Mobile User Interfaces for Efficient Verification of Holograms

Andreas Hartl*
Graz University of Technology

Jens Grubert†
Graz University of Technology

Christian Reinbacher‡
Graz University of Technology

Clemens Arth§
Graz University of Technology

Dieter Schmalstieg¶
Graz University of Technology

Abstract

Paper documents such as passports, visas and banknotes are frequently checked by inspection of security elements. In particular, view-dependent elements such as holograms are interesting, but the expertise of individuals performing the task varies greatly. Augmented Reality systems can provide all relevant information on standard mobile devices. Hologram verification still takes long and causes considerable load for the user. We aim to address this drawback by first presenting a work flow for recording and automatic matching of hologram patches. Several user interfaces for hologram verification are presented, aiming to noticeably reduce verification time. We evaluate the most promising interfaces in a user study with prototype applications running on off-the-shelf hardware. Our results indicate that there is a significant difference in capture time between interfaces but that users do not prefer the fastest interface.

Keywords: Document inspection, holograms, augmented reality, user interfaces, mobile devices.

Index Terms:
H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, augmented, and virtual realities, H.5.2 [Information Interfaces and Presentation]: User interfaces—Input interaction styles, H.4.9 [Image Processing and Computer Vision]: Applications—, I.7.m [Document and Text Processing]: Miscellaneous—.

1 Introduction

While trained professionals can identify the majority of fake documents or holograms within a few seconds[1], most lay people inspect holograms on security documents just by looking for changes in appearance or the pure presence of rainbow colors, which has no particular value w.r.t. security [20]. First level inspection of holograms is currently based on printed guides which are often issued by public authorities. They usually show distinct patterns visible within the hologram area. However, they often lack an indication on the viewing direction and do not specify requirements on the lighting conditions. Consequently, the inspection may be tedious for the untrained user. Also, in real-world situations manuals are likely not always at hand, so users fall back to solely looking for appearance changes.

We propose user interfaces for efficient verification of holograms targeting laymen (see Figure 1). Specifically, our target audience consists of individuals who did not receive advanced training for checking holograms. We first address the mobile recording and matching of hologram patch data as a basis for automatic verification systems running in real-time. The problem is subsequently treated as an alignment task, for which we present a constrained navigation interface. This process finally leads to the design of a hybrid user interface. We compare these interfaces in a user study involving original and modified documents, informing a detailed discussion on the usefulness of these interfaces.

Figure 1: User interfaces for hologram verification: Constrained navigation (top-left), alignment (top-right) and hybrid user interfaces (bottom-left) are designed, implemented and evaluated within a user study. They allow to reliably capture image data suitable for automatic verification. Results are presented to the user in a summary (bottom-right).

2 Related Work

Hologram verification is a complex task that demands repeatable image capture conditions and robust matching with reference information. In case of mobile setups, this also requires suitable user guidance to efficiently capture relevant information.

Hyuk-Joong and Tae-Hyoung perform automatic inspection of holograms [11, 14] in a stationary setting. They illuminate the document using infrared LEDs located on a hemisphere. Images are captured with a CCD camera at controlled illumination angle, followed by frequency correlation matching for verification of hologram patterns. They subsequently extend the system with correction of hologram rotation angles and evaluate it with two Korean banknotes. Pramila et al. [15] perform detection of a watermark embedded in a dual-layer hologram. They place the hologram on uniform background and use a stationary, well-aligned image acquisition setup with controlled illumination. Janucki et al. [10] use a Wiener filter to quantitatively assess holograms. Their setup uses LED illumination for hologram capture, followed by registration of the background and the optical device.

The feasibility of capturing and verifying holograms in a mobile AR setting has been demonstrated using an off-the-shelf mobile device [6]. By using the built-in flashlight of the device as a dominant light-source, the appearance of reference patches can be reproduced in a mobile context. Still, this approach requires manual matching of recorded patches by the user. Compared with a refer-

*e-mail:hartl@icg.tugraz.at
†e-mail:grubert@icg.tugraz.at
‡e-mail:reinbacher@icg.tugraz.at
§e-mail:arth@icg.tugraz.at
¶e-mail:schmalstieg@icg.tugraz.at

[1] according to a domain expert consulted

ence manual, this approach takes a lot of time and involves heavy physical and cognitive load for the operator. The user is guided towards the required poses using a complex alignment-based user interface, where most parts are augmented onto the target. Proper alignment requires pointing at the base of the reference viewing ray, looking into its direction (iron sights) and adjusting the distance using two appropriately scaled circles at the base and top of the ray. Finally, the orientation around the viewing ray must be matched (virtual horizon). In contrast, we use a different alignment sequence and visualization along with automatic image capture and matching, which should make the process more efficient.

The user guidance component required for mobile hologram capture and verification deserves special attention in the context of this work. User guidance can be approached by visualization of the view alignment error concerning a given reference pose. Examples are surgical scenarios in telemedicine, where colored augmented coordinates are used for easier navigation of the end effector [3]. Pyramidal frustums can also serve as a means of guidance for navigation. This can be seen as a geometric representation of the camera at the time of capture [17]. This approach is used for real-time visual guidance for accurate alignment of an ultrasound probe by Sun et al. [19]. After tracking artificial skin features for probe localization, visual guidance for 6 DoF alignment is provided via an augmented virtual pyramid. Such a pyramidal representation is also related to the Omnidirectional Funnel [2], which is useful for calling attention. For practical purposes, view alignment can be conducted in several steps (see [6]). Bae et al. [1] use visual guidance for re-photography. They analyze the camera image to determine, if a sufficiently similar image was captured. Then, three visualizations are presented for alignment. First, a 2D arrow indicates the required direction of movement w.r.t. a top-down camera viewpoint. Second, this information is also indicated concerning a back-front camera viewpoint. Finally, they visualize edges for adjustment and feedback of the current camera orientation. Heger et al. [9] perform user-interactive registration of bone with A-mode ultrasound. The pointer is mechanically tracked and a 2D-indicator is used to provide visual feedback about the deviation from the surface normal during alignment of the transducer to the local bone surface.

Alternatively, guidance can be achieved by visualization of a constrained navigation space. Shingu et al. [16] create AR visualizations for re-photography tasks. They use a sphere as a pointing indicator along with a half-transparent cone having its apex at the sphere as an indicator of viewing direction. Once the viewpoint is inside the cone, it is not visible anymore. The sphere then changes its color when it is fully visible. This corresponds to a valid recording position. Sukan et al. [18] propose a wider range of look-from and look-at volumes for guiding the user to a constrained set of viewing positions and orientations, not counting roll (ParaFrustum). This can be realized as an in-situ visualization or via non-augmented gauges. In the in-situ variant, the transparency of volumes is modulated depending on the distance and orientation of the current pose. In addition, the general representation of the look-at volume is also changed. Although our approach for constrained navigation is similar, the mobile capture of holograms requires the user not only to enter, but to explore such space in order to get suitable image data. For reasons of efficiency, this must be accounted for in the corresponding visualization together with the requirement to work in small workspaces and on small screen sizes.

3 Mobile Hologram Verification

Hologram verification can be seen as a subtask within a document verification process. In the following we describe a setup for capturing reference data from holograms along with a matching approach that can be used for automatic verification at runtime. Such information can be used in a mobile pipeline for interactive document verification, which performs classification, tracking and augmentation of relevant information (Figure 2). This allows to select the correct reference information for hologram verification and to make sure, that the element is observed from the correct viewpoint.

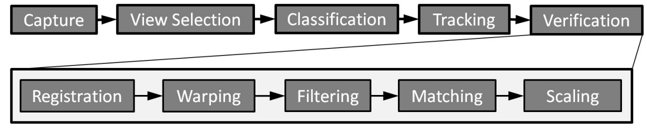

Figure 2: Overview of our mobile document verification pipeline. For verification, images are first registered. Then, the extracted hologram patches are warped and filtered. Finally, the intermediate results of different similarity measures are fused using precomputed scaling coefficients.

3.1 Preprocessing

Capturing Reference Data With moderate ambient light, the appearance of a hologram is largely dominated by using the LED flashlight of mobile devices. This essentially means that the workspace consists of a hemisphere centered at the hologram on the document. In contrast to previous work [6], we use an industrial robot (Mitsubishi MELFA) for capturing all relevant appearances of a view-dependent element (see Figure 3). This allows reliable sampling of holograms and eliminates undesired human influence. We spatially sample a hemispherical space using the robot and remotely control the device. We capture the current video image and the corresponding pose for each position on the hemisphere.

We assume the hologram to be planar and project its bounding box into the image using the recorded pose. We estimate an image transformation with respect to the hologram region on the undistorted template and subsequently warp the sub-image containing the hologram. For increased accuracy, we perform an additional registration step using the template of the document before extraction and rectification of the corresponding patch. The result is a stack of registered image patches that represent all observable appearances of the current hologram.

View Selection For successful verification, a series of representative views must be selected using reference information available from the manufacturer or by systematic recording of the hologram and thorough analysis of the captured image data. The choice of reference poses obviously depends on the hologram (e.g., number of transitions) and is constrained by the particular setup being used. We exclude all data recorded $< 5^o$ and $> 55^o$ away from the orthogonal view in order to avoid artifacts caused by orthogonal views and tracking failure.

From the perspective of security it seems reasonable to select very different patches having small distances in space. For reasons

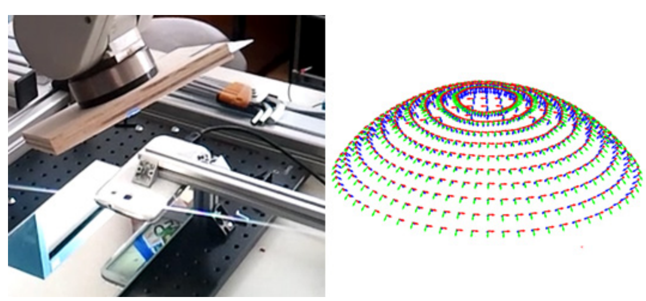

Figure 3: We sample the view-dependent element on a document using an industrial robot and an off-the-shelf mobile phone (left). Due to using a single dominant light source, the element is sampled from viewpoints situated on a hemisphere (right).

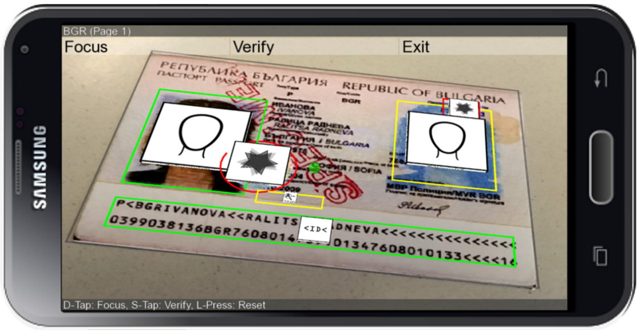

Figure 4: Mobile document verification system tracking an instance of a passport. Security features are augmented onto the document. Detailed information about an element can be triggered using a focus point at the center of the screen.

of practical usability, a small amount of stable views seems preferable, because the user will move. As a lower bound, at least two visually different views recorded from sufficiently different viewing directions are required.

3.2 Runtime processing

Classification and Tracking The first step in document verification, the identification of the document type, can be achieved by manual selection or by computing the class of the document using images from the current instance to be verified. We use visual search running entirely on the mobile device. This avoids the transfer of sensitive information across networks and reduces latency [7]. Then, associated reference data relevant to the verification process can be loaded. Tracking works in real-time directly on the mobile device, using natural features obtained from an exemplary template selected during the previous process [22]. In order to allow the verification of slightly bent documents, tracking poses are smoothed in a small ring buffer. Available reference information can be represented by an initial augmentation, providing instant feedback on the presence and location of relevant security features for manual verification (Figure 4).

Matching Verification of selected reference data demands a suitable similarity measure, which may then also be used on-line for verification. In the work of Hartl et al. [6], normalized cross correlation (NCC) is evaluated for matching, giving reasonable results for the majority of views. However, certain holograms (e.g., rainbow) show a large amount of different colors, which leads to noisy measurements. From our experience, pre-filtering (Gauss, Median) with a reasonable kernel size improves robustness for this type of holograms. On the other hand, stereograms tend to produce more distinct patches, which lead to strong edge responses. Consequently, it seems reasonable to use both intensity and edge information in the process. We propose shape-matching using the modified Hausdorff distance [4] and weight the contributions according to Eq. 1,

$$\begin{aligned} score = (s_{NCC} * k_{NCC} + d_{NCC}) f \\ + (s_{MHD} * k_{MHD} + d_{NHD})(1-f) \end{aligned} \quad (1)$$

where s_{NCC} and s_{MHD} denote the corresponding individual results, k and d denote individual scaling coefficients, and f is a weighting factor. The individual scaling coefficients are computed using all the recorded reference data and are used to perform appropriate scaling at runtime (see Figure 2). We use integral images and a sliding window to avoid a costly registration step. With these modifications, matching runs in real-time on mobile devices.

Viewpoint and flashlight act as triggers for different appearances. From our experience, not using the flashlight as a dominant light source does not give repeatable results. According to informal tests, there is limited invariance with different devices, depending on camera properties, flash intensity and relative position.

4 User Guidance for Hologram Verification

In order to get reasonable input data for verification, the user should be supported throughout the image capture process. In general, manual interaction such as tapping on the screen is not desirable for reasons of accuracy. Consequently, image capture should be triggered automatically, when the user is in a suitable position.

We observed that many holograms feature similar appearances in very different locations. Consequently, one could think of rejecting the entire pose information during matching and just taking care that the user is pointing towards the hologram. However, this does not seem feasible, since users cannot be expected to sample the hologram space (hemisphere) without guidance. This is backed up by an informal user study we conducted, in which a given hologram had to be sampled as deemed appropriate by the user. We obtained reasonable matching scores, but a very low spatial coverage. Users did not know when to stop the process. Although in case of originals, task completion time could be reduced by an early-exit based on the matching score, this is not feasible with fakes. It is mandatory to consider the viewing direction in order to get a good coverage of the pose space and provide a reliable exit mechanism. Using information about the viewing direction when matching provides additional verification security.

An obvious approach is to guide the user to align the mobile device with exactly those view points, which are associated with the selected reference data. Alternatively, a portion of space can be visualized for sampling by the user, which requires coverage of a larger region instead of given positions. Combining both approaches leads to a hybrid variant, which uses a comparatively small region for sampling relevant data. In the following, we cover the design of these approaches in more detail. In favor of usability, we decided to omit an explicit check of the in-plane rotation of reference views during matching. This is motivated by the fact that when views are placed on a hemisphere, reasonable results can be achieved by just rotating the target.

4.1 Alignment Interface

Sampling holograms can be treated as an alignment task, where users have to point at the center of the element, align with the viewing direction using iron sights, match the rotation along the viewing ray using a virtual horizon and take care of the recording distance [6]. Although this causes a lot of strain, we believe that a careful design in conjunction with automatic recording and matching, can lead to considerable gains in efficiency. This could make the alignment approach a strong competitor for constrained approaches.

We propose an improved alignment interface, which was designed in an iterative process involving continuous user feedback. A sketch of the elements involved can be seen in Figure 5. We observed that users often had trouble matching the overall orientation/rotation of the document and the device with the original approach. Not being able to do so, makes the overall alignment process more tedious. Consequently, the revised approach starts with coarse alignment of document-device orientation. We project the camera center and the top point of a reference pose down on the target surface and compute the relative angle as a rough indicator of initial alignment. This can be visualized as a color-coded indicator within a circle around the element. Depending on the sign of the computed error, arrows are placed on the circle to indicate the required movement of the target. Upon successful alignment (within a certain range), we proceed with more accurate indicators for the viewing direction. We use animated rubber bands as indicators for

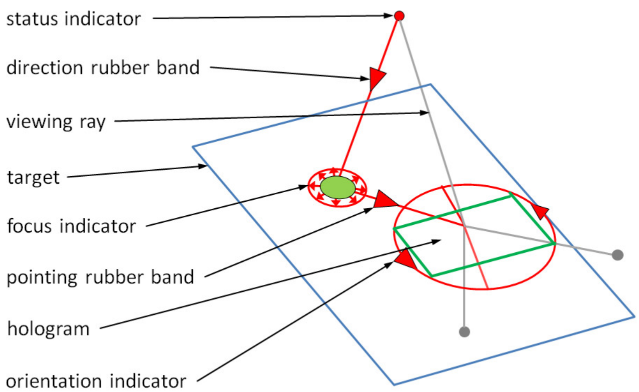

Figure 5: Geometry of the revised alignment approach. Matching takes place by alignment of target rotation and pointing with the indicator at the element. Finally the viewing direction is refined using the direction rubber band at an acceptable viewing distance.

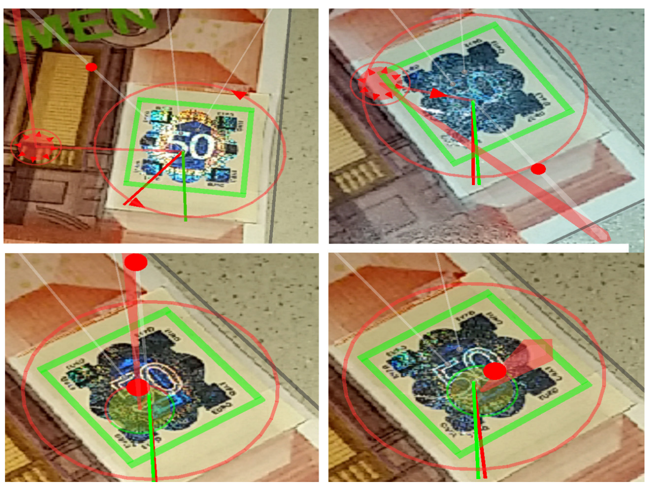

Figure 6: Exemplary alignment sequence: Not aligned (top left). Aligning target rotation (top right). Pointing at target (bottom left). Aligning viewing direction along hemisphere arc (bottom right).

pointing at the element, but also for the vertical angle on the hemisphere. In both cases, the goal is to follow the animated arrows in order to shrink down the rubber band into a point (see Figure 6). Finally, a focus indicator is realized as a scaled sphere placed at the base point of the current viewing direction on the target. Animated, directed arrows indicate the required direction of movement. Note that we perform an initial focus operation at the first view to be aligned and keep this setting throughout the process.

Views are captured sequentially, with feedback on the overall progress of the operation. This aims to reduce visual clutter for the user. Upon successful alignment, several frames are recorded from the live-video stream and automatically matched against prerecorded reference data. From these measurements, the one having the highest matching score is selected as the result patch for the user. During the process, we provide guidance towards the desired direction, but also feedback regarding the quality of alignment. Similar to the previous approaches, we aim to minimize the required movements for the user by automatic selection of the nearest view. A live-view of the rectified hologram patch is constantly displayed during spatial interaction in order to provide visual feedback of the changes in appearance with varying recording position (see Figure 6 for an exemplary alignment sequence).

After recording each of the views, a summary including the current overall decision (genuine/fake) is presented to the user (see Figure 1). The user may skim through the captured views and compare them side-by-side with the expected reference data. If the system suggestion is revised by the user, an overall similarity score is recomputed, which eventually changes the final decision. The user may also re-record certain views in order to get a better basis for the final decision. This can be done in the summary for the current view and works for all the approaches described in this paper.

4.2 Constrained Navigation Interface

The task can also be treated within a constrained navigation framework. The idea is to guide the user to sample larger portions of space instead of aligning with single views. By giving more freedom to the user, this can reduce workload and task completion time.

The initial step is to guide the user to point at the hologram as required by the recording setup. We provide guidance using an animated rubber band, which shows a moving arrow, once outside a given radius away from the element (see Figures 7, 8). Then, the capture distance needs to be adjusted as a starting point for an autofocus operation, so that the assumption about the flashlight being the dominant light-source holds. For this purpose, we scale the entire widget and require the user to adjust the distance, so that the outer ring of the widget stays within the given distance bounds.

Although the robot recording operates on a hemisphere, it does not seem reasonable to apply this concept directly. An augmented hemisphere would certainly lead to coverage of the entire space, but not necessarily in the shortest possible time. With an augmented hemisphere, the most obvious movement is to scan hull slices and then rotate the document for the next slide. We empirically verified that changing orientation from an orthogonal starting point (conic) is much faster than target rotation with slice-scanning.

In favor of efficiently treating both originals and fakes, the user should be guided towards different viewing directions or ranges. We propose a 2D orientation map (projection of the conic space) [9] for this task. It is divided into slices that are aligned on one or more tracks. The current position on the map is visualized by a cursor, and the current slice is also highlighted. The cursor position is corrected by the target orientation, so that the movement direction always corresponds to the orientation of the device (see Figures 7, 8). In general, it is not sufficient to just capture a single shot inside each slice. We record several shots per slice, that differ at least by a given angle threshold. The exact amount is automatically calculated, taking into account the area of the slice. Consequently, the user can move freely inside the pie slices during the process. The tiny arrows around the cursor serve as movement indicators. Whenever the user remains static inside a non-completed slice, flashing

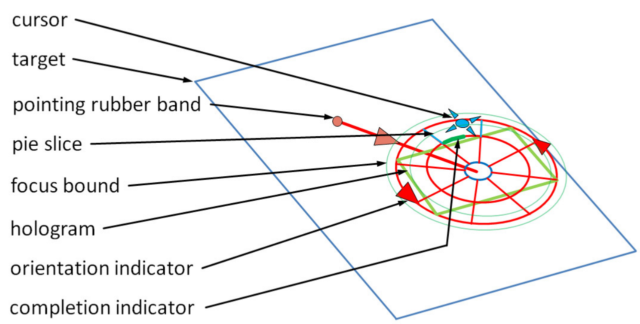

Figure 7: Geometry of the proposed constrained navigation approach for sampling the hologram. The user is guided to point at the element and a cursor is controlled by the 2D orientation on an augmented pie, divided into slices and tracks.

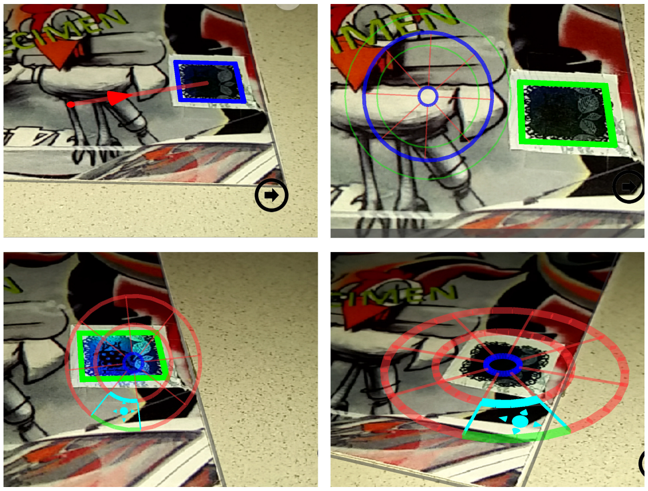

Figure 8: We guide the user to point at the element using an animated rubber band (top-left). Focus adjustment showing the layout of the orientation map and green distance bounds (top-right). Constrained navigation UI with pie slices (bottom-left). Augmentation directly onto the document/element (bottom-right).

arrows remind to move on. The upper arc defined by a (sub-)slice is used as a completion indicator, which switches from red to green with increasing slice coverage. The orientation map is realized as a widget placed in the screen plane (2D-CON) or augmented onto the target (AR-CON).

In a pilot study, we tried using either no visual information on the capturing procedure or a progress bar without any orientation information. Using no visualization at all gave the best completion time, but also the lowest spatial coverage. In the following, we dropped the interface without guidance and the progress bar. It must be noted that even with the AR-CON interface, not all participants sampled the entire hologram. Consequently, we went to incorporate slightly more guidance with the goal to only check pie slices containing a reference view (see Figure 9).

4.3 Hybrid Interface

The location of reference views cannot be mapped straightforward to pie slices. It may be necessary to associate several pie slices with a single reference view, increasing the amount of slices to be checked. Since the number is generally much lower than the total number of pie slices, we use small regions on the augmented map around reference locations, which also serve as local completion indicators (AR-HYB, Figure 9).

These two UIs were evaluated in another pre-study, this time involving a demonstration phase. According to the results obtained, AR-HYB had a much lower task completion time compared with AR-CON. Users were able to complete the task using both approaches (perfect coverage of interesting slices/regions) and obtained reasonable patch-matching scores. Users generally gave very positive ratings concerning the type of guidance and overall usefulness of the application, with a clear preference for AR-HYB. Motivated by user demand and our own reasoning, this clearly moved the approach more in the direction of an alignment task. As we consider our informal studies only suitable for guiding the design process, we conducted a more detailed evaluation.

5 Evaluation

We evaluated the most promising candidate for constrained navigation (CON) and the hybrid approach (HYB, see Figure 9) against

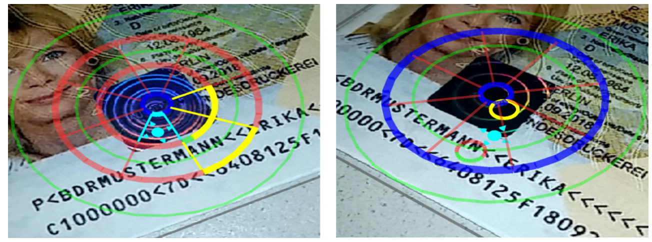

Figure 9: AR UIs with guidance for interesting subspaces. Either pie-slices (AR-CON, left) or circular regions (AR-HYB, right) are indicated for sampling by the user.

the alignment UI (ALI, see Figure 6). After image capture, a summary is presented to the user (see Figure 1) independent of the UI used for capture. The global system decision is communicated via a colored square (green...valid, yellow...unsure, red...invalid) to the user. Each reference has its own page, showing the reference data on the left side of the screen and the best recorded match on the right side, along with a local rating by the system, which can be changed by the user in case of doubt. It must be noted that we also monitored distance as capture condition, so that the users had to stay within the allowed distance range for the CON and HYB interfaces. We manually selected two reference views per hologram with a visually equal spatial distribution. We consider two views to be the minimum for verification of view-dependent elements.

5.1 Study Design and Tasks

According to a domain expert we consulted, professionals can identify most fake documents or holograms within a few seconds. The focus of the following study is on laymen without advanced domain knowledge or experience, using an off-the-shelf smartphone for hologram inspection. In contrast to the work of Hartl et al. [6], we do not compare a printed manual to an AR-System, but we seek to improve upon the long inspection time of AR-based hologram verification.

We designed a within-subjects study to compare both the performance and user experience aspects of the three aforementioned user interfaces for hologram verification.

The study had two independent factors: interface and hologram. The independent variable of main interest was interface (with three levels: ALI, CON, HYB). We modeled hologram as fixed effect (four level), since the holograms were deliberately selected (and not randomly sampled from a population) in order to represent intensity-dominated and shape-dominated samples including common mixtures.

For each of the four holograms, we selected the corresponding reference views with the goal of minimizing the variance an individual hologram could have on the results. Dependent variables of interest were task completion time (both capture and decision time), system performance (how well the system could verify the validity of the hologram), user performance (how well the user could verify the validity of the hologram), and user experience measures (usability, workload, hedonic and motivational aspects).

For each interface, the actual verification procedure started upon pointing the center of the screen at the element and tapping on it. For the ALI interface, the user had to align the rotation of the document with the current reference view (azimuthal angle), point at the center of the hologram and adjust the viewing direction (polar angle) along with the capture distance. In case of the CON interface, the user had to point at the element, following the base rubber band. Then, the orientation cursor had to be moved inside the indicated (connected) pie slices by changing the azimuth and inclination angles through device movement and monitoring the operating distance. The HYB interface had to be operated in a similar way.

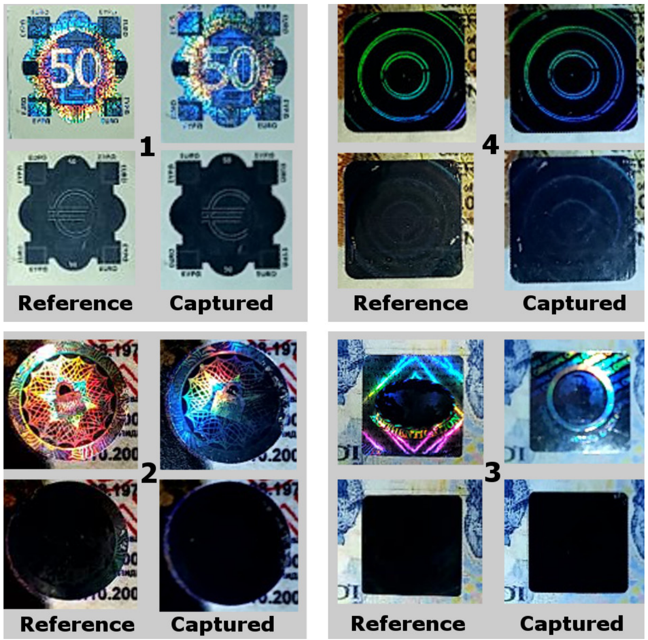

Figure 10: Samples used in our study. We evaluated all user interfaces with two original (no. 1, 4 - top row) and two fake (no. 2, 3 - bottom row) holograms, where each was placed on a different document template. Reference information recorded with the robot setup is used by the system for matching, while the other images are exemplary recordings during verification by the user.

However, the cursor had to be aligned and moved inside small circular regions. Upon successful sampling, the system summary/system decision was presented to the user.

5.2 Apparatus and Data Collection

We conducted the study in a lab with illumination from the ceiling enabled (fluorescent lamps). In order to minimize variations induced by daylight changes, we kept the blinds of the room closed throughout the entire study.

All user interfaces were integrated into a single Android application running on the Samsung Galaxy S5 mobile phone (Android 4.4.2) and using the built-in camera with LED flashlight enabled. Reference data for verification was recorded with our robot using the same device (see Section 3.1).

We used four holograms as shown in Figure 10, each on a different base document. With our choice of samples and reference data, we aimed to address the non-trivial case of hologram substitution, since that is rather common according to a document expert we consulted for our study. Although some of the views we selected (i.e., black patches) may not resemble the typical appearance of holograms for the public, we believe that the large visual difference w.r.t. the other image in the pair justifies their use.

We collected data for evaluation through automatic logging on the test device itself, questionnaires and interviews. For data analysis, we used Matlab, R, and SPSS. Null hypothesis significance test were carried out at a .05 significance level, if not otherwise noted.

5.3 Procedure

Each participant was informed about the study purpose and the approximate length prior to the start of the study. The participants filled out a demographic questionnaire and then conducted the Vandenberg and Kuse mental rotation test [21]. They were informed that they would test a total of 12 holograms with three user interfaces (four holograms per interface). Although 12 holograms were shown to the participants as a stack, only a subset of four holograms was used for all interfaces (see Figure 10).

The following procedure was repeated for all three user interfaces. A training phase with both a correct and a fake document (not appearing in the actual study) was conducted. This also included an explanation of application controls along with document classification and tracking. Participants could test the interface as long as they liked (on average less than five minutes). After feeling comfortable with the interface, participants were asked to use the current interface to capture four holograms, one at a time. After capturing a single hologram, the system presented its decision on the validity of the single views and an overall decision (valid, unsure, invalid). After seeing the system decision, the participants were asked to fill out a post-task questionnaire, in which they were asked to assess the validity of the hologram on their own (5-item bipolar scale: I am totally sure that the hologram is fake ... neutral ... I am totally sure that the hologram is valid). After validating four holograms with the current interface, the users filled out a post-interface questionnaire (5-item Likert scale, ease-of-use and time items of the After Scenario Questionnaire [12]), the NASA TLX questionnaire (with weighting of items) [5], the AttrakDiff [8] and Intrinsic Motivation Inventory questionnaires [13].

After having conducted this procedure for all three interfaces, the participants filled out a final questionnaire, in which they should choose their preferred interface (overall preference, which interface was fastest to use, which interface was easiest to use). Finally, they were asked about the reasons for their choices. Participants received a voucher worth 10 EUR for their time.

The starting order of both interface and hologram was counterbalanced. The tasks where blocked by interface. While each participant was exposed to each hologram three times we took care to make them believe it was a separate hologram (by showing a staple of several documents and hiding from them which document was drawn out of the staple). Also each participant was exposed to individual interface-hologram combinations exactly once during the study. The whole procedure took on average 90 minutes. Participants could take a break anytime they wanted.

5.4 Participants

19 volunteers (2 female, age $M = 26.8, SD = 4.46$) participated in the study. All except one participant owned at least one smartphone or tablet, where the majority (16) had been using it for at least one year. In general, participants reported to be interested in technology. Thirteen participants had already used an AR application at least once. Seven participants had never attempted to verify a hologram before. In the mental rotation test, the majority of participants scored reasonably ($M = 0.8, SD = 0.14$). With 19 participants assessing 4 holograms with 3 interfaces, we obtained 228 samples.

5.5 Hypotheses

Based on our observation and the insights gained during prestudies, we had the following hypotheses: *H1:* The hybrid UI will be the fastest among all interfaces. *H2:* The alignment UI will be the most accurate one, but slow. *H3:* The constrained navigation UI will be the easiest to use.

The hybrid interface combines desirable elements from alignment (accurate end position) and constrained navigation (marked interaction space). With a small number of reference views, checking should be very fast (*H1*). The revised alignment interface should assure the most accurate capture positions and consequently has the best prospects for accurate matching and verification (*H2*). This might come at the cost of increased capture time. The constrained navigation approach gives most freedom to the user. The pie slice layout could be familiar to users, although accuracy w.r.t.

single reference views might not be as good and by design, a bigger space needs to be sampled (*H3*).

5.6 Findings

We performed an analysis of task completion time, user and system performance and user experience aspects for hologram verification.

Task Completion Time For capture time (the time from start of the task until the presentation of system results), a two-way within-subjects analysis of variance showed a significant main effect for interface, $F(2,36) = 3.60$, $p = .038$, *partial* $\eta^2 = .17$ and a significant main effect of hologram, $F(3,54) = 4.04$, $p = .012$, *partial* $\eta^2 = .18$. The interaction between interface and hologram was not significant.

Multiple pairwise post-hoc comparisons with Bonferroni correction for interface revealed that the mean score for capture time (in seconds) for the hybrid interface ($M = 37.22, SD = 38.20$) was significantly different compared to alignment ($M = 57.01, SD = 55.77$) ($t(75) = 3.44, p = .001$), but not compared to constrained navigation ($M = 44.43, SD = 20.70$). Also, there was no significant difference between constrained navigation and alignment.

Multiple pairwise post-hoc comparisons with Bonferroni correction for hologram revealed that the mean score for capture time (in seconds) for hologram 2 ($M = 39.61, SD = 29.39$) was significantly different compared to hologram 4 ($M = 55.19, SD = 53.51$), $t(56) = -3.23, p = .002$, but not compared to hologram 1 ($M = 45.58, SD = 40.97$) or hologram 3. Also there were no other significant differences between holograms. Furthermore, there where no learning effects for either interface or hologram as indicated by a within-subjects analysis of variance.

The decision time (the time spent in the summary screens) over all interfaces was on average 18.45 seconds ($SD = 15.32$). A two-way within-subjects analysis of variance showed no significant main effect for interface but for hologram $F(3,54) = 3.233, p = .029,$ *partial* $\eta^2 = .152$. However, multiple pairwise post-hoc comparisons with Bonferroni correction for hologram did not indicate any significant pairwise differences (hologram 1 $M = 17.7, SD = 12.72$, hologram 2 $M = 23.7, SD = 19.54$, hologram 3 $M = 15.53, SD = 8.85$, hologram 4 $M = 16.98, SD = 17.54$). The interaction between interface and hologram was not significant.

To summarize, the capture time using the hybrid interface was significantly faster than the alignment interface and for hologram 2 compared to hologram 4. For decision time, no pairwise significant differences could be found. There were no learning effects for interface or hologram.

User and System Performance Over all participants and holograms, 79.6% of the users' decisions were correct (treating both items 'I am totally sure that the hologram is [in]valid' and 'I am sure that the hologram is [in]valid' as correct answers). For 12.5% of the decisions, the users where unsure if the hologram was valid or fake. An investigation of the effects of the predictors interface and hologram on the dichotomous dependent variable 'correctness of user decision' using logistic regression was statistically not significant. Note that we had to exclude one participant from this sub-evaluation due to incomplete data.

73.1% of the system decision were correct. The system was unsure if the hologram is valid or fake in 11% of all cases. As for user decision, we used logistic regression to investigate the effects of interface and hologram on the dichotomous dependent variable 'correctness of system decision'. The logistic regression model was statistically significant $X^2(5) = 58.83, p < .0001$, explained 37.5% (Nagelkerke's R^2) of the variance in system decision and correctly classified 81.5% of the cases. The Wald criterion demonstrated that hologram made a significant contribution to prediction (*Wald* $X^2(3) = 20.80, p < .0001$), but interface did not. The system only made correct decisions in 50.0% for hologram 1 (neutral: 27.8%, hologram 2 correct: 100%, 3 correct: 94.4%, 3 neutral: 0.04%), 4 correct: 74.1%, 4 neutral: 13.0%).

To summarize, users were able to correctly validate (decide if the hologram is valid or false) in 80% of the cases, but the system only in 73%. Hologram was a significant predictor for system decision with a validation performance for hologram 1 of only 50.0%.

User Experience We investigated ease of use and satisfaction with task duration with the ASQ, cognitive load with the NASA TLX, and hedonic and motivational aspects with AttrakDiff and Intrinsic Motivation Inventory questionnaires, after each participant had finished using a single interface.

A one-way Friedmann ANOVA by ranks did not indicate a significant effect of interface on ease-of-use. Similarly, for satisfaction with task duration (over all 4 holograms per interface), there was no significant effect of interface. Note that we had to exclude one participant from this sub-evaluation due to missing data.

For cognitive workload as measured by NASA TLX, one-way Friedmann ANOVAs by ranks did not indicate significant effects of interface on the subscales (mental demand, physical demand, temporal demand, performance, effort, frustration) or the overall measure. Due to space reasons and the non-significance of the omnibus tests, we will not report further statistics here.

Similar, for pragmatic quality (PQ), hedonic quality - identity (HQI) and hedonic quality - stimulation (HQS), as measured by AttrakDiff, and for value-usefulness and interest-enjoyment as measured by the Intrinsic Motivation Inventory, one-way Friedmann ANOVAs by ranks did not indicate significant effects of interface.

In the final questionnaire, 47% of the participants indicated that CON was easiest to use (ALI: 21%, HYB: 32%), 42% indicated that CON was fastest to use (ALI: 16 % HYB: 42%) and 47% favored CON overall (ALI: 26.5%, HYB: 26.5%).

In summary, the statistical analysis could not indicate significant effects of the interfaces on usability, workload, hedonic qualities or intrinsic motivation. Still, about half of the participants preferred CON overall and indicated that it was easiest to use.

6 Discussion

Our analysis did not fully confirm hypothesis H1. The hybrid interface was the fastest one, taking roughly 40 s for image capture, being significantly faster than the alignment interface (which took around one minute for verification). However, the hybrid interface was not significantly faster than the constrained navigation interface (ca. 45 s).

While this is a significant improvement over related work ([6], but using up to six views), this is still a long time span and probably not feasible for a quick check in a real-world situation. However, as most checked documents will be originals, an early exit for such samples could further decrease checking time. As decision time did not vary significantly between interfaces, they are all suited to recording data for verification.

Around 73% of the system decisions were correct, which may seem rather low. As there was no significant effect of any interface, hypothesis H2 does not hold in this regard. If we only neglect wrong decisions (i.e., combine positive and neutral decisions), the system performance would still be below the combined rate for user decisions (system: 84% correct vs. user: 92% correct). It seems that users either came up with their own (more invariant) similarity measure during the study, or they used additional appearance information gathered through the sampling process for their decisions, which was not available to our system (e.g., due to non-matching viewing direction). However, most of the neutral system decisions (around 63%) were caused by hologram 1 (50 EUR banknote, see Figure 10). This hologram shows rainbow colors, which is a very difficult case for our matching approach. Together with the rather conservative parametrization of our system (avoiding false positives), and the encouraging results of hologram 4 (around 90%

combined rate), we speculate that the type of hologram has considerable influence on its verifiability with the proposed approach.

While the statistical analysis did not indicate significant effects of interface and user experience measures, we obtained a large number of comments in the post-hoc interviews throughout the study. The HYB UI, being the fastest one, was described four times as 'intuitive', 'good to use' or 'easy' (CON: 7, ALI: 3). However, four participants reported that the movements required were initially not clear (CON: 4, ALI: 5). With the CON UI, four users recognized the freedom in movement. For the slow ALI UI three users expressed their interest in that UI ('interesting', 'cool idea', 'visually best'). One user stated that it was 'easy to spot, what to do, but difficult to accomplish'. Two users also gave positive comments about the usefulness of the summary.

For the CON and HYB interface, one user suggested to always display the pointing rubber band, even when the widget is perfectly at the screen center. For CON, one user suggested an additional completion indicator for pie slices involving the pie region itself instead of the border. The same user also suggested to use additional indicators for viewing ray alignment in the ALI UI.

Despite being the fastest one (around 40 s), the hybrid user interface did not receive the same degree of user consent as the CON interface when taking into account the comments. Users explicitly criticized the final alignment stage involved. As a take away, it seems that the most efficient interface does not necessarily reflect the general preference of the user. Such awareness should be considered for real-world deployments of mobile AR user interfaces requiring fine-grained maneuvering.

7 Conclusion

We designed, implemented and evaluated several different user interfaces for checking holograms within a mobile AR framework for document inspection with the goal to considerably reduce checking time for the user. Alignment, constrained navigation and hybrid approaches using automatic matching were compared in a user study. Although the hybrid interface had the fastest completion time, users preferred the constrained navigation interface over the other two according to the comments received. As each of the interfaces served equally in the capture of verification data, the choice of UI (constrained navigation or hybrid) might depend on user preference or previous training. Still, verification performance of the system should be improved. Besides parameter tuning, another approach would be to use more reference data for verification or to neglect the relationship between appearance and recording position during matching. However, this would decrease the security of the current verification approach, since elements having the same appearances at different viewing positions could not be differentiated anymore.

For future work, we plan to improve document tracking in order to increase overall robustness and usability. As the hologram type seems to be a critical factor for robust and efficient checking, we also want to further analyze our pipeline regarding different types of holograms and measure their effect on checking time and system decisions. We have also started experimenting with embedding the verification interface in a mobile game which makes sure through game objectives that interesting space gets covered. This might be particularly useful in educational and training settings.

Acknowledgments

This work is supported by Bundesdruckerei GmbH.

References

[1] S. Bae, A. Agarwala, and F. Durand. Computational rephotography. *ACM Trans. Graph.*, 29(3):24:1–24:15, July 2010.

[2] F. Biocca, A. Tang, C. Owen, and F. Xiao. Attention funnel: Omnidirectional 3d cursor for mobile augmented reality platforms. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1115–1122, 2006.

[3] K. Chintamani, A. Cao, R. Ellis, and A. Pandya. Improved telemanipulator navigation during display-control misalignments using augmented reality cues. *Systems, Man and Cybernetics, Part A: Systems and Humans*, 40(1):29–39, 2010.

[4] M.-P. Dubuisson and A. Jain. A modified hausdorff distance for object matching. In *Pattern Recognition, 1994. Vol. 1 - Conference A: Computer Vision amp; Image Processing., Proceedings of the 12th IAPR International Conference on*, volume 1, pages 566–568 vol.1, Oct 1994.

[5] S. G. Hart and L. E. Staveland. *Human Mental Workload*, chapter Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. North Holland Press, Amsterdam, 1988.

[6] A. Hartl, J. Grubert, D. Schmalstieg, and G. Reitmayr. Mobile interactive hologram verification. In *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 75–82, 2013.

[7] A. Hartl, D. Schmalstieg, and G. Reitmayr. Client-side mobile visual search. In *Proceedings of the 9th International Conference on Computer Vision Theory and Applications*, pages 125–132, 2014.

[8] M. Hassenzahl, M. Burmester, and F. Koller. AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In *Mensch & Computer 2003: Interaktion in Bewegung*, pages 187–196, Stuttgart, Germany, 2003. B. G. Teubner.

[9] S. Heger, F. Portheine, J. A. K. Ohnsorge, E. Schkommodau, and K. Radermacher. User-interactive registration of bone with a-mode ultrasound. *Engineering in Medicine and Biology Magazine, IEEE*, 24(2):85–95, 2005.

[10] J. Janucki and J. Owsik. A wiener filter based correlation method intended to evaluate effectiveness of holographic security devices. *Optics Communications*, 218(4-6):221–228, 2003.

[11] H.-J. Kwon and T.-H. Park. An automatic inspection system for hologram with multiple patterns. In *SICE, 2007 Annual Conference*, pages 2663–2666, 2007.

[12] J. R. Lewis. Psychometric evaluation of an after-scenario questionnaire for computer usability studies: The asq. *SIGCHI Bull.*, 23(1):78–81, Jan. 1991.

[13] E. McAuley, T. Duncan, and V. V. Tammen. Psychometric properties of the intrinsic motivation inventory in a competitive sport setting: A confirmatory factor analysis. *Research quarterly for exercise and sport*, 60(1):48–58, 1989.

[14] T.-H. Park and H.-J. Kwon. Vision inspection system for holograms with mixed patterns. In *Automation Science and Engineering (CASE), 2010 IEEE Conference on*, pages 563–567, 2010.

[15] A. Pramila, A. Keskinarkaus, E. Rahtu, and T. Seppnen. Watermark recovery from a dual layer hologram with a digital camera. In A. Heyden and F. Kahl, editors, *SCIA*, Lecture Notes in Computer Science, pages 146–155. Springer, 2011.

[16] J. Shingu, E. Rieffel, D. Kimber, J. Vaughan, P. Qvarfordt, and K. Tuite. Camera pose navigation using augmented reality. In *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 271–272, 2010.

[17] N. Snavely, S. M. Seitz, and R. Szeliski. Photo tourism: Exploring photo collections in 3d. *ACM Trans. Graph.*, 25(3):835–846, July 2006.

[18] M. Sukan, C. Elvezio, O. Oda, S. Feiner, and B. Tversky. Parafrustum: Visualization techniques for guiding a user to a constrained set of viewing positions and orientations. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, pages 331–340, 2014.

[19] S.-Y. Sun, M. W. Gilbertson, and B. W. Anthony. Computer-guided ultrasound probe realignment by optical tracking. In *Biomedical Imaging (ISBI), IEEE International Symposium on*, pages 21–24, 2013.

[20] R. L. van Renesse. *Optical Document Security*. Artech House, third edition, 2005.

[21] S. G. Vandenberg and A. R. Kuse. Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, 47(2):599–604, Oct. 1978.

[22] D. Wagner, G. Reitmayr, A. Mulloni, T. Drummond, and D. Schmalstieg. Real-time detection and tracking for augmented reality on mobile phones. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 16(3):355–368, 2010.