Interdisciplinary Undergraduate Research with Focus on
Hyperspectral / Multispectral Imagery

Stefan A. Robila
Department of Computer Science
Montclair State University, Montclair, NJ
robilas@mail.montclair.edu

ABSTRACT

The paper introduces a framework for the development of an interdisciplinary research approach in Computer Science / Computer Engineering education. Given the difficulty level required in many current research directions as well as the curriculum constraints on the number and type of courses that need to be taken by CS / CE undergraduate students, it is a challenge to attract them to participate in interdisciplinary research projects. At the same time, developments throughout scientific disciplines rely on computation based solutions. Our approach is to identify several factors that need to be addressed when designing an interdisciplinary project that relates to CS / CE, such that it corresponds to the current curriculum restrictions. Based on these factors we present several examples of projects related to hyperspectral images, a type of data increasingly used in geosciences. We believe the experience described here can be extended to other areas of interdisciplinary research.

1. INTRODUCTION

In an ever-changing world, computing sciences have to reassess their role and actively adapt to the needs of the society. Today, computers are used with significant success in all sciences. The progress brought by relatively inexpensive and efficient computing power has given birth to new sciences on their own such as bioinformatics, medical informatics, informational linguistics, etc. In this context, Computer Science and Computer Engineering encounter a growing pressure in redefining themselves. They have been interdisciplinary sciences from the beginning, being thought as supporting activities in computation intensive problems. While they have since evolved into science of their own, they continue to stay at the basis of most of the interdisciplinary projects. While performing interdisciplinary research leads to faster progress across disciplines, it raises new challenges in formulating valuable projects for undergraduate research. On one hand, the CS / CE student is faced with an increasingly tight curriculum in order to assimilate the growing base of knowledge in the field. On the other hand, interdisciplinarity implies that the student will need to become exposed to topics outside Computer Science and Engineering, most probably not covered anywhere within the allowed sequences of courses.
To overcome these challenges, we suggest an analysis of the scientific environment with regard to interdisciplinary. The interdisciplinary approach in solving problems is encouraged both by higher education institutions as well as by granting agencies such as National Science Foundation. In this paper we address these trends and then analyze how they relate to computer education. We further develop a group of factors that can be used to characterize a research project and apply them on examples from our active research. As a case study, we present some considerations on how hyperspectral image processing, a research area from geosciences, can constitute an active research topic in CS / CE and how undergraduate students can be involved. Developed in the last two decades, hyperspectral image technology has brought significant improvements in remote sensing\(^1\). However, the relatively large data size, as well as the often slow iterative optimization processes involved leaves room for considerable contributions.

The paper is organized as follows. In the next section we provide a brief review of the concept of interdisciplinarity and the major driving forces behind it and we discuss several factors that are used to characterize a research project. In Section 3 we present a case study of our work on hyperspectral imagery and the undergraduate research projects that have resulted. The conclusions are presented in Section 4 followed by the references.

2. INTERDISCIPLINARY RESEARCH IN COMPUTING SCIENCES

Research is characterized as interdisciplinary if it involves the participation of two or more fields of study. While the idea of interdisciplinary work has always existed throughout history, as science fields have become more and more advanced, its importance has only increased. Interdisciplinary science has been shown to contribute to speedup in discovery and is gaining an increased visibility within the general population\(^2\). Funding agencies, advisory groups, and professional organizations also support the trend. In March 2003, the National Academy of Sciences (NAS) has formed the Committee on Facilitating Interdisciplinary Research, to “examine the scope of interdisciplinary research and provide findings, conclusions, and recommendations as to how such research can be facilitated by funding organizations and academic institutions”\(^3\). The initiative emphasizes the commitment of NAS towards a systematic support for interdisciplinarity and will likely produce an important reference document used by government agencies and academic institutions.\(^4\) The National Science Foundation (NSF) also recognizes the importance of interdisciplinary research by funding numerous cross-directorate initiatives aimed at fostering cooperation among disciplines.

Computing sciences such as CS and CE are interdisciplinary fields. They draw characteristics from Mathematics, Physics, and Chemistry, etc. At the same time, the computing field has matured on its own and requires highly specialized skills. To reflect this, the CS / CE curriculum is continuously modified (usually through addition) making it increasingly difficult to leave space for introduction of topics from other disciplines. This affects the ability of professionals to acquire a broad scientific knowledge and negatively impacts on their involvement in interdisciplinary research.
Various solutions have been proposed to address this problem. One approach is to design specialized programs such as science informatics (bioinformatics, geoinformatics, information science). Through them, the students are exposed to a variety of science courses and learn to integrate computing skills in various disciplines. Alternatively, non computing departments have started to increase the number of computer science courses in their curriculum or have included computing components in pre-existing science courses. While these approaches have certain values, they also tend to reduce the level of proficiency.

We suggest that interdisciplinary research should be seen supporting the Computer Science or Engineering curriculum through undergraduate research projects. These can be either part of classes or can take the shape of independent study or summer programs. They are meant to enrich the undergraduate experience without changing the main focus, a solid CS / CE background. In devising interdisciplinary undergraduate projects we suggest that the following issues need to be addressed:

**Background knowledge needed** In order to attract students at an early stage in their studies it is preferable to devise projects that would require relatively basic computing knowledge. This means that the student has taken the first computer courses in the sequence (such as CS1, CS2). Some projects require basic image understanding or limited knowledge of calculus and differential equations. In many instances, it is unlikely that the student has any prior background in the science involved in the project. The faculty guiding the project also needs to have a basic understanding in the field. This can be achieved through independent research, auditing several courses, or participation in joint projects with non-computer faculty. In addition, several grant programs provide funding for faculty to specialize in new areas. One example is the Directorate of Mathematical Sciences’s IGMS program that provides funding up to $100,000 for a faculty to become knowledgeable in a new field.

**Time needed to complete the project** When the effort to accomplish the task goes beyond several weeks of work, the project becomes more appropriate for a semester long independent study or for a summer research project. If the effort can be limited to several weeks, the project is deemed appropriate for class projects.

**Knowledge that needs to be acquired** This knowledge has to conform to CS / CE curriculum while providing the student an interdisciplinary experience. The knowledge should also preferably offer the student the capability to further develop as a professional.

**Outcomes of the project** There are two important types of outcomes: for the student, and for the faculty’s research project. At the end of the project, the student must have the ability to have performed original work involving design and coding. An outcome will constitute the resulting application accompanied by a written report detailing the obstacles encountered by the student as well as the solutions designed to overcome them. As a second desired outcome I list the preparation and presentation of short research posters or papers to local and national professional meetings such as ACM, IEEE student meetings and symposiums. Finally, a valuable outcome is the student exposure to active cutting edge research preparing her / him for a possible future career. In terms of the relevance to the faculty’s work, the undergraduate research project work must be a meaningful contribution. Given the time spent on direct work with the students, the
benefits of the undergraduate project must overcome the benefits that would have been achieved by working alone on the problem.

Logistics of the project (hardware and software) Often, organizing projects in new fields require significant investments for specialized software packages, data, and equipment. It is desired that the college or the university provide support to these endeavors as part of strategic plans to encourage interdisciplinary work.

3. CASE STUDY – HYPERSPECTRAL IMAGERY

3.1. Background
Remote sensing is generally described as the measurement, from a distance, of spectral features of the Earth’s surface and atmosphere. These features are recorded by satellite- and aircraft-carried instruments and are usually stored as digital data. Remote sensing applications include meteorological modeling (through satellite acquired data), geological surveys (reflectance in specific wavelengths is important for mineral exploitation), environment monitoring (water pollution management using observations on algae development), agriculture (crop development and yields), forestry (forest classification), etc. In general, the sensed data is collected as images (spectral images or spectral bands), with each image corresponding to intervals of wavelengths. Each element from the image (pixel) is associated with a certain area of the scene surveyed and with its spectral response. A collection of spectral images over several wavelength intervals for the same scene is called a multispectral image.

Figure 1 presents examples of spectral images for distinct wavelength intervals ranging from visible to middle infrared (bandwidth 10nm). The image represents a residential area. For vegetation, we note the bright pixel intensity in the middle infrared bands and the low intensity for the visible bands. The roads tend to have the same level of intensity across the visible bands as well in the near-infrared band. Also, the rooftops show considerably high intensity values in all the bands. This is due to the high reflectance of the material. In processing multispectral data, it is a common practice to define pixel vectors as the vectors formed of pixel intensities from the same location, across the bands as represented in Figure 2. Since each pixel corresponds to a certain region, a pixel vector will represent the spectral information (collected in the multispectral image) for that region. The number of bands produced by a multispectral sensor is at most of the order of tens. This leads to an inherent limitation. If the goal is to capture the spectral reflectance of the classes over a large bandwidth, the multispectral sensors will have spectral bands each covering large bandwidths. Large bandwidths make differentiation between classes more difficult.

Hyperspectral images are remotely sensed data sets where the spectral measurement is performed using hundreds of narrow contiguous wavelength intervals. Usually, hyperspectral sensors cover wavelengths from the visible range (0.4μm-0.7μm) to the middle infrared range (2.4μm). Due to the narrow bandwidth and the abundance of observations, the pixel vector for each pixel location resembles a continuous function of wavelengths. This function describes the reflectance of the material for wavelengths within the frequency interval covered by the sensor. Figure 3 provides examples of these functions (also called spectra) for vegetation, roads, buildings, and glass. The
A hyperspectral image was produced using the Hyperspectral Digital Image Collection Experiment (HYDICE) sensor. The HYDICE instrument is an airborne sensor with 210 spectral bands, 10nm bandwidth covering wavelengths in the range 0.4μm-2.4μm. The size of the images produced is 320x320 pixels and the spatial resolution 0.75m. The data were provided by Spectral Information Technology Application Center (SITAC).

Most of the practical applications for hyperspectral imagery are derived from similar applications based on multispectral imagery. Two very important applications are classification and target detection. In the case of classification, the goal is to group the pixel vectors that share common characteristics in the same class. It is hoped that these classes correspond to different materials present in the scene, and that, this grouping would allow automated delimitation and quantification of them. When information (in the form of pixel sets already assigned to classes) is provided for training, the classification is called supervised. Both supervised as well as unsupervised classification techniques use a wide array of measures (distances, estimation of the probability distributions of the classes, etc.) and assumptions (Gaussianity, limits on number of pixels, number of classes, values of the standard deviation for a class within each band).

![Figure 1. Examples of spectral bands of the same scene. (a) 0.6μm (visible-green), (b) 0.74μm (visible-red), (d) 2.25μm (middle infrared).](image1)

![Figure 2. (a) Example of a multispectral image. (b) Organization of a multispectral image. Pixel vectors are formed of the pixel values for the same coordinates.](image2)
Figure 3. (a) Hyperspectral image. Spectra (pixel vectors) for various classes present in the images (b) road, (c) vegetation, (d) building, (e) glass panel.

In target detection, the goal is to search for pixel vectors that either indicate the presence of a certain material or present characteristics that differentiate them from the surrounding background. Although it can also be considered as classification, the major difference in target detection is that the target class is sparsely populated\textsuperscript{11}. Moreover, the target or portions of the target often occupy areas smaller than the one covered by a single pixel, leading to difficulty in separation.

Work in hyperspectral imagery often requires only small financial commitment for software and data. While commercial software packages can be used, many freely available packages also exist. Table 1 lists several of the packages. We note, however, that many of the projects do not require specialized projects but rely on C, C++, or Java implementations. Hyperspectral data is also freely available through various sources. Hyperspectral cameras are still relatively expensive, with the lowest price currently being in the range $30,000.
3.2. Projects

Many of the target detection or classification methods involving hyperspectral imagery require a
experience with optics, multivariate statistics, and differential equations, apart from relatively
solid programming skills. Thus, at first sight, studying this type of data may prove to be too
difficult as topics for undergraduate research in Computer Science or Computer Engineering.
However, one must remember the main concerns related to computer processing of hyperspectral
images. Some relate to data size, high computational costs or communication and
synchronization of hyperspectral sensors with the computer system. All these are valid
computing questions.

Based on them, we have devised several activities that can be accomplished either as class
projects or through independent study. The laboratory setup is minimal, involving several
workstations, programming language of choice (Java or C, C++), and image sets available
through various websites (such as IEEE Geosciences and Remote Sensing Society).

<table>
<thead>
<tr>
<th>Product</th>
<th>Company / Web address</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENVI</td>
<td>Research Systems (KODAK)</td>
<td>Cost</td>
</tr>
<tr>
<td>Imagine</td>
<td>ERDAS (Leica)</td>
<td>Cost</td>
</tr>
<tr>
<td>Geomatica</td>
<td>PCI Geomatics</td>
<td>Cost</td>
</tr>
<tr>
<td>Multispec</td>
<td>Purdue University</td>
<td>Free</td>
</tr>
<tr>
<td>Hypercube</td>
<td>US Army Corps of Engineers</td>
<td>Free</td>
</tr>
<tr>
<td>MicroMSI</td>
<td>National Geospatial Intelligence Agency</td>
<td>Free</td>
</tr>
</tbody>
</table>

Table 1. List of software packages for hyperspectral imagery processing

Image Processing

a) Anomaly Detection. One major area of study in hyperspectral imagery is target detection.
Unlike supervised target detection that requires access to libraries of signatures and
understanding of distance measures between vectors, unsupervised target detection deals with
identifying “anomalies”. The anomaly identification can be done through edge or clutter
detection. Figure 4 shows the results of such application where the data has been just filtered
using an application written in C. The targets are rows of panels, undistinguishable through
human eye but visible in certain spectral bands. The attractiveness of such project relies in the
fact that most of the target detection is based on inspection of single images. Thus, projects
involving anomaly detection reduces mainly to basic image processing. The project is
appropriate for students that are or have taken the CS II course. While it can be used as class
project in an Image Processing course, it can also be used as stand alone project. For the student,
the outcomes of this project are an improved understanding of the image structure, and of the GUI and file processing settings. For the hyperspectral imagery research, this type of activity helps the scientist study new techniques faster (by assigning implementation and testing to the student).

b) Color Composite Images. Human-Centered Color Mapping deals with assigning colors to various image bands ensuring a standard representation of the human color space and subsequently converting this representation to RedGreenBlue values than can be used to drive a color display. A large number of color spaces have been proposed in the literature of color vision. While some have become standardized, nothing has been done for emerging hyperspectral technologies. The project requires a clear understanding of the color composition. A preliminary project involving deduction of possible color schemes is ongoing. The project was originally suggested by a CS student with interest in graphics and painting. It is currently used in CS1 to stimulate students in understanding the GUI interface offered by Java. In its current format, students design an applet with buttons that allows them to vary the intensity of the fundamental colors and displays the resulting color. Students are able to decrease or increase the amount of each of the colors.

File Processing

Given even a modest resolution (such as 320 x 320 pixels), with 210 spectral images and integer representation of the pixel intensity, a hyperspectral data set requires over 80MB of storage space. It is thus, of importance to analyze efficient ways of storing the data. Several standards exist (Band Interleaved By Line, Band Interleaved By Pixel, Band Sequential) that would be beneficial for one method or another. As new technologies emerge, it will be useful to develop better ways of storage. The project is a good example for processing large data sets. We are also using it in discrete mathematics classes to emphasize on the complexity of algorithms and on the importance for efficient processing.

Distributed Processing

The distributed version of the algorithm partitions the hyperspectral data $S$ into subcubes that are processed separately (see Figure 5). Each subcube consists of a set of pixel vectors. Following this phase, each of the subcubes can be processed relatively independent of each other. This approach allows us to take advantage of the benefits of distributed processing environment. Once
the data are partitioned, a separate process (worker) is started to work on each subset. A master process collects the data and composes the transform. The transform is then communicated back to the individual processes that apply it to the corresponding subcubes yielding the final result. The project is open to undergraduate students interested in learning distributed processing. The topic can be either part of a related course or constitute a stand-alone project. As a research value the project constitutes a significant contribution to the field. From the point of view of remote sensing this is still a relatively new research direction. While most of the work has focused on improved processing methods, little has been done for achieving speedup through distributed versions.

**Real Time Processing and Data Acquisition**

Recently, the amount and the accuracy of the hyperspectral data have increased significantly. In addition to imagery provided by sensors installed on aircraft (HYDICE, AVIRIS) or on satellites (Hyperion), hyperspectral sensors have been made available as off-the-shelf cameras. While the applications provided with the cameras allow for basic image capture and saving, they do not provide any option related to on-the-fly processing or direct interaction of the researcher with the sensor. As such, while the image acquisition capability exists, the benefits of using such an instrument are limited by the rather rigid computer applications that accompany it. Currently, the manufacturer’s provided camera software allows only for static image capture using the sensor. Because of this, experiments testing highly efficient solutions such as distributed processing, subset selection, etc. are conducted only on static data and cannot directly benefit the availability of the sensor. A possible project consists in the development of new computer tools for hyperspectral data acquisition using the off-the-shelf sensors. The application will be written in a high level programming language (C++, or Java) and provide a dynamic interface that can be used in current and future research. By creating the in-house application for data acquisition, practical applications that use hyperspectral imagery can be integrated seamlessly with the sensor, leading to very efficient computing times and even real time. Additionally, this can lead to development of new ways of using the sensor.

Figure 5. Partition of hyperspectral data into subcubes. The illustration shows a partition into four subcubes
The development of such application, involves fairly extensive coding and testing. In addition, it requires training in basic notions of remote sensing, optics, and image processing as well as reasonable programming skills. Talented undergraduate students can meet all these qualifications. The project is currently underway as University funded Student-Faculty Research project. The equipment was purchased as the college’s initial support for new research laboratories. Figure 6 shows the research setting with the hyperspectral camera pointed towards a floral arrangement. The student has designed data extraction components for the software that allow the separation of a band or of a pixel vector (see Fig 7). This required understanding of remote sensing technologies, knowledge in computer graphics and in connecting code written in different programming languages. In this case, the original application was written in C++ while the display methods added are written in Java. The effect on student’s focus has been significant; he has gained an understanding of the hyperspectral imagery by designing various lighting scenarios and investigating the features of the equipment. While the project is winding down, the student has indicated the intention to apply to the Graduate School and to continue work in a similar direction.

Figure 6. Laboratory setting for real time hyperspectral data acquisition.

Figure 7. Hyperspectral imagery application using multiple programming languages. The pixel vector collection (right side) was developed as part of an undergraduate research project.

4. CONCLUSIONS

In the process of developing an interdisciplinary research program, one should always remember the great importance of involving students in research endeavors. Unfortunately, as research projects involve interdisciplinary knowledge, the possibility of student contribution may decrease.
In this paper we have investigated how the need for interdisciplinary work influences Computer Science or Computer Engineering education. While changes in curriculum are often advocated possible, we suggested that the undergraduate student can also benefit from research projects included within regular CS classes. As example, we have shown how a relatively advanced topic, hyperspectral imagery, can generate interesting undergraduate projects. Irrespective of the research area, in devising undergraduate research projects, we should be looking at several aspects: the student background, required knowledge, time needed to complete the project, usefulness of the project both for the research activity as well as the educational value. A balance of these factors ensures a high quality activity and satisfaction of all parties involved.

REFERENCES


BIOGRAPHICAL NOTE:

STEFAN ROBILA received the B.S. in Computer Science in 1997 from University of Iasi, Romania, the M.S. in Computer Science and Ph.D. in Computer Information Science in 2000 and 2002 respectively both from Syracuse University. Currently he is an Assistant Professor in Computer Science and the director of the newly formed Center for Imaging and Optics at Montclair State University. His interests lie within pattern recognition with applications in computer security (steganography) and image processing (in particular multispectral and hyperspectral imagery).