Happy Measure: Augmented Reality for Mobile Virtual Furnishing

Rahul Swaminathan[‡], Robert Schleicher[‡†],
Simon Burkard[†], Renato Agurto[†], Steven Koleczko[†]
[‡]Telekom Innovation Laboratories and [†]Technische Universität Berlin
Ernst-Reuter-Platz 7, 10587 Berlin, Germany

ABSTRACT

We present a novel vision based augmented reality system called *Happy Measure* to facilitate the measurement, 3D modeling, and visualization of furniture and other objects using a smartphone or mobile device equipped with a camera. We also study the concomitant interaction metaphors that enable interactive 3D model capture and manipulation in augmented environments. The proposed system allows for interactive measurement of an object's size and the creation of primitive based 3D models from a single photograph. The appearance of the furniture (color textured model) is also captured easily by the system using the underlying (or multiple) images taken by the user. The system thus allows the user to capture textured 3D models of furniture or other objects and place and manipulate them virtually elsewhere for visualization purposes.

The methods to compute textured 3D models and measure real world objects from a single image are strongly related to the interactions they permit. We compare two interaction metaphors used to capture 3D textured models of object to ensure easy interaction while still obtaining accurate measurements in a user test. Results suggest that one is superior in terms of measurement accuracy and also subjective user experience as it allows for continuous touch interaction on the whole screen. Virtually placing a modeled object in another location is another aspect of the presented system and we explore a novel interaction paradigm to perform this task along with initial user tests.

Keywords: Augmented reality, image based rendering, camera calibration, visual metrology, mobile interaction

INTRODUCTION

Increasing power of mobile devices and smartphones has significantly contributed to the progress of augmented reality (AR) applications becoming mobile. The first ever mobile AR system, called *The Columbia Touring Machine* (Feiner, MacIntyre, Hollerer, & Webster, 1997), had to use a wearable laptop together with GPS and Orientation sensors housed in a backpack. In comparison, current day mobile and hand-held devices such as smartphones and tablet PCs contain similar sensors and computing power and are capable of much the same, if not more.

Mobile Augmented Reality

In general, mobile AR applications can be classified as being applicable globally at any location in the world, or being local to the current position of the user within a local coordinate frame. Globally relevant (applicable) AR applications usually depend on GPS based location and

orientation together with a known 3D model of the world around them to determine what the user is looking at. Many mobile AR systems ignore the three dimensional nature of the world around the user while providing information at the horizon (see e.g. (―Layar,‖ n.d.)). In such applications, there is no precise registration of the user's view and the surrounding 3D world. As a consequence it is impossible to augment the world with objects that are scaled in proportion to the user's position and context in their surroundings.

On the other hand most *local* AR applications make use of *2D* markers (M. Fiala, 2005; Hirokazu Kato & Billinghurst, 1999; Wagner, Reitmayr, Mulloni, Drummond, & Schmalstieg, 2008) to visually localize the user in a local coordinate frame rather than GPS. The marker thus provides the required registration of the captured image (view) and the surrounding 3D world. As a consequence, a much better localization (pose) of the viewer is obtained together with potentially the exact scale and size of objects in the world. Recent improvements in natural feature detection, tracking and recognition have also lead to natural features being used instead of the traditional marker to estimate pose and motion of a mobile device (see (Bay, Ess, Tuytelaars, & Van Gool, 2008; Kurz & Ben Himane, 2011; Wagner et al., 2008)).

The need for determining location of the user/viewer is usually a prerequisite to be able to correctly augment the viewed world with additional virtual information or content. There is a wide range of domains to which augmented reality is effectively used including entertainment, education/training, interaction with virtual objects (see (Chang, Koh, & Been-Lirn Duh, 2011a; Olsson & Salo, 2011; de Sa, Churchill, & Isbister, 2011). In all cases, the content needs to correctly be augmented onto the video captured by the camera and displayed to the user coherent with the perspective of the user. For this purpose, using a marker is typically a standard robust approach to determine the correct perspective of the user.

However, AR goes beyond simply augmenting the visual experience of the real world with virtual content. Many AR applications require interaction of the user with the virtual content (Chang et al., 2011a; Chang, Koh, & Been-Lirn Duh, 2011b; Gjosaeter, 2009; H. Kato, Billinghurst, Poupyrev, Imamoto, & Tachibana, n.d.; Olsson & Salo, 2011; Papagiannakis, Singh, & Magnenat-thalmann, 2008). In this context, the metaphors used for interaction, the enabling technologies and display techniques, all play a key role in overall user experience. The variation of interaction in the mobile AR context is large, and beyond the scope of this paper (see (Kolsch, Bane, Hollerer, & Turk, 2006; Papagiannakis et al., 2008) for an overview of interaction with AR systems and enabling technologies). We shall instead focus on AR with hand-held devices such as mobile phones and particularly those relevant to the application of modeling and manipulation of 3D objects using AR for virtual furnishing.

Interactive AR based Furnishing

We present *Happy Measure*, a novel system that allows the user to photograph furniture or other objects using their mobile phone camera and directly measure their sizes from the image alone. We are interested in the overall size of objects rather than detailed measurements of individual parts of the object. We therefore define the size of the object in terms of its bounding box as shown in Figure 1.(top right). While manually measuring furniture and objects would provide very accurate results, it does not provide the sense of space occupied by an object nor how it would look when placed there. This issue is further complicated if the object in question is heavy and cannot be moved around easily or worse still, not at hand.
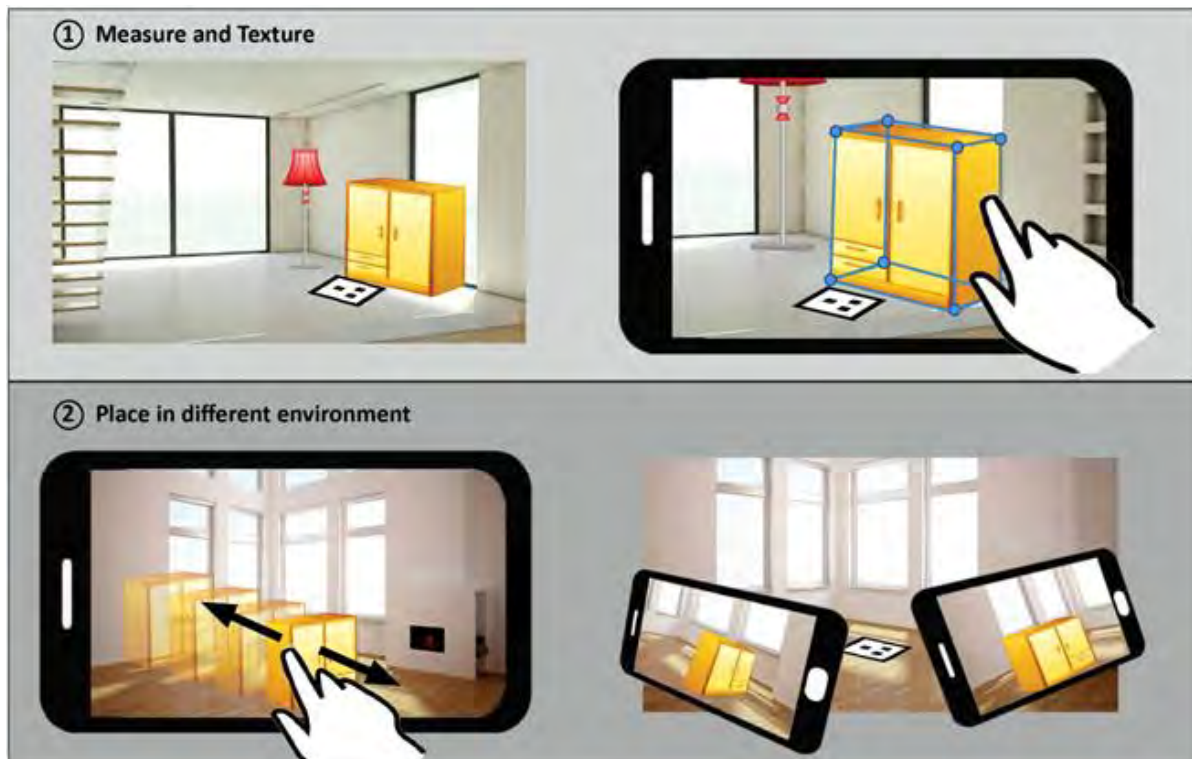
Figure 1: Happy Measure: the marker based AR mobile furnishing system allows users to interactively define a bounding box of an object and thereby determine its size. The system uses multiple images to compute a textured 3D model that can be manipulated and virtually placed in another room. Happy measure is available for android devices at www.happymeasure.de

The system we propose (see Figure 1) enables users to interactively create 3D models of real world objects such as furniture. Furthermore, once measured, the system allows the user to walk around the object, take more photographs and automatically compute image based textures for the 3D models. Finally, using a marker to determine the user's perspective, the furniture is rendered in real-time in a new position/environment and can be manipulated in 3D on the smartphone or tablet. While the idea of displaying intricate *authored* 3D models is not novel (see (Olsson & Salo, 2011)), none of the other AR systems allows for capturing nor the augmentation with user generated 3D content of real-world objects.

Interactions with Virtual 3D Objects

A key challenge is to provide an interface that allows the user to manipulate AR elements in a convenient way, given the constraints of a small screen and limited spatial accuracy of movement gestures via touch. Ideally, this manipulation should be achievable with on hand while the other hand is holding the device.

In general, there has been a lot of work in the field of interactive 3D AR using various technologies including specialized gloves and tracking and gesture recognition (Piekarski, 2006; Piekarski & Thomas, 2002). However, this requires the user to wear specialized head mounted see-through displays together with specialized gloves and is in general not suited for the average user with a smartphone. Yet another approach to model 3D objects used tactile sensors (Anabuki

& Ishii, 2007) in order to compute accurate 3D models of objects. Again, this approach although innovative and accurate, does not lend itself to easy and everyday use due to the complicated use of tactile sensors. Yet another system utilized numerous ARtags (M. Fiala, 2005), to control a 3D modeling software like 3D Studio Max in order to create piece-wise polygonal 3D model (P. Fiala & Adamo-Villani, 2005). While being very flexible, the system is tedious and requires the use of multiple 2D markers and fine control in order to create polygonal models.

Apart from these ―multi modal‖ utilizing gloves, markers or tactile sensors, there has been significant work of measuring and modeling from images alone. These methods, that employ visual metrology (A. Criminisi, I. Reid, 2000; Sturm & Maybank, 1999) require human input to provide certain constraints from the image. The typical constraints used in single or multiple view metrology include the interactive determination of scene planes, vanishing points, vanishing lines, amongst others in the image. Recent improvements in dense feature tracking and matching have also been employed for 3D reconstruction and rapid model acquisition of hand held objects (Pan, Reitmayr, & Drummond, 2009). This approach assumes the object to be Lambertian[1], and requires the user to ―scan in‖ the object using video while the system simultaneously computes and refines a 3D model. This is a challenge for large and typically non-Lambertian objects such as polished wooden or glass objects.

Happy Measure – Interactive Visual Metrology, Modeling and Manipulation

Our aim is to create simple, quick and intuitive interaction metaphors that assist user in both, creating basic yet satisfying 3D textured (realistically colored) models of furniture and other objects as well as manipulating them virtually for visualization purposes using a smartphone or tablet PC. We shall focus on the interaction metaphors used for the measurement or modeling process as well as the virtual model manipulation essentially. Details of how our system also automatically extracts textures from multiple images to compute an effective colored model of objects is detailed in Appendix 1 so as not to detract from the main focus of the paper.

We begin with two proposed interaction metaphors to determine the size and basic 3D model of the objects and describe the underlying mathematical models as well. We explore the theoretical limitations of both approaches as well as constraints on interaction. We then present results from a user study we performed comparing each of these two interaction metaphors. Details of the experimental process, the methods used, and their relation to the overall user experience are discussed.

Finally, we explore a novel interaction paradigm to manipulate virtual objects in a room using a mobile device enabled with augmented reality. We present details of this interaction as well as early results of initial user tests done using this approach.

We end the paper with discussing our observations and key results and explore possible avenues to further develop our system and possible directions for future research.

---

[1] Lambertian surfaces ideally have diffused reflection and their appearance and apparent brightness is the same regardless of the observer's angle of view. They do not exhibit reflection or highlights as with metallic surfaces.
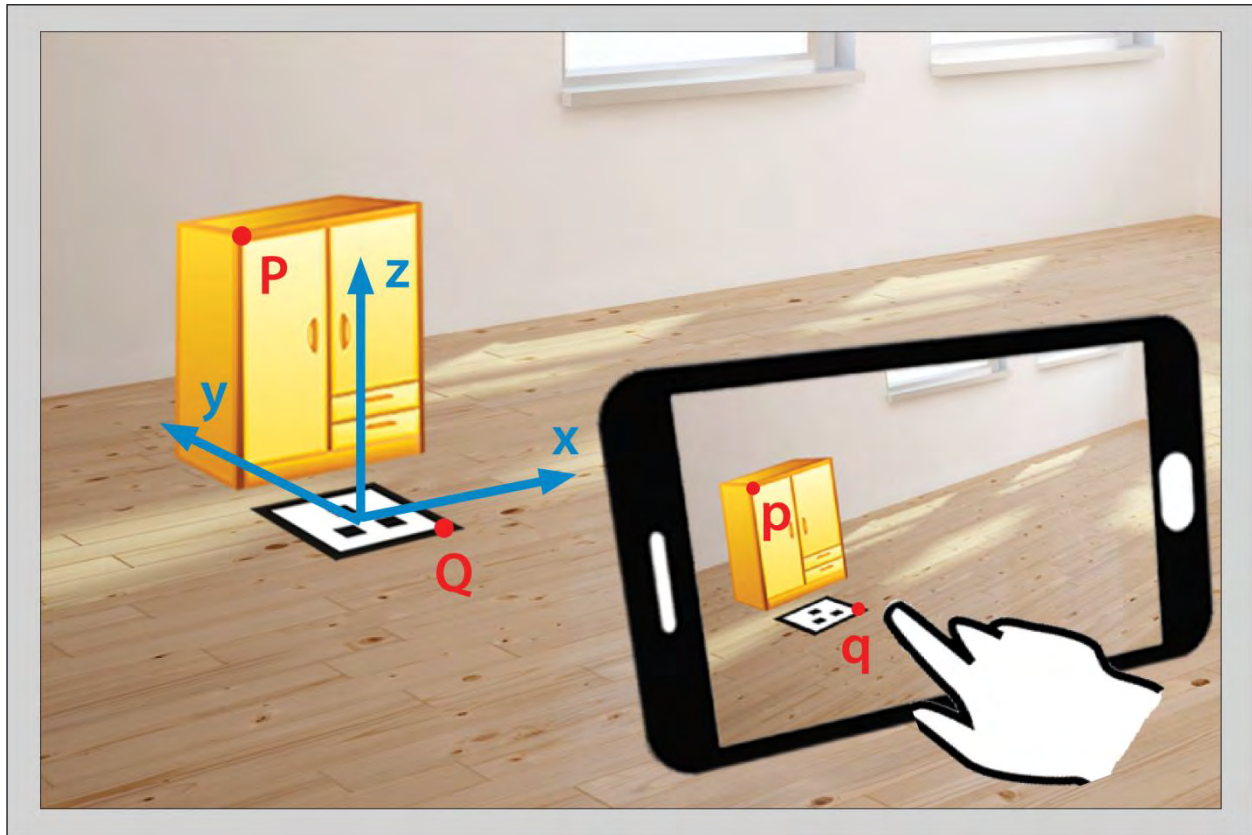
*Figure 2: The 2D marker on the floor defines the world coordinate frame as illustrated above. A point **q** in the image can then be directly related to its corresponding position **Q** on the ground in the plane of the marker by a homography .The marker provides the system all geometric information as regards the camera's pose (rotation and translation) wrt. the world coordinate frame via the acquired image.*

INTERACTIVE VISUAL METROLOGY

We now introduce the basic requirements for visual metrology from a single image. We describe the underlying geometric constraints and requirements as well as introduce the two main interaction paradigms proposed and studied in this paper.

Visual Metrology

In general, an image is simply a collection of pixels and does not explicitly provide any further information about the scene in terms of metric geometric relationships. There is no sense of scale, depth nor angels or field of view from an image alone. Thus, from a single image alone, it is impossible to know how far objects are, their relative sizes, nor their true shape. To do so, we require two key additional pieces of information:

- **Imaging model:** Under the perspective imaging model, which most cameras adhere to, the imaging model parameters are represented by the *3x3* camera matrix **K** which can be calibrated for automatically (see (Zhengyou Zhang, 1999)).
- **Reference object:** One must have a reference object in the scene of known size so as to make geometric inference of other objects in the world. This reference object is usually a 2D marker (M. Fiala, 2005; H. Kato & Billinghurst, 1999).

In particular, to make metric and geometric analysis of the image we must know the camera's pose (rotation R and translation t) with respect to the scene. Furthermore, this information is also critical in order to visually augment the user's view of the world from the correct perspective. This pose information in fact together with scale of objects in the scene can be derived from the image of the marker together with the object being modeled.

The corners of the marker in the real-world together with the corresponding image points in the image provide a relationship between points in the image (q) and points in the plane of the marker (Q) called the *Homography* given by the relation:

$$Q \cong H \cdot q \hspace{4cm} Eq.(\ 1)$$

Given any still image, the system automatically detects the marker and the corresponding corners of the marker to then determine the underlying perspective relationship (given by the homography (Hartley & Zisserman, 2003)) between the image plane and the marker plane. Next we utilize the known camera matrix K, to determine the camera's pose in terms of its rotation R and translation t (see (Yi Ma, Stefano Soatto, Jana Kosecká, 2004) Chapter 5 for details) with respect to the world coordinate frame defined by the marker as shown in Figure 2.

In general, once the pose (R | t) of the camera with respect to the marker has been determined, we can relate any other 3D point (P) to its image point (p) under perspective projection as:

$$p \cong K \cdot (R|t) \cdot P \hspace{4cm} Eq.(\ 2)$$

It should be pointed out that given the image point (p) it is still impossible to determine the corresponding world point (P) as depth information is still missing.

In order to measure objects in the scene from an image, we also need to know where exactly the object of interest lies within the image. Since automatic segmentation of the object of interest is relatively challenging, especially for the purpose of robust measurement, we require user input to define the extent of the furniture/object being measured.

We shall now present two key interaction paradigms developed specifically with touch displays of smartphones in mind. While such devices offer the possibility to touch any area on their screen, and thereby manipulate an AR object directly, such an interaction may sometimes appear too novel for users used to a classic "windows, icons, mouse & pointer" (WIMP) setup. The two interaction metaphors are:
- **Face-based Model Interaction**: A rather conservative approach, where the user first chooses the face of the model cube she wants to extend explicitly via a button, and then extends or shrinks the cube along that dimension by moving her finger along the ruler at the bottom of the screen as shown in Figure 3.
- **Corner-based Direct Interaction**: A rather innovative approach that exploits the touch screen displays of current mobile devices, namely adjusting the size of the model cube by snapping its corners to the object corners. It allows for direct selection inside the image, i.e. touching corners from the reference cube and moving them directly on the display as shown in Figure 6.

Figure 3: Model manipulation with face-based interaction: Here the user can choose the specific face of the bounding box to be moved. Then a simple swipe of the finger to the left or right on the ruler extrudes the box along the selected face. Since the actual size of the 3D model is internally maintained, the visual overlay provides feedback in the image as to its fit to the object. When satisfied, this directly provides the size and position of the object/furniture with respect to the marker.

Face-based Model Interaction

This paradigm requires neither geometric analysis nor computation from the device being used but rather relies on the user's perception of 3D space under perspective projection (in the image). The user is tasked with manipulating (extruding) the sides of a place-holder box augmented into the image, so as to make it visually circumscribe the object.

The user places the marker on (or next to) the object so that it is precisely aligned with the edges of the object. Usually aligning the marker with the edges is an effortless act. Furthermore, discarding the necessity to rotate the face-based model reduces the number of interaction steps needed and therefore reduces complexity as well as duration of the measuring task.

The user then takes a snapshot image which is processed in order to detect the marker and thereupon determine the camera pose. As shown in Figure 3, the system then displays a control-box of known size on the marker. The user is provided with buttons in a control-panel to select specific faces of the box to *extrude*, thereby moving and resizing it. The user can select one face at a time, and swipe left or right along the provided *ruler* to extrude each face of the box inwards (making the box smaller) or outwards (making it bigger). These operations allow the user to resize as well as translate the box in the plane of the marker.

Since the size and shape of the box is internally maintained (graphical model), rendering the box in the correct perspective is then trivial. The system provides real-time feedback to the user by augmenting the acquired image with the correct perspective view of the manipulated box as the user attempts to circumscribe the furniture or object being measured. When the box exactly fits the furniture, it directly determines the size of the furniture.

It should be noted that the interaction forces either the top or the bottom of the control box to lie in the plane of the marker. This ensures avoiding any scale-depth ambiguities i.e. ambiguities between a close by small object and a distant large object, in the image.

An example usage
We now illustrate the face-based interaction via buttons and ruler on an Android based smartphone. As shown in Figure 4, the user points the smartphone camera at the table with the marker placed below it, aligned with its front edge. The system then detects the marker and computes the pose of the camera (user's perspective) with respect to the marker on the floor, and accordingly augments the view with a control-box (in cyan).

On the right, the control-panel allows the user to select various faces of the box. The user first aligns the rear-face of the cube with the back of the table Figure 4(b) using the ruler. Next, starting with the right, left, front, and top, the user extrudes (out or in) as required the various faces to visually determine the bounding box in the real world. Note that, although the object is not truly rectangular, it can still be measured accurately. Figure 5 shows the object to be measured (LxHxW) as 138x73x59cm while the true dimensions are 140x72x60cm In most cases measurements tend to be within a couple of centimeters of their true sizes.

Constraints
As noted before, the model manipulation is to some extent restrictive, since the user relies on feedback in the perspective view. Rotation control was omitted in order to avoid it's complexities under perspective projection. As a consequence, the marker must be precisely aligned with the edge of objects. Failure to align the marker results in errors in measurement and subsequent steps including texture computation (described in Appendix 1).
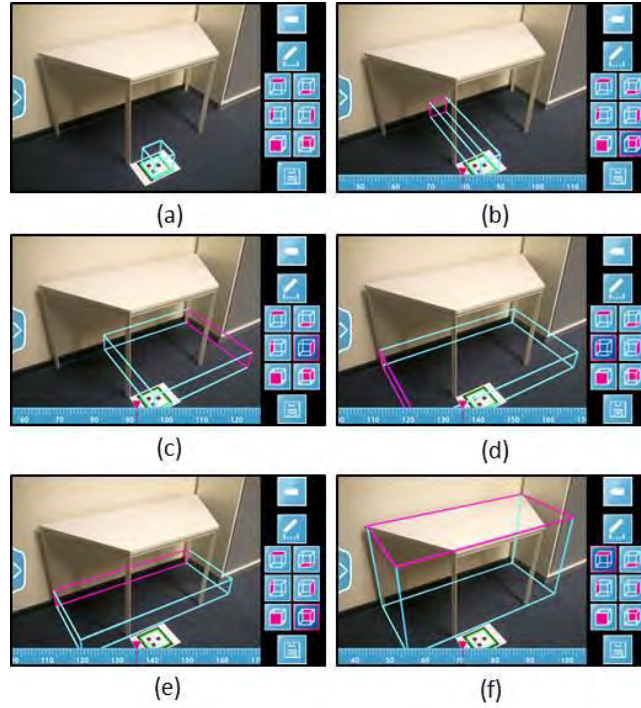
Figure 4: Android smartphone screenshots of model manipulation via face-based interface for visual metrology. The user selects faces (b-f): rear, right, left, front and top via buttons, in order to manipulate a virtual bounding box to fit its view in the image via moving a slider along the horizontal ruler at the bottom of the screen. The control-box provides the object size and position.
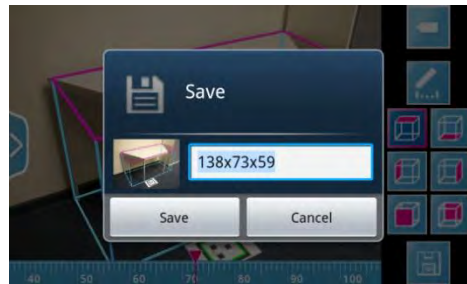

Figure 5: Result of measurement using Happy Measure together with the face-based interaction metaphor. Errors in measurement were within a few cm of their true measures.

Corner-based Direct Interaction

The face-based model manipulation approach in the previous section offered a well-established interaction scheme, but required considerable number of interaction steps to align the control-cube with the extent of the object. Typically, the user would need to iterate over the faces by adjusting and readjusting the faces (depending on the perspective) to eventually obtain the 3D size of the object.

However, since the size of an object is determined by purely its length, width and height, all we need are three corner points along the corresponding *diagonals* to compute its size. This

indicates that the user should ideally be able to provide all necessary data by simply indicating the three points along the three diagonals of the object, with three ―touch/clicks" on the screen.

Five point Interaction

We now explore a new and direct interaction metaphor requiring the user to only position corners on the object. As shall be demonstrated, this does not require the marker to be aligned with the furniture edges, unlike before, providing greater flexibility with more robust estimates.

As shown in Figure 6, the user again takes a snap-shot of the object with the marker. As before we estimate the pose of the camera with respect to the marker based world coordinate frame. These parameters are again used to augment the control-box virtually on the marker. Now, instead of adjusting various box faces, all the user must do is to drag-and-drop five corners of the box onto the corresponding corners of the object. Note that this correspondence is defined in the image and not in 3D. Furthermore, it should be noted that although we need three points minimally, this minimal set of points must define the three diagonals of the object. For example, three points in a plane (one face of the object) do not provide the necessary constraints to determine its 3D size. By requiring the user to provide us five points, we are assured of sufficient data needed to measure the object without demanding a deeper understanding of 3D modeling from the user, which we describe in detail in the following section.

Measurement by Reshaping

We now develop the mathematical model used to estimate the size of furniture. The corners indicated by the user provide a correspondence between the control-box corners and those of the object. We pose the problem of determining the size of the object as that of determining the parameters needed to reshape the control-box and move it in 3D so that the resulting (resized) corners of the cube align with those of the object as indicated by the user. The reshaping consists in two transformations: scaling and then rotation and translation.

Consider the control-box in Figure 7 (top left). In order to fit the furniture it must first be resized appropriately. The resizing is modeled by a diagonal matrix $S = \text{diag}(sx, sy, sz)$, where (sx, sy, sz) correspond to the length, width and height of the furniture. The next step is to appropriately rotate and translate the resized control-box (see Figure 7 (bottom)) so as to match the furniture's real 3D location. Only then will the reshaped box corners align with the points provided by the user.
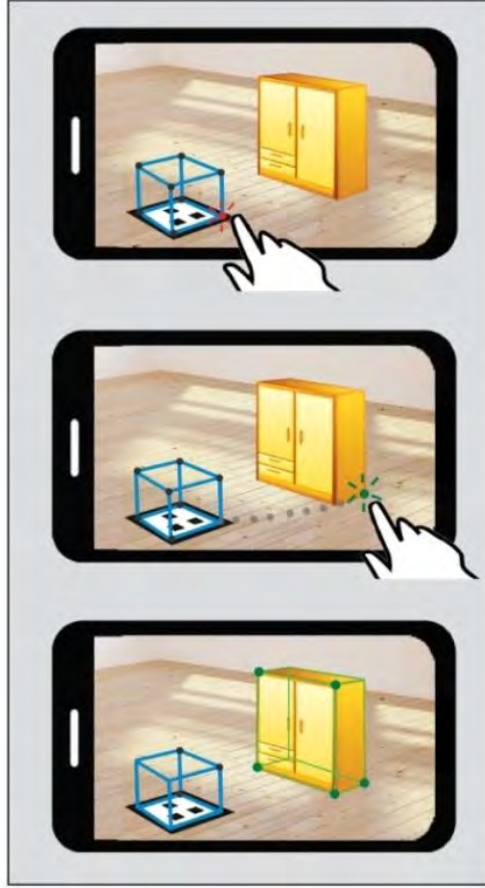
Figure 6: Corner-based direct interaction - Exploiting the fact that the whole screen is touch-sensitive, one can directly point out to our system the correspondences between the unit box shown on the marker to those of the furniture. The system then computes directly the size and pose of the object in the room relative to the marker.
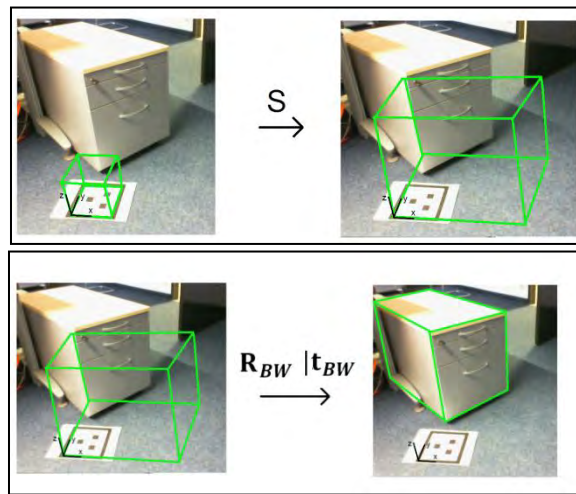

Figure 7: Modeling the bounding box in terms of a unit control-box and the shape transform M. The shape transform models the resizing matrix S that defines the size of the bounding box explicitly as well as the rotation and translation required to align it with the real world object.

Rotation is only in the ground plane (about the z axis) and can be defined by:

$$\mathbf{R}_{BW}(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

where θ is the rotation angle. The translation vector is:

$$\mathbf{t}_{BW} = (t_x \quad t_y \quad t_z)^T$$

It should be added that when the marker is on the floor, we can define translation only in the ground plane, thereby $t_z = 0$, a giving:

$$\mathbf{t}_{BW} = (t_x \quad t_y \quad 0)^T.$$

Using S, R_{BW} and t_{BW} we define the reshaping matrix M, that warps every point of the unit box to the world 3D coordinate of the furniture. The reshaped corners of the control-cube are then given by:

$$Q_W^i = \mathbf{M} \cdot Q_B^i , \text{ where } \mathbf{M} = [\mathbf{R}_{BW} \cdot \mathbf{S} \,|\, \mathbf{t}_{BW}] \text{ giving:}$$

$$\mathbf{M} = \begin{pmatrix} \cos\theta \cdot s_x & -\sin\theta \cdot s_y & 0 & t_x \\ \sin\theta \cdot s_x & \cos\theta \cdot s_y & 0 & t_y \\ 0 & 0 & s_z & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

In order to linearize the above constraint (note trigonometric elements in M), we can define M as a matrix with 7 degrees of freedom (*d.f.*) given by:

$$\mathbf{M} = \begin{pmatrix} m_{11} & m_{12} & 0 & m_{14} \\ m_{21} & m_{22} & 0 & m_{24} \\ 0 & 0 & m_{33} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \qquad\qquad Eq.(\,3)$$

Reshaped image constraints

Finally we need to define the relationship between the world 3D coordinates Q_W^i and the 2D coordinates q^i of the furniture in the image (provided by the user). Under perspective projection, see Eq.(2), the image of any world point Q_W^i is:

$$q^i \cong \mathbf{K} \cdot [\mathbf{R_C}|\mathbf{t_C}] \cdot Q_W^i$$

Where $\mathbf{K}$ is the intrinsic camera matrix, and $\mathbf{R}$ and $\mathbf{t}$ denote the pose. Combining these matrices, we define the projection matrix $\mathbf{D}$ as:

$$\mathbf{D} = \mathbf{K} \cdot [\mathbf{R_C}|\mathbf{t_C}]$$

Now we define the relation between any 3D point Q_B^i of the unit control-box to a 2D point q^i in the image after reshaping using Eq.(3) as:

$$q^i \cong \mathbf{D} \cdot \mathbf{M} \cdot Q_B^i$$

where $m_{11}, m_{12}, m_{21}, m_{22}, m_{33}, m_{14}$ and m_{24} are unknown.

Every pair of points $q^i = (x^i, y^i, 1)$ and $Q_B^i = (X_B^i, Y_B^i, Z_B^i, 1)$ provides two constraints. Thus, four corresponding corners would suffice to solve linearly for the reshaping parameters, thereby determining the size of the furniture. Our use of five user provided points provides an over determined system of equations resulting in a robust estimation of size.

Insofar, we described two interaction paradigms that rely on different theoretical underpinnings to determine the object's size. Theoretically, the *corner-based direct interaction* allows more flexibility and is more tolerant to errors in marker placement. We therefore expect it to be more robust in measurement. However, no statement can be made on the quality of the user input in relation to each method up to this point. This will be explored in the following section.

USER EXPERIMENTS

The previous section described two interactions and their mathematical background. We now present results from our user study experiment exploring their operability as well as corresponding user experience. We chose a current smartphone model, the Samsung Galaxy S3 running Android 4.0.4 where we logged various performance data. Additionally, participants had to specify their mental effort (Eilers, Nachreiner, & Hänecke, 1986; Sauro & Dumas, 2009) as well as the general user experience (Hassenzahl & Monk, 2010) of both interfaces.

Experimental Setup

The experiment took place in one of our laboratory rooms under constant lightening conditions. Six colleagues from our lab were recruited as participants (3 female, 3 male, average age 31.5 years, ranging from 27-36 years) for usability evaluation. None of them received compensation for participation. Their task was to measure a suitcase (51.5 x 37 x 19 cm) photographed from different perspectives, whereby the picture was taken by the experimenter and subjects only had to determine the size of the objects using the corresponding interfaces.

Procedure

After welcoming the test subjects, they were introduced to the general task and the two interface variants by the experimenter. This blended into a training phase where the users could get accustomed with both interface types for measurement (face-based vs. corner-based measurement) on their own. Once they felt comfortable with the software, they were seated, and the experiment began.

The experimenter took a photo of the suitcase and handed the mobile phone to the participant, who pressed a button to start measuring. The button press also served to start logging on the phone. The subjects were instructed to measure until they were satisfied with the result and then save the measurement. Pressing save ended one trial. This procedure was repeated five times, whereby the perspective of the picture and the position of the marker sheet varied, while the measurement interface type stayed the same. Once this block was completed, subjects filled in questionnaires to assess the previous interaction condition, and could also give their suggestions in a free-text field. Then the second part started, where they had to complete the same five trials using the other interaction metaphor (or condition). Sequence of conditions varied across subjects. After filling in the questionnaires for the second interaction block, the experiment was finished and participants were dismissed. Altogether one session lasted around 30 minutes. For one subject (Nr. 6) only four trials per condition were accomplished due to battery issues.

Recorded Data

The following parameters were logged on the phone during trials:

- **Trial duration**: defined as the time in seconds from the subject starting to measure until saving the model.
- **Number of interactions**: defined as the number of touch events in the corner-based measurement mode. For the face-based measurement mode, pressing a button to select a face together with sliding the ruler was counted as one interaction instead of two to avoid bias by design.
- **Touch movement time**: defined as the time from touch start to touch release for an interaction on the main screen. For the corner-based method this corresponds to time moving the model cube corners, for the face-based method this corresponds to the time manipulating the ruler as we ignored button presses when calculating this variable.
- **Relative measurement error**: is defined as the average deviation (in percentage) in size of the object as measured by the subject, from ground truth measurement across all three dimensions of the suitcase in percent.

Subsequent to each block, the following self-assessments were collected:

- **Perceived mental effort**: the Subjective Mental Effort Questionnaire [SMEQ, SEA in German, (Eilers et al., 1986)] is a ruler-like vertical scale that varies in its German version from 0 to 220 with additional text labels varying from 'hardly strenuous' (=20) to 'exceedingly strenuous' (=205). A discussion of the English version can be found in (Sauro & Dumas, 2009).
- **Joy of use**: To get a simple overall assessment of the overall joy of use for each interface, we used the smiley scale developed by (Jäger, 2004). It shows five smileys ranging from a frowning face (coded as -2) to a smiling face (codes as +2) with a neutral face in between (coded as 0).
- **User experience:** The AttrakDiff questionnaire is a semantic differential that measures various aspects of product quality by presenting pairs of opposite adjectives the user has to describe a product or service on (Hassenzahl, M. Burmester & Koller, 2003). We used the AttrakDiff mini that contains 10 items (Hassenzahl & Monk, 2010). The ratings vary from 1 to 7 where higher values indicate better evaluation for the corresponding dimension. The underlying dimensions are:
  Pragmatic quality (PQ): to what extent is the product assessed as useful to fulfill the given task. This dimension is closest related to classic usability criteria like effectiveness and efficiency.
  Hedonic Quality – Identity (HQ-I): does the product help me to communicate values relevant for my self-esteem?
  Hedonic Quality – Stimulation (HQ-S): Is the product perceived as original or creative and challenging in a positive way?
  Attractiveness°(ATT): How attractive is the product rated overall?

Results

We first present the analyses of the log data during interaction, then the self-assessment subsequent to the interactions, and finally relate these two.

Log-Data analysis

As the number of participants (n=6) and trials per condition (t=5) is manageable, we will always present one plot with values averaged across subjects for each trial (left side of the figures), as well as one plot that shows averages per user across the trials she/he accomplished (right side of the figures) for each parameter to provide the reader a complete insight into our results. Statistical analysis for comparing mean differences, e.g. regarding the two interface types, was done using the SPSS$^{®}$ function Mixed Linear Models (MLM). The results are comparable to a classic analysis of variance (ANOVA), but MLM can also handle subjects with missing trials (Quené & van den Bergh, 2004), i.e. subject 6 that only accomplished 4 trials per condition. In addition, correlated error terms in repeated measurements within one subject are accounted for (SPSS, 2005). As a consequence of this, the degrees of freedom reported for the *F*-values may vary (SPSS, 2005).
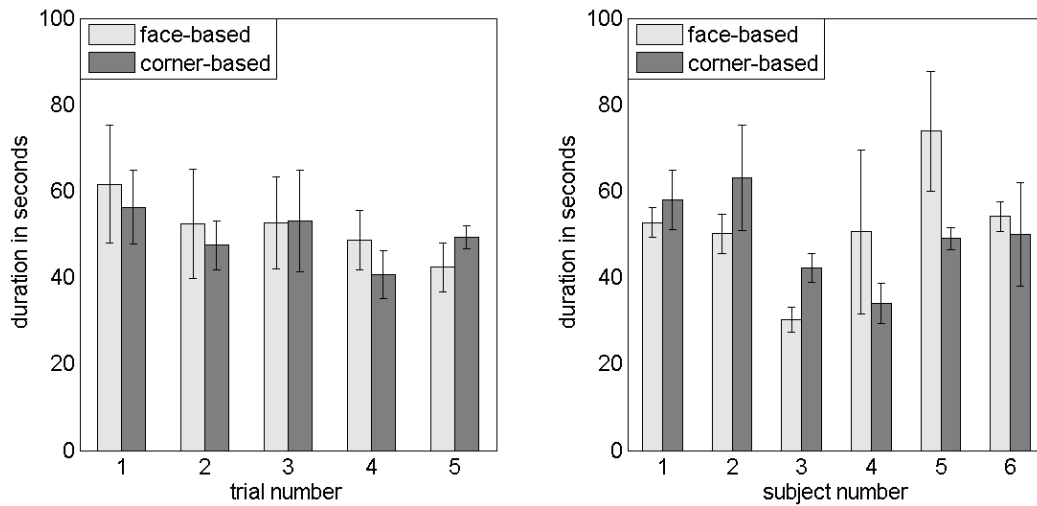

Figure 8: Duration for each interface type (face-based vs. corner-based interaction) averaged over trial number (left) and subject number (right). Whiskers denote standard errors.

Figure 8 shows the average duration of all subjects for each trial (left) as well as the average duration of each subject across all trials (right) for both interface types. As it can be seen, there is no unequivocal trend for one interaction (face-based vs. corner-based measurement) to be faster across all trials or all subjects. Respectively, there is no statistically significant effect of interface type on trial duration ($F_{1,16}$=0.209, p=0.654). Although trial duration tends to decrease across trials, the effect on trial number is also not significant ($F_{4,33}$=0.894, p=0.479), and neither a interaction of interface type*trial number, which would mean that trial duration significantly changed for only one interface type ($F_{4,32}$=0.151, p=0.961). The same, no significant main effect of interface type, holds for the logged number of interactions ($F_{1,16}$=4.142, p=0.059, see Figure 9) and movement time ($F_{1,17}$=3.083, p=0.097, see Figure 10), whereby number of interactions appears to be slightly lower for the corner-based method (Figure 9), and the touch movement time slightly higher as compared to the face-based interaction mode (Figure 10).
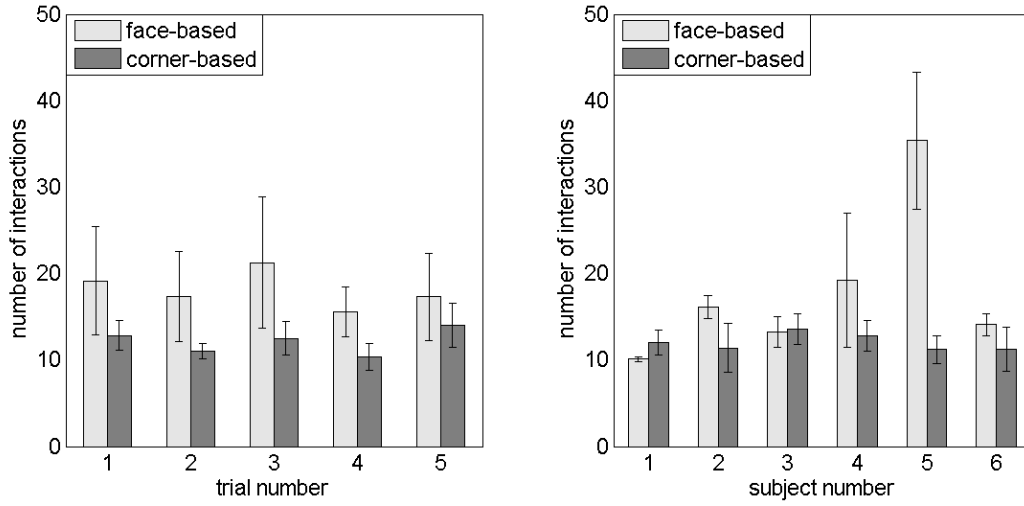
Figure 9: Number of interactions for each interface type (face-based vs. corner-based interaction) averaged over trial number (left) and subject number (right). Whiskers denote standard errors.
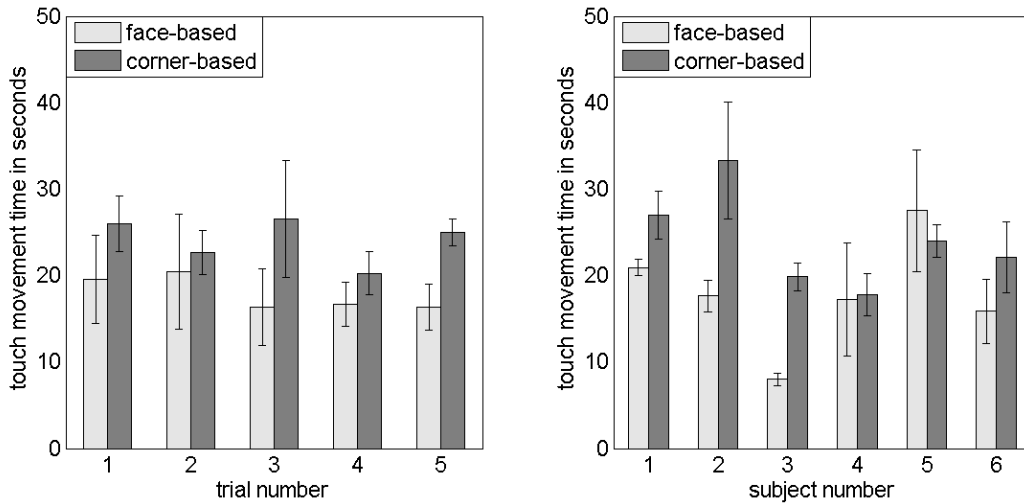

Figure 10: Touch movement time for each interface type (face-based vs. corner-based interaction) averaged over trial number (left) and subject number (right). Whiskers denote standard errors.

Looking at the relative measurement error (expressed in percent of object size), the face-based measurement interface apparently leads to higher errors as compared to the corner-based measurement interface (around 10% for face-based vs. around 5% across all trials in Figure 11). This effect is statistically significant ($F_{1,14}$=15.496, p=0.002) and also persists if subject 4, the one with the highest error (see Figure 11 right), is taken out of the sample ($F_{1,10}$=9.690, p=0.011). Again, there is no effect of *trial number* ($F_{4,34}$=1.634, p=0.188), and also no interaction of *interface type*trial number* ($F_{4,33}$=0.690, p=0.604) on measurement error.
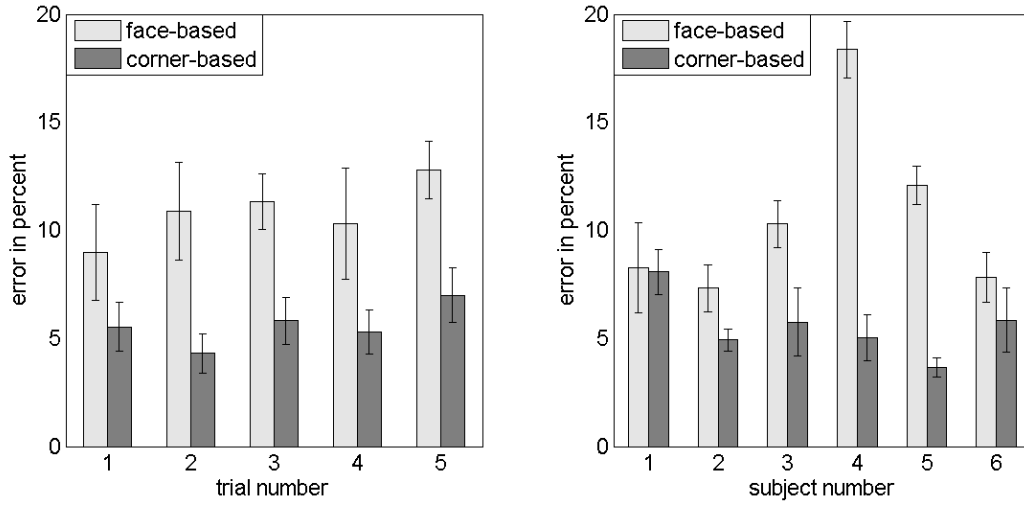
Figure 11: Measurement error for each interface type (face-based vs. corner-based interaction) averaged over trial number (left) and subject number (right). Whiskers denote standard errors.

To summarize, while both interfaces do not vary significantly with regard to variables that describe the process of measurement (duration, number of interactions, touch movement time), the outcome appears to be affected as the measurement error is lower for the corner-based measurement method. This variable is also only moderately correlated to number of interactions (r=.337, p=0.001), but neither to duration (r=.100, p=0.457) nor movement time (r=-.119, p=0.357).

Questionnaire data analysis
Table 1 shows the mean ratings for both interfaces regarding perceived mental effort (SEA scale), joy-of-use (Smiley scale), and assessment according to German school grades that vary from 1 as 'very good' to 6 as 'fail' with 2 meaning 'good' and 3 'satisfactory'.

	Condition			
	face-based		corner-based	
	Mean	Standard Deviation	Mean	Standard Deviation
SEA scale	33.17	24.50	32.33	28.44
Smiley scale	0.17	0.98	0.83	0.98
School grade	3.17	1.17	2.38	1.17

Table 1: Average rating of both interface types for perceived mental effort (SEA scale), joy of use (smiley scale), and according to German school grades.

Although the average given grade is slightly better for the corner-based (2.38) than for the face-based (3.17) interface, this difference is not statistically significant ($F_{1,5}$=4.395, p=0.09). With regard to overall joy of use expressed on a five point smiley scale ranging from a frowning face (coded as -2) to a smiling face (codes as +2) with a neutral face in between (coded as 0) the face-based interaction method is rated as almost neutral (0.17 in table), the corner-based method gets a slightly positive rating (0.87). However, this difference is also not statistically significant ($F_{1,5}$=4.000, p=0.102). Experienced mental effort rated on a scale from 0 to 220 showed little

difference between both interfaces (see Table 1) with a value of 32-33 residing between 'hardly strenuous' (=20) and 'a little bit strenuous' (=40).

The AttrakDiff mini measures the user experience on four dimensions that describe the perceived pragmatic (PQ) and the hedonic qualities (HQ-I & HQ-S) as well the overall attractiveness of a product (Hassenzahl & Monk, 2010). Although the corner-based interaction mode appears to get slightly higher values here (see Figure 12), a multivariate analysis of variance (MANOVA) with the values on all four dimensions as dependent variables yielded no significant difference between the two interface types ($F_{4,5}$=3.229, p=0.115).
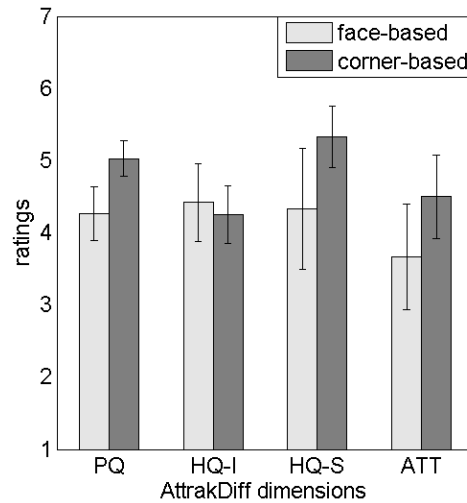

Figure 12: Average ratings for both interface types on the dimension of the AttrakDiff mini. PQ = Pragmatic Quality, HQ-I = Hedonic Quality – Identity, HQ-S = Hedonic Quality – Stimulation, ATT = Overall attractiveness. Whiskers denote standard errors.

To assess the relation between the user performance and subsequent questionnaire ratings, we calculated correlations between the various ratings of each subject and the average values of the interaction parameters across all five trials per condition as well as their maximum during such an experimental block (see Table 2). The maximum was used because it is known that retrospective subjective assessment is biased towards extreme (negative) incidents (Miron-Shatz et al. 2009), which is also discussed for the case of user experience evaluation (Wechsung et al. 2012).

The values in Table 2 appear to confirm this: while subjective mental effort (SEA scale) correlates significantly with average trial duration and number of interactions, the pragmatic quality (AttrakDiff's PQ dimension) that describes as how useful a service or device is perceived, correlates mostly with maximum values. As these correlations with measurement error, duration, and number of interactions are all negative, this means that mainly the largest errors, longest durations, and highest numbers of interactions lead to lower ratings. Hedonic qualities are less affected by interaction performance, the exception being the dimension 'Stimulation' (HQ-S),

which increases with average touch movement time. Apparently, this type of interaction is assessed as innovative and challenging, and will thus lead to a higher rating.

Interaction Parameter	SEA Scale	Smiley-Rating	School grade	AttrakDiff dimensions			
				PQ	HQ-I	HQ-S	ATT
Measurement error mn	.068	-.492	.333	**-.844**	.015	-.223	-.427
Measurement error max	-.120	-.391	.308	**-.773**	.036	-.261	-.328
Duration mn	**.578**	-0.284	.293	-.194	.331	.294	-.005
Duration max	.434	-.513	.400	**-.583**	.030	.193	-.250
#Interactions mn	.336	-.357	.218	-.501	.184	.033	-.033
#Interactions max	.395	-.455	.264	**-.677**	.07	-.035	-.484
Touch time mn	**.580**	.105	-.079	.205	.354	**.603**	.337
Touch time max	.558	-.261	.176	-.239	.060	.353	-.050

Table 2: Pearson correlation coefficients of the average (mn) as well as maximum (max) values of interaction parameters and questionnaire results. Correlations that are significant on the 5%-level are indicated in bold. For further explanation see text.

To summarize the results, we found that there was a significant improvement in the accuracy of measuring objects using the corner-based interaction without confounding issues related to higher mental effort or lower joy of use. Moreover, the improvement in accuracy was found not to be caused by increased interaction time or any other related performance measure. Surprisingly, increased touch interaction was perceived as stimulating/exciting by the users.

Having explored the measure and model capture[2] process we shall next address aspects of model manipulation and placement in augmented environments.

VIRTUAL OBJECT PLACEMENT AND MANIPULATION

The key step to the augmented reality application for virtual furnishing after measuring is the ability to virtually place, move and manipulate objects in an environment. To do so, we again employ the marker to determine the camera's pose. Then we render a realistic view of the furniture model in true size depending on the pose of the user.

By manipulate we mean that the user is allowed to move the object virtually along the ground and rotate the object in the plane of the marker. The interaction metaphors are again in line with those used in Visual Metrology. We have two interaction paradigms which we shall elaborate briefly. We also provide snapshots of the system in action as well as results comparing real objects to captured and modeled objects using our system.

The topic of virtual object translation and rotation via gestures is mostly discussed in the context of digital tabletops, that also allow the user to manipulate the objects displayed on their surface. Here, the main question is the number of touch input points that are possible and the degree of freedoms of object movements that are desired (Hancock, Carpendale, & Cockburn, 2007).

---

[2] The presented app also allows to capture surface textures of a determined model. A detailed description of this feature can be found in the appendix.

The task of rearranging furniture allows us to narrow the interaction space to 4 degrees of freedom: translation in all three dimensions, i.e. moving a piece of furniture in the marker plane (2 dimensions) plus up and down, i.e. above and below the marker plane. Rotation is limited to the spinning around the *Z*-axis, i.e. changing the side that is facing the user, for example a column of drawers. Rotation around the *X* or *Y* axis are unlikely, as most pieces of furniture have a well-defined underside with e.g. feet, that prevents changes in these dimensions.

Regarding touch points, we aimed at relying on only one touch point to avoid unnecessary occlusions, in the context of gestural interfaces sometimes also called the "fat finger problem" (Widgor & Wixon, 2011), and to account for the fact that not all mobile devices may offer multi-touch.

Analogous to the interface for measuring an object, there are again two ways of interaction:
- **Sliding Faces:** Selecting the dimension the object is supposed to be moved on via a button, and then slide the object on that dimension via a swipe along a ruler on the screen.
- **Direct manipulation:** Moving the object directly via touch without prior pre-selection of the dimension of interest.

The latter version requires a clear mapping of gestures to movement dimensions, and a differentiation of movement (translation) and rotation. We will describe the two interactions in more detail in the following.

Sliding Faces Approach

Figure 13 shows the interface for button-based manipulation similar to the buttons for the face-based measurement mode (see Figure 4). The main advantage is that the user is explicitly made aware of the dimension she will be manipulating, the drawbacks being that the buttons occupy a part of the screen, and that the increase/decrease movement on the ruler does not always correspond to the dimension the object is moved on, e.g. a horizontal touch movement to initiate a vertical object movement.

The object is by default placed initially centered on the marker. The system assumes this to be the default position with *zero* translation or rotation. Whenever the user selects a face (via a button), and then swipes across the ruler, an internal counter increments or decrements the translation to the object in the said direction. This is displayed to the user by changing the augmented image in real-time. The same is also done for rotation by simply incrementing or decrementing the angle about the *Z*-axis by which the object should be rotated. This allows all manipulations to be immediately visible on the main screen, i.e. no delays when translating slider movements into object movements.

Direct Manipulation Approach

Analogous the measurement mode, the sliding faces approach places all interactions outside the virtually augmented image and thus only allows for indirect manipulating of the virtual object via buttons and ruler. As a first attempt to also offer a direct touch-screen based interaction scheme for placing, we chose the following mapping of gestures (illustrated in Figure 14) to movement dimensions:

- **Drag object**: Touching the *augmented* object, and dragging it to another location in the image corresponds to moving the furniture in the marker plane, i.e. along dimensions *X* and *Y* in Figure 2.
- **Vertical swipe movement**: Touching the screen (outside of the object), and swiping the screen vertically up or down, results in a vertical translation to the object in 3D, i.e. along the *Z*-axis in Figure 2.
- **Horizontal swipe movement:** Not touching the object and swiping across the screen horizontally spins the object in the plane of the marker, i.e. a rotation about the *Z*-axis in Figure 2.
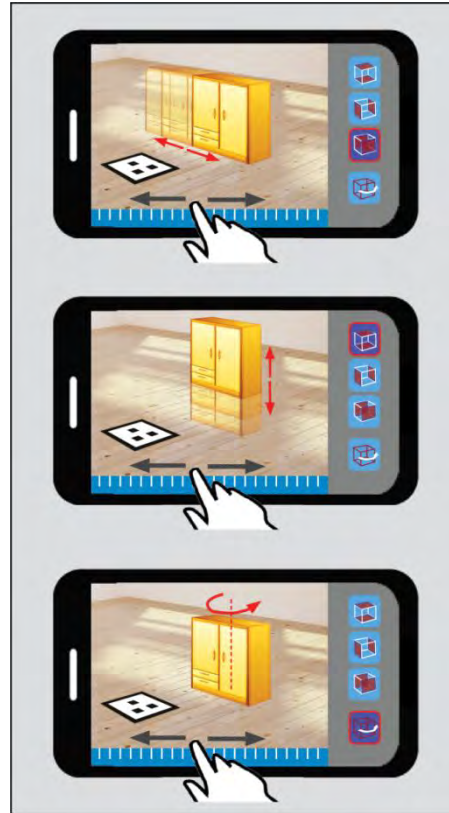

Figure 13: Object manipulation with button-based interaction. Here the user can choose the dimension the object is supposed to be moved on via button. Then a simple swipe of the finger to the left or right on the ruler translates the box in the selected direction or rotates it around the Z-axis if the rotate button (lowest) is chosen.

As with the sliding-faces approach, all movements are simply expressed as adding or subtracting values (translation or rotation) from the object position in the three dimensions. The only difference is the drag object interaction, which was planned as being closer direct push-pull of the object along the floor. When the user touches the object on the screen, the system recognizes this act. Then, the point in the image under the finger is mapped directly to the corresponding point in 3D on the floor (marker plane) via the homography relationship in Eq.(1). Note that, depending on how the camera was held, it is possible for the user to drag the object to a point that lies very far away (mathematically) or above the horizon. To avoid such annoyances, we restrict the movement of objects to within *10* meters of the marker.
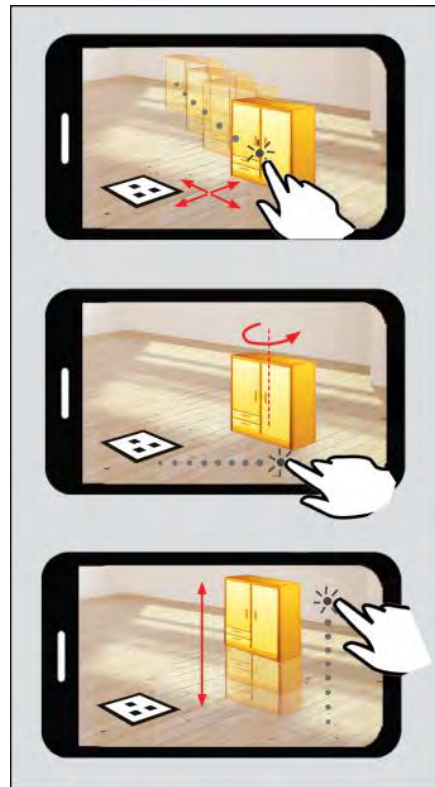
Figure 14*: Direct object manipulation - exploiting the fact that the whole screen is touch-sensitive, the user can move the virtual furniture directly in the marker plane by touching and then dragging it. Rotation around the Z-axis is achieved by horizontal swipe gestures outside the object. Lifting by vertical finger movement outside the object. The system then computes directly the position of the object in the room relative to the marker.*

Informal user test – preliminary results

As we were not sure whether users would understand our proposed mapping of gestures to object movements described in the previous paragraph, especially the gesture required for rotation, and to what extent the option of direct manipulation would be beneficial at all here, we conducted a small user test where we offered a mixed interface, i.e. rotation via horizontal swipe on the screen as well as via a button. Similar to the user test described before, six colleagues from our lab had to complete a task five times in a row, this time placing an object in various positions. Unlike the previous test, the main purpose was not to compare two variants, but rather to see whether the users would pick up the direct manipulation approach at all, or would still use a button for rotation. In addition, we were mostly interested in qualitative feedback, less in quantitative data.

Results were rather mixed: while the rotate button was only used in 6 trials out of 30, whereas the on-screen rotation in 26/30, users still described the gestures as not intuitive without further information on the screen. Moreover, average on-screen rotation duration was highly correlated to negative school grades (r=.922, p=.009) and low ratings in pragmatic quality (r=-.893, p=.016), indicating that the interface was perceived as less useful.

These preliminary results convinced us that the direct manipulation still needed some improvement regarding interaction metaphor and mapping of gesture to manipulation type. So we decided to keep the button-based object placing interface for the time being in the version that is currently available in the Google's Play Store. It is accessible via www.happymeasure.de.

DISCUSSION AND CONCLUSIONS

This paper described the challenges of taking augmented reality applications to the mobile phone with a specific application, virtual furnishing. In the first part, we showed how the task of deriving measurements of a three dimensional object from a two dimensional image using a single paper marker could be achieved via two different approaches by either extending the faces of a virtual control-box or by snapping the corners of the same control-box to the corners of the object to-be-measured. Either approach was linked to a different interaction concept: the first located all resizing interactions outside the actual image of the object, the second allowed for manipulations directly in the image.

From a mathematical point of view, the corner-based direct approach models the box fitting problem completely and allows for errors in marker placement, etc. Thus, we expect it to be more robust to measurement process compared to the face-based approach. The practicality of positioning corners accurately enough however, depends on the usability of such an interface and therefore need not lend to more accurate measurements. We were glad however, that the user test also showed that the corner-based interaction allowed people to measure more precisely. In addition, there was no speed-accuracy tradeoff, meaning no general relation between measurement error and time spent on measurement ($r=.1$).

The results also confirmed the usefulness of self-assessment in addition to performance data. Next to obvious findings like the perceived mental effort being directly related to task duration ($r=.578$), also more subtle aspects were revealed, for example that people judged the usefulness of the app under investigation more on their outcome (correlation of pragmatic qualities to measurement error $r=-.844$) than on the process of measuring (correlation to mean duration $r=-.194$). The low correlations of hedonic qualities and overall attractiveness to performance data once again confirmed the validity of the Attrakdiff dimensions (Hassenzahl & Monk, 2010). One exception being the average on-screen touch time: on-screen touch movements were perceived as more exciting (correlation to HQ-S). Apparently users appreciated the attempt to exploit the possibilities of mobile touch screens. However, this statement is not valid under all circumstances as was shown with our preliminary results regarding the placing mode, where the on-screen (touch-based) rotation correlated with lower ratings for usefulness. Thus, touch manipulation is not beneficial per se, but has to be accompanied by a meaningful gesture mapping, which is perhaps not that surprising at all.

The test we conducted for the place mode was of rather informal nature, and thus the results have to be interpreted with care. We mention them nevertheless to let the reader benefit from these insights, and at the same time show possible room for improvements.

In a similar vein, the first user study (regarding the measurement mode) had a limited sample size of $n = 6$, and differences between the two interface types concerning in number of interactions and touch movement time that were already present in the data might have become

significant with larger *n*. However, that was not our main intention – we wanted to see whether the two interfaces showed differences of practical relevance, not statistical significance. That we could confirm, namely the lower measurement error for corner-based measurement. At the same time, this interaction type is perceived as more stimulating by the users due to the direct touch interaction.

Future research of ours will surely include a satisfying mapping of gestures on 2D surface onto object manipulation in 3D for the place mode. Finding appropriate interaction concepts for rotation and translation is a non-trivial task, as illustrated by (Froehlich, Hochstrate, Skuk, & Huckauf, 2006; Hancock et al., 2007), who both deal with stationary devices. For mobile devices, the issue may even be enhanced. On the other hand, the availability of additional sensors in this device class may offer additional possibilities.

REFERENCES

A. Criminisi, I. Reid, A. Z. (2000). Single View Metrology. *International Journal of Computer Vision*, *40*(2), 123–148. doi:10.1023/A:1026598000963

Anabuki, M., & Ishii, H. (2007). AR-Jig: A Handheld Tangible User Interface for Modification of 3D Digital Form via 2D Physical Curve. *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality* (pp. 1–10). IEEE. doi:10.1109/ISMAR.2007.4538826

Bay, H., Ess, A., Tuytelaars, T., & Van Gool, L. (2008). Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding*, *110*(3), 346–359. doi:10.1016/j.cviu.2007.09.014

Chang, Y.-N., Koh, R. K. C., & Been-Lirn Duh, H. (2011a). Handheld AR games — A triarchic conceptual design framework. *2011 IEEE International Symposium on Mixed and Augmented Reality - Arts, Media, and Humanities* (pp. 29–36). IEEE. doi:10.1109/ISMAR-AMH.2011.6093653

Chang, Y.-N., Koh, R. K. C., & Been-Lirn Duh, H. (2011b). Handheld AR games — A triarchic conceptual design framework. *2011 IEEE International Symposium on Mixed and Augmented Reality - Arts, Media, and Humanities* (pp. 29–36). IEEE. doi:10.1109/ISMAR-AMH.2011.6093653

Eilers, K., Nachreiner, F., & Hänecke, K. (1986). Entwicklung und Überprüfung einer Skala zur Erfassung subjektiv erlebter Anstrengung. *Zeitschrift für Arbeitswissenschaft*, *40*(4), 215–224.

Feiner, S., MacIntyre, B., Hollerer, T., & Webster, A. (1997). A Touring Machine: Prototyping 3D Mobile Augmented Reality Systems for Exploring the Urban Environment. *97 Proceedings of the 1st IEEE International Symposium on Wearable Computers*.

Fiala, M. (2005). ARTag, a Fiducial Marker System Using Digital Techniques. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'05)* (Vol. 2, pp. 590–596). IEEE. doi:10.1109/CVPR.2005.74

Fiala, P., & Adamo-Villani, N. (2005). ARpm: An Augmented Reality Interface for Polygonal Modeling. *Mixed and Augmented Reality, IEEE / ACM International Symposium on*, *0*, 196–197. doi:http://doi.ieeecomputersociety.org/10.1109/ISMAR.2005.13

Froehlich, B., Hochstrate, J., Skuk, V., & Huckauf, A. (2006). The GlobeFish and the GlobeMouse : Two New Six Degree of Freedom Input Devices for Graphics Applications. *CHI 2006*.

Gjosaeter, T. (2009). Computer Supported Collaborative Design Using Augmented Reality. *2009 International Workshop on Social Informatics* (pp. 35–40). IEEE. doi:10.1109/SocInfo.2009.21

Hancock, M., Carpendale, S., & Cockburn, A. (2007). Shallow-Depth 3D Interaction: Design and Evaluation of One-, Two- and Three-Touch Techniques.

Hartley, R., & Zisserman, A. (2003). *Multiple View Geometry in Computer Vision* (2nd ed.). New York, NY, USA: Cambridge University Press.

Hassenzahl, M., & Monk, A. (2010). The inference of perceived usability from beauty. *Human-Computer Interaction*, *25*(3), 235–260.

Hassenzahl, M. Burmester, M., & Koller, F. (2003). (2003). AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In J. Ziegler & G. Szwillus (Eds.), *Mensch & Computer 2003: Interaktion in Bewegung* (pp. 187–196). Stuttgart: B. G. Teubner.

Jäger, R. (2004). Konstruktion einer Ratingskala mit Smilies als symbolische Marken. *Diagnostica*, *50*(1), 31–38.

Kato, H., & Billinghurst, M. (1999). Marker tracking and HMD calibration for a video-based augmented reality conferencing system. *Proceedings 2nd IEEE and ACM International Workshop on Augmented Reality* (pp. 85–94). IEEE Comput. Soc. doi:10.1109/IWAR.1999.803809

Kato, H., Billinghurst, M., Poupyrev, I., Imamoto, K., & Tachibana, K. (n.d.). Virtual object manipulation on a table-top AR environment. *Proceedings IEEE and ACM International Symposium on Augmented Reality (ISAR 2000)* (pp. 111–119). IEEE. doi:10.1109/ISAR.2000.880934

Kato, Hirokazu, & Billinghurst, M. (1999). Marker Tracking and HMD Calibration for a Video-based Augmented Reality Conferencing System. *International Workshop on Augmented Reality*.

Kolsch, M., Bane, R., Hollerer, T., & Turk, M. (2006). Multimodal interaction with a wearable augmented reality system. *IEEE Computer Graphics and Applications*, *26*(3), 62–71. doi:10.1109/MCG.2006.66

Kurz, D., & Ben Himane, S. (2011). Inertial sensor-aligned visual feature descriptors. *CVPR 2011* (pp. 161–166). IEEE. doi:10.1109/CVPR.2011.5995339

Layar. (n.d.). Retrieved from http://www.layar.com/

Olsson, T., & Salo, M. (2011). Online user survey on current mobile augmented reality applications. *2011 10th IEEE International Symposium on Mixed and Augmented Reality* (pp. 75–84). IEEE. doi:10.1109/ISMAR.2011.6092372

Pan, Q., Reitmayr, G., & Drummond, T. (2009). ProFORMA: Probabilistic Feature-based On-line Rapid Model Acquisition. *Procedings of the British Machine Vision Conference 2009* (pp. 112.1–112.11). British Machine Vision Association. doi:10.5244/C.23.112

Papagiannakis, B. G., Singh, G., & Magnenat-thalmann, N. (2008). A survey of mobile and wireless technologies for augmented reality systems (pp. 3–22). doi:10.1002/cav

Piekarski, W. (2006). 3D Modeling with the Tinmith Mobile Outdoor Augmented Reality System. *IEEE Computer Graphics and Applications*, *26*, 14–17. doi:http://doi.ieeecomputersociety.org/10.1109/MCG.2006.3

Piekarski, W., & Thomas, B. H. (2002). The Tinmith system: demonstrating new techniques for mobile augmented reality modelling. *Australian Computer Science Communications*, *24*(4), 61–70.

Quené, H., & van den Bergh, H. (2004). On multi-level modeling of data from repeated measures designs: a tutorial. *Speech Communication*, *43*(1-2), 103–121. doi:10.1016/j.specom.2004.02.004

SPSS. (2005). *Linear Mixed-Effects Modeling in SPSS: An Introduction to the MIXED Procedure*. Chicago, IL.

Sauro, J., & Dumas, J. S. (2009). Comparison of Three One-Question, Post-Task Usability Questionnaires. *Proceedings CHI 2009*.

Sturm, P., & Maybank, S. (1999). A Method for Interactive 3D Reconstruction of Piecewise Planar Objects from Single Images. *British Machine Vision Conference* (pp. 265–274).

Wagner, D., Reitmayr, G., Mulloni, A., Drummond, T., & Schmalstieg, D. (2008). Pose tracking from natural features on mobile phones. *2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality* (pp. 125–134). IEEE. doi:10.1109/ISMAR.2008.4637338

Widgor, D., & Wixon, D. (2011). *Brave NUI World - Designing Natural User Interfaces for Touch and Gesture*. Burlington, MA: Morgan Kaufmann Publishers.

Yi Ma, Stefano Soatto, Jana Kosecká, S. S. S. (2004). *An Invitation to 3-D Vision*. Springer.

Zhengyou Zhang. (1999). Flexible camera calibration by viewing a plane from unknown orientations. *Proceedings of the Seventh IEEE International Conference on Computer Vision* (pp. 666–673 vol.1). IEEE. doi:10.1109/ICCV.1999.791289

de Sa, M., Churchill, E. F., & Isbister, K. (2011). Mobile augmented reality. *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services - MobileHCI '11* (p. 749). New York, New York, USA: ACM Press. doi:10.1145/2037373.2037504

APPENDIX I

TEXTURE COMPUTATION FOR IMAGE BASED RENDERING (IBR)

Accurate measurement alone does not provide a full sense of the space occupied and the appearance of furniture. For this purpose, our system is also capable of automatically extracting textures from the object's images. The textures are then used for image based rendering of furniture virtually placed in the scene.

Automatic Textures for Image based Rendering

While being able to compute the size of furniture is interesting, it alone does not convey the full impression of how that furniture would fit in another environment. In order to get a better and more complete sense of space occupied it is necessary to be able to capture the furniture and place it virtually in another location. The user should be able to virtually walk around it and ―feel" its presence. This is only possible if the virtual placed furniture not only has the size of the original object but also the appearance.

We employ image based rendering to compute textures for the object model. Once the object's position and size is precisely measured via visual metrology, the user is then capable of capturing multiple still-shots of the furniture. For every image acquired the system automatically processes it to extract relevant texture information. The details of this process including ways to combine multiple images shall be described in the subsequent sections.

After the 3D location of the box in the world is known, we can extract the face textures from different images. Using the coordinates of the furniture corners in 3D (obtained after measurement) together with the projected image corners we estimate multiple homographies per bounding box' face. Then, for every possible face the following steps determine texture computation:

- **Check Visibility**: For every image determinate which faces are visible to the camera and which ones are occluded.
- **Extract textures**: Create a texture image of a face, using the available pixels in the original image. Since each texture is a rectangular image, resampling is done using the computed face homographies.
- **Combine textures**: Of all extracted textures of a face, create one consistent texture image.
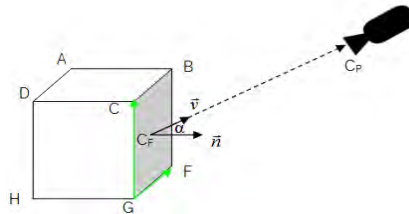

Figure 15 : Visual representation of view vector v and surface normal n of face:CBFG. A positive dot product determines whether the face is visible from the given camera position.

Visibility check

From every image taken by the user it is possible to extract 1 to 3 faces, depending on the camera position. So, first we need to determinate which faces are visible to the camera. As shown in Figure 15, we can compute the unit vectors from the center of the face to the camera ($\vec{v}$) and the normal to the face ($\vec{n}$), given the 3D location of every vertex of the furniture and the camera position. Finally, since ($\vec{v}$) and ($\vec{n}$) are unit vectors, their dot product ($\vec{v} \cdot \vec{n}$) reveals whether the face is visible ($\vec{v} \cdot \vec{n} > 0$) or hidden ($\vec{v} \cdot \vec{n} < 0$).

Texture extraction via face-homographies

To create a texture, we need to fill every pixel in the texture (q_i^t) with a pixel (q_i) obtained from the source image. To do that, we need the homography H, which maps any point of the texture to a point in the source image as shown in Figure 16. The homography is computed by relating the four vertices of the textures correspond to the four vertices of the face in the source image.
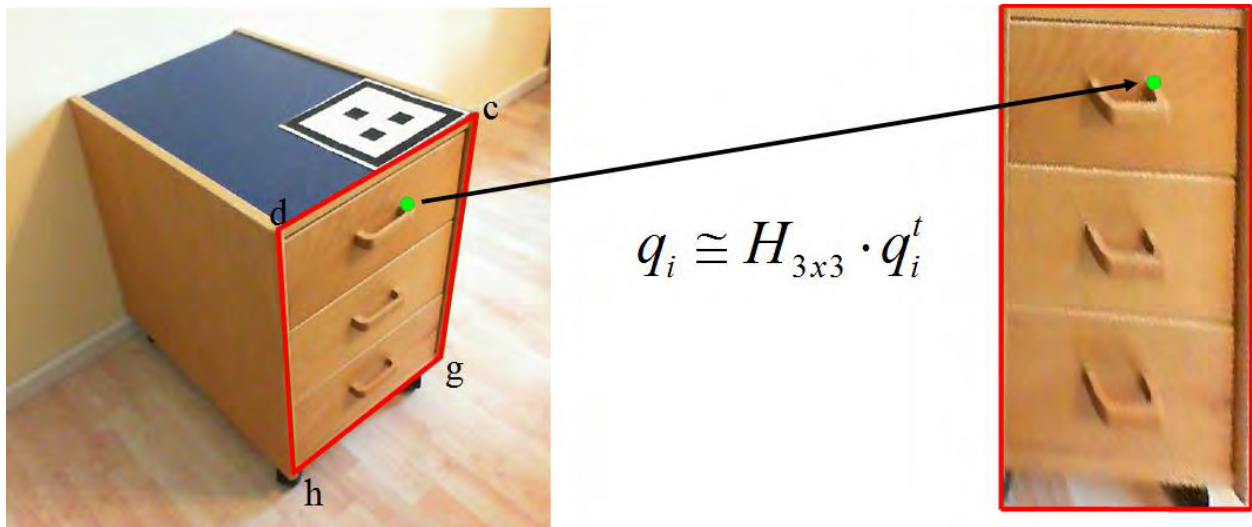

$$q_i \cong H_{3x3} \cdot q_i^t$$

Figure 16: A homography is used to find the corresponding pixel q_i in the source image and map it to the texture q_i^t, defined as: $q_i \cong \boldsymbol{H} \cdot q_i^t$.

Combining textures

After extracting all possible faces from every source image it is possible to have various textures for the same face. Therefore, we need to have a criterion to select a final texture. To do this, we need to calculate for every texture a weight that represents how good the texture represents the real face of the object.

To calculate the weight of a texture two factors are considered:
- **Resolution or area:** The resolution or area in pixels of the quadrilateral region in the source image, from where the texture pixels where mapped. The bigger the area, the better the resolution of the texture.
- **Viewing angle:** The viewing angle of the face. This means that a frontal view of the face is preferred to an oblique one.

If the four vertices of the face are (x_a, y_a), (x_b, y_b), (x_c, y_c) and (x_d, y_d) as shown in Figure 17 the area resolution A_i^f for face f and source image i can be calculated as:

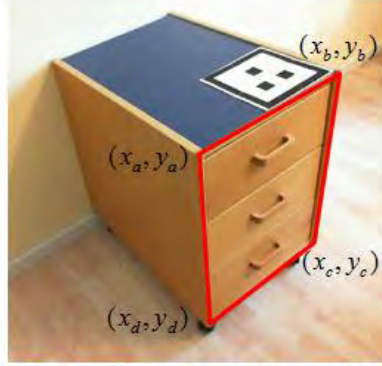$$A_i^f = \frac{1}{2} \cdot \left| (x_c - x_a) \cdot (y_d - y_b) - (x_d - x_b) \cdot (y_c - y_a) \right|$$

Figure 17: The 2D vertices a,b,c,d of face ABCD are used to calculate the area of that face regarding a specific texture.

Together with the visibility factor $c_i^f = \vec{v_I} \cdot \vec{n_f}$, we define the weight of a texture as:

$$w_i^f = A_i^f \cdot C_i^f$$

Heuristics for combining weighted face images.

Once we have the weight information as criterion of how good a texture is, two different heuristics to get a final texture are possible.

Weighted average texture.

This heuristic combines the pixel values of all extracted textures of a face, giving priority to those textures with higher weights.

After N textures have been processed, the value final value for a pixel P_N for every pixel in the final texture is given by:

$$P_N = \frac{\sum_{i=1}^{N} w_i \cdot p_i}{\sum_{i=1}^{N} w_i}$$

Where p_i is the value of the pixel in the i-th extracted texture with weight w_i. Notice that N is the number of extracted textures for a certain face and not the total number of source images.

Best texture.

This heuristic aims to store always the best extracted texture of a face. For every image acquired we compute the visible faces and first determine their weights. If the currently extracted face weight is better than what was previously computed, we replace the previous texture with that extracted from the current image.

We notice this greedy approach leads to sharper images as seen in Figure 18 and has a much higher user acceptance than normalized weighted textures.

Figure 18: Comparing the virtually augmented object after image based rendering (with cyan colored frame) with the original object. The placement was performed using the face-based manipulation paradigm.