0	0	0
7	0	0	0	0	0	0
8	0	0	0	0	0	0

Table 2. Accuracy [% out of 499 runs] of hand detection with DBScan on synthetic data. Overall accuracy is 67% (18185/26946).

Hands\Fingers	Random 1-5	1	2	3	4	5
random 1-8	68	98	58	57	64	60
1	25	100	0	0	0	0
2	46	100	29	31	33	40
3	75	100	59	67	72	73
4	83	98	77	77	83	83
5	85	96	83	81	85	84
5	83	95	80	81	81	74
7	79	90	76	75	76	71
8	70	84	69	65	65	63

We tested all three algorithms on the same dataset. Our algorithm was implemented in Java with the MT4j framework [24], while for DBScan and XMeans we used Weka's implementations [25].

The goal of these algorithms is hand detection, which in turn has two goals: assessment of the number of hands on the screen and mapping fingers to hands. A correct detection happens when both goals are met; the number of hands is assessed correctly and all fingers are correctly mapped to the hands. Tables 1-3 show the accuracy (in %) of the algorithms on the same dataset for XMeans, DBScan and HDCMD respectively. Table 4 shows the results for HDCMD on a different dataset, where hands were allowed to completely overlap one another without any restrictions.

Table 3. Accuracy [% out of 499 runs] of hand detection with HDCMD on synthetic data. Overall accuracy is 97% (26121/26946).

Hands\Fingers	**Random 1-5**	**1**	**2**	**3**	**4**	**5**
random 1-8	100	100	100	100	100	100
1	100	100	100	99	100	100
2	99	99	99	99	98	100
3	98	97	98	98	98	100
4	96	97	97	95	97	100
5	95	97	95	94	95	100
5	95	94	92	92	90	100
7	90	88	89	91	92	100
8	100	100	100	100	100	100

Table 4. Accuracy [% out of 499 runs] of hand detection with HDCMD on synthetic data with completely overlapping hands. Overall accuracy is 93% (25150/26946).

Hands\Fingers	**Random 1-5**	**1**	**2**	**3**	**4**	**5**
random 1-8	100	100	100	100	100	100
1	99	99	99	99	99	100
2	98	98	97	98	97	100
3	96	95	96	96	95	100
4	93	96	93	92	90	100
5	89	89	87	87	85	100
5	88	84	83	86	86	100
7	85	82	78	77	78	99
8	100	100	100	100	100	100

5.2 HDCMD – display size, hand detection accuracy and the number of hands detected

The simple heuristics behind HDCMD, inspired by human hand anatomy, imply a tight connection between hand detection accuracy, display size and the maximum number of hands that can be detected successfully. Qualitatively speaking: for a fixed display size, an increase of the number of hands we want to detect should result in lower detection accuracy. To confirm and quantify this assumption, we prepared data for Fig. **1** and Table 5.

Fig. **1** shows HDCMD’s hand detection accuracy for different display sizes. For each display size the accuracy was calculated from 10000 snapshots, where a snapshot consisted of a random number of hands from the interval [1,8], each with a random number of fingers.

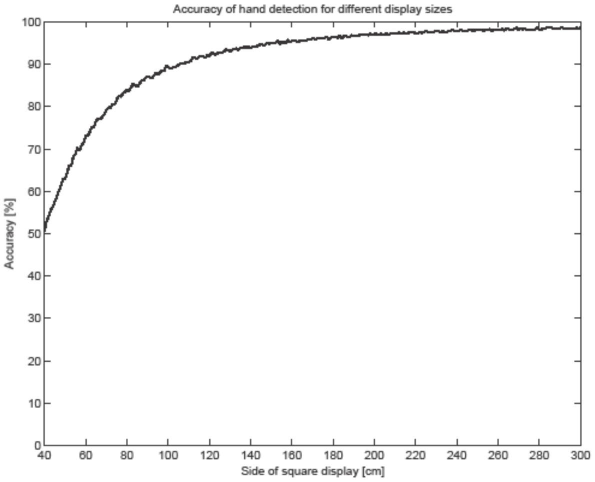


Fig. 1. The accuracy of HDCMD for different display sizes calculated from 10000 snapshots with a random number of hands and fingers from the intervals [1,8] and [1,5], respectively.

Table 5. Maximum number of hands for different display sizes and hand detection accuracies.

Accuracy [%]		**90**	**95**	**96**	**97**	**98**	**99**
Size [*cm*]	**Area [*cm*^{2}]**	**Maximum number of hands**					
20	**400**	2	2	2	2	2	2
40	**1600**	3	2	2	2	2	2
60	**3600**	5	3	3	3	2	2
80	**6400**	6	4	4	3	3	2
100	**10000**	7	5	5	4	4	3
120	**14400**	9	6	5	5	4	3
140	**19600**	10	7	6	6	4	3
160	**25600**	11	8	7	6	5	4
180	**32400**	13	9	8	7	5	4
200	**40000**	14	9	9	8	6	5
220	**48400**	15	11	9	8	7	5
240	**57600**	17	11	10	9	7	5
260	**67600**	18	13	11	10	8	6
280	**78400**	19	13	12	10	8	6
300	**90000**	20	13	13	11	9	6

Table 5 investigates the relation between display size, hand detection accuracy and number of hands on the display in more detail. Data was gathered as follows: for each display size, we started by calculating the accuracy of hand detection for the maximum number of hands maxHands set to one. If the accuracy was above the selected threshold (90%, 95%, 96%, 97%, 98% or 99%), we increased maxHands, generated new data and calculated accuracy again. We repeated this until accuracy dropped below the selected threshold. Each time the accuracy was calculated from 10000 snapshots, prepared with TDG, and each snapshot consisted of a random number of hands from the interval [1, maxHands], each hand with a random number of fingers.

6 Discussion

The goal of this research was to establish whether clustering can be used as a means for hand detection on multitouch displays or, in other words, if clustering can determine how many hands are touching the display and which finger, represented by a touchpoint, belongs to which hand. We found out that available clustering algorithms suitable for the task fail to provide sufficient accuracy, while the algorithm presented in this paper performs significantly better. Furthermore, we performed tests that show the limitations of the proposed HDCMD algorithm.

6.1 Hand Detection Accuracy

Table 1 shows that XMeans proved useless in terms of hand detection. The main factor for the poor overall accuracy of the algorithm (21%) is that XMeans tends to underestimate the number of clusters due to the Bayesian information criterion used in determining the number of clusters [22]. This also explains why it only performs well, when there are only one or two hands on the screen as there is a smaller chance for underestimating the number of clusters. When the number of hands (i.e. actual clusters) increases, XMeans always detects less hands than are actually present.

DBScan performs considerably better than XMeans, but still not satisfactorily. Although we are reporting results with the chosen parameters that yield the best overall accuracy (67%), this accuracy is not high enough for the algorithm to be useful in hand detection for multitouch interaction. The main problem is that this choice of minP and Eps results in the algorithm detecting more hands, when there is actually only one hand with more fingers on the screen. Table 2 also shows that DBScan works better when the number of touchpoints on the screen is such that the differences in densities are more distinctive.

To find the best possible results, we experimented with different DBScan and XMeans parameters (in the interest of space, the data is not shown here). Despite this, the aforementioned algorithms performed significantly worse than our human-anatomy inspired algorithm. Due to their poor overall accuracy DBScan and XMeans both proved unsuitable for hand detection for multitouch interaction. In contrast, HDCMD boasts an overall accuracy of 97%. This makes it a suitable means for hand detection. Table 3 shows that, in case all hands are touching the screen with 5 fingers, the accuracy rises to 100%. This can be attributed to the incremental nature of the algorithm that implicitly transforms a snapshot of contemporary touchpoints to a series of touchpoints. This reduces the errors caused by two hands that are close together. For example, consider a combination of 2 hands represented with 4 fingers each. The algorithm correctly maps the four fingers of the first hand and can then make a mistake when mapping the first finger of the second hand by mapping it to the first hand. In the case of hands represented with five fingers this could not happen, because the first hand would already have 5 fingers mapped to it.

6.2 HDCMD's Limitations

Another point observable from Table 3 is that the accuracy of hand detection is consistent across various combinations of hands and fingers; the only noticeable trend is the decrease of accuracy with the increasing number of hands on the screen. Our assumption that this is due to the increased chance of hands being close together was confirmed by Fig. **1**, where the accuracy of hand detection is plotted for different display sizes. As expected, accuracy increases with the increase of the display's size.

In Table 5 we give the maximum number of hands for different sizes that still resulted in an accuracy over the next thresholds: 90%, 95%, 96%, 97%, 98% and 99%. Display size is given as the length of the side of a square display in *cm* and its respective area in cm^2. The smallest display we tested was 20 *cm* wide as less than this can be considered as a single user display and the biggest display was 300 *cm* wide. We would like to point out that, although a square display of this size (300 *cm*) is of little practical use (because the center of such tabletop or of such wall is hardly reachable for users) the results apply also for displays with the same area but with a different format. In this sense, the main goal of Table 5 is to show to what extent our algorithm for hand detection can be used: how many hands can it detect on how big a display. For example, when developing an application for 3 users interacting bimanually, Table 5 tells us that the application should run on a display with an area of at least 6400 cm^2 for 90% accuracy or 14400 cm^2 for 95% accuracy. Clearly, a lower number of hands on the display andClearly, a lower number of hands on the display and/or a larger display have a positive effect on the accuracy of hand detection. In other words, HDCMD compromises between the maximum number of hands we can detect and the accuracy of the detection. This is due to HDCMD's simplistic nature.

Our algorithm makes errors, when hands are close together. Figure 1 shows these errors; we can see that the errors can be classified in two types. The first type of error is when fingers from two different hands are detected as fingers from a single hand and the second type of error is when fingers of two hands are incorrectly mapped to two separate hands. In the first case both the number of hands detected as well as the mapping of fingers to hands are incorrect, while in the second case only the mapping of fingers to hands is incorrect.

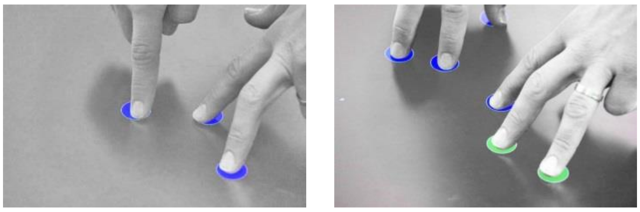


Fig. 2. HDCMD's errors: fingers from different hands are detected as if they were all from the same hand (left) and fingers from hands close together are incorrectly mapped (right). For illustrative purposes we desaturated all the colors in the pictures except blues and greens.

6.3 Use Cases & Future Directions

The aforementioned limitations of HDCMD influence its possible use cases and the directions for future work. In the limited information context we are exploring (x and y coordinates of touchpoints) there is no way of knowing if two hands belong to the same user. This poses a limit to HDCMD - it can only detect hands, but not users. In other words, HDCMD can be used to build applications that are hands-aware, but not user-aware. Nevertheless, in Section 2 we mentioned a valid use-case for hands-aware applications, namely "cooperative gestures for co-located groupware [5]." In these applications, some commands can only be invoked by gestures performed collectively by multiple group members. As a result, in this scenario, the system is only interested in knowing that all users have taken part (hand detection) in the collaborative hand gesture rather than with the identities of the users. Another possible use-case is connected to the notion of territoriality described in [26], where Scott et al. "conducted two observational studies of traditional tabletop collaboration in both casual and formal settings" and found out that "collaborators use three types of tabletop territories to help coordinate their interactions within the shared tabletop workspace: personal, group, and storage territories." Personal territories belong to a specific user and all interaction that occurs in them can be attributed to that user. Within personal territories, HDCMD can be used to support the implementation of bimanual interaction of a single user.

User studies of applications using HDCMD in the abovementioned and other scenarios are one possible direction for future research. Another option originates from the fact that clustering on snapshots is a more difficult problem than clustering dynamic data, because the dynamics of interaction can also be an aid in determining which hand the finger belongs to. For example, the speed of a touchpoint can discriminate it from touchpoints from a nearby hand. As speed is a feature calculated from touchpoint coordinates it does not reduce the general applicability of HDCMD.

7 Contribution and Conclusion

Our work shows that, in contrast to some assumptions, clustering can successfully be exploited for hand detection on multitouch surfaces. Our main contribution is showing how and to what extent this can be achieved. We present an incremental clustering algorithm based on simple heuristics stemming from the anatomy of the human hand. The algorithm determines the number of hands on the screen and maps each finger to its hand with an accuracy of 97%, tested on synthetic data. The features used for clustering are the x and y coordinates of the touchpoints on the screen, which means that the algorithm can be used on all multitouch displays regardless their construction.

Acknowledgements. This research was funded in part by the European Union, European Social Fund, Operational Program for Human Resources, Development for the Period 2007-2013 and by the Slovenian Research Agency.

References

1. Buxton, W.: Multi-Touch Systems That I Have Known and Loved, http://www.billbuxton.com/multitouchOverview.html Last accessed: 25.1.2013.
2. Benko, H., Morris, M. R., Brush, A. J. B., & Wilson, A. D.: Insights on Interactive Tabletops : A Survey of Researchers and Developers. *research.microsoft.com* (2009)
3. Leganchuk, A., Zhai, S., & Buxton, W.: Manual and cognitive benefits of two-handed input: an experimental study. ACM Transactions on Computer-Human Interaction (TOCHI), 5(4), pp. 326–359, ACM (1998)
4. Kin, K., Agrawala, M., DeRose, T.: Determining the benefits of direct-touch, bimanual, and multifinger input on a multitouch workstation. In: Proceedings of Graphics interface, pp. 119–124, Canadian Information Processing Society, Ontario (2009)
5. Ringel Morris, M., Huang, A., Paepcke, A., Winograd, T.: Cooperative Gestures: Multi-User Gestural Interactions For Co-Located Groupware. In: Proceedings of the ACM Chi Conference on Human Factors in Computing Systems, pp. 1201–1210, ACM (2006)
6. Terrenghi, L., Kirk, D., Sellen, A., Izadi, S.: Affordances For Manipulation of Physical Versus Digital Media on Interactive Surfaces. In: Proceedings of the Sigchi Conference on Human Factors in Computing Systems, Chi ’07, pp. 1157–1166, ACM (2007)
7. Rogers, Y., Lim, Y. K., Hazlewood, W. R., Marshall, P.: Equal Opportunities: Do Shareable Interfaces Promote More Group Participation Than Single Users Displays? Human-Computer Interaction, 24(1-2), pp. 79–116, Taylor & Francis (2009)
8. Malik, S., Laszlo, J.: Visual Touchpad: A Two-Handed Gestural Input Device. In: Icmi ’04: Proceedings of the 6th International Conference On Multimodal Interfaces, pp. 289–296, ACM, New York (2004)
9. Wu, M., Balakrishnan, R.: Multi-Finger and Whole Hand Gestural Interaction Techniques For Multi-User Tabletop Displays. In: Proceedings of the 16th Annual ACM Symposium on User Interface Software and Technology, Uist ’03, pp. 193–202, ACM, New York (2003)
10. 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: Proceeding of the Twenty-Sixth Annual Sigchi Conference on Human Factors in Computing Systems, Chi ’08, pp. 1285–1294, ACM, New York (ACM)
11. Dietz, P., Leigh, D.: DiamondTouch: a multi-user touch technology. In: Proceedings of the 14th annual ACM symposium on User interface software and technology, pp. 219–226. ACM (2001)
12. Schmidt, D.: Know Thy Toucher. In: CHI ’09 Workshop Multitouch and Surface Computing, Boston (2009)

13. Schmidt, D., Chong, M. K., Gellersen, H.: Handsdown: Hand-Contour-Based User Identification For Interactive Surfaces. In: Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries, Nordichi ’10, pp. 432–441, ACM, New York (2010)
14. Dohse, K. C., Dohse, T., Still, J. D., Parkhurst, D. J.: Enhancing multi-user interaction with multi-touch tabletop displays using hand tracking. In: Advances in Computer-Human Interaction, 2008 First International Conference on, pp. 297–302, IEEE (2008)
15. Echtler, F., Huber, M., Klinker, G.: Shadow Tracking on Multi-Touch Tables. In: Proceedings of the Working Conference on Advanced Visual Interfaces, Avi ’08, pp. 388–391, ACM, New York (2008)
16. Marquardt, N., Kiemer, J., Greenberg, D.: What Caused That Touch?: Expressive Interaction With A Surface Through Fiduciary-Tagged Gloves. In: ACM International Conference on Interactive Tabletops and Surfaces, Its ’10, pp. 139–142, ACM, New York (2010)
17. Ringel Morris, M., Huang, A., Paepcke, A., Winograd, T.: Cooperative Gestures: Multi-User Gestural Interactions For Co-Located Groupware. In: Proceedings of the ACM Chi Conference on Human Factors in Computing Systems, pp. 1201–1210, ACM (2006)
18. Han, J. Y.: Low-Cost Multi-Touch Sensing Through Frustrated Total Internal Reflection. In: Proceedings of the 18th Annual ACM Symposium on User Interface Software and Technology, Uist ’05, pp. 115–118, ACM, New York (2005)
19. Çetin, G., Bedi, R., Sandler, S.: Multi-touch Technologies, 1st edition (2009)
20. Kaltenbrunner, M., Bovermann, T., Bencina, R., Costanza, E.: Tuio: a Protocol for Table-Top Tangible User Interfaces. In: Proceedings of the 6th International Workshop on Gesture in HumanComputer Interaction and Simulation GW, pp. 1–5 (2005)
21. Kotsiantis, S., Pintelas, P.: Recent advances in clustering: A brief survey. WSEAS Transactions on Information Science and Applications, *1*(1), pp. 73–81 (2004)
22. Schmidt, D., Chong, M. K., Gellersen, H.: Handsdown: Hand-Contour-Based User Identification For Interactive Surfaces. In: Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries, Nordichi ’10, pp. 432–441, ACM, New York (2010)
23. Ester, M., Kriegel, H. P., Jörg, S., Xu, X.: A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases With Noise. In: Proceedings Of The ACM Sigkdd International Conference On Knowledge Discovery And Data Mining, pp. 226–231, AAAI Press (1996)
24. Laufs, U., Ruff, C., Zibuschka, J.: Mt4j – A Cross-Platform Multi-Touch Development Framework. In: Engineering Patterns for Multi-Touch Interfaces, Workshop of the ACM Sigchi Symposium on Engineering Interactive Computing Systems (2010)
25. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.: The Weka Data Mining Software: an Update. Special Interest Group on Knowledge Discovery and Data Mining Explorer Newsletter, 11(1), 10–18 (2009)
26. Scott, S. D., Sheelagh, M., Carpendale, T., Inkpen, K.M.: Territoriality in Collaborative Tabletop Workspaces. In: Proceedings of Cscw ’04, pp. 294–303, ACM (2004)

4 A Personal Perspective on Photowork: Implicit Human Computer Interaction for Photo Collection Management

In this chapter, the paper (Blažica et al., 2013b) titled "A personal perspective on photowork: implicit human computer interaction for photo collection management" by Bojan Blažica, Daniel Vladušič and Dunja Mladenić is presented. The paper is published in the Personal and Ubiquitous Computing[1] journal.

The paper explores the possibility to use implicit human computer interaction to aid personal photo collection management. Other state-of-the-art approaches to the problem rely on meta data analysis, on image content analysis, temporal clustering, tagging, crowdsourcing etc. The problem is that none of them is capable of taking into account the user's personal relationship with a single photo; whether the photo is of particular importance to the user. We call this personal relationship the user's *affinity* for a photo. The experiments in the paper reveal how affinity is a relative measure that can be used to sort photos by comparison and that affinity itself is correlated with the time a user spends viewing a picture. Furthermore, by looking at viewing times, it is also possible to distinguish the task a user is currently performing - whether he/she is searching for a photo, browsing through them or making a selection.

The presented paper is another example of contextual information acquisition, but at the same time, it also provides a valid use case for context-awareness. The basic idea is that the way we interact with natural user interfaces can implicitly disclose additional (contextual) information that can be exploited by a context-aware system to better understand the user.

The first author of the paper conceived the basic idea in the paper, i.e. using implicit human-computer interaction for photo organization management. He provided an operational definition of the key term of the paper *affinity* and implemented the tablet application needed to carry out the experiments. Coauthors helped in the definition of the experiments and in the analysis of the gathered data. They also contributed to the final structure and clarity of the paper.

[1] IF 2011 = 0.938; ET - computer science, information systems: 2 quartile; YE - telecommunications: 2 quartile

Pers Ubiquit Comput
DOI 10.1007/s00779-013-0650-6

ORIGINAL ARTICLE

A personal perspective on photowork: implicit human–computer interaction for photo collection management

Bojan Blažica · Daniel Vladušič · Dunja Mladenić

Received: 31 August 2012 / Accepted: 5 February 2013

Abstract In the age of digital photography, the amount of photos we have in our personal collections has increased substantially along with the effort needed to manage these new, larger collections. This issue has already been addressed in various ways: from organization by meta-data analysis to image recognition and social network analysis. We introduce a new, more personal perspective on photowork that aims at understanding the user and his/her subjective relationship to the photos. It does so by means of implicit human–computer interaction, that is, by observing the user's interaction with the photos. In order to study this interaction, we designed an experiment to see how people behave when manipulating photos on a tablet and how this implicitly conveyed information can be used to aid photo collection management.

Keywords Photowork · Personal information management · Photo collection organization · Implicit human–computer interaction

1 Introduction

Photo collection management and consumption has been studied for decades. Part of Richard Chalfen's seminal work in the field [1–3] dates back to as early as 1975 [3]. Since then, practices concerning photowork (personal information management concerned with photos) have changed, most notably with the advent of digital photography—for example, in 1972 the estimated number of photos taken in a year was 4.75 billion [3], while in 2011, 6 billion photos were uploaded each month on Facebook alone [4]. Whittaker et al. [5] explored how this abundance of photos effects people's ability to recall the photos and retrieve them and its effect on their corresponding memories regarding the photos after not viewing them for a longer period of time. They find that due to the large amount of photos and lack of meaningful organization and storage of photo collections, people fail to find 40 % of their photos. The phrase included in the title of their paper "easy on that trigger dad" compellingly describes the problem at hand: In terms of money spent, digital photos have become inexpensive and can be made easily and quickly, yet in terms of the effort needed to organize and maintain them, they are rather expensive. In a sense, digital photography simply transformed the problem of deciding whether or not to take a photo to the problem of organizing large collections of photos. According to Whittaker et al. in many cases, the transformed problem, photo collection management, never gets solved properly.

There are many solutions to the problem of photo collection management and organization based on various approaches like semantic annotations [6], image recognition [7], (time-based) clustering [8, 9], social networks [10] and various combinations of the above [11]. However, photos are "seemingly of a very high subjective nature" [5]

B. Blažica (✉) · D. Vladušič
XLAB Research, Pot za Brdom 100, Ljubljana, Slovenia
e-mail: bojan.blazica@xlab.si

B. Blažica · D. Mladenić
Jožef Stefan International Postgraduate School,
Jamova 39, Ljubljana, Slovenia

D. Mladenić
Artificial intelligence laboratory, Jožef Stefan Institute,
Jamova 39, Ljubljana, Slovenia

Published online: 10 March 2013

(p. 39), and an organizational method should take this into consideration. For this reason, we looked at photo collection management from a different, more personal perspective. Instead of organizing photos according to, for example, time, we try to organize them according to the user's affinity. The main question that we are addressing is how to capture and describe the subjective relationship between the user and a photo. The answer we are proposing lies in the realm of implicit human–computer interaction: "Implicit human computer interaction is an action performed by the user that is not primarily aimed to interact with a computerized system, but which such a system understands as input" [12] (p. 2). In the scenario of photo collection management, the main action performed by the user is the consumption of photos (e.g., browsing, viewing with friends, etc.) which, at the same time, serves as input for a photo collection management system capable of implicit human–computer interaction.

According to Petrelli et al. [13], photos (and other mementos) should not be viewed only from a life-logging perspective that uses technology to exhaustively capture a person's life. Instead, we should use technology to support active remembering and the consumption of mementos, as they are actively consumed in the course of ongoing social activity. Similarly, when prioritizing user requirements for photo-sharing technologies, Frohlich et al. [14] (p. 173) found that users would like to use photos "more extensively as catalysts for conversation in extended family and friendship contexts." For example, O'Hara et al. [15] elaborated on the possibility of using photos in the everyday context of sharing a meal, and Hilliges et al. [16] explored photo browsing, organizing and sharing using a wall display and interactive tabletop in the future living room. At the heart of both examples, there is rich and diverse user interaction with photos—a good prospect for implicit human–computer interaction.

The goal of this paper is to explore new possibilities in photo collection management that arise from implicit human–computer interaction. The intent is not to replace current photo management solutions but to complement and improve them by addressing photo management from a user's subjective perspective, which has so far been neglected. The basic information that we intend to extract by means of implicit human–computer interaction is the subjective relationship between the user and a photo. We call this relationship the user's *affinity* for a photo. The meaning of the word affinity here extends beyond its common understanding of natural attraction or liking. A user's affinity for a photo may be caused by the photo's aesthetics or by the memories the photo evokes. These memories may be pleasant or unpleasant, as affinity does not necessarily tell us anything about whether the user's relationship with the photo is positive or negative, but rather that the user's subjective relationship with the photo is strong, which in turn means that the photo is important to the user. In other words, the user's *affinity* for a photo concurs with Chalfen's [3] (p. 2) notion of *importance* of a photo in the home-mode: "And although artists, art historians, and art critics frequently speak of 'important' and 'valuable' images, we are dealing with a different notion of importance here. In the home-mode, images are indeed important in an intimate context, and these images are valued by small groups of biologically and socially related people." In this sense, good and bad photos are not discriminated based on photo quality (in terms of composition, exposure, lighting, etc.), but rather on affinity; a good photo is a photo with high affinity—a photo that is important to the user.

We first examined this novel approach of addressing photo collection management from a personal perspective in a previous paper [17], where we presented an algorithm for photo collection management, a photo visualization technique for multitouch tabletops and preliminary experiments that tested whether the time spent viewing a picture can be considered as a measure of the user's affinity for that picture. These experiments were conducted on a multitouch tabletop with public images from Flickr. In this paper, we further narrow the focus on *personal* photo collection management—experiments were conducted on a tablet, and participants were asked to bring their own photos. The experiments described in [17] and those presented here also differ in the tasks that the participants were asked to perform: Previously, the participants first browsed photos scattered on the tabletop and were then asked to make 10 pairwise comparisons, while here the participants first browsed the pictures one by one, then rated them and finally made some selections. In this paper, we also attempted to identify the basic photowork task that the user was performing.

Conducting the experiments on a tablet is another novelty in the field. To the best of our knowledge, no study has yet been carried out that would examine user behavior during photowork on a tablet device, despite the fact that tablets have been identified as devices with great photowork potential [18]. It has to be noted that we will only be examining a small subset of basic photowork-related tasks. The basic photowork tasks identified by Kirk et al. [18] are sorting, selecting and filtering.

This paper is structured as follows. In Sect. 2 we present the experiment conducted to explore the fundamentals of implicit photo collection management. The results of the experiment presented in Sect. 3 are discussed along with options for future work in Sect. 4. In Sect. 5, we conclude the paper by briefly stating the contribution of our work.

2 Methods

The aim of this work is to elaborate on implicit information extraction from user interaction in the context of photo collection organization and to show that implicit human–computer interaction can aid basic photowork tasks. To explore user behavior while interacting with photos, we developed an experimental tablet application. Data gathered from real users using this application enabled us to develop and validate the following hypotheses:

H1 The time a user spends viewing a photo (viewing time) can be used as a measure of the user's affinity for that photo;

H2 Viewing times in combination with other data (available without explicit user intervention) can be used to predict ratings that the user would assign to the photos;

H3 It is possible to identify which basic photowork task the user is performing by observing photo viewing times.

We used a tablet application for gathering data as it presents a natural tool for personal photo management [18] (different from photo organization for photography professionals) and offers richer interaction due to the fact that the user directly manipulates photos (and other objects) on the screen with his/her hands, without a "middleman" like a mouse or a keyboard. The workflow of the experiment supported with the application was the following. The user was first asked to import either a collection of personal photos from a recent event (e.g., trip, family holiday, party, etc.) that have not yet been looked at or photos that have not been looked at for some time from an older event. Then he/she had to perform four tasks, with these instructions:

- Step 1: "Browse through your photos as you would normally do. Tap 'Done' in the action bar after you've finished."
- Step 2: "Please rate the photos you've just browsed based on how much they mean to you. Rate on a scale from 1 to 10, where 10 means that the photo is really important to you."
- Step 3: "Select a set of photos that you would show to your friends and family. To select or unselect a photo tap on it. Tap 'Done' in the action bar after you've finished."
- Step 4: "Select ONE photo to represent this collection of photos. For the album cover, Facebook profile, wallpaper … Tap 'Done' in the action bar after you've finished."

At any time during the experiment, only one photo was present on the screen and the user could move through the photos by swiping his/her finger left and right. The steps of the experiment were designed to mimic ordinary browsing behavior (step 1), selecting behavior (step 3) and searching behavior (step 4) and to explicitly obtain feedback from the participant (step 2). The rationale behind the condition that photos had to be either newly acquired or not seen for a long time is that both evoke a similar and strong emotional reaction that we attempted to capture through interaction patterns. This reaction is stronger in a browsing scenario and weaker while selecting or searching photos. The focus of our experiment was observing this original reaction to photos and determining how this can be exploited for implicitly aiding basic photowork, while a longitudinal study on how the reaction changes with time and how it effects implicit photo collection management is out of the scope of this paper and left for future work. To respect their privacy, participants were given the option either to erase the images after the experiment or to "donate them to science" for further research.

The application was developed on the Android platform (API level 11); the tablet used was a Lenovo ThinkPad tablet. Figure 1 shows two screenshots of the application: browsing during step 1 on the left and rating in step 2 on the right.

The data we collected with the application included the following:

- Viewing times: Each time a photo was displayed, we measured and stored its duration on the display;
- Viewing traces: A viewing trace consists of the indices of the viewed photos along with associated viewing times (one trace per experimental step);
- Ratings on a scale from 1 to 10 for all images;
- User data: age, gender, self-assessed level of photo enthusiasm and how the user stores and shares his/her photos (possible answers included "Facebook (or other social media)," "Flickr (or other specialized photo service)," "Via mail," "Print," "I organize photo-viewing events");
- Meta-data about photos stored in EXIF tags, for example: the date the photo was taken, the model of the camera used, shooting mode, aperture and shutter settings, ISO speed, use of flash, etc.

In total, we collected data from 10 subjects, seven males and three females, between the ages of 26 and 33 (see Table 1 for more details about the participants). The duration of the experiment was not limited, and the subjects were free to bring as many photos as they liked. After the experiment, five users added tags to the photos viewed in the experiment. They were free to create their own tags as long as the tags were content related. The tags the users added described the main theme of the photo (nature, castle, museum, etc.), the people on the photo or the event related to the photo (concert, strolling, mountaineering, etc.). One tag was assigned to each photo.

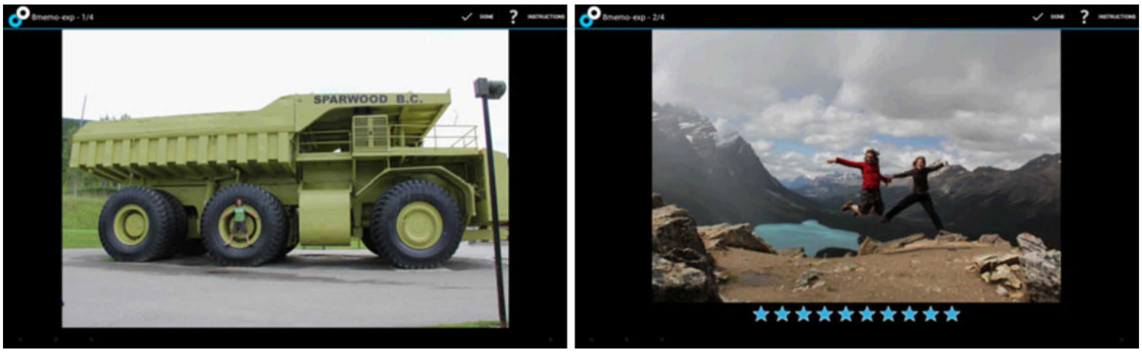


Fig. 1 Screenshots from the experimental tablet application: browsing on the *left* and rating on the *right*

Table 1 Data about the participants involved in the experiment

User	Age	Gender	Photo enthusiasm	Number of photos viewed	Rating range	Average rating	Sharing habits	Donated photos	Tagged photos
1	26	Male	2	100	1–10	5.43	Mail	Yes	Yes
2	30	Female	4	95	1–10	7	Mail	Yes	Yes
3	28	Male	5	289	1–10	4.42	Events	Yes	Yes
4	28	Male	4	124	3–10	6.68	Events	Yes	No
5	26	Female	3	38	1–10	4.68	Facebook	Yes	Yes
6	31	Male	1	105	1–10	5.99	Mail	Yes	Yes
7	33	Male	2	89	1–10	3.18	Events	No	No
8	26	Female	2	300	1–8	2.94	Events	Yes	No
9	30	Male	4	57	1–10	6.18	Facebook	Yes	No
10	30	Male	5	96	1–10	7.34	Facebook	Yes	No
All	28.8		3.2	129	1.2–9.8	5.38		9/10	5/10

Once the data were gathered, we analyzed it according to two main directions: behavior identification and rating prediction. Behavior identification consists of time series analysis in order to identify the behavior and thus the task the user is performing (browsing, selecting and searching), while rating prediction is based on viewing times and EXIF data for single photos (taken from step 1) and aims at automatically assigning a rating to photos. In the cases where users donated their photos, we also used histograms for analysis.

3 Results

The analysis of the data gathered in the experiments is presented in six tables. Table 1 presents the participants, their age, gender, self-reported photo enthusiasm, number of photos viewed in the experiment, the ratings the participants used to rate the photos, the average rating they assigned, how they share photos and whether they donated and/or tagged photos. In total, 1,293 photos were involved in the experiment.

3.1 Time as a measure of photo affinity (H1)

The goal of Table 2 is to validate hypothesis H1, which states that viewing time is a measure of the users' affinities for individual photos; we made all possible pairs of photos and compared the photos in each pair to see whether a longer viewing time corresponds to a higher rating given by the user. The results of these comparisons are presented in Table 4 (for each user individually in rows 1–9 and for the entire population in the last row) along with the correlation coefficients. The last row confirms the hypothesis. Additionally, we compared the 300 "best" to the 300 "worst" photos for the entire population according to viewing time. Pairs were selected randomly from these two pools, and in 78.7 % of cases, the photo with a longer viewing time was also the photo with a higher rating.

3.2 Predicting ratings (H2)

Next, we tried to learn a model that would be able to predict the ratings given by the users. Two different models were trained for each user: a regression model that outputs

Table 2 Pairwise comparison of photos and the percentage of comparisons where a longer viewing time corresponds to a higher rating given by the user and the correlation coefficients between viewing times and ratings

User	Number of comparisons	Percentage of comparisons where a longer viewing time corresponds to a higher rating (%)	Correlation coefficient
1	4,950	61	0.20
2	4,465	45	0.11
3	41,616	67	0.67
4	7,626	57	0.36
5	703	67	0.57
6	5,460	52	0.21
7	3,916	61	0.63
8	44,850	47	−0.09
9	1,596	42	−0.18
10	4,560	58	0.35
All	835,278	60	0.32

a rating prediction on a scale of 1–10 and a classification model that classifies photos into three bins. For the latter, photos were previously divided into bins (terciles) representing three categories: bad, average and good (in terms of affinity as discussed in the introduction). This division was performed individually for each user.

Each photo was represented with the following feature vector: the order (index) of the photo in the set, cumulative viewing time, minimal viewing time, maximum viewing time, viewing frequency, histogram of the photo divided into 10 bins (where available), aperture, shutter speed, orientation of the photo, flash settings, ISO speed, date and time the photo was taken, the rating of the photo and the tag given by the participant (where available). To build the regression and classification models, we used Weka's [19] implementation of support vector machines (SVM) with a polynomial kernel. The models were trained on normalized data. SVM were chosen for this experiment because they handle classification as well as regression. Preliminary experiments with different algorithms have also shown that SVM perform better on this task, although a thorough analysis regarding which algorithm is optimal for the prediction of ratings is out of the scope of this paper.

The ability to predict ratings was evaluated with stratified tenfold cross-validation. The results of these evaluations are presented in Table 3; rating prediction by regression is evaluated using the mean absolute error of the predicted ratings and the relative-absolute error, while rating prediction by classification is compared with the default accuracy of a model that assigns all photos to the majority class and with the percentage of photos that made a "2 bin jump," that is, bad photos that were predicted as good and vice versa. The results do not confirm hypothesis H2.

3.3 Identifying current basic photowork task (H3)

Tables 4, 5 and 6 examine the behavior of users while performing three different tasks: browsing, selecting and searching (tasks similar to basic photowork tasks). Table 4 shows the average time that users spent viewing a single photo and the average view frequency for a single photo. In general, it can be observed that the average viewing time decreases in the order of browsing, selecting and searching (viewing frequencies increase, respectively), and it is therefore possible to discriminate between tasks based on average viewing times. These average viewing times were calculated a posteriori from the whole set of photos (for a user), which implies that the user has finished viewing the photos. In a real-world application, the information which task the user is performing is needed while the task is being performed. Therefore, we also calculated the average moving averages of viewing times for subsets of three and six photos. For example, when viewing the 19th photo, the moving average for a subset of size 3 is calculated as the average viewing time of photos 16, 17 and 18. This means that when the user is viewing the fourth photo, we are already able to identify the task he/she is performing. Table 5 presents the average moving average viewing time for each user individually, while (for clarity) Table 6 summarizes the same data for all users. Tables 4 and 6 show that moving averages of viewing times are consistent with the overall average viewing times for browsing and that they are better at discriminating between selecting and searching.

4 Discussion and future work

Based on the data presented in Sect. 3, this experiment confirms hypotheses H1 and H3 ("viewing time can be used as a measure of the user's affinity for a photo" and "it is possible to identify what basic photowork-related task the user is performing by observing photo viewing times") and rejects hypothesis H2 (viewing times in combination with other data (available without explicit user intervention) can be used to predict ratings that the user would assign to the photos).

4.1 Viewing time as a measure for photo affinity (H1)

The low correlation between viewing times and ratings (Table 4) shows that viewing time is not an absolute measure of photo affinity, but rather a relative measure of photo affinity as shown by the high percentage of pairwise

Table 3 Evaluation of rating prediction by regression and by classification

User	Regression		Classification				
	Mean absolute error	Relative-absolute error (%)	Terciles limits		Accuracy (%)	Default accuracy (%)	2 Bin jumps (%)
			Low	High			
1	2.67 (2.73)	109.13 (111.69)	3.5	7.5	34 (38)	38	18 (13)
2	1.99 (1.94)	108.86 (105.7)	5.5	8.5	28.42 (33.68)	44.21	4.21 (9.47)
3	1.51 (1.54)	66.75 (67.88)	2.5	5.5	63.67 (64.36)	38.4	5.19 (4.50)
4	1.26	90.72	5.5	7.5	46.77	43.55	9.68
5	2.72 (2.57)	88.66 (83.70)	2.5	7.5	36.84 (52.63)	39.47	21.05 (13.16)
6	1.97 (2.06)	88.62 (92.88)	4.5	7.5	41.9 (43.81)	36.19	22.86 (18.10)
7	1.42	72.02	1.5	4.5	59.55	39.33	2.25
8	0.64	63.45	2.5	3.5	51	39	14.00
9	1.43	86.30	5.5	7.5	52.63	40.53	14.04
10	2.04	114	6.5	8.5	35.42	36.46	16.67
All	1.77	88.85	4	6.8	45.02	39.51	12.79
Taggers	2.17 (2.17)	92.37 (92.4)	3.7	7.3	46.50 (40.97)	39.25	24.26 (11.65)

Values in brackets represent results of predictions with the use of tags (where available). In the last two rows, the average results are reported (the last row includes only results from users that tagged their photos)

Table 4 Average viewing times and viewing frequencies per photo for three different tasks: browsing, selecting and searching

User	Browsing		Selecting		Searching	
	Time (ms)	Frequency	Time (ms)	Frequency	Time (ms)	Frequency
1	5,851	1.02	2,642	1.37	770	1.01
2	4,034	1.01	1,406	1.09	752	1.03
3	2,875	1.04	1,500	1.13	751	1.10
4	4,431	1.02	1,823	1.04	649	1.09
5	3,191	1.00	2,649	1.39	6,768	3.25
6	2,003	1.00	1,584	1.01	872	1.04
7	3,385	1.00	2,401	1.33	1,405	1.09
8	1,899	1.00	1,250	1.06	835	1.04
9	1,895	1.14	2,105	1.19	2,494	2.40
10	3,985	1.00	2,182	1.09	2,183	1.04
All	3,355	1.02	1,954	1.17	1,748	1.41

photo comparisons where a longer viewing time corresponded to a higher rating (Table 4). Additionally, the analysis where we compared the best and worst 300 photos according to viewing times showed that viewing time is especially suited to discriminate between extremes, between good and bad photos. The following example will illustrate how viewing times can fruitfully be used to aid photo collection management. Researchers consistently report [5, 18] that people try to reduce the amount of photos they have to manage by erasing them from the camera (10 %) and on the computer during import or, later, by filtering and selection (8 %). This amounts to 17 % of the taken photos being permanently erased. We could do this automatically based on viewing times, which can be confirmed by looking at the gathered data; the 17 % of photos with lower viewing times have an average rating of 3.7 (SD = 2.4), while the other 83 % of photos have an average rating of 5.2 (SD = 2.8). In psychometric evaluations, moderate correlations (with absolute values as small as 0.30 or 0.40) are large enough to justify the use of psychometric instruments [20]. From this point of view, we can say that the observed correlation further confirms that viewing time is a measure of photo affinity.

4.2 Predicting ratings (H2)

Although we can see in Table 3 that in the classification part of the experiment we were able to improve the default classification accuracy (on average and in 7 out of 10 cases), this improvement and the absolute classification accuracy is too small to be considered a usable tool for automatic rating. A fact further supported by the “2 bin jump” metric is that on average, 12.79 % of the photos that are bad/good get classified as good/bad, which is unacceptable for such a system (from a usability point of view). It is interesting to notice the vast differences in classification accuracy between users; the highest accuracy was 72 % (user 3) and the lowest accuracy 47 % (user 6). One possible explanation for this could be that the features used for classification (viewing times, frequencies, EXIF metadata and photo histogram) are only capable of predicting photo ratings for a certain type of user. Both prediction models performed well on photos viewed by users 3, 7 and 8, and the characteristics that these three users have in common are low average ratings and photo-sharing habits

Table 5 Average moving averages for the last three and six photos viewed by each user during browsing, selecting and searching

User	Subset size = 3 (ms)						Subset size = 6 (ms)					
	Browse		Select		Search		Browse		Select		Search	
	Avg	Std	Avg	Std	Avg	Std	Avg	Std	Avg	Std	Avg	Std
1	5,728	3,296	1,917	778	758	347	5,726	2,181	1,914	574	755	248
2	3,922	1,000	1,247	362	733	146	3,919	708	1,232	240	740	104
3	1,868	448	1,180	288	784	351	1,860	316	1,178	228	757	247
4	1,897	804	1,541	418	818	179	1,858	530	1,529	318	792	105
5	3,088	1,045	1,863	688	2,163	397	3,047	809	1,857	515	2,128	162
6	2,730	1,199	1,328	563	674	300	2,711	968	1,325	464	666	246
7	3,917	2,630	1,989	802	2,111	1,419	3,947	2,328	1,948	612	1,930	973
8	1,563	625	1,715	691	973	720	1,534	401	1,670	532	946	473
9	4,358	1,557	1,755	800	547	396	4,366	1,287	1,758	591	496	267
10	3,329	934	1,779	615	1,255	596	3,313	714	1,759	487	1,166	313
All	3,240	1,354	1,631	600	1,081	485	3,228	1,024	1,617	456	1,038	314

Table 6 Average moving averages for all users for two subset sizes: 3 and 6

Subset size	Browse		Select		Search	
	Avg	Std	Avg	Std	Avg	Std
3	3,240	1,354	1,631	600	1,081	485
6	3,228	1,024	1,617	456	1,038	314

(organizing photo-viewing events). However, the sample of users and the amount of personal data collected is insufficient for any conclusions to be drawn. Similarly, we can say that tags do not improve rating prediction accuracy. In order to confirm this, more data are needed.

4.3 Identifying current basic photowork task (H3)

At the beginning of step 4, a user spontaneously stated "I know exactly which photo I will search for." Although systematic gathering and analysis of qualitative data such as this comment are out of the scope of this paper, we can nevertheless say that this anecdotal comment supports our assumption that step 4 of the experiment, where users were asked to choose one photo from the collection, is an example of searching behavior. Furthermore, in this step all of the users viewed only a subset of images.

Viewing times differ considerably across the users, ranging from 1.9 to 5.9 s during browsing, dropping to a range of 1.3–2.6 s while selecting and finally ranging from 1 to 3.2 s while searching. From Table 4 we can also see that browsing, selecting and searching can be distinguished by looking at the viewing times and frequencies. On average, participants spent 3.4 s looking at each photo when browsing, 2 s when selecting and 1.4 s when searching for a photo. On the one hand, viewing times decreased between tasks, while on the other hand, viewing frequencies increased from 1 view per photo during browsing to 1.2 views while selecting and 1.4 views per photo during searching. There were, however, two exceptions to this rule, user 5 and user 9; user 5 spent the most time viewing a photo while searching and was otherwise consistent with the majority of users, while user 9 behaved in exactly the opposite manner with regard to the other users in terms of viewing times (but not viewing frequencies). This user later admitted that he was distracted during the experiment and was trying to finish as quickly as possible. These two participants, however, help to make a precious point: A system for implicit photo collection management must be capable of handling noisy data. One way of designing such a system is to abandon the goal of fully *automatizing* photo collection management in favor of *aiding* the user to manage his/her collection of photos. This "*aid over automatize*" guideline is also supported by Frohlich and Fennell [21], who found that there is a great need for interfaces that help compare and sort photos. Kirk et al. [18] identified sorting, selecting and filtering as the basic tasks in photowork. The tasks observed in our experiments (browsing, selecting and searching) are similar to these activities. As can be observed in Tables 5 and 6, these tasks can be distinguished by means of moving averages of viewing times as soon as the user views a small number of photos. It is therefore possible to automatically customize the user interface based on the task at hand, for example, revealing and hiding tools for photo selection and editing or differentiation of photo caching to achieve faster interaction when photo quality is less important, that is, while searching.

4.4 Future work

One possibility for future work is to go beyond viewing times and explore other clues for implicit photo collection management, for example, smile detection using front-facing cameras on tablets. Next, when additional implicit clues are discovered, rating prediction could be made possible. At that point, a thorough study of different algorithms and features for learning prediction models would be imperative. Another possibility is to perform a longitudinal study of how viewing times and viewing frequencies behave in time and whether observing them for a longer period brings additional useful information for rating prediction. During the experiment, some participants spontaneously commented on the task they were performing. A future study could systematically collect and analyze these types of comments in order to shed additional light on the topic. As mentioned in Sect. 2, the focus of this paper was to capture the user's original reaction to photos and confirm that this information can be used to aid basic photowork tasks. Similarly, experiments could be conducted that aimed at identifying the differences between single- and multiuser implicit interactions with photos. Future research should also deal with the implementation of features based on implicit human–computer interaction that aid the user in managing his/her photo collection. A simple example would be correcting the orientation of a photo (if for some reason the information about the photo's orientation usually stored in EXIF data is lost); the user is viewing photos on a tablet in landscape mode until he/she encounters a portrait photo that is not rotated automatically. He/she will rotate the tablet to take a look at the photo; this information will be picked up by the motion sensors in the tablet and interpreted by the photo collection management software as a signal of incorrect photo orientation.

As mentioned in the introduction, this is, to the best of our knowledge, the first study that examines user behavior while engaged in basic photowork on a tablet device. We believe that the obtained results also apply to photowork on smartphones and cameras, although this should be verified along with whether and how these results apply to desktop computers—another possible direction for future research.

5 Contribution and conclusion

Current solutions to the problem of storing and maintaining large photo collections are based on various state-of-the-art technologies including clustering, social network analysis, image recognition, photo meta-data analysis and others. Each of them brings a valuable contribution to the puzzle of photo collection management but, at the same time, lacks the ability to consider the user—the owner of the photos and its corresponding memories—and his/her personal affinity for them—his/her subjective relationship to each photo and the memory it evokes. In this sense, the main contribution of this work is to provide the foundation for a different perspective on photo collection management. This perspective is user-centered and draws on implicit human–computer interaction. It is not intended as a replacement for current methods of photo collection management, but as an addition that tries to bring "the personal perspective" to photo collection management. As a cornerstone, we have shown that photo viewing time can be interpreted as a measure of the user's affinity for a particular photo and that it is possible to imply whether the user is currently browsing, selecting or searching for a photo.

Acknowledgments The authors would like to thank their friends and colleagues who contributed to the experiments. This research was funded in part by the European Union, European Social Fund, operational program for human resources, development for the period 2007–2013 and the Slovenian Research Agency.

References

1. Chalfen R (1987) Snapshot versions of life. Bowling Green State University Popular Press, Bowling Green, OH
2. Chalfen R (1979) Photograph's role in tourism: some unexplored relationships. Ann Tour Res 6(4):435–447
3. Chalfen R (1975) Introduction to the study of non-professional photography as visual communication. In: Folklore forum bibliographic and special series. Saying cheese: studies in folklore and visual communication, vol. 13, pp. 19–25
4. Good J (2011) How many photos have ever been taken? 1000 memories blog. Available at: http://blog.1000memories.com/94-number-of-photos-ever-taken-digital-and-analog-in-shoebox (Accessed December 2012)
5. Whittaker S, Bergman, O, Clough P (2010) Easy on that trigger dad: a study of long term family photo retrieval. Pers Ubiquit Comput 14. 1 (January 2010). 31–43. doi:10.1007/s00779-009-0218-7
6. Latif K, Mustofa K, Min Tjoa A (2006) An approach for a personal information management system for photos of a lifetime by exploiting semantics. In: Bressan S, Küng J, Wagner R (eds) Proceedings of the 17th international conference on Database and Expert Systems Applications (DEXA'06). Springer-Verlag, Berlin, Heidelberg. pp 467–477. doi:10.1007/11827405_46
7. Corcoran P, Costache G (2005) Automated sorting of consumer image collections using face and peripheral region image classifiers. IEEE Trans Consum Electron 51. 3 (August 2005) pp 747–754. doi:10.1109/TCE.2005.1510478
8. Graham A, Molina HG, Paepcke A, Winograd T (2002) Time as essence for photo browsing through personal digital libraries. In JCDL '02: Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries. ACM. pp 326–335
9. Platt JC, Czerwinski M, Field BA (2003) Phototoc: automatic clustering for browsing personal photographs. In: Information. Communications and Signal Processing. 2003 and the Fourth Pacific-Rim Conference on Multimedia. Proceedings of the 2003

Joint Conference of the Fourth International Conference on. pp 6–10 Vol.1

10. Rabbath M, Sandhaus P, Boll S (2011) Automatic creation of photo books from stories in social media. ACM Trans Multimedia Comput Commun Appl 7S. 1. Article 27 (November 2011). 18 pages. doi:10.1145/2037676.2037684
11. Girgensohn A, Adcock J, Cooper M, Foote J, Wilcox L (2003). Simplifying the management of large photo collections. In Human–Computer Interaction INTERACT 3: 196–203
12. Schmidt A (2000) Implicit human computer interaction through context. Personal technologies, 4(2–3), 191–199. Springer. Retrieved from http://www.springerlink.com/index/10.1007/BF01324126
13. Petrelli D, Whittaker S, Brockmeier J (2008) Autotopography: what can physical mementos tell us about digital memories? In: Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems (CHI '08). ACM, New York, NY, USA, 53–62. doi:10.1145/1357054.1357065
14. Frohlich D, Kuchinsky A, Pering C, Don A, Ariss S (2002) Requirements for photoware. In: Proceedings of the 2002 ACM conference on Computer supported cooperative work (CSCW '02). ACM, New York, NY, USA, 166–175. doi:10.1145/587078.587102
15. O'Hara K, Helmes J, Sellen A, Harper R, Bhömer M, Van den Hoven E (2012) Food for talk: phototalk in the context of sharing a meal. Hum Comput Interact 27(1–2):124–150
16. Hilliges O, Wagner M, Terrenghi L, Butz A (2007) The living-room: browsing, organizing and presenting digital image collections in interactive environments. Intelligent Environments, 2007. IE 07. 3rd IET International Conference on In Intelligent Environments, 2007. IE 07. 3rd IET International Conference on (2007), pp 552–559
17. Blažica B, Vladušič D, Mladenić D (2011) Shoebox: a natural way of organizing pictures according to user's affinities. Human–computer interaction. Towards mobile and intelligent interaction environments, Springer, Berlin, pp. 519–524
18. Kirk D, Sellen A, Rother C, Wood K (2006) Understanding photowork. Proceedings of the SIGCHI conference on Human Factors in computing systems. Held 2006 in Montreal, Quebec, Canada. ACM Press, New York, pp 761–770
19. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH (2009) The WEKA data mining software: an update; SIGKDD explorations 11(1)
20. Nunnally JC (1978) Psychometric theory. McGraw-Hill, New York, NY
21. Frohlich D, Fennell J (2007) Sound, paper and memorabilia: resources for a simpler digital photography. Pers Ubiquit Comput 11(2):107–116

5 Summary and Conclusion

This chapter concludes the thesis by summarizing the main results of the articles presented in Chapters 2–4 and discussing their role in relation to the hypotheses listed in the introduction. A short description of where the data and the software used in the experiments can be found is given. Finally, this chapter outlines some possibilities for future work and gives an overview of how the presented results could impact the use cases described in Section 1.1.5.

5.1 Results Summary

This thesis advocates the use of natural user interfaces to design and build context-aware systems. It focuses on *context acquisition and understanding* as one of the key challenges in the field of context-awareness and presents some solutions – all based on multitouch displays. More specifically, the thesis addresses four research hypotheses (presented in Section 1.2). The first states that NUIs are inherently context-aware, the second is that contextual information can further increase the expressive power of NUIs, the third that MT displays provide enough information to perform user identification, and the fourth that implicit human-computer interaction with NUIs is also rich with information and can be used as a source of contextual information. What follows is a brief summary of the results presented in the previous chapters and how they relate to the research hypotheses.

5.1.1 Research hypotheses - revisited

Hypothesis #1: Natural User Interfaces (NUIs) and Multitouch (MT) displays are, to some extent, inherently context-aware; the information they carry is sufficient to build context-aware systems. This is indirectly confirmed by the confirmation of the other hypothesis listed, as in all the experiments we used multitouch displays as they are – 'out of the box', without turning to additional hardware. The key thing to understand here is that the hardware that is used in NUIs - not only multitouch displays, but also depth sensors, mobile phones, etc. - is rich with information and that this information can and should be explored and exploited beyond simple gestural interaction. An example is this thesis' review of NUIs from a context-awareness perspective that shows how NUIs are a viable way towards context-aware systems.

Hypothesis #2: Understanding and exploiting context-awareness further extends the expressiveness of NUIs. Pointing devices like mice and joysticks represent the core of the graphical user interface and its WIMP paradigm of interaction. In (Buxton, 2009), Bill Buxton states that doing everything by manipulating just one point around the screen with a pointing device "gives us the gestural vocabulary of a fruit fly" and continues "we can not only do better, but as users, deserve better. Multi–touch is one approach to accomplishing this – but by no means the only one, or even the best." The next logical step is to try to further extend the expressiveness of MT displays and NUIs. This is exactly what we achieved by looking at NUIs as context-aware systems. For example, in Chapter 3, we present a clustering algorithm for hand detection on MT displays that together with a review of related methods shows how a feature like hand detection acts as an enabling factor for use cases such as cooperative gestures for co-located groupware. The context addressed here is represented by the knowledge of the number of hands and/or users involved with an application. Similarly, the MTi method described in Chapter 2 allows application designers to make use of identity, while Chapter 4 shows how NUIs expressiveness can also be extended with information that is conveyed implicitly between the user and the system.

Hypothesis #3: Multitouch displays provide enough information to perform user identification. To test this hypothesis, an identification method, the MTi, was developed and tested on a small (34 users) dataset obtained on a multitouch LLP display. The scalability of the method was evaluated on a larger hand-geometry (Dutağacı et al., 2008) database of 100 users. Additionally, a usability study with toy example applications was performed to illustrate the use of the MTi method and how users react to it. On both databases the method achieved an accuracy of around 94 %, while its SUS score of 79 suggests an above-average usability. The MTi method is based on a set of 29 features that transform input data (touch coordinates) so that identification is made possible with different off-the-shelf classifiers. Besides providing an important aspect of context, i.e. user identity, the MTi method serves as an additional proof of the potential of multitouch data and as an example of what can be achieved with this potential.

Hypothesis #4: The way we interact with NUIs can implicitly disclose additional (contextual) information that can be exploited by a context-aware system to better understand the user. The test for this hypothesis was the series of experiments in the scenario of photo organization with real users presented in Chapter 4. Observing viewing times and viewing frequencies proved to be useful to identify which photowork task the user is currently performing (searching, browsing or selecting) and as a measure for the user's affinity for a particular picture. As viewing times and viewing frequencies are not specific to multitouch displays or NUIs, it is possible to extend these findings to all types of interfaces. Furthermore, it is plausible to expect that exploring MT specific interaction cues such as flick gesture direction, speed and/or acceleration will disclose additional contextual information.

In terms of **goals and expected contributions** of the dissertation, it can be said that all were successfully met. First, Section 1.1 in the introduction provides a *thorough overview of the literature of the fields of context-awareness and natural user interfaces* with a special emphasis on multitouch displays. The summary of various definitions of natural user interfaces is of particular importance as, to the best of our knowledge, there is no such overview of this novel and emerging research field. Second, in Chapter 3, a *clustering algorithm for hand detection* on multitouch displays is defined and evaluated. The emergence of different solutions to the problem of hand detection on multitouch displays during the making of this thesis also highlights the importance of the problem. These solutions are also carefully reviewed in Chapter 3. The next contribution of the thesis is the definition of MTi — *a biometric method for user identification on multitouch displays.* Next, in Chapter 4, we present a proof-of-concept *model for implicit extraction of information from user interaction.* All the mentioned algorithms and concepts were not just theoretically envisioned, but an *implementation and evaluation on artificial and/or real world data* is also provided. Finally, this thesis' last contribution is also the *publicly available database for further research on user identification on multitouch displays* presented in the next section.

5.2 Software and Data Availability

In preparing the papers that form the core of this thesis care has been taken to assure a high level of repeatability of the studies concerning hand detection with clustering (Chapter

3), photo collection management with implicit human computer interaction (Chapter 4) and identification on multitouch displays (Chapter 2). Repeatability of the latter can be further improved by making the datasets and software used in the studies publicly available.

5.2.1 MTi datasets and software

In the article “MTi: a method for user identification on multitouch displays” two datasets and a sample application for the usability study were used. They are freely available from the following repository:

- Demo software: `https://bitbucket.org/bblazica/mti`
- MTi dataset: `http://goo.gl/TVNRZ`
- Bosphorus dataset: `http://goo.gl/k3Cs0`

The contents of both datasets are described in detail in the paper (Chapter 2), while a formal description of how the data is stored and how to use it can be found in the ‘readme.txt’ file accompanying both databases. The databases are available in CSV format. The Bosphourous database is partly re-published with kind permission from Bulent Sankur of the Boğaziçi University. These databases can be used to test improved versions of the proposed identification method or to evaluate completely new methods, provided that they only require the coordinates of touchpoints. Another possible direction of future use for the MTi database is a comparative study of how the MTi method behaves on multitouch displays of different construction (the MTi database was obtained on an LLP multitouch display).

The available software is a demo implementation of the method developed with Weka, the MT4j (multitouch for java) framework (Laufs et al., 2010) and uses Chang and Lin’s implementation of Support Vector Machines (Chang and Lin, 2011) for performing identification. The method’s implemented consists of an extension of the class AbstractCursorProcessor, a custom identification event (extends class MTGestureEvent) and a graphical component that supports the process of identification with visual feedback. Additionally, a component for enrollment of users in the identification model is provided. Finally, the toy applications described in Chapter 2 are also available. The software can help replicate the usability study carried out in Chapter 2 or can be used to develop other test cases for the MTi identification method.

5.3 Conclusions and Future Work

To conclude, we will look again at the use cases for multitouch interaction presented in Section 1.1.5 and see how the work presented in this thesis affects them. These use cases are: learning, problem solving and planning, information visualization, tangible programming, entertainment, play and edutainment, music and performance, social communication, and tangible reminders and tags. Possibilities for future work will also be discussed along each use case.

The article presented in Chapter 2 in its motivation section presents some studies concerning the direct application of multitouch displays in a *learning* environment. Perhaps the most valuable example is given by Rogers et al. (Rogers et al., 2009) who report that tangible interfaces have a positive impact on the participation of children with learning disabilities (and others who find it hard to talk or are incapable of verbal communication). Another example is the study of a public multitouch wall in the centre of Helsinki where Peltonen et al. (Peltonen et al., 2008) report that during interaction with the wall, some form of role-taking is needed in order to let users learn the opportunities for interaction from their peers. The MTi identification method presented in Chapter 2 as well as the hand detection method presented in Chapter 3 serve as a facilitator in the abovementioned learning-connected examples. Regarding implicit human-computer interaction, a possible way for future research is to explore various interaction cues to see if similar associations as viewing time and affinity can be found in the context of learning. For example, when a user answers a quiz-like question, the time taken and the type of gesture used to answer, or how the gesture was performed (direction, speed, acceleration) could disclose some information about how certain the user is about the answer.

Problem solving and planning and *information visualization* are perhaps two categories where the application of this thesis' findings is the most straightforward. The MTi method can be used for interface and visualization personalization, while for hand detection methods we have already identified cooperative gestures for co-located groupware as their most likely scenario of application. Some examples of this category are urban planning applications, command and control rooms for strategic planning, media walls used in news, etc. Obviously, the example used to explore the possibilities of implicit human-computer interaction - photo collection management - also falls under this category.

In the introduction we said that TUIs have long been used in *music and performance* applications and that one of the features that makes them so appropriate for this use case is their support for collaboration and sharing of control. Exactly the same feature is addressed by the hand detection methods presented in Chapter 3. It is also not difficult to imagine a personalization of a TUI-based instrument built with the MTi method or the involvement of an audience by means of implicit human-computer interaction in a live musical performance; similarly to YVision's [1] Audienceentertainment system that adopts direct interaction.

On the other hand, the contribution of this thesis to use cases such as *tangible programming*, *entertainment*, *play and edutainment*, and *social communication* is less pronounced, although some examples can still be found. We already mentioned the use of tabletops in collaborative tasks, which also falls under the social communication category and how hand detection and identification can support collaborative tasks (i.e. co-located groupware). Both these features can also be used to enrich the user experience in augmented traditional board games and interactive installations in museums.

Finally, the use case for multitouch displays connected to the category of *tangible reminders and tags* are vacation souvenirs that, when placed on an interactive surface, open an associated photo collection. In this example, the use of findings from Chapter 4 is straightforward: each time a collection is displayed, viewing times for photos and the organization of the collection can be updated. These principles of photo collection management by implicit human-computer interaction can also be applied directly to vacation souvenirs or other entities that form some sort of collection. Similarly as some photos make up a photo collection, some souvenirs can make up a souvenir collection. In this example, a souvenir collection can be understood as a collection on a higher hierarchical level. This means that a collection can be treated as an entity and the other way around, an entity can be treated as a collection. In the photo example discussed, a photo could represent a collection of regions of interest or pixels. The question open for future research is to define how exactly a collection should be managed by implicit human-computer interaction on different hierarchical levels and in different application domains. For example, knowing which region of the image is looked at the most could be used to automatically enhance the image by sharpening the most important region and blurring out the others (an effect often used by photographers to emphasize a part of the photo and make it more attractive).

[1] `http://www.audienceentertainment.com/portfolio-post.php?portfolioid=24`

Section 1.2 states that the purpose of this thesis is to show that natural user interfaces are a viable way towards context-aware systems. Based on the above-mentioned review of research hypotheses, how they were tested, the results obtained, and the use cases listed, it is plausible to conclude that the thesis met its purpose. The proposed methods for context-acquisition on multitouch displays bring additional functionalities and possibilities to application developers and thus further broaden the already rich interaction vocabulary that multitouch displays, and natural user interfaces in general, provide.

On a higher level, future research involving natural user interfaces and context-awareness could continue along the path traced by this thesis, i.e. exploring possibilities of context acquisition that arise from rich data available from NUIs. Once we know how to reliably acquire different aspects of context, research efforts should shift towards providing support for building context-aware systems and thus simplifying their development. This will lead to a more widespread adoption of such systems, which will in turn make researching their influence on human-computer interaction and evaluation, both, possible and important.

6 Acknowledgements

Thanks to my *supervisors* for introducing me to the world of research, for their advice, help, support and correct relationship. Thanks to my *friends* for listening to my whining about the occasional pointlessness of academic endeavours. Thanks to some *anonymous reviewers* for reminding me of the beauty of science. Thanks to my *colleagues* for their ideas and camaraderie. Thanks to *XLAB* for giving me the possibility and the freedom to conduct the research described in this thesis.

Finally, special thanks to my *family* – my wife, my parents and my brother – for all the above and more: for their love, support, encouragement ...

This research was funded in part by the European Union, European Social Fund, operational program for human resources, development for the period 2007-2013.

7 References

Aarts, E.; Wichert, R. Ambient intelligence. In: Bullinger, H.-J. (ed.) *Technology Guide.* 244–249 (Springer-Verlag Berlin, Heidelberg, 2009).

Abowd, G. D.; Mynatt, E. D.; Rodden, T. The human experience. *IEEE Pervasive Computing* **1**, 48–57 (2002).

Amerika, M. *Meta/Data: A digital poetics* (The MIT Press, Cambridge, 2009).

Atzori, L.; Iera, A.; Morabito, G. The internet of things: A survey. *Computer Networks* **54**, 2787–2805 (2010).

Benko, H.; Wilson, A. D.; Baudisch, P. Precise selection techniques for multi-touch screens. In: *Proceedings ACM CHI 2006: Human Factors in Computing Systems.* 1263–1272 (ACM, New York, 2006).

Bettini, C.; Brdiczka, O.; Henricksen, K.; Indulska, J.; Nicklas, D.; Ranganathan, A.; Riboni, D. A survey of context modelling and reasoning techniques. *Pervasive and Mobile Computing* **6**, 161–180 (2010).

Blake, J. *Natural User Interfaces in .NET* (Manning Publications Co., New York, 2013).

Blažica, B. *Izdelava večdotičnega zaslona in njegova uporaba v geoinformacijskem sistemu* (B.Sc. thesis, Faculty of electrical engineering, University of Ljubljana, Ljubljana, 2009).

Blažica, B.; Vladušič, D.; Mladenić, D. Shoebox: A natural way of organizing pictures according to user's affinities. In: Jacko, J. (ed.) *Human-Computer Interaction. Towards Mobile and Intelligent Interaction Environments.* 519–524, Lecture Notes in Computer Science (Springer-Verlag Berlin, Heidelberg, 2011). http://dx.doi.org/10.1007/978-3-642-21616-9_58. Accessed: May 2013.

Blažica, B.; Vladušič, D.; Mladenić, D. *HDCMD: a Clustering Algorithm to Support Hand Detection on Multitouch Displays* (Springer-Verlag Berlin, Heidelberg, 2013). In press.

Blažica, B.; Vladusič, D.; Mladenić, D. Ubiquitous personalization of a smartphone, used as a universal controller. In: *Workshop on Location-Based Services in Smart Environments (LAMDa'12).* 21–22 (2012). http://goo.gl/5nd3A. Online, accessed: May 2013.

Blažica, B.; Vladušič, D.; Mladenić, D. Mti: A method for user identification for multitouch displays. *International Journal of Human-Computer Studies* **71**, 691–702 (2013a). http://www.sciencedirect.com/science/article/pii/S1071581913000372. Accessed: May 2013.

Blažica, B.; Vladušič, D.; Mladenić, D. A personal perspective on photowork: implicit human-computer interaction for photo collection management. *Personal and Ubiquitous Computing* 1–9 (2013b). http://dx.doi.org/10.1007/s00779-013-0650-6. Accessed: May 2013. In press.

Brown, P. J.; Bovey, J. D.; Chen, X. Context-aware applications: from the laboratory to the marketplace. *Personal Communications, IEEE* **4**, 58–64 (1997).

Buxton, B. *Multi-Touch Systems that I Have Known and Loved* (2009). http://www.billbuxton.com/multitouchOverview.html. Online, accessed: May 2013.

Buxton, B. *Making user interfaces natural* (2010). http://research.microsoft.com/en-us/about/feature/nui-video.aspx. Online, accessed: March 2013.

Buxton, W.; Myers, B. A study in two-handed input. In: *ACM SIGCHI Bulletin.* **17**, 321–326 (ACM, New York, 1986).

Chalmers, D. *Sensing and Systems in Pervasive Computing: Engineering Context Aware Systems* (Springer-Verlag, London, 2011).

Chalmers, D.; Dulay, N.; Sloman, M. Towards reasoning about context in the presence of uncertainty. In: *Proceedings of Workshop on Advanced Context Modelling, Reasoning And Management at UbiComp 2004* (2004). http://goo.gl/GurEc. Online, accessed: May 2013.

Chang, C.-C.; Lin, C.-J. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology* **2**, 1–27 (2011).

Chen, H. *An Intelligent Broker Architecture for Pervasive Context-Aware Systems* (Ph.D. thesis, University of Maryland, Baltimore County, Catonsville, 2004).

De Nardi, A. *Grafiti - Gesture Recognition management Framework for Interactive Tabletop Interfaces* (B.Sc. thesis, Universita di Pisa, Pisa, 2008).

Dey, A. K. *Providing architectural support for building context-aware applications* (Ph.D. thesis, Georgia Institute of Technology, Atlanta, 2000).

Dourish, P. What we talk about when we talk about context. *Personal Ubiquitous Comput.* **8**, 19–30 (2004). http://dx.doi.org/10.1007/s00779-003-0253-8. Accessed: May 2013.

Dutağacı, H.; Sankur, B.; Yörük, E. Comparative analysis of global hand appearance-based person recognition. *Journal of electronic imaging* **17**, 011018–011018 (2008).

Engelbart, D. C. *Augmenting human intellect: A conceptual framework* (1962). http://www.dougengelbart.org/pubs/augment-3906.html. Online, accessed: May 2013.

Epstein, B. *Digital living room* (1998). http://tinyurl.com/bp95roa. Online, accessed: March 2013.

Fitzmaurice, G.; Ishii, H.; Buxton, W. Bricks: Laying the foundations for graspable user interfaces. In: *Proceedings of the ACMSIGCHI Conference on Human Factors in Computing Systems.* 442–449 (ACM Press/Addison-Wesley Publishing Co., New York, 1995).

Foehrenbach, S.; König, W. A.; Gerken, J.; Reiterer, H. *Natural interaction with hand gestures and tactile feedback for large, high-res displays* (Bibliothek der Universität Konstanz, Konstanz, 2008).

Frohlich, D. M. The history and future of direct manipulation. *Behaviour & Information Technology* **12**, 315–329 (1993).

Greenfield, A. *Everyware: The Dawning Age of Ubiquitous Computing* (New Riders, Berkeley, 2006).

Han, J. Low-Cost Multi-Touch Sensing through Frustrated Total Internal Reflection. In: *Proceedings of the 18th Annual ACM Symposium on User Interface Software and Technology.* 115–118 (ACM, New York, 2005).

Hay, S.; Newman, J.; Harle, R. Optical tracking using commodity hardware. In: *7th IEEE/ACM International Symposium on Mixed and Augmented Reality, 2008. ISMAR 2008.* 159–160 (IEEE, Washington, 2008).

Hayes, P. J.; Reddy, D. R. Steps toward graceful interaction in spoken and written man-machine communication. *International Journal of Man-Machine Studies* **19**, 231–284 (1983).

Hewett, T. T.; Baecker, R.; Card, S.; Carey, T.; Gasen, J.; Mantei, M.; Perlman, G.; Strong, G.; Verplank, W. *ACM SIGCHI Curricula for Human-Computer Interaction* (2009). http://old.sigchi.org/cdg/cdg2.html. Online, accessed: March 2013.

Hornecker, E. I don't understand it either, but it is cool - visitor interactions with a multi-touch table in a museum. In: *TABLETOP 2008. 3rd IEEE International Workshop on Tabletop Horizontal Interactive Human Computer Systems.* 113–120 (IEEE, Washington, 2008). http://dx.doi.org/10.1109/TABLETOP.2008.4660193. Accessed: May 2013.

Hutchins, E. L.; Hollan, J. D.; Norman, D. A. Direct manipulation interfaces. *Human-Computer Interaction* **1**, 311–338 (1985).

Indulska, J.; Sutton, P. Location management in pervasive systems. In: *Proceedings of the Australasian information security workshop conference on ACSW frontiers 2003-Volume 21.* 143–151 (Australian Computer Society, Inc., Darlinghurst, 2003).

Ishii, H.; Ullmer, B. Tangible bits: towards seamless interfaces between people, bits and atoms. In: *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems.* 234–241 (ACM, New York, 1997).

Jain, K.; Low, T. *Dj touch - an ftir touchscreen device* (2011). http://goo.gl/0iWPJ. Online, accessed: May 2013.

Jordà, S.; Geiger, G.; Alonso, M.; Kaltenbrunner, M. The reactable: exploring the synergy between live music performance and tabletop tangible interfaces. In: *TEI '07: Proceedings of the 1st international conference on Tangible and embedded interaction.* 139–146 (ACM, New York, 2007).

Kaltenbrunner, M. *Tangible musical interfaces* (2013). http://modin.yuri.at/tangibles/. Online, accessed: March 2013.

Korpipää, P.; Mäntyjärvi, J. An ontology for mobile device sensor-based context awareness. In: Blackburn, P.; Ghidini, C.; Turner, R.; Giunchiglia, F. (eds.) *Modeling and Using Context.* **2680**, 451–458 (Springer-Verlag Berlin, Heidelberg, 2003).

Kurfess, F. *Cpe\csc 486: Human-computer interaction* (2013). http://goo.gl/LEBc0. Online, accessed: March 2013.

Laufs, U.; Ruff, C.; Zibuschka, J. *Mt4j-a cross-platform multi-touch development framework* (2010). http://goo.gl/Ewob4. Online, accessed: May 2013.

Licklider, J. C. R. Man-computer symbiosis. *IRE Transactions on Human Factors in Electronics* **1**, 4–11 (1960).

Makkonen, J.; Avdouevski, I.; Kerminen, R.; Visa, A. Context awareness in human-computer interaction. *Human-Computer Interaction* (2009). http://goo.gl/prqoM. Online, accessed: May 2013.

Mauney, D.; Howarth, J.; Wirtanen, A.; Capra, M. Cultural similarities and differences in user-defined gestures for touchscreen user interfaces. In: *CHI '10 Extended Abstracts on Human Factors in Computing Systems.* 4015–4020 (ACM, New York, 2010). http://doi.acm.org/10.1145/1753846.1754095. Accessed: May 2013.

Maybury, M. Intelligent user interfaces: an introduction. In: *Proceedings of the 4th international conference on Intelligent user interfaces.* 3–4 (ACM, New York, 1998).

McAvinney, P. *The Sensor Frame - A Gesture-Based Device for the Manipulation of Graphic Objects* (Carnegie-Mellon University, Pittsburgh, 1986).

Morris, M. R.; Huang, A.; Paepcke, A.; Winograd, T. Cooperative gestures: multi-user gestural interactions for co-located groupware. In: *Proceedings of the SIGCHI conference on Human Factors in computing systems.* 1201–1210 (ACM, New York, 2006).

Motamedi, N. Hd touch: multi-touch and object sensing on a high definition lcd tv. In: *CHI'08 Extended Abstracts on Human Factors in Computing Systems.* 3069–3074 (ACM, New York, 2008).

Norman, D. A. Natural user interfaces are not natural. *Interactions* **17**, 6–10 (2010).

NUI Group Community. *What is a 'natural user interface'?* (2009). http://nuigroup.com/faq/. Online, accessed: March 2013.

Orel, R.; Blažica, B. 3fmt-a technique for camera manipulation in 3d space with a multi-touch display. In: *2011 IEEE Symposium on 3D User Interfaces (3DUI).* 117–118 (IEEE, Washington, 2011).

Oviatt, S.; Cohen, P. Perceptual user interfaces: multimodal interfaces that process what comes naturally. *Communications of the ACM* **43**, 45–53 (2000).

Partridge, G. A.; Irani, P. P. Identtop: a flexible platform for exploring identity-enabled surfaces. In: *CHI'09 Extended Abstracts on Human Factors in Computing Systems.* 4411–4416 (ACM, New York, 2009).

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: *Proceedings of the SIGCHI conference on human factors in computing systems.* 1285–1294 (ACM, New York, 2008).

Petersen, N.; Stricker, D. Continuous natural user interface: Reducing the gap between real and digital world. In: *8th IEEE International Symposium on Mixed and Augmented Reality, 2009. ISMAR 2009.* 23–26 (IEEE, Washington, 2009).

Rogers, Y.; Lim, Y.-k.; Hazlewood, W. R.; Marshall, P. Equal opportunities: Do shareable interfaces promote more group participation than single user displays? *Human–Computer Interaction* **24**, 79–116 (2009).

Rosenfeld, R.; Zhu, X.; Toth, A.; Shriver, S.; Lenzo, K.; Black, A. W. *Towards a universal speech interface* (PA school of computer science, Carnegie-Mellon University, Pittsburgh, 2000).

Saha, D.; Mukherjee, A. Pervasive computing: a paradigm for the 21st century. *Computer* **36**, 25–31 (2003).

Saponas, T. S.; Tan, D. S.; Morris, D.; Balakrishnan, R.; Turner, J.; Landay, J. A. Enabling always-available input with muscle-computer interfaces. In: *UIST '09: Proceedings of the 22nd annual ACM symposium on User interface software and technology.* 167–176 (ACM, New York, 2009).

Schilit, B.; Adams, N.; Want, R. Context-aware computing applications. In: *First Workshop on Mobile Computing Systems and Applications.* 85–90 (IEEE, Washington, 1994).

Schmidt, A. *Ubiquitous computing-computing in context* (Ph.D. thesis, Lancaster University, Lancaster, 2003).

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* (Technical University of Munich, Munich, 2008).

Shaer, O.; Hornecker, E. Tangible user interfaces: past, present, and future directions. *Foundations and Trends in Human-Computer Interaction* **3**, 1–137 (2010).

Shneiderman, B. The future of interactive systems and the emergence of direct manipulation. *Behaviour & Information Technology* **1**, 237–256 (1982).

Strang, T.; Linnhoff-Popien, C. A context modeling survey. In: *Proceedings of Workshop on Advanced Context Modelling, Reasoning And Management at UbiComp 2004* (2004). http://goo.gl/1Uvz1. Online, accessed: May 2013.

Ğetin G., S. S., Bedi R. *Multi-touch Technologies* (2009). http://tinyurl.com/d5g7qok. Online, accessed: March 2013.

Weiser, M. The computer for the 21st century. *Scientific American* **265**, 94–104 (1991).

Weiser, M. Some computer science issues in ubiquitous computing. *Communications of the ACM* **36**, 75–84 (1993).

Weiser, M.; Brown, J. S. The coming age of calm technology. In: *Beyond calculation.* 75–85 (Copernicus New York, New York, 1997).

Wigdor, D.; Forlines, C.; Baudisch, P.; Barnwell, J.; Shen, C. Lucid touch: a see-through mobile device. In: *UIST '07: Proceedings of the 20th annual ACM symposium on User interface software and technology.* 269–278 (ACM, New York, 2007).

Wigdor, D.; Jiang, H.; Forlines, C.; Borkin, M.; Shen, C. Wespace: the design development and deployment of a walk-up and share multi-surface visual collaboration system. In: *CHI '09: Proceedings of the 27th international conference on Human factors in computing systems.* 1237–1246 (ACM, New York, 2009).

Wigdor, D.; Wixon, D. *Brave NUI world: designing natural user interfaces for touch and gesture* (Morgan Kaufmann, Burlington, 2011).

Wikipedia. *Physical computing – Wikipedia, the free encyclopedia* (2013a). http://en.wikipedia.org/wiki/Physical_computing. Online, accessed: March 2013.

Wikipedia. *vi – Wikipedia, the free encyclopedia* (2013b). http://en.wikipedia.org/wiki/Vi. Online, accessed: March 2013.

Wixon, D. *The challenge of emotional innovation* (2008). http://vimeo.com/2893051. Online, accessed: March 2013.

Wolpaw, J. Brain-computer interfaces for communication and control. *Clinical Neurophysiology* **113**, 767–791 (2002).

Wood, R.; Ashfield, J. The use of the interactive whiteboard for creative teaching and learning in literacy and mathematics: a case study. *British Journal of Educational Technology* **39**, 84–96 (2008). http://dx.doi.org/10.1111/j.1467-8535.2007.00703.x. Accessed: May 2013.

List of Figures

List of Tables

Appendices

Appendix A: Bibliography

Publications related to the dissertation

1.01 Original scientific article

- Blažica, B.; Vladušič, D.; Mladenić, D. Mti: A method for user identification for multitouch displays. *International Journal of Human-Computer Studies* **71**, 691–702 (2013a). http://www.sciencedirect.com/science/article/pii/S1071581913000372. Accessed: May 2013.

- Blažica, B.; Vladušič, D.; Mladenić, D. A personal perspective on photowork: implicit human-computer interaction for photo collection management. *Personal and Ubiquitous Computing* 1–9 (2013b). http://dx.doi.org/10.1007/s00779-013-0650-6. Accessed: May 2013. In press.

1.08 Published scientific conference contribution

- Blažica, B.; Vladušič, D.; Mladenić, D. *HDCMD: a Clustering Algorithm to Support Hand Detection on Multitouch Displays* (Springer-Verlag Berlin, Heidelberg, 2013). In press.

- Blažica, B.; Vladušič, D.; Mladenić, D. Shoebox: A natural way of organizing pictures according to user's affinities. : Jacko, J. () *Human-Computer Interaction. Towards Mobile and Intelligent Interaction Environments.* 519–524, Lecture Notes in Computer Science (Springer-Verlag Berlin, Heidelberg, 2011). http://dx.doi.org/10.1007/978-3-642-21616-9_58. Accessed: May 2013.

Other publications

1.08 Published scientific conference contribution

- Blažica, B.; Vladusič, D.; Mladenić, D. Ubiquitous personalization of a smartphone, used as a universal controller. : *Workshop on Location-Based Services in Smart Environments (LAMDa'12).* 21–22 (2012). http://goo.gl/5nd3A. Online, accessed: May 2013.

- Orel, R.; Blažica, B. 3fmt-a technique for camera manipulation in 3d space with a multitouch display. : *2011 IEEE Symposium on 3D User Interfaces (3DUI).* 117–118 (IEEE, Washington, 2011).

Appendix B: Biography

Bojan Blažica was born on May 13, 1984 in Šempeter, Slovenia.

In 2008, he started working at XLAB Research. He received his B.Sc. in 2009 from the Faculty of Electrical Engineering, University of Ljubljana, with the thesis "Construction of a multitouch display and its use in a geographic information system". In 2009, he received funding for post-graduate studies for the Junior Researcher programme of the Ministry of Higher Education, Science and Technology of the Republic of Slovenia. As a PhD candidate he continued his work on tangible interfaces and natural user interfaces in general at the Jožef Stefan International Postgraduate School.

His initiative to connect Slovenian HCI researchers and practitioners led to the founding of a web-page dedicated to the Slovenian HCI community (`www.hci.si`), the organization of the HCI-SEE workshop and HCI-IS conference in 2013.