SIP (2017), vol. 6, e12, page 1 of 14 © The Authors, 2017.

doi:10.1017/ATSIP.2017.10

INDUSTRIAL TECHNOLOGY ADVANCES

Learning in the 21st century cyber-physical age

CHANDRAKANT PATEL[1], YANG LEI[1], LEI LIU[2*], RARES VERNICA[1], JIAN FAN[1], BRAD SHORT[3], JERRY LIU[1] AND STEVEN J. SIMSKE[4]

The learning tools necessary to prepare the next generation of students must be shaped by the socio-economic needs of the 21st century. The needs of the 21st century – from rebuilding city scale physical infrastructure to personalized healthcare – not only require learning from the wealth of global information available on the Internet, but also the building of a strong grounding in fundamentals. History has shown that the depth in fundamentals has been achieved through conventional books. Indeed, authoritative books in physical fundamentals have been penned in the 19th century and early 20th century. We present a 21st century cyber-physical learning platform that combines the best of physical books with information systems. The systemic instantiation of the platform combines modern optics and computing to view books, scan objects, and enable interactive learning – while simultaneously benefitting from the vast pool of information on the Internet. The hybrid learning platform preserves the best of the past to pave the way for the future. It also enables future research such as meta-data and descriptive tagging of the large number of images available on the web.

Keywords: cyber physical systems, education, immersive learning

Received 24 February 2017; Revised 18 August 2017

I. INTRODUCTION

Technology and science in the 19th and early 20th centuries were largely about the industrialization of electro-mechanical (physical) systems like the steam engine and the utility grid. The latter half of the 20th century has been about information (cyber) systems and the Internet. The 21st century will be about the integration of these two, namely cyber-physical systems (CPS).

The 19th century and early 20th century – or the machine age – led to authoritative contributions in the fundamentals of physical systems. A wealth of knowledge was created, and chronicled with great depth, in books on subjects such as physics, chemistry, thermodynamics, fluid mechanics, solid mechanics, structural dynamics, energy technologies, power generation, etc.

In the latter half of the 20th century, termed herein the cyber age, the rise of the Internet led to the proliferation of access devices such as e-readers and a correspondingly large pool of distributed information readily available on the Internet. Often, the information created by a wide range of contributors lacked corroboration and verification. Regardless, a profound change occurred in the ways in which we consumed information to meet our learning needs. Schools provided their students with tablets and encouraged the online search of information and use of e-books. Printed books were literally considered a physical burden.

The 21st century needs – driven by social, economic and ecological "megatrends" – are cyber-physical. In light of resource constraints, there is a need to maximize the utilization and improve the efficiency of conventional physical systems such as oil refineries, power plants, power distribution systems, city scale infrastructure, transportation infrastructure, etc. New civil infrastructure such as roads and bridges needs to be built given the increase on demand resulting from the long-standing increase in population and rapid urbanization. Resource constraints are motivating the need for novel methods in energy sourcing and consumption in order to reduce the energy used across the lifetime of products. This is the notion of "least lifetime Joules products". Such sustainability considerations require great depth of knowledge of physical fundamentals, and of information management and analysis techniques.

Printed Books serve as Indelible Metadata of Important Information

The rise of 21st century CPS requires students to have an in-depth understanding of the physical fundamentals and of information systems. In that vein, the learning has to follow suit as well.

Routinely accessing Internet-based information using information technology devices, and assimilating

[1]HP Labs, Palo Alto, California, USA
[2]Huawei Technologies, Santa Clara, California, USA
[3]HP Inc., San Diego, California, USA
[4]HP Labs, Fort Collins, Colorado, USA
*This work was done when the author was with HP Labs.

Corresponding author:
Y. Lei
Email: ylei@hp.com

knowledge of physical systems without understanding the physical fundamentals in depth, will not work in the long term, as the decisions made will become unanchored with the physical realities of the devices. By the same token, learning of fundamentals using physical textbooks, without corresponding researching of the wealth of information on the Internet, will not suffice. In the 21st century cyber-physical age, the learning must also be cyber-physical – physical textbooks combined with information gathered and availed through a range of information technology devices. An undergraduate civil engineering student, learning the mechanics of solids using a classic book such as the one by Egor Popov, must also be able to quickly view the state-of-the-art automated bridge construction being applied in China by availing herself of the video content available on the Internet [1].

Furthermore, a key differentiator of physical books is the lasting impact in one's memory [2]. Printed physical books create a metadata of information in one's memory that can be lifelong. One of the authors, Chandrakant Patel, recounts below a few of the instances in his own career.

- As a high school student in India in the early 1970s, the author recalls with great fondness his Physics book. The study of work, energy and power availed from the book entitled "O-Level Physics" by A.F. Abbott cemented his understanding of the laws of thermodynamics – particularly the 2nd Law of Thermodynamics. Thirty years after the high school textbook experience, given the depth in understanding, the author was able to create an entire research program based on the lifetime use of available energy or exergy (2nd Law of Thermodynamics) as the measure [3].
- Likewise, when the use of an induction motor in a modern cyber physical vehicle such as Tesla Model S needed to be understood, the author went back to his visual memory of the Physics book by Abbott. The color, the font, and the pictures served as metadata to recollect information that is more than four decades old.
- In college, in 1982, the author learnt about the collapse of the Tacoma Narrows Bridge in a course on mechanical vibrations. The bridge – in Tacoma, Washington – collapsed as a result of wind induced vibrations. Today, 35 years later, the author can go to the page in the physical textbook that explains the cause in minutes. The book, on Mechanical Vibrations, was originally written by J.P. Den Hartog, Professor of Mechanical Engineering at MIT, in 1934. It was followed by several revised editions. The feel of the paper, the font, and the *vivid sketch* that explained the cause, and the location of the sketch in the book, have become indelible metadata.

The sketch in Den Hartog's book shows that the original bridge had a road bed with an "I" type cross-section. The wind blowing across the "I" cross-section created eddy currents, which induced a series of "pushes" that resulted in a torsional or twisting motion, as shown in Fig. 1. The frequency of the wind induced pushes on the bridge's cross-sectional span, which coincided with one of the natural frequencies. The pushes aligned or tuned with the natural movement of the bridge. Therefore, akin to a swing, and given the alignment, each push added to the amplitude of

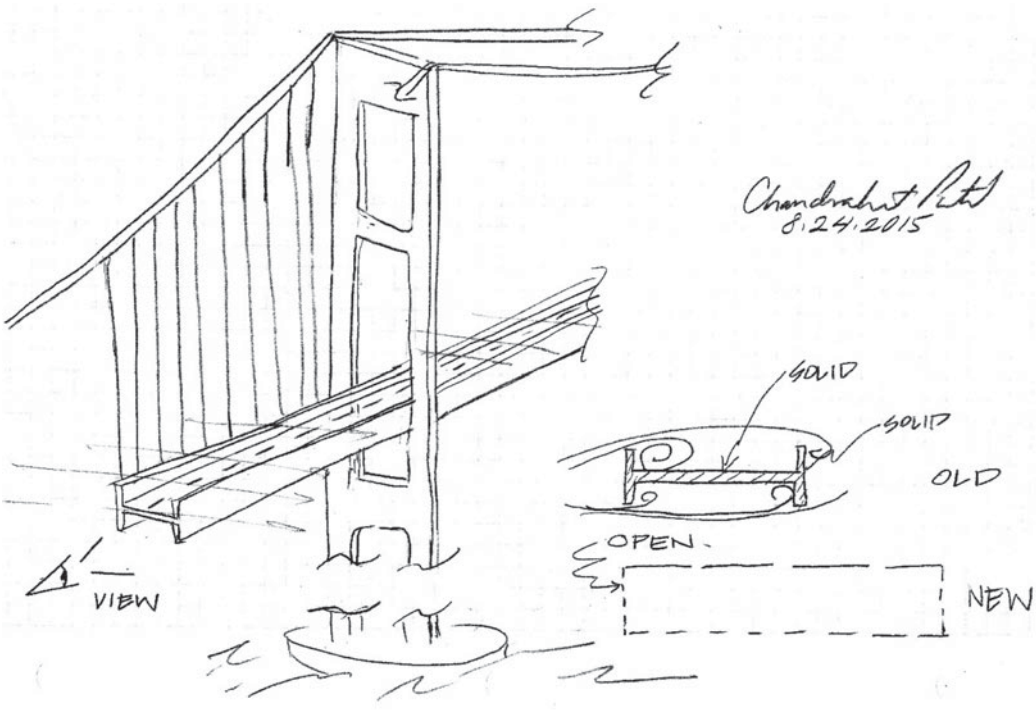


Fig. 1. A sketch of the Tacoma Narrows Bridge and its road bed designs.

https://doi.org/10.1017/ATSIP.2017.10 Published online by Cambridge University Press

vibration, resulting in an amplified condition called resonance. The amplification from the focused energy caused the catastrophic failure. The sketch in the book shows that the new bridge has a box construction with gaps that allows air to flow through; and, the box design is stiffer from a torsional or twisting point of view.

In 1982, the author spent half a day in search of the film of the Tacoma Narrows Bridge collapse in the library. Today, on the Internet, within seconds one could avail many videos from the cyber repository in the "cloud". Deep learning garnered through the careful study of primary source material is effectively augmented by the breadth of media available online.

A cyber-physical platform is needed to preserve the best of both worlds – the understanding of fundamentals and of readily accessing the actual content from the Internet. This is not a break with the past, but in fact a logical extension of the history of education. Since the first hieroglyphics on clay tablets, educational materials have been used to augment human intelligence. As time and technology continue to move forward, the nature and extent of the cybernetic augmentation of the physical intelligence of humans will continue to deepen. Will there be a "feed" (M.T. Anderson, "Feed", Candlewick Press, 2002, ISBN 978-0-7636-2259-6) in us all? Current trends in electroencephalographic interfaces certainly indicates as much. The main difference from the clay tablets is the immediacy of the content augmentation.

II. INTRODUCING A CYBER-PHYSICAL LEARNING PLATFORM FOR THE 21ST CENTURY

We propose a cyber-physical hybrid learning system, Meaningful Education and Training Information System (METIS) that provides an easy digital-to-print-to-digital content creation and immersive learning experience [4, 5]. A key aspect of METIS is in allowing students to direct their own learning experience and to control the amount of mark-up to be added to the baseline materials. METIS incorporates technologies for document image retrieval (DIR), handwritten annotation extraction, layout, image recommendation, personalization, co-creation, and assessment [6–9]. These features facilitate and, in many cases, significantly simplify common teacher/student tasks. Our system has been demonstrated at several international education events, partner engagements, and pilots with local universities and high schools. It is targeted for secondary education, where learning process includes content creation and collaboration as significant components. We present the system and discuss how it enables hybrid learning.

In this section, we use the following terms: *Article*, *Book*, *Page*, *Note*, and *Cheat Sheet*. These terms are defined as follows.

- Article: a collection of text, images, and other multi-media elements related to a particular topic.
- Book: a collection of articles in a set order.
- Page: a section of a book, as delimited by the layout engine.
- Note: a section of a page as delimited by the user.
- Cheat sheet: a collection of notes from one or more books.

The METIS system supports cyber-physical learning in many ways. The platform described in Section 2.1 is accessed through a Web browser and runs on a cloud infrastructure. As such, it has minimal hardware requirements. Essentially, any device with Internet connection and a Web browser can be used to access the platform. Moreover, a browser extension is used to access the content when the device is offline. In cases where the Internet connection is unreliable, the entire platform can be deployed on premises (e.g. a server running on the school site). Additional hardware components, such as a projector-camera imaging system, is used to enable the immersive cyber-physical learning experience, which is described in detail in the rest of the paper. Section 2.2 describes the physical system requirements in details.

A) METIS architecture

The entire METIS platform is built as a web application using the Play Framework running on a Java server. The high-level system architecture of METIS is provided in Fig. 2.

Rendering content in METIS uses a custom built engine called AERO, which provides custom publishing capabilities [6]. AERO takes as input the content to be rendered and a set of templates. From these inputs, AERO generates a layout of the content within the given templates while optimizing an aesthetic scoring function. For example, a scoring function minimizes the amount of left-over white space left in the template or measures the distance between text containing a reference to a figure and the figure itself.

AERO runs entirely in the client browser using JavaScript. At start-up, a suite of layout templates are loaded, followed by the content. Using a greedy algorithm for each

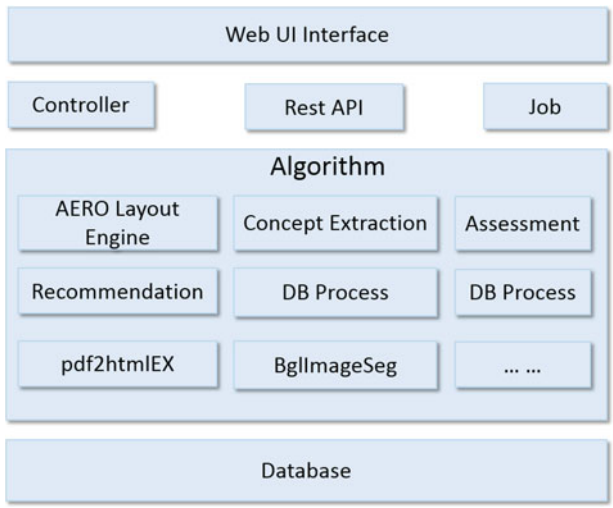


Fig. 2. System architecture of METIS.

https://doi.org/10.1017/ATSIP.2017.10 Published online by Cambridge University Press

page, a template is selected from the suite and filled with different chunks of content. The score is measured at each step and the best score, template, and content combination is remembered. In this regard, a METIS Book is simply a collection of Articles in a predefined order. Thus, making long or complex Book is a trivial extension of the base rendering engine. To optimize the rendering efficiency, we save the output of the Book and can leverage this cache instead of AERO to re-render the content every time.

For generating the PDF version of the Book, we extend the PhantomJS library. PhantomJS acts as a headless browser and runs the AERO engine. Once the content is rendered (stripped of excess content and/or reformatted), PhantomJS saves the HTML output in PDF format.

1) Markup & accessibility

By virtue of METIS Articles being based on HTML, content can contain text, images, video and other elements supported by HTML. As screen-reader accessibility is an important aspect of education, METIS denotes headings (H1 to H6 tags), figures (using the HTML5 FIGURE tag), and other markup. Moreover, to evaluate METIS accessibility, we use and conform to the WAVE tool.

2) Facilitating interaction with digital reader

For a better reading experience, AERO uses the zoom.js library to zoom in the content of the page and adapt it to the browser window current width and height. This allows AERO to adapt elegantly to different screen resolutions as Learners switch devices and platforms, rather than "naively" scaling up content.

3) Notes architecture

When created, each Note is associated with a user, a Cheat Sheet, a section in the Cheat Sheet (if applicable), a book, a page on that book, a start element ID and an end element ID (like paragraph or figure), and a start and end offset in the start and end elements, respectively. A Note also contains the selected text in rich HTML format, an assigned color for the color bar, a creation date, and a modification date. Using the cell phone camera or a regular webcam, the Learner can capture a picture of a book page or Note (on which she took notes) and send the picture to a predefined email address. The system then extracts the image and applies a suite of color and geometrical adjustments to the image in order to improve clarity. We perform QR code recognition and identify the page or Note. Combining the email address and ID in the QR code, the enhanced image is appended to the Cheat Sheet, which was last used to take notes on the same book.

4) Offline support

We recognize the need for offline support in a learning platform. To address this need, we provide a browser add-on fully integrated with the online platform. The add-on is developed for Google Chrome and Mozilla Firefox browsers, leveraging the local storage in the browser extensions. By using local storage in browsers that have cross-browser sync, we are able to keep multiple browsers for a given user in sync with offline content and preferences.

B) Physical design of the platform

An immersive cyber-physical learning platform requires a computing system and form factor that expands beyond conventional touch screen, webcam, computing platforms. Traditional computing systems lack the visual and physical capture sensing capabilities and configurations necessary to provide real-time feedback for immersive physical learning. A new computing platform is needed that can "see" both the user and physical objects and respond with superimposed digital content interacting in the same shared space. An immersive computing platform enables a user experience we call direct augmented reality or "blended reality" ideal for aiding a kinesthetic learning process (Fig. 3).

Touch screen interfaces with integrated world facing cameras have become prevalent, and such low latency tactile and optical feedbacks are valuable for natural user interaction and traditional video and photo capture. However, to enable immersive physical learning a computing platform needs intelligent awareness of the physical content coexisting with the digital interface. Specialized projection display technology to project onto physical objects and specialized sensing hardware oriented to look back at the display are needed to enable this real-time blended reality experience. This configuration of a digital display projection engine shining coincident, with camera sensors connected in a

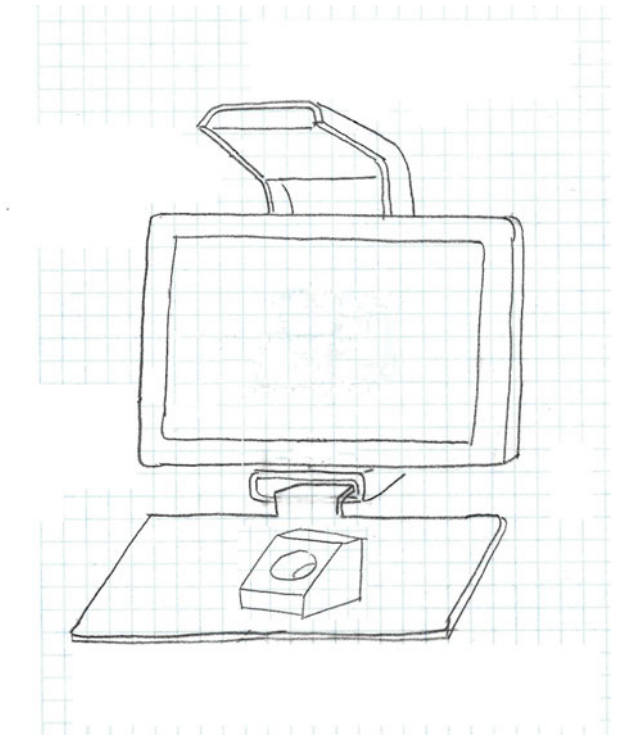


Fig. 3. One example of many possible setups of the computing platform for immersive learning.

https://doi.org/10.1017/ATSIP.2017.10 Published online by Cambridge University Press

computational feedback loop, defines a projector-camera or Pro-Cam configuration, which is an ideal closed-loop interface system for building an immersive computing platform. One example of such immersive computing systems available in the market is Head-Mounted Display (HMD), such as VR headset. Another example is the HP Sprout All-in-One desktop PC [10], which is used to prototype the immersive experience for our cyber-physical learning solution.

III. IMMERSIVE LEARNING

In this session, we will introduce the immersive learning experience enabled by METIS and the computing platform described in Session II. The proposed immersive learning system can find corresponding pages in the database by capturing an image of the textbook on the working surface, augment meaningful digital content on the screen or on top of the print, and let users interact with the content by touching the physical textbook.

Our advanced document retrieval and handwritten annotation extraction algorithm are built in the system to enable this learning experience. When users are reading textbooks in front of Sprout, they may log into their METIS account and put the book on Sprout Touch Mat. A document retrieval algorithm is used to analyze the document image, find its archived digital version, and bridge the physical textbook with meaningful digital information.

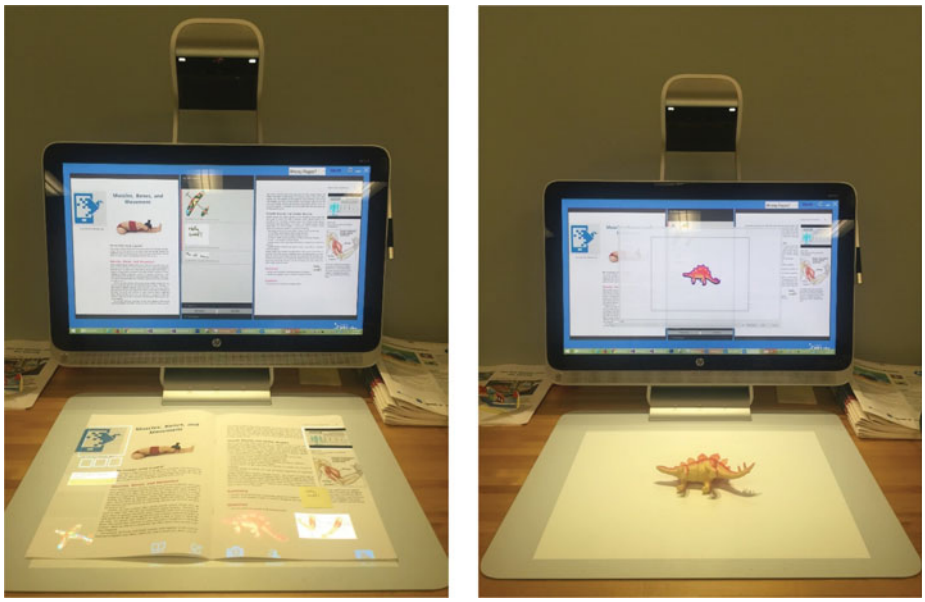
Educators and Learners both can create interactive digital content to augment any page in a book through either the web interface or immersive learning system, which we call "annotating" the physical book. Digital content examples are external links, images and videos, text comments, and quizzes. When students read specific pages, these interactions are projected on top of the physical book, as shown in Fig. 4(a). Actions are triggered through touching the augmented content on the physical book; for example, opening a webpage, playing a video, or taking some notes.

When users write notes on their print book, the handwritten annotation extraction algorithm extracts the notes from book and groups them into sentences or paragraphs. Then they are archived in the database and can be shared easily with team members. One feature of our system is that it remembers where in the page the note was written. When the notes are shared, other users can visualize it through a projection on top of the print book.

A big challenge for modeling a Learner's behavior is the difficulty of tracking offline learning activities, which occurs when user is learning through a physical book. Our system actually enables offline learning activity logging. With DIR, annotation extraction, and the digital–physical linkage, we know exactly which page of which book the user is reading, what notes were taken, and how many interactions the user had with this book. These data are logged into our database in real-time and serve as feedback to the METIS system. These data are provided to the educator and also can be used to model the Learner's behavior.

Feedback can be viewed as transforming learning into a control system with feedback, where the controlled variable is the learning. The feedback is the student's attentiveness, as assessed by eye gaze, electroencephalographic activity, pulse rate, electrocardiographic analysis, etc. This input offers a potentially less arbitrary means of assessing the effectiveness of a learning approach than test scores, since it is independent of the material. It also allows the Learner to learn at a pace and rate of repetition best suited to his learning style.

One example use case of the hybrid learning system is the co-creation experience. The best way to learn is to teach. Our co-creation feature allows both educators and Learners to add their understanding of the core knowledge



(a) (b)

Fig. 4. Demonstration of the hybrid learning system. (a) Physical textbook serves as an index to digital information. (b) Scanning and adding an object to the physical textbook.

https://doi.org/10.1017/ATSIP.2017.10 Published online by Cambridge University Press

and share with others. In addition to sharing knowledge, it is a good way for collaboration and peer-teaching. Our unique contribution here is blending the digital and physical techniques of social learning together through web-based and immersive learning solutions.

Imagine a group of students are working on a team project to build an autonomous car prototype. The fundamental physics and mechanics knowledge are in their textbooks. In addition, they need to figure out what parts to order, how to build the software, and how to make sure they work together perfectly. Each member of the team is responsible for part of the process, and thus their work should be shared within the group.

With METIS, the textbook serves as the index to the digital resources online. All information needed, such as the parts to order, instructional videos, software packages, etc., are linked to specific locations on certain pages of the book. Students can also write down notes on the print book if preferred; the notes will be digitized and saved in the system. Any physical objects they already have can be two-dimensional (2D) or 3D scanned with the projector-camera system. Figure 4(b) demonstrates the 2D scanning capability. Each person's contribution will be shared with the group. Everything a person may need is in one place with easy access.

Figure 5 shows the overall system architecture for co-creation. Users can create and manage annotations from both web-based and immersive learning. The *web-based learning* is responsible for creating annotations from the recommendation service or cheat sheet, or creating annotations directly at any position on the page. *Immersive Learning* is used to retrieve and locate pages when the Learner puts the book on the Sprout Touch Mat, creates the 2D/3D visual content for annotation, or extracts the hand-written notes from printed pages. The annotations that are created by either process are synchronized in real-time, as all these resources are managed in the cloud. The Learner (Fig. 5, top center) has the option to co-create annotations and share any created annotations with others in both web-based and immersive learning environments. The details of each component will be discussed in the following sections:

Web-based Learning: The web-based learning module has the following three major components:

- *Annotation via Recommendation*: While reading, Learners may have problems understanding a portion of the content. METIS allows the Learner to select content that he fails to understand or wants to explore more at length, thereby triggering the recommendation service to discover external learning resources in multiple formats, such as video, image and web content (step 1 of Fig. 6). These recommended resources are used to create the annotation by simply dragging and dropping to the position where this query was initiated (step 2 of Fig. 6). METIS has the capability to extract annotation attributes, such as title, URL, resolution, dimensions, etc. (step 3 of Fig. 6). The Learner accesses the annotation at a later time by simply clicking the "Extra Content" button in the

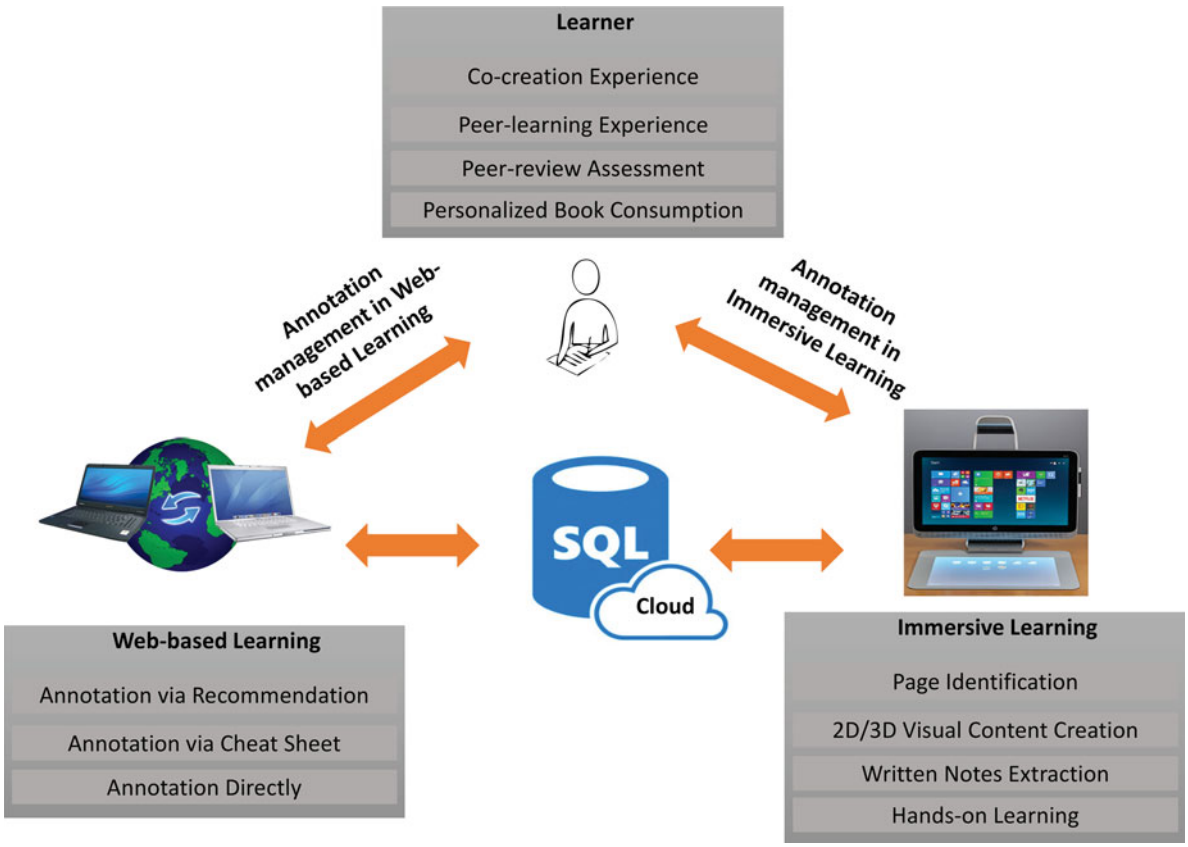


Fig. 5. System architecture for co-creation.

https://doi.org/10.1017/ATSIP.2017.10 Published online by Cambridge University Press

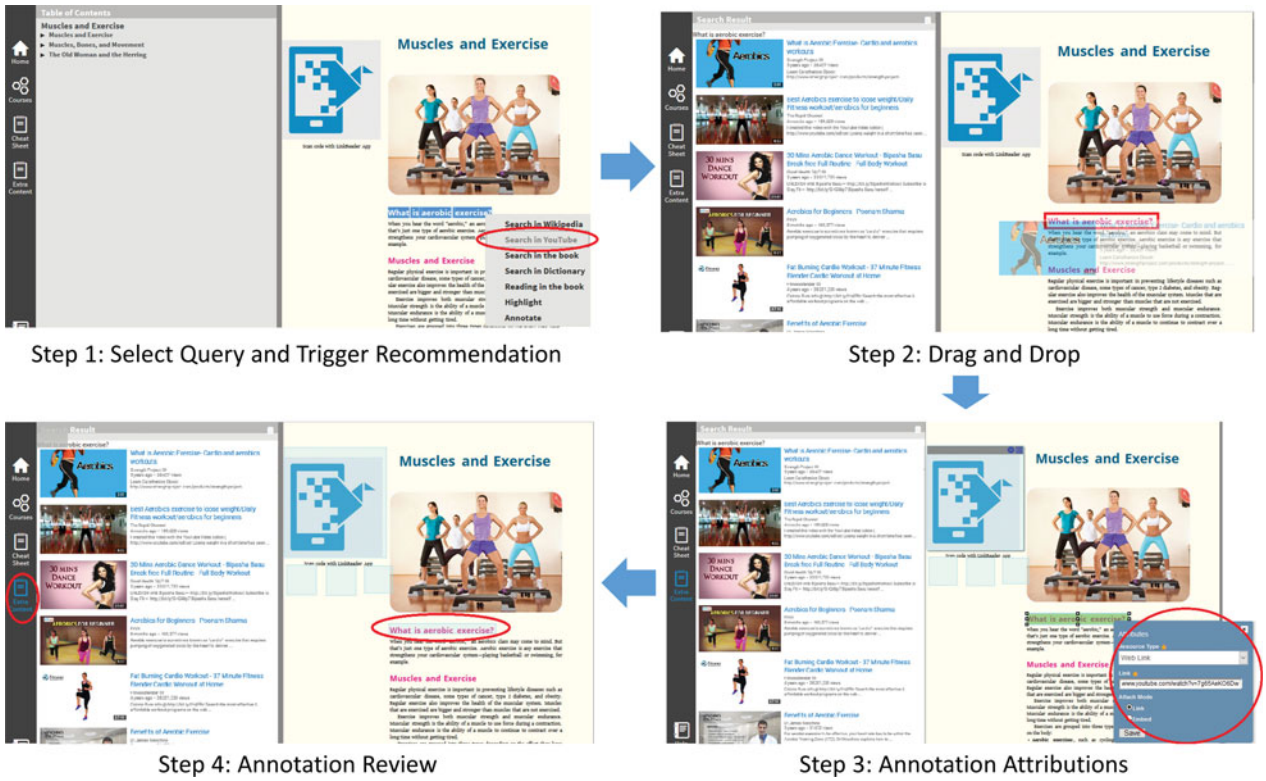


Fig. 6. Illustration of creating annotation via recommendation.

left sidebar, whereby the previously built annotation area will appear on the current reading page. By simply clicking the annotation area, the previously saved annotation resources will be available (step 4 of Fig. 6).

- *Annotation via Cheat Sheet*: Similarly, Learners can create the annotation by dragging and dropping the resources from Cheat Sheet notes. The content in these notes come from the scanned or extracted hand-written images from Sprout, the content notes created from the reading book, or any notes/comments that one or more Learners have saved previously.
- *Annotation Directly*: The Learner also can move the mouse to any position where he wants to create an annotation and right click the mouse to trigger the annotation option.

Immersive Learning: The immersive learning module has the following four major components:

- *Page Identification*: While reading the printed book, the Learner may want to create an annotation on a specific page. He can place this page on the Sprout Touch Mat, and the Sprout camera will capture an image of the page. Our image retrieval algorithm identifies and retrieves the page of the book currently being read while showing the digital version on the screen.
- *2D/3D Visual Content Creation*: To create visual content for annotation purposes, suppose that the Learner places a "HP mouse" on the mat and wants to get a 2D screen image added to current reading page. Our algorithm first identifies the page of the book in the above step then uses the identified book image as background. After scanning the object, METIS automatically determines the object location and size on the printed book, and then adds it to the same page of digital book, with the same size and location. Learners can customize with their preferences as default settings. As the last step, Learners can customize the attribute list of the annotation they wish to create, and save it (similar to step 3 of Fig. 6). Figure 7 is an example of creating a 2D/3D visual content for annotation from Sprout.
- *Written Notes Extraction*: It is easier to take hand-written notes while Learners are reading the printed materials. To enable a hybrid annotation creation experience, METIS has the ability to extract hand-written notes from the printed pages/books. The extracted hand-written notes are available for creating annotations in an image format. Figure 8 is an example of the extracted hand-written notes saved in a Cheat Sheet (available for annotation creation at a later time).
- *Hands-on Learning*: The immersive computing platform (HP Sprout) features a well-calibrated projector-camera system that is ideal for structured light 3D scanning. This feature is incorporated in our immersive learning system and supports the co-creation of 3D models. Furthermore, when a 3D printer is connected, customized part design and (re)production is a very engaging hands-on experience for learning, which we call "Art to Part".

Learner: The Learner module has following four major components:

- *Co-creation Experience*: Different Learners reading the same book page can create and share their personal annotations with others. Also, annotations created from either

https://doi.org/10.1017/ATSIP.2017.10 Published online by Cambridge University Press

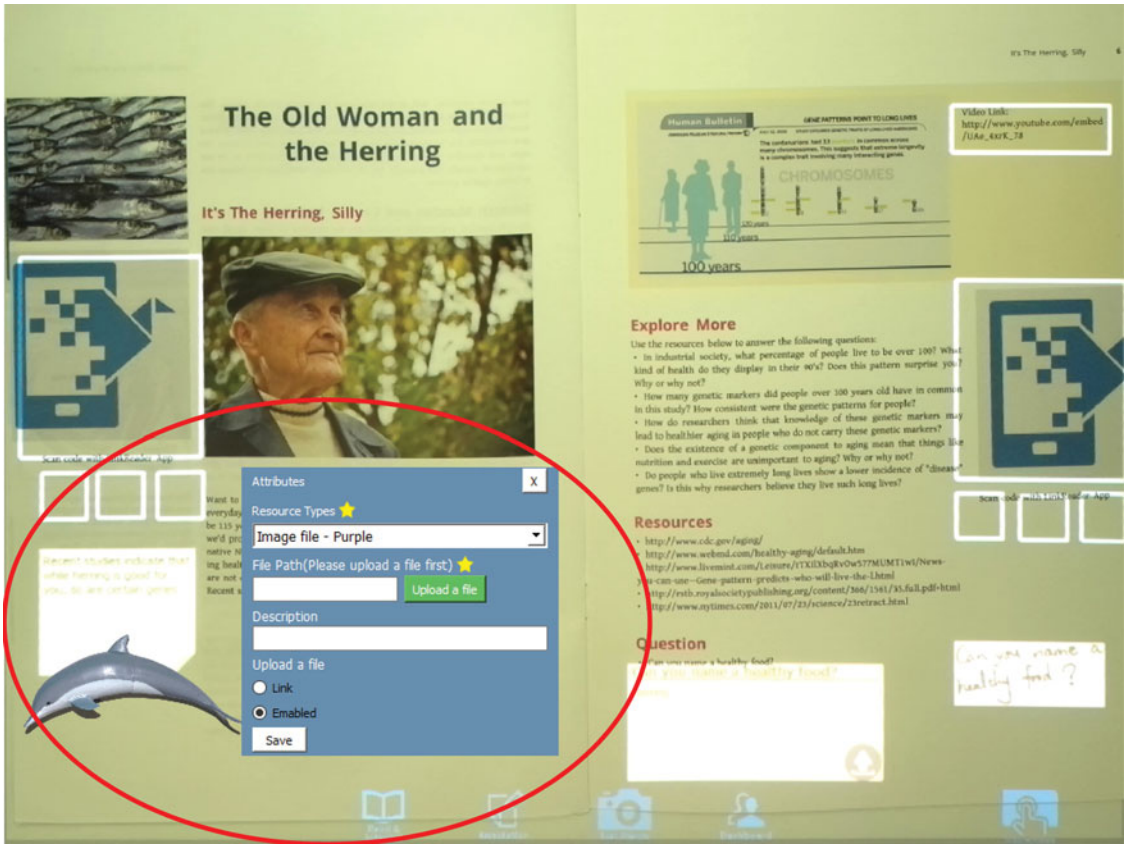


Fig. 7. Illustration of 2D/3D capture for creating annotation from sprout.

Fig. 8. Example of hand-written notes extraction.

a web-based learning system (e.g. METIS) or an immersive learning system (e.g. Sprout) will be synchronized in real-time, which allows the Learners a real-time co-creation and discussion experience.

- *Peer-learning Experience*: METIS supports real-time annotation co-creation, which makes real-time discussion possible. Also, Learners can teach other peer-Learners using our system. For example, two students can work

https://doi.org/10.1017/ATSIP.2017.10 Published online by Cambridge University Press

in front of two immersive computers (Sprout), with one raising a question and the other teaching the content step by step with our real-time hand-written extraction and sharing function.

- *Peer-review Assessment*: METIS first automatically extracts key concepts from learning content based on machine learning technology. Then all of the extracted key concepts are fed to Learners to allow the Learners to build the connections and define the pairwise relationship based on their understanding. This collection of terms and pairwise relationships is a concept map. Every Learner would not only build his own concept map based on his understanding, but also review the concept maps defined from other Learners and assign the review a score and comments. At the end, each Learner may check the score and comments from other reviewers, and then give feedback to the reviewers by rating their comments as helpful or not. This feedback system adds a level of social accountability and process fidelity. In total, every Learner is evaluated for her understanding of the concepts (average review ratings from peer-reviewers), as well as her ability to provide meaningful and constructive feedback to others. This peer-review assessment function is assessed from either web-based learning system (e.g. METIS) or immersive learning system (e.g. Sprout)
- *Personalized Book Consumption*: Every Learner learns differently. Learning outcomes can benefit greatly if such personal learning characteristics are considered. Based on this observation, we developed a new technology to generate a personalized book which provides the same book view to all the students. At the same time, we consider each Learner's profile to embed the most appropriate learning resources, which are helpful to aid her in comprehending the content or exploring more resources. On the digital side (from both web-based or immersive learning systems), a Learner can click on the small icon we created for her to explore the additional learning resources recommended based on her learning profile (reading history, test records, learning behavior, learning preference, etc.). On the print side, she can scan each page with a smart device or HP Sprout; then, the embedded personalized learning resources will be provided in a pop-up page. As an aside, if two students scan on the same page, they will be provided with different supplementary learning resources as our system could automatically recognize them and build their learning profile.

IV. METHODOLOGIES

A) Document image retrieval

Image recognition and retrieval has been an active research field over the past several decades [11–18]. Based on different query inputs to the image retrieval system, most of this research focuses on two approaches: image meta search [15] and content-based image retrieval (CBIR) [16]. Image meta search requires descriptive keywords or text as input and then searches for the images, which most accurately display the semantic meaning of keywords or text. The difficulty of image meta search is how to do indexing and build the database efficiently, especially when database is huge and real-time is required [15], so that retrieval can be done fast and accurate. In contrast, CBIR takes a query image as input and searches for images in the database based on visual similarity. CBIR serves as the prerequisite for many applications such as annotation extraction, reader evaluation, and supplementary content embedding.

Most CBIR techniques use a feature vector to represent the query image, and then compare this feature vector with other feature vectors in the database to obtain similarity. Traditionally, people consider three types of features for image retrieval: color features, texture features and/or shape features [17]. However, all of these approaches suffer some limitations. Color features cannot distinguish between document images, which carry different semantic meanings. Texture features typically require transformation into a frequency domain for similarity comparison, which ignores spatial information. Shape features usually need the existence of certain objects in the image, which is not always guaranteed in practice [18]. In [19], the author proposes a similarity ranking method based on nearest-neighbor distance, which outperforms support vector machine-based image retrieval method.

We are specifically interested in DIR. Normally, print textbooks are in the form of documents, which may include text, drawings and pictures. We call these mixed-content documents. In the context of cyber-physical learning, what we have are printed textbooks and visual sensors, say cameras, to help us identify the current page. This leads to an image-based instance retrieval problem. The input is an image, and we want the exact match from our database.

Our DIR algorithm employs a hierarchical structure, which applies Bag-of-Words (BoW) [20] first, followed by spatial verification. The BoW stage utilizes SIFT [21] features and applies K-means clustering [22] to generate centroids in feature space. If the BoW finds a highly significant matching image in our database, our IR engine terminates the retrieval process and output this significant result. Otherwise, the BoW stage will output the 10 most likely images in the database according to the feature-space similarity score and feed them to a subsequent spatial verification stage. In other words, BoW is applied first to narrow down the potential matching candidates, and then the more expensive, from a processing standpoint, spatial verification conducts point-wise matching between the captured and candidate images (Fig. 9).

1) Bag-of-words training

Instead of matching each individual point in two images, which is computationally intensive, BoW feature reduction can significantly improve the retrieval speed. Similar to a traditional BoW approach, we first need to train a code book in the feature descriptor space, which consists of N centroids in the feature descriptor space. There are some

https://doi.org/10.1017/ATSIP.2017.10 Published online by Cambridge University Press

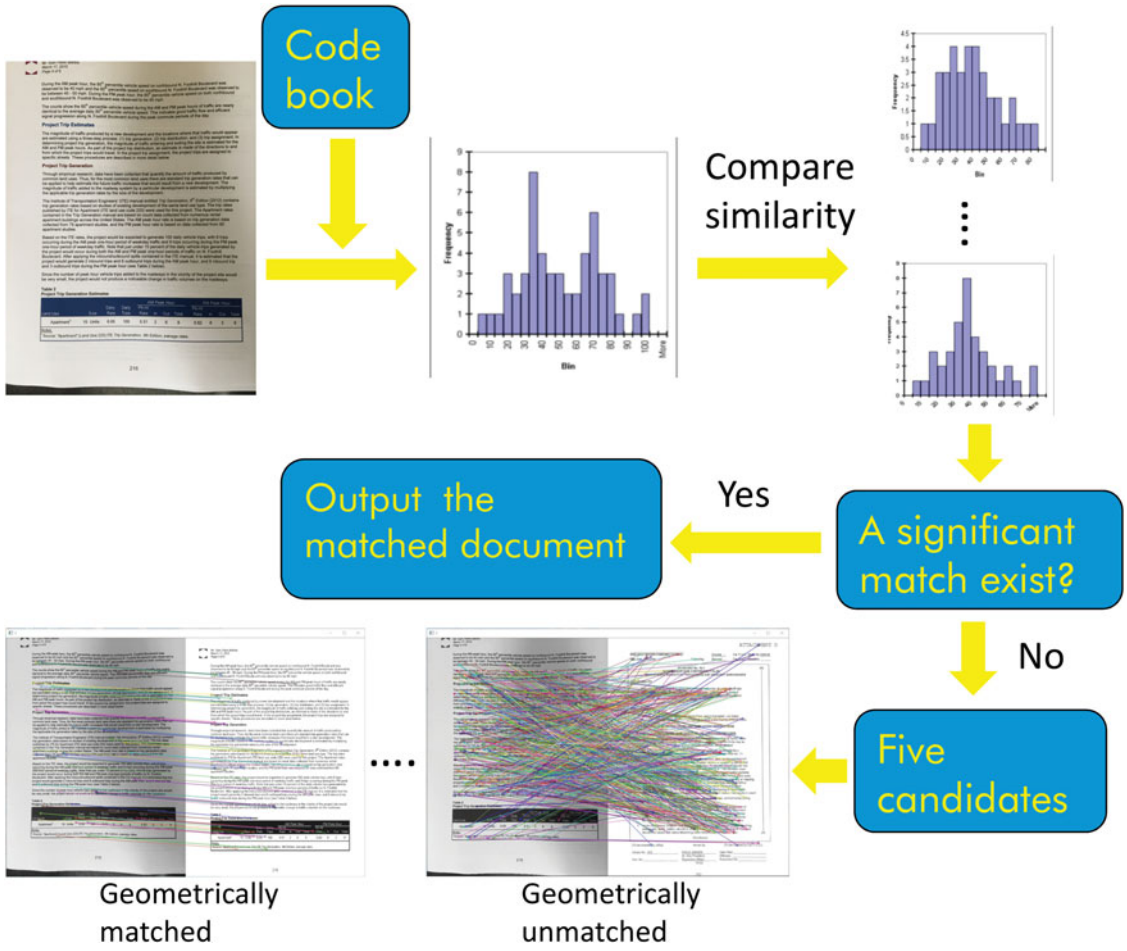


Fig. 9. Flowchart of the Document Image Retrieval Algorithm.

studies about how to pick optimal N for different applications. In our case, we choose $N = 400$ empirically. The SIFT descriptor is used to build the BoW code book.

In order to build the code book, we extract points of interest of all images in the database and then find the descriptors associated with all these points of interest. In our case, these descriptors are 128-dimensional vectors. K-means clustering is then used to get the code book, which leads to a histogram vector representation of each image in the dataset. Every bin in this histogram represents the number of points assigned to this centroid. In our case, image i is represented by a 400-bin histogram vector V_i.

2) BoW retrieval

In the retrieval stage, we extract all points of interest and associated descriptors in the captured image c, then follow the code book that was built during training to generate its histogram representation V_c. All the images in the dataset are then ranked by the cosine similarity S_i between their histogram vector and V_c.

$$S_i = \frac{V_c V_i}{|V_c||V_i|}. \quad (1)$$

The top two highest similarity score S_{i^*} and $S_{i^{**}}$ are then compared. If $S_{i^*} - S_{i^{**}} > T_1$, we will output image i^* as the exact match. Threshold T_1 is experimentally determined in a conservative way, i.e. we err on it being larger rather than smaller. Otherwise, the top 10 candidates are selected to continue to the next stage.

3) Weighted spatial ranking

The BoW solution described above has two potential issues, especially with document images. First, it ignores the spatial information of the image. For example, two images may have same content but are organized in a spatially different fashion. Thus, they may have very high similarity score, but they are not a good match. Secondly, cosine similarity only evaluates the angle of two vectors and ignores their magnitudes.

To address the above issues, we introduce the *Weight Similarity Ranking* (WSR) approach. Instead of searching for the feature vector V_i in the database which has the largest cosine similarity, we also consider the spatial consistency which is weighted by the number of matched points used for generating homography.

For each candidate image, we conduct point-wise matching with the query image in the feature descriptor space and find a group of points that have the smallest distances. These points are considered possible matches in the feature descriptor space. However, we need to verify them spatially.

https://doi.org/10.1017/ATSIP.2017.10 Published online by Cambridge University Press

With the group of matched points and their coordinates, we can build a homography H, which represents the possible perspective relation between these two images. We can expect that a real matched candidate in the database should have strong homography. In contrast, false candidates may have good similarity in BoW, but will show a poor homography with query image.

To evaluate the quality of H_i for candidate i, we apply single value decomposition (SVD) to H_i and get its eigenvalue ratio R_i. A smaller R_i indicates a high quality homography. In addition, we look at the number of points N_i that are used for generating H_i, and introduce the Weighted Spatial Ranking Score (WSRS) as follows:

$$WR_i = \frac{R_i}{N_i} \tag{2}$$

Then the top 10 candidates are re-ranked based on their WSRS values. Note that a lower WSRS value actually indicates a preferred Homography. If a significant preferred WR_{i^*} can be found such that $|WR_{i^{**}} - WR_{i^*}| > T_2$, where i^{**} is the second preferred result, we terminate the process and output image i^* as the exact match. Otherwise, we will look into the Homography for affine transformation.

Remember that our imaging system has a fixed, pre-calibrated downward facing camera. Therefore, the image captured has been corrected for perspective distortion. That means the Homography between captured image and its matching image should be very close to similarity transformation, which includes scaling, rotation, and translation. Therefore, we obtained the affine score A_i from H_i.

$$A_i = |H_i(3,1)| + |H_i(3,2)|. \tag{3}$$

Then the candidate with smallest affine score will be our final exact matching image.

WSR helps significantly to identify the document image with similar templates with the natural images that include repeating information. Some of our test cases are images with identical content but located in different places, or include same dominant features. In these cases, affinity test shows good discriminative power.

B) Handwritten annotation extraction

Extracting handwritten annotations in document images is an important research topic in document imaging. Assume the document is from a known database, image retrieval technology will find the original image. With removal of the pre-printed content, it can lead to a higher data compression rate, make content sharing, organizing, and archiving much easier. In METIS, the handwritten annotation extraction algorithm enables digitizing handwritten notes, content sharing, and co-creation experience mentioned in the previous section.

Most previous work has been focusing on document image obtained by flatbed scanners, assuming there is no perspective distortion [23]. Also, much work has been on form dropout rather than general annotation extraction

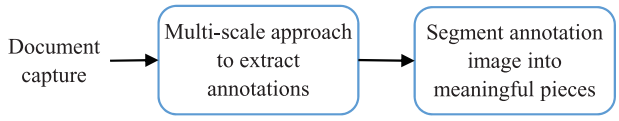


Fig. 10. High-level Flowchart of Handwritten Annotation Extraction Algorithm.

[24, 25]. Form dropout systems normally make assumptions about the special structure in the document and annotations. Also, they usually only consider global registration by matching preprinted lines. However, for general documents with lots of texts, this is not enough. Some other work that addresses general annotation ex-traction calculate local displacement vectors for image blocks on a uniform grid in order to achieve better registration results [26]. But this may break the annotations into parts and result in artifacts.

There are lots of work on general image registration as well [27, 28]. They are mainly based on feature point registration. While it is good for getting an initial alignment between two images, annotation extraction requires higher registration accuracy than general image registration.

We propose a multi-scale approach to extract handwritten annotations from the document images in a meaningful way [29]. It compares the rectified captured image with the original image, which can be obtained through image retrieval technology at various resolutions in order to remove the noise caused by non-uniform distortions, such as camera lens distortion and document surface curvature, while still preserving the true annotations. It also provides users substantial flexibility with a voting scheme and potentially different weights at different resolution levels. In addition, we analyzed the final annotation image and found the meaningful pieces out of it, such as an image patch of a handwritten paragraph. This broadens the application of annotation extraction, and makes it easier to share the notes.

With the fixed, pre-calibrated camera in our system, we can reliably assume no perspective distortions on the captured document image, and other types of non-linear distortion are very small, such as camera lens distortion and surface curvature. Also, we assume the images are captured under reasonable lighting conditions. Given the current digital cameras and normal reading/school environment, these are reasonable assumptions. Our system handles general document annotation extraction, using both feature points and image content comparison. The workflow is summarized in Fig. 10. An example of the experimental results is shown in Fig. 11.

When we integrated the METIS Immersive Learning system on Sprout, we designed an interactive experience that allowed users to combine the extracted notes, and to choose to upload into our system or share with the team. Figure 12 shows the interface that supports this experience.

V. CONCLUSIONS AND FUTURE WORK

In light of the societal and economic trends, we asserted that the 21st century needs require us to be strongly grounded

https://doi.org/10.1017/ATSIP.2017.10 Published online by Cambridge University Press

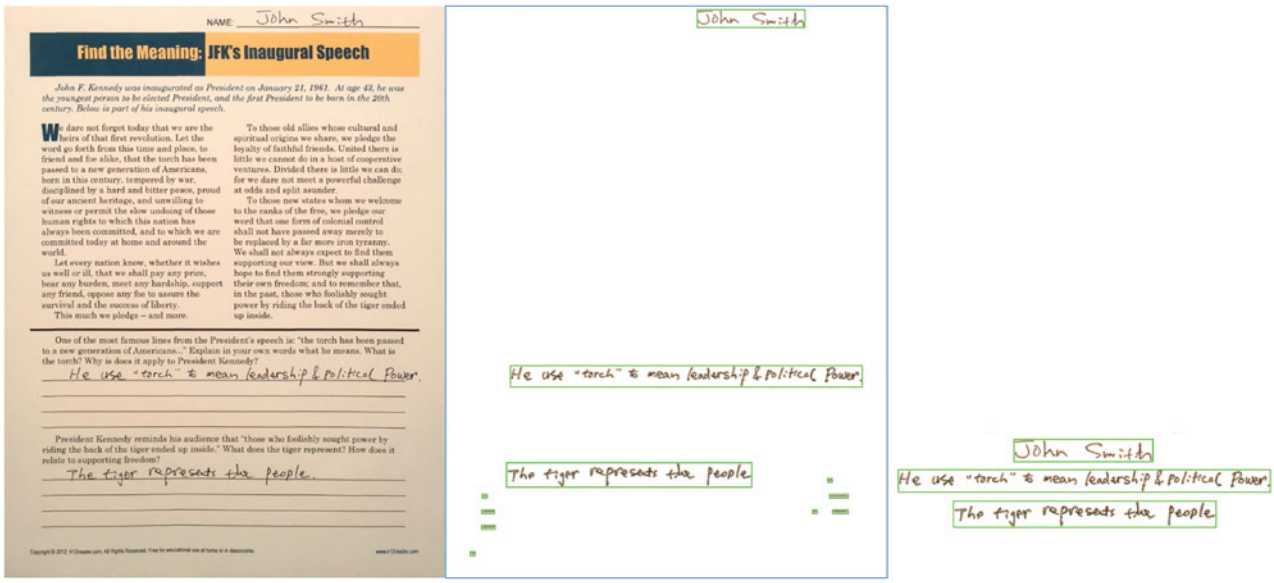


Fig. 11. Experimental results. Left: captured image. Middle: extracted meaningful handwritten notes. Right: zoomed-in results.

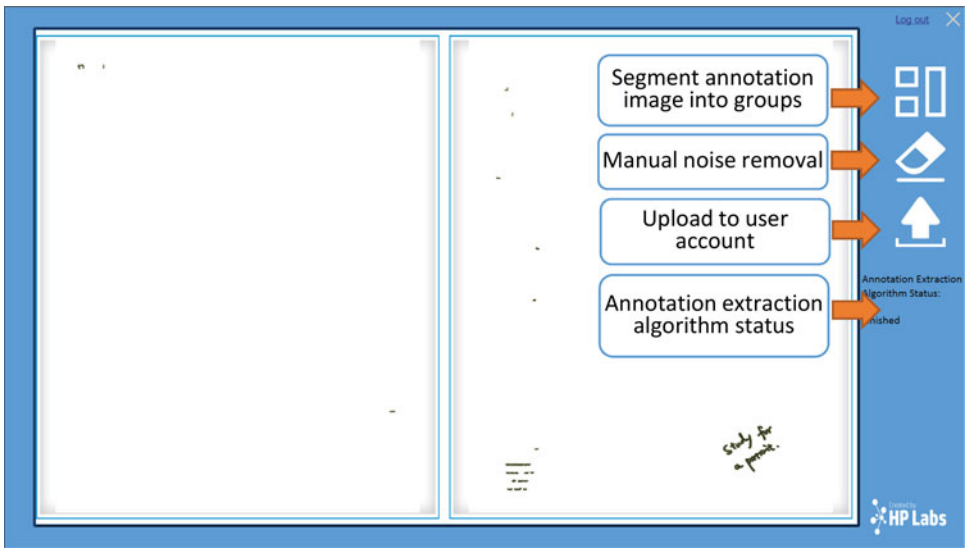


Fig. 12. Interface for the interactive experience of handwritten annotation extraction.

in fundamentals while taking advantage of the wealth of information available on the Internet. In that context, we motivated a cyber-physical hybrid learning system that combined traditional book based learning with web based information sharing and collaboration. We presented the architecture and physical design of a cyber-physical learning platform and articulated the solutions students can use, given the capabilities to scan books and objects, avail information from the Internet, and collaborate with others. Among the many features of the system, annotation of books was shown as an important feature. Not only can the annotation capabilities of authoritative content in physical books serve as metadata for future reference, they can also serve as useful information to the community. As an example, while studying the root cause of catastrophic failure of Tacoma Narrows Bridge in an authoritative book such as Den Hartog's Mechanical Vibrations, the user can annotate the information. And, in parallel, the user may combine the content with selected videos from the large body of content available on the Internet. This overall body of information is useful as future reference to the user, her peers, and the community. Furthermore, a classroom full of students doing the same exercise can lead to broader community contribution under the supervision of the teacher.

We believe that the flexible, configurable cyber-physical learning platform we have demonstrated paves the way for a number of application and research threads:

- Support assessment as a feedback to devise personalized learning for students.
- Lead to revisions in textbooks, and novel ways of learning, by teacher curation of annotation by students.
- Apply the platform to textually represent images and thereby create a database of images that can be retrieved with text search.

https://doi.org/10.1017/ATSIP.2017.10 Published online by Cambridge University Press

- Novel ways of collaborating through the scanning and sharing physical objects, particularly in light of the growth in 3D printing.
- Creating video tutorials of complex devices while referencing authoritative texts and Internet-based content.

REFERENCES

[1] Pittman, K.: Meet China's Bridge Building Robot. www.Engineering.com, October 27, 2015.

[2] Myrberg, C.; Wiberg, N.: Screen vs. Paper: what is the difference for reading and learning? Insights, 28 (2015), 49–54.

[3] Patel, C.: Sustainable ecosystems: enabled by supply demand management, in Int. Conf. on Distributed Computing and Networking (ICDCN 2011), Bangalore, India, 2011, 12–28.

[4] Liu, L. *et al.* METIS: a multi-faceted hybrid book learning platform, in The 16th ACM Symp. on Document Engineering (DocEng 2016), Vienna, Austria, 2016, 31–34.

[5] Hailpern, J. *et al.* To print or not to print: Hybrid learning with METIS learning platform, in The 7th ACM SIGCHI Symp. on Engineering Interactive Computing Systems (EICS 2015), Duisburg, Germany, 2015, 206–215.

[6] Vernica, R.; Damera Venkata, N.: AERO: an extensible framework for adaptive web layout synthesis in The 2015 ACM Symp. on Document Engineering (DocEng 2015), Lausanne, Switzerland, 2015, 187–190.

[7] Liu, L.; Liu, J.; Wu, S.: Image discovery and insertion for custom publishing, in The 9th ACM Conf. on Recommender System (RecSys 2015), Vienna, Austria, 2015. http://dblp.org/db/conf/recsys/posters2015

[8] Liu, L.; Koutrika, G.; Wu, S.: LearningAssistant: a novel learning resource recommendation system in The 31st IEEE Int. Conf. on Data Engineering (ICDE 2015), Seoul, Korea, 2015, 1424–1427.

[9] Wang, S.; Liu, L.: Prerequisite concept maps extraction for automatic assessment in The LILE Workshop in 25th Int. World Wide Web Conf. (WWW 2016), Montreal Canada, 2016, 519–521.

[10] Sprout Pro by HP G2. http://sprout.com.

[11] Rui, Y.; Huang, T.S.: Image retrieval: current techniques, promising directions, and open issues. J. Vis. Commun. Image Represent., 10 (1999), 39–62.

[12] Swets, D.L.; Weng, J.J.: Using discriminant eigenfeatures for image retrieval. IEEE Trans. Pattern. Anal. Mach. Intell., 18 (1996), 831–836.

[13] Gudivada, V.N.; Raghavan, V.V.: Content based image retrieval systems. Computer, 28 (1995), 18–22.

[14] Schmid, C.; Mohr, R.: Local gray value invariants for image retrieval. IEEE Trans. Pattern Anal. Mach. Intell., 19 (1997), 530–534.

[15] Lawrence, S.R.; Giles, C.L.: Meta search engine. US Patent 6,999,959, February 14, 2006.

[16] Smeulders, A.W.; Worring, M.; Santini, S.; Gupta, A.; Jain, R.: Content-based image retrieval at the end of the early years. IEEE Trans. Pattern Anal. Mach. Intell., 22 (2000), 1349–1380.

[17] Yang, M.; Kpalma, K.; Joseph, R.: A survey of shape feature extraction techniques, in Pattern Recognition Techniques, Technology and Applications, InTech, 2008, 43–90.

[18] Deb, S.; Zhang, Y.: An overview of content-based image retrieval techniques, in The 18th Int. Conf. on Advanced Information Networking and Applications (AINA 2004), Fukuoka, Japan, 2004, 59–64.

[19] Giacinto, G.: A nearest-neighbor approach to relevance feedback in content based image retrieval, in The 6th ACM Int. Conf. on Image and Video Retrieval, Amsterdam, The Netherlands, 2007, 456–463.

[20] Csurka, G.; Dance, C.; Fan, L.; Willamowski, J.; Bray, C.: Visual categorization with bags of keypoints, in The workshop on Statistical Learning in Computer Vision, ECCV, Prague, 2004, 1–22.

[21] Lowe, D. G.: Object recognition from local scale-invariant features, in The 7th IEEE Int. Conf. on Computer Vision (ICCV), Kerkyra, Greece, 1999, vol. 2, pp. 1150–1157.

[22] MacQueen, J.: Some methods for classification and analysis of multivariate observations, in The 5th Berkeley Symp. on Mathematical Statistics and Probability, Oakland, CA, USA, 1967, vol. 1, no. 14, pp. 281–297.

[23] Dubois, E.; Pathak, A.: Reduction of bleed-through in scanned manuscript documents, in IS&T Image Processing, Image Quality, Image Capture Systems Conf., Montreal, Canada, 2001, 177–180.

[24] Safari, R.; Narasimhamurthi, N.; Shridhar, M.; Ahmadi, M.: Form registration: a computer vision approach, in IEEE Int. Conf. on Document Analysis and Recognition (ICDAR), Ulm, Germany, 1997, vol. 2, pp. 758–761.

[25] Mao, J.; Mohiuddin, K.: Form dropout using distance transformation, in IEEE Int. Conf. on Image Processing (ICIP), Washington, DC, USA, 1995, vol. 3, pp. 328–331.

[26] Ye, M.; Bern, M.; Goldberg, D.: Document image matching and annotation lifting, in IEEE Int. Conf. on Document Analysis and Recognition (ICDAR), Seattle, WA, USA, 2001, 753–760.

[27] Isgro, F.; Pilu, M.: A fast and robust image registration method based on an early consensus paradigm. Pattern Recognit. Lett., 25 (2004), 943–954.

[28] Yang, J.; Blum, R.S.; Williams, J.P.; Sun, Y.; Xu, C.: Non-rigid image registration using geometric features and local salient region features, in IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR), New York, NY, USA, 2006, vol. 1, pp. 825–832.

[29] Lei, Y.; Fan, J.; Liu, J.: A multi-scale approach to extract meaningful annotations from document images, in IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, 2016, 1951–1955.

Chandrakant Patel is currently HP's Chief Engineer and Senior Fellow. He has led HP in delivering innovations in chips, systems, data centers, storage, networking, print engines and software platforms. He is a pioneer in thermal and energy management in data centers, and in application of the information technology to drive available energy management at city scales. Chandrakant is an ASME and an IEEE Fellow, and has been granted 151 patents and published more than 150 papers. An advocate of return to fundamentals, he has served as an adjunct faculty in engineering at Chabot College, U.C. Berkeley Extension, San Jose State University and Santa Clara University. In 2014, Chandrakant was inducted into the Silicon Valley Engineering Hall of Fame.

Dr. Yang Lei is a Senior Research Scientist at HP Labs, Palo Alto, California, leading the imaging and 3D vision systems design and development in multiple projects. Her research interests are in digital imaging and computer vision systems and their applications in the industry, including large-scale image recognition and understanding, 3D object scanning and retrieval, etc. Yang has published multiple articles in IEEE and ACM conferences, such as IEEE ICIP,

https://doi.org/10.1017/ATSIP.2017.10 Published online by Cambridge University Press

ICASSP, ACM DocEng, and Journal of Imaging Science and Technology (JIST). She was invited to serve as Program Co-Chair, organizing committee member, or program committee member in major international conferences in imaging and multimedia field, for example IEEE BigMM, IEEE ISM, etc. Yang received her Ph.D. in Electrical and Computer Engineering from Purdue University, West Lafayette.

Dr. Lei Liu is currently a principle research scientist in Huawei, USA. Leading the architecture design and development of machine learning systems in multiple projects. He received his Ph.D. degree in Computer Science and Engineering at Michigan State University. His research interests cover multiple aspects of large scale data mining and machine learning, including large multi-class classification, zero-day malware detection, recommendation system, data driven education and data mining with wearable devices. Lei has published over 30 referred articles in top conferences and journals, including ICDE, SDM, INFOCOM, WWW, CIKM, RecSys, ASONAM, etc. He was invited to serve as Chair, Senior PC or PC for international conferences in data mining and machine learning, such as WWW, CIKM, ECML-PKDD, ASONAM, ICHI, etc. He is an inventor of more than 30 granted or applied US patents.

Rares Vernica is a Senior Research Scientist at HP Labs. He is part of the Print Adjacencies & 3D Lab where he works on developing new systems and prototypes for enabling seamless access to digital and physical information. RaresŠ main area of expertise is data management and system scalability. He has architected and led various efforts in HP Labs around user analytics, multimedia analytics, scalable book publishing, hybrid learning, and rendering engines. Rares Vernica received a Ph.D. in Computer Science from University of California, Irvine in 2011 and a B.Sc. from Politehnica University of Bucharest, Romania in 2004.

Jian Fan is currently a principal research engineer at HP Labs. Jian holds a Ph.D. degree in computer engineering from the University of Florida. His research interests include image processing, document image processing and computer vision.

Brad Short is an HP Distinguish Technologist and currently the Chief Technologist & Product Experience Architect in the Immersive Computing Group at HP. He is responsible for developing future generation computer products, interactive experiences, human-computer interfaces, and the required technologies. Brad has received over 45 patents (over 20 granted, and over 25 more pending). Brad earned a Master of Science degree in Mechanical Engineering and a Bachelor of Science degree in Applied & Engineering Physics from Cornell University, Ithaca, NY.

Jerry Liu is currently a Senior Research Manager at HP Labs, Palo Alto, California. His research interests are in data analysis and sensor systems, with over 20 issued patents in these fields. At HP Labs, Jerry manages teams of scientists and engineers working on life science, computer vision, and predictive analytics, working to develop and drive differentiating technologies to market. Jerry earned his Master of Engineering degree and Bachelor of Science degree in Electrical Engineering from Cornell University, Ithaca, NY.

Steve J. Simske is an HP Fellow and a Research Director in HP Labs. He has led HP in delivering innovations in algorithms, multi-media, labels, brand protection, imaging, 3D printing, analytics and life sciences. He is a long-time member of the World Economic Forum Global Agenda Councils, leads the Steering Committee for the ACM DocEng Symposium, and is currently President of the Imaging Science and Technology professional organization. Steve is an IS&T Fellow, and has more than 160 granted US patents and more than 400

https://doi.org/10.1017/ATSIP.2017.10 Published online by Cambridge University Press