A REAL-TIME PRODUCTION TOOL FOR ANIMATED HAND SKETCHES

J. Loviscach[1]

[1] Fachhochschule Bielefeld (University of Applied Sciences), Wilhelm-Bertelsmann-Str. 10, 33602 Bielefeld, Germany
joern.loviscach@fh-bielefeld.de, Fax +49 521 106 7160

Abstract

The look of hand-drawn sketches has become fashionable in video production. This paper introduces a software tool to produce corresponding videos in real-time during lectures or presentations, at the desk or in the studio. Currently, two styles are available. In the first style, a hand seems to draw on a whiteboard. In the second style, the presenter seems to stand behind a transparent board, which is simulated with the help of a camera. In both cases, the input stems from a standard graphics tablet. The image of the lecturer's arm is synthesized from photographs and animated through inverse kinematics; the sound of the pen and the eraser is synthesized from recordings. Auxiliary functions include a ghosted script for the presenter and drag&drop of graphical elements prepared in advance.

Keywords: Pen-based computing, video production, inverse kinematics, matting

1 Introduction

Sketching and writing shown in progress—either as stop-motion animation or as actual real-time recording—have gained considerable prominence in visual media. In particular, the "RSA Animates" [13] series by Cognitive Media has attracted tens of millions of viewers. These are high-caliber talks illustrated with a hand drawing on a whiteboard; no talking head is shown. This can be considered a cleaned-up version of the style of thousands of mathematics education videos recorded by enthusiasts and uploaded onto YouTube. One mode of the tool presented in this paper creates this look, see Figure 1.

Another incarnation of the "a writing/drawing hand" topos can be seen in the transparent boards that populate police offices and forensics labs in the movies. In actual use, such boards would be highly impractical, as the uncontrollable background makes the writing hard to read. In reality, transparent boards stem from command centers where intelligence staff collects data and writes them on the board in mirror script, which commanding officers read from the *opposite* side to do the planning. Hence, transparent boards even make no sense on the bridge of a science-fiction starship; the data could better be displayed on an opaque surface.

The movie "A Beautiful Mind" has brought another instance of writing onto transparent backgrounds in the audience's mind: a mathematician who scribbles equations on window panes. This may be the model for scientists who use the glass walls of their offices as boards, see Figure 2.

Despite their practical limitations, the high-tech look or—in the "A Beautiful Mind" variant—the nerdy look of transparent boards does not seem to vanish from movie production. In the context of this paper, a *simulated* transparent board actually has clear advantage over a standard board: The presenter does not occlude the image, neither with his or her body not with his or her hand, see Figure 3. In addition, the virtual transparent board supports fine-grained control of the background so as to diminish the detrimental effect it may have on legibility. The second mode of the tool presented in this paper grabs a real-time camera image of the presenter and applies an automatically generated matte to separate the presenter's image from the background, which is replaced by black color. This and a strong reduction of the brightness of the image of the lecturer ensure high legibility of the strokes in the foreground. The virtual transparent board enables doing this without the presenter having to write in mirror script.

This paper describes a tool developed to create visual media in the two styles described above—and potentially others, as the definition of a style is a pluggable extension attached to the software's core. The use of the software is threefold: First, it provides real-time visuals and sound for the audience at a live event or for a broadcast, e. g. the weather report. Second, it can record these for archival or for distribution on the Internet. Third, it can serve as a tool to speed up production in the studio, for instance by replacing stop-motion animation workflows. In all these situations, using an electronic board and synthesizing the hand as opposed to actually photographing simplifies recording drastically, in particular as no sophisticated lighting and camera setup—if any at all—are required and the lettering always has perfect sharpness and contrast.

This paper and the tool it describes are a largely extended version of [10]. The paper is structured as follows: Section 2 describes related work; Section 3 introduces the visual elements that make up the final result. The synthesis of the image and the animation of the hand and the arm are explained in Section 4, whereas Section 5 provides details on the synthesis of the sound that accompanies the pen and eraser

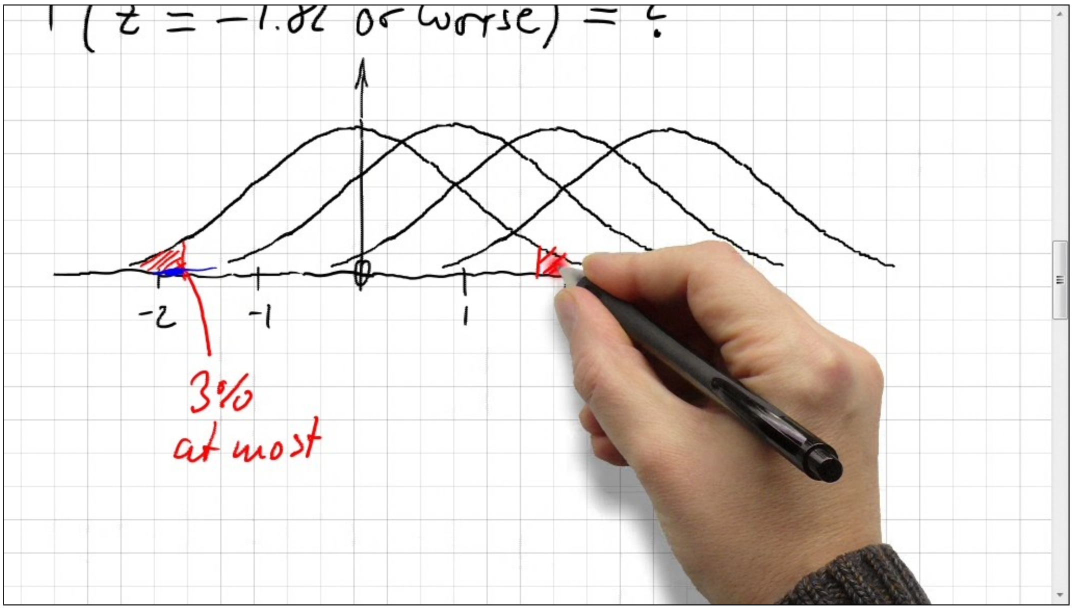

Figure 1: The first style offered by the prototype consists of simulated hand writing on a whiteboard.

Figure 2: Writing on glass panes has found its way from Hollywood into regular offices (snapshot taken by the author).

tools. Section 6 presents the ideas behind the user interface aimed at live action recording. Further functionality provided for comfort is outlined in Section 7. Section 8 covers the implementation of the prototype; Section 9 concludes this paper and gives an outlook on future developments.

2 Related work

Illustration-based presentation techniques are becoming mainstream, as can be seen from bestselling books on the subject [11]. The combination of hand sketches with the automatic processing offered by a computer has long been explored for at least a decade, see e. g. [9]. In addition, internet video has popularized the genre of speed drawing. Here, a time-lapse screencast of graphics software shows the process of an artist creating an image. Web sites such as `www.queeky.com` already offer complete online support to create and watch speed drawings.

Semi-transparent interfaces had first been studied for icon toolboxes [4]. Today, their application in experimental systems for computer-supported cooperative work is not uncommon. ClearBoard [5], a seminal work, used a complex camera setup to enable users to share a virtually transparent board between two locations. Over the course of time, this idea has been reimplemented with the current tools of the trade, see e. g. [12].

To reduce clutter, one may extract only the arms of the collaborators from a camera's image stream and superimpose them over a shared document [15]. Another option is to extract the image of the upper part of the collaborator's body [8]. In a classroom situation, the camera may capture the lecturer and parts of an electronic blackboard. Hence, extracting the image of the lecturer becomes more involved [2]. As discussed in that paper, the benefit of compositing the lecturer's image onto the screencast of the board is that the viewer does not need to split his or her attention: In many lecture videos, the lecturer distractingly appears in a second frame.

Background removal and edge enhancement [3] can further simplify the image over which the main content is superimposed. Taking this one step further, the full-color image may be replaced by a shadow, which also helps to lower the required data bandwidth [1]. On a related note, projected virtual shadows may be used for interaction with wall-sized displays that are taller than a human [14].

Graphics tablets enable intuitive real-time animation input. For instance, one may redefine the timing of an animation through strokes on a tablet [16] or use two styluses on one tablet to generate walking motion [7]. Even more parameters can be controlled in real time through multitouch interaction [6].

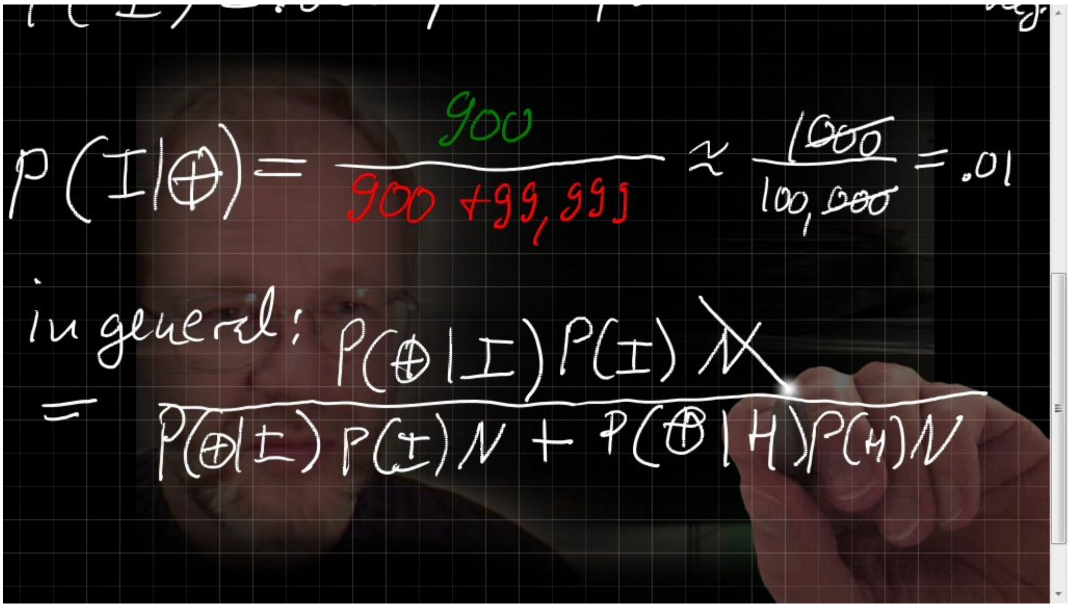

Figure 3: The second visual style offered by the prototype simulates a transparent board onto which the presenter—who is visible in the background—seems to be drawing from the opposite side.

3 Visual elements

The most prominent visual aspect of the software is the ink layer. This can be extended vertically to virtually arbitrary length. A scroll bar appears on the right—as an interface element for the presenter, but at the same time as visual feedback to the viewer to highlight the operation of scrolling. A square grid scrolls in sync with the ink layer. It acts as an aid for writing and drawing, provides further clues during scrolling, and—in the second style—emphasizes the effect of the transparent board, as it appears in front of the presenter's image.

In the first style (whiteboard), the arm with the pen or eraser and the hand appear in front of the ink layer, together with a soft semitransparent shadow. In the second style (transparent board), the arm and the hand appear behind the ink layer. In that style, there is no specific image of the hand with an eraser tool; rather, the palm of hand serves to remove ink. The currently selected drawing color is indicated by a soft halo around the tip of the pen image. When the presenter touches the tablet with the stylus, this halo becomes more opaque.

The windows of other software can be dragged in front of the sketch. The corresponding changes of the mouse cursor are recorded in the video, see Figure 4. Note that the bare display memory does not contain the mouse cursor.

In the second style—the transparent board—a camera is used to capture the presenter. The pixel region that forms his or her image is found through background subtraction. This requires a static background and switching off the camera's automatic exposure control and focusing. To better cope with the imprecision and the jagged edges occurring in the process, the result is blurred to create a soft matte and superimpose the presenter onto a black background.

The camera image is mirrored left-to-right to give the presenter the impression of looking into a mirror. A non-mirrored image would be too unfamiliar and hence confusing. In a similar vein, the virtual writing hand is the *left* hand of the person visible behind the screen. Otherwise, the virtual arm would reach into the video frame from its left side, but the presenter—if right-handed—would confusingly see his or her real arm extending to the right side of the graphics tablet.

4 Synthesis of the hand

The inputs to the simulation of the writing hand are the following: the screen position of the stylus, whether the stylus is held normally (as a pen) or inverted (as an eraser), and whether or not the stylus is in the sensing range of the tablet. The latter indicates a fundamental issue when driving a simulation with a graphics tablet: When the stylus is lifted by about one centimeter, the tablet can no longer detect it and hence does no longer provide location data. The simulation, however, requires a steady stream of location data. Hence, in such a case the system resorts to inventing a plausible location from ballistic motion: The reference point—which otherwise is linked to the tip of the stylus—falls to the bottom of the visible region with constant acceleration downward, starting at zero speed.

The reference point drives the inverse kinematics of two rigid components: the hand and the arm. The hand can rotate by 60 angular degrees around the wrist joint. Any other motion causes the arm to be dragged along with the hand.

The arm upward from the wrist and the hand are each represented by fixed images, see Figure 5; the image of the hand changes according to whether the pen or the eraser is currently being used. The transition between the images of the hand and the stump of the arm is concealed by the sleeve (whiteboard style) or by feathering the hand around the wrist (transparent board style). To create the illusion of a shadow in the former style, two additional images are coupled to the hand and the arm and are rendered below them, blurred and in semitransparent black. The overall size of the hand and the arm as well as their color can be adjusted interactively.

5 Audio synthesis

To add a touch of realism, sound is synthesized for the pen and the eraser tool. This is based on actual recordings of the "fffsh" sound produced with a light stroke of a felt pen on a whiteboard and the "eeek" sound generated by a strong stroke. For each of these two kinds, a dozen of such sounds has been recorded and edited into a seamless loop of approximately one second's length, which turned out to be far enough to hide the repetition when the loop repeats during a long stroke.

The control input to the audio synthesis stems from the graphics tablet. It outputs the current position and speed of the stylus and the pressure currently sensed by its tip. The speed and the pressure control the overall volume and the blending of the two base sounds. In addition, the speed has a small influence on the pitch of the sounds: The higher the speed, the higher the pitch. The position of the stylus is used to control the left–right placement in the stereo panorama, see Figure 6. All these mappings from the control input to the physical parameters of the sound are adapted through hand-tuned functions. For the eraser tool, these create different settings that focus on the "eeek" part.

During recording, the synthesized sound is only stored in an audio file, but not fed to the computer's speakers. Otherwise, it would crosstalk into the signal of the microphone used to record the voice.

***Figure 5:** The whiteboard style uses five image to synthesize the hand, the arm and their shadow (top three rows); the transparent board requires three images (bottom two rows).*

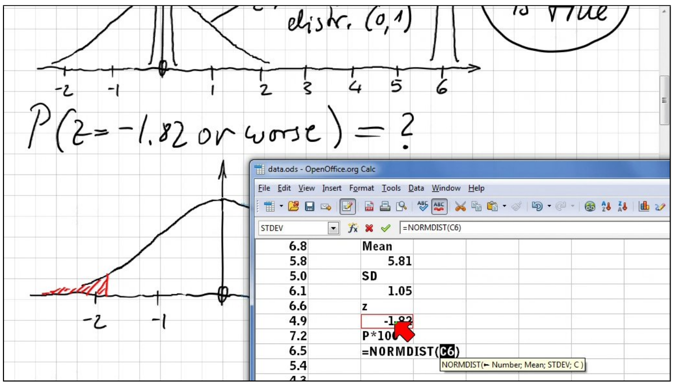

***Figure 4:** Windows of other software can be brought in front. They are recorded with the appropriately shaped mouse pointer.*

These sounds are inspired by the felt pen sound effects in the title sequence of the "RSA Animates" [13]. Using them continuously throughout a video requires keeping them at subtle audio levels. The comments that the author's first video with this kind of synthesized pen/eraser sound received on YouTube were very vocal about this.

6 User interface

The user interface is tailored to live action. There are no menus; every function is invoked by pressing a single specific key. Even actions such as load/save and undo/redo that regularly require two simultaneous key presses (Control+...) are handled with a single key each, thus enabling the presenter to operate the keyboard with one hand while keeping the stylus in the other.

The stylus has four modes of operation: draw; erase (using the opposite end of the stylus; pressing the space bar switches from erasing locally to immediately removing every stroke touched by the tool); select and translate/scale (pressing the Shift key and using the stylus as a lasso for selection and then to adjust the position and the size of the selection); select the strokes, remove them from the board and insert them into a new compartment on an indefinitely long visual clipboard (pressing the Control key and using the stylus as a lasso for selection).

The clipboard is placed below the lower end of the recorded part of the screen and hence remains invisible in the video. Shapes stored here can be dragged onto the visible image, see Figure 7. The clipboard may be used for mathematical charts and curves as well as maps and symbols required for instance for a weather report. Inserting such "clipart" as hand-drawn strokes instead of clean bitmap or vector graphics helps to save the look of a sketch. To not occlude the shape dragged from the clipboard by the image of the hand, the shape is grabbed at its rightmost point.

7 Helper functions

Regular presentation software shows notes in a separate frame on the presenter's private screen. This causes another instance of the "split attention" effect, in this case affecting the presenter, not the audience. In addition to this effect, notes shown as text do not help much when the task is to sketch an illustration in real time and while talking. This paper proposes to solve this issue by a kind of teleprompter: The presenter can prepare text, drawings and stage directions and see them superimposed on the actual content during the live performance. These notes appear in a ghosted fashion, see Figure 8, scroll in sync with the content and are automatically removed from video recording.

Another function added for convenience is to set the drawing window to a semitransparent state, see Figure 9. This way, the user can trace diagrams, symbols etc. with the stylus, to create

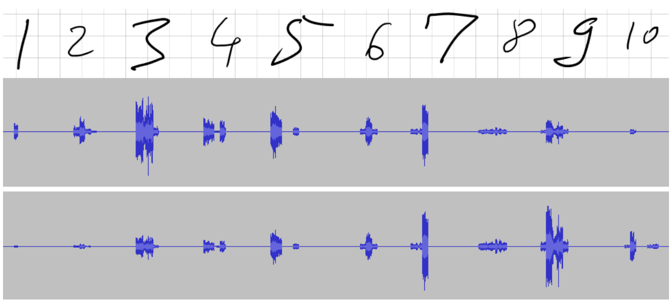

Figure 6: Digits written from left to right at different pressure levels show the changes in audio output (mid/lower row: waveform of the left/right audio channel).

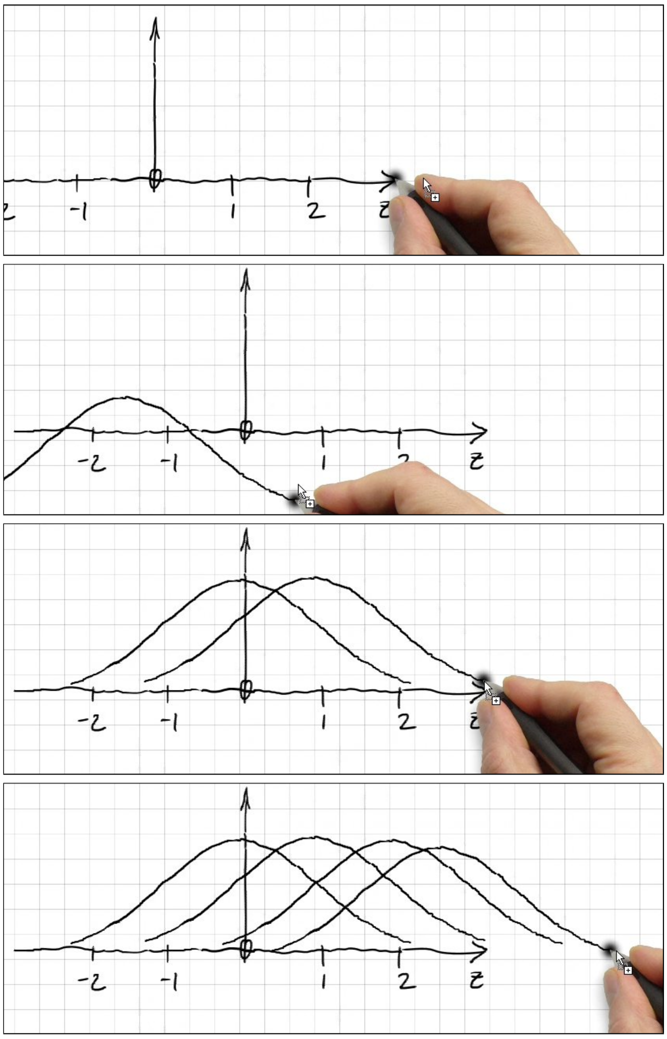

Figure 7: The user can drag prepared groups of strokes from a clipboard that remains invisible in the video and drop them onto the board.

templates for the "teleprompter" or to create artwork for the clipboard in an appropriate hand-sketched style.

8 Implementation and hardware

A webcam, a USB microphone and a Wacom Bamboo Fun graphics tablet are used as input devices. Experiments with webcams connected through USB and with webcams that are integrated into the bezel of a notebook computer's screen worked out well. Integrated webcams have a comparatively low vantage point. With them, the room's ceiling becomes visible in the image; columns and other architectural structures lead to converging lines in the background. These effects could be confusing to the viewer, but are, however, mostly removed by background subtraction.

Some TabletPCs—that is, Windows notebook computers with built-in graphics tablet screen—possess an integrated webcam. This is of limited use, as every motion of the stylus on the

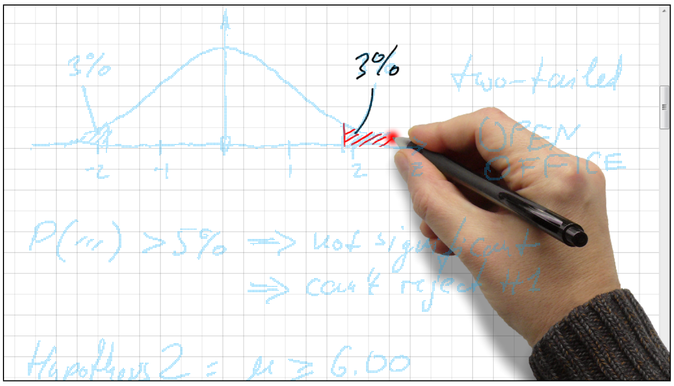

Figure 8: The ghosted view of notes prepared upfront turns the board into a teleprompter.

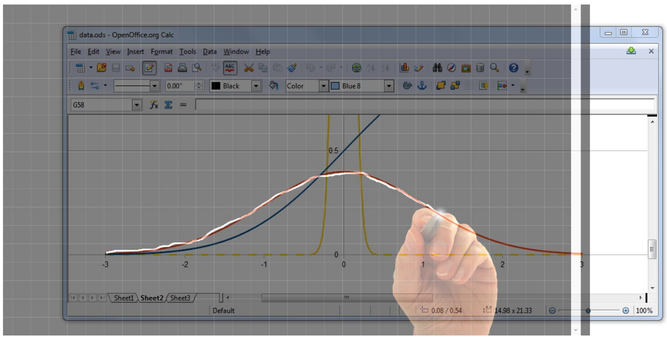

Figure 9: With the software's window switched to transparent mode, the user can trace the content of other windows.

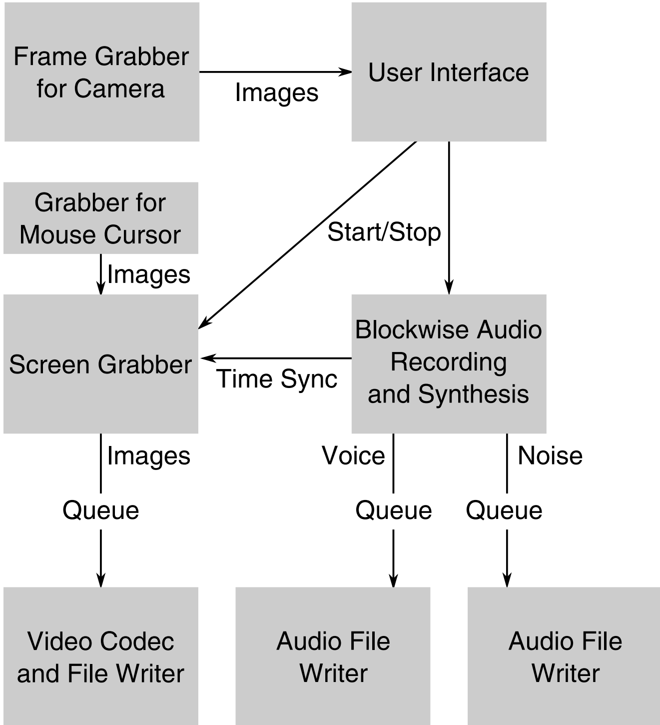

Figure 10: The workload is distributed onto eight threads (shaded rectangles). In particular, compressing and saving data is delegated to low-priority threads fed by long, non-blocking queues.

screen makes the screen and hence the camera shake. The user may even bring the stylus inadvertently into the camera's field of view.

The prototype software is implemented for Microsoft Windows 7 in the language C#. It makes heavy use of the amenities of the Microsoft .NET 4.0 framework, in particular the Windows Presentation Foundation (WPF) library for graphical user interfaces and stylus input.

To fully leverage a quad-core processor, the software operates in a highly multithreaded fashion, see Figure 10. On a computer powered by an Intel Core i7 2.8 GHz quadcore processor, the prototype can record 20 frames per second without dropped frames or dropped audio samples at a resolution of 1024×576 pixels. Currently, the frame rate seems to be limited by the rendering speed of the WPF library, possibly due to subtle efficiency issues with some of the classes used. Even though the WPF library already employs the graphics processor (GPU) for compositing and other effects, this signifies that the rendering should be handled more directly on the GPU.

A well-known (and well-feared) issue in recording a computer's screen on the same machine is losing the synchronization between audio and video. The prototype makes sure that at each point in time the number of video frames written is at least two less than the number that would be required as computed from the number of audio samples written.

9 Conclusion and outlook

This paper has presented a system to create video animations of a drawing and writing hand in real time. The ghosted notes that act as a teleprompter drastically reduce the need for rehearsing. The transparent board possesses a vital benefit for communication: The presenter's body and hand do not obstruct the view of the board.

This tool is particularly suited for visual explanations, for which mathematics lectures are just one example. Science TV shows, financial news and weather reports could equally benefit from illustrations in progress. For these uses, the clipboard may prove vital: Icons, company logos and other graphical elements can be prepared upfront to be brought onto the board quickly and with no clashing in the style.

Future work can improve on the details of the hand, for instance by using more than two elements and/or by deforming them in addition to translating and rotating them. A necessary color correction of the hand could be determined automatically rather than leaving Hue/Saturation/Value settings to the user. The background subtraction employed to bring out the face of the presenter may be made more robust by face recognition.

Gestures executed with the stylus to invoke commands could be offered in addition to keystrokes. A multitouch panel may be used to show both hands or even to animate several fingers independently.

Further visual elements can be integrated, for instance background videos—for instance satellite imagery for a weather report—and real-time 3D-sketching. The latter could consist in drawing a 2D profile curve which automatically is converted to a body of revolution. This would be rendered non-photorealistically with sketched strokes to fit to the hand-sketched content. Nonetheless, the 3D object could be rotated freely with the stylus.

A ruler to draw lines and a compass to draw circles can be added. These, too, would have to produce strokes that look hand-sketched. Likewise, the frames and the user interface elements of other windows brought in front of the board could be distorted to create a hand-made or even a sketched look. The look of strokes dragged from the clipboard onto the visible board would benefit from random variation. Currently, each new copy of such a stroke looks exactly like the original one. This is implausible for a hand sketch. (Moving the strokes by dragging them is not only im*plausible*, but would even be im*possible* in real life. However, this effect does not appear to be offensive to the viewer.) Finally, one could integrate an automated animation of hand-sketched diagrams similar to [9].

References

[1] Mark Apperley, Laurie McLeod, Masood Masoodian, Lance Paine, Malcolm Phillips, Bill Rogers, and Kirsten Thomson. Use of video shadow for small group interaction awareness on a large interactive display surface. In *Proceedings of the Australasian User Interface Conference AUIC '03*, pages 81–90, 2003.

[2] Gerald Friedland and Raul Rojas. Anthropocentric video segmentation for lecture webcasts. *J. Image Video Process.*, pages 9:1–9:10, 2008.

[3] Karl Gyllstrom, Dorian Miller, and David Stotts. Techniques for improving the visibility and "sharability" of semi-transparent video in shared workspaces. In *Proceedings of the ACM Southeast Regional Conference ACM-SE 45*, pages 425–430, 2007.

[4] Beverly L. Harrison, Gordon Kurtenbach, and Kim J. Vicente. An experimental evaluation of transparent user interface tools and information content. In *Proceedings of the ACM Symposium on User Interface and Software Technology UIST '95*, pages 81–90, 1995.

[5] Hiroshi Ishii and Minoru Kobayashi. ClearBoard: a seamless medium for shared drawing and conversation with eye contact. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems CHI '92*, pages 525–532, 1992.

[6] Michael Kipp and Quan Nguyen. Multitouch puppetry: creating coordinated 3D motion for an articulated arm. In *Processdings of the ACM International Conference on Interactive Tabletops and Surfaces ITS '10*, pages 147–156, 2010.

[7] Philipp Kolhoff, Jaqueline Preuß, and Jörn Lovisach. Walking with pens. In *Eurographics 2005 Short Presentations*, pages 33–36, 2005.

[8] Andreas Kunz, Thomas Nescher, and Martin Kuchler. Collaboard: A novel interactive electronic whiteboard for remote collaboration with people on content. In *Proceedings of the International Conference on Cyberworlds CW '10*, pages 430–437, 2010.

[9] Joseph J. LaViola, Jr. and Robert C. Zeleznik. MathPad^2: a system for the creation and exploration of mathematical sketches. *ACM Trans. Graph.*, 23:432–440, 2004.

[10] Jörn Loviscach. Turning a graphics tablet into a transparent blackboard. SIGGRAPH 2011 Poster, 2011.

[11] Dan Roam. *The Back of the Napkin: Solving Problems and Selling Ideas with Pictures*. Portfolio, Penguin, NY, 2008.

[12] Ian Robinson, Kar Han Tan, Ramin Samadani, Bruce Culbertson, and John Apostolopoulos. A laser-based system for through-the-screen collaboration. In *ACM SIGGRAPH 2010 Talks*, page 49:1, 2010.

[13] Royal Society for the Encouragement of Arts, Manufactures and Commerce. RSA Animates. `http://comment.rsablogs.org.uk/videos/`, last checked 2011-06-27, 2011.

[14] Garth Shoemaker, Anthony Tang, and Kellogg S. Booth. Shadow reaching: a new perspective on interaction for large displays. In *Proceedings of the ACM Symposium on User Interface Software and Technology UIST '07*, pages 53–56, 2007.

[15] Anthony Tang, Carman Neustaedter, and Saul Greenberg. VideoArms: Embodiments for mixed presence groupware. In *Proceedings of the British HCI Group Annual Conference HCI 2006*, pages 85–102, 2006.

[16] Sílvio César Lizana Terra and Ronald Anthony Metoyer. A performance-based technique for timing keyframe animations. *Graph. Models*, 69:89–105, March 2007.