Project Number: 900260

Project Start: Jan. 1, **2006** **Project Duration (in years):** **3**

Please indicate what type of project is being proposed (mark only one).

Feasibility/Scoping [Less than 100 days effort]	Development[1]	Research	Infrastructure[1]	Education	Time Series
	x				

Project Manager/PI: Duane Edgington

Project Title: **Automated Visual Event Detection and Classification Tool**

Lead Scientist: Jim Barry, Steve Haddock, Bruce Robison, Edie Widder
Lead Engineer: Danelle Cline
Project Team: Jim Barry, Andrew Chase, Danelle Cline, Dan Davis, Duane Edgington, Kevin Gomes, Steve Haddock, Linda Kuhnz, Bruce Robison, Brian Schlining, Rob Sherlock, Nancy Jacobsen Stout, Susan von Thun.

[1] Development and infrastructure projects must complete the risk matrix (see item #5 at the end of the proposal form).

ABSTRACT (new projects only)

This new project is follow-on to the three year project AVED: Automated Visual Event Detection (900260). Building on the success of AVED, this project will develop a complete end-to-end system that can be used by trained science personnel (as opposed to computer specialists), is time-efficient, and can be trained and qualified for a new target category (organism) relatively simply and easily.

This project advances oceanography by simplifying and speeding the critical process of consistently annotating quantitatively large amounts of video, and by empowering the scientist as the end user of the system. These annotations, targeted to specific categories (i.e. species or groups of organisms), enable exploitation of the vast collection of instrumented video that has been collected over the decade at midwater and benthic sites. In turn, this facilitates the mining of the video database to answer questions on the carbon cycling within these habitats, seasonal and episodic variations, and distribution and abundance of key organisms.

The scope of the project is to:

- Complete AVED work to build an integrated video processing system that can seamlessly process from Digital Betacam tape without manual steps and import the AVED data into the VARS system.
- Deliver an integrated system with the capacity to quantitatively analyze a large volume of video data using an integrated classification system which identifies individual organisms.
- Develop graphic display interfaces to enable efficient review and qualification of AVED's annotations (detections and classifications).
- Enhance system to additionally process HDTV video (purchase HDTV interface card, modify system as required).
- Improve classification[2] accuracy by refinement of AVED's classification system.
- Develop system to collect existing human annotations of selected organisms efficiently and easily to build training sets for AVED's classifier.
- Prototype modification of AVED to handle fixed observatory cameras, including fixed HDTV cameras and Eye-in-the-Sea.

We will continue to coordinate with related MBARI engineering and science projects, including the proposed quantitative camera project (Reisenbichler) and Precision Controls (Rock).

State of the art: The AVED prototype system developed at MBARI is state of the art for processing underwater video from standard video cameras. No commercial systems exist for this application. Commercial annotation systems do exist for broadcast television data with motion triggered detection that can, for example, generate key frames, but those systems do not support subimage object detection, tracking and classification as needed

[2] the ability to identify individual organisms

for our applications. AVED builds on research classification systems targeted at face recognition, surveillance, and biomedical classification problems. No other system exists for our application of midwater or benthic moving camera video data, or for observatory camera data. No commercial systems have the ability to manage classification databases or automate classification. We continue to closely monitor research and commercial video processing systems looking for technology and solutions to apply to our unusual application. Many other organizations would benefit from and are interested in this technology.

Success: Our ultimate measure of success for the project is putting the system into the hands of MBARI scientists and seeing it used for new science investigations that rely on existing and future video data.

Impact: Automated video analysis, a major goal of this project, would be a very significant advance in engineering, supporting not only ocean science, but also terrestrial and space sciences. It is easy to imagine that several industrial applications could emerge as a technology transfer. If automated video analysis was easy, it would be available now, as it has been in high demand for decades. AVED, and the activities proposed, here represent a significant step in the long-term development of an automated video analysis system.

Acronym: AVEDACT (pronounced äve-dak)

1) PROPOSAL

a) *The problem addressed (new projects only)*

Statement of the problem

One of several areas where unmanned remotely operated underwater vehicles (ROVs) have proven to be extremely useful tools in oceanography is the capture of video and visual images of biological creatures and benthic features. The great success of ROVs in collecting high-resolution video during dives encourages us to look for ways to increase the amount of information and knowledge we gather from underwater cameras. We see tremendous potential benefit in automating portions of the analysis of video, and to extend image collection to cabled video cameras deployed as part of ocean observatories, and to unmanned autonomous underwater vehicles (AUVs) and other underwater platforms.

ROVs are natural platforms for collecting broadcast quality high-resolution video, since high-speed telemetry from vehicle to surface and electrical power for camera and lights are available via a cable tether. Some analysis or adaptive sampling is possible during the dive via surface video monitors and annotation systems, and outline or detailed post-dive analysis is routinely done in the video lab using sophisticated tools such as the MBARI

Video Annotation and Reference System (VARS). There are two bottlenecks in collecting and analyzing this data that we hope to address in this project:

- Video annotation is demanding and tedious work; highly trained scientists can work at most four hours a day annotating video; the capacity of the video lab and research staff to post-process dive footage is a historical bottleneck in reaping the benefit of these expensive dives. One of our goals is to develop systems and methods to augment the productivity of video annotation analysts who process video from ROV dives. The demands of quantitative video analysis, and the tedium of making accurate counts of very frequent events (organisms or features) would be addressed using these by annotation automation aids.

- As understanding unfolds over the years, there is a frequent need to reexamine previously annotated video for visual objects of new interest, or to reclassify annotations. The demands of keeping up with current dives are in natural conflict with reexamining previous annotations. We see automation as a way to focus human talent on areas where creativity and adapting to changing conditions are key, while releasing repetitive activities and quality control tasks to machine approaches. Furthermore, automation will allow us to fully exploit the immense amount of video data already collected that remain underutilized.

The goal of this project is to implement high-payoff technologies that will enable detection and classification (identification[3]) of a visual event in the presence of "noise" in the visual field. We have investigated both research work and commercial systems that do visual image processing and image capture. We apply and implement selected best technologies for our specific problem areas.

This development project will be done with participation and input from

- Science users (Robison group, Barry group, Haddock group, Widder group, Jacoben Stout group, and others),
- Video lab personnel and science researchers who analyze video (expertise on features of interest, "noise" in the visual field, and illumination requirements),
- 600034 Steve Rock's group (expertise on automated video processing and controls), outside organizations deploying cameras on AUVs, and the 600034 project team,
- Research organizations developing neuromorphic vision systems and chips (Kwabena Boahen, [U. Penn], Tobi Delbruck, [ETH Zurich]),
- External collaborators, with their own funding (Dirk Walther [Caltech], Christof Koch [Caltech], Laurent Itti [USC], Pietro Perona [Caltech], Marc'Aurelio Ranzato [NYU], Andres Castano [NASA JPL]).

It is our distinct intent not to do machine vision research as part of this project. Rather, we focus on applying best technologies and proven robust systems that are being

[3] In machine vision sometimes referred to as "recognition".

researched and implemented in the top machine vision research labs. We partner with those labs to offer our compelling application and video database as test bed for their implementations. It is our intent to develop a unique compelling system with wide applicability to marine studies that is well engineered and crisply executed.

Why it is important

MBARI has significant investment in ROV video image capture technology and scientific investigation tied to video analysis. This project implements technology that would leverage the investment in staff and equipment to answer scientific questions. In addition, technologies to add autonomous vehicles and autonomous cameras to the scientific arsenal will be extremely valuable in increasing our ability to sample more thoroughly, more often and in more places, economically.

The ability to autonomously detect and identify an organism could open up new lines of scientific inquiry. In interviewing end-users, we find that the ability to analyze archived video for a key group or species emerged as a disruptive technology[4] that would greatly increase scientific value and use of this expensive asset

Technology to aid analysis of video from cabled observatory cameras will be critical for those cameras to be cost effective instruments. The experience from the H2O observatory, and other efforts placing cameras on observatories, is that it is simply too expensive to manually analyze the continuous (or time-lapse) video from such deployments over long deployments.

Finally, implementation and ongoing investigation of alternative image processing and detection techniques could suggest follow-on innovative work, such as optical-flow based navigation, feature following, habitat identification, extension to acoustic imaging and analysis and so on.

Present state of art/understanding

Visual image processing is a huge field with a rich history of investigation. Two areas we discuss here are 1) feature detection, and 2) classification.

Automatic visual feature detection has been a research and engineering topic in a number of disciplines, including medical image analysis, surveillance, robotics (for example, place recognition) and industrial control. There have been major advances, particularly in the last 10-15 years, which move machine vision from being an esoteric research area to being a field that is delivering robust, mathematically based engineering solutions. Here

[4] The invention and adoption of word processing systems is an example of a disruptive technology, which dramatically changed the way people created and edited documents, compared to the way they were generated and modified when the typewriter and whiteout were the tools of text entry and editing. Here, we mean a capability that changes the kind of questions and the number of questions that can be economically asked.

we present highlights of appropriate technology to motivate the proposed implementation tasks.

There are two classical approaches to visual feature detection. The first is a parametric approach, where a captured image is processed to detect edges, lines, corners, junctions and so on, and then these parameters are used to match some model of the feature of interest. The images and parameters can be 2D, or 3D images can be constructed and analyzed. Video image processing can add multi-frame images to the analysis, comprehending blurring, rotation, noise, and shifts between frames due to motion. This enormous field of research and engineering has yielded surprisingly few robust general systems, but there are systems of focused application that are successful in feature detection and visual event detection. An example of such as system is the rock detection and recognition system deployed on MARS Exploration Rovers Opportunity and Spirit, developed at the Jet Propulsion Laboratory by Andres Castano.

The second active area of research, which has yielded some successful deployed systems, is in the area of surveillance of natural scenes (i.e. outdoors under natural lighting). In modern successful systems Gabor filtering is often used to do noise filtering and feature detection. Gabor filters are a mathematical center excitation-surround inhibition (center-surround) operation that has direct parallels to information processing in mammalian retina and early vision processing in the brain. The experience is that this approach is more general and applies very well to natural scenes, where lighting is less controlled and there is more visual clutter. This approach is a cornerstone of the AVED system implemented to date.

Classification: Many vision systems include learning algorithms, to tune the systems to the particulars of the environment. Modern systems use Bayesian network probabilistic learning systems, extensions of the classical artificial neuronal networks (ANNs) that have had extensive research and practical application. Bayesian learning is a good model for how classification is done in mammalian visual systems. MBARI is in an unusually favorable position to apply learning algorithms to video image analysis, given our wealth of fully annotated video (some 12,000 hours worth) and some 1,000,000 annotated events available as a teaching set.

Perona and colleagues at Caltech have developed the state of the art classification system to recognize objects in natural scenes. This system, in its most advanced form published at the July 2004 CVPR conference, has been demonstrated to recognize 101 classes of objects, being trained with only 6-8 examples of each object. The performance of the system exceeds 90% correct classification. The input to the system, implemented in Matlab, consists of objects segmented from the scene. Segmentation using the AVED system has been successfully demonstrated, and integration of segmentation from a system similar to ours with the Perona “Constellation” system was also presented at CVPR 2004 (Rutishauser, Walther, Koch, Perona). Perona student Marc’Aurelio Ranzanto has developed a more traditional classification system based on features extracted via Local Jets and classification via a Gaussian Mixture Model of learned feature distribution. Ranzato applied this system to images of biological “particles” (microscope images of pollen and biological fluids), which in some ways resemble our transect images, where the objects of interest are relatively small in the scene, and the objects are not oriented. We have applied this system to our video with good success, and we are continuing to integrate and adapt that system as part of our extended project.

Thus, state of the art understanding in machine vision, along with robust programs implementing these algorithms, presents for the first time the opportunity to integrate robust systems to detect, classify and recognize marine organisms from video taken in their natural environment. The potential impact of this integration on ecological work at MBARI, and ecological work in oceanography and fisheries research, could be enormous.

How it relates/contributes to MBARI's mission and strategic plan

This proposal attempts to incorporate leading-edge research and proven robust mathematically based engineering approaches problems in oceanography through visual image processing. Our application of these technologies is to our knowledge novel and new. The project would offer breakthrough approaches to analyzing video from ROVs, AUVs, manned submersibles and towed cameras for species and abundance surveys that would have wide applicability and interest to oceanography. The project would yield critical technology to make use of real time video from cabled observatory cameras on MARS, ORION and other observatory efforts. This high-reward project would unlikely be undertaken at other oceanographic institutions, but is yielding approaches and products of great interest to those researchers and engineers.

b) Approach to the problem (new and continuing projects)

If the proposed project includes multiple individual projects, please list each of these projects as a subheading in the proposal. (For example, Benthic Biology and Ecology subheadings include Biology of deep-sea CO_2 disposal and Cold seep biology – see example on PCO web.) For each subheading, please address the following:

Status of project

We have completed the tasks attempted in 2003, and 2004, and are on track to complete the tasks planned for 2005, with the exception noted below. These tasks include:

- Extending the "vision" program from USC to comprehend the often-sparse data of our underwater video, where many individual frames are "boring" with no visual events to be detected.
- Extending the program to comprehend motion as equivalent to space in video collected from a moving camera, identifying related events in each frame to the object responsible for each event.

- Migrating from proof-of-concept NTSC video copies and frame grab hardware to full resolution frame grab hardware fully integrated with the digital time code-based annotation database.
- Testing the extended program on multi-CPU clusters for video analysis of long video segments.
- Comparing the performance of the program with expert human annotators for both midwater and benthic quantitative transects.
- Demonstrated and integrated a proven classification system to enable classification of single targeted species.

We have not completed the task of Extension to bioluminescent zooplankton (Automated Bioluminescence Video Detection). This task was assigned to Tim Meese, who was not able to allocate significant effort to this task during 2005. His recent resignation makes it highly unlikely that we will be able to address this task in 2005, since this was Tim's area of expertise. We have not completed the task of integrating motion analysis into the event detection module. We have concluded that the technology is not sufficiently developed for us to implement this capability without extensive technology development beyond the scope of our charter. We seek external, self-funded collaboration on that engineering research/development task. We have shifted priority to the tasks enumerated below for 2005 which we think focuses our efforts on high-yield capability, setting up the proposed work for 2006.

Results to date

1. We have developed a process flow to enable automated comparison of automated event detection with human annotation. We have demonstrated algorithms to separate "interesting" events from "boring" events detected by the saliency algorithm and compared results with human annotation. Our collaborators have enhanced the core saliency algorithms to improve detection of long string-like animals such as siphonophores, and have implemented tracking of detected events.
2. We have developed an approach to stream video to our system and capture the results continuously. We have demonstrated the software system running on our compute cluster processing complete video transects and long video segments.
3. We have developed detailed requirements and use cases from MBARI scientists for specific missions that drive publication quality research.
4. We have published, presented and demonstrated our work at the IEEE Computer Vision and Pattern Recognition Conference (CVPR), presented to the MBARI board of directors at MBARI Day of Science and Technology, presented a poster at the MTS/IEEE Oceans Conference, presented a poster at ASLO/TOS and ASLO summer meetings to the marine ecology community, and presented a poster at the Fall AGU session on *artificial intelligence applications in the Geosciences* and a presentation at the spring Joint Assembly meetings session on *How Emerging Technologies Can Advance the Frontiers in the Geoscience*s. Highpoints of our presentations and publications to date include 1) our 2004 CVPR paper, which was double blind peer reviewed by machine vision experts

and selected as one of only 54 out of over 1000 submissions (5% acceptance rate) for oral presentation at CVPR in plenary presentation to an audience of over 500 machine vision and pattern recognition specialists, 2) our concurrent day-long demonstrations of the system at CVPR which was seen by hundreds of machine vision experts, and 3) Rob Sherlock's posters at 2004 and 2005 ASLO summer meetings which were very well attended and received.

5. Analysis of benthic and midwater quantitative video transects for event detection.
6. We have quantified the performance of our detection system: over 80% detection of annotated events in midwater transects, and over 90% detection in benthic transects at Smooth Ridge, with some species detected at the 98% level.
7. We have added a robust object classification module to our system, and demonstrated 85-90% correct classification of benthic organisms examined to date including *Rathbunaster californicus* in Smooth Ridge quantitative video transects.
8. We have completed a requirements analysis of automated event detection, including work flow, and reviewed those requirements with science users and video lab personnel. We have written a draft requirement analysis for classification user interface, which we are reviewing.

   The system has proven to be robust and consistent, requiring no tuning to adapt to changes in lighting, camera settings, turbidity, or other natural conditions. We have set up scripts for different environments (tracking conditions under different camera motions) for midwater transects, benthic transects, or arbitrary movement of the vehicle.

Tasks remaining for 2005:

1. Hold a users meeting, to which the Management Team will be invited, similar to those organized for GULPER and the Multibeam AUV, to gauge the commitment of the user community. This is in direct response to the MT feedback on our 2006 abstract, namely:

"The original AVED project had a well defined goal, made significant technical advances, yet the results were not adopted by the user community (ITD) identified at the outset. Thus the Management Team is concerned whether the user community is really going to adopt the results of this project. Please organize a meeting, to which the MT would be invited, similar to those organized for GULPER and the Multibeam AUV, to gauge the commitment of the user community."

We have met with the user community for our requirements reviews and ongoing discussions several times throughout the year. For example, the video lab staff have identified the capabilities in AVED as a critical component of their vision of the future for their work, as identified in the *Ocean Research and Biodiversity Explorer (ORBE)* initiative idea that they developed (without involvement of the AVED team) in April 2005.

We will call this additional user meeting, but we were unable to schedule it before phase 1 submission, due to cruises and other schedule issues of key participants. We plan to compete this meeting early in the phase 1 period.

2. Integrate HD capture card into Beowulf cluster. - 10 days. We have identified an HDTV linux-compatible capture card, from the same vendor as our current SD linux compliant capture card.

3. Classification/recognition data management and GUI development. - 15 days (D.Cline only), 20 days (A. Chase), 10 days (B. Schlining)

This task enables the AVED team to more easily build and test classes needed for the recognition system, as well as modifications to the recognition algorithms. Is also an incremental step towards making recognition system usable by a non-expert user; this first step is proposed as a modest interface, usable by AVED team as a stand-alone application on our compute server (*Nanomia)*.

This task has four components: (1) install a database to manage the collections of images required for classification (these collections are tied to concepts in the VARS Knowledge Base and reside on the same database server VARS is hosted on), (2) design a simple GUI to enable a non-expert user to manage image collections and run the classification/recognition system, (3) Provide API to the classification subsystem Matlab code for the GUI, and (4) systems engineering and task coordination to tie all components together. Subtask 1 and 2 are to be done by Brian and Andrew respectively; 3 and 4 by Danelle Cline.

It currently requires days of effort to collect, organize and stage a training set for a particular target organism. Most of this effort is computer work, involving VARS identification of candidate annotations, identifying specific frames to capture, running capture scripts, and organizing the resulting files. Some of the work requires science analysis to identify the organisms and to check that all of the images are in fact representative of the target species. The amount of that work varies with the quality of the images and the ease of human identification.

4. Install remote VCR control software for digital data recording (DDR) - 5 days (D. Cline)

This task enables recording of SD video directly into our RAID for processing, thus reducing the time to transcode to the correct format. Completion of this task will enable more data to be processed by reducing the overhead in transcoding, as well as providing a digital recording database to better manage video data, instead of the file/directory organization which is currently used.

The task is to install remote VCR software to allow user to set IN/OUT timecodes and control VCR from the DDR software. This provides remote control of the VCR connected directly to the AVED capture workstation *(Ephyra)* from the digital data recording software on the Beowulf cluster.

5. AVED database installation - 15 days (Brian Schlining) Task is to build and install AVED database to store AVED metadata. We plan that the VARS database will eventually include a tie to this metadata database. This task is to install the database and provide data access objects to access the data (e.g. from another application like a user interface). The task is an incremental step towards providing AVED metadata to the MBARI community, or external users, for query through VARS.

Publications (from the last 5 years)

1. Edgington, Duane R., Ishbel Kerkez, Danelle E. Cline, Dorothy Oliver, Marc'Aurelio Ranzato, and Pietro Perona, Detecting and Tracking Animals in Underwater Video. *Proceedings IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, San Diego.CA, 2005. (peer reviewed).

2. Edgington, D.R., Kerkez, Ishbel, Oliver, Dorothy, Cline, Danelle, Sherlock, Rob, Robison, Bruce, Kuhnz, Linda, Walther, Dirk, Marc'Aurelio Ranzato, Perona, Pietro. Detection and Classification of Selected Midwater and Benthic Organisms in Underwater Video. in *International Ocean Research Conference Proceedings*. June 2005. Paris, France: UNESCO.

3. Edgington, D.R., Kerkez, Ishbel, Oliver, Dorothy, Cline, Danelle, Kuhnz, Linda, Walther, Dirk, Marc'Aurelio Ranzato, Perona, Pietro. Detection and Classification of Rathbunaster Californicus in Underwater Video. in Eos. Trans. AGU, 86(18), Jt. Assem. Suppl., Abstract U24A-05, 2005.

4. Edgington, D.R., Walther, D., Cline, D., Sherlock, R., Koch, C. Automated Annotation of Midwater Video Transects: Taking a Look at the Distribution of Poeobius meseres in Monterey Bay During and After an El Nino. in *ASLO Summer Meeting*. June 2005. Santiago de Compostela, Spain.

5. Edgington, D.R., Kerkez, I., Oliver, D., Kuhnz, L., Cline, D., Walther, D., Itti, L. Detecting Benthic Megafauna in Underwater Video, Eos Trans. AGU, 85(47), Fall Meet. Suppl., Abstract OS43B-0551, pg. F1086, 2004.

6. Walther, Dirk, Duane R. Edgington, Christof Koch, Detection and Tracking of Objects in Underwater Video. *Proceedings IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, Oral Session 2B, Washington, D.C., 2004. (peer reviewed).

7. Edgington, Duane R., Dirk Walther, Danelle E. Cline, R. E. Sherlock, and Christof Koch, Detecting and Tracking Animals in Underwater Video. *Proceedings IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, demonstration, Washington, D.C., 2004.

8. Edgington, D. R., D. Walther, D. E. Cline, R. Sherlock, and C. Koch. Detecting and tracking animals in underwater video using a neuromorphic saliency-based attention system. American Society of Limnology and Oceanography (ASLO) Summer Meeting, Savannah, Georgia, USA, June 13-18, 2004.

9. Duane R. Edgington, Dirk Walther, Danelle E. Cline, Rob Sherlock, Karen A. Salamy, Alexis Wilson, and Christof Koch. Detecting and tracking animals in underwater video using a neuromorphic saliency-based attention system. American Society of Limnology and Oceanography (ASLO)/The Oceanography Society (TOS) Ocean Research Conference, Honolulu, Hawaii, USA, February 15-20, 2004

10. Walther, D., Edgington, D. The art of seeing jellies. *The Neuromorphic Engineer* (2004) **1**, 6-6

11. Edgington, D. R., D. Walther, D. E. Cline, R. Sherlock, K. A. Salamy, A. Wilson, and C. Koch, Detection of Visual Events in Underwater Video Using a Neuromorphic Saliency-based Attention System, *Eos Trans. AGU, 84*(46), Fall Meet. Suppl., Abstract H11F-0912, 2003

12. Walther, D., Edgington, D., Salamy, K., Risi, M., Sherlock, R.E., Koch, C. (2003). Automated video analysis for Oceanographic Research. IEEE Computer Vision and Pattern Recognition Conference (CVPR) Proceedings, Madison Wisconsin, June 16-22, 2003. (peer reviewed).

13. Edgington, D., Walther, D., Salamy, K., Risi, M., Sherlock, R.E., Koch, C. (2003). Automated Event Detection in Underwater Video. *MTS/IEEE Oceans 2003 Conference Proceedings*, San Diego, CA, September 2003. IEEE Press.

14. Edgington, D. R. and D. Walther (2002). Detection of visual events in underwater video using a neuromorphic saliency-based attention system. *Report on the 2002 Workshop on Neuromorphic Engineering*, pages 25 – 29, Telluride, CO, July 2002.

15. Shih Chii Liu, Jorg Kramer, Duane Edgington, Tobi Delbruck. Making direction selective cells using AER retina and spiking neurons. *Report on the 2001 Workshop on Neuromorphic Engineering*, pages 94 – 96, Telluride, CO, July 2001.

Plans for 2006

Science benefits

- The identification of individual organisms in existing and future video would be of significant use for analyzing animal distribution and for habitat characterization. The identification of target species will eventually lead to the ability to characterize whole communities
- Event detection will be of great use in analyzing large volumes of video from stationary observatories

Science Applications:

1) A key science application of our system that has emerged is the quantitative analysis of a large set of videos for single species of interest. In each case, being able to fully analyze the large video archive that MBARI has collected for a species or group would lead to multiple publications of impact. For example:

A. Radiolarians have recently emerged as a species of interest to Dr. Haddock, with a hypothesis that they could be major contributors to carbon cycling in the deep-sea. The AVED system with recognition capability for radiolarians would be a way to quantify their impact. The alternative is to re-annotate much of the 11,000 hours of video MBARI has collected, since radiolarians have only recently been routinely annotated, and have not been in the "top 10" list of creatures that the midwater annotators have focused their annotations on. The radiolarians are in the video, but re-annotation, by machine or by humans, will be required to develop the scientific story.

   To quote and paraphrase from Dr. Haddock's 2004 proposal, radiolarians are an example where many potential breakthroughs can come from answering basic questions and from quantifying organism occurrences in midwater video. As with most jellies, these protists can only be collected using specialized sampling methods, so that live radiolarians have received almost no scientific attention. This lack of effort is in stark contrast to their importance in many ecosystems: as was learned during blue-water dives on the Hawaii transit, polycystines can be dominant members of the ecosystem in oligotrophic waters. These numbers, when considered along with their siliceous skeletons, suggest that they may replace diatoms as the major factors in silica cycling outside of coastal upwelling zones.

   In the deep sea, Haddock and colleagues have discovered that, contrary to a century of literature, all of the tuscaroid radiolaria examined are colonial. This morphology is key to their method of feeding, allowing them to act as "marine Velcro", accumulating many times their own biomass of detritus from the water column. Because of their ability to accumulate and repackage marine snow, they will play a role in carbon transport in the deep sea. Quantitative annotation of their distribution will enable calculation of contribution to carbon cycling in the deep ocean. That contribution could be significant.

B. *Poebius mesmeres* has emerged as an important midwater species. In early studies midwater ecologists did not appreciate how widely spread they are in the water column. *Poebius mesmeres* is now thought to be important in the ecology of eastern North Pacific. The distribution and the effect of El Nino events on their abundance is unknown. A particular question of interest in the Robison group is whether they have a bimodal distribution above and below the oxygen minimum layer. Ironically, they are hard to annotate because they are so abundant. Like radiolarians, they have not been annotated quantitatively until recently in transect data. The task to study these creatures is to carefully annotate midwater transects for *P. meseres*, and to correlate the annotations with depth and environmental conditions. An automated *P. meseres*,classifier with predictable performance would directly yield scientific results.

C. Jane Lee, a Ph.D. student in adjunct Bill Hamner's lab, is studying the giant siphonophores *Praya spp.* and *Apolemia spp*. She is not using the video database for her work because these animals are not quantitatively annotated. Again, an automated re-annotation of video transects for these large animals would directly enable scientific results.

D. Dr. Kevin Raskoff at CSUMB is working with the MBARI midwater group to study *Aegina spp*., an extremely abundant midwater genus. Appreciation of their wide distribution is relatively recent. An automated re-annotation of video transects for these abundant animals would directly enable scientific results.

E. Sergestid shrimp belong to another group that has been inconsistently annotated. Their distinctive color, shape and motion make them another candidate for automated re-annotation of video transects of these abundant animals that would directly enable scientific results.

F. Dr. Barry has told the AVED team that without equivocation his scientific investigations would make heavy use of a system that could predictably recognize even a few specific benthic species or groups. An example is Ophiuroidea, a large group of echinoderms that includes the brittle stars (Ophiurida) and basket stars (Euryalida). While we have found that benthic transect video is very amenable to automated event detection and classification, it is nonetheless very tedious and time-consuming for humans to annotate quantitatively. The applications for benthic ecology include habitat identification and classification. Areas for us to explore include classifying environmental and geological features in the video in addition to the animals, for habitat characterization. In the future, real-time classification would be very powerful for benthic transects, to enable zooming in of a slaved camera to enable measurement of size, reproductive state, substructure (e.g. larval forms to be released), and parasites (such as those found on *Rathbunaster*).

G. A vision that emerged from the MBARI Video Lab staff (ITD) in their discussions in April 2005, while exploring MBARI initiative ideas, is the following: Because it is not possible to identify, much less quantify or describe, every biological and geological observation in the vast MBARI video archive, our primary focus will be to enable automated data mining on the existing ROV video archive. In order to extract structured ecological community information, we first will train our system to automatically identify and quantify key species observable in a selected habitat. The resulting summary of target organisms as well as a still image or moving images would be presented to users for review, validation, and further analyses. We envision this technology being useful in both fine- and rough-scale community structure characterizations. For example, the speed at which analysis of quantitative video transects (QVTs) in the water column are analyzed would be greatly increased if the most common species, such as the polychaete *Poeobius meseres* or the narcomedusa *Aegina citrea*, could be automatically quantified. To enable statistical analyses at a larger scale, species distribution and abundance could be determined for higher taxonomic groups. For example, enumeration of fish, asteroids, holothurians, urchins, corals, and anemones over long periods of time would provide valuable ecological modeling data for selected sea floor communities. To further enrich these results, benthic habitat types could be identified by classifying the seafloor's geologic features such as boulder, rippled sand flat, or talus slope. Another potential application of this technology would be to provide temporal and spatial data for fisheries use. For example, being able to track the abundance and distribution of the squid *Doscidicus gigas* would provide valuable information for local and foreign fisheries. Similar data could also be provided for benthic fisheries if, for example, *Sebastes spp*. (rockfish) could be distinguished from other species and enumerated.

Once mining of the existing data archives has been proven, the vision is to adapt the system for use during real-time video data collection. The general scenario is that a user annotates in real-time using VARS while AVED runs concurrently in very near real-time. Since it is impossible for researchers to focus on every recorded observation in real-time, we will use the same data mining technology to assist scientists in detecting those objects that will inevitably be missed by human annotators. Such a tool would allow scientists to immediately begin their data analyses and possibly adjust the goals of the current mission to best take advantage of the situation.

Engineering/technology development

1. Integrate summer student Jerome Mariette's segmentation algorithm improvements. - 15 days. Summer student Jerome Mariette worked on improving the AVED segmentation algorithm and the results of his work were promising. This task involves porting his Matlab implementation to faster c/c++ implementation within the AVED code. This will improve segmentation of translucent and low contrast animals over the current segmentation algorithm.

2. Design and implement seamless integrated AVED tool. Running the AVED system currently requires significant expertise in Linux. AVED is also a complex, distributed system with two compute servers, one Beowulf compute cluster, and two RAID storage

devices. Three steps are currently required to use the AVED system: 1) video is digitized into the correct format for processing, 2) the video is then processed with AVED event detection software, and lastly 3) objects from the AVED event detection are tested with the AVED Classification software. Each step requires moving files around, executing scripts, and/or running matlab code. And each step does not necessarily run on the same machine, making it even more difficult to use.

Toward the goal of implementing a more seamless workflow, we have made incremental improvements to the workflow in 2005 and have identified the necessary steps to achieve a fully integrated tool in 2006. This task is to complete these steps required to build a seamless system, and handover AVED to the MBARI science users. The initial application targets both quantitative work and Dive Annotation Assistant for the video lab.

Subtasks include:

2a. Design and test AVED control and edit GUI. 40 days
Task involves designing a user interface to execute the AVED event detection software and edit the results. This task is to put a front-end GUI to the job controller and customized scripts used to execute the AVED software. It is also to build a tool to preview and edit the output of the AVED system.

2b. Extend AVED classification/recognition GUI. 20 days
Task includes extending work from 2005 to allow user to run classification system remotely from other desktop machines. This will allow video lab or other science staff to use this tool for quantitative work.

2c. Complete AVED-VARS integration. 15 days
Task includes completing the database link between AVED data and VARS data. Will also involve some work with VARS to import and query AVED data. This task would be ideally suited for Brian Schlining to participate in because of his special expertise in VARS.

3. Project management and system design. This project is managed as a traditional engineering project, with requirements specifications, implementation specifications and WBS, and weekly project team meetings. Project management tasks are divided between the project manager and lead engineer as in traditional software engineering projects. This task acknowledges the time required for project management and system design.

4. Integrate, test, debug. This task comprehends effort to work with users on specific uses of the system, conduct tests, analyze results, and address any problems or needed enhancement revealed by the tests.

5. System improvement. This task comprehends effort to continuously improve the system, by addressing improvements needed in performance, accuracy, or user experience as identified by users during the development period.

6. Eye-in-the-Sea exploration. In this task, we explore application of AVED to fixed camera ocean observatory applications. In particular, we explore the problem of identifying and classifying events of interest from continuous streaming video, or time lapse clips. An example is the case of identifying frames where a fish is in motion in front of the video camera, so long as that fish is not a hagfish (which are visual noise in Eye-in-the-sea science use cases). This task will be coordinated with adjunct Edie Widder, to further develop science use cases. We expect the output of this task to be detailed development specifications, based on explorations and requirements analysis, for proposal in follow-on years.

Coordination with external grant applications:

The AVED team has submitted two proposals to NSF for potential external funding:

The first, 701556 (Robust marine Habitat Classification), a collaborative proposal with Prof. Itti's lab at USC, proposes to apply new technology developed in Itti's lab to MBARI underwater video. This method extracts key signatures of scene characteristics from spatial spectral information in video images of the scene. The proposal applies this method to underwater video, in concert with the AVED system of event detection and object classification, to classify images into habitat types in benthic video. That proposal complements, but does not supplant, any of the tasks proposed in this MBARI proposal.

The second, 701557 (Mining Video Library), with no funded external collaborators, funds technology development to 1) improve the detection module, including motion saliency, feature-map enhancements for long range interactions within the scene (example long thin animals such as *Nanomia bijuga*), and learning of feature sets for optimal event detection, 2) improve the classification module, and 3) integrate the event detection and classification modules, enabling feed forward and feedback signaling between the modules to enhance detection and classification, as well as improve the quality of segmentation of objects of interest. This proposal enhances the work in this MBARI proposal, but again does not duplicate it.

Both of these external projects would fund at the end of 2005, and run 3 years currently.

Field programs

None planned.

Data (availability outside project team, uniqueness)

The team members will do data reduction and analysis as part of the project. Additional analysis of dive video (for example, evaluating the output of the visual event detection system, and special annotations of selected videos marking frames at the coordinates of events (animals or rocks) will be requested of our video lab project support members or MBARI science department members. The data will be stored in the VARS database along with human annotations from video analysis and environmental data from other sources

Significant changes from previous project proposal (continuing projects only)

n/a

c) ***Specific deliverables (new and continuing projects)***

- **Products from research and development (instruments, tools, documentation, publications, presentations, etc.):**

1. The main product from this project will be a working implementation of the visual image event detection and classification system processing digital BetaCam and HDTV archived videos on shore at full resolution integrated with the VARS derived time-coded annotations, essential to individual scientists to produce publication quality results on the occurrence of specific organisms in analyzed video.
2. Collaborations and explorations will be documented in whitepapers placed on the Visual Image Event Detection website and project presentations.
3. Results of the research and development will be presented at selected conferences and workshops, including the MTS/IEEE Oceans Conference, ASLO, IEEE/ACM Computer Vision and Pattern Recognition Conference. Results from the development will be published in selected peer reviewed IEEE publications as appropriate.

- **How success will be measured**

1. By working implementation of shore-side visual event detection and classification system, useful to science staff.
2. Documentation of explorations and collaborations.
3. Accepted presentations and publications at appropriate venues.
4. Successful external and/or internal follow-on project proposals.

d) ***Project team (new and continuing projects)***

Jim Barry, Andrew Chase, Danelle Cline, Dan Davis, Duane Edgington, Kevin Gomes, Steve Haddock, Linda Kuhnz, Bruce Robison, Brian Schlining, Rob Sherlock, Nancy Jacobsen Stout, Susan von Thun.

Project Manager: Duane Edgington
Lead Scientist: Jim Barry, Steve Haddock, Bruce Robison, Edie Widder
Lead Engineer: Danelle Cline

External Collaborators, with their own funding: Dirk Walther, [Caltech], Christof Koch, [Caltech], Laurent Itti, [USC], Pietro Perona [Caltech], Marc'Aurelio Ranzato [NYU], Andres Castano [NASA JPL]

e) *Project lab space (new and continuing projects)*

We require lab space of about 24 feet by 15 feet, with typing-height surfaces all around, for our video-lab test setup (computer, video capture card, video monitor, tape deck, computer monitor), 3 workstations, and notebook computers. Our current project lab space in A244 is adequate for the project as currently scoped, with 3-4 people working on the project at any one time. We require fiber optic cable access to our beowulff cluster in the building A computer room, which interfaces to our capture deck and capture control computer.

3) BUDGET JUSTIFICATION (new and continuing projects)

1.	Project computer disks and upgrades.		$3,000
2.	Production cluster system includes:		$55,000
	Beowulf cluster	$35,000	
	HDTV interface card	$5,000	
	high speed interconnect	$5,000	
	high speed disk storage	$10,000	
3.	Matlab simulink video processing toolkit		$1,000
4.	Presentation, collaboration and participation in Neuromorphic Engineering workshop		$3,500
5.	Lab supplies and stock		$2,000
6.	Team travel to collaborators (Caltech, USC, NYU, JPL)		$5,000
7.	Travel to present work (ASLO, AGU, CVPR)		$7,000
8.	Misc software tools		$3,000
9.	Compression software tools		$2,000
10.	Materials and supplies for summer intern		$1,000

11. Project reference books	$900
12. Matlab Image Processing Toolbox (shared license)	$4,000
13. UI tools	$1,000
14. fiber-to-SDI/SDI-to-fiber converters and fiber cables	$3,000

We are requesting an additional, new compute cluster. Our current development system is 3 years old; new cluster hardware is about 2x higher performance for the same cost as our original system. As we move to production use of our system for science, we will need a development system (our original cluster) to test new enhancements and bug fixes, while leaving a production system, of highest performance, available for uninterrupted science use. We will share streaming RAID disk array, data storage RAID disk array, and other components where possible. The current processing bottleneck in our system is core algorithm execution speed, and inter-processor communication speed. The proposed new system would directly address these two bottlenecks.

The Matlab Image processing toolbox, to be installed on the MBARI Matlab share, directly supports development (under Linux) of improvements in classification for the project. The compression software tools are for commercial tools to do high-quality video compression to explore potential performance and storage efficiency boosts in the production system. The UI tools item is a commercial package for user interface development, in support of that task. Other items in the budget are for recurring expenses.

4) MILESTONES (new and continuing projects)

	Milestone 1	**Milestone 2**	**Milestone 3**	**Milestone 4**	**Milestone 5**		**Total Resources**[5] for each resource category for all milestones listed here
Milestone Date	Q2 2006	Q4 2006	n/a	n/a	Q3 2006		
Milestone Description	Integrate improved detection and segmentation	Integrated AVED tool	Project Management	Integration, test, debug	Eye-in-the-Sea exploration		
EE							
SE	15	65	35	40	20		175
ME							
Machine Shop							
ITD Video lab		20	10	10			40

[5] This number should be the same number you entered on the resource requests for this particular resource.

5) RISK MATRIX (development and infrastructure projects only)

Please carefully read Section 5.2.3 in the instructions for internal proposals for an explanation regarding the risk matrix. This table must be completed for **ALL** development and infrastructure proposals (new and continuing). The instructions can be found at http://mww.mbari.org/resources/2006_Proposal_Process/Internal-only_Instructions_2006.htm.

	Technical Risk	Organizational Risk	Tech Transfer Risk	Labor Estimate Risk
High				
Medium	x			x
Low		x	x	

The technical risk of AVED, initially quite high, has been largely addressed in the previous 3 years of effort on 900260, where we addressed the high technical risk elements through proof of concept demonstrations and risk-driven development. We see the proposed work as being of moderate technical risk, building on our previous work and requiring straight-forward software engineering tasks. There is no technology creation ("research") in the proposed tasks.

The organizational risk is low: we have built a strong base of enthusiastic internal adopters for our planned system. We have developed the expertise in house for the proposed work.

The tech transfer risk is also low: through our extensive presentations, presence at international science conferences, and publications we have built a large external community of enthusiastic adopters for our planned system. We have several requests in hand from external ocean science and engineering groups who want to use our system when we make it available. At this time there is no competing solution.

The labor estimate risk is moderate: we have not done these exact technical tasks before, and we have not developed detailed WBS plans. However, the moderate technical risk mitigates concern about estimate errors; the tasks are similar to those in other projects accomplished by the experienced development team members.

Questions we have addressed:

Will the project allow a broad segment of the community to answer important oceanographic questions?
How unique will the contribution be?
Does the project play to MBARI's strengths & directions?
Does the proposal demonstrate a good understanding of the prior art and foundation technologies?
Does the project have strong backing of, and involvement from, both Science and Engineering?
Has the team demonstrated the ability to champion developed systems to the larger community?
Is there a plan for responsible project management in keeping with the size of the effort?
How effective is the project in leveraging off existing capabilities (rather than reinventing the wheel)?
What is the timeliness of the project?
Is there clear and realistic set of functional requirements or method to derive functional requirements?
Are risks understood, is the project well scoped, and has a cost benefit analysis been performed?
How adequate is the project plan (WBS, schedule, risk mitigation plan, schedule of design reviews)?
What are the suitability of the project team expertise and the availability of key staff?
Has this team been productive with prior MBARI support?
Has the team demonstrated cost consciousness, clear budget justification, and careful budget management?
Does the assembled team have the requisite skills and skill level for the proposed research?
Has the project team demonstrated enthusiasm and commitment?
Are requests for shared MBARI resources (ship time, personnel, etc.) well explained and justified?
Does the project have a transition plan, especially to the larger community?