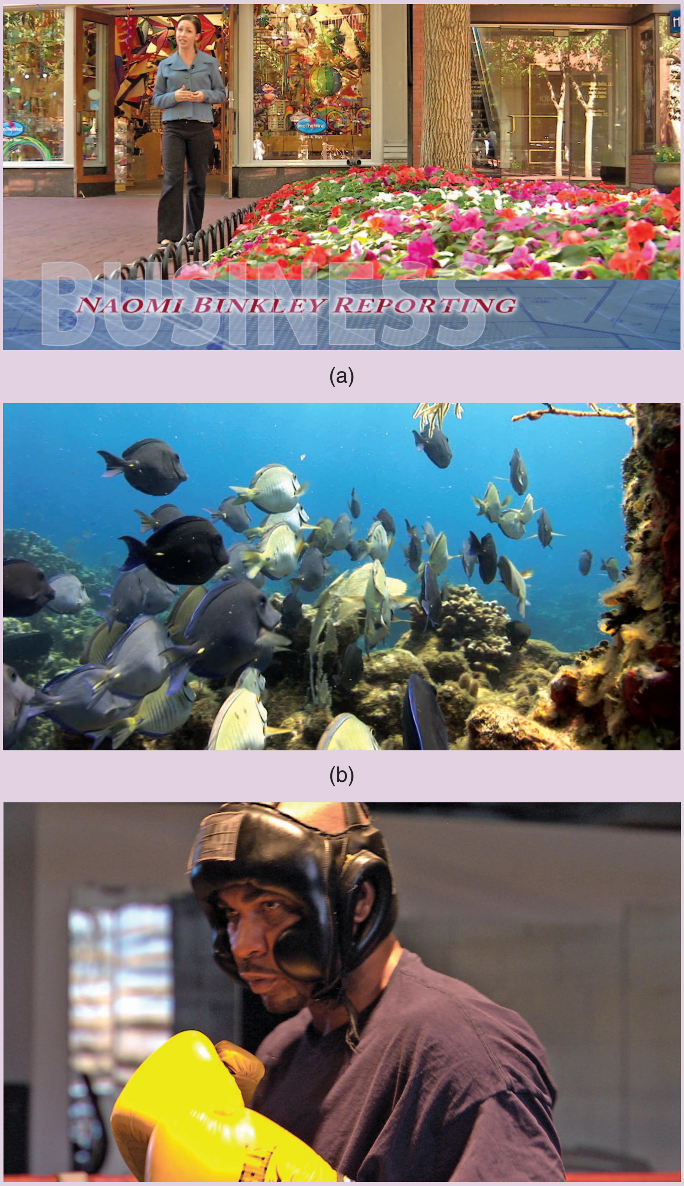
Margaret H. Pinson

The Consumer Digital Video Library

Video is a booming industry: content is embedded on many Web sites, delivered over the Internet, and streamed to mobile devices. Content providers own vast quantities of studio-quality video (i.e., produced to the quality standards of a television studio), but legal contracts between actors, producers, and owners limit how and where others can use such video. As a result, finding and getting rights to use relevant video remains an obstacle to addressing relevant research problems. The Consumer Digital Video Library (CDVL) Web site (http//:www.cdvl.org) attempts to address this obstacle. The CDVL makes high-quality, uncompressed video clips available for free download. Content owners have granted permission for use, and a use agreement protects owners' rights. The clips are ideal for use by the education, research, and product development communities. Developing objective video quality models and testing emergency telemedicine systems are two applications enabled by CDVL content.



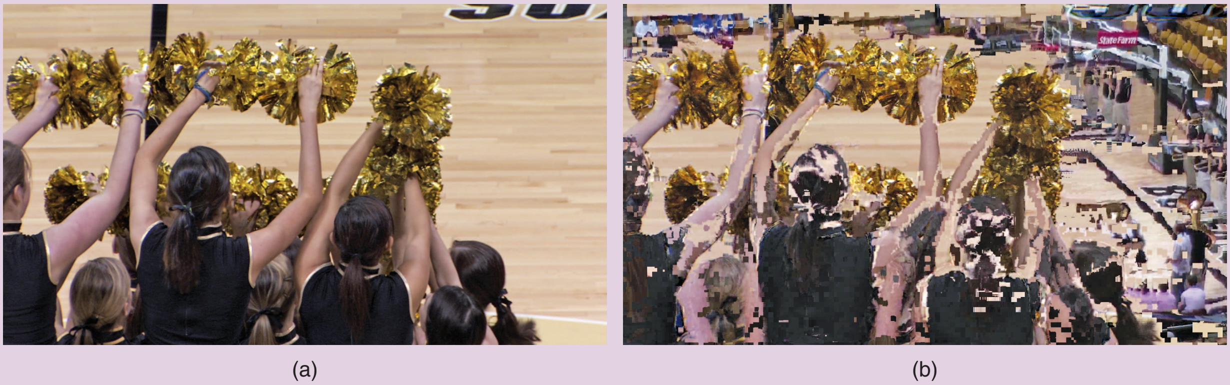
(a) (b) (c)

[FIG1] Examples of CDVL studio-quality footage, available royalty free for research and development: (a) simulated news with animated overlay and simultaneous pan and zoom, (b) underwater footage with unpredictable motion, and (c) a boxing promotional piece with very fast motion. (Figure 1 used with permission. Parts (a) and (c) were commissioned by ITS and filmed by Fireside Productions. Part (b) is an outtake from a copyrighted work that was submitted to the CDVL by Liquid Assets.)

VIDEO QUALITY DATA SETS

The features that make the CDVL an important resource for researchers in video quality are high-quality source video and royalty-free use. The CDVL emphasizes broadcast quality high-definition television (HDTV) clips with a variety of characteristics (see Figure 1). The CDVL sequences can be played at technology research and development events, and sample frames can be published in technical papers. The intent is to enable well-balanced experiment designs [1].

Digital Object Identifier 10.1109/MSP.2013.2258265
Date of publication: 12 June 2013

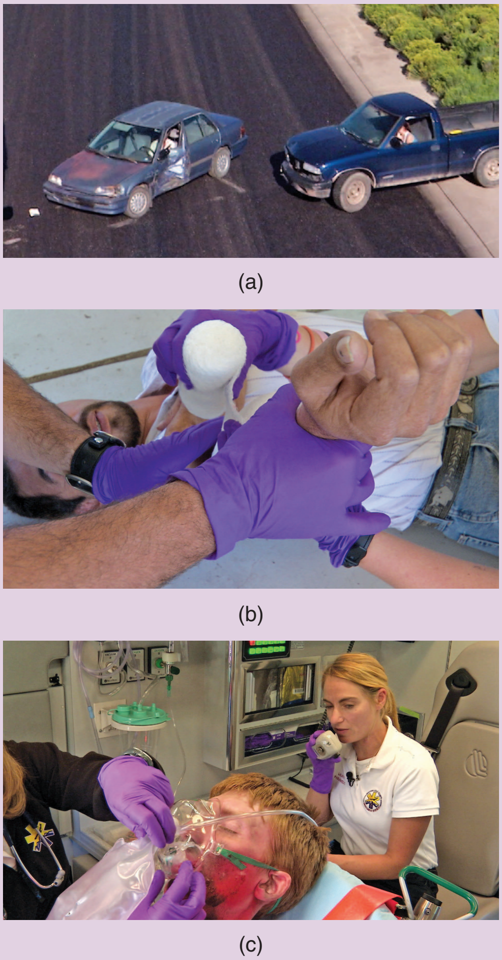


[FIG2] **Sample frames from (a) an original video and (b) MPEG-2 at ≈10 Mb/s with burst errors generated by a decoder that mixed content from different frames in response to the transmission errors. Video was filmed by ITS with permission from the University of Colorado.**

The CDVL hosts five subjective video quality data sets (vqeghd) [2] from HDTV tests conducted by the Video Quality Experts Group (VQEG). The VQEG evaluates objective video-quality models. The goal of these models is to mimic a person's opinion of the quality of a short video clip. For example, a person might rate the quality as "excellent," "good," "fair," "poor," or "bad." Objective video-quality models rank clips on a numerical scale that follows the same progression (e.g., "bad"=1, "excellent"=5).

To test these models, the VQEG performs subjective video-quality experiments. Each experiment includes carefully selected video clips. The clips are coded in a variety of ways, such as MPEG2 at a rate of 10 Mb/s with bursts of bit errors (see Figure 2). A panel of viewers rates the quality of each clip. This yields a mean opinion score (MOS) for each clip. ITU Rec. P.910 describes testing procedures that ensure accuracy [3]. Taken together, the original clips, impaired clips, and MOS form a subjectively rated video-quality data set.

In general, industry wants an objective video-quality model that uses only the impaired video to predict MOS [4]. This approach is called no-reference (NR) and could be deployed on any video stream. By contrast, full reference (FR) models need full access to the original video. Reduced reference (RR) models need some information from the original video. The FR and RR model approaches are robust, because the original video provides "truth data," (i.e., what the video is supposed to look like). Since the original video is seldom available, an NR model would be more useful. However, good NR models are extremely difficult to design, as they require accurate models of human vision, object recognition, and quality judgment. Based on the VQEG model validation reports, the ITU approved recommendations that include FR and RR models. By contrast, NR models have shown poor performance in VQEG validation testing [4], and no ITU recommendations were approved. This absence has stimulated interest in bit stream models and hybrid perceptual bit stream (hybrid) models. Bit stream models examine only the encoded bit stream; and hybrid models examine both the impaired video and the encoded bit stream. These models may be easier to develop than NR models since more information is available. However, the drawback is reduced flexibility because each



[FIG3] **Simulated emergency medical services (EMS) video samples: (a) security camera view of an accident crash site, (b) paramedics wrapping an arm laceration in gauze, seen from a paramedic's helmet camera, and (c) hospital call-in of a burn patient, from inside an ambulance. (Figure 3 is used with permission and was commissioned by PSCR and filmed by Fireside Productions.)**

video coding algorithm needs its own model. In October 2012, bit stream models were standardized in ITU-T Recommendations P.1201 and P.1202. Within the VQEG, the hybrid project is evaluating hybrid models and the Joint Effort Group (JEG) is cooperatively working to develop a hybrid perceptual bit stream model (see http://www.vqeg.org for details). The CDVL expects to host some of the hybrid subjective quality databases when they become available. Vqeghd2 includes bit stream data, so it can be used to investigate hybrid models.

VIDEOS FOR EMERGENCY TELEMEDICINE

EMS personnel can improve patient care by using video (Figure 3): an emergency room doctor could watch a video of an automobile crash site to understand the forces involved in the accident, watch paramedics respond to a call in real time through the paramedic's helmet camera, and provide specialized care advice. The CDVL contains two sets of video clips of EMS calls. First, Hutchinson Community College donated 74 video clips, where each clip was 19 s in duration. The Hutchinson footage was filmed during paramedic training exercises. Second, EMS personnel and the Public Safety Communications Research (PSCR) program (www.pscr.gov) choreographed and filmed 40 incidents with durations ranging from 20 s to 8 min. The PSCR footage demonstrates the most common EMS calls, such as cardiac arrest, car crash triage, burn patient, laceration, patient transfer, abdominal pains, and a broken femur. Two paramedics carry out realistic treatments and discussions, occasionally assisted by firefighters or medevac personnel. Moulage artists applied mock wounds to simulated patients that span a wide range of ages, gender, and races. The variety of skin tones and lighting conditions allows investigation into the interactions between skin tone, codec, and display.

AUTHOR

Margaret H. Pinson (mpinson@its.bldrdoc.gov) is with ITS, a research and engineering branch of the U.S. Department of Commerce.

REFERENCES

[1] M. H. Pinson, K. S. Boyd, J. Hooker, and K. Muntean, "How to choose video sequences for video quality assessment," presented at the *7th Int. Workshop Video Processing and Quality Metrics for Consumer Electronics (VPQM)*, Scottsdale, AZ, Jan. 2013.

[2] Video Quality Experts Group. (2010, June 30). Report on the validation of video quality models for high definition content [Online]. Available: http://www.vqeg.org

[3] ITU-T Rec. P.910, "Subjective video quality assessment methods for multimedia applications," International Telecommunication Union, Geneva, Switzerland, Apr. 2008. [Online]. Available: http://www.itu.int

[4] D. Hands, D. Bayart, A. Davis, and A. Bourret, "No reference perceptual quality metrics: Approaches and limitations," in *Proc. 16th SPIE Human Vision and Electronic Imaging*, San Jose, CA, Feb. 2009, vol. 7240, p. 72400.

[SP]