

DDXVIDEO: A Lightweight Video Framework for Autonomous Robotic Platforms

Elliot Duff

CSIRO ICT Centre

PO Box 883, Kenmore, Qld 4069, AUSTRALIA

Email: firstname.lastname@csiro.au

Abstract

This paper describes the implementation of a lightweight video framework for autonomous robotics platforms. It is based upon DDX (Dynamic Data eXchange) which is our third generation real-time publish-subscribe software (event-based middleware). Computational resources on autonomous robotic platforms can be very limited, and thus it essential that mechanisms for access to video data have as little impact upon the computation load as possible.

1 Introduction

We are particularly interested in the navigation of autonomous outdoor robotic platforms: including submarines, helicopters and ground vehicles. With advances in computer hardware over the last few years, we have become increasingly interested in systems that are capable of real-time video processing. As with most research organizations, we have a number of researchers who are performing different task on the same platform: i.e. optical flow, stereo, scene segmentation, beacon localization etc. To solve the task of navigation, many of these tasks need to be solved in parallel. Therefore it is important that we establish a common framework for video acquisition and processing. Here is list of desirable features for such a framework:

1. ability to decouple video acquisition from video processing and display;
2. support different video sources (analog or digital camera, pre-recorded or processed video);
3. ability to synchronize video with other sensors and control data;
4. ability to record video without compression or dither;
5. stream video to remote locations over wireless link; and
6. be simple and transparent with very low overheads.

The traditional method has been to develop a monolithic process, with separate threads that adhere to a strict API and are

given appropriate resources. This is the approach taken by the Player/Stage [Vaughan *et al.*, 2003] platform with pluggable software modules. Although this has a number of advantages (i.e code reuse) we would prefer a more decentralized approach. There are a number of alternative robotic software platforms that could be used: YARP [Metta *et al.*, 2005] and ORCA [Brooks *et al.*, 2005] using TAO CORBA [Schmidt, 2004] or CARMEN [Montemerlo *et al.*, 2003] using IPC [Simmons and James, 2004].

The TAO CORBA A/V [Mungee *et al.*, 1999] streaming service is based upon a distributed system which has been modified to provide local support. For several reasons discussed in [Corke *et al.*, 2004], we would prefer the reverse: a system that is based upon local support with distributed features added on. Given this preference and the overall complexity of CORBA, we decided to investigate the possibility of modifying our own software to support video.

The remainder of this paper is structured as follows. Section 2 describes the implementation of supporting video in DDX and the modifications that were made to increase its efficiency. Section 3 describes some of the experimental results, and Section 4 concludes and presents future directions.

2 Implementation

For over 10 years we have been building complex robotic systems. Over time we have developed a series of software platforms (middleware) to facilitate this endeavor. Our third generation middleware software is called DDX (Dynamic Data eXchange). This is event-based middleware that provides distributed real-time publish/subscribe access to sensor and control data. It does this through an efficient shared memory mechanism managed by a **store**. The **store** provides a naming service to data in shared memory. Stores on multiple machines can be linked by means of a global **catalog** and data is multicast between the stores when the data is requested (subscribed to). DDX also has a number of service applications, including the ability to log and replay data in real-time. The implementation of DDXVIDEO consists of an agreed format in the **store** and a suite of applications (see Figure 1).

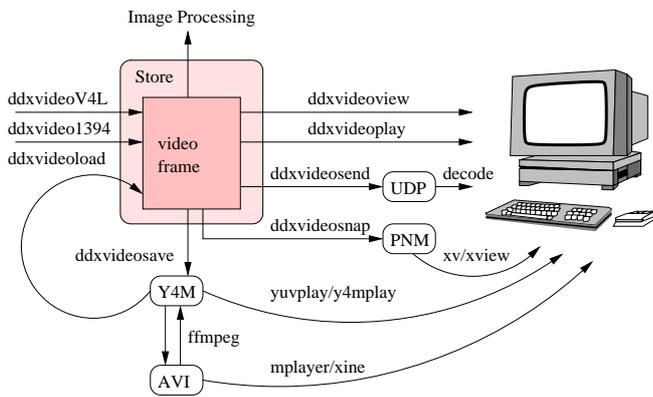


Figure 1: DDXVIDEO applications.

2.1 Video Format

Typical video formats can either be packed or planar, use RGB or YUV colourspace, and have different levels of sub-sampling. Analogue frame grabbers (such as the BT878 chipset) processed all incoming video as YUV422. In this packed colourspace the U and V components (chroma) are down-sampled 2 to 1 in the horizontal direction. This means that if an RGB colourspace is requested, the image will only contain colour information that existed in the original YUV422 image.

Another video format is YUV420. In this colourspace, the U and V (chroma) are downsampled 2:1 in both the vertical and horizontal direction. This differs from YUV411, which is downsampled 4:1 in the horizontal direction. Whilst the YUV420 format has half the chroma resolution of the YUV422 format, it is the format preferred by JPEG and MPEG compression routines and by XVideo to display video under X11. This is important factor when the speed of encoding and decoding will have an impact upon the computational load. Perhaps the most important factor was the requirement to be able record and playback the video stream without compression. The most popular uncompressed video format is YUV4MPEG [<http://mjpeg.sourceforge.net>], Y4M for short. Whilst not perfect, the YUV420P format is a reasonable compromise between resolution, compatibility and speed, and thus was chosen as the default format for video in the **store**.

The shared memory data structure is shown in Figure 2. There are three data structures in the **store**. The first is DDX_VIDEO, which contains a header and three video planes (Y,U,V). The Y is the luminance, whilst the U and V make up the chrominance at half the resolution of the Y plane. To accommodate multiple video streams that need to be synchronized (such as stereo) each video stream is placed on top of one another in memory. This is represented by the number of fields in each frame. This is shown in Figure 2 where there are three video fields per video frame.

The data structure is defined in DDX by the following dec-

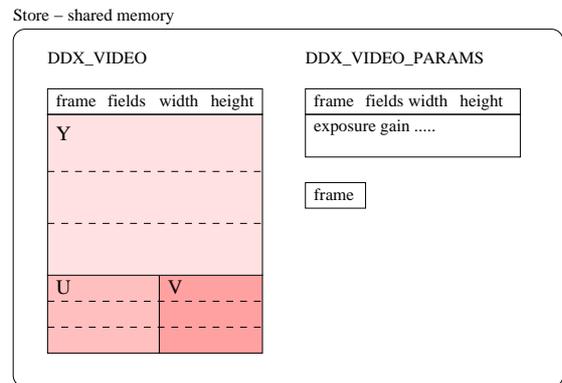


Figure 2: Video data format in store.

laration:

```
DDX_STORE_TYPE( DDX_VIDEO,
struct {
    int frame; // Frame sequence number
    int fields; // Number of fields in frame
    int width;
    int height;
    char y[MAX];
    char u[MAX/4];
    char v[MAX/4];
} );
```

The second data structure contains camera parameters which are updated whenever the camera settings are modified. Typically, the camera is controlled by a simple RPC request to the corresponding application. The third data structure is a simple integer frame counter. This is used to synchronize normal DDX data with video data.

2.2 Access to Video data

Access to the video data is provided by standards DDX calls to the **store**:

```
DDX_STORE_ID * storeId = NULL;
DDX_STORE_ITEM * itemPtr = NULL;
storeId = ddx_store_open(NULL, port, timeout);
itemPtr = ddx_store_lookup_item( storeId, "video", NULL, 0);
```

where **ddx_store_open** is used to make a connection to the local **store**, and **ddx_store_lookup_item** is used to return a handle to a named item. In this case to *video*. The item handler is then used to read the video data.

```
DDX_VIDEO video;
ddx_store_read(itemPtr, &video, &ts, 1.0, 1);
```

This function can be used in blocking or non-blocking mode. In either case, a local copy is made of the current video frame. To guarantee the integrity of data, the store places a mutex around the read and writing of data to shared memory. This prevents one application writing to memory whilst another is reading. This behaviour is critical for most sensor and control data, however if we only wish to “spy” on the current data, then this locking can be excessive. To prevent unnecessary locking and copying of data, functions to DDX were introduced that provide direct access to the store data. This is done with the following commands:

```
DDX_VIDEO * video;
video = ddx_store_var_pointer(itemPtr);
ddx_store_read_direct(itemPtr, &ts, 10.0, 1);
```

where, **ddx_store_var_pointer** provides a pointer to the actual video shared memory. The **ddx_store_read_direct** acts as a semaphore informing the application that someone has just finished writing to the store. The direct access to video memory is used by applications (such as display applications) that will not be significantly affected if the video memory were altered whilst displaying the image. This type of behavior can reduce the computational load. Of course, the fact that we now have direct access to the current video frame is not without risk, and it is up to the application writer to exercise caution.

2.3 Video Applications

There are a number of application that have been written that can load video into the **store**:

ddxvideoV4L use V4L (video4linux) drivers to acquire images from analog framegrabber.

ddxvideo1394 use dc1394 (firewire) drivers to acquire images from firewire bus.

ddxvideoSVS use SVS libraries (SRI's Small Vision Software) to acquire range images from stereo camera. <http://www.videredesign.com/>

ddxvideoload read Y4M (yuv4mjpeg) file and load into store at specified time-steps.

And a number of applications that display or save video from the **store**:

ddxvideoplay display video stream locally

ddxvideoview display video image remotely (step though each frame)

ddxvideosave write videostream as Y4M file

ddxvideosnap write single frame as PPM file

ddxvideosend use libavcodec to encode image and send out through UDP port.

decode to decode UDP packets and display on remote screen. Record AVI file if requested.

Here is an example that will acquire video at 5Hz in CIF resolution.

```
catalog &
store -m 8M &
ddxlog -c logger.conf &
ddxvideoV4L -camera 2 -skip 5 -size CIF -name video &
ddxvideoview -name video &
ddxvideosave -name video -output /tmp/test.y4m &
```

Here the **catalog** and **store** are started with sufficient shared memory to handle video. Video is then acquired on a specified channel, with a specified rate and size, and copied into the store with a specified name. In this

case it is called *video*. The current *video* frame can be displayed with **ddxvideoview**. In theory, one could use **ddxlog** to record the video stream. However, this log file is only compatible with DDX. The preferred option is to use **ddxvideosave** to record a Y4M file. Once recorded the Y4M file can be displayed with **yuvplay** or **mplayer**. It can also be transcoded (converted to a number of video formats) with **ffmpeg** or **mencoder**.

One of the features of the Y4M file is that it is possible to record the time step between each video frame. This time-step is used to replay the video back to the store in real-time, which can be used to test the performance of the image processing routines. For example:

```
ddxvideoload /tmp/test.y4m
```

Furthermore, if the frame counter is recorded in a log file along with named sensor data, it is possible to synchronize log data with video data.

2.4 Video Streaming

Although the store can be configured to handle large data structures (such as images), the sharing of data between stores is currently restricted to the size of a single UDP packet (8k). Therefore to send the video sequence to another computer it is appropriate to bypass the store and broadcast the video directly. Unfortunately an uncompressed PAL video stream consumes approximately 256Mbps, which is far beyond the capabilities of the WiFi network that we have between our autonomous platforms. Although there are many applications that can compress and stream video (FFmpeg, Mbone, Mash, VideoLan, Darwin, RealSystems) they are all designed to synchronize the video and audio streams.

To achieve synchronization it is important that the packets arrive in the correct order and at the correct time. Since neither of these conditions can be guaranteed on the Ethernet, buffering is required at the receiving end to reorder and request that lost packets be resent. This buffering can have a significant effect upon latency, especially in wireless environments which can have a low bandwidth and high packets loss. Since is our intention to use the remote video stream to control the robot it is critical that the latency be as small as possible (less than 100ms). Fortunately, since we are not interested in the audio stream, it possible to write our own streaming software without buffering.

```
ddxvideosend -port 8000 &
```

In this example, the video stream can be decoded and viewed on another computer with an application called **decode**. Since, by default, the video is multicast on a UDP socket, any number of users on site can "tune-in" into port 8000 to watch the streaming video. This is a very convenient way to keep track of what is happening on our autonomous platforms without imposing additional load on the system. Thus we need to choose a video CODEC that provides reasonable quality and robustness, with low latency and bandwidth, without consuming too many computational resources. Motion



Figure 3: Ground vehicle platform : Automated ride-on lawn mower fitted with GPS, IMU, laser rangefinder and a number of Firewire cameras. The vision processing and control computers are fitted behind the driver's seat.

JPEG is one of the first streaming technologies to be used. It is robust because there is no interframe dependence, ie. frames can be lost without significant impact, however significant compression is required to get below the 8K threshold of UDP transmission. Higher compression can be achieved with interframe compression - where there is compression from one frame to another (ie. MPEG 1,2 & 4). The most appropriate for streaming is MPEG4 because it has been designed to withstand significant packet loss. Since MPEG4 is robust to packet loss, no attempt is made to recover dropped frames; reorder frames that are out of sequence, or even to impose a time-step between frames. As each frame is received it is decompressed and displayed. In this way latency is kept to an absolute minimum. This application also has the ability to record the received video sequence to an AVI file.

3 Experimental Platforms

We have a number of experimental robotic platforms that require real-time video processing.

3.1 Ground Vehicle

The ground-vehicle platform (see Figure 3) is a small ride-on lawn mower that is fitted with five digital cameras: two fish-eye, two stereo, and an omni-directional mega-pixel camera [Usher *et al.*, 2003]. These cameras are connected via Firewire to a Pentium-M miniITX computer. The output from the omni-directional cameras is shown in Figure 4. Colour segmentation of the UV plane is used to track artificial beacons and segment the road.



Figure 4: Omnidirectional megapixel image taken from ground vehicle showing road with a footpath on the left.

3.2 Underwater Platform

The underwater platform (see Figure 5) is a hybrid AUV designed for endurance, manoeuvrability and functionality in shallow waters (<100m) for the task of environmental monitoring. Along with GPS and INS, it is fitted with two stereo pairs of analog cameras [Dunbabin *et al.*, 2005]. The stereo cameras are line interlaced and fed into a standard BT878 frame grabber on a 800MHz PC104 Crusoe processor. In `ddxvideoV4L` the video is de-interlaced into the respective fields and placed on top of one another. This is shown in Figure 6, where a number of rocks on the bottom of a swimming pool can be seen. The disparity between the images is used to estimate height, whilst the disparity in time is used to estimate velocity. Simultaneously, another pair of forward looking cameras are used for collision avoidance.

3.3 Aerial Platform

Over the years, the CSIRO Robotics Team have expended considerable energy into the development of a low cost autonomous air vehicle in the form of a 60 size RC helicopter [Buskey *et al.*, 2003]. As a means of reducing development costs we have developed a simulated aerial platform that consists of a sensing pod that is driven by four computer controlled wires (see Figure 7). The pod is fitted with a number of interchangeable Firewire cameras, connected to a self-powered wireless Pentium-M miniITX [Usher *et al.*, 2005]. In one application, the aerial platform is "flown" over a scale model power line. The results from the SVS stereo camera is shown in Figure 8, where the range image is inserted between the left and right image. The range image is used to track the power lines and to identify a clear landing area.

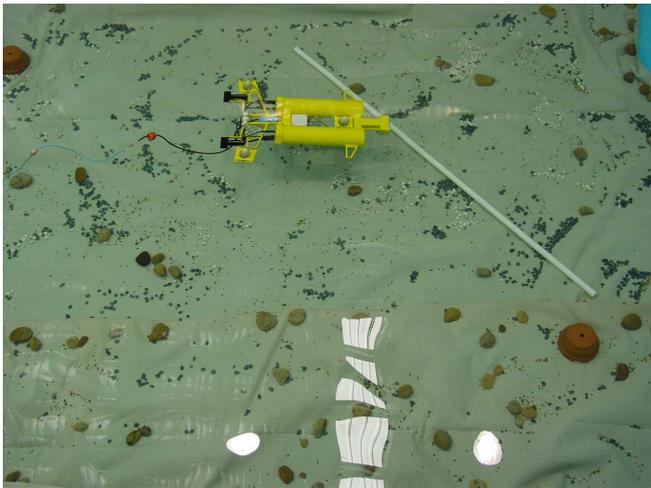


Figure 5: Underwater platform in swimming pool.

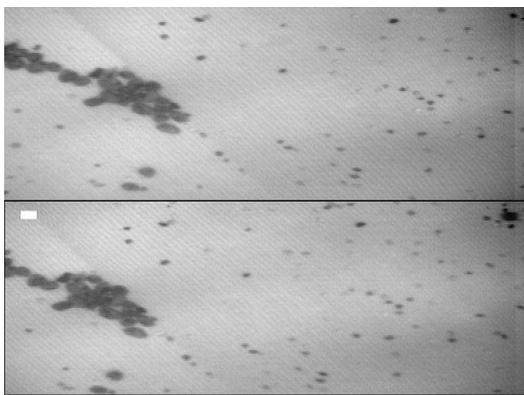


Figure 6: Stereo image taken from downward facing cameras showing horizontal disparity in rock positions.

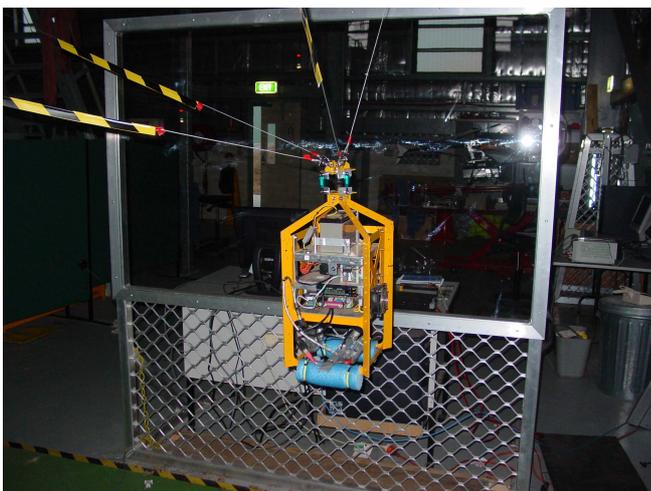


Figure 7: Simulated aerial platform suspended by four cables, in front of remote control station.



Figure 8: Tertiary image taken from aerial pod of scale-model power lines. The disparity between the right (top) and left (bottom) image is used to create a range image (middle).

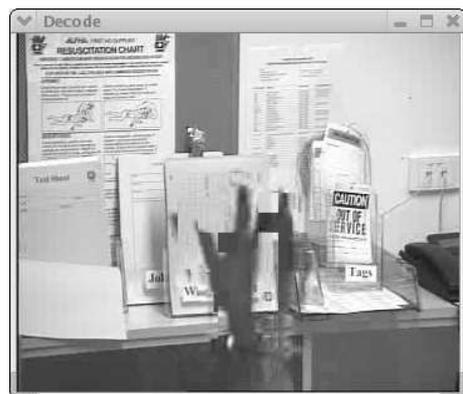


Figure 9: Single image taken in lab of waving hand showing the effect of MPEG4 compression at 200kbps with 30% packet loss.

3.4 Performance

One key requirement of DDXVIDEO is the ability to stream video quickly and robustly over a wireless network. The results of MPEG4 compression at 200kbps over a IEEE 802.11b network is shown in Figure 9, where a packet loss of 30% has been simulated by deliberately dropping every third packet. In this figure, a hand is waived in front of the camera to demonstrate MPEG4's ability to rebuild the image. Note the black squares around the edges of the hand. Similar robustness was exhibited after re-ordering the packets, mixing packets from other streams, and dropping GOP frames. On the aerial platform, the compression software only consumed 5% of the available CPU time.

Another requirement of DDXVIDEO is the ability to record uncompressed video at appropriate frame rates. On the underwater platform a frame rate of 5Hz was required to test optical flow algorithms. The performance of the 800MHz Crusoe with a laptop hard drive was rather disappointing, with significant delays of up to 0.8s every minute or so. This delay resulted in the loss of video frames that could not be recovered with conventional buffering. Fortunately, the delays were found to be related to the EXT3 file system used under Linux, and were avoided by disabling the file system's journaling function. This problem highlighted the fact that video acquisition is still at the "bleeding edge" of technology for embedded computer systems, and the interaction between the operating system and hardware cannot be taken for granted.

4 Conclusion

DDXVIDEO was originally developed to assist with the development of image processing routines on autonomous platforms. Since it is now being routinely used on all of our autonomous platforms (air, land and water) it would be reasonable to conclude that its implementation has been a success. Its success lies in the fact that:

- It provides multiple access to the same video stream with insignificant overhead (just pointers).
- It can store multiple "fields" in the same video frame with guaranteed synchronization.
- It can save video streams in format that is compatible with popular video players.
- It is possible to replay video streams in real time, with exactly the same data.
- It uses MPEG4 compression which is efficient and requires minimal bandwidth.
- It uses UDP multicast which has low overheads and very low latency.

It is interesting to note that other robotic platforms have adopted new transport layers based upon shared memory and UDP: e.g. CRUD in the ORCA platform, and Gazebo in the Player/Stage.

In future work, we intend to support alternative colourspace (compatible with Intel's IPP) and support off-the-shelf streaming software (VideoLan's VLC).

Acknowledgment

The author would like to thank the rest of the CSIRO Autonomous Systems Laboratory who have contributed ideas and have thoroughly tested this software. In particular, Pavan Sikka for his help with DDX and Cedric Pradalier for his help with the 1394 code.

References

- [Brooks *et al.*, 2005] Alex Brooks, Tobias Kaupp, Alex Makarenko, Anders Oreback, and Stefan Williams. Towards Component-Based Robotics, 2005. <http://orca-robotics.sourceforge.net>.
- [Buskey *et al.*, 2003] Gregg Buskey, Peter Corke, Jonathan Roberts, Peter Ridley, and Gordon Wyeth. The CSIRO autonomous helicopter project. In Bruno Siciliano and Paulo Dario, editors, *Experimntal Robotics VIII*, volume 5 of *STAR*. Springer-Verlag, 2003.
- [Corke *et al.*, 2004] Peter Corke, Pavan Sikka, Jonathan Roberts, and Elliot Duff. DDX: A Distributed Software Architecture for Robotics Systems. In *Proceedings of the Australian Conference on Robotics and Automation*, Canberra, Australia, 2004.
- [Dunbabin *et al.*, 2005] Matthew Dunbabin, Jonathan Roberts, Kane Usher, and Peter Corke. A hybrid auv design for shallow water reef navigation. In *Proceedings of ICRA 2005*, 2005.
- [Metta *et al.*, 2005] Giorgio Metta, Paul Fitzpatrick, and Lorenzo Natale. YARP: Yet Another Robot Platform, 2005. <http://yarp0.sourceforge.net>.
- [Montemerlo *et al.*, 2003] Michael Montemerlo, Nicholas Roy, and Sebastian Thrun. CARMEN Carnegie-Mellon Robot Navigation Toolkit, 2003. <http://www.cs.cmu.edu/carmen>.
- [Munsee *et al.*, 1999] Sumedh Munsee, Nagarajan Suren-dran, and Douglas Schmidt. The Design and Performance of a CORBA Audio/Video Steaming Service. In *Proceedings of the HICSS-32 International Conference on System Sciences, Multimedia DBMS and the WWW*, Hawaii, USA, 1999.
- [Schmidt, 2004] Douglas Schmidt. Real-time CORBA with TAO, 2004. <http://www.cs.wustl.edu/schmidt/TAO.html>.
- [Simmons and James, 2004] Reid Simmons and Dale James. *Inter-Process Communication*. Carnegie-Mellon University, 2004. <http://www.cs.cmu.edu/ipc>.
- [Usher *et al.*, 2003] K. Usher, P. Ridley, and P.I. Corke. Visual servoing of a car-like vehicle — an application of om-

nidirectional vision. In *Proc. IEEE Int. Conf. Robotics and Automation*, pages 4288–429, Taipei, September 2003.

[Usher *et al.*, 2005] Kane Usher, Graeme Winstanley, Peter Corke, and Dirk Stauffacher. Air vehicle simulator: an application for a cable array robot. In *Proceedings of the International Conference on Robotics and Automation*. IEEE, April 2005.

[Vaughan *et al.*, 2003] R. T. Vaughan, B P Gerkey, and A Howard. On device abstractions for portable, reusable robot code. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2421–2427, Las Vegas, USA, 2003.