Researchers at the Advanced Digital Sciences Center (ADSC) have achieved a major advance in the state of the art for real-time, robust video object cutout, and demonstrated this advance in a lightweight video chat system.

This work is part of ADSC’s broader program for achieving low-cost, low-power, realistic, and flexible audiovisual telepresence to support natural and seamless communication between individuals through a wide variety of media. This problem is a grand challenge in itself, as acknowledged by the inclusion of virtual reality in the U.S. National Academy of Engineering’s list of 14 grand challenges for engineering in the 21st century [NAE11].

1 Audiovisual Telepresence Challenges
When Alice and Bob attend a meeting in a virtual meeting room, in fact Alice may be seated in a coffee shop and Bob may be at home. To Bob, it must appear as though Alice is in the meeting room, not in the coffee shop. In other words, the telepresence system must remove the busy restaurant from the background of Alice’s camera image, and replace it by the appropriate part of the virtual meeting room. Although it is easy for humans to ignore a visual backdrop, no known algorithm can do a decent job of cutting out Alice from her backdrop in real time, using any existing camera and a conventional processor – not even with hardware assistance in the form of a GPU or FPGA – in spite of intense interest in both the industrial and research communities. For realistic telepresence, a videoconferencing system must solve this classical theoretical problem, known as *bi-layer video segmentation*.

Given this and the other challenges that must be addressed, it is not surprising that state-of-the-art teleimmersive systems, such as the National Tele-Immersion Initiative [NTII11] and the TEEVE project at the University of Illinois at Urbana-Champaign [TEEVE11], are typically very bulky, demand high bandwidth, and are too computationally intensive to run on an ordinary PC or laptop. These barriers seriously hinder their wide deployment in daily practice.

2 ADSC’s Real-time, Robust Video Object Cutout
As shown in Figure 1, ADSC has created and demonstrated a video object cutout technique that automatically segments a foreground layer from a live VGA or HD video stream in real time, using only a single webcam and a commodity desktop or laptop CPU [Lu11a]. A video of the cutout system in action can be accessed from ADSC’s web site, [http://adsc.illinois.edu](http://adsc.illinois.edu).

Grounded on a solid inference basis, the cutout algorithm probabilistically fuses different cues together into a unified optimization framework. The cutout approach includes many novel and robust design choices, which are key to achieving the high accuracy and good performance shown in the prototype implementation of the cutout algorithm. Further, the overall solution framework and some of its key components are applicable for real-time video object cutout with stereo cameras or depth cameras, another key area of interest at ADSC [LMPD11, MLD11a, MLD11b].

ADSC’s new approach to video object cutout advances the state of the art for this classical computer vision problem in three major ways.

- **Reliable segmentation with a single webcam.** ADSC’s approach works reliably with a single webcam, yet produces accurate results whose quality is comparable to those obtained with more costly or complex setups, such as stereo cameras, depth sensors, and flash and no-flash image pairs. As shown in Figure 2, ADSC’s approach is robust against intensity and color variations in the subject’s
appearance and a variety of background changes present in real-life applications, all of which previous methods find very challenging to handle: background and foreground with similar colors, dramatic movements in the foreground, a non-stationary background with moving objects, illumination changes, and shaky cameras.

• **Real-time performance.** ADSC’s cutout approach extracts foreground objects at unprecedented speeds on an ordinary PC, e.g., processing 18 frames per second at a 1280×960 resolution on a mainstream laptop CPU, with a CPU load of less than 40%. This contrasts with previous methods, which process even 160×120 or 320×240 low-resolution videos at only 7-15 frames per second, or entirely offline.

• **Ease of use.** As shown on the right-hand side of Figure 4, ADSC’s cutout approach requires simple and minimal user participation during system initialization, and is completely automated thereafter. These advantages make ADSC’s cutout approach appealing and practical for wide application in video conferencing, e-learning, and virtual reality.

3 ADSC’s Lightweight Teleimmersive Multiparty Video Chat System

ADSC has created and demonstrated a practical, lightweight teleimmersive video chat system with truly minimal equipment requirements: any webcam, PC, and the public Internet. Figure 3 shows a screen shot from this system, called CuteChat [LNNS11]. A video of CuteChat in action is available at ADSC’s web site, [http://adsc.illinois.edu](http://adsc.illinois.edu).

CuteChat leverages ADSC’s new video object cutout technology to provide a radically new video chat experience by merging two or more video conferencing participants onto a shared background environment. With the addition of a depth camera, CuteChat employs gesture recognition to allow the participants to interact with each other or with the selected background content in a natural and integrated manner, such as paging through presentation slides or online photo albums with a wave of the arm. With the goal of making the system easily accessible to the typical consumer, CuteChat addresses network bandwidth challenges, in addition to the need for single-webcam real-time video object cutout discussed above. To provide low-latency video communication over the public Internet, CuteChat employs novel
techniques for object-based video coding, which reduces bandwidth requirements and speeds up the video coding process, both by a factor of 3 to 4.

The current CuteChat implementation performs bilateral, Internet-based immersive video conferencing between two or more remote users, with VGA-sized (640×480) video object cutout, object-based encoding and decoding, and immersive video composition, all in real time on a commodity laptop or desktop PC. CuteChat clearly demonstrates unique advantages over current state-of-the-art telemersive systems, which, as discussed earlier, are typically bulky, require high Internet bandwidth, and are too computationally intensive to run on an ordinary PC. In addition, CuteChat is easier to use than previous systems, and provides better video quality under most conditions.

As shown on the left-hand side of Figure 4 and in Figure 5, ADSC has also created methods that allow users to perfect and beautify the appearance of their skin, or to have objects outlined in black and rendered in a limited range of colors, providing a cartoon-like or paint-by-numbers appearance. These
Figure 3 The CuteChat videoconferencing system. (a) One participant’s actual setting. On the screen, the participants are cut out and repositioned against a shared virtual backdrop, in this case a photo album that they are paging through using mouse clicks. (b) In this CuteChat screen view, one participant has been cut out and repositioned against a shared real background.

Figure 4 Left: CuteChat allows participants to beautify their skin. Right: The CuteChat initialization procedure simply requires users to align their bodies with an adjustable silhouette.

Figure 5 Top left: The input image and the same image with the background automatically blurred in real time. Top right: Colors smoothed and objects outlined. Bottom left: Foreground smoothed and outlined, with backdrop natural. Bottom middle: Foreground natural, with backdrop outlined and colors smoothed. Bottom right: Outline drawing for foreground, with backdrop natural.
facilities, called QuickToon [Lu11b], are incorporated into the CuteChat system and can be mixed and matched in the foreground image and the backdrop image.

About Us
The Advanced Digital Sciences Center (ADSC), established in 2009 and located in Singapore, is a wholly-owned subsidiary of the University of Illinois at Urbana-Champaign. ADSC’s research program in interactive digital media is funded by the Agency for Science, Technology, and Research (A*STAR).

Jiangbo Lu is a research scientist at ADSC and the lead researcher for ADSC’s video cutout work. Two of his inventions have been protected by Exploit Technologies (the commercialization arm of A*STAR) as trade secrets, with licensing and commercialization schemes under active development. Before joining ADSC, Jiangbo worked for IMEC, Microsoft Research Asia, and VIA-S3 Graphic. His Ph.D. is from Katholieke Universiteit Leuven in 2009.

Minh Do is an associate professor of electrical and computer engineering at the University of Illinois at Urbana-Champaign. Minh received a CAREER Award from the National Science Foundation. At Illinois, he has been a Beckman Fellow at the Center for Advanced Study and received a Xerox Award for Faculty Research. He is a coauthor of four papers that have won major awards at their respective venues, and is an associate editor for IEEE Transactions on Image Processing. Minh received the Dr.Sc. degree from EPFL in 2001.

Viet-Anh Nguyen is currently a postdoctoral fellow at ADSC. Before joining ADSC, he worked in the School of Electrical & Electronic Engineering, NTU as a Research Fellow from May 2008 to July 2010. His research interests include image and video processing, media compression and delivery, computer vision, and color imaging. Viet-Anh received his Ph.D. degree from Nanyang Technological University in 2010.

Johan Vu is a software engineer at ADSC, where he works closely with Dr. Jiangbo Lu in the video telepresence project. Johan’s work focuses on improving performance with state-of-the-art techniques and making the system available on multiple platforms (Windows, iOS, Android, etc). Johan received his B.S. degree from National University of Singapore in 2010.

References

