



CALIFORNIA STATE SCIENCE FAIR 2012 PROJECT SUMMARY

Name(s) Colin C. Aitken	Project Number S1401
Project Title On the Theory of Functions and Collatz-Like Conjectures	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals The modern field of mathematics is a widely varied and diverse subject, with subfields ranging from statistics to homological algebra. Due to this variety, different tools have been developed in each area, which means that discoveries in one area rarely give insights into other areas. This project's original goal was to find a structure which could be used to algebraically study a seventy-five-year-old conjecture known as the Collatz Conjecture, but soon expanded to the goal of uniting all of mathematics, which would provide a powerful new tool for making progress on any mathematical problem.</p> <p>Methods/Materials This project focuses on the new idea of a functionally-closed set, which consists of a semigroup of functions (a set of functions which is closed under composition) that map a given set to itself, along with the set itself. The inclusion of the set itself leads to novel notions of when two functionally-closed sets are homomorphic, or have similar structures. The resulting structure, to the best of the author's knowledge, has not been studied previously.</p> <p>Results After fleshing out the basic theory of these structures, a way was found to map almost any mathematical structure onto certain functionally-closed sets, and vice versa. This provides a framework by which one can transform problems in one area of mathematics, such as algorithms, into a completely different area of mathematics, such as field theory. This was applied to transform the Collatz Conjecture, an unsolved problem about integers, into a statement about topology (a field of math related to geometry), which allowed the solution of a class of problems similar to the Collatz Conjecture. This was also extended to produce results in computer science, most notably a new method of generating problems unsolvable by computers as well as an algorithm allowing computers to perform the transformations between areas of math in certain areas.</p> <p>Conclusions/Discussion This research provides a powerful new tool for mathematical work and makes significant progress on the Collatz Conjecture, and has applications throughout the field of mathematics and computer science. One practical application is the project's usefulness in Artificial Intelligence, particularly the in the theory of Automated Theorem Provers, which underlie the safety of NASA craft and other systems.</p>	
Summary Statement This project provides a method of transforming problems from one area of mathematics to others, allowing tools from one area of math to be used to solve problems in other areas.	
Help Received Teacher helped with entry and science fair procedures, Dr. David Roe (PIMS fellow at the University of Calgary) helped review the final result for accuracy (no research was done at the university or by Dr. Roe, he only reviewed the final result)	



CALIFORNIA STATE SCIENCE FAIR 2012 PROJECT SUMMARY

Name(s) Christine Chang; Joseph Chang	Project Number S1402
Project Title A Novel Approach Applying Image Processing Techniques on Retinal Images for Vessel Extraction	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals Develop and implement an accurate image processing algorithm to address challenges in vessel extraction for retinal images in order to help the diagnosis and prevention of retina-related blindness.</p> <p>Methods/Materials Major steps include edge detection, false edge removal, vessel junction restoration, and vessel labeling. (1) Edge detection: Apply Canny Edge Detector to the input image to establish a baseline upon which our algorithm will further improve. (2) False edge removal: Remove misidentified segments by Canny Detector. Our program creates a histogram of frequency vs. pixel-count, which looks like a $y=1/x$ curve whose vertex point is used as the threshold to remove objects with fewer pixels. (3) Vessel junction restoration: Fix broken junctions introduced by Canny Detector. At a broken junction, track and find the direction of a vessel. Extend the vessel in the opposite direction for a certain length. If another vessel is found, bridge the gap and restore the vessel junction. Our experiments produce good results with a length extension up to nine sigma (the standard deviation used in the directive of Gaussian). (4) Vessel labeling: A typical vessel is represented as two parallel edges. This step fills the interior pixels of a vessel. The challenge is to distinguish the area within a vessel and the area between two different vessels that are parallel to each other. Our algorithm traces the opposite direction of the gradient vector (the inward direction) to find the pairing edge of the same vessel.</p> <p>Results We run our algorithm in MATLAB against the twenty test images in the DRIVE database. The average accuracy is 92.78% with a standard deviation of 0.54%. The accuracy rate ranks 3rd among six published methods, including supervised methods (i.e. methods that require training). Excluding the supervised methods, our algorithm's accuracy ranks second. In addition, our algorithm runs efficiently and takes 8 to 10 seconds per image on a laptop with an Intel Core i5 at 2.3 GHz.</p> <p>Conclusions/Discussion This project develops an effective junction restoration algorithm and a novel vessel labeling method. Combined with Canny Edge Detector, our algorithm produces a high average accuracy rate of 92.78%, and the accuracy rate is very consistent from image to image (with 0.54% standard deviation). Our algorithm also runs efficiently, taking less than 10 seconds per image on a home computer.</p>	
Summary Statement This project develops an effective junction restoration algorithm and a novel vessel labeling method to accurately and efficiently extract vessels from retinal images to help the diagnosis and prevention of retina-related blindness.	
Help Received Dr. Chang advised on the basic principles behind Canny Edge Detector. Parents purchased MATLAB licenses for students to use.	



**CALIFORNIA STATE SCIENCE FAIR
2012 PROJECT SUMMARY**

Name(s) Karen Chee	Project Number S1403
Project Title The Table of Secrets	
Objectives/Goals To study numbers of the form $X^2 + 5(Y^2)$ and see why they are so strange in comparison to the rest of the numbers of the form $X^2 + D(Y^2)$ where D is a whole number. I am following the path that Fermat took when studying Diophantus's old work, but I am branching off and making my own way. My hypothesis: While all of the $X^2 + Y^2$ tables have numbers that are divisible by prime numbers (and upon dividing, resort into patterns of similar looking multiples), the $X^2 + 5Y^2$ has numbers that appear in the table for unknown reasons. I hypothesize that there is some of outside numbers that corresponds with the $X^2 + 5Y^2$ table.	
Abstract	
Methods/Materials A posterboard, Excel spreadsheets, paper, and glue. Also, lots of pencils and lots more thinking! I created the tables via the Excel computer program. I looked for patterns and colored them in, and pieced together boxes to create a new table.	
Results The multiplicative patterns in the data sets suggest a product law (or really more than one such law) on pairs of integers which has all the usual arithmetic features (commutative, associative, distributive, etc.) But prime factorization of pairs of integers is NOT unique. This has far-reaching consequences, and calls for some modification of the theory of these number pairs that might restore the uniqueness of prime factorization in some sense.	
Conclusions/Discussion Fermat's mysterious remarks point to a need to ensure that the "generalized numbers" arising naturally in number theory should factor uniquely into primes and we exhibit on broadly applicable way to do this.	
Summary Statement I am studying numbers of the form $X^2 + 5Y^2$ and creating a new, previously hidden table from the original's entries.	
Help Received Laurens Gunnarsen helped create the Excel spreadsheets for me because I do not have the Excel program on my computer).	



CALIFORNIA STATE SCIENCE FAIR 2012 PROJECT SUMMARY

Name(s) Richard Cho	Project Number S1404
Project Title Exploring Quadratic Residues and Their Potential Applications	
Abstract	
Objectives/Goals Objectives: 1) Explore the distribution of quadratic residues and find patterns or explore randomness. 2) Number Theory is quite heavily used in modern-day cryptography. Are there possible applications of quadratic residues?	
Methods/Materials Methods: Use the computer to generate and visualize data. Then, with these leads, proceed mathematically until a suitable explanation is reached, proving some relations. (In the parts of the project emphasizing math.) Also, programs can verify explanations and proofs with cast amounts of empirical data. I used a combination of C++ programs and Mathematica notebooks to generate and process data. Materials: Computer, Pen/pencil, Paper, Code::Blocks C++ IDE, Mathematica 8 for Students.	
Results This project was a multifaceted one. Several interesting patterns in the distribution of quadratic residues were explored and proven. As for the application into cyprography, a novel way was found for factoring semi-primes, which can revolutionize the current state of cryptography.	
Conclusions/Discussion Through a long exploration through quadratic residues, we learned that some very interesting things. Interesting paterens about the distribution and number of quadratic residues were proven, usually modulo a prime. Mainly C++ programs were deploed to generate quadratic residues, count the number of distinct elements, and count the number of consecutive pairs and triples. Mathematica was used, with its powerful graphical functions, to help us in discovering pattern on the distribuion of quadratic residues modulo primes. To end, there was a hint into solving one of the biggest unsolved problems in cryptography, how semi-primes can be factored easily, using the number of quadratic residues. This would be ground breaking, as such a discovery would shatter RSA and would revolutionize the Internet and computer driven world we live in today.	
Summary Statement An investigation into quadratic residues, where interesting patterns were found as well as a competely new way for factoring semi-primes.	
Help Received Dr. Ali Gurel answered some of my mathematical questions about quadratic residues.	



**CALIFORNIA STATE SCIENCE FAIR
2012 PROJECT SUMMARY**

Name(s) Cooper G. Frye	Project Number S1405
Project Title A Comparison of Test Scores Using Modeling and Bayesian Statistics	
Abstract Objectives/Goals The objective of this project was to find a good way to compare scores from tests with varying numbers of questions. Methods/Materials A model of test-taking was created where every test-taker had a fixed but unknown chance to answer any of the questions correctly. Bayesian statistics were then used to find the probability distribution of this chance based on a certain test score. The expected value of this chance was then calculated. This number was used to compare among scores. Results The expected value for someone's chance to answer a question correctly was the number of questions right plus one, divided by the number of questions total plus two. Conclusions/Discussion This shows that there is a good way to directly compare scores from tests with different numbers of questions. My objective was completed successfully.	
Summary Statement Comparing different test scores to each other.	
Help Received Mother helped print out and attach papers, Project adviser read initial work and gave advice	



CALIFORNIA STATE SCIENCE FAIR 2012 PROJECT SUMMARY

Name(s) Sean S. Haas	Project Number S1406
Project Title Comparison of File Systems for the Megalithic Kernel	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals The goal of this project is to implement two RAM disk file systems and test their performance under Dreckig OS. This operating system uses the experimental megalithic kernel where all software is within the kernel for efficiency and security. This architecture makes traditional file systems impractical, so I have designed these file systems for speed and efficiency with this architecture. The two file systems, as well as Dreckig OS, were designed and written by me.</p> <p>Methods/Materials Two file systems, BFS and bFAT, were implemented in x86 assembly language inside of the Dreckig OS v0.005.8 kernel. BFS stores needed information about a file in the file itself, and bFAT stores file information in a small file allocation table at the start of the RAM disk. To be able to test how fast each file system could write and read files I wrote a timer that makes use of the 8253 PIT. I then put each file system through a series of tests. I had each system make a 10, 20 and 30 byte file; this tested each file systems writing speed. Both file systems then searched for each file to test for read speed. These tests were repeated multiple times for validity. I controlled the computer system, code base, and operating system so that all tests could be performed in the same environment.</p> <p>Results The BFS file system was implemented in 89 lines of code, while the bFAT system was implemented in 98. It took BFS, on average, 55.3 ticks to write a file, and 4633.3 ticks to read a file. It took bFAT, on average, 72.6 ticks to write file, and 4644.2 to read a file. BFS was able to write a file 17.3 ticks faster than bFAT and read a file 10.9 ticks faster, while being implemented in 9 less lines of code.</p> <p>Conclusions/Discussion The BFS file system was able to operate faster than bFAT, while using less code. Currently the bFS file system is being further implemented and tested under Dreckig OS. A computer running Dreckig OS will be displayed at the fair.</p>	
Summary Statement Designed, implemented and tested performance of two file systems for use with the megalithic kernel architecture.	
Help Received	



**CALIFORNIA STATE SCIENCE FAIR
2012 PROJECT SUMMARY**

Name(s) Andrew C. Haden	Project Number S1407
Project Title The Effect of Parallel Processing on MD5 Brute Force Efficiency	
Abstract Objectives/Goals My project aimed to find what level of parallel processing, quantified in threads, results in the best performance of an MD5 brute force algorithm determined by permutations tested per second and CPU load. Methods/Materials In my project, I created a program in the C# language that executes a brute force algorithm with provided numbers of threads. I ran this program on a Windows-based computer with a four-core processor. Results The permutations tested per second and CPU load (%) increased from 54,958 and 14% at one thread to at 141,687 and 53.6% at four threads, the peak. A significant performance decline was observed with only 34,935 permutations/second and 17.6% CPU load at eight threads and 25,588 permutations/second and 15.4% CPU load at 16 threads. Conclusions/Discussion The performance of the MD5 brute force peaked at four threads with 141,687 permutations tested per second and 53.6% CPU load. The performance was negatively affected when the algorithm was run with 8 and 16 threads. From this, I concluded that the performance of an MD5 brute force algorithm is greatest when parallelized with the same number of threads as the host computer has CPU cores. I assume this also applies to other computationally-heavy tasks that are parallelized, however that would have to be confirmed in a separate study.	
Summary Statement My project found what level of parallel processing, in threads, results in the best performance of an MD5 brute force algorithm.	
Help Received None	



CALIFORNIA STATE SCIENCE FAIR 2012 PROJECT SUMMARY

Name(s) William Hang; Frank Su	Project Number S1408
Project Title Traffic Light Detection and Tracking in Video for the Prevention of Automobile Accidents	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals This project aims to develop an algorithm to recognize and track traffic lights in order to warn distracted drivers of their presence. This algorithm hopes to achieve a 94% recall rate and 92% precision rate for red traffic lights and a 96% recall rate and 94% precision rate for green traffic lights.</p> <p>Methods/Materials The algorithm was written in C and used libraries from OpenCV. This algorithm first detected traffic lights within video frames by performing connected component analysis to isolate distinct regions within an image. These regions were then filtered by three criteria: size, aspect ratio, and circularity. The resulting regions are then tracked by opening a search area around their coordinates within the next frame to look for a region of similar characteristics. These regions were found with connected component analysis. The algorithm then draws a colored box over detected traffic lights and issues a beep to alert the distracted driver. This process is iterated for the frames within each video. This algorithm was run on a bank of test videos. After noting deficiencies, methods within the algorithm, thresholds, and other values were changed.</p> <p>Results The algorithm successfully detected and tracked 71 out of 76 total traffic lights in a total of 44 test videos. It incorrectly detected three objects as traffic lights and was unable to detect five traffic lights. The algorithm achieved a precision rate of 93.10% and a recall rate of 93.10% for red lights, and a precision rate of 97.78% and a recall rate of 93.62% for green lights.</p> <p>Conclusions/Discussion The developed algorithm achieved a very reasonable recall and precision rate. However, it did not achieve the recall goal because the algorithm was too strict and excluded traffic lights if conditions were less than perfect. The next steps from now would be to improve upon the algorithm itself by using different methods of detection with shape detectors and machine learning. Tracking can also be improved with LK-Optical Flow. We also hope to develop the algorithm to work in various driving conditions such as overcast, rainy, and nighttime. Most importantly, this algorithm will later be developed to work in real-time.</p>	
Summary Statement To develop an algorithm that recognizes and tracks traffic lights in daylight in order to warn inattentive drivers of their presence.	
Help Received Dr. Serge Belongie and Dr. Brendan Morris provided valuable technical feedback and support; Mrs. Elaine Gillum and Mrs. Erin Schumacher provided valuable feedback on notebook and presentation.	



**CALIFORNIA STATE SCIENCE FAIR
2012 PROJECT SUMMARY**

Name(s) Yosira Hernandez	Project Number S1409
Project Title Inappropriate Use of the Lotka-Volterra Predator-Prey Equation	
Abstract Objectives/Goals The purpose of this study was to demonstrate that the Lotka-Volterra differential equations may not be the appropriate equations to describe certain predator-prey models, which indeed may not be a true cyclical predator-prey relationship, but rather competition models. Methods/Materials In my examination of the Lotka-Volterra equation, I duplicated the data used in Gause's 1934 competition experiments between <i>Paramecium caudatum</i> and <i>Paramecium aurelia</i> , and his classic yeast competition experiments. I also used student data replicating Park's 1954 competition experiments with two <i>Tribolium</i> species. Data was graphed and examined for any sort of a classic cyclical predator-prey interaction. Both Gause and Park had used the Lotka-Volterra equation to describe a predator/prey relationship. Results I found the data sets and graphs do not indicate a classic cyclical predator-prey model and that the parameters for interaction and carrying capacities required by the Lotka-Volterra equation, do not fit into the equation. Conclusions/Discussion I concluded that the Lotka-Volterra equation may not be the appropriate equation for the competition math models in cases where the competition is not cyclical, but rather demonstrative of some sort of kill mechanism. Both Gause (in 1934) and Park (1954) used the Lotka-Volterra cyclical predator-prey equation to describe their work, when at the same time my duplicated experimental data and their data describe a simple competition kill mechanism that does not fit the parameters of the equation. The equations cannot be set up, nor solved.	
Summary Statement The purpose of this study was to demonstrate that the Lotka-Volterra differential equations may not be the appropriate equations to describe certain predator-prey models, which indeed may not be a true cyclical predator-prey relationship, b	
Help Received	



**CALIFORNIA STATE SCIENCE FAIR
2012 PROJECT SUMMARY**

Name(s) Mark Holmstrom; Theresa McLaughlin	Project Number S1410
Project Title Neighbors with Prescribed Prime Factors	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals To formulate a new method that is more efficient than that of Lehmer and still creates a relatively thorough solution set.</p> <p>Methods/Materials An initial set is augmented with new solutions found by a simple procedure. We then repeat this procedure to further augment the set until no new solutions arise. The process ends when no more solutions are found through combination of any two numbers in the set.</p> <p>Results We performed our process with a maximal prime factor of 163 where as Lehmer's method only reached a prime factor of 41. We found 115,207 solutions to our equation. The largest solution we found was 19,316,158,377,073,923,834,000, though it may be possible to find a larger solution if we used a higher maximal prime.</p> <p>Conclusions/Discussion Our new method creates solutions to Lehmer's equation that he was unable to find through his exhaustive methods. Work on the ABC Conjecture relates back to many other conjectures and theorems, such as Fermat's Last Theorem, Roth's Theorem, and the Mordell Conjecture. These findings are specific to set of numbers that fit the limitations of the ABC Conjecture and Lehmer's equation.</p>	
Summary Statement In our project, we developed a more efficient method to solve an equation in comparison to a possible solution proposed by D. H. Lehmer in 1964.	
Help Received Dr. Conrey of the American Institute of Mathematics helped us in the technical side of our project and allowed us to use his house as a research base.	



CALIFORNIA STATE SCIENCE FAIR 2012 PROJECT SUMMARY

Name(s) Andrew S. Hong	Project Number S1411
Project Title The Art of Digital Steganography: A New Algorithm That Can Process Any Amount of Information in Various Formats Intact	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals The purpose of the project was to develop an algorithm that can conceal data of a variety of uncompressed files of different formats and recover the hidden data without loss of information.</p> <p>Methods/Materials The algorithm was implemented in Mathematica, a programming language. All materials used to test the algorithm belong to the student himself. There were four different cases to test the algorithm, and each case was tested numerous times to show that no information in the hidden and cover files was altered or lost. The first task was to conceal an image file inside another image file by using different numbers of LSBs (least significant bits) for concealing data to show the resulting image. The second task was to conceal a music file inside an image file. The third task was to conceal an image file inside a video file by splitting the data of the image, and the last task was to conceal a text file with over 40,000 words inside a video file.</p> <p>Results In all test cases, the implemented algorithm was successful in concealing and recovering vast amounts of information in various uncompressed digital formats without loss of information. Using fewer numbers of LSBs to conceal information resulted in better image quality, and there were no visible changes in the cover files when 1 bit was used to conceal information.</p> <p>Conclusions/Discussion The algorithm shows that there is a viable way of hiding a large amount of information without triggering suspicion and recovering information without any loss. The results also show that the algorithm works for files in a variety of uncompressed formats, including sound, image and text. Using one bit to hide information neither alters the hidden information nor creates noticeable differences in the cover file. Most importantly, if given multiple cover files, data to be hidden no longer has to be smaller than the data of the cover file because hidden information can be spread across multiple cover files. This algorithm can be extended to a wide variety of applications, such as digital security and intelligence. Digital steganography can be used with encryption for watermarks and other security uses. Finally, reverse steganography can be used to crack the code and reveal information in the area of counter-terrorism.</p>	
Summary Statement To improve current methods of digital steganography, I developed encoder and decoder algorithm to conceal data of a variety of uncompressed files of different formats.	
Help Received Dr. James Choi guided me in Mathematica programming language and digital image processing. My parents constantly gave me support and encouragement.	



**CALIFORNIA STATE SCIENCE FAIR
2012 PROJECT SUMMARY**

Name(s) Davis Huang	Project Number S1412
Project Title Fixing the Graymodel	
Abstract Objectives/Goals A population dynamics model of gray whales by a biologist John Brandon does not "work" for unknown reasons. My goal was to discover why and attempt to fix this "graymodel." Methods/Materials The populations model is online at graymodel.stanford.edu . I used Processing, a sub language of Java, to model some of Brandon's equations and Microsoft Excel for graphs. I analyzed Brandon's equations and data that he used to find out which equations or what data could be modified, kept, thrown out, etc. to obtain a better model. Mainly, my analysis focused on the kill count of gray whales. Results There was a major flaw in the model, a discrepancy between the numbers modeled and the data given. From there, I took certain equations and data and "fixed" and re-graphed the data, which should now more accurately model the kill count of the whales. Conclusions/Discussion There is now new data for the model; however, there is no way to measure the accuracy or precision of these obtained numbers from my program. In addition, many of Brandon's equations were confusing and difficult to understand. There were several attempts to contact John Brandon through email and personal visits, but no reply was given. Without contact from the original programmer, progress has been slow, starting from ground zero. However, I do believe that the data is more accurate and could potentially help in the process of fixing this population model. The results will be sent to John Brandon and hopefully future contact will be made to continue this project.	
Summary Statement My project focuses on analyzing a pre-existing gray whale population dynamics model by biologist John Brandon to find the error in the model and to fix it.	
Help Received Diana Herrington provided background information on the model; Dr. Ke Wu helped explain some of the mathematics; Mother helped glue the display	



**CALIFORNIA STATE SCIENCE FAIR
2012 PROJECT SUMMARY**

Name(s) Abraham P. Karplus	Project Number S1413
---	---------------------------------------

Project Title Machine Learning Algorithms for Cancer Diagnosis

<p style="text-align: center;">Abstract</p> <p>Objectives/Goals Machine learning algorithms can be used for cancer diagnosis, but which are best? This project compares four algorithms: Decision Tree, Majority, Nearest Neighbors, and Best Z-Score (my own design; a slight variant of the Naive Bayes algorithm) at diagnosing cancer type in two datasets: breast cancer and colorectal cancer.</p> <p>Methods/Materials Both datasets were gene expression levels from tumor cells. For breast cancer, the algorithms predict basal or luminal (breast cancer types). For colorectal cancer, they predict the p53 mutation. A cross-fold validation program split the data into training and testing sets: the algorithms trained on 80% of the samples and were tested on the rest for performance and time. This train-test split was done 150 times for each algorithm on each dataset.</p> <p>Results For the breast cancer dataset, the Best Z-Score algorithm did best. All three levels of Decision Tree were adequate but slow. Majority was fast but did terribly. Nearest Neighbors was perfect using few neighbors, but as bad as Majority when using many neighbors. For the colorectal dataset, the Best Z-Score algorithm again performed best. All three Decision Trees performed comparably and only slightly worse than Best Z-Score, but took 50-140x longer. Majority was again terrible. Nearest Neighbors was 4x slower than Best Z-Score and had a performance about that of Decision Tree (a bit worse than Best Z-Score); its best performance used a medium number of neighbors.</p> <p>Conclusions/Discussion In summary, Best Z-Score did very well on all tests. Nearest Neighbors did extremely well on easy tasks and acceptably on hard ones. My guess that Decision Tree would work best was not vindicated, as it performed reasonably but took over an order of magnitude longer to train than the others. I learned a lot about the field of machine learning, especially as I implemented all of the algorithms myself, and even designed the most effective of them. To continue the project, I will implement the algorithms Random Forest, Support Vector Machine, and Naive Bayes, some of the most popular currently in use. I am especially curious how Naive Bayes would perform, since my current best, Best-Z Score, was based on Naive Bayes. Another expansion is to implement feature selectors and other dimensionality reducers, which could greatly speed up the algorithms (especially Decision Tree) and improve their performance.</p>

<p>Summary Statement</p> <p>How well do several different machine learning algorithms do at diagnosing cancer from gene expression levels?</p>

<p>Help Received</p> <p>Father suggested this project and connected me to Cancer Machine Learning group at UCSC. Cancer Machine Learning group provided cancer data and project suggestions. Mother helped with the writing and time management.</p>



CALIFORNIA STATE SCIENCE FAIR 2012 PROJECT SUMMARY

Name(s) Madhu V. Krishnan; Mike H. Wu	Project Number S1414
Project Title Autonomous Navigation and Robust Object Detection through Edge-based Optical Flow	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals The main goal is to create an algorithm that efficiently differentiates stationary objects from the background using the optical approach and limiting the system to one sensor. Our calculation goals include finding the relative position of objects, and finding their time-to-collisions (TTCs). The algorithm was based off of Horn-Schunck Optical Flow and Sobel Edge-Detection.</p> <p>Methods/Materials The materials needed were intentionally minimized to a video camera and source code developed with a combination of Matlab and Java. The implemented Optical Flow algorithm was adjusted with a convolution approach, comparing the luminosities of a group of pixels in consecutive image frames. The output of this algorithm is a gradient map, with each vector representing the motion of a pixel on the image frame. The objects were detected through an enhanced Sobel Image, which creates a binary image of discrete objects. Subsequently, this binary image and gradient map were overlaid to create distinct #blobs# of vectors, which are then grouped with a bounding-box algorithm. Finally, the overlaid gradient map is used to calculate the TTC for each object using a calculation based off of multidimensional calculus and similar triangles.</p> <p>A further material used in this project includes a robot, designed to test the algorithm, and based off of the LEGO Mindstorms kit. The robot operates under a simple object-avoidance system, turning away from the closest object a specified number of degrees that depends on the TTC.</p> <p>Results Each trial has outputs of binary images, gradient maps, and TTCs that are saved in movie format. In addition, videos of the robot running the object-avoidance system from both its perspective and a spectator's perspective have been recorded. These video trials will be viewable during the fair. Since most of the project is conceptual, the TTCs constitute the only form of quantitative data.</p> <p>Conclusions/Discussion In each trial, the robot was able to accurately, and in real time, detect the presence of an object and turn away to avoid a collision. This algorithm can be ported into a variety of autonomous vehicle applications, including object retrieval systems and mapping systems. The optical approach developed is advantageous for providing real time implementation, but having the ability to provide more information than radar-based systems despite having only one sensor.</p>	
Summary Statement To differentiate ambiguous stationary objects from a background and find their respective time to collisions through one sensor.	
Help Received	



**CALIFORNIA STATE SCIENCE FAIR
2012 PROJECT SUMMARY**

Name(s) David K. Legg	Project Number S1415
Project Title Learning to Walk: Building and Programming a Virtual Robot	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals My objectives are to: 1) create a virtual bipedal robot consisting of two manipulators and a two-jointed connecting bar. 2) write a control program to investigate the viability of such a design in a virtual environment.</p> <p>Methods/Materials A Windows desktop computer with anyKode Marilou Physics Editor (Home Edition) was used to build the virtual robot and the test environment, and Microsoft Visual C# 2010 Express Studio, installed on the same computer, was used to write the control program, including the logical functions and servo commands for the robot. The library Marilou.Mods from anyKode was used to interface between C# and Marilou Exec (the robot simulator).</p> <p>I began with a very basic robot consisting of two boxes joined by a motorized hinge. I continually refined and enhanced my design, building up to my current design.</p> <p>Results I was able to create a virtual robot consistent with my first objective. I also wrote a control program that enabled the robot to move around in its environment per user requests. In addition, I added multiple cameras to give users a robot's-eye view of the environment.</p> <p>Conclusions/Discussion The robot design described above is viable. It has the ability to accurately move around, and to reach out with a manipulator to interact with the environment. As a prototype, it is a promising design with possible applications as a general-purpose working robot, a search-and-rescue tool, and as a spacewalking unit to perform tasks outside a spacecraft.</p>	
Summary Statement This project is the building and programming of a bipedal, virtual robot as a proof of concept to investigate the design's capabilities.	
Help Received Father helped teach me C#	



**CALIFORNIA STATE SCIENCE FAIR
2012 PROJECT SUMMARY**

Name(s) Deborah E. Ma	Project Number S1416
Project Title The Effect on Software Stability by Targeted and Random Function Removal	
Abstract Objectives/Goals My objective was to understand how Linux Version 1.0 Kernel behaved under targeted and random removal, and to see whether Linux Version 1.0 Kernel behaved more like the SF or ER model. Methods/Materials The materials used in this project included a computer, Wolfram Mathematica, Linux SourceCode Version 1.0 Kernel, and Microsoft Excel. For targeted and random removal, the function call graph of Linux Version 1.0 Kernel was loaded onto Wolfram Mathematica. The duplicate function calls were deleted, and the largest SCC was found. For targeted removal, the most connected SCC function was removed until the software system disintegrated. For random removal, a random SCC functions was removed until the software system disintegrated. The results were then graphed on Microsoft Excel. Results Linux Version 1.0 Kernel disintegrated approximately 5.875 times faster with targeted function removal than random function removal. For targeted removal, the software system disintegrated when 0.125 of the largest Strongly Connected Component (SCC) original functions. On average, for random removal, the software system disintegrated when 0.734375 of the original number of SCC functions were removed. For targeted removal, the deviation was zero, and for random removal, the average deviation was 3.98. When compared to targeted and random removal graphs of SF and ER, Linux Version 1.0 Kernel's graph was more like the SF graph. Conclusions/Discussion My results supported my hypothesis. Linux Version 1.0 Kernel disintegrated faster with targeted removal than with random removal, and behaved more like the SF model than the ER model. Linux Version 1.0 Kernel, however, behaved more like an ER than expected. The results show that for a software system to be immune to random attacks (e.g. viruses and bugs), the software system should be modeled after the SF network model.	
Summary Statement Linux Version 1.0 Kernel's software stability disintegrates more with targeted than with random removal, and behaves more like the SF than the ER model.	
Help Received Was given article that inspired the project and was taught how to program in Wolfram Mathematica by mentor, Mr. Gene Ma	



CALIFORNIA STATE SCIENCE FAIR 2012 PROJECT SUMMARY

Name(s) Ajay Mandlekar; Jingyu Oh	Project Number S1417
Project Title A Novel Approach to Sound Steganography through Pseudorandom Sequences and Recombination Techniques	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals To produce a novel device that would allow secure and protected transmission of a secret audio message between a sender and a receiver through the use of a publicly available sound file. While everyone would be able to access the sound file with ease, only the intended recipients would have the ability to access the secret audio message. This device could be used for many important tasks and make secret communication easier than ever before. Some potential users include the CIA/FBI, corporate divisions, or any other sector requiring secure and discrete communication.</p> <p>The publicly available sound file must not be audibly suspicious in any way (must sound like the original music file). The intended recipient must be able to receive a comprehensible secret message from the sound file. The device must be able to work with common sound file formats. Security should be maximized through various means.</p> <p>Methods/Materials The first few stages of development involved actually writing the code for the device. Matlab was used to both write the code and analyze results, and flac files were used for their capability to record at high sampling frequencies. After the program demonstrated success, variables used in the program such as sampling frequency or "ampdec" were optimized through experimentation. After optimal values were found, the device was refined in various ways, such as adding a passcode or correcting for bad user inputs.</p> <p>Results Preliminary tests on the prototype code revealed that although the theory behind the work was solid, additional measures needed to be taken to ensure success of the device. Hence, "ampdec", or an amplitudinal decrease, was implemented into the device. After the device demonstrated success, the code was run with different values for sampling frequency, "ampdec", and "randmax". It was found that the optimal values for the above three variables were 100,000 Hz, 50, and 25, respectively. Afterwards, the code was refined furthermore, with the result of higher security and ease of use.</p> <p>Conclusions/Discussion The goal of this project was to create a device that could encrypt a secret message into a music file and have the ability to decrypt such files as well. The device made accomplished these goals, and the project was a success. All six design criteria were met. This device paves the path for countless future developments and applications in the field of communications.</p>	
Summary Statement The goal of the project is to create a novel device that would allow secure and protected transmission of a secret audio message between a sender and a receiver by encrypting the message in a public music file.	
Help Received Dr. Inhwon Oh provided software and helped with conceptual understanding of signal processing techniques; teachers at school gave advice and feedback from a layman's perspective.	



CALIFORNIA STATE SCIENCE FAIR 2012 PROJECT SUMMARY

Name(s) George D. Morgan	Project Number S1418
Project Title Efficient Operating System Design and Development for the ARM Architecture	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals The objective of my project was to determine if a mobile operating system could be developed to run on a wide variety of ARM based mobile processors while remaining portable to any new core revision, efficient, consuming minimal power, and keeping a short instruction runtime.</p> <p>Methods/Materials ~ ARM based development platforms. Platforms of choice: Atmel's SAM7S128 (ARMv7TDMI) ARM based micro controller and Qualcomm's Scorpion with the Snapdragon S1 chipset with QSD8650 (ARMv7TDMI): (found in many HTC phones). Each processor architecture will be emulated and is not specifically mapped during compiling. This ensures that the code will run with the same efficiency level on any revision of the processor. Each of these are just an example of what platform the mobile operating system can be run on. ~ An ARM emulator. Emulator of choice: QEMU, a powerful open source processor emulator that can be adapted to many applications, and used to simulate the execution of ARM code on many processors. Allows the emulation of the generic ARM infrastructure so we can test run the code before it goes onto a platform itself. ~ An ARM development toolchain (complete with an assembler, compiler, linker, and debugging tools). Toolchain of choice: the GNU ARM Toolchain for Linux. Allows the compiling and linking of ARM assembly and C into a binary executable by the generic ARM infrastructure - the basis of emulation with QEMU.</p> <p>Results The initialization code developed using static instruction set commands, when compiled, is not only fully backwards compatible with all the previous versions of the ARM architecture, but executable on future revisions of the processor with maximum efficiency. This proves that the initialization code and the various components of the OS written in assembly will be versatile enough to be executed on any member of the ARM family, old or new; this confirms that the operating system will remain universal.</p> <p>Conclusions/Discussion By analyzing the variation of different processor core revisions from a couple of ARM family members and by undertaking a series of development procedures that consisted of writing the various parts of a mobile operating system, observing their interaction with each other when linked, and simulating the execution of the OS in its entirety on the basic infrastructure of the chosen architectures, it became clear that it was indeed possible.</p>	
Summary Statement My project is about the development of a basic mobile operating system to accompany the wide variety of ARM processor revisions on the market today while retaining portability and efficiency.	
Help Received None. All of the work done on this project was my own.	



**CALIFORNIA STATE SCIENCE FAIR
2012 PROJECT SUMMARY**

Name(s) Jacqueline L. Prager	Project Number S1419
Project Title The Die is Cast: The Height at Which the Drop of a Die Switches from Deterministic to Random Results	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals The purpose of this experiment was to see at what height a die that was dropped on a firm textbook became random.</p> <p>Methods/Materials A Mindstorms robotic Lego machine was constructed to drop a casino-quality die at heights of one-centimeter intervals starting at zero. At each centimeter, the robot dropped the die a hundred times in an even manner.</p> <p>Results At heights from zero to six centimeters above the ground, the number of times the number one face landed upright was more than any of the other number faces. The sides of the die, faces two through five, became more frequent at heights of six to eight centimeters. By eight centimeters, the bottom of the die (face six) had increased to the frequency of the other faces. Using the chi-square statistical analysis test, the die drop results were found to be random when the die was dropped from at least twelve centimeters above the ground, showing the hypothesis to be correct.</p> <p>Conclusions/Discussion The die became random at higher heights because the farther the die was from the ground, the more potential energy was being converted to kinetic energy when it contacted the ground. The die, therefore, then bounced more. This project used a uniform surface, a textbook, throughout the entire experiment. Future experiments could be performed to see if dropping the die on different surfaces can affect the height at which the die becomes random. The die could be dropped on plastic, wood, metal, carpet, foam, and stone, from the height of twelve centimeters.</p>	
Summary Statement This experiment will investigate the relationship between how high a die will have to be dropped on a textbook until the die produces random results.	
Help Received Step-brother helped build the robot	



**CALIFORNIA STATE SCIENCE FAIR
2012 PROJECT SUMMARY**

Name(s) Bianca Ray Avalani	Project Number S1420
Project Title An Intelligent Driver Assist System Based on Multimodal Sensor Fusion	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals This research is aimed at developing a driver-assistance system based on multimodal sensor fusion to help drivers in navigating their cars safely. It includes elderly drivers and those with disabilities including poor reflex, eyesight and hearing.</p> <p>Methods/Materials I have developed a computer-based system for a car, integrated with an array of digital cameras, omnidirectional microphones, accelerometer and gyroscope.</p> <p>In the first phase of this project, I have focused on the problem of changing lanes and merging on-ramp on a highway. By analyzing video data from these sensors using optical flow and morphological techniques, this system computes motion vectors of objects in the scene. Next, it performs iterative feature extraction on this vector space, and applies a camera handoff algorithm to identify cars in motion. It also uses array-processing techniques for audio source-localization and tracking, and analyzes power spectrum of audio signals using spectrograms based on Short-Time Fourier Transforms to identify surrounding vehicles. Data from all the sensors are combined to give the driver a final Go / No-Go signal for lane change.</p> <p>Results I collected 300+ data-samples under different traffic, lighting and road conditions. My overall system correctly provided alerts in 92.8% of these test-cases.</p> <p>Conclusions/Discussion Unlike previous researchers who used expensive lasers, LIDARs or time-of-flight cameras, I have applied cost-effective audio and video sensors to address this lane change problem, just as humans utilize their eyes and ears while driving. My system could also detect if the driver was distracted due to phone conversation or weaving in and out of lanes while driving.</p>	
Summary Statement This project is about developing an intelligent driver-assist system based on multimodal sensor fusion to help drivers drive safely.	
Help Received My parents drove the car to help me collect data.	



CALIFORNIA STATE SCIENCE FAIR 2012 PROJECT SUMMARY

Name(s) Jeffrey N. Rosenberg	Project Number S1421
Project Title On Mathematical Growth Models: Representing Accurate, Meaningful, and Continuous Change with Dynamic Limiting Factors	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals Growth models have been quintessential in mainly economic and biological applications, but are also practical and could be explored more for medicine, chemistry, industry, and sociology. Those models however, are only applicable to constant or closed systems. The objective is to derive more accurate growth models that account for changes in limiting factors, made manifest by a dynamic carrying capacity, with whatever pattern might be most appropriate.</p> <p>Methods/Materials Basic forms of the logistic and exponential growth models were used as the basis of the growth models. The logistic was then built upon using other mathematical functions; basic forms of constant, step, linear, quadratic, trigonometric, series, sequences, and sigmoidal functions. Arithmetic and calculus based methods were also utilized, such as limits, differentiation, integration, first and second derivative test, etc. Data was also collected from published and public information to ensure real-world practicality of the new growth models produced. The data was compiled using a TI-84 silver edition graphing calculator and a Microsoft excel model.</p> <p>Results The three major applications provided with data points were a model of Japanese population growth (n=140, sigmoidal), cumulative oil future sales (n=56, phase-wise), and Lynx-Hare (predator-prey) population relationship (n=60, sinusoidal). All models were projected against the corresponding set of data points, and the error never exceeded 5%. Other theoretical models were derived to match estimated growth in sales with seasonality, diffusion of innovation with induced and planned obsolescence, tumor growth, symbiotic biological relationships, and more.</p> <p>Conclusions/Discussion Growth models with stagnate carrying capacity are unrealistic and can only show short-term growth. These new models with dynamic limiting factors can measure and predict growth accurately, long-term. Human population continues to expand the resources available and decreases the effects of limiting factors, which require dynamic models. The projected sales of a product based upon seasonality or a diffusion of innovation make these models practical. Application branch into medicine, chemistry, industry, and most likely, nearly everything that changes. Everything is growing and decaying, so a major step in understanding the world is to look first to how things grow and change.</p>	
Summary Statement Over several months, I derived a novel mathematic growth model to better represent growth accounting for dynamic limiting factors over a continuous and differentiable function.	
Help Received None	



CALIFORNIA STATE SCIENCE FAIR 2012 PROJECT SUMMARY

Name(s) Eric Sauer	Project Number S1422
Project Title PDFClearance: Developing Software to Detect Malicious PDFs	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals PDFs are a globally used file format making them a new chosen vector for malware attacks. Because PDFs can run JavaScript, hackers have learned how to exploit these files and hide their attacks deep within the code. Currently there is no anti-virus software that states clearly whether an unopened PDF, containing a new or altered malicious PDF exploit, is malicious. The purpose of this research project was to develop a new type of software called PDFClearance that detects the behavioral intent of an unopened PDF file by analyzing the JavaScript methods within its text.</p> <p>Methods/Materials Over 1100 malicious and benign PDFs were analyzed by comparing their uses of JavaScript and seven malicious uses were determined. PDFClearance looks for these malicious uses of JavaScript and determines whether an unopened PDF file is most likely benign, possibly malicious, or most likely malicious. The new software was tested for its performance with the previously analyzed collection of 1100 PDFs to confirm successful identification of the seven malicious JavaScript uses and a random selection of 1000 PDFs from a not yet analyzed PDF data base to determine its accuracy in correctly identifying malicious PDFs from benign ones.</p> <p>Results It was hypothesized that if malicious uses of JavaScript in PDFs can be determined, then it is possible to create software that can detect whether PDFs are malicious with at least 95% accuracy. From the data obtained PDFClearance was successful in identifying a malicious PDF 96% of the time and a benign PDF 97% of the time, supporting the research hypothesis. When identifying PDFs from the random collection of benign and malicious PDFs, PDFClearance correctly identified PDFs 95% of the time.</p> <p>Conclusions/Discussion PDFClearance was successful in correctly identifying the behavioral intent of an unopened PDF. This new software offers a significant improvement for cyber defense against malicious PDFs and helps protect the information and systems we rely on every day contributing to the cyber security of our nation. Future development will involve incorporating PDFClearance into current PDF-readers and also creating a web browser plug-in to check PDFs as they are downloaded.</p>	
Summary Statement This project analyzes a collection of malicious and benign PDFs and develops new software that can detect the behavioral intent of an unopened possibly malicious PDF.	
Help Received Mentor Dr. Thomas Kroeger from Sandia National Laboratory answered questions I had about cyber defense issues.	



**CALIFORNIA STATE SCIENCE FAIR
2012 PROJECT SUMMARY**

Name(s) Aakash N. Shah	Project Number S1423
Project Title Single and Multi-Modal Biometrics Based Secure Authentication and Personal Verification	
Objectives/Goals As the number of security breaches and fraud are on the rise, the need for highly secure authentication and personal verification technologies are vital. This project studies and summarizes extensive single and multi-modal Biometrics technologies based on physiological and anatomical characteristic unique to an individual, and are confidential and not intrusive. The project also provides the key advantages/disadvantages of each in the industry where they are being applied.	
Abstract Methods/Materials 1. Thoroughly research each type of Biometrics technologies, research how they work and where they are used. 2. Review how these are applied and identify pros/cons of each based on level of security they will provide. 3. Depending on industry, the usage model and required degree of secure authentication, recommended a single or multi-modal Biometrics technology. 4. Suggest best multi-modal combinations and highlight new Biometrics technologies to investigate.	
Results My results explain and prove my hypothesis because it is clear that certain Biometrics systems can't work everywhere due to size, cost and relevance. See research report for details.	
Conclusions/Discussion The numbers identity theft is increasing day by day, which drives the need for highly secure Biometrics. This project summarizes extensive single and multi-modal Biometrics technologies based on physiological and anatomical characteristic unique to an individual, are confidential and not intrusive. I believe that the selection and modality of the Biometrics technology will depend on the cost, feasibility and ease of usage. My results are the following: Fingerprints are used in a variety of places, frequently used in forensics and the FBI database. The FBI database is said to hold the fingerprints over 51 million criminal record subjects and over 1.5 million civil (non-criminal) fingerprint records. Cost of Biometrics is low. Palm prints tend to be used in banks and different police forces, and cost is low to medium. Iris recognition is primarily used in airports, used to facilitate entry into the US for pre-approved, low-risk travelers, and used private companies/businesses, cost is high. Voice recognition is used in many medical research facilities and in different military bases across the world and the cost is medium to high.	
Summary Statement Research Single and Multi-modal Biometrics - Feasibility and Improvements	
Help Received My father guided me for research.	



CALIFORNIA STATE SCIENCE FAIR 2012 PROJECT SUMMARY

Name(s) Yousuf M. Soliman	Project Number S1425
Project Title GoAI: Creating an Artificial Intelligence to Play Go	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals The purpose of this project was to develop an artificial intelligence to play the ancient Chinese board game, Go. The game is played on a nineteen by nineteen tiled board and players alternate placing pieces on the board, a players goal is to capture the opponents pieces by surrounding them with his own colored pieces. Go takes minutes to learn, but years to master. Go is one of the few games in which a computer to date, has not been able to successfully beat a human player without uncertainty.</p> <p>Methods/Materials The only material necessary for this project is a computer running a Linux based operating system. One must first set up the development environment. Once set up, one can begin to program the artificial intelligence. After it is fully written and debugged, it must be tested. The AI will run one million games against 1 Dan professional human players on the KGS Go Server, Fuego GOAI, and GNUGO.</p> <p>Results The artificial intelligence ran on the online KGS GO Server, one million games. Out of these one million games, the 'Go' artificial intelligence won a whopping 893917 games. The win to lose ratio for this artificial intelligence is .893 and an 89.3% win. The artificial intelligence is playing at the 1 Dan level and is winning a majority of the games against the humans it plays. Once the artificial intelligence developed in this project was completed, it was pitted against the other leading artificial intelligence in the game of Go. Against Fuego, one of the most powerful Go artificial intelligence systems available, GoAI was able to win approximately 74% of the time. And against the second most powerful, GNUGo, GoAI was able to win an enormous 93% of the time. As the artificial intelligence is a little bit more powerful than Fuego, and much more powerful than GNUGo, the artificial intelligence is currently in a very good standing.</p> <p>Conclusions/Discussion As the application was able to compete and play against the human player, along with that the application uses artificial intelligence to determine its move, therefore it would be a success. The application is very advanced in the terms of the artificial intelligence. The biggest changes include a much more powerful opening sequence artificial intelligence, as when there are many positions opoen and without any immediate dangers the computer needs to be able to anticipate many moves into the future.</p>	
Summary Statement The purpose of this project, GoAI, is to create an artificial intelligence to play the game of Go.	
Help Received I would like to thank my computer science teacher, Mrs. Najwan, for teaching me the basics of code. Along with that I would like to thank my mother, Wafaa Eldereiny, for helping me with my board and driving around to get me the materials.	



CALIFORNIA STATE SCIENCE FAIR 2012 PROJECT SUMMARY

Name(s) Christian J. Stewart	Project Number S1426
Project Title Low-Cost Autonomous Civilian Unmanned Aerial Systems	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals Unmanned Aerial Systems pose a wide variety of civilian applications. However, the majority of current systems were developed for military applications and are expensive to build, operate, and maintain. The objective of the project was to design and develop a low-cost Autonomous Unmanned Aerial System (AUAS) capable of meeting the needs of civilian applications.</p> <p>Methods/Materials A list of requirements for the UAS was made. A design was created, and a prototype built to follow the design. Commercial costs of each component were recorded, and the total cost of the system calculated. Software was developed to offer user-friendly point-and-click control of the system. An experiment to verify the system's capability to execute a sample application, Aerial Imaging for Survey, was devised. Interest areas were selected by an operator with the custom Mission Control software, and the system was deployed under full autonomous control. Data measuring every aspect of the plane's flight, primarily locational, airspeed, and altitude was electronically recorded. This test procedure was repeated with five differing interest areas. Custom software was developed and used to determine the area imaged by the on-board camera, and was subsequently compared with the initial interest area of each flight. The resulting percent interest area imaged was recorded.</p> <p>Results The system imaged over 100% of the interest area for every test. The system, if manufactured for commercial use, costs \$594.36 per unit, \$24,405 less than a comparable drone currently in use, which costs \$25,000.</p> <p>Conclusions/Discussion It was concluded that the final design was capable of meeting the needs of a common civilian application at a significantly reduced cost to a comparable drone currently in use.</p>	
Summary Statement I designed, developed, and built a prototype Autonomous Unmanned Aerial System capable of fulfilling the needs of a common civilian application, and wrote software to provide simple point and click control of the system.	
Help Received Mother helped procure board elements; Jim Hall donated airborne computer; Father helped with transport; Advise from Damon Russell, Doc. Student at Caltech	



**CALIFORNIA STATE SCIENCE FAIR
2012 PROJECT SUMMARY**

Name(s) Prem M. Talwai	Project Number S1427
Project Title A Geostatistical Solution to the Inverse Problem in Groundwater Modeling Using Stochastic Partial Differential Equations	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals The objective of this project is to develop a new stochastic groundwater flow model by characterizing the hydrogeological variables (including hydraulic head and transmissivity) as stochastic processes. This will enable hydrogeologists to better characterize heterogeneous aquifer properties while illustrating groundwater flow patterns.</p> <p>Methods/Materials A new variable, the logarithm of the inverse transmissivity, was introduced into the deterministic groundwater flow equation, and the resulting equation was transformed into a stochastic partial differential equation through the implementation of small-perturbation methods. The fluctuations of the hydrogeological variables about their expected values were described by cylindrical Wiener processes on infinite-dimensional Banach space. Through the elimination of pure deterministic terms (which could be estimated using numerical finite difference and finite element methods), a second-order linear stochastic partial differential equation relating the fluctuations was obtained. Stochastic convolution methods were then used to establish a definite formula for calculating the fluctuation of the hydraulic head given deterministic information of the transmissivity.</p> <p>Results The new formula allows for the easy quantification of the random nature of the hydraulic head through the simple substitution of known deterministic information. The new methodology was extensively tested and yielded results that very closely characterized the random moments of the simulated datasets.</p> <p>Conclusions/Discussion My new stochastic groundwater flow model allows hydrogeologists to accurately incorporate the diverse physical properties of aquifers while characterizing flow patterns. Hydrogeologists no longer need to make unjustified assumptions regarding the homogeneity of aquifers. They can now truly understand the groundwater flow patterns apparent in heterogeneous porous media. By accurately implementing my stochastic groundwater formula, practitioners can better illustrate the complex process of contaminant transport and groundwater flow occurring in the subsurface.</p>	
Summary Statement I developed a novel stochastic groundwater flow model that allows hydrogeologists to better characterize the heterogeneous physical properties of aquifers while describing groundwater flow patterns.	
Help Received	



CALIFORNIA STATE SCIENCE FAIR 2012 PROJECT SUMMARY

Name(s) Nathan H. Xu	Project Number S1428
Project Title Optimal Trajectory Planning for an Unmanned Helicopter: Bifurcations in the Presence of Obstacles	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals The main objective is to create a computer program that is capable of guiding an unmanned helicopter from point A to point B in minimum time without hitting the obstacles between them.</p> <p>Methods/Materials Different MATLAB programs are developed and validated to enable the critical capability of obstacle avoidance in the nonlinear optimal trajectory planning of an unmanned aerial vehicle (UAV) helicopter, named HeLion. Numerical simulations are performed to explore and discover the characteristics of the bifurcations in the presence of obstacles. The materials and software used in my project are as follows: a Window 7 based desktop PC; a UAV helicopter nonlinear dynamic model; a Pseudo-Spectral (PS) optimization package; a p-norm based mathematical model for obstacle avoidance; a student version of MATLAB.</p> <p>Results The results illustrate that the obstacle configuration plays a key role in determining the verifiable optimal trajectory, such as flying over or around obstacles, for the UAV helicopter - HeLion flying between Point A and B in an obstacle-rich environment. An innovative procedure designed and implemented in this project, to iteratively search for a bifurcation point (BP) in the presence of an obstacle is quite accurate. The procedure also significantly reduces the computational requirement to find the additional BPs when one needs to study the impact of various obstacle configurations on bifurcations.</p> <p>Conclusions/Discussion Based on hundreds of simulations, I conclude that the new-created computer program, using the 'inequality state constraints' of a p-norm based mathematical obstacle avoidance model, is an accurate and efficient tool for off-line UAV helicopter optimal trajectory planning with the critical capability of obstacle avoidance. The benefits from this research include but not limited to the following: (1) enabling the critical capability of the 'obstacle avoidance' for HeLion optimal trajectory planning, (2) providing a tool for the validation of the efficiency of future autopilot designs adding samples to the collection of HeLion's optimal trajectories, (3) applying the knowledge obtained from this research to optimal trajectory planning of UAVs other than HeLion.</p>	
Summary Statement What would one do to guide an unmanned helicopter from point A to point B in minimum time without hitting the obstacles between them?	
Help Received Prof. Wei Kang of the Naval Postgraduate School (NPS) provided me with the Pseudo-Spectral (PS) optimal control package and a numerical model for HeLion. My parents provided me with a Desktop PC and bought me a student version of MATLAB.	



**CALIFORNIA STATE SCIENCE FAIR
2012 PROJECT SUMMARY**

Name(s) Samyukta Yagati	Project Number S1429
Project Title Detecting Duplicate Content in Text Documents Using N-Gram Indexing	
Abstract Objectives/Goals The goal was to develop a novel test for accurately identifying duplicated content in a large text corpus and build an efficient program to detect duplicate content based on the test. Methods/Materials The basic algorithm is to break each document in the collection into n-grams (which are consecutive runs of n words), compile an index of these n-grams, and search for clusters of documents that share a large number of n-grams. Then, a second, smaller paragraph-level index is assembled from these clusters to ascertain the proximity of shared n-grams. I experimented with different n-gram sizes, document types, and document sizes to maximize the effectiveness of the program. The program was implemented in the Java language using DrJava IDE and JDK 6.0 on an iMac and a MacBookPro. Results The program successfully identified duplicate paragraphs in large documents (70 to 150 pages in length). It also pinpointed duplicate content in hundreds of news articles from the web and identified duplicate content within a single, large document through the paragraph indexing analysis. Through my experiments, I discovered that an n-gram size of three provides the best balance between storage space and accuracy. Conclusions/Discussion I developed a simple, effective criterion for finding duplicate content in document sets of moderate size, which I implemented into a fast, easy-to-use, accurate stand-alone program that allows the user to check for duplicate content in a group of documents.	
Summary Statement I built a duplicate content detector for large text document corpora using a simple trigram overlap test.	
Help Received My father helped me to learn to use the IDE and find relevant information about Java language constructs on the Internet.	



CALIFORNIA STATE SCIENCE FAIR 2012 PROJECT SUMMARY

Name(s) Andrew H. Zhang	Project Number S1430
Project Title An Automated Process of Calculating Bone Mineral Density Using 3D Image Processing Algorithms	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals An abnormal bone density leads to fractures and/or osteoporosis. Currently, the way to measure patient's bone density is to take a bone mineral density (BMD) measurement. This process requires a patient to take a CT scan, which is analyzed by image-processing software. Currently, the most frequent method for BMD is Quantitative Computed Tomography (QCT), where the doctor draws three rectangles for each vertebral unit, capturing the area of the trabecular bone. From this, the software calculates the mean density. However, this process has several problems. The rectangles do not cover the entire area of the bone and inexact sections will be drawn for each scan because the doctor cannot perfectly determine the correct position. The objective of this project was to write new software that would calculate a more accurate region of the trabecular bone by using 3D image processing algorithms and minimizing human labor.</p> <p>Methods/Materials The materials used were an Macintosh computer, Objective-C language, Osirix, and 158 samples of data. The software was written from scratch using the language Objective-C, as a plugin for the Imaging Software Osirix. The user interface allows the doctor to click points in each vertebral unit. 158 samples (one sample consists of one patient's CT scan) were experimented on by testing both the new software, and the QCT method. The BMD's outputted by both softwares were computed.</p> <p>Results The new software generated BMD's that were on average about 3% greater than the results from the QCT method. The individual BMD's of patients ranged from a difference of -10% to 10% between the result generated by the new software, and the result generated by the QCT software.</p> <p>Conclusions/Discussion From the pictures, the new software gets a better approximation of the area of trabecular bone. The areas which the new software approximated were more accurate than those using the QCT method, which probably resulted in the differing results. By having a more accurate and faster process of measuring BMD, doctors can give better diagnosis and prevent the diseases caused by an abnormal BMD.</p>	
Summary Statement The project is about increasing the accuracy of calculating Bone Mineral Density from CT scans by having the software calculate the areas of trabecular bone rather than having a radiologist draw the area of trabecular bone	
Help Received Used computer and data samples at Los Angeles Biomedical Research Institute under the supervision of Chris Dailing	



**CALIFORNIA STATE SCIENCE FAIR
2012 PROJECT SUMMARY**

Name(s) Kevin Zhang	Project Number S1431
Project Title RAF Sets and the Origin of Life	
Abstract Objectives/Goals To determine the probability that life arises using randomized sets of "molecules", "reactions", and "catalysts". This further relies on the necessity of a RAF set (as defined in the 2010 study by Nan, Andersen and Kjaersgaard) for the arisal of life. Methods/Materials Computer, Java, BlueJ, Excel Results Sets with a larger number of molecules exhibited a smaller probability of containing a RAF set, and sets with larger numbers of reactions exhibited a larger probability of containing a RAF set. Conclusions/Discussion This experiment proved that smaller molecules such as amino acids and RNA are more likely to form from a set of random molecules and reactions than proteins and DNA, respectively. Most importantly, I was able to determine a mathematical basis for approaching the problem of the origin of life. My models allow me to determine the approximate probability that a certain molecule was formed given a basic set of parameters, which include the number of molecules involved, the number of reactions involved, the probability that an individual molecule catalyzes a specific reaction, and whether each reaction takes single or multiple reactants/products. Furthermore, when graphing these probabilities versus the number of reactions, I found that the slopes of each best-fit line (each line was for a given number of molecules and a given catalysis probability) satisfied a curve that almost seemed biological in nature: for small numbers of reactions, the initial growth in the slope of the graph was small, while there was a period of rapid growth near the middle of my graph, from 25-35 reactions. This rapid growth decreased sharply when approaching high numbers of reactions, approaching a limiting value (which, of course, theoretically should be 1).	
Summary Statement My project is to model the origin of life using a computer and a given set of parameters, further allowing me to approximate the probability that a specific molecule arises under these parameters.	
Help Received (none)	