# Context-Aware Video Compression for Mobile Robots

Daniel A. Lazewatsky, Bogumil Giertler, Martha Witick, Leah Perlmutter,
Bruce A. Maxwell, and William D. Smart

***Abstract*— Operating robots across networks with unknown, bandwidth, latency and other conditions presents difficulty when the operation depends on real-time feedback and control. Standard video compression methods do a good job compressing arbitrary video, but do not take domain knowledge into account when more information about the video is known beforehand. We have incorporated robot odometry into the video pipeline, allowing video quality to be selectively reduced at times when odometry suggests that such a reduction will not adversely affect task performance of human operators. We found that selectively reducing video quality significantly reduced bandwidth usage, increasing the robot's responsiveness and controllability, while having no measurable effect on task performance.**

## I. Introduction

Teleoperation of robot systems (both direct control and shared autonomy) often rely on video to provide feedback to the human operator. Although this video is likely to be low-resolution, it may be transmitted over networks with unknown, nondeterministic conditions [3], causing unknown latency. Standard compression techniques, such as H.246, Theora, VP8, etc. are good general-purpose compressors, but do not take into account extra information that may be known about the video or how it is being used.

Furthermore, it has been estimated that 600 pixels suffices for humans to read text [4], and 200 pixels suffices to recognize simple objects [6], suggesting that streaming video to remote operators at standard resolutions of $320 \times 240$ or $640 \times 480$ uses orders of magnitude more bandwidth than is necessary for the operator to adequately perform many tasks.

Using less bandwidth for video transmission has several benefits, including lower resource usage, lower latency, and faster frame rates. When remotely operating a robot, latency and frame rate are essential to making the system responsive and controllable. The less autonomous a robot is, the more important low latency and high frame rate becomes.

For some tasks, however, such as using the robot as a remote avatar to explore a highly visual environment such as an art museum, a high quality video stream is essential to making the robot a useful tool. While users may be able to effectively move and control the robot with low quality video, if they cannot adequately view their environment the robot cannot act as an effective avatar. Picking a single setting to balance low latency and a fast frame rate with high quality images will inevitably compromise one or both constraints.

The robot's own motion, camera intrinsics, the human visual system, and the scene itself, all affect the amount and utility of information conveyed in a particular video frame. If the information content of an image can be estimated in real-time, either by direct or indirect estimation, the robot can adjust the compression parameters on a frame-by-frame basis in order to optimize the balance between latency, frame rate, and overall image quality.

We have found that the largest factor affecting video quality is motion-blur caused by robot motion. Once an estimate of the usefulness of a particular frame is known, the frame can be preprocessed before being given to a general-purpose compressor and sent over the network.

To demonstrate the concept, we have developed systems to evaluate both direct and indirect estimation of information in a video frame. We use image entropy as a direct estimate of information content, and robot motion commands as an indirect estimation of information, since most motion blur is caused by the robot's own movement.

We begin by presenting an analysis of information content in images under various conditions in order to derive a method for approximating the usefulness of an image in real time, and go on to show that under certain circumstances, drastically increasing the compression in specific ways does not affect the usefulness of the video, while significantly reducing bandwidth usage.

We then present a user study that dynamically adjusts video compression based on movements commands to the robot (indirect estimation). The study asked users to explore and view a simulated art museum, and task that requires both robot motion and viewing of the environment. The results showed that users found the dynamically adjusted video to provide a more responsive and controllable robot system without significantly compromising the overall video quality.

Overall, dynamic video compression based on either direct or indirect estimation of video information more effectively adapts the video stream quality to the current actions of the user, enabling the system to more effectively achieve both high image quality and high frame rates in an action-dependent manner.

This work was partially supported by the National Science Foundation, under awards OCI-0753340 and OCI-0753364 and by a research gift from Willow Garage, Inc. At the time of writing, Smart is on a paid sabbatical at Willow Garage, Inc.

Daniel A. Lazewatsky is with the Department of Computer Science and Engineering at Washington University in St. Louis, St. Louis, MO 63130, USA. dlaz@cse.wustl.edu

Bogumil Giertler, Martha Witick, Leah Perlmutter, and Bruce A. Maxwell are with the Department of Computer Science at Colby College, Waterville, ME 04901, USA. bmaxwell@colby.edu

William D. Smart is with the Department of Computer Science and Engineering at Washington University in St. Louis, St. Louis, MO 63130, USA. At the time of writing, he is on sabbatical at Willow Garage, 68 Willow Road, Menlo Park, CA 94025, USA. wds@cse.wustl.edu

Fig. 1. Comparison of original (a), motion blurred (b), scaled (c) and histogram compression (d).

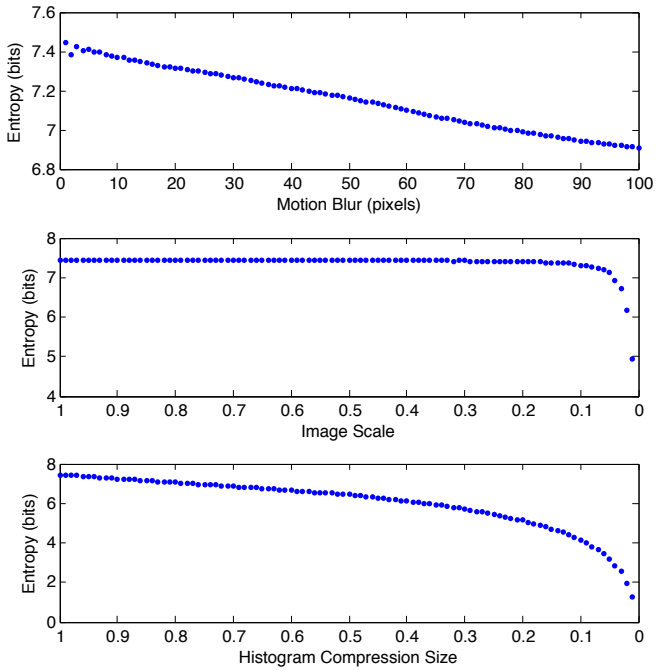



Fig. 2. Entropy with motion blur (top), image scaling (middle), and histogram compression (bottom).

## II. Information Content of Images

A series of tests have been performed to quantify how different conditions affect the information content of images. Images of size $m \times n$ pixels were treated as vectors, $X$ of size $mn$, and the entropy of $X$ was calculated as

$$H(X) = -\sum_{i=1}^{mn} p(x_i) \log_2 p(x_i) f$$

where $p(x)$ is taken from the normalized intensity histogram of the image.

### A. Motion Blur

Artificial motion blur was added to grayscale images over blur sizes ranging from 0 to 100 pixels. Motion blur was added by convolving the image using a uniform, 1-dimensional uniformly distributed convolution kernel with length equal to the number of pixels of motion blur. An example of an unblurred image, and an image with 50 pixels of motion blur is shown in figure 1 (a, b). The results, of varying amounts of motion blur, shown in figure 2 (top), show that as the amount of blur increases in an image, the entropy in the image decreases, indicating that the amount of information carried by a pixel is inversely proportional to the amount of blur in an image.

### B. Image Scaling

A test image was resized over scales ranging from 1 (original size) to 0.1. An example of an image scaled to one tenth the size of the original is shown in figure 1 (c), after being rescaled to take up the same amount of screen space as the original, using nearest-neighbor interpolation to preserve pixelation effects. The associated entropy is shown in figure 2 (middle). The image entropy changes very little before a certain point, suggesting that images can be scaled down by almost an order of magnitude without significant loss of information.

### C. Color Space

Using the same test image as above, the image's histogram was compressed by dividing every pixel by a constant scale factor, $s$:

$$I_s(p_i) = \lfloor I(p_i)/s \rfloor$$

for $s$ ranging from 1 (the original image) to 0.1. An example image with a histogram compressed to one twentieth the size of the original is shown in figure 1 (d). The results are shown in 2 (bottom). Although the entropy does not drop off as sharply after a certain point as with scaling, there is a similar relationship, suggesting that a threshold could be determined allowing histogram compression which maintains enough information.

## III. Entropy in Images from a Robot System

To analyze the entropy in images from a real robot system, image data was taken from the Moving People, Moving Platform Dataset (MPMP), made available by Willow Garage, Inc [8], as well as data collected locally using a WU Telepresence Robot [7]. The data consist of approximately 2.25 hours of monochrome images (40,838 frames in total), originating from one camera of a stereo pair on the pan/tilt head atop a PR2. Full coordinate transform data, consisting of $x, y, z$ linear velocity and $x, y, z$ angular velocity for each frame, was available at a frequency of approximately 40Hz [9]. Reported velocities included motion of the robot base, and any movement of the pan/tilt head.

Linear motion was found to have little effect on image entropy. There was a very small, but significant ($p < 0.05$) correlation between the robot side-to-side motion and the

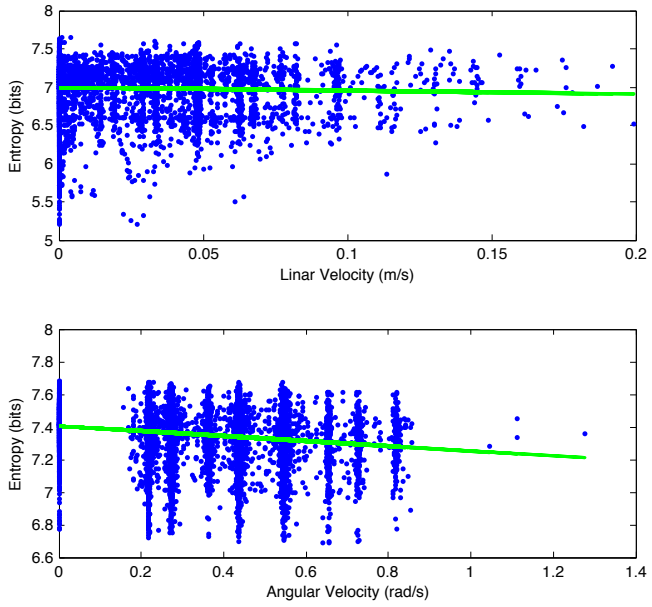



Fig. 3. Image entropy shown for translational motion (top) and angular motion (bottom).

image entropy ($\rho = -0.02$, using a Pearson product-moment correlation), however, not a strong enough correlation to serve as an appropriate predictor for variable compression. Rotations of the camera affected image entropy as predicted in II-A.

Data containing angular motion show a small ($\rho = -0.04$), but significant ($p < 0.001$) correlation between angular velocity and entropy, shown in figure 3. Unlike the ideal case, described in II-A where the image entropy depends only on the original image and added blur, there are many factors which affect entropy in images captured from a real robot moving about the world, including, but not limited to scene composition, lighting, and independent motion. All of these are potential indicators of reduced image entropy, but are not directly measurable and beyond the scope of this work.

## IV. Visual Perception

With mobile robot platforms, motion blur presents a significant obstacle for successful teleoperation, especially while turning. The amount of motion blur is dependent on the camera exposure time, the speed of the robot and the distance to the object. As the robot's speed increases, the amount of blur also increases, and as shown previously, decreases the entropy of the image. As a result, these blurred images contain less information and are less useful to humans.

However, the blur does not affect different types of information equally. When turning, it is still possible for humans to estimate angular velocity using optic flow, a feature which is preserved in motion blur. Because of this, if variable compression is performed only when turning, it is possible to perform large amounts of compression under the condition that the correct information is preserved.

## V. Variable Compression

As described in II-B and II-C, it is possible to discard image data without significantly altering the information content. Because of this, we are able to use lossy compression techniques. There are two aspects of the video stream we can adjust to perform lossy compression:

- **Spatial resolution**: decreasing the overall size of the image by shrinking its dimensions. Reducing the image size achieves significant bandwidth reduction: reducing each dimension by half reduces the data stream by a factor of four.
- **Color resolution**: reducing the total number of colors or shades in the image. Decreasing color resolution can also achieve significant bandwidth reduction: converting colors to a 5/6/5 bits per color channel (red/green/blue) compresses the image to two bytes per pixel instead of three. Color compression also allows a general purpose image or video compression algorithm to use fewer bits when using Huffman Coding or run-length encoding.

We can adapt the video compression using simple or complex functions. In the simplest case, we can use binary adaptation where the robot has two levels of compression and chooses between them depending upon whether it is moving.

We can also adapt the video compression in a more continuous manner. In section II we described how different operations affect the information content of images. Using these relations, we can tie robot motion, provided as odometry (more specifically angular velocity) to compression.

For example, we can define two linear functions based on the angular velocity $\omega$, a minimum image resolution $r$, and a minimum color resolution $d$. The function $S(\omega)$ defines the spatial resolution, and $D(\omega)$ defines the color resolution of the image passed on to the general compression algorithm.

$$S(\omega) = -|\omega| \frac{1-r}{\omega_{max}} + 1$$
$$D(\omega) = -|\omega| \frac{1-d}{\omega_{max}} + 1$$

Note that we use $|\omega|$ so compression depends only on the magnitude of the motion. Using a continuous function to manage the adaptive compression has the potential to provide a smoother user experience.

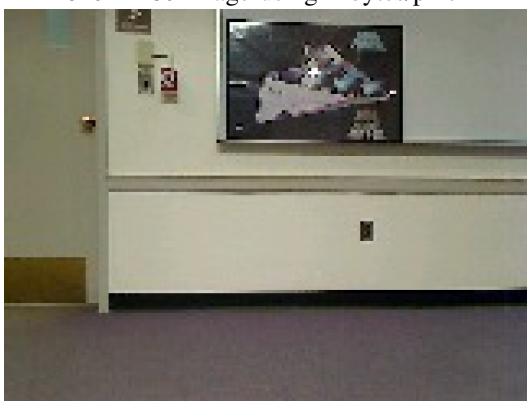
## VI. Experiments

### A. Experimental Setup

We conducted a field test with novice users to evaluate the impact of adaptive compression on their ability to remotely control a robot and complete a simple task. We used a simple binary adaptive video compression technique, switching the video quality based on whether the robot was moving.

The robot system used for testing was an iRobot Magellan Pro base, equipped with sonar, laser, and IR sensors for obstacle avoidance, and an EeePC netbook equipped with a Logitech QuickCam Pro 9000 for communication, control, and video processing. The user interface ran on an iPad (1), with both systems running on a wireless G network.

640 x 480 image using 2 bytes/pixel



160 x 120 image using 2 bytes/pixel

Fig. 4. Comparison of compressed and uncompressed images

For the experiment, we kept the user interface simple. Users viewed the robot's video presented full screen on the iPad. A control strip overlaid along the bottom of the screen contained a stop button. Users could move the robot forward and backward using up and down swipes. Users could turn the robot left and right using horizontal swipes. While our interface contained many other features–such as a map and clicking to send the robot to a location–we disabled those features for the experiment to require the user to be continuously engaged in controlling the robot. Note that our system did not give the users direct control over the robot, and the robot system would override the user's commands if it sensed an obstacle in the way.

The experiment was set up as a direct comparison of completing a task with adaptive video compression and without. For both conditions, the task was to navigate to and view two posters, similar to the task of visiting two pieces of art in a museum. Figure 4 shows one of the posters.

After a description and demonstration of the user interface, users were given the opportunity to practice with the system with only the base compression for up to 10 minutes, or until they were comfortable controlling the robot, in order to reduce any learning effect during the experiment.

The robot was then placed in one of two starting locations, selected randomly, with the pieces they needed to visit behind and to the right of the robot, between 3-5m away. In both cases, the robot's starting location was the same distance and orientation away from the two pieces. Users were then told the relative orientation of the two pieces and asked to navigate the robot to good viewing locations in front of them.

The interface–with compression or without compression–was randomly ordered so that half the time users made the first run with dynamic compression, and half the time users made the first run without it. Both interfaces used a 640x480 image with color compression from 3 bytes to 2 bytes/pixel, followed by run-length-encoding as the base video signal. The interface without compression used this format continuously, the interface with compression used it only when the robot was not moving.

For the compressed signal, the size was reduced to 160x120 while the robot was moving, with the color compression and RLE-encoding remaining identical. The amount of bandwidth required for the compressed signal was $1/16$ of the base video signal. Examples of the same scene at the large resolution and small resolution, captured from the iPad are given in Figure 4. We used a binary adaptation strategy and an extreme amount of spatial compression in order to emphasize the difference between the two conditions.

After the first run, users filled out a questionnaire with five questions which asked users to rank different aspects of the robot interface on a scale from 1 (excellent) to 5 (poor). After the second run, users filled out a second, identical questionnaire. The five aspects were the following.

1) Overall video quality (VQ)
2) Responsiveness of the interface (RI)
3) Clarity of images in the video (CI)
4) Ability to control the robot (AC)
5) Ability to explore the area (AE)

## VII. Results

### A. User Study

Seven users completed the study. The users were college age students, balanced between men and women, with normal, or corrected normal vision.

To analyze the results on the five rankings, we calculated the means for the runs with and without adaptive compression and then executed a two sample t-test with unequal variances. The means, standard deviations, and z-scores are shown in Table I. Questions 1, 3, and 5 showed no significant difference between the two cases, meaning that users found no real difference between the dynamically compressed or uncompressed video when asked about overall video quality, image clarity, or their ability to explore the area. However, questions 2 and 4 found significant differences, at 99% and 95%, respectively, using a 2-sample t-test with unequal variances; the users' rankings of the responsiveness of the interface and their ability to control the robot were significantly higher with the dynamically compressed video.

### B. Compression

In order to determine the bandwidth affects of the techniques described in V, a test image was compressed over a range of input angular velocities. The image was saved in Portable Network Graphics (PNG) format [2] (the same

TABLE I

USER STUDY RESULTS

| Question | VQ | RI | CI | AC | AE |
|---|---|---|---|---|---|
| Mean (static) | 2.57 | 3.71 | 2.43 | 2.71 | 2.14 |
| Stdev (static) | 0.98 | 0.76 | 0.98 | 0.49 | 0.38 |
| Mean (dynamic) | 3.00 | 2.57 | 3.00 | 2.14 | 2.29 |
| Stdev (dynamic) | 1.15 | 0.53 | 1.00 | 0.69 | 0.95 |
| z (static - dynamic) | -0.75 | **3.27** | -1.08 | **1.79** | -0.37 |

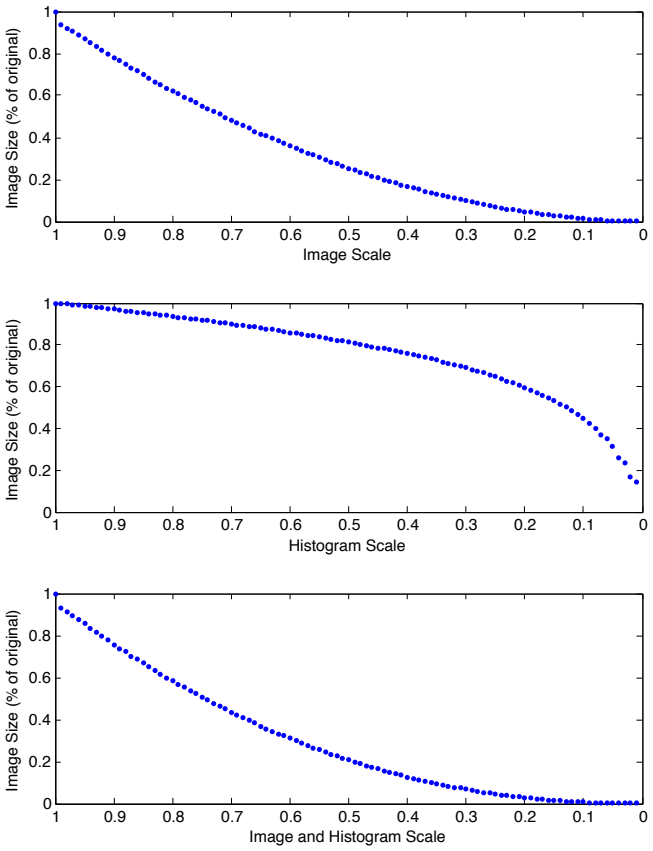



Fig. 5. Compression of a single image over a range of scales for image resolution (top), histogram compression (middle) and combination (bottom). The image size is shown as a percentage of the size of the unmodified image.

format as the source image), keeping the PNG encoder settings constant across scale factors.

Reducing the resolution and colorspace both have noticeable effects on the file size of the output image, as shown in figure 5. Spatial resolution has a stronger effect on PNG compression. Color resolution can more strongly affect encoding methods such as run-length-encoding, where compressed colors tend to increase the length of runs. Task-dependent parameters for minimum resolution and minimum colorspace size can also significantly reduce bandwidth usage without affecting task performance.

## VIII. DISCUSSION

### A. *User Study*

The results of the user study show that reducing bandwidth of the video stream–which increases the frame rate because it is a fixed size pipe–improves the robot's responsiveness and the user's ability to control the robot. This finding, by itself, is not surprising, as a higher frame rate gives users faster and more immediate feedback. However, users also did not rank the video quality or image clarity significantly lower with the dynamic interface, likely because the image quality was identical to the uncompressed interface when the robot was stopped. Combined with motion blur and the speed with which the image is changing during motion, the low quality video stream is either less noticeable or acceptable to users because it still provides the essential information.

For situations where the quality of the image is important, such as when visiting or exploring a museum remotely, the dynamic compression automatically provides a balance between responsiveness and image quality. In our experiment we used an extreme level of resolution compression, sending $1/16$ as much data, yet users felt the overall visual quality was similar to using no compression. The adaptive compression works well, because the image quality is important primarily when the robot is still.

For situations where video quality has more importance during motion, one could adjust the level of compression to the point where the compression level is not the limiting factor to visibility, but motion blur and the speed of change limits the user's ability to identify details. Overall, however, while the robot is moving, frame rate and responsiveness increase in importance, and dynamic video compression is one way to achieve overall better performance in a remote robot exploration task.

## IX. Future Work

As presented in this paper, image entropy has been considered only for entire images. However, the entropy is likely to vary greatly within a single image. For example, if a large portion of an image consists of a blank wall, that region of the image is unlikely to be useful to a human and consequently can be compressed more. JPEG2000 allows regions within an image to be compressed differently [1], providing an easy entry point for such work.

We have argued that optical flow is a major component of the image information used by humans when remotely operating robots. Better understanding the human perceptual system in relation to robot task performance would help guide compression in this context, and help develop compact scene representations which still carry the information necessary for human operators (something which has been touched upon by Brooks and Ince[5]).

## References


[1] ISO/IEC 15444-5:2003. *Information technology – JPEG 2000 image coding system: Reference software*. ISO, Geneva, Switzerland.

[2] ISO/IEC 15948:2004. *Information technology – Computer graphics and image processing – Portable Network Graphics (PNG): Functional specification*. ISO, Geneva, Switzerland.

[3] K. Brady and T. Tarn. Internet-based remote teleoperation. In *Robotics and Automation, 1998. Proceedings. 1998 IEEE International Conference on*, volume 1, pages 65 –70 vol.1, May 1998.

[4] G.S. Brindley. The number of information channels needed for efficient reading. *J Physiol*, 177:44, 1965.

[5] T.L. Brooks and I. Ince. Operator vision aids for telerobotic assembly and servicing in space. In *Robotics and Automation, 1992. Proceedings., 1992 IEEE International Conference on*, pages 886 –891 vol.1, May 1992.

[6] H.G. Vaughn Jr. and H. Schimmel. *Visual prosthesis: the interdisciplinary dialogue*, chapter Feasibility Of Electrocortical Visual Prosthesis, pages 65–79. New York, Academic Press, 1971.

[7] Daniel A. Lazewatsky and William D. Smart. An inexpensive robot platform for teleoperation and experimentation. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA 2011)*, 2011.

[8] Caroline Pantofaru. The Moving People, Moving Platform Dataset. `http://bags.willowgarage.com/downloads/people_dataset/`.

[9] Caroline Pantofaru. User observation & dataset collection for robot training. In *Proc. of Human-Robot Interaction (HRI)*, Lausanne, Switzerland, 03/2011 9998. ACM Press.