FullHand: Markerless Skeleton-based Tracking for Free-Hand Interaction

Srinath Sridhar, Gilles Bailly, Elias Heydrich, Antti Oulasvirta, Christian Theobalt

MPI–I–2016–4–002 September 2016

Authors' Addresses

Srinath Sridhar
Max Planck Institute for Informatics
Campus E 1 4
D-66123 Saarbrücken, Germany

Gilles Bailly
Group VIA
46 rue Barrault
75 013, Paris, France

Elias Heydrich
D-66123 Saarbrücken, Germany

Antti Oulasvirta
Electrical Engineering Building
Department of Communications and Networking
Otakaari 5, 13000 Aalto University, Finland

Christian Theobalt
Max Planck Institute for Informatics
Campus E 1 4
D-66123 Saarbrücken, Germany

Acknowledgements

This work was supported by the ERC Starting Grant CapReal. We would like to thank Monika Singh.

Abstract

This paper advances a novel markerless hand tracking method for interactive applications. *FullHand* uses input from RGB and depth cameras in a desktop setting. It combines, in a voting scheme, a discriminative, part-based pose retrieval with a generative pose estimation method based on local optimization. We develop this approach to enable: (1) capturing hand articulations with high number of degrees of freedom, including the motion of all fingers, (2) sufficient precision, shown in a dataset of user-generated gestures, and (3) a high framerate of 50 fps for one hand. We discuss the design of free-hand interactions with the tracker and present several demonstrations ranging from simple (few DOFs) to complex (finger individuation plus global hand motion), including mouse operation, a first-person shooter and virtual globe navigation. A user study on the latter shows that free-hand interactions implemented for the tracker can equal mouse-based interactions in user performance.

Keywords

free-hand interaction; gesture input; finger input; hand tracking; skeleton tracking

FullHand: Markerless Skeleton-based Tracking for Free-Hand Interaction

Srinath Sridhar[1], Gilles Bailly[2], Elias Heydrich[3], Antti Oulasvirta[4], Christian Theobalt[1]

[1]Max Planck Institute for Informatics
{ssridhar,theobalt}@mpi-inf.mpg.de

[2]LTCI, CNRS, Telecom ParisTech
gilles.bailly@telecom-paristech.fr

[3]eliashartz@yahoo.de

[4]Aalto University
antti.oulasvirta@aalto.fi

(a) (b) (c) (d)

Figure 1. (a) FullHand can track one or both hands with all fingers. (b, c) FullHand enables free-hand interactions for many applications such as virtual globes and first-person shooters. (d) We envision miniature multi-camera setups for hand tracking in the future. The blue cylinders represent cameras.

ABSTRACT

This paper advances a novel markerless hand tracking method for interactive applications. *FullHand* uses input from RGB and depth cameras in a desktop setting. It combines, in a voting scheme, a discriminative, part-based pose retrieval with a generative pose estimation method based on local optimization. We develop this approach to enable: (1) capturing hand articulations with high number of degrees of freedom, including the motion of all fingers, (2) sufficient precision, shown in a dataset of user-generated gestures, and (3) a high framerate of 50 fps for one hand. We discuss the design of free-hand interactions with the tracker and present several demonstrations ranging from simple (few DOFs) to complex (finger individuation plus global hand motion), including mouse operation, a first-person shooter and virtual globe navigation. A user study on the latter shows that free-hand interactions implemented for the tracker can equal mouse-based interactions in user performance.

Author Keywords

free-hand interaction; gesture input; finger input; hand tracking; skeleton tracking;

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

INTRODUCTION

Exploiting the exceptional dexterity of the human hand for computer input has been a prime goal for research on input devices and interaction techniques. *Hand articulation* refers to the coordinated movement of the 27 bones controlled by 38 muscles in the hand and the forearm [12]. Fingers are the most precisely controllable parts of the body in spite of high angular velocity in their movement. Although all degrees of freedom (DOFs) cannot be independently controlled, individuation of finger control becomes virtually perfect with practice [12]. However, common input devices used today, such as the mouse, tap only into a fraction of the hand's capacity.

Several tracking methods have been proposed to capture the articulation of the hand for interactive applications. They can be classified into two categories. (1) *Contact-based* methods measure joint angles with instrumented gloves, or they use fiducial markers on the skin tracked by cameras [31, 27]. However, these methods restrict free motion of the hand, and they can be uncomfortable and unpractical for users. (2) *Non-contact* methods, typically based on computer vision, do not require contacting sensors. However, existing methods have limitations related to the set of DOFs they capture or interactive performance. For instance, the Leap Motion tracks only salient points like fingertips, and only succeeds under a constrained range of hand orientations. This restricts designers to a narrow set of free-hand interactions.

In this paper, we present *FullHand*, a system for hand motion tracking and interaction. FullHand tracks the motion of the hand using a *kinematic skeleton* that captures the major rotational and translational degrees of freedom of the hand. FullHand has several advantages over previous methods: (1) it captures the motion of the hand with all fingers, (2) achieves a framerate of 50 fps for one hand, (3) has a low latency,

(4) achieves high levels of precision, (5) can reliably recover from tracking errors, (6) supports two-handed interaction and (7) enables rapid development of interaction techniques by offering an abstraction (skeleton).

Markerless tracking of finger motion (articulations) for HCI is a challenging problem because of the absence of discriminating image features, rapid motions, self-occlusions, the large number of possible poses and homogeneous colour distribution. At the same time, tracking fingers is essential for enabling free-hand interactions. Previous approaches have avoided this problem by using ad-hoc solutions to directly detect gestures without tracking fingers [30]. In order to find a more principled and robust solution to this problem, we extend a previous hybrid algorithm for hand tracking that captures finger articulations as a skeleton [24]. This method combines *generative* hand pose estimation with a *discriminative* one. The input to our method are RGB images from a calibrated camera video setup, monocular time-of-flight depth data and a hand model adapted to a person. The output are the global pose and joint angles of the hand as a skeleton. The kinematic skeleton provides a means for rapid design of free-hand interactions.

We describe how the tracker was developed to allow (1) low latency, (2) high precision, (3) coverage of typical motions in HCI, and (4) two-handed interaction. Results from a technical evaluation show an accuracy of 87% on a dataset of 19 annotated video sequences. Results from a *gesture elicitation* study to confirm the usefulness of FullHand for interaction tasks. FullHand allows users to perform gestures and multi-finger controls that were not possible with previous systems [28].

After presenting the technical contribution, we discuss the design of free-hand interactions using the kinematic skeleton. We build on previous work in 3D interaction and human factors to derive guidelines for free-hand interactions that exploit finger articulations. We designed and implemented free-hand interactions for navigation in 3D scenes, simulation of input devices, mid-air menu techniques and games. The designed interactions explore different capabilities of hand motion including finger articulations (upto 8 fingers) and global hand motion. For example, we demonstrate a mid-air menu selection technique that uses several fingers and terrain level flying with global hand motion (Figure 1). To critically assess if such interactions can be tracked and be beneficial for user performance, we conducted a study of virtual globe navigation.

To sum up, the primary contributions of this paper are:

- An extension to a previous hybrid approach for skeleton-based hand tracking for interactive applications.
- The design and implementation of interactive applications demonstrating the use of FullHand for hand and finger motion controls.
- A user study and a gesture elicitation study to validate the proposed approach.

RELATED WORK

Free-hand tracking for interaction is an old topic dating back to as early as 1979 [5, 6]. Several initial approaches, e.g. for virtual reality or robotics, were based on gloves [31, 27] to ease the problem of hand tracking. However, users may be reluctant to put gloves, especially during long work sessions or when they have to switch with other devices such as the mouse or the keyboard.

Markerless capture of free-hand motion and gestures with non-contact tracking is more challenging. As a result, only a few gesture sets have been proposed and most of the interaction techniques [2, 10, 30] are limited to pinching with one or two hands [30, 10]. This posture can easily be recognized even with RGB cameras but is sensitive to hand orientations and occlusion and does not exploit rich finger coordination.

With the introduction of infrared-based depth sensors like the Kinect, it has become easier and more robust to detect hand gestures. It has been used in large variety of applications such as tabletops [11], distant displays [3], and 3D desktops [15]. For instance, Keskin *et al.* [14] proposed a method for recognizing finger spelling in depth data. While these methods work well for application-specific hand interactions, they do not generalize and capture the full range of hand motions.

Markerless high DOF free-hand motion tracking for interaction has only recently been explored by Wang *et al.* [28] for 3D CAD modelling. However, this method uses only part of the hand motion space for interaction (6 DOF). Articulated hand motion tracking continues to be a challenging computer vision problem which has restricted its application in interaction scenarios. Techniques for hand tracking can be divided into *generative* and *discriminative* methods [9]. Table 1 shows a comparison of state-of-the-art hand tracking methods with respect to their usefulness in interactive applications.

Generative methods employ a hand model (*e.g.* kinematic skeleton) and synthesize a *pose* for the model that best explains the input (e.g. [17, 22]). For instance, Oikonomidis *et al.* [21] used a depth sensor and a method based on particle swarm optimization to achieve a frame rate of 15 fps. Other generative approaches [4, 25, 16] suffer from large computations times and are thus unsuitable for interaction.

Discriminative methods use prior knowledge about hands (*e.g.* pose database) and try to explain the input images based on this knowledge. One such method that uses a pose database was proposed by Athitsos and Sclaroff [1]. This idea was further explored by Wang and *et al.* in both color glove-based [29] and markerless variants [28].

Recently, a hybrid method for single hand tracking in a motion capture setting was proposed by Sridhar *et al.* [24]. This method was able to track one hand at 10 fps which is insufficient for interactive applications. In this paper, we extend their hybrid method to realtime (50 fps), single and bimanual hand motion tracking. We also demonstrate our method for interaction on a wide range of applications.

HAND MOTION TRACKING

In this section, we describe our method for articulated hand tracking that is inspired by the hybrid approach of Sridhar *et*

Method	Speed	Type	Accuracy	No. of Cameras	Applications
Leap Motion	200 fps, Low latency	Salient points, multiple hands	-	1	Navigation, games, etc.
FORTH [21, 22]	15 fps, High latency	Primitive model, 2 hands	10mm	1	Object interaction
6D Hands [28]	20 fps, Low latency	Skeleton, 2 hands	-	2	3D CAD Modelling
MPI [24]	10 fps, High latency	Skeleton, 1 hand	13mm	6	-
Intel [17]	60 fps, unknown latency	Unknown, 2 hands	-	1	Virtual object interaction, etc.
Ours	50 fps, Low latency	Skeleton, 2 hands	$<$ 15mm, 87% times	$5+1$	Virtual globe, games, etc.

Table 1. Comparison of various hand tracking techniques for interaction.

al. [24]. We chose this particular hybrid approach because it is well suited for interaction applications. The generative component of the hybrid method lends itself for fast optimization which is suitable for interaction but prone to local optima issues leading to wrong hand and finger pose. But when combined with a discriminative component this issue is alleviated leading to better hand and finger pose. We now describe our setup, briefly summarize the hybrid method and explain specific extensions that we have made to enable fast bimanual tracking.

Physical Setup

Hand Modelling: Before capturing hand articulations we create a kinematic model for the hand consisting of 32 joints. We model the hand with 26 degrees-of-freedom (DOFs) (20 joint angles, 3 global rotations and 3 global translations), a common approach in motion tracking and computer graphics [21, 26]. The joint angles of the model are limited to a fixed range taken from anatomical studies [23]. The final output of hand tracking are the joint angle parameters, Θ. Since the size of hand differs for each person, we incorporate 3 scaling parameters which allows us to customize the skeleton for each participant. This step is currently not automatic, but simultaneous calibration of hand shape and pose in a calibration step is possible [26].

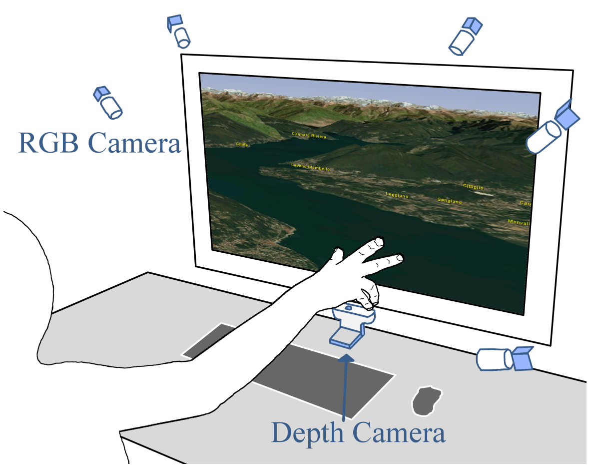


Figure 2. Our tabletop setup requires 5 **RGB cameras and** 1 **depth sensor.**

Figure 2 shows the physical setup for hand motion tracking and interaction. It consists of 5 RGB cameras and 1 depth sensor. The image data from RGB cameras provides high visual accuracy for tracking. The complementary single-view depth data helps us to retrieve poses effectively. The setup also consists of a large television screen for interaction and visual feedback. The setup requires calibration of the cameras for intrinsic and extrinsic camera parameters.

While we realize that such a setup is currently cumbersome to setup, we believe that in the future, miniature cameras (see Figure 1) and ambient cameras in homes and offices will become widely available. Moreover, there is also some work in the computer vision literature on reducing the number of cameras needed and eschewing camera calibration completely.

Tracking Algorithm

Markerless optical hand tracking is our approach of choice as it requires no interference with or instrumentation of the hand in any form. However, it is an inherently hard problem because of the large number of DOFs, fast motions, homogeneous skin color distribution and self-occlusions. In the past, numerous approaches for hand tracking have been proposed, which can be roughly classified into generative and discriminative methods. However, both classes of methods in isolation suffer from issues that make them unsuitable for interaction tasks. Generative methods optimize a 3D model-to-image consistency measure, E. Fast generative trackers use local optimization of this energy that tends to converge to erroneous local pose optima, e.g. leading to *sticky fingers* – two fingers overlapping each other on the image. Discriminative methods aim to infer hand poses from a learned space of plausible poses by means of extracted features. In this context, many approaches index into the hand pose space, and suffer from scaling problems due to exponential database sizes for high DOF models. In this work, we adopt a hybrid approach which combines generative and discriminative tracking, and which exploits their non-congruent failure modes for mutual benefit [24].

Estimation of the hand pose parameters (see Figure 3), Θ, at a time step of video is performed by running two tracking strategies in parallel. The first strategy is a generative tracker that uses multi-view color images, and that relies on a Sum-of-Gaussians scene representation, originally introduced by Stoll *et al.* [26]. It represents the hand in 3D by a kinematic bone skeleton, to the bones of which a discrete set 3D Gaussian functions are attached. Each Gaussian function is assigned a color, too. Similarly, each 2D image is decomposed into regions of similar color by means of a quad-tree

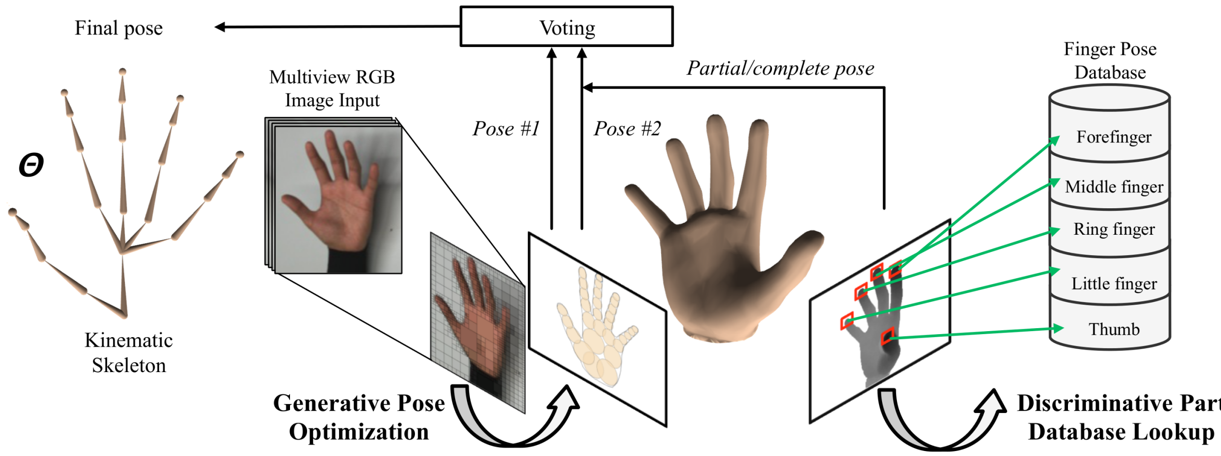


Figure 3. The tracking algorithm is a combination of a generative and discriminative method.

decomposition, and to each region a 2D Gaussian with associated average color is fitted. The hand pose is found by optimizing the overlap between the 3D hand SoG model with all 2D image SoG models. The SoG representation enables the definition of a 3D-2D consistency measure that has analytic derivatives. In addition, the consistency measure can be defined as a smooth function, lends itself to efficient parallelization, and can be effectively optimized with a fast conditioned gradient ascent solver that is initialized with an extrapolated solution from preceding time steps.

The second strategy is a discriminative pose estimation algorithm that uses images from the depth camera. It relies on a part-based strategy that estimates the pose of each finger separately rather than the full pose simultaneously. This is achieved by extracting fingertips on the depth image using a linear support vector machine (SVM) classifier, and by using the detected positions to find the closest match in multiple exemplar finger pose databases. Having separate databases for each finger has several advantages. The part-based strategy enables compartmentalization of the database and effective indexing into a much more densely sampled pose space than with a database storing full hand poses. Further on, with our method even partial hand poses can be found, for instance if some fingers are occluded.

Both tracking strategies yield a pose hypothesis for the hand. The final pose hypothesis is either (1) the solution from generative tracking, or (2) the solution from generative tracking initialized with the outcome of discriminative pose estimation. A final voting step selects the best solution based on the generative consistency measure, E.

Fast Bimanual Tracking in a Tabletop Setting

We have improved the above tracking strategy in several ways to enable fast one and two handed tracking. First, we enable realtime, low latency tracking by exploiting the algorithmic design of the tracking. Second, we enable two handed tracking which captures the articulations of all fingers. Finally, we show that the hybrid method can be optimized to work well in a tabletop setting instead of the controlled studio environment that was used by Sridhar *et al.* [24].

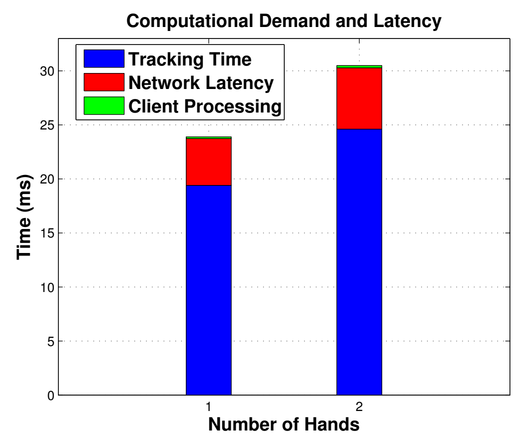


Figure 4. Plot of the computation times for one and two hands.

Both the generative and discriminative components of the algorithm lend themselves well for parallelization which we exploit. For the generative method, we use the structure of the consistency measure that allows parallel computation during pose optimization. The discriminative method detects fingertips on the depth image using the *sliding window* technique. We run multiple sliding windows on non-overlapping parts of the image in parallel which leads to lower computations times. Moreover, the two instances of the generative method run in parallel for even more gains. Overall, our average computation times were 3 to 4 times better than those reported by Sridhar *et al.* Figure 4 shows a plot of the computation times of the tracker averaged over 3, 1000 frame runs with a user performing slow and fast hand motions. The average time to process one frame was 19 ms (50 fps).

For bimanual tracking, we created a kinematic skeleton for both hands which together consist of 65 joints and 53 DOFs. Since the computation times are proportional to the number of DOFs of the hands and the fingertips to be detected on the depth image, our computational performance reduces to 20-40 fps. However, this is still sufficient for realtime interaction. Figure 4 shows a comparison of our computational performance for both single and two hands. For interacting with applications we send the tracked hand (along with gestures which are described later) over the network on a WebSocket protocol. Figure 4 shows the network latencies along with the tracking performance.

Finally, we setup our cameras in a tabletop setting (Figure 2) to match real world conditions. By tuning the parameters of the Sum-of-Gaussians representation we were able to achieve tracking performance comparable to Sridhar *et al.* Section shows a plot of tracking accuracy for our gesture elicitation study. Figure 6 show sample tracking results with one and two hands.

GESTURE ELICITATION STUDY AND ACCURACY ASSESSMENT

In order to understand the kind of gestures that users prefer for interaction and the tracker's capability in covering these, we conducted a gesture elicitation study. In this context we define a gesture to be a semantically meaningful motion of the hand within a given temporal period (e.g. pinching). We chose 6 student volunteers to participate in this study. All participants were right-handed males with a mean age of 29.2 (SD = 5.0). None of the participants had prior experience using or developing free-hand gestures.

Method

We prepared static images of interaction scenarios representative of the four interaction sub-tasks.

1. Navigation: Participants were presented with images of a virtual globe in both space and terrain viewpoints. They were asked to visualize navigating to cities and flying through buildings and valleys.
2. Selection: Participants were presented with images of a grid menu with 16 items and instructed to simulate selection of three highlighted items.
3. Manipulation: Three primitive objects were shown at random positions on the screen. The participants were instructed to simulate selecting and moving these objects so that they aligned vertically.
4. System Control: Participants were shown images of window switching and photo flipping and were asked to simulate this using hand gestures.

We presented static images instead of video sequences because we found in a pilot study that the interaction technique used in the video (eg. mouse for navigation) biased the kind of gestures that participants elicited. We gave participants 3-5 minutes to think of the gesture that they wanted to perform for each task. They were then asked to orally explain their gesture. Finally, we recorded them performing that gesture using our multi-camera setup.

Results

Participants were allowed to use global hand motion, all finger motion and both hands. When participants repeated the same gesture for two tasks they were asked to perform a different one. In all, we recorded a total of 28 sequences consisting of 22061 multi-view image frames. Table 2 shows a classification of the elicited gestures based on the type, number of hands and fingers that participants used for each task. In Table 2 we summarize the results of the elicitation study based on the number of hands and fingers that participants used. Users performed gestures that included pointing for navigation, finger waving for the selection, *swipe*-like gesture for manipulation and wrist rotation for system control.

Task	One Hand	Two Hands	Avg. No. of Active Fingers
Navigaton	3	3	1.5
Selection	6	0	2.3
Manipulation	5	1	2.0
System Control	5	1	1.2

Table 2. Results from the elicitation study showing number of participants who used one or two hands.

Since we recorded all elicited gestures, we also gained a large multi-view image sequence corpus as a dataset for evaluating the accuracy of hand tracking. While a few datasets exist for measing hand tracking performance, our dataset is specifically of users performing gestures for interaction tasks. In order to show that our tracking method is able to track the gestures elicited, we manually annotated (fingertip and palm locations) the elicited gestures for 3 out of the 4 tasks. Because of the large size of our dataset, we subsampled the data by annotating once every 10 frames. We adopted the tracking error of Oikonomidis [21]. We then measured the tracking accuracy to be the percentage of total frames in a sequence that had an error of less than 15 mm. A plot of this measure averaged over all datasets for each user is given in Figure 6.

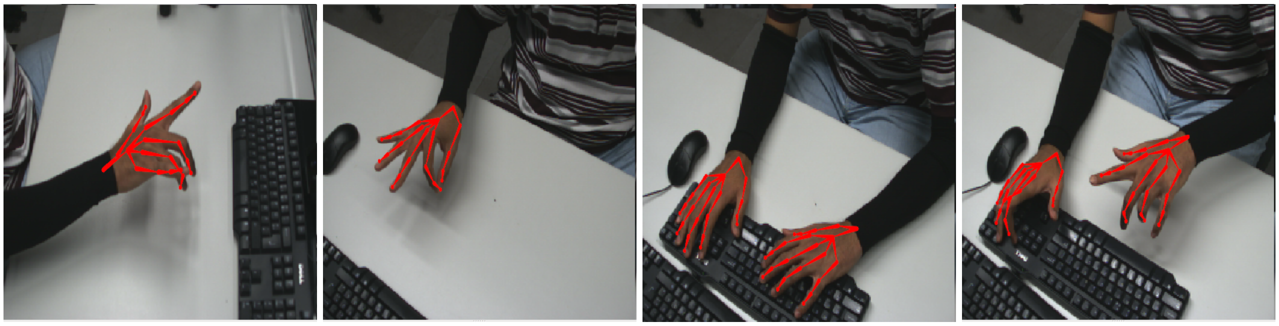
We were able to track an average of 86% of the total frames at an accuracy 15 mm or better (after subsampling). The dataset that we have recorded is useful both from the user perspective and the tracking perspective. To our knowledge, such a large dataset with specific free-hand gestures for markerless free-hand tracking is not currently available.

DESIGNING FREE-HAND INTERACTIONS

Skeletal representation of hand motion provides a rich and flexible means for designing free-hand interactions. This section outlines the design problem and collects guidelines from previous literature. We then present multiple examples of interactions designed for FullHand using these heuristics and guidelines to demonstrate the capability of hand tracking and the effectiveness of the skeleton-based approach.

The problem in designing free-hand interactions is that the motion space is large and there are multiple ways to map them. Based on previous literature, the design problem can be split into four sub problems: Task Description, Gesture Definition, Gesture Mapping (assigning of a gesture to a task) and parameter optimization.

First, a *task* can be split into multiple sub-tasks. Previous work suggests splitting each sub-task into two- or three-dimensional tasks [13, 18]. Each sub-task, in turn, can address Navigation, Selection, Manipulation, or System con-



(a) Tracking a hand pointing. (b) Tracking a hand pinch. (c) Tracking two hands typing on a keyboard. (d) Another example of tracking two hands.

Figure 5. Sample results from the hand tracking algorithm.

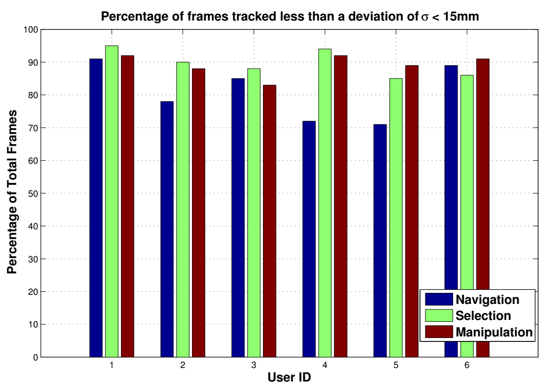


Figure 6. Plot of accuracy defined as the percentage of frames with error < 15 mm.

trol [7]. Second, designers should define the set of *gestures* they want to use. Selection of gesture sets depends on many factors including ergonomic considerations and technical constraints of the gesture recognizer. Third, the designer has to *map* appropriate gestures to a UI control task. Finding the right assignment of gestures to tasks and sub-tasks is not an easy problem. Different users use different kinds of interactions for the same task and one way to find commonality is through elicitation studies such as the one we conducted .

Finally, once a mapping has been defined, designers need to optimize the technique and choose appropriate transfer functions between hand motion and virtual motion for each subtask. A small amplitude gesture can trigger a small or large displacement on the screen. This requires user trials and constant improvement by the designer. To further inform design choices, we collected several guidelines from previous literature on hand interaction, human hand functioning, and motor control. Table 3 presents several guidelines under these categories.

In the above discussion we have not mentioned the effect of the hand tracking or gesture recognition component in designing interactions. Often, limitations in hand motion tracking or gesture recognition leads designers to come up with gestures that are easier to detect rather than easier for users.

Finger individuation
F1. The principal motions of the digits of the hand are extension/flexion, apposition/opposition of the index and the thumb, and the abduction/adduction of digits [19]
F2. Use index finger and thumb for independent controls [12]
F3. Avoid simultaneous control by middle, ring, and little finger [12]
F4. Allow tremor [12]
Motor control
L1. For higher skill, favor motions that are familiar [12]
L2. Only use the necessary maximum of degrees of freedom [28]
L3. Choose memorable gestures [28]
L4. Directions of motion should be congruous between hand and VE
L5. Performance increases when shoulder muscles can contribute to control [20]
Ergonomics
E1. Avoid hyperextension of fingers
E2. To minimize muscular loading, reduce global motion [28]
E3. Avoid continuous isometric tension of large muscles [8]
E4. Provide a rest for elbow and forearm [28]
E5. Elbow angle should be around 90 degrees [8]
E6. Place the display for comfortable body posture [28]

Table 3. Guidelines for free-hand interaction design from previous literature.

In this context, FullHand offers more flexibility because we track the continuous skeleton motion of the hand and detect gestures on the tracked skeleton. In our current work we adopt a heuristics-based approach which is quick to implement and robust enough to enable interactions. For instance, to detect pinch gestures, we use the position of the thumb tip and the fore finger tip as a measure.

FREE-HAND INTERACTION APPLICATIONS

In order to demonstrate the capability of the tracker and the skeleton-based approach for interaction, we show applications that (1) span different kinds of tasks (navigation, manipulation, selection and system control) and (2) employ fingers, one hand and bimanual input for interaction. Table 4 lists the applications based on the type of control task and the number of hands and fingers involved. We now discuss each in turn.

Interactive Application	Navigation	Selection	Manipulation	System Control	No. of Hands	No. of Fingers
Space Invaders	●	●	○	○	1	2
Menu Selection	○	●	○	○	2	8
Mouse	●	●	●	●	1	2
Virtual Globe	●	○	○	○	1	3
FPS	●	●	◒	○	1	3

● fully covered ◒ partially covered ○ not covered

Table 4. Comparison of different interaction applications based on the sub-tasks involved and biomechanical simulation against traditional instruments for physical ergonomics costs.

Navigation + Selection: Space Invaders

Space Invaders, a popular arcade game, combines a one dimensional navigation (maneuvering) and discrete selection (shooting) task. We use a pinch gesture similar to that shown in Figure 8 where it is used as a discrete selection event to shoot. To move the spaceship on the screen, we use the raw hand position data we receive from the tracker. Qualitative tests of this interaction technique showed that users were able to successfully complete the game.

Two-Handed Interaction: Menu Selection

In this application we show that users are able use both their hands for interacting for a *menu selection* task. We simulate a menu consisting of 8 items and use a pinch gesture recognizer to detect pinching of all fingers with the thumb. Each pinch gesture is a discrete event and is mapped to one item on the menu. The technique demonstrates two-handed interaction for selecting commands without requiring the visual modality.

Emulation of Input Devices: Mouse

FullHand can also be used to emulate existing input devices such as the keyboard or the mouse which capture less DOFs. Using virtual input devices have the advantage of reducing the cost for switching from one device to another one. Check the Imaginary devices advantages. By capturing slightly exaggerated versions of typical hand and finger motions required for *e.g.* a mouse, FullHand is able to stand-in for that device's functionality. Moreover, FullHand provides more degrees of freedom than existing hand trackers such as the Leap motion, making it possible to emulate this input device.

3D Navigation: Virtual Globe

Virtual globes such as Google Earth or NASA WorldWind [1] are an example of a 3D navigation task. They benefit from free-hand control because of the nature of the task involving multiple degrees of freedom. In this example, we used NASA WorldWind and connected it using WebSocket to the obtain the raw joint angle parameters and recognized gestures.

We divide virtual globe navigation into two distinct *viewpoints* and propose two techniques to control navigation in each viewpoint. Although they are different techniques, they are compatible with each other.

[1] `http://worldwind.arc.nasa.gov/`

Space Viewpoint: This mode is active when the camera is 4 km or above the globe's surface. In this viewpoint there are three parameters that are controllable – the latitude, longitude and altitude. To control *altitude* (zooming) users perform a *pinch gesture* as shown in Figure 8. The distance between the thumb and the forefinger on the tracked skeleton defines a rate based control of zooming. A *dead zone* (a region where motion is ignored) of 30 mm centered around the natural arched distance between thumb and forefinger is used when no control is wished. The pinch gesture is one of the principal hand motions and is easy to perform for users.

To control *latitude* and *longitude* (panning), users can choose between two gestures – one that involves clutching and one that does not. The clutch-based gesture uses the flexion angle of the middle finger as a *delimiter* that enables panning relative to the current position of the hand. We observed that this gesture is a good delimiter since it can be moved without affecting the fore finger and the thumb and is seldom performed accidentally by users. For a comfortable flexion angle, a pilot study shows that 45 degrees is good compromise between robutness and comfort. For the clutchless gesture, the position of the hand on the table relative to a predefined center indicates both the direction of the pan and the speed as shown Figure 8. Furthermore, we introduced a circular dead zone of 200 mm diameter which worked well for many users. In designing this interaction technique for the space viewpoint, we followed several of the guidelines introduced earlier in designing this interaction (F1, L1, L3 and E1).

Terrain Viewpoint: This viewpoint is automatically activated below 4 km and has 7 camera parameters that are controllable (pitch, roll, yaw, latitude, longitude, heading and altitude). Figure 9 shows the gestures for controlling the camera parameters. The pitch, roll and yaw are controlled by the same metaphor as a *flying vehicle* which is familiar to many users. However, we also allow users to fly forwards and backwards by means of a delimiter which is the flexion of the thumb.

This interaction choice was a direct result of a pilot study that we conducted that showed that the *flying vehicle* metaphor was the most natural for users.

We provide users with a visual cue by means of a smooth camera transition when the 4 km mark is reached. The user can then seamlessly switch from one technique to another. We refer the reader to section for a user study conducted using the interaction techniques described here and the supplementary material for videos showcasing this interaction.

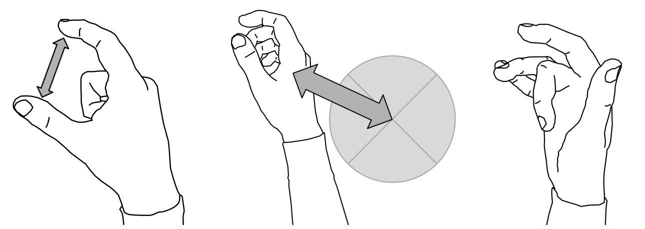
Multiple Controls: First-person Shooter

In order to demonstrate that we are able to track more complex tasks that involves navigation and selection in a time-critical environment, we created free-hand interactions for a first-person shooter game [2]. General movement of the character was performed by isometric hand motions similar to the Virtual Globe's space viewpoint. For instance, once the middle finger is clutched moving the hand to the left would cause the character to sidestep to the left. Aiming was performed akin to the Virtual Globe's terrain viewpoint, for shooting a

[2] `http://bytonic.de/html/jake2.html`

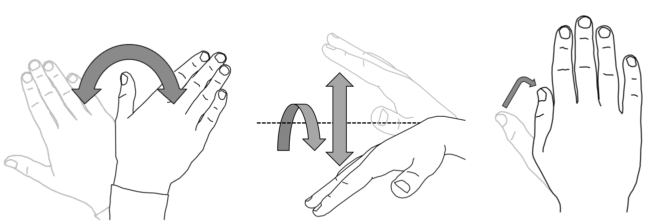
(a) Virtual globe, space viewpoint (b) Virtual globe, terrain viewpoint (c) FPS (d) Space Invaders

Figure 7. Examples of interaction applications made possible by FullHand.



(a) Pinching to zoom (b) Panning (c) Clutching

Figure 8. Hand motions for interacting with a virtual globe in *space viewpoint*.



(a) Yawing (b) Pitching (c) Moving

Figure 9. Hand motions for interacting with a virtual globe in *terrain viewpoint*.

pinching gesture analog to Space Invaders was used. Figure 7 shows screenshots from many of the above examples.

Please see the supplementary video for sample results from all the above applications.

STUDY OF VIRTUAL GLOBE NAVIGATION

To evaluate the capability of the tracking approach for finger articulations in interaction, we conducted a user study with the virtual globe application. We compared performance in four navigation tasks against the default mouse-based interaction option in WorldWind. The mouse controlled virtual globe navigation through the left, right and middle buttons along with motion. Free-hand interactions are pinching for zooming, hand motion with clutching for panning and palm orientation for orientation as in Figures 8 and 9.

We chose the mouse as the baseline, because it provides a hard benchmark. Most computer users have thousands of hours of experience in mouse pointing, including uses for navigation tasks and 3D environments. To our knowledge, this is the first comparative user study using a *markerless* approach for articulated hand tracking.

Method

The participants were six postgraduate student volunteers, all male and right handed, with a mean age of 29.5 years (SD = 4.93 years). All participants confirmed that they use the mouse on a daily basis. The four navigation tasks, illustrated in Figure 11, were:

1. Cities: Flying between cities in different continents with city-sized target circles of size 1 km. The route length was in the order of 20000 km. This task was repeated 5 times.
2. Continents: Moving between continents in the space viewpoint where the entire globe is visible. The circle target size was of the order of 1000 km. The route length was of the order of 15000 km. This task was repeated 10 times.
3. Villages: Moving between regional towns. The average route length was 50 km. This task was repeated 10 times.
4. Terrain: Moving along valleys and rivers at the terrain level. The average route length was 150 km. This task was repeated 3 times.

In tasks 1-3, the user had to move the camera viewpoint through a predefined sequence of areas that were highlighted as circles on the globe's surface. Task 4 involved moving the camera through ring-shaped posts at a terrain level where natural formations like mountains and rivers serve as visual assists. The sizes of the areas ranged from continent-sized to about one kilometer radii. A waypoint was considered selected when a crosshair in the center of the display was brought on top of it. Since the users had no previous experience with hand tracking, each task was repeated multiple times with both interfaces. To eliminate order effects, half of the participants performed the tasks with the mouse first, while the other half started with the tracker. The order of Tasks 1-4 was randomized.

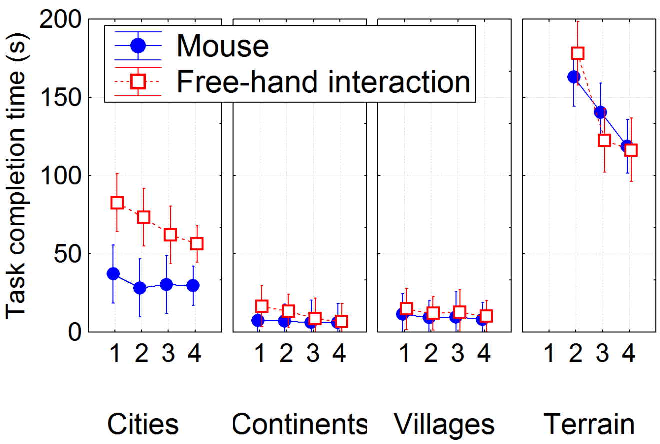


Figure 10. Development of task performance for mouse vs. free-hand interactions in four navigation tasks with the Virtual Globe. Free-hand performance becomes virtually equivalent with the mouse in 3 out of 4 tasks after a few trials. Vertical bars denote 95% confidence intervals.

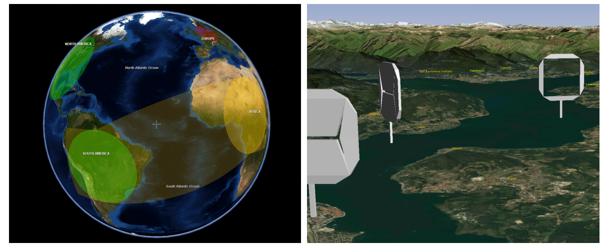
Results

The analyzed dataset has altogether 327 trials. For statistical testing, we performed a 4 (Task) $\times$ 24 (Interface) repeated measures ANOVA.

Figure 10 provides an overview of the trends with 95% confidence intervals. Not surprisingly, the effect of Task was significant, $F(1, 319)=302.5$, $p<0.001$. We also obtained a significant effect of Interface, $F(1, 319)=11.7$, $p=0.001$. Alas, performance with the mouse was better. However, a closer analysis of the tasks showed that this difference is attributable to Task 1. The interaction effect Task $\times$ Interface was significant, $F(3, 319)=7.5$, $p=<0.01$. Figure 10 suggests that user performance in Tasks 2-4 was equal with the mouse in the latter half of the repetitions. In contrast, in Task 1, performance with the mouse was always better. A Post Hoc comparison (Bonferroni) against the two showed a statistically significant difference between the mouse and free-hand interactions only for Task 1 ($p<0.001$).

To sum up, parallel performance was achieved for 3 out of 4 tasks. Given the small number of trials and the lack of previous experience with hand tracking, we consider this result promising. Furthermore, we learned that the poor performance with the tracker in Task 1 is due to hand tremor caused by the absense of an arm rest.

CONCLUSION AND FUTURE WORK



(a) Continents Task (b) Terrain Task

Figure 11. Examples of tasks used in the study of Virtual Globe.

FullHand extends a computer vision method to track hand articulation, and especially finger articulation, for interactive applications. It follows a hybrid approach and uses a multicamera setup to track the skeletal motion of 26 degrees of freedom with a low latency. Whereas previous trackers have shown point designs without critical evaluation, we subjected the method to both technical and empirical assessments. Results from a motion elicitation study suggest that combining finger articulation with global hand motion is natural to users. The hand tracking algorithm had an error of < 15 mm in 87% of the datasets that we collected. A broad range of interactive techniques were designed to further explore this capability. Our examples range from menu selection that uses multiple finger motion of two hands to first-person shooter where 3 fingers and global hand motion are simultaneously used for playing.

We developed one of the interaction techniques further to be used in a real application, a 3D virtual globe. Results from a controlled user study show that although interaction was difficult at first, users' performance in three out of four tasks rapidly developed to a level comparable with the mouse. Although the study has a limited sample size, it demonstrates that the capability of the tracker can be actually used for free-hand interactions. To our knowledge, it is the first controlled study of interactive applications of markerless hand articulation tracking that report objective measures of user performance.

Previous markerless free-hand interaction technologies imposed constraints on designers regarding the type of interactions that they could create due to technical limitations. Since we track a kinematic skeleton new interaction techniques can quickly and efficiently be detected and used for interaction.

We regard these results favorable to the idea of using the hybrid tracking approach for HCI. In the future, we will improve the capability by allowing the use of fewer cameras. Presently, our hand model creation process is semi-automatic, and we plan to improve this by adopting automatic methods. We also require users to wear a black sock for image segmentation purposes. We contribute to the research community by releasing the tracker and annotated datasets.

REFERENCES

1. Athitsos, V., and Sclaroff, S. Estimating 3D hand pose from a cluttered image. In *2003 IEEE Computer Society Conference on Computer Vision and Pattern*

Recognition, 2003. Proceedings, vol. 2 (June 2003), II – 432–9 vol.2.

2. Bailly, G., Müller, J., Rohs, M., Wigdor, D., and Kratz, S. Shoesense: a new perspective on gestural interaction and wearable applications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, ACM (New York, NY, USA, 2012), 1239–1248.

3. Bailly, G., Walter, R., Mller, J., Ning, T., and Lecolinet, E. Comparing free hand menu techniques for distant displays using linear, marking and finger-count menus. In *Human-Computer Interaction INTERACT 2011*, P. Campos, N. Graham, J. Jorge, N. Nunes, P. Palanque, and M. Winckler, Eds., no. 6947 in Lecture Notes in Computer Science. Springer Berlin Heidelberg, Jan. 2011, 248–262.

4. Ballan, L., Taneja, A., Gall, J., Van Gool, L., and Pollefeys, M. Motion capture of hands in action using discriminative salient points. In *Computer Vision ECCV 2012*, A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid, Eds., vol. 7577 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg, 2012, 640–653.

5. Baudel, T., and Beaudouin-Lafon, M. Charade: remote control of objects using free-hand gestures. *Commun. ACM 36*, 7 (July 1993), 28–35.

6. Bolt, R. A. "put-that-there": Voice and gesture at the graphics interface. In *Proceedings of the 7th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '80, ACM (New York, NY, USA, 1980), 262–270.

7. Bowman, D. A., Kruijff, E., LaViola, J. J., and Poupyrev, I. An introduction to 3-d user interface design. *Presence: Teleoperators and Virtual Environments 10*, 1 (Feb. 2001), 96–108.

8. CARTER, J. B., and BANISTER, E. W. Musculoskeletal problems in VDT work: a review. 1623–1648. PMID: 7957019.

9. Erol, A., Bebis, G., Nicolescu, M., Boyle, R. D., and Twombly, X. Vision-based hand pose estimation: A review. *Computer Vision and Image Understanding 108*, 12 (Oct. 2007), 52–73.

10. Fukuchi, K., Sato, T., Mamiya, H., and Koike, H. Pac-pac: pinching gesture recognition for tabletop entertainment system. In *Proceedings of the International Conference on Advanced Visual Interfaces*, AVI '10, ACM (New York, NY, USA, 2010), 267273.

11. Hilliges, O., Izadi, S., Wilson, A. D., Hodges, S., Garcia-Mendoza, A., and Butz, A. Interactions in the air: adding further depth to interactive tabletops. In *Proceedings of the 22nd annual ACM symposium on User interface software and technology*, UIST '09, ACM (New York, NY, USA, 2009), 139148.

12. Jones, L. A., and Lederman, S. J. *Human Hand Function*, 1 ed. Oxford University Press, USA, Apr. 2006.

13. Karam, M. *PhD Thesis: A framework for research and design of gesture-based human-computer interactions*. phd, University of Southampton, Oct. 2006.

14. Keskin, C., Kirac, F., Kara, Y., and Akarun, L. Real time hand pose estimation using depth sensors. In *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)* (Nov. 2011), 1228 –1234.

15. Lee, J., Olwal, A., Ishii, H., and Boulanger, C. SpaceTop: integrating 2D and spatial 3D interactions in a see-through desktop environment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, ACM (New York, NY, USA, 2013), 189192.

16. MacCormick, J., and Isard, M. Partitioned sampling, articulated objects, and interface-quality hand tracking. In *Computer Vision ECCV 2000*, D. Vernon, Ed., no. 1843 in Lecture Notes in Computer Science. Springer Berlin Heidelberg, Jan. 2000, 3–19.

17. Melax, S., Keselman, L., and Orsten, S. Dynamics based 3D skeletal hand tracking. In *Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*, I3D '13, ACM (New York, NY, USA, 2013), 184184.

18. Ni, T. *A Framework of Freehand Gesture Interaction: Techniques, Guidelines, and Applications*. PhD thesis, VirginiaTech, Sept. 2011.

19. Nierop, O. A. v., Helm, A. v. d., Overbeeke, K. J., and Djajadiningrat, T. J. P. A natural human hand model. *Visual Comput 24*, 1 (Jan. 2008), 31–44.

20. Oakley, I., Sunwoo, J., and Cho, I.-Y. Pointing with fingers, hands and arms for wearable computing. In *CHI '08 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '08, ACM (New York, NY, USA, 2008), 32553260.

21. Oikonomidis, I., Kyriazis, N., and Argyros, A. Efficient model-based 3D tracking of hand articulations using kinect. British Machine Vision Association (2011), 101.1–101.11.

22. Oikonomidis, I., Kyriazis, N., and Argyros, A. Full DOF tracking of a hand interacting with an object by modeling occlusions and physical constraints. In *2011 IEEE International Conference on Computer Vision (ICCV)* (2011), 2088–2095.

23. Simo Serra, E. *Kinematic Model of the Hand using Computer Vision*. PhD thesis, Institut de Robtica i Informtica Industrial, 2011.

24. Sridhar, S., Oulasvirta, A., and Theobalt, C. Interactive markerless articulated hand motion tracking using RGB and depth data. In *2013 IEEE International Conference on Computer Vision (ICCV), (to appear)* (Dec. 2013).

25. Stenger, B., Thayananthan, A., Torr, P. H. S., and Cipolla, R. Model-based hand tracking using a hierarchical bayesian filter. 1372–1384.

26. Stoll, C., Hasler, N., Gall, J., Seidel, H., and Theobalt, C. Fast articulated motion tracking using a sums of gaussians body model. In *2011 IEEE International Conference on Computer Vision (ICCV)* (Nov. 2011), 951 –958.

27. Sturman, D., and Zeltzer, D. A survey of glove-based input. 30–39.

28. Wang, R., Paris, S., and Popović, J. 6D hands: markerless hand-tracking for computer aided design. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, UIST ’11, ACM (New York, NY, USA, 2011), 549558.

29. Wang, R. Y., and Popović, J. Real-time hand-tracking with a color glove. 63:163:8.

30. Wilson, A. D. Robust computer vision-based detection of pinching for one and two-handed gesture input. In *Proceedings of the 19th annual ACM symposium on User interface software and technology*, UIST ’06, ACM (New York, NY, USA, 2006), 255258.

31. Zimmerman, T. G., Lanier, J., Blanchard, C., Bryson, S., and Harvill, Y. A hand gesture interface device. 189192.

Below you find a list of the most recent research reports of the Max-Planck-Institut für Informatik. Most of them are accessible via WWW using the URL `http://www.mpi-inf.mpg.de/reports`. Paper copies (which are not necessarily free of charge) can be ordered either by regular mail or by e-mail at the address below.

Max-Planck-Institut für Informatik
– Library and Publications –
Campus E 1 4

D-66123 Saarbrücken

E-mail: `library@mpi-inf.mpg.de`

MPI-I-2016-5-002	A. Mishra	?
MPI-I-2016-5-001	D. Gupta, K. Berberich	Diversifying Search Results Using Time
MPI-I-2015-4-002	J. Tompkin	Interactive Ranking to Define Criteria on High-dimensional Visual Data
MPI-I-2014-5-002	A. Anand, I. Mele, S. Bedathur, K. Berberich	Phrase Query Optimization on Inverted Indexes
MPI-I-2014-5-001	M. Dylla, M. Theobald	Learning Tuple Probabilities in Probabilistic Databases
MPI-I-2014-4-002	S. Sridhar, A. Oulasvirta, C. Theobalt	Fast Tracking of Hand and Finger Articulations Using a Single Depth Camera
MPI-I-2014-4-001	K.I. Kim, J. Tompkin, C. Theobalt	Local high-order regularization on data manifolds
MPI-I-2013-RG1-002	P. Baumgartner, U. Waldmann	Hierarchic superposition with weak abstraction
MPI-I-2013-5-002	F. Makari, R. Gemulla, R. Khandekar, J. Mestre, M. Sozio	A distributed algorithm for large-scale generalized matching
MPI-I-2013-4-001	M. Granados, K.I. Kim, C. Theobalt	HDR reconstruction of cluttered dynamic scenes
MPI-I-2013-1-001	S. Ott	New results for non-preemptive speed scaling
MPI-I-2012-RG1-002	A. Fietzke, E. Kruglov, C. Weidenbach	Automatic generation of inductive invariants by SUP(LA)
MPI-I-2012-RG1-001	M. Suda, C. Weidenbach	Labelled superposition for PLTL
MPI-I-2012-5-004	F. Alvanaki, S. Michel, A. Stupar	Building and maintaining halls of fame over a database
MPI-I-2012-5-003	K. Berberich, S. Bedathur	Computing n-gram statistics in MapReduce
MPI-I-2012-5-002	M. Dylla, I. Miliaraki, M. Theobald	Top-k query processing in probabilistic databases with non-materialized views
MPI-I-2012-5-001	P. Miettinen, J. Vreeken	MDL4BMF: Minimum Description Length for Boolean Matrix Factorization
MPI-I-2012-4-001	J. Kerber, M. Bokeloh, M. Wand, H. Seidel	Symmetry detection in large scale city scans
MPI-I-2011-RG1-002	T. Lu, S. Merz, C. Weidenbach	Towards verification of the pastry protocol using TLA+
MPI-I-2011-5-002	B. Taneva, M. Kacimi, G. Weikum	Finding images of rare and ambiguous entities
MPI-I-2011-5-001	A. Anand, S. Bedathur, K. Berberich, R. Schenkel	Temporal index sharding for space-time efficiency in archive search
MPI-I-2011-4-005	A. Berner, O. Burghard, M. Wand, N.J. Mitra, R. Klein, H. Seidel	A morphable part model for shape manipulation
MPI-I-2011-4-003	J. Tompkin, K.I. Kim, J. Kautz, C. Theobalt	Videoscapes: exploring unstructured video collections
MPI-I-2011-4-002	K.I. Kim, Y. Kwon, J.H. Kim, C. Theobalt	Efficient learning-based image enhancement : application to compression artifact removal and super-resolution
MPI-I-2011-4-001	M. Granados, J. Tompkin, K. In Kim, O. Grau, J. Kautz, C. Theobalt	How not to be seen inpainting dynamic objects in crowded scenes
MPI-I-2010-RG1-001	M. Suda, C. Weidenbach, P. Wischnewski	On the saturation of YAGO
MPI-I-2010-5-008	S. Elbassuoni, M. Ramanath, G. Weikum	Query relaxation for entity-relationship search

MPI-I-2010-5-007	J. Hoffart, F.M. Suchanek, K. Berberich, G. Weikum	YAGO2: a spatially and temporally enhanced knowledge base from Wikipedia
MPI-I-2010-5-006	A. Broschart, R. Schenkel	Real-time text queries with tunable term pair indexes
MPI-I-2010-5-005	S. Seufert, S. Bedathur, J. Mestre, G. Weikum	Bonsai: Growing Interesting Small Trees
MPI-I-2010-5-004	N. Preda, F. Suchanek, W. Yuan, G. Weikum	Query evaluation with asymmetric web services
MPI-I-2010-5-003	A. Anand, S. Bedathur, K. Berberich, R. Schenkel	Efficient temporal keyword queries over versioned text
MPI-I-2010-5-002	M. Theobald, M. Sozio, F. Suchanek, N. Nakashole	URDF: Efficient Reasoning in Uncertain RDF Knowledge Bases with Soft and Hard Rules
MPI-I-2010-5-001	K. Berberich, S. Bedathur, O. Alonso, G. Weikum	A language modeling approach for temporal information needs
MPI-I-2010-1-001	C. Huang, T. Kavitha	Maximum cardinality popular matchings in strict two-sided preference lists
MPI-I-2009-RG1-005	M. Horbach, C. Weidenbach	Superposition for fixed domains
MPI-I-2009-RG1-004	M. Horbach, C. Weidenbach	Decidability results for saturation-based model building
MPI-I-2009-RG1-002	P. Wischnewski, C. Weidenbach	Contextual rewriting
MPI-I-2009-RG1-001	M. Horbach, C. Weidenbach	Deciding the inductive validity of $\forall\exists^*$ queries
MPI-I-2009-5-007	G. Kasneci, G. Weikum, S. Elbassuoni	MING: Mining Informative Entity-Relationship Subgraphs
MPI-I-2009-5-006	S. Bedathur, K. Berberich, J. Dittrich, N. Mamoulis, G. Weikum	Scalable phrase mining for ad-hoc text analytics
MPI-I-2009-5-005	G. de Melo, G. Weikum	Towards a Universal Wordnet by learning from combined evidenc
MPI-I-2009-5-004	N. Preda, F.M. Suchanek, G. Kasneci, T. Neumann, G. Weikum	Coupling knowledge bases and web services for active knowledge
MPI-I-2009-5-003	T. Neumann, G. Weikum	The RDF-3X engine for scalable management of RDF data
MPI-I-2009-5-003	T. Neumann, G. Weikum	The RDF-3X engine for scalable management of RDF data
MPI-I-2009-5-002	M. Ramanath, K.S. Kumar, G. Ifrim	Generating concise and readable summaries of XML documents
MPI-I-2009-4-006	C. Stoll	Optical reconstruction of detailed animatable human body models
MPI-I-2009-4-005	A. Berner, M. Bokeloh, M. Wand, A. Schilling, H. Seidel	Generalized intrinsic symmetry detection
MPI-I-2009-4-004	V. Havran, J. Zajac, J. Drahokoupil, H. Seidel	MPI Informatics building model as data for your research
MPI-I-2009-4-003	M. Fuchs, T. Chen, O. Wang, R. Raskar, H.P.A. Lensch, H. Seidel	A shaped temporal filter camera
MPI-I-2009-4-002	A. Tevs, M. Wand, I. Ihrke, H. Seidel	A Bayesian approach to manifold topology reconstruction
MPI-I-2009-4-001	M.B. Hullin, B. Ajdin, J. Hanika, H. Seidel, J. Kautz, H.P.A. Lensch	Acquisition and analysis of bispectral bidirectional reflectance distribution functions
MPI-I-2008-RG1-001	A. Fietzke, C. Weidenbach	Labelled splitting
MPI-I-2008-5-004	F. Suchanek, M. Sozio, G. Weikum	SOFIE: a self-organizing framework for information extraction
MPI-I-2008-5-003	G. de Melo, F.M. Suchanek, A. Pease	Integrating Yago into the suggested upper merged ontology