cAR: Contact Augmented Reality with

Transparent-Display Mobile Devices

1st Author

1st author's affiliation  
1st author's E-mail address

2nd Author

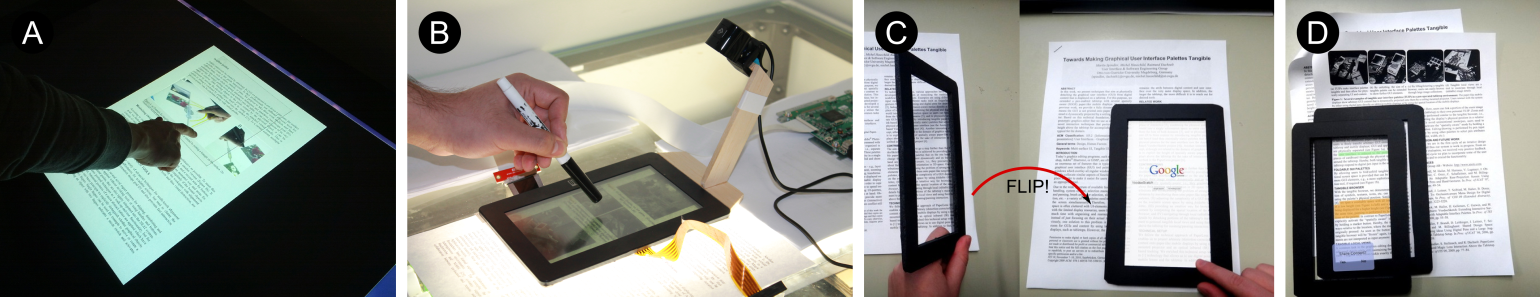


Figure 1: Transparent portable devices implementing Contact Augmented Reality for active reading tasks: A- Tabletop prototype. B- tPad prototype. C- Flipping to online-search for selected content (mock-up). D- Stacking for content sharing (mock-up).

2nd author's affiliation  
2nd E-mail

3rd Author

3rd author's affiliation  
3rd E-mail

**ABSTRACT**

We present Contact Augmented Reality (cAR), an approach to augmented reality where a transparent display mobile device rests on top of the augmented object. By following an iterative and user-centric design approach we identified interaction techniques for cAR devices and organized them in three categories: contact-based, content-based and off-contact. We built two prototype cAR devices and explore their usage for the sample domain of active reading. A first low-level touch-enabled prototype uses a tabletop and transparent tangibles and allowed us to iteratively design and test interaction techniques. The second and higher-level prototype, called a tPad, is a transparent 7” touch LCD display which can be placed on top of a back-lit paper document. The tPad uses an external camera to identify the document, and determines its location and orientation via feature tracking. We report on user feedback and elaborate on the technical challenges for this new class of contact-augmented reality devices.

**Categories and Subject Descriptors**

H.5.2 Information Interfaces and Presentation: User Interfaces: Input Devices and Strategies, Interaction Styles.

**General Terms**

Design, Experimentation, Human Factors.

**Keywords**

Contact Augmented Reality, Transparent Devices, Tablet Augmented Reality, Active Reading.

# INTRODUCTION

Mobile Augmented Reality (AR) overlaps virtual content on images of the real-world captured using a built-in camera. New trends in transparent display technologies allow users to view both virtual content and physical objects simultaneously. However, most conceptual demonstrations of this new technology [8, 9, 16, 27] appear to be a direct evolution of mobile AR.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

*Conference’10*, Month 1–2, 2010, City, State, Country.

Copyright 2010 ACM 1-58113-000-0/00/0010 …$15.00.

We propose Contact Augmented Reality (cAR), a relatively unexplored and novel form of interaction with mobile transparent devices. cAR renders virtual content over a physical artefact placed directly underneath and in direct contact with the display, such as a map, a form or a book. This direct contact provides spatial alignment between the display and object at a very short distance, mitigating the *registration* and *rendering* complexities of traditional AR displays [5]. In cAR, *registration* is reduced to identifying the current object and finding the relative 2D location and orientation of the cAR device on top. With cAR *rendering* no longer requires perspective corrections.

cAR brings to users the benefits of interacting with a digital medium without losing the affordances of the physical artifact such as when planning a trip or for active reading [1, 2]. A user browsing a physical foldout of a map can place a cAR-enable device on top of it to highlight points of interest, draw routes and make notes directly on the device, without affecting the paper map. The cAR-enabled device can be lifted-off so that the user can continue browsing the physical map, flipping parts of it and checking side legends without loss of context. The user can further re-examine virtual content, such as to videos or images associated with a specific region of interest on the map, by resting the device back on the map. This retrieves and displays all the content previously created as the user reexamines previously visited and annotated regions.

In this paper we introduce cAR, identify interaction techniques for cAR devices and show how such techniques can be applied for an application area, Active Reading [1], leveraging the affordances of paper [19] and the benefits of digital system [14]. We built two cAR prototypes. The first prototype is a tabletop-based emulation using transparent tangibles (see Figure 3), enabling the rapid prototyping and testing of the interaction techniques. The second prototype is a 7” transparent LCD mobile device called the tPad (see Figure 4), allowing us to address rendering and registration.

Our contributions are at the conceptual, interaction design and implementation levels. First, we present Contact Augmented Reality and differentiate it from existing AR approaches. Second, we identify a series of interaction techniques for cAR devices. Finally, we present two cAR prototypes detailing their software and hardware, and show how cAR can be applied to active reading, a task that benefits from the affordances of a physical artifact and its associated digital content.

# CONTACT AUGMENTED REALITY

Our conceptualization and implementation of cAR is guided by the vision that printed material is information rich. Wall posters, newspapers, book pages, are all associated with far more content, and in much diverse formats (multimedia), than is possible to etch in ink. With cAR, such associated content can be retrieved by simply placing a transparent-display mobile device directly on top of the object, be it a poster, map, or newspaper. This vision acknowledges the affordances of physical objects; in the case of paper documents the naturalness and convenience of reading, scribbling, folding or manipulating printed paper is unmatched by its digital counterparts. While existing mobile devices already offer access to digital information by means of mobile AR, cAR is based on the notion that interactions with the digital are enriched by the tangibility of physical moving a device on and off the augmented object.

We identify three categories of interaction techniques for cAR devices: First, contact-based interactions, e.g. placing the device on a newspaper could retrieve additional data about that object such as audio or video; second, content-aware interactions, e.g. tapping on unknown words triggers translation; and third, off-contact interactions, e.g. information between devices can be easily exchanged by stacking one on top of another. Other interactions within these categories include content extraction, scribble triggers, orientation to content, and flipping for dual-sided access.

For cAR to operate, the fundamental requirement is to establish a frame of reference between the device and the object to augment (a coordinate system). cAR interactions require knowing the 2D location of the device relative to the origin of the object’s coordinate system. This implies that, for example, using a cAR device on a book while in bed or while sitting on a table makes no difference when determining its relative location. An important consequence of the spatial alignment between the transparent display of the cAR device and the augmented object is that digital content is rendered on a virtual plane parallel to the object surface; this means that no homographic transformations are needed for aligning the digital augmentation.

We envision cAR devices that augment only when in contact with the target artefact, acting as normal devices when they are not resting on top of objects. Therefore, augmenting an object does not require the user to hold the cAR device in front or above the object for an extended period of time. This mitigates the potentially adverse effects on physical demand [16]. Another implication is that augmentation is triggered by means of implicit interaction; simply placing the cAR device on top of the target object. Once the object is identified (registration), the device launches the application associated with it. A final implication is that a cAR device can augment an object without being the center of attention all the time; acting at times as an ambient or a secondary display, at others allowing undisturbed view to the underlying object due to its transparency.

In summary cAR integrates virtual and physical worlds by:

* augmenting physical objects upon contact,
* preserving the affordances of physical objects,
* integrating display and input functionalities, and
* simplifying registration and rendering to two dimensions.

# RELATED WORK

cAR builds on work in augmented reality, magic lenses and transparent portable devices.

## Augmented Reality

*Augmented Reality* (AR) enhances the real world by embedding digital content onto it. Bimber and Raskar list three basic AR challenges: display technology, registration and rendering [5]. The display technology determines the complexity of the other two. *Traditional AR* relies on mobile displays carried by the users (e.g. smartphones, pico-projectors, HMDs) allowing the augmentation of any object within the display’s field-of-view but requiring complex operations for registration (i.e. 3D object recognition) and rendering (i.e. field-of-view and perspective calculations). Moreover, mobile displays present limitations in terms of resolution, focus, lighting and comfort. A complete reference to AR technologies and applications can be found here [4, 7].

On the other hand, *Spatial AR* (SAR) relies on displays fixed in the environment (e.g. projections, transparent LCDs, etc) [6]. Knowing the exact location of the display and the augmented object provides SAR applications with *spatial alignment*, a linear correspondence between virtual content and real world objects. Spatial alignment facilitates the creation of AR applications because as simpler registration and rendering operations are required. To preserve the spatial alignment, both the display and augmented object should remain spatially fixed, limiting SAR to non-mobile applications [5].

Contact Augmented Reality (cAR) incorporates elements from both traditional and spatial AR. From traditional AR it maintains the vision of a mobile device that augments any object and is carried around by the user. From spatial AR it incorporates the property of spatial alignment, thus the knowledge about the location of both display and object and their correspondence. In brief, cAR is both mobile and spatially aligned.

## Magic Lenses and Tangible Views

cAR is inspired by Bier et al.’s Toolglass and Magic Lenses [5]. Toolglass and magic lenses for WIMP interfaces sit between the application and the cursor to provide rich operations and visual filters on the digital content. For example, a toolglass widget can have different areas each with unique operations, such that by clicking the target object through the toolglass the digital content is modified in different ways. Similarly, the magic lens widget can hide or show details of an underlying digital object by placing the widget on top of it. Moving beyond WIMP, Mackay et al.’s implemented a tool-glass and magic lens approach to augmenting a biology laboratory book [22]. Researchers built physical magic lenses using transparent acrylic plates with fiducial markers [20] and 3D head tracking [29]. Similarly, tangible views [30] provide secondary displays for tabletop computers.

cAR, a concept developed for transparent mobiles, encapsulates ideas from toolglasses and magic lenses. With cAR, the physical object is visible and modifications happen on its digital model. cAR advances the concept introduced by the A-book in several ways: we use feature tracking registration, explore off-contact and content-aware interactions, and rely on transparent displays, none of which have been explored before.

## Transparent Handheld Devices

Transparent handheld devices are the subject of popular design concepts covering smartphones and tablets [8, 9]. Such concepts are instrumental in proposing novel interactions (some of them similar to the ones we explore); however they do not discuss usage contexts and technical limitations. Nonetheless, a few devices are reaching the public (e.g. Lenovo S800) and production of such displays is scheduled for the near future (see Futaba Corporation [[link](http://www.oled-info.com/futabas-oled-road-map-amoleds-2014-transparent-and-flexible-oleds-cars-2015)], Fujitsu [[link](http://www.fujitsu.com/be/Images/Workplace_of_the_Future.pdf)]). While such devices will become commercially available, we possess limited understanding of the breadth of interaction techniques they afford. One explored aspect is their support for touch interaction on the back of the device. LucidTouch [30] and LimpiDual [25] studied back-the-device touch to overcome the fat-finger and finger occlusion problems. Lee et al. [21] studied the binocular parallax problem.

Our work moves beyond such problem-oriented focci to include a broad a range of techniques for augmenting objects upon contact.

# cAR Interaction Techniques

We present cAR interaction techniques in three categories: *contact-based,* *off-contact* and *content-aware*. Contact-based and off-contact interactions are what distinguish cAR from other forms of augmented reality. Off-contact techniques such as flipping or dual-side interactions bear resemblance to those from other areas, such as tangible views [31], however, we have included them for the sake of completeness. Figure 2 illustrates the basic interaction techniques for cAR devices.

## Contact-based

Contact-based interactions are what primarily distinguish cAR from other forms of AR approaches. These techniques are manipulations of the cAR devic in relation to the object upon which it rests.

*Placing/Removing* – The basic interaction for cAR devices is placing the device on top of the augmentable object. Upon contact, the device tries to identify the object below and respond to it. In simple cases the device can respond to simple properties like color or type (e.g. text, drawing, skin, paint, etc.). In complex cases, such as for a map, the device needs access to a model of the underlying object (e.g. in the cloud). Conversely, removing the cAR device from the object can be interpreted as an implicit interaction technique.

*Translation* – A user can translate a cAR device horizontally (x-y directions). The device can interpret movement in each axis as input. An application can use translation in X and Y to translate a virtual layer accordingly and maintain a correspondence between it and the real object.

*Rotation* – The device can also interpret the rotation of the device as input. This rotation can either be relative to the original placement or to the object’s “north”. An application can use changes in rotation to, for example, change display settings like opacity and zoom factor.

*Freezing* – The device might freeze its current state by ignoring changes in translation and rotation. By freezing the current state, the user can move the device freely around while preserving the current state of the application, e.g. the current view in the virtual plane. This could for example be a video being displayed in the center of the device regardless of its position (see Figure 3B).

*Shaking* – Shaking the device while is still lying on the augmented object. A shaking gesture can be understood as a command to the cAR application to modify or remove the displayed content. Shaking can be combined with translation and rotation.

## Off-Contact

Off-contact interaction techniques do not require the cAR device to rest on top of the augmented object. These techniques are unique to the transparency affodances of cAR devices.

*Flipping* – A cAR device can be flipped around and have the other side of the screen on top. The device can apply different modifications on the user interface upon flipping like, for example, a 2x zoom, an inverse color filter, a translate feature, or launching a secondary application for the actual document.

*Stacking* – A cAR device can also be stacked on top of another one. Given that both displays are transparent, the digital content of both devices could be visible at once. A natural usage of stacking is content sharing, where digital content created in one device can be pulled up by the device on top or pushed to the device below.

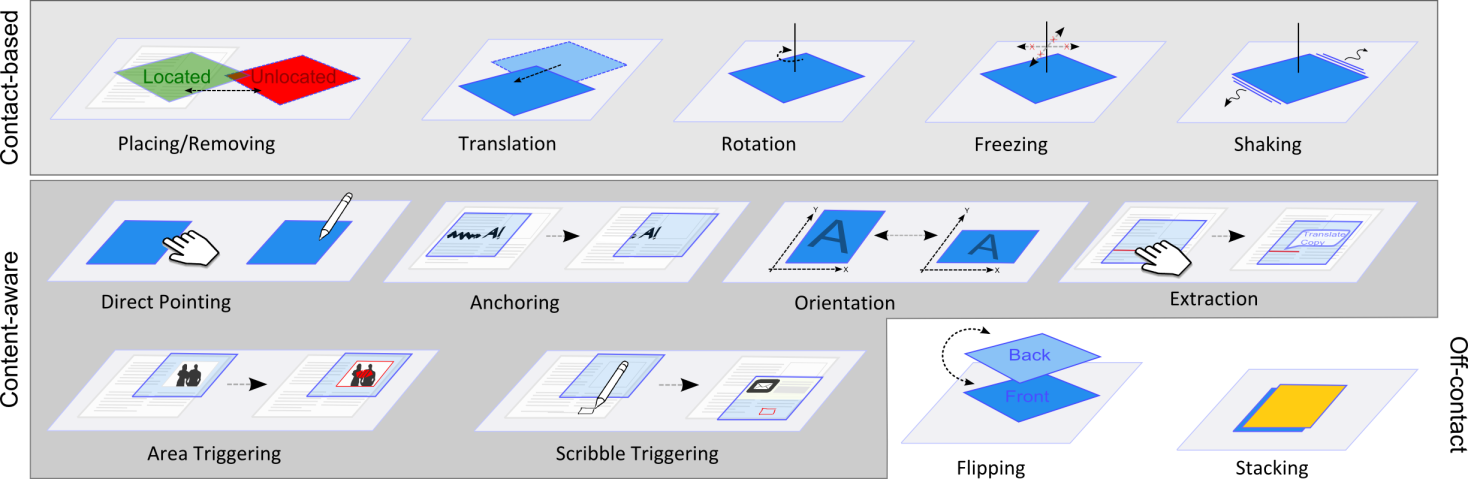
## Content-Aware

Content-aware interaction techniques leverage knowledge about the underlying physical object.

*Direct Pointing* – Direct pointing allows the user to interact with the content displayed on the cAR display either by touch or pen. By means of direct pointing users can interact with user-interface elements like menus or issue touch gestures. Users can also interact with spatially-aligned digital content visible at the time.

*Extraction* – Users can interact with elements of the digital representation of the augmented object. For example, a cAR magazine application allows users to select and look up word defi-nitions and translations, and their occurrences in the document.

Figure 2. cAR interaction techniques grouped by the three identified categories



*Triggers* – Triggers are areas in the physical object that activate special responses by the cAR device when placed on top of them. Triggers can be area-based or scribble-based. Area-based triggers are special zones statically defined in the document being augmented, such as an image on a newspaper used for triggering a video associated with the article. Scribble-based triggers are physical modifications on the physical object which are read and interpreted by the device; e.g. a hand-drawn square launches the browser application by moving the cAR device on top of it.

*Anchoring* – Anchoring is the basic interaction to attach digital content to a particular location of the physical object. Anchored content exists on a virtual layer spatially aligned with the physical object. For example, scribbles are created anchored to specific locations of a paper document.

*Orientation* – A cAR application can adjust the orientation of its user-interface based on the coordinate system of the augmented object. This technique resembles the change of orientation of mobile phones according to the way users hold them (portrait vs. landscape). For example, a cAR application’s user-interface can adjust to the direction of the text and orientation of the paper document.

# Sample Application: Mapping cAR to Active Reading

To demonstrate the novel interactions provided by cAR, we chose to apply our implementation to Active Reading [ref.]. Active reading happens in tasks where people read to gain knowledge and understand all aspects of a specific topic like reading to self-inform, reading to cross-reference or reading to support discussion [2]. Although our goal is not to create a cAR system for active reading that outperforms existing approaches computers [21, 26, 23, 17, 33], we use this application area to explore the range of cAR techniques. Some basic features for active reading include: outlining, underlining, highlighting, searching, scribbling, digital annotations, note-taking, non-sequential navigation, information seeking, quoting, comparing, and content sharing.

Figure 1 (front page) shows prototypes and sketches with different active reading features and the supporting interaction techniques. For example, users can add hand-written notes using a stylus (Figure 1B), access a color version of an image by tapping on the “Figure X” text on the document (Figure 1A), perform an online search on a selected word by simply flipping the device (Figure 1C), and share content by stacking one device on top of the other (Figure 1D). Table 1 shows the complete set of mappings between interaction techniques, and active reading features.

## Implementation Requirements

To implement the proposed interaction techniques and build a cAR application, an implementation should fulfill the following technical requirements:

**RQ1** – Registration: a cAR device should identify the object upon which it rests, and determine its position and orientation in relation to such object. For example, document id, page number, x-y position and rotation.

**RQ2** – Touch and Pen Input: a cAR device should allow users to interact with digital content via touch (for coarse interactions like pushing buttons or accessing content) and pen (for fine interactions like scribbling text).

**RQ3** – Dual Side Interaction: a cAR device can flip over, thus interaction should happen on either side of the screen.

**RQ4** – Device Integration: cAR devices should be able to identify one another when stacked, and seamlessly integrate for content sharing purposes.

**RQ5** – Object Model: An object model for the augmented object should exist, containing meta-data for the cAR device to interpret its current location (frame of reference, orientation, and contents).

# Tabletop Prototype

We built a tabletop prototype to support the design process by allowing for faster prototyping and testing of design alternatives without the technological limitations of a high-fidelity prototype.

Conceptually, the physical documents (books or sheets of paper) are substituted with the interactive surface of the tabletop which also provides touch input capabilities (RQ2). The cAR device itself is simulated by a transparent tangible that is spatially tracked on the tabletop via fiducial markers (RQ1). Different markers on both sides enable flipping (RQ3). A document viewer shows the augmented document, aligns the created content to the actual page, and extracts the words users tap on for further interaction (RQ5). The UI for the simulated device is shown at the location and orientation of the tangible, giving the impression of a translucent display that can be moved freely on top of a document.

## Implementation

We implemented the prototype on a Samsung SUR40 tabletop. We attached Microsoft ByteTags to a 7” acrylic glass mockup device using IR reflective foil to minimize their obtrusiveness.

|  |  |  |
| --- | --- | --- |
| **Int. Technique** | **Active Reading Feature** | **Prot** |
| Placing/ Removal | Document recognition, reading | Both |
| Translation/ Rotation | Browsing virtual content anchored to locations in the document. | Both |
| Freezing | Ignores translation and rotation, thus maintaining the current digital view. | tPad |
| Shaking | Undo for highlights and scribbles | tPad |
| Direct Pointing (hand and pen) | UI interaction, creating and manipulating digital contents | Both |
| Anchoring | Adds notes and scribbles to fixed locations of the physical document. | Both |
| Orientation | Adjust the device’s UI to the text orientation | tPad |
| Extraction | Selecting words from the text for the purpose of in-document search, online search, and translation. | Both |
| Area/Scribble Triggers | Starting a video when hovering an image, and launching app when hovering a particular glyph. | tPad |
| Flipping | Full screen online-search of selected word, and magic-lens color filter. | TT |
| Stacking | Content sharing from bottom device to top device. | tPad |

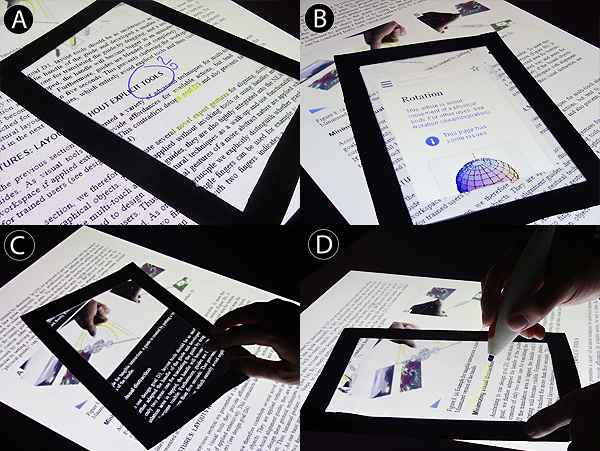
Table 1. Mapping between cAR interaction techniques, active reading features and the prototypes. TT: Tabletop prototype.

Figure 3 shows this tabletop prototype. The prototype supports touch and pen input with an IR pen (Figure 3C-D). We implemented the following features: the user can write free-hand annotations or highlight text (Figure 3A), tap on figures to show an overlay with additional information, e.g., a video, or tap on references to show the corresponding bibliographic entry. Additionally, flipping the display after selecting a word, switches to a browser view showing an online encyclopedia’s entry for the word (Figure 3B). If nothing is selected, a color-inverted view is shown, illustrating how content can be presented in different ways (Figure 3C).

### Limitations

While this prototype is well suited for rapid prototyping it is also limited. Its lacks a display on the transparent device when removed from the tabletop and cannot accurately simulate the haptics of working with real paper or other physical objects. Finally, this prototype does not support device integration (*stacking* - RQ4). However, the benefits of using such a quick form of feedback mechanism further confirmed and aided in improving the value of cAR techniques.

Figure 4. A cAR tabletop prototype: A) highlights and scribbles, B) online-search of selected content, C) inversion lens when flipping without selection, and D) pen input.



## User Feedback

We used the prototype to gather early user feedback. Eight users (6 male, 2 female, on average 28 years old) participated. After introducing participants to active reading and the background of the project they had the chance to try out the system. The session consisted of a semi-structured interview of about 30 minutes during which the participants were able to explore the prototype, guided by the interviewer.

### cAR and Interaction Techniques

In general users appreciated the cAR concept and its applicability to active reading situations. Particularly, users highlighted the value of getting access to information not already included in the text (e.g. video or color images) as well as the benefits of having highlights and annotations in digital format for later use. Some users indicated it would be better suited for books (rather than for short documents) and for situations where a table is available to limit user fatigue from holding the device against a poster wall, for example.

Users easily grasped the value of the *translation*, *rotation*, *direct pointing*, and *anchoring* interaction techniques, and their effects on the contents on the device (e.g. menus) and on the virtual layer (e.g. scribbles and notes). Similarly, they appreciated the possibilities offered by content *extraction*, and suggested other usages like translation and social media sharing. On the other hand, *flipping* received mixed reactions. It was perceived to be laborious by six of the participants. This may have been influenced by the form factor and limitations of the prototype. However, it further emphasizes the need for careful mapping between interaction techniques and application tasks. Thus, complicated gestures should be used only for complex mode switches or rare tasks.

### Active Reading Support

Feedback was mixed for both highlighting and annotating: For three participants, these were the most important features of the prototype, while the others did not see a clear advantage of combining digital annotations with physical, printed documents or preferred techniques similar to PDF readers like word-based marking and comment boxes instead of free-hand marking and scribbling. These opinions might have differed if our tests included maps, as such documents are often marked heavily when used [ref.]. Two participants mentioned the importance of keeping track of where annotations are, highlighting the need for overviews of the content or off-screen markers. Finally, users frequently mentioned the possibility to export such annotations and extracted content, ranging from simple clipboard functionality to integration of some form of social network for sharing comments about specific parts of a document.

Six users especially liked the concept of linking text and pictures in the physical document to additional media (e.g., videos) or metadata (e.g., reference list entries). Users proposed translating individual words or looking up terms in an online encyclopedia, even before they were shown this feature, thinking of it as “convenient” and “quite cool”. Hence, this feature should be considered essential. Moreover, completely replacing content was also well received. Four participants mentioned zooming text for reading assistance as a useful feature, half of the participants also proposed automatic translation of the text under the device. However, this contradicts the notion of see-through augmentation.

### Other Feedback

Participants also mentioned the offset and a possible loss of context when working through the device due to its thickness and border frame. Finally, several users mentioned the flexibility of paper as a problem when combined with the inflexible prototype. This is aggravated when the user has to hold the device and a paper stack in the same hand. Working on a desk or other surface, like three users suggested, would eliminate this problem.

# cAR-Pad Prototype

Our second prototype, the *cAR-Pad*, is a high-fidelity prototype which helped us to further explore the proposed interaction techniques and the technical challenges with building a self-contained cAR device.

Our prototype uses a LCD display on top of a light table and the documents to augment are printed single sided on white paper. The light table acted as back light necessary for our LCD-based prototype. Future transparent devices, with OLEDs would not require such a setup. Currently, we use a camera attached to the display for registration (RQ1). A touch-overlay allows the display to receive touch and pen input (RQ2), while an accelerometer and magnetic sensors enable flipping (RQ3) and device integration (RQ4) respectively. The cAR-Pad runtime holds a PDF version of the document with added meta-data as object model (RQ5).

Figure 4 shows the cAR-Pad prototype (top) and the runtime system components. The device runtime is designed as an application container with an application launcher (DashboardApp), a general purpose application (CalculatorApp), and a cAR application for active reading (ActiveReader). On startup, the DashboardApp lists the installed applications for the user to select.

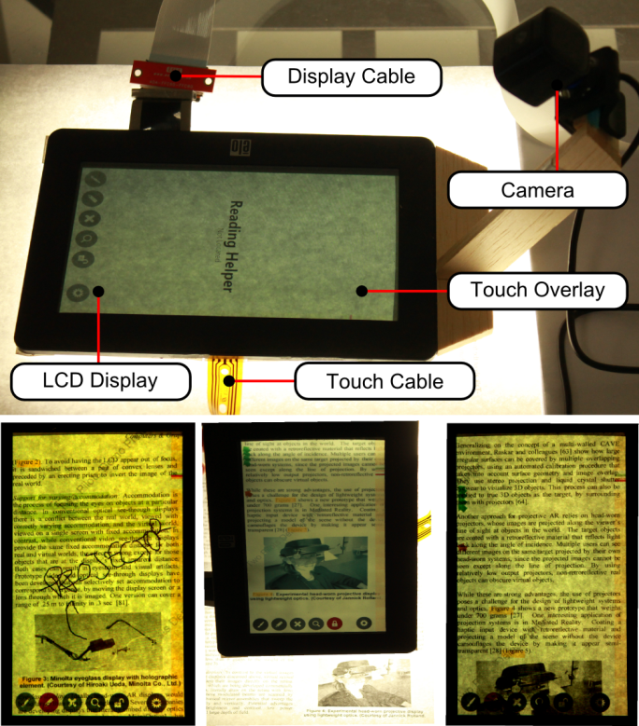
The ActiveReader supports all the features listed in Table 1, except flipping due to the protrusion of the camera. The tPad includes a soft-keyboard to support text entry, and uses *rotation* to control opacity and zooming levels. When presented with the configuration screen, the user can rotate the cAR-Pad to control the transparency of the virtual layer. Users can also zoom into the virtual layer to “create” more space for scribbles. The device introduces off-screen markers to indicate the location of anchored digital content.

We explored two novel and simple interaction techniques explored, *orientation* and *freezing*. For orientation the ActiveReader reorganizes its menus according to the text flow, so that the menus are away from the main reading and interaction area thus avoiding obstruction. This strategy also reduces the presence of fingers and stylus in the image captured for registration. Moreover, users can freeze the cAR-Pad on a particular location and the visible digital content will remain fixed regardless of the device’s movements; a user could then move to a different page or pass the device to another person while having the digital content visible at that particular location.

The cAR-Pad also introduces two complex interaction techniques: *stacking* and *triggers.* We implemented stacking via magnetic switches and magnets embedded on the device’s frame. When physically stacked, the magnets of the device on top align with the magnetic sensors of the device below, which starts a networked pairing process. Upon pairing, users can see both the physical document and the digital content of both displays. At this point, the device on top can “pull content” from the device below according to three strategies: *pull all*, *pull current page*, and *pull selection*. *Pull all* transfers all annotations, scribbles and highlights created in the device below for all pages for the current document. Likewise, *pull current page* limits the transfer to the current page. When *pull selection* is chosen the user manually picks the particular highlighted objects to transfer. Stacking finishes either by explicitly selecting the un-pair button, or by physically separating the devices.

Our implementation supports both model and user-defined triggers. The device uses scribble-based triggers to launch specific apps; in this implementation a hand-drawn square launches the CalculatorApp when the cAR-Pad is placed over the square. We use the calculator simply as a demonstration of this capability. User-defined triggers are device-wide and independent of the currently running application. To showcase model-defined triggers, we created meta-data for a sample document such that a video is automatically played when the cAR-Pad is placed on top of a particular image.

Figure 3. Top – tPad system components. Bottom – tPad at runtime: scribbles, lock-in function and off-screen markers.



## Implementation

### Hardware and Software Architecture

Figure 4-top shows the hardware components used for the cAR-Pad prototype. We re-purposed a 7’’ inch LCD resistive-touch USB display by removing the backlight and extending its display and touch bus cables. The LCD rests on top of the physical paper (one sheet of paper at a time) which in turn rests on a custom-built D65 light table (glass table with fluorescent lights underneath). Display and touch overlay are connected to the original display controller board. We added a Microsoft LifeCam 6000, and an Arduino Pro Micro controller board at 5V with 4 reed-switches and a ADXL335 3-axis accelerometer. The display controller board, camera and Arduino are all connected to a computer running Windows 7. We use C# and Microsoft WPF for authoring and rendering, and C++ and OpenCV 2.4.3 for image processing and feature matching. Network messages for content sharing are JSON-encoded and sent via UDP in the local network. The ActiveReader uses the TallComponents PDF kit ([link](https://www.tallcomponents.com/pdfkit4.aspx)) for accessing pixel-level location information of the PDF contents.

### Camera-based 2D registration

To determine location and orientation of the cAR-Pad within the document (a.k.a registration) we use the camera attached to the device to capture the device’s screen and the underlying surface from above and a feature-matching algorithm. The camera captures the cAR-Pad screen and the underlying surface from above, and the registration algorithm processes it against potential known documents. The algorithm detects the position of the captured image within a digital version of the physical document by matching features from the captured image with features from the document. The location of the image (page number, x-y coordinates, and rotation) maps the location of the actual cAR-Pad to the printed version of the document.

The features, also known as keypoints, efficiently describe image patches and should be invariant to rotation, noise and scale. To detect significant keypoints we use FAST (Features from Accelerated Segment Test), a fast and high quality corner detector [26, 27]. Once keypoints are located, we need a robust and compact keypoint descriptor. We use Fast Retina Keypoint (FREAK), a descriptor inspired by the human retina which is generally faster to compute and more robust to scale, rotation and noise than conventional keypoint descriptors like SIFT, SURF or BRISK [3]. The cAR-Pad maintains a database of documents with it can augment. The device stores the digital content, graphical representation, and pre-computed FAST-features for each document.

Since the feature matching algorithm is not invariant to perspective transformation, for each captured image a correction of the perspective distortion is applied by using a static warping matrix prior to the feature calculation. If the algorithm finds the corrected captured image within the document, a transformation matrix between both is computed and used to derive location and orientation of the device.

Figure 5 shows the results of our registration algorithm tested at the center of the 10 pages of our sample document [4] at 9 different angles and under three conditions: without screen, with screen (no digital content shown) and partly occluded (with a finger pointing to the middle of the display). The initial results show that our implementation works efficiently for most angles, with performance decreasing as the quality of the image decreases (with screen and partially occluded conditions) particularly for the angles between 45 and 90 degrees. These results demonstrate the feasibility of a self-contained camera-based cAR device, noting that further investigations are necessary in different light conditions and with other technologies like transparent OLED displays.

### Technical Limitations

Even though hand-held and mobile, our cAR-Pad is limited by the light table it needs in order to be a see-through display. Nonetheless, both hardware and software architecture were designed for a self-contained device, meaning it all could work with minimal modification when transparent OLED displays become available. On the other hand, registration works at only 10 FPS approximately. The prominence of the camera keeps us from exploring flipping, and touch and pen input are limited to a single side (the cAR-Pad does not implement RQ3). Finally, the nature of LCD displays kept us from really exploring stacking as the display on-top did not receive enough backlight.

# Discussion

This section presents the main challenges we identified for designing and implementing cAR devices.

## Object Model

A model of the object being augmented (RQ5) is a fundamental piece for cAR because it supports multiple interaction techniques (e.g. anchoring, orientation, extraction, triggers) and compelling application features (e.g. search, video playback, etc). The question remains on how such model should be created and potentially distributed to a cAR device. In the case of document-based applications like active reading such a model could be made available by, for example, the publisher of the physical document, either as a self-contained cAR device application or as a file formatted for a general purpose reader. In any case the basic contents, meta-data, and media files should be bundled by a third party and distributed to the device. We envision a scenario where the device, upon touching a document for the first time, tries to locate itself within a list of known documents, in a local or cloud.

In another approach, the cAR device attempts to create the model of the document based on the images captured by its camera. A basic image-stitching algorithm could provide the basic frame of reference for *contact-based* interaction techniques, *direct pointing* and *anchoring*. More elaborate techniques like *orientation* and *extraction* would need user assistance or some degree of artificial intelligence (text flow detection, extract text from images, etc).

## Camera-based Registration and Optimizations

A major aspect of our implementation was dedicated to the camera-based registration algorithm (RQ1), allowing us to identify its flaws and propose several optimizations.

The main flaw we found in this approach is its impact on the device itself. Given that the camera needs to capture a good portion of the screen, it needs to be elevated from the display plane. Our feature matching algorithm is sensible to lighting conditions and device orientation, and dependent on the number of observable features. To ameliorate the hardware requirements the device could use wide-angle lenses (with the added image distortion that needs to be compensated), multiple smaller cameras distributed around the display, or pixel-level capture similar to Microsoft’s PixelSense technology for tabletops.

The other limitations can be handled through software optimizations. For example, registration could use a multi-level approach to reduce computational load: if the current page is already known, the device tries to detect its new location by searching only on the actual page. In case of a failure a search is performed on the adjacent pages, and ultimately in the complete document. However, when the document is large enough (e.g. a 700 pages chemistry book) such search could take several seconds. In this case, the cAR device could calculate the camera features locally and delegate the search to an online service. For better energy efficiency, a cAR device could only call the registration algorithm when a movement is detected through the accelerometer or upon request by pressing a physical button.

## Color Mixing

A major usability obstacle when using the cAR-Pad prototype was the interference of colors between the paper document and the content on the LCD display, affecting the legibility of the digital content. This phenomenon has been studied to some extent [13, 31], and also occurs in transparent OLED displays. Handling color mixing is an important issue for consumer-ready cAR devices and we propose three potential solutions. The simplest solution tries to re-locate UI elements to areas where color mixing is minimal (similar to [32]). A more elaborate solution relies on concepts of color theory to predict color mixing and compensate the displayed color (similar to [24]). An optimal solution uses a display technology that integrates the advantages of LCD and OLED displays: to show digital content the LCD layer shows a black pixel to block the light from behind, while the OLED layer shows the colored pixel. For transparent areas the LCD shows a white pixel while the OLED shows nothing.

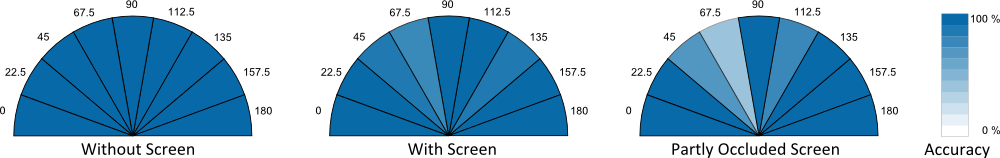
## Device Ecologies

A recurring topic during the design and user-feedback session was the integration of content created in the cAR device with other computational tools including other cAR devices (RQ4). If we consider active reading as part of the creative process, the content from the cAR device should be accessed later for archival or reference. A cAR device should then provide mechanisms for integrating with other computing devices like traditional PCs.

# Conclusions

In this paper we proposed the notion of Contact Augmented Reality, presented a series of interaction techniques for cAR devices, and implemented two cAR prototypes to support active reading. Rather than proposing an optimal tool for active reading support, we aimed at exploring how users would interact with cAR devices in this sample domain and at evaluating the feasibility with existing technologies.

Figure 5. Registration accuracy based on rotated tPad performance. Left: baseline without the tPad screen, Middle: with screen, Right: when placing a finger on top of tPad, giving partial occlusion.



Our first low-fidelity prototype uses a tabletop computer and acrylic tangibles. A semi-structured user feedback session showed that participants understand the cAR concept and the interaction techniques, and appreciated the opportunities offered by *extraction* such as online sharing, translation and saving for later. Participants also highlighted the possibilities it opens for active reading such as rich-media (video), content search and references lookup. We used the insight gained from users of the tabletop prototype to design the cAR-Pad, a prototype with all of the elements necessary for a self-contained device. Building the cAR-Pad confirmed that a camera-based approach can efficiently determine location in a document, and identified the hardware and software elements necessary to support off-contact interaction techniques (*flipping* and *stacking*). Building both prototypes provides the basis to discuss the importance of an object model, possible registration optimizations, color mixing and device ecologies.

Our future work includes building a complete self-contained cAR-Pad with improved image capture, batteries, and an embedded-computer without the need of a light table. We will extend our exploration to other application areas like map navigation, music playback from sheet music, and support for hand-written forms. We also want to explore the multi-level registration, ad-hoc model acquisition, and different model distribution strategies.

# REFERENCES

1. Adler, M. J. and Van Doren, C. How to Read a Book. Revised edition, Simon and Schuster, New York, 1972.
2. Adler, A., Gujar, A., Harrison, B., O'Hara, K., and Sellen, A. A diary study of work-related reading: design implications for digital reading devices. In Proc. CHI '98.
3. Alahi, A. and Ortiz, R. and Vandergheynst, P. 2012. FREAK: Fast Retina Keypoint, IEEE Conference on Computer Vision and Pattern Recognition, 2012.
4. Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., and MacIntyre, B. 2001. Recent Advances in Augmented Reality. IEEE Comput. Graph. Appl. 21, 6.
5. Bier, E. A., Stone, M. C., Pier, K., Buxton, W., and DeRose, T. D. 1993. Toolglass and magic lenses: the see-through interface. In Proc. SIGGRAPH '93.
6. Bimber, O. and Raskar, R. 2005. Spatial Augmented Reality: Merging Real and Virtual Worlds. A. K. Peters, Ltd., Natick, MA, USA.
7. Carmigniani, J., Furht, B., Anisetti, M., Ceravolo, P., Damiani, E., and Ivkovic, M. 2011. Augmented reality technologies, systems and applications. Multimedia Tools Appl. 51, 1 (January 2011), 341-377.
8. Concept Phones. (2013, March 17). Transparent Phones. [Website]. Retrieved from http://www.concept-phones. com/tag/transparent-phone/.
9. Corning Incorporated. (2013, March 17). A Day Made of Glass 2: Unpacked. The Story Behind Corning's Vision. (2012) [Video file]. Retrieved from http://www.youtube. com/watch?v=X-GXO\_urMow.
10. Dachselt, R. and Al-Saiegh, S. Interacting with printed books using digital pens and smart mobile projection. In Proc. of the Workshop on Mobile and Personal Projection (MP2) @ ACM CHI 2011, 2011.
11. Futaba Corporation (2013, April 03). Futaba's OLED road map from 2012, Jan. 19 Website <http://www.oled-info.com/futabas-oled-road-map-amoleds-2014-transparent-and-flexible-oleds-cars-2015>
12. Fujitsu (2013, April 03)Addressing Trends and Business needs from (2012, June 14) Website http://www.fujitsu. com/be/Images/Workplace\_of\_the\_Future.pdf
13. Gabbard, J.L., Swan, J.E., Zedlitz, J., and Winchester, W.W. 2010. More than meets the eye: An engineering study to empirically examine the blending of real and virtual color spaces. In Proc. of VR '10. IEEE.
14. Golovchinsky, G. Reading in the office. In Proc. 2008 ACM workshop on Research advances in large digital book repositories, BooksOnline '08, pp 21-24
15. GSM arena. (2013, March 19). Lenovo S800 – Full phone specifications. (2011) [Website]. Retrieved from <http://www.gsmarena.com/lenovo_s800-4862.php>.
16. Hincapié-Ramos, J.D., Guo, X., Moghadasian, P., and Irani, P. 2014. Consumed Endurance: A Metric to Quantify Arm Fatigue of Mid-Air Interactions. In Proc. CHI 2014. ACM. *To Appear.*
17. Hinckley, K., Bi, X., Pahud, M. and Buxton, B. Informal information gathering techniques for active reading. In Proc. CHI '12. 2012. ACM.
18. Iwabuchi, M., Kakehi, Y., and Naemura, T. 2008. LimpiDual touch: interactive limpid display with dual-sided touch sensing. In ACM SIGGRAPH 2008 posters (SIGGRAPH '08). ACM, New York, NY, USA.
19. Japan’s transparent display vending machine. <http://techcrunch.com/2011/12/20/video-japan-vending-machine-2-0/>.
20. Kim, K., and Elmqvist, N. 2012. Embodied lenses for collaborative visual queries on tabletop displays. Information Visualization, 11(4), 319-338.
21. Lee, J.H., Bae, S.H., Jung, J., and Choi, H. 2012. Transparent display interaction without binocular parallax. In Adjunct Proc. UIST '12. ACM.
22. Marshall, C. C., Price, M. N. Golovchinsky, G., and Schilit, B. N. Designing e-books for legal research. In Proc. of the 1st ACM/IEEE-CS joint conference on Digital libraries, JCDL '01, 2001. ACM.
23. Mackay, W.E., Pothier, G., Letondal, C., Bøegh, K., and Sørensen, H.E. 2002. The missing link: augmenting biology laboratory notebooks. In Proc. UIST '02. ACM.
24. Matulic, F., and Norrie, M. C. Supporting active reading on pen and touch-operated tabletops. In Proc. AVI '12. 2012. ACM
25. Menk, C., and Koch, R. 2011. Interactive visualization technique for truthful color reproduction in spatial augmented reality applications. In Proc. of ISMAR '11.
26. Ohtani, T., Hashida, T., Kakehi, Y., and Naemura, T. 2011. Comparison of front touch and back touch while using transparent double-sided touch display. In ACM SIGGRAPH 2011 Posters (SIGGRAPH '11). ACM, 2011.
27. Price, M. N., Schilit, B. N. and Golovchinsky, G. 1998. Xlibris: the active reading machine. In Proc. CHI 98.
28. Rosten, E. and Drummond, T. 2006. Machine learning for high-speed corner detection, European Conference on Computer Vision, vol. 1, pp. 430-443, May. 2006
29. Schilit, B., Golovchinsky, G., and Price, M. Beyond paper: supporting active reading with free form digital ink annotations. In Proc. CHI '98, pp 249-256.
30. Schmalstieg, D., Encarnação, L. M., and Szalavári, Z. 1999. Using transparent props for interaction with the virtual table. In Proc. I3D '99. ACM.
31. Spindler, M., Tominski, C., Schumann, H., and Dachselt, R. 2010. Tangible views for information visualization. In Proc. ITS '10. ACM.
32. Sridharan, S.K., Hincapié-Ramos, J.D., Flatla, D., and Irani, P. 2013. Color Correction for Optical See-Through Displays Using Display Color Profiles. In Proc. VRST '13. ACM.
33. Tanaka, K., Kishino, Y., Miyamae, M., Terada, M., and Nishio, S. 2008. An information layout method for an optical see-through head mounted display focusing on the viewability. In Proc. ISMAR '08. IEEE.
34. Tashman, C. S. and Edwards, W. K. Liquidtext: a flexible, multitouch environment to support active reading. In Proc. CHI '11. 2011. ACM.
35. Wigdor, D., Forlines, C., Baudisch, P., Barnwell, J., and Shen C. 2007. Lucid touch: a see-through mobile device. In Proc. UIST '07. ACM.
36. Wu, C. H., Robinson, S. J. and Mazalek, A. 2008. Turning a page on the digital annotation of physical books. In Proc. TEI '08. ACM.

All material on each page should fit within a rectangle of 18 × 23.5 cm (7" × 9.25"), centered on the page, beginning 1.9 cm (0.75") from the top of the page and ending with 2.54 cm (1") from the bottom. The right and left margins should be 1.9 cm (.75"). The text should be in two 8.45 cm (3.33") columns with a .83 cm (.33") gutter.

# TYPESET TEXT

## Title and Authors

The title (Helvetica 18-point bold), authors' names (Helvetica 12-point) and affiliations (Helvetica 10-point) run across the full width of the page – one column wide. We also recommend phone number (Helvetica 10-point) and e-mail address (Helvetica 12-point). See the top of this page for three addresses. If only one address is needed, center all address text. For two addresses, use two centered tabs, and so on. For more than three authors, you may have to improvise.

## First Page Copyright Notice

Please leave 3.81 cm (1.5") of blank text box at the bottom of the left column of the first page for the copyright notice.

Table 2. Table captions should be placed above the table

|  |  |  |  |
| --- | --- | --- | --- |
| **Graphics** | **Top** | **In-between** | **Bottom** |
| Tables | End | Last | First |
| Figures | Good | Similar | Very well |

## References and Citations

Footnotes should be Times New Roman 9-point, and justified to the full width of the column.

Use the “ACM Reference format” for references – that is, a numbered list at the end of the article, ordered alphabetically and formatted accordingly. See examples of some typical reference types, in the new “ACM Reference format”, at the end of this document. Within this template, use the style named *references* for the text. Acceptable abbreviations, for journal names, can be found here: <http://library.caltech.edu/reference/abbreviations/>. Word may try to automatically ‘underline’ hotlinks in your references, the correct style is NO underlining.

The references are also in 9 pt., but that section (see Section 7) is ragged right. References should be published materials accessible to the public. Internal technical reports may be cited only if they are easily accessible (i.e. you can give the address to obtain the report within your citation) and may be obtained by any reader. Proprietary information may not be cited. Private communications should be acknowledged, not referenced (e.g., “[Robertson, personal communication]”).

# FIGURES/CAPTIONS

Place Tables/Figures/Images in text as close to the reference as possible (see Figure 1). It may extend across both columns to a maximum width of 17.78 cm (7”).

Captions should be Times New Roman 9-point bold. They should be numbered (e.g., “Table 1” or “Figure 2”), please note that the word for Table and Figure are spelled out. Figure’s captions should be centered beneath the image or picture, and Table captions should be centered above the table body.

# ACKNOWLEDGMENTS

Removed for blind review.

# REFERENCES

1. Bowman, M., Debray, S. K., and Peterson, L. L. 1993. Reasoning about naming systems. *ACM Trans. Program. Lang. Syst.* 15, 5 (Nov. 1993), 795-825. DOI= <http://doi.acm.org/10.1145/161468.16147>.
2. Ding, W. and Marchionini, G. 1997. *A Study on Video Browsing Strategies*. Technical Report. University of Maryland at College Park.
3. Fröhlich, B. and Plate, J. 2000. The cubic mouse: a new device for three-dimensional input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (The Hague, The Netherlands, April 01 - 06, 2000). CHI '00. ACM, New York, NY, 526-531. DOI= <http://doi.acm.org/10.1145/332040.332491>.
4. Tavel, P. 2007. *Modeling and Simulation Design*. AK Peters Ltd., Natick, MA.
5. Sannella, M. J. 1994. *Constraint Satisfaction and Debugging for Interactive User Interfaces*. Doctoral Thesis. UMI Order Number: UMI Order No. GAX95-09398., University of Washington.
6. Forman, G. 2003. An extensive empirical study of feature selection metrics for text classification. *J. Mach. Learn. Res.* 3 (Mar. 2003), 1289-1305.

Columns on Last Page Should Be Made As Close As Possible to Equal Length