Aerial Video Processing Project

Concept of Operations (CONOPS)

Sponsor

Dorotha Biernesser

Alex Johnson, Capt

George Mason University

SYST ENG 699

April 2013

Michael Gibb

Brian Jones

Mingming Lei

Minh Truong

This page intentionally left blank.

Table of Contents

1. Introduction

In 1965, the cofounder of Intel, Gordon Moore, predicted that the number of transistors on integrated circuits would double approximately every two years. History demonstrated that the doubling time was actually approximately every 18 months. During this same period of time, the space per unit cost of hard drive memory doubled even faster, at every 14 months. The price of one Gigabyte of storage shrank from $300,000 in 1971, to $.10 in 2010. In the 20th century, the pace of Gordon's law has contributed the two necessary ingredients, processor speed and cheap data storage, for what will be the 21st century's primary technological horizon – Big Data.

The pace of development in the sensory industry has reflected the pace of technology in general. With the burgeoning demand for data that these systems can deliver, presents the daunting task of how to handle that data. The struggle to process, analyze, and most importantly glean an enhanced insight from massive amounts of "Big Data" is a commonality between many diverse industries in developed countries around the world. Every day, 2.5 quintillion bytes of data is accumulated from everywhere; climate sensors, posts to social media, digital media, purchase transaction records, cell phone GPS coordinates just to name a few. In fact, 90% of the data that exists today has been created in the last two years alone. This leads us to the conclusion that optimal "Big Data" analytics will be one of the primary obstacles that humanity will try to overcome in the 21st century.

The Air Force is struggling with several manifestations of the "Big Data" problem. One manifestation is the rapid collection of full motion video recordings. The Air Force's current solution is to have UAVs gather video up to 12 hours in length. This video is watched by human analysts at the UAVs home base and checked for areas of interest. If anything of significance is found, the human analyst follows the government protocols to declassify the video segment and then follows procedures to send it up the chain of command who then distributes it to those that need it such as the Red Cross and the Federal Emergency Management Agency (FEMA). The main part of our project's endeavor is to examine a couple ways to bypass the AF's current practice, by designing a versatile video system designed to process up to 12 hours of video feed, automate identification of anomalies at near real-time, and rapidly transfer video down to human analysts on the ground for more precise and thorough analysis of system identified anomalies.

1.1. Problem Statement

The United States Air Force (USAF) currently lacks the capability to quickly process large quantities of video data in support of disaster relief. Unmanned Aerial Vehicles (UAVs) fly over terrain following natural disasters such as hurricanes, earthquakes, tornados, and floods to gather up to 12 hours of video in a single flight. Human analysts must watch this video continuously for the duration of the footage to spot area of interest. The problem that exists is this laborious task of watching 12 or more straight hours of video is known as "death by video" for the analysts. This is not only expensive to have the man hours required, but it also delays the disaster relief efforts.

1.2. Scope

This project will focus on defining a concept of operations (CONOPS) for a software system to process and analyze twelve hours of video for the purpose of disaster relief. The main deliverable for the project will be a CONOPS document that will recommend an appropriate image analysis algorithm for executing the task. The project is only concerned with defining how the software system should work. No recommendations for hardware solutions, network nodes, communication link, or video sensors will be provided under the project scope. The project team plans to deliver a robust CONOPS document to document how the system will function at a high-level. This will allow the team to come up with two feasible alternatives which will go into more details about the flow of data and potential algorithms. The alternatives will be evaluated using a predetermined and sponsor approved criteria model and the best recommendation will be provided.

1.3. Stakeholders

The primary stakeholders will be the Secretary of the Air Force's Acquisition (SAF/AQ) department. Mrs. Dorotha Biernesser is the program monitor element for this department and has expressed the interest for the team to complete the project. Mrs. Biernesser plans to share the analysis with the A2C department which oversees short range planning for the Air Force. A system that automates video processing benefits the Air Force by speeding up the time to process video data and freeing up analysts to work on more pressing work. Both of these parties have a vested interest in the details and satisfactory completion of the project.

Additionally, disaster response and disaster relief groups such as the Red Cross, The Department of Homeland Security, and the Federal Emergency Management agency are stakeholders in the project. These individuals will depend on the information provided by the final system when it is deployed to provide accurate and timely processing of video footage to develop disaster response and relief strategies.

The Systems Engineering and Operations Research (SEOR) department at George Mason University (GMU) is a stakeholder in the project because they perform the role of the umbrella corporation overseeing our small team's project. GMU has a fine tradition of delivering quality engineering products to customers and project sponsors alike. Their stake in the project will be to ensure reliable and intelligent systems engineering practices have been executed on the project to provide a feasible and well thought out solution for the sponsor to uphold the reputation of the department and university.

2. Capability Need

2.1. Business Need(s)

The USAF needs a system or system-of-systems that can speed up the amount of time it takes to process large amounts of video. Up to 12 hours of video needs to be viewed and analyzed to detect changes on the ground following natural disasters such as hurricanes, floods, tornados, and earthquakes. The USAF needs the system to quickly analyze video and alert users of anomalies found on the ground to give the necessary information to first responders and government agencies.

2.2. Business Need Capability Gap

The current capability gap that exists within the USAF today is the lack of an automated system that can analyze video feeds and determine structural damage on the ground. The current capabilities rely solely on human operators to watch video feeds and spot anomalies. When UAV missions come in with up to 12 hours of video, it can be stressful and tedious for the human analyst to watch all 12 hours. This length of work for a single human analyst leads to problems with missing important information in the video. The first problem comes from fatigue. Human are prone to errors when overworked. The second problem is related to human factors engineering. Long video feeds will sometimes have little changes in them and the lack of changes leads to type II errors in humans by missing the changes when they happen. Boredom and fatigue in the human analyst will add up to misses in the data feed that can have dramatic effects on the recovery process from natural disasters. To correct for this gap in capabilities, an automated system is proposed in this CONOPS to aid the human analyst in the analysis process. The automated is not intended to completely replace the human in the loop, but rather to aid and compliment the skill sets of the human analyst.

Not only will the automated system have significant impacts to efficiency improvements in the area of disaster relief, the proposed solutions in this CONOPS can have far reaching effects into other military operations for the USAF or border patrols for the DHS. The use of automated algorithms and processing systems will inevitably lead to improvements in the technology that makes them possible. Human analyst capabilities will remain the same as long as they stay in use. Automated algorithms will help define the future capabilities by representing current capabilities and their use will help guide the requirements for systems of the future in ways that human capabilities could not. This investment in technology and computing power will benefit the USAF by keeping them on the cutting edge of technology in the field of automated image and video processing.

2.3. Current Situation

The Air Force's current solution is to have aerial vehicles gather video up to 12 hours in length. This video is watched by human analysts at the UAVs home base and checked for areas of interest. If anything of significance is found, the human analyst follows the government protocols to declassify the video segment and then follows procedures to send it up the chain of command who then distributes it to those that need it such as the Red Cross and the Federal Emergency Management Agency (FEMA). The main part of our project's endeavor is to examine a couple ways to bypass the AF's current practice, by designing a versatile video system designed to process up to 12 hours of video feed, automate identification of anomalies at near real-time, and rapidly transfer video down to human analysts on the ground for more precise and thorough analysis of system identified anomalies.

3. Operations and Support Description

This section is used to identify and explain the business needs, user groups, organizations, environment, interdependencies and other circumstances in which the solution must operate.

3.1. Missions (Primary/Secondary)

The primary mission the automated algorithm(s) will contribute to is disaster relief. Natural disasters such as tornados, hurricanes, floods, and earthquakes have the potential to destroy large man-made structures and devastate the infrastructure of densely populated areas. The primary mission of the automated solution will be to aid and enhance the capabilities of the human analyst in detecting structural damage on the ground following such a natural disaster. The USAF will gather the information and disseminate it to those that need it most, like first responders (fire departments, police departments) and government agencies such as FEMA and the Red Cross. These organizations will use the information to provide disaster support to those affected and also locate areas of significant damage that may prevent evacuations and rescues.

3.2. Policies, Assumptions and Constraints

3.2.1. Policy Assumptions

National Geospatial-Intelligence Agency (NGA) standards will apply to data transmissions of video and image processing. NGA standards apply to video transmission, image transmission, timestamp allocation, and aerial surveillance standards. The Motion Imagery Standards Board (MISB) is the policy writing body of NGA responsible for writing requirements and industry standards. These standards are meant to serve as a best practices guide for industry to follow for interoperability between government and Department of Defense (DoD) standards.

One such standard is 0301.5, the MISB profile for Aerial Surveillance and Photogrammetry Applications (ASPA). To promote interoperability between file servers and systems, all systems should use the advanced authoring format (AAF). This format was designed for the television industry but has since been carried over for digital production, archiving, and distribution. This format will be important for sharing video feeds between government agencies and passing video feeds between DoD and USAF servers. Material eXchange Format (MXF) is an alternate multimedia file format for the exchange of program material between file servers. Together, these two format methods can help the DoD and USAF community to specify commercial off the shelf (COTS) products for motion imagery processing, exploitation, and distribution functions.

Timestamp allocation will be an important part of vide analysis to determine not only when the video occurred but also where it occurred. By flying the UAVs at predetermined navigational wavepoints, timestamps will help analysts reference the flight path to determine where anomalies were detected by triangulating the UAV's position against the time of the video and its navigational path. MISB standard 0605.3 details the standards for inserting time stamps and metadata in high definition uncompressed video. Uncompressed high definition video formats have the capacity to handle large amounts of data in the Vertical Ancillary Data Space (VANC). By mapping the metadata to the VANC, it is possible to align metadata to a specific frame. The functionality allows for the use of timestamps to be assigned. It should be noted that NGA does not currently have a standard timestamp format but they have noted that it is a future objective for the MISB to provide one.

Certain algorithms will work with a multitude of video formats including electro-optical, white and black scale, and even infrared. MISB standard 0404 defines the standards for the compression of infrared motion imagery. Infrared video is sometimes chosen because of its high contract capabilities and its widely available high definition formats of 720 or 1080p. UAV video feeds will need to follow MISB standard 0404 because of the 14-bit dynamic range pixel interface. The compression algorithms being used to transmit and analyze video are using 14-bit which greatly enhances the retention of range information that is usually lost when utilizing the old 8 or 10 bit systems. The software algorithms and hardware systems will need to specify 14-bit transmission interfaces to ensure optimal quality and interoperability with current DoD and USAF best practices.

Table 1. Infrared Compression Options

IR Compression Options	Compliance Description	Input Color Format	Codec Implementation
Level A	Fully Compliant	14-bit 4:0:0	JPEG2000 or H.264 FRExt Profile IDC Level 244, High 4:4:4
Level B	Partially Compliant	10-bit 4:2:2	Scaled and converted to 10-bit 4:2:2 color and compressed with H.264
Level C	Less Compliant	8-bit 4:2:0	Scaled and converted to 8-bit 4:2:0 color and compressed with H.264
Level D	Minimally Compliant	8-bit 4:2:0	Scaled and converted to 8-bit 4:2:0 color and compressed with MPEG2

3.2.2. Assumptions

This project will operate under several assumptions. These assumptions are designed to provide boundaries and guidance for the concept of operations and the feasibility study. The first assumption is that all videos incoming will be following NGA MISB standards. This means the system will not need to account for any unknown video formats or interfaces. Users can provide additional inputs to the system to provide more information on the video itself while the video is being loaded. The system can receive the 12-hour video in its entirety or simultaneously while it is being captured.

It is assumed that any before-disaster videos taken of an area are of the same format and taken at the same altitude as the post-disaster videos. This means that pre and post imagery should align without additional processing. It is also assumed that obtaining and storing aerial video of any particular geographical area is legally feasible.

We will also assume that humans will be viewing the results of the video processing software and will need to validate the identified video anomalies. This also means that users will not likely rewatch the videos to catch any anomalies that the software has missed. We will also assume the technology exists for the software components and algorithms to function as intended and that the USAF possesses or will possess the necessary hardware to provide the needed computational power.

3.2.3. Constraints

Physical constraints related to processing power will be limited by the USAF's hardware infrastructure. Their investment in hardware, servers, and video algorithm research will determine their operational effectiveness if they choose to implement one or more of our algorithm recommendations. System interfaces and interoperability requirements will be determined by DoD and USAF agencies that must deal with other legacy systems and stick to strategic planning for their service. These constraints will be out of the control of this project team and the stakeholders at the acquisition agency for the USAF.

The sharing of information will be constrained by current DoD practices of UAV video feeds. Currently, all UAV video feeds are classified SECRET upon transmission receipt since they are often used to surveillance battlefields around the world. The sweeping classification means that even video feeds of structural damage in a city will be classified SECRET since it is taken from a UAV using classified hardware. Relaying these video feeds to outside organizations requires them to be formally declassified before allowing them to be viewed by others. The decision to declassify video feeds and the time it takes to complete that task is beyond the control of the project team and the stakeholders in the USAF. Of note, is if mission is flown by DHS asset, then privacy constraints might apply, but classification restraints probably will not apply.

Lastly, physical constraints will be placed on which algorithms will work best by whether or not pre-recorded video was able to be obtained. Many of the algorithms rely on a previous video feed to be on hand for comparison. This allows the images to be separately broken apart and compared to previous images to detect structural changes. Certain natural disasters like earthquakes will not give warning signs like an approaching hurricane. Therefore, previous images may not be readily available to use. The ‘before’ videos will need to be taken and stored well before natural disasters strike to ensure they are available should something happen. If this does not happen it will constrain and prevent the use of some algorithms, which will be discussed in the later sections.

The following are the high-level constraint requirements for our system:

- The system shall accept digital video.
- The system shall be able to accept video feeds containing 30 frames per second resolution.
- The system shall be able to accept video feeds of 2048 by 2048 pixel view (200 meters by 48 meters).
- The system must be able to process up to 12 hours of video.
- The system shall accept electro optical (EO) and infrared (IR) video feeds.
- The system shall be able to detect anomalies in video footage.
- The system shall be capable of alerting the user of anomalies found in video footage.

3.3. Operational Description

3.3.1. Operating Concept (OpCon)

Figure 1 below is a description showing the major, interactive participants, systems, subsystems and their interrelationships of the operating concept for aerial video data processing. Aerial video of various geographical terrains are captured by an air vehicle. These terrains may include areas affected by a flood, hurricane, earthquake, tornado, fire, and so on, while the air vehicle may be a fixed-wing aircraft, helicopter, UAV, etc. The recorded video is then uplinked to a communication satellite so that the video could be relayed to a disaster management center. Depending on the hardware capabilities, the video data may be simultaneously transferred during capturing, directly transmitted to a ground station via a line of sight link, or may be physically transferred from the plane to the center in a data storage medium. After the video data have reached the center, it is then automatically processed by at least one of the algorithms outlined in this CONOPS to determine if any areas-of-interest, or damaged geographical areas, exist in the video footage. The result of the video analysis will be outputted to a human analyst for a final review. The analyst then determines the appropriate action needed to take in response to any geographical damages detected.

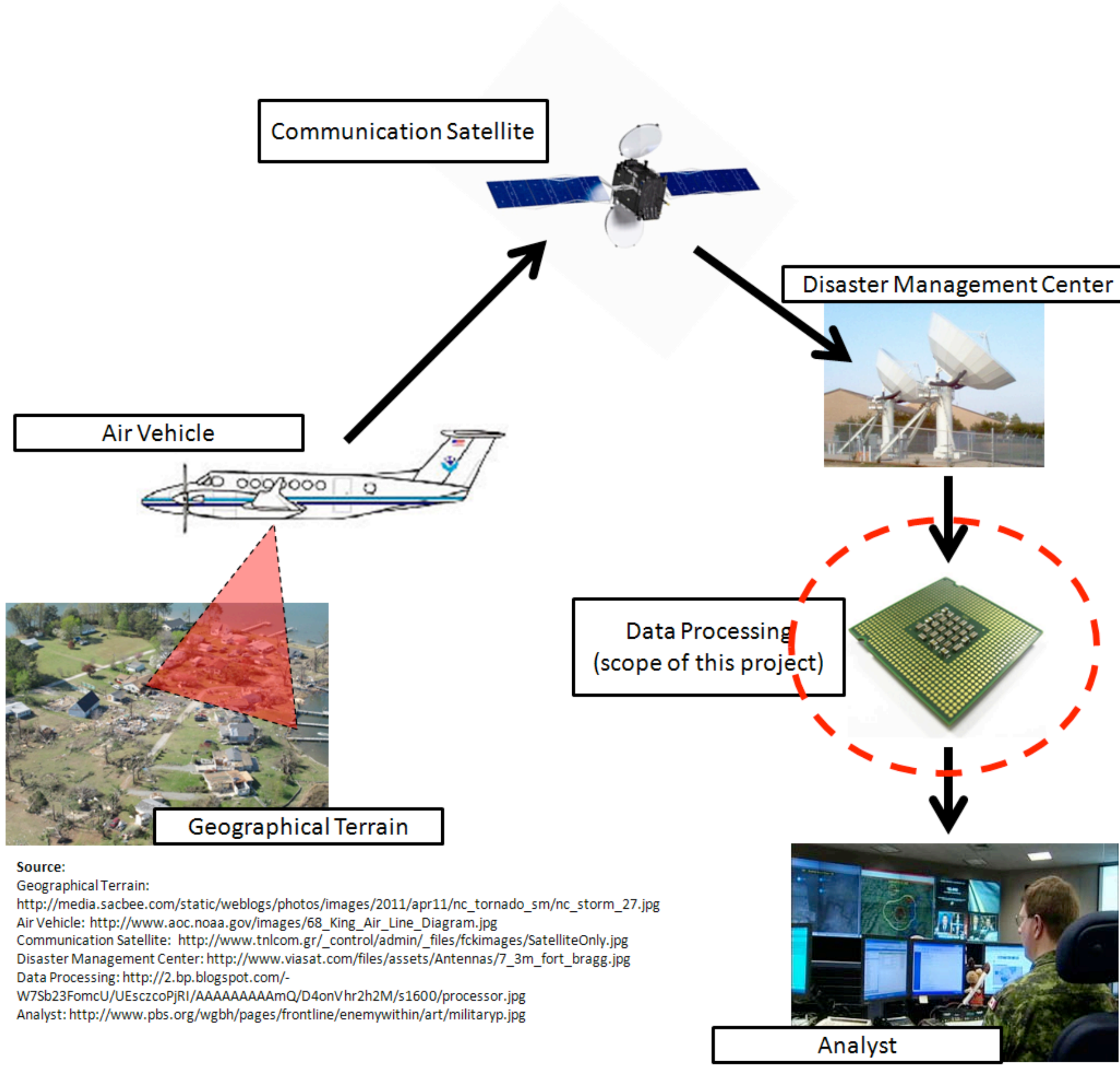

Figure 1. Concept of Damage Management System

The dashed circle in the figure above defines the system boundary of this project. That is, the system receives raw video footage, process and analyze the video to determine if there are any damages shown, then outputs the results to the analyst.

3.3.2. Environmental Conditions

The environmental conditions of the video processing steps will take place in a secure indoor laboratory. The software programs that need heavy computational power will require vast computing power, servers, and most likely have hardware requirements that detail environment temperature and humidity. Those specific environmental requirements related to physical hardware aspects of this project have been scoped out by the team and sponsor. As part of the video or image collection process, it should be noted that clear weather conditions should will provided optimal results of video images. Clouds and rain will obstruct the view of cameras unless infrared video is being used.

3.3.3. Interoperability with Other Elements

The proposed algorithms will be integrated into the current hardware solutions that the USAF currently possesses. Algorithm development can be completed in house or can be readily bought in a COTS product depending on the nature of the algorithm. Some code is also available for free online. A receiver system will be integrated with the necessary hardware and software to implement the algorithm. Some algorithms require significantly more processing power than others to run efficiently in real time; this is something the receiver system may not have. It should be noted that hardware design is beyond the scope of this project. The video feeds will always follow NGA MISB standards to ensure interoperability with other systems and to ensure interfaces will function properly when transmitting video feeds to and from servers and systems.

3.4. Product Support Description

The support for the proposed solution will be minimal in nature because it is a standalone software solution designed to aid the human operator in the task of analyzing video. Hardware and/or software engineers may be needed to perform periodic maintenance to the system or to apply updates to the system as requirements changes, security patches get issued, or parts need replacing. The human analyst will handle the other support missions of operating the system and using it to analyze video feeds. Operational support in the field should be minimal once the system is up and running.

3.5. Potential Impacts

Potential impacts can occur through all phases of the operational spectrum and can range from minor inconveniences to major delays in processing time and mission failure. Minor inconveniences may include improper algorithm set up and configuration. For example, the Canny Edge detection algorithm explained in section 3.6.2 requires preset parameters to properly blur an image and create edges for image comparison. It takes trial and error to properly configure these parameters for a specified image. The parameters will need to be set according to the resolution of the video feed and the desired accuracy of the algorithm which will affect processing time. This may take some time to properly configure and if done improperly, will require new parameter values to be assigned and the video feed to be re-analyzed.

NGA MISB standards may require some interfaces to be changed and upgraded to comply with the latest standards. These standards will help ensure system interoperability between USAF and other government agencies and organizations. New procedures for executing video analysis will need to be defined by the USAF to set up standard operating procedures and training will need to take place to train operators on how to effectively use the new algorithms to aid in their work. It is envisioned that the USAF will keep their current requirements for data retention and dissemination but those may be required to change if MISB standards are updated to reflect industry best practices.

A major potential impact will occur if new hardware configurations are required to optimally run the proposed algorithms. Current infrastructure limitations and budgetary concerns may prohibit the necessary hardware systems from being provided which will then render the algorithms useless. This poses a major risk to the solutions implementation and realization.

3.6. Consideration of Alternatives

Figure 2 and Figure 3 below show the Use Cases for detecting areas-of-interest using an IR sensor and an EO sensor, respectively. The IR sensor is ideal for detecting a thermal bloom such as a heat signature from a burning building, a person stranded in the middle of the ocean, or even a malfunctioning nuclear power plant. It is a possibility that the system can be configured to only identify thermal characteristics within a particular range of temperature to be able to distinguish between multiple elements emitting heat.

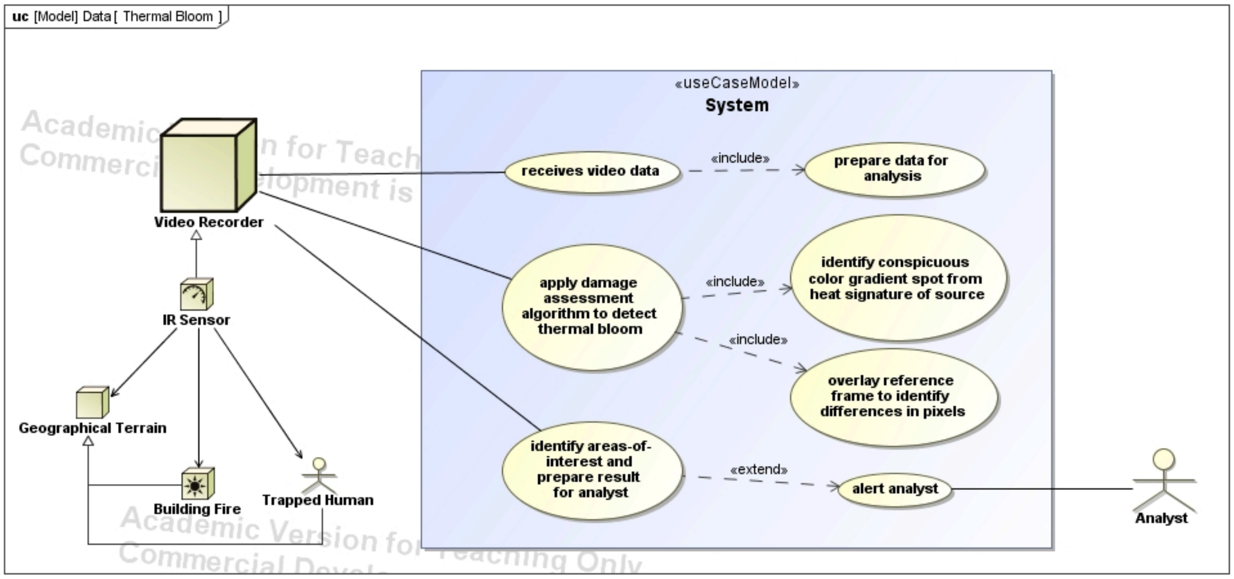

Figure 2. IR Use Case

The EO sensor is ideal for detecting structural damages such as a destroyed road, collapsed building, or a collapsed bridge.

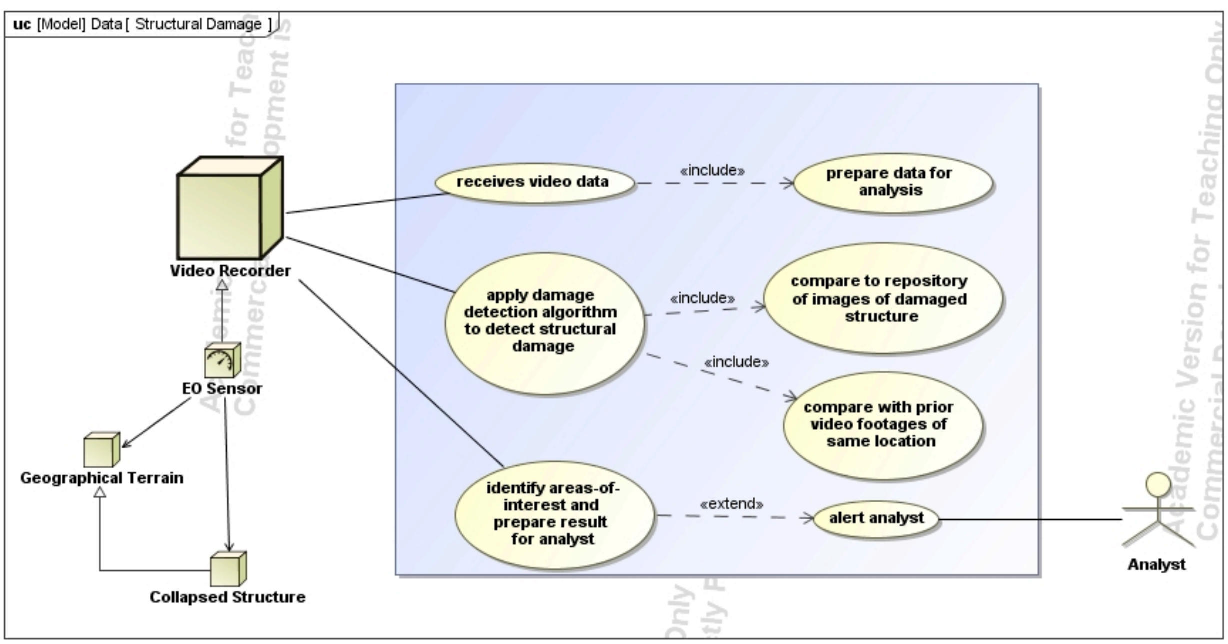

Figure 3. EO Use Case

3.6.1. Genetic Algorithm

Genetic algorithms can be used to identify objects in imagery. Genetic algorithms are based off genetic principals such as selection, crossover, and mutation and are used to find the solution that is "most fit." By defining general guidelines on what an object should look like and how it relates to other objects, it's possible to identify objects in an image. There are three key steps to making this possible. The first step is to define an object ontology. The second step is to process the images into distinct regions of objects to be identified. And the final step is to run the regions of objects through a genetic algorithm to recognize what that object most likely is.

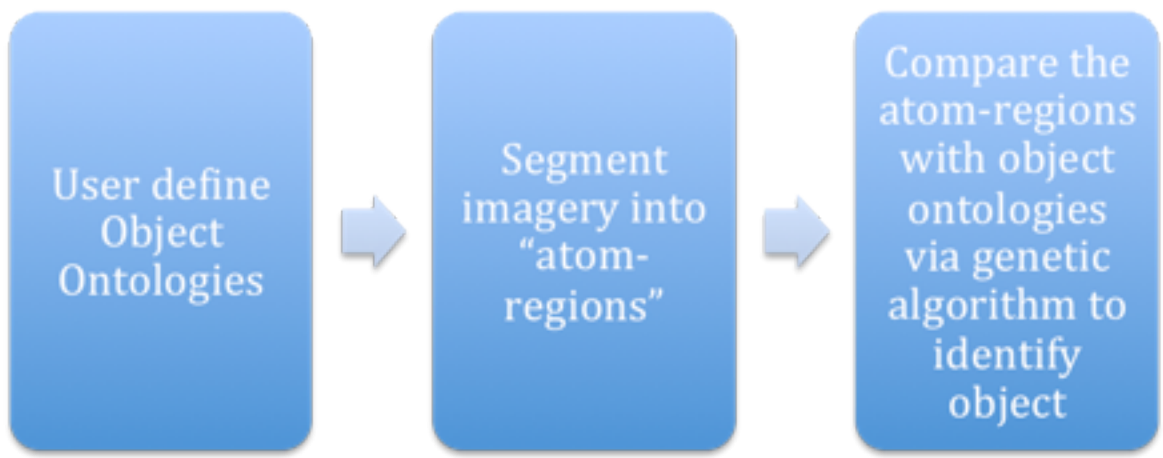

Figure 4. Object Recognition through Genetic Algorithm

Object ontologies define the characteristics of an object and how that type of object can relate to other objects. This allows an algorithm, such as the genetic algorithm, to know what characteristics it should be looking for in order to classify something as a given object. The object ontology defines the characteristics of the object itself. This includes the color, the shape, and the motion of the object. The ontology also defines how the object of interest relates to other objects per the imagery. For example, an object ontology for a plain wooden chair would characterize it as brown, four legs, and motionless. It would also likely be sitting on the ground with walls around it.

In order to identify objects in an image, images can be segmented into what are called atom-regions. Atom regions are generated from color and motion segmentation algorithms that block out regions of an image based on these colors and motions. Each of these atom-regions are essentially segments of the image that can be a unique object and need to then be processed and identified.

The genetic algorithm utilizes the object ontologies that have already been defined to identify what objects the processed atom-regions or segments of the image are. It does this by running each atom-region through a fitness function composed of two parts: one that measures the closeness of the atom-region with the object-only characteristics with the object ontology; another that calculates a relationship distance between two atom-regions or segments of the image to identify how well the region fits with the relationship portion of the object-ontology. The atom-region can then be identified as the object that generates the highest values for the fitness equation. Should there be no object that the atom-region best fits, it can be classified as an unknown object. If adjacent segments of an image are classified as the same object, the segments are merged into one atom-region.

Utilizing the genetic algorithm, there are two approaches to take for identifying objects of interest. The first is to define object ontology to identify objects of interest and notify user when these objects are identified. This would mean defining what damaged infrastructure or thermal blooms would appear as in imagery. The system would then be required to notify users when any of these objects of interest are identified from the imagery. The alternative approach is to define and object ontology based on "normal" objects in aerial videos (define objects as roads, bridges, etc) and notify user when an unknown object is processed. A third alternative would be combine the two approaches. Users can define object ontologies for objects of interest and normal objects and have the system notify them if any objects of interest or unknown objects are identified.

The strongest point of using a genetic algorithm to help identify video anomalies is its independence from prior image. Because users define the objects through ontologies, the algorithm can identify any object that fits that defined ontology. The genetic algorithm approach also presents a flexibility in defining objects of interest through the ontologies so notifications can provide more information to users. However, extensive work is required to properly and fully identify any and all objects that the system should be able to recognize. Additional work may be needed to refine the ontologies depending on the landscape of the area; for example, ground would be a different color in a desert environment compared to a more forested area. Another shortfall of the system is that its accuracy to identify objects is unknown for black and white imagery given that a portion of the fitness equation is based off color matching. Without color, hue and saturation information is lost and the algorithm's accuracy may be reduced. Below is an example of how this image processing technique could work; the left column shows the before-image; the center column shows the image after it's been segmented into its atom-regions; and the right column shows how the atom-regions have been identified based on the object ontologies defined.

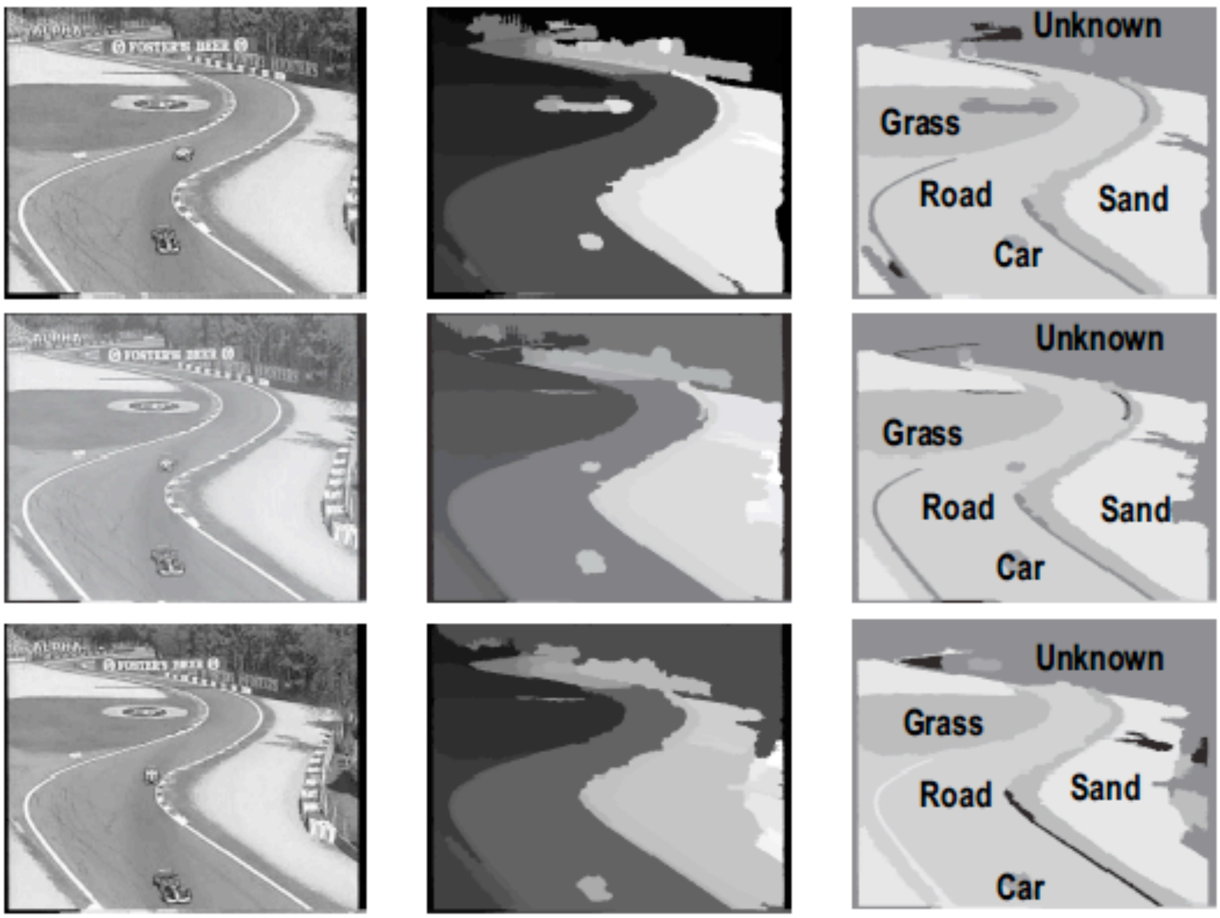

Figure 5. Genetic Algorithm Object Identification Example

3.6.2. Canny Edge Detection Algorithm

John Canny developed one of the best, if not the best, edge detection algorithms in 1986. He performed his work during the early days of computer vision technology but the Canny edge detector is still a state-of-the-art detector. John Canny considered the mathematical problem of an optimal smoothing filter given the criteria of detection, localization, and minimizing responses to a single edge. He showed that the filter given these assumptions is a sum of four exponential terms. He also showed that this filter can be well approximated by first order derivatives of Gaussians. Because the Canny algorithm is susceptible to noise present in the raw image, it uses a filter based on a Gaussian (bell curve) , where the raw image is convolved with a Gaussian filter. It is important to note that the Canny Algorithm is only good for providing the edges of an image. A COTS product for image analysis will still need to be used to compared previous Canny images with current Canny images to determine if damage is present in the structure. Of course this limitation implies that a small crack in the foundation of a building can be missed because there is no significant damage to the edge of the structure.

Different parameters will affect the computation time and effectiveness of the Canny algorithm. The first is the size of the Gaussian filter. This smoothing filter used in the first stage directly affects the results of the algorithm. Small filters cause less blurring which allow detection of small sharp lines. The large filter causes more blurring and smears the value of a pixel across a bigger distance in the image. This makes it better at detecting larger smoother edges like a bridge or building. Since the USAF will be processed large scale aerial video, a large filter would be better but trial and error will be needed to get the parameters just right for the image resolution and the extent of the video feed which is predetermined by how high the UAV flies above a given target.

Threshold values for hysteresis pose a major problem to the application. A threshold set too high can miss important information while a low threshold value can identify irrelevant information. It is highly difficult to identify a generic threshold that works for all images and no tried and tested solution exists at this point. Trial and error will also be needed to get this value just right for the USAF's intended application.

This algorithm works best with any type of video feed but must use a previously loaded reference video. A UAV would need to fly a preprogrammed path along known GPS points and angle its sensor in predetermined positions. A second flight path over the same area would allow the video to be broken down into still images. The algorithm would then take a still frame image and blur the image and apply the Gaussian filter. That image would be overlaid with the same image from the previous run. The differences in the images would be shown on the image and significant differences will be relayed to the user video audio or visual alerts. Areas of significant change can be highlighted certain colors and use colored boxes to alert the user to the position on the screen. Audio alerts such as bells and whistles can be programmed into the algorithm in case the program is running in the background and the human analyst is not constantly monitoring the screen.

The Canny algorithm works best when Graphics Processing Units (GPU) are utilized. While hardware design is out of the scope of this project, it is mentioned here as a caveat to enabling the successful implementation of the Canny algorithm. The use of a GPU allows the Canny algorithm sufficient processing power to run in near-real time. Multiple digital signal processors (DSPs) are recommended because of the multiple number of processes needed to be performed on the image. A long video stream also complicates the processing time because of the constant stream of incoming images. For this reason, the Canny Algorithm can be designed to run on every frame by the image comparison tool that compares current images with previous images can be designed to run at different intervals. A suggested interval may be every ½ second or every 1 second depending on the speed of the flight and how long a given structure stays in frame during an aerial flight. This image comparison interval can help ease the burden on processing power and speed up the processing time.

The main stages of the Canny Algorithm are: noise reduction by filtering using a Gaussian blurring filter, determining the gradient of an image, relating the edge gradients to directions that can be traced, tracing valid edges, and hysteresis thresholding to eliminate breakups of edge contours. The image below shows the processing steps needed to transform an image from its input stage to the final Canny Edge Detection output. Notice the two images of the woman in lower left hand corned of the figure. The image on the left is before the Gaussian blurring and the image on the right is the after image. The edges are blurred so the computer algorithm can detect only sharp changes in edge detection. Without Gaussian blurring, every edge would seem very prominent and when determining changes in each image, false positives would become a problem as the software thinks every major edge has changed significantly.

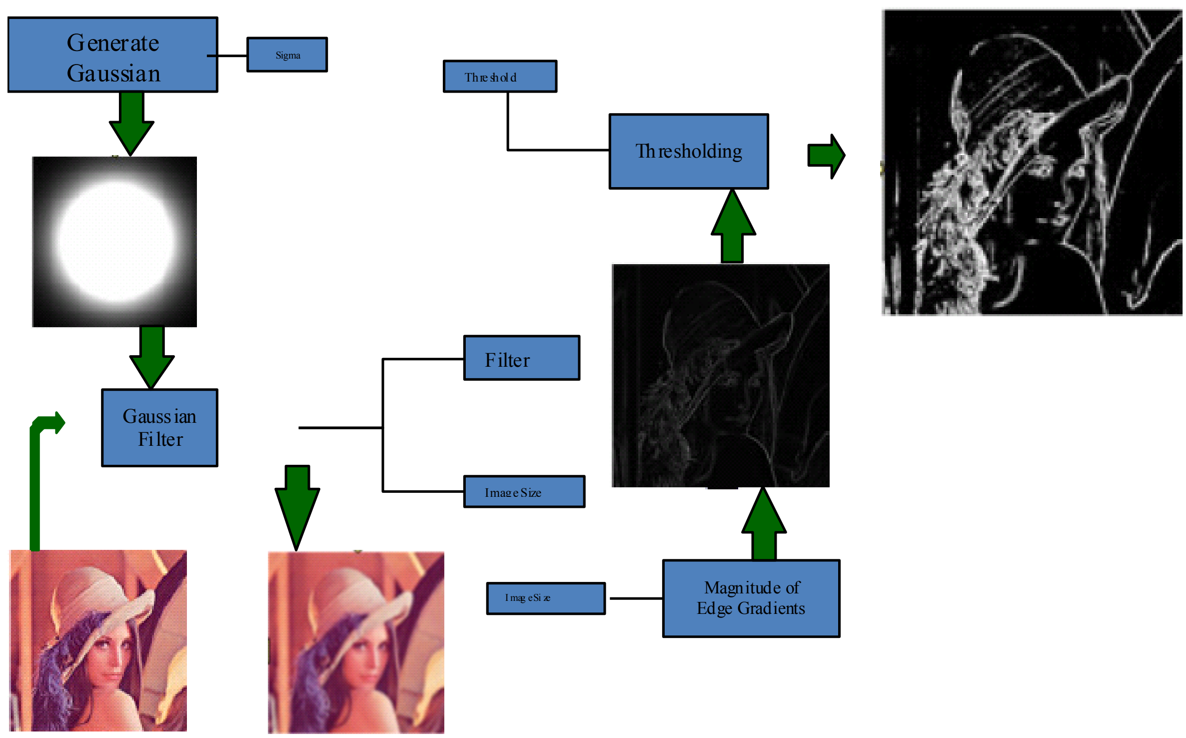

Figure 6. Canny Edge Detection Steps

3.6.3. Boundary Tracking and Segmentation Algorithm

Humans can discriminate amongst more than 30,000 optical groupings, and can distinguish objects in the span of a few hundred milliseconds. The field of computational object identification that attempts to augment this undertaking is nowhere near human performance, however, significant advancement in the previous few years has been made. Part of the work our team focused on was to ascertain a couple types of trailblazing algorithms that attempt to achieve human tier performance. Image segmentation is the partitioning of an image or video stream into groups of pixels that correspond to objects or parts of objects. This process takes cues from similarity of pixel brightness, color, texture and motion, as well as other information derived from predetermined object classifications. A boundary detection algorithm, takes the cues from these large datasets of human derived objects to calibrate the cues and learn optimum cue grouping schemes to establish boundaries of identified segments of an image or video stream.

We separate the problem of boundary detection from what is characteristically discussed as edge detection. A boundary is a contour in the image plane that denotes a transformation in pixel state from one object or shape to another. In contrast, an edge is frequently expressed as a sudden variation in some image feature such as brightness or color. Edge detection is thus one practice that is usually applied toward the objective of boundary detection. Another technique would be to distinguish objects in the image or video stream and use that evidence to infer the boundary.

To measure accuracy of an established boundary detection and image segmentation algorithm, it is recommended that the precision-recall curve is used to capture the trade off between accuracy and noise as the detector threshold varies. Precision-recall curves are a standard appraisal technique for evaluating boundary detection and image segmentation algorithms. Precision is the likelihood that the detector’s signal is valid, and recall is the probability that the human derived data was detected.

The two steps of cue optimization are to first optimize each of the cues like brightness, color, texture, etc independently from human derived benchmarks; then combine cues to develop a single algorithm by using logistic regression which may or may not eliminate redundant cues. This finally incorporates all value added cues into a single measurable metric.

The computational times required for training and evaluating the algorithm varies by several orders of magnitude. For training, only several minutes is necessary assuming human derived stock photos and video have already been accumulated. The boundary estimation typically can consume many hours depending on the length of video or number of pixels that are introduced. The hardest task in utilization of this algorithm would be accumulating a large data set of human-labeled boundaries in natural images, to help formulate the task of cue combination for boundary detection.

3.6.4. Dynamic Processing Algorithm

Advancing technology enables large amount of video data to be collected. However, the amount of data has grown much faster than the resource power required to process them. Sophisticated algorithms, like the Canny algorithm or facial detection algorithm, are typically known for their high accuracy in processing videos, but requires a large amount of computational resources. Thus it is not time efficient when processing 12-hour videos having 30 frames-per-second (fps). While less accurate algorithms, like the genetic algorithm or background detection algorithm, can process the video at a faster rate, they often fail to identify major details such as destroyed bridges or damaged buildings. These inaccurate algorithms are herein considered as *cheap* and the accurate ones are herein considered as *expensive* referring to the amount of processing power they required. Chen et al., authors of *Dynamic Processing Allocation in* Video, assumed that the expensive algorithm is over 90% accurate and may be in a form of a human analyst. To reap the benefit of both speed and accuracy, Dynamic Processing Algorithm (DPA) is a method that employs both the cheap and expensive algorithms along with an inference algorithm to manage processing power. In other words, processing capability is improved by determining where in the relevant part of the video to apply the expensive algorithm. Figure7 below shows the general logic of DPA.

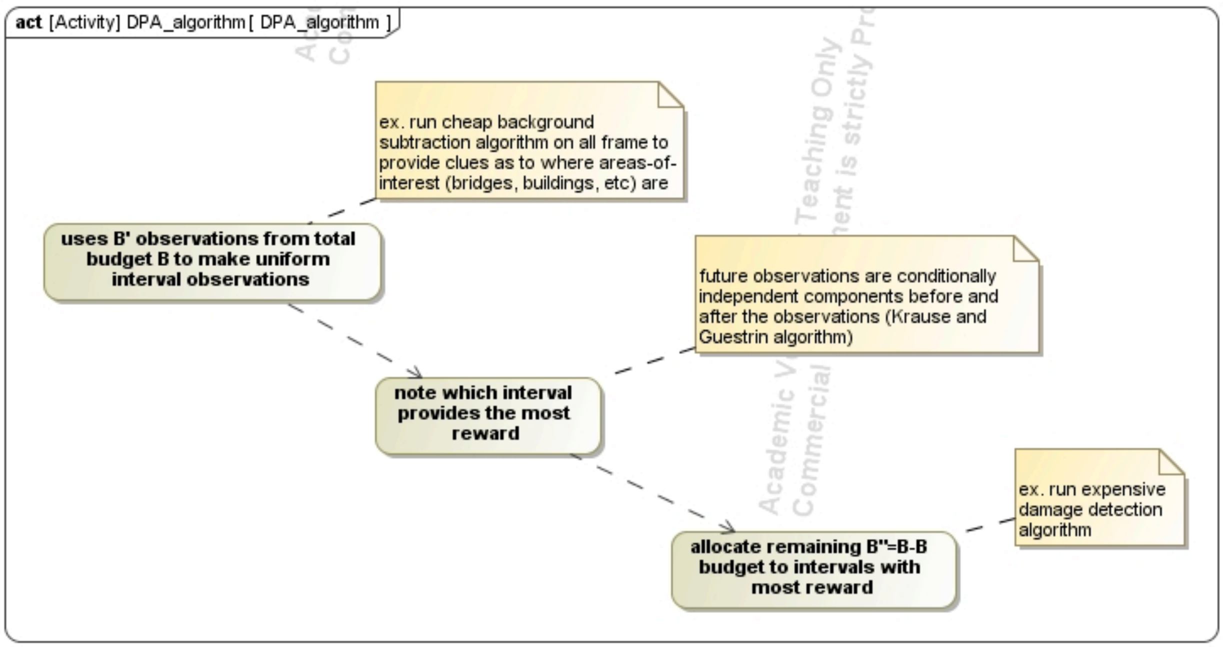

Figure 7. General Logic of DPA

A cheap algorithm is applied to each frame of the video, and depending on the desirability of the output, DPA then directs an expensive algorithm to the desired frame for future data processing and analysis. A second-order Markov model is used to determine two observations for each frame: the output of the cheap algorithm and potentially the output of the expensive algorithm. The example used by the authors is first applying a cheap background subtraction (Figure 8) to provide a clue of whether people are in each frame. If so, an expensive face detector (Figure 9) is applied to that frame. We can apply the same logic when searching for damages in a video. A cheap background subtraction algorithm is first applied to determine if a man-made structure, for example, exist in the video frame. If so, then an more robust algorithm is applied to that frame to determine if there are any damages.

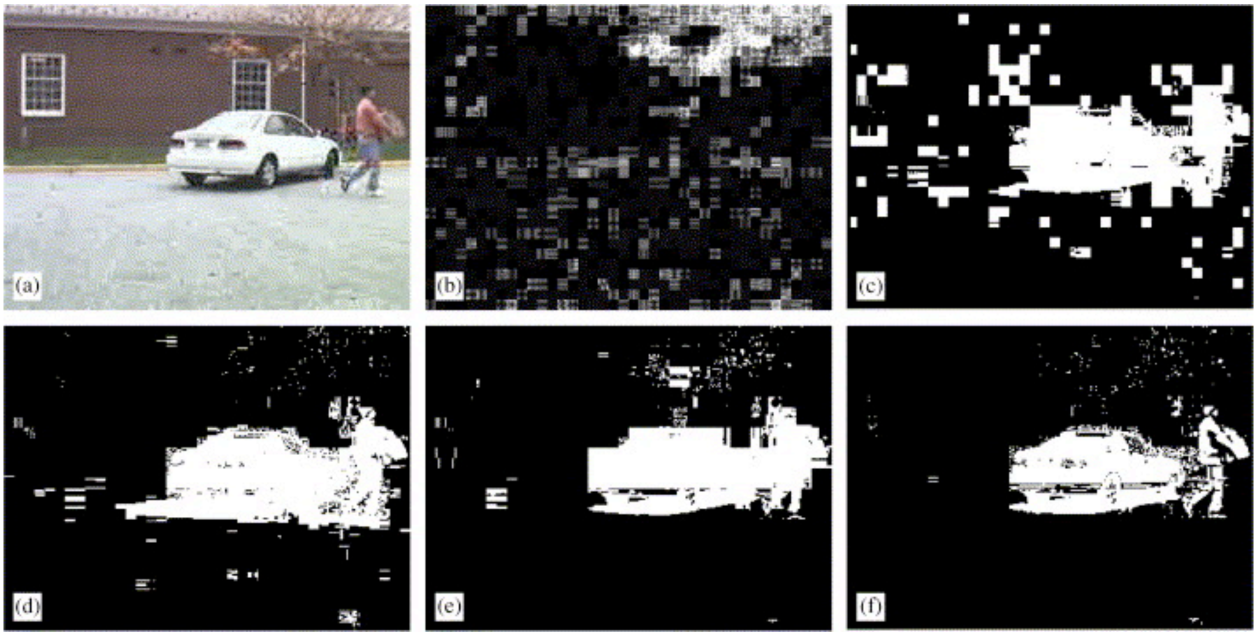

Figure 8. Background detection results on a compressed video: (a) original image, (b) standard deviations, (c) unimodal model , (d) MOG, (e) Kernel, (f) CB

Table 2. Examples of methods in face detection - the same can be applied in detecting structural damage

Methods of Face Detection
1. Top-down knowledge-based method – determines relationship between structural features such as eyes, nose, and mouth.
2. Bottom-up feature-based method – seek invariant structural features such as eyebrows, hair texture, and skin color.
3. Template-based method – compare faces with input image.
4. Appearance-based method – use statistical analysis and machine learning techniques.

DPA is beneficial in that it uses multiple algorithms to achieve a balance between accuracy and speed in video data processing. This is also a concern because having a plurality of algorithms complicates the logic required to detect areas-of-interest. DPA can be applied to videos with large number of frames because the processing resources are managed to distribute the expensive algorithm sparingly throughout the entire video. As such, DPA is not meant for simultaneous capture and processing; it requires an end point to determine where to evenly allocate the expensive algorithms. Although it is not within our scope to identify a time efficient algorithm, as the processing speed relies heavily on the hardware capabilities, DPA provides an alternate view into video data processing with a major consideration into time constraint.

3.6.5. End-to-End Approach to Damage Detection

Figure 9 below depicts a generalized diagram of the approach employed for damage assessment using color indices, edge intensities and local variances. There is no specific name to the algorithm outlined below. Note that its high-level logic may overlap with the algorithms in the previous section. This section is primarily based on Derya Ozisik's *Post-earthquake Damage Assessment Using Satellite and Aerial Video Imagery.*

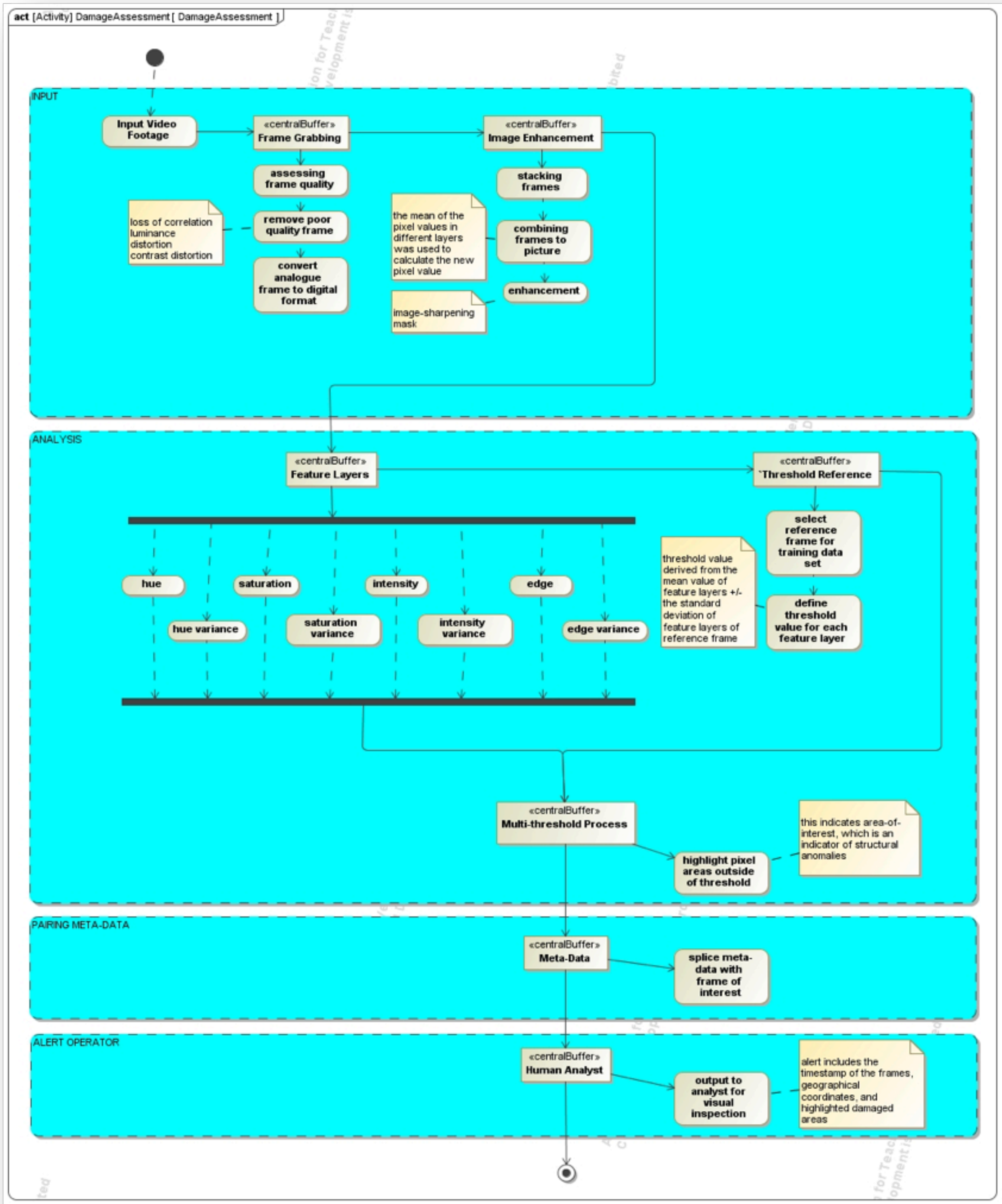

Figure 9. General Approach SYSML Diagram

The input stage receives video footage and frame quality is assessed. The poor quality frames are removed and the better quality frames are converted to digital format. The frames are rated on loss of correlation, luminance, distortion, and so on. An enhanced frame is created by *stacking* digitize frames together. The mean pixel values in the different digitized layers are used to calculate the new pixel value of the enhanced frame (see Figure 10 below).

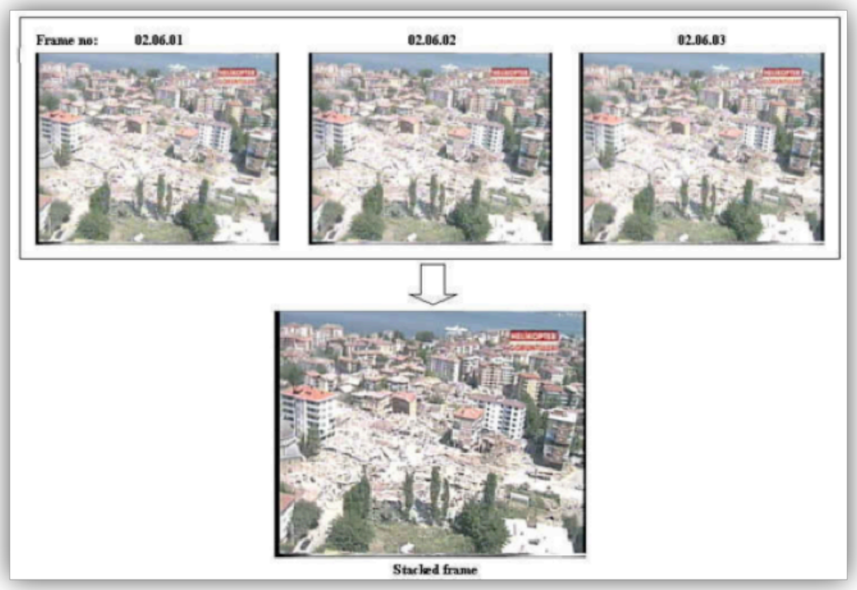

Figure 10. Frame Enhancement

In the processing stage, 8 featured layers of the frame are created based on hue, saturation, intensity, edge, hue variance, saturation variance, intensity variance, and edge variance. Figure 11 below shows these featured layers. A reference frame is selected to define the threshold value for each featured layer; the threshold value is derived from the mean value of the feature layers +/- the standard deviation of the feature layers of reference frame. In the multi-threshold process, the featured layers are compared to the threshold value of the reference frame to determine the location of the areas-of-interest, as shown in Figure 12.

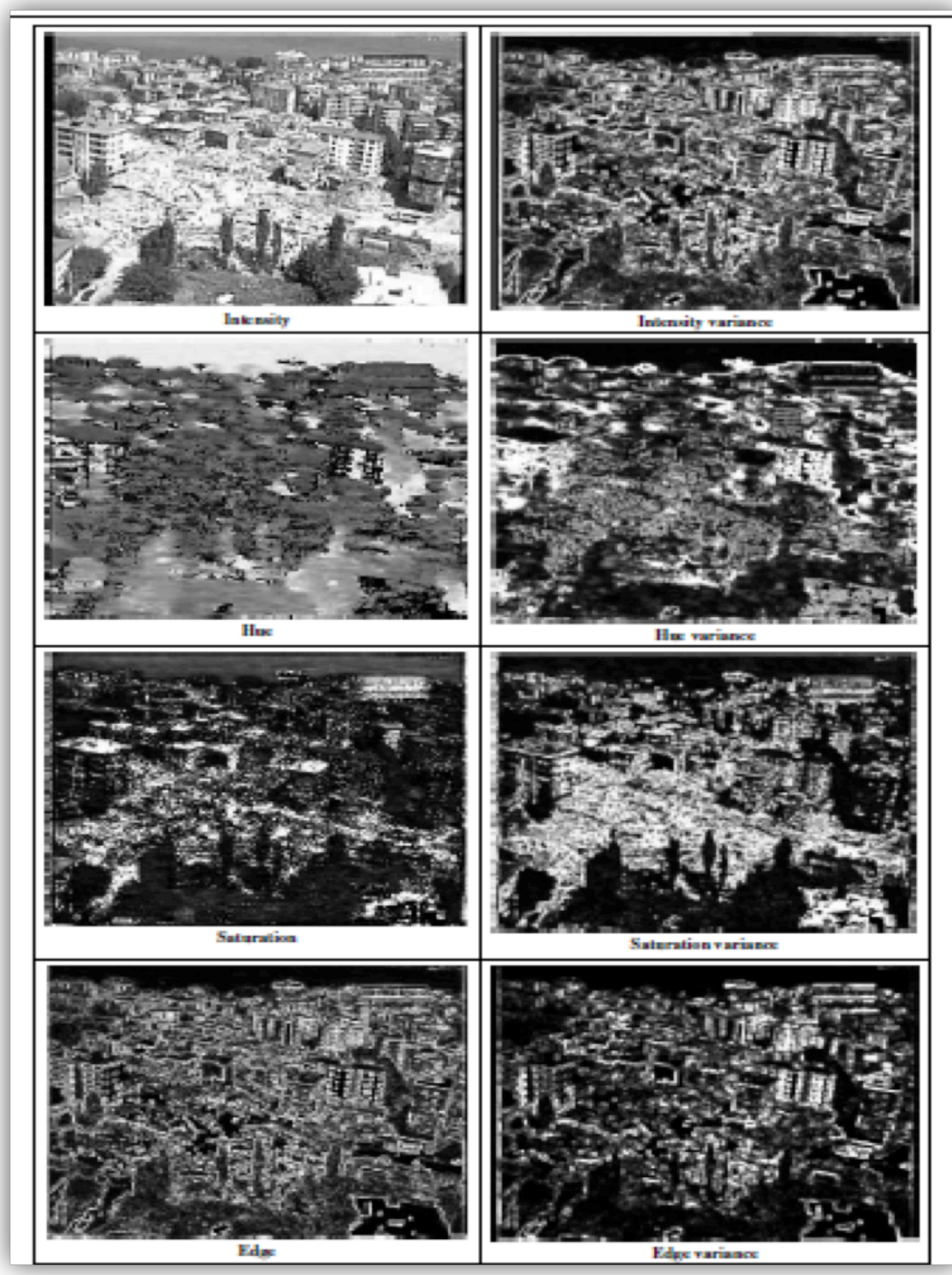

Figure 11. Featured Layers

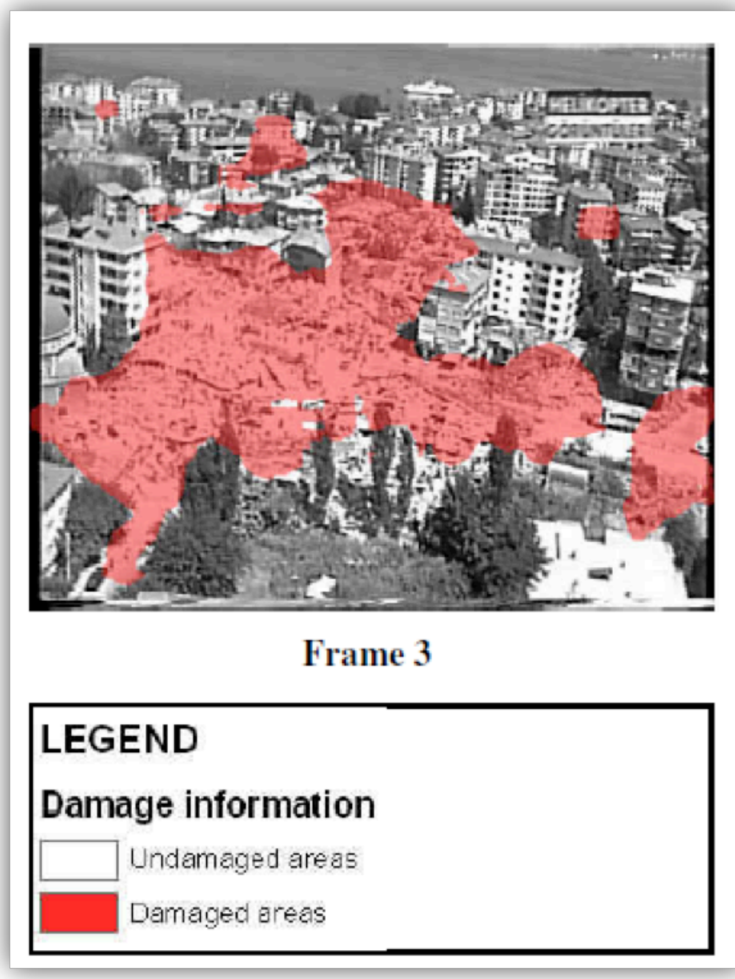

Figure 12. Damaged Areas

Meta-data such as coordinate location or time lapse of the video footage are spliced with the frames deemed to contain the area-of-interest. The areas-of-interest, in this instance, are highlighted red in the frame to output to the operator. Other means of system output may be a circular border or an audio alarm is activated to alert the operator.

3.6.6. Functional Capabilities Needed

The functions needed to process video information are as follows:

- The system will have an input module that will allow up to twelve hours of video to be loaded for processing
- In the event a before-image is needed for comparison, the system should be able to retrieve before image comparisons for the video surveillance area
- The system shall be able to take the video that has been loaded and go through each frame to identify any infrastructure damage and thermal bloom situations
- The system shall generate a notification or alert for each video segment where an anomaly is detected
- The system should allow users to verify whether an anomaly detection is of interest for disaster relief
- The system should manage records of all anomaly detections as well as which have been verified as valid versus not valid

4. Functional Capabilities

Figure 13. Four Stages for Damage Detection

4.1. Operations

The operational approach to detecting areas-of-interests, such as structural damages, can be categorized into four stages:

1. **Input stage**: when the system receives aerial video footage and the system gathers image/video data from other sources corresponding the geographical areas being assessed.
2. **Processing stage**: when the system analyzes the data collected.
3. **Pairing meta-data stage**: when the system pairs the meta-data corresponding to what the processing stage outputs as the areas-of-interest.
4. **Alert stage**: when the system alerts an operator of the areas-of-interest.

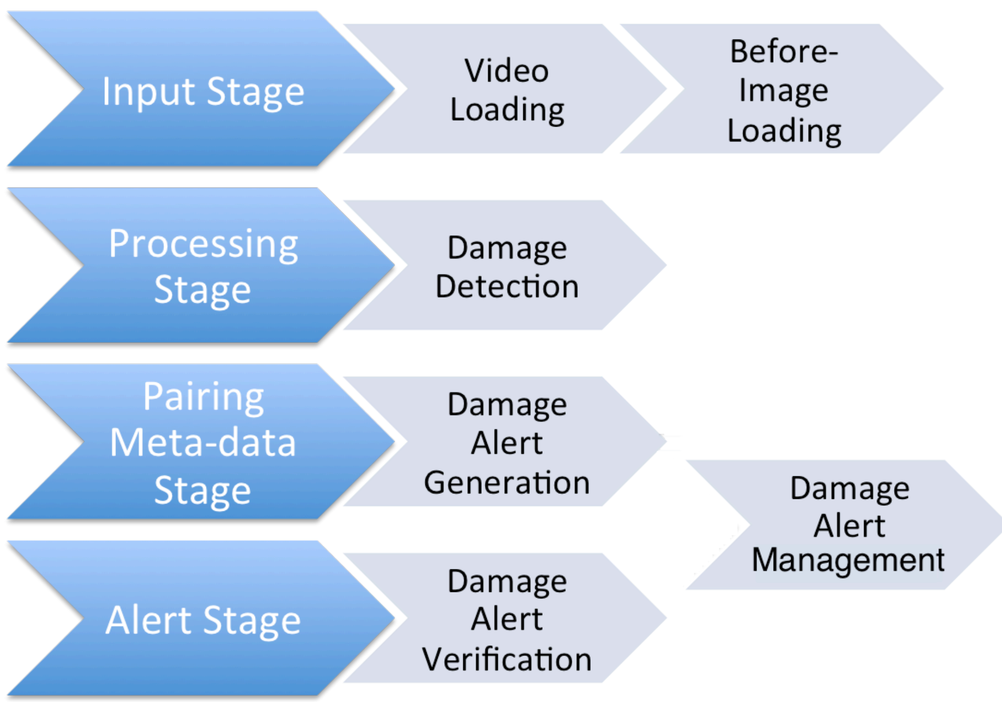

Figure 14. Correlation of Stages to Functional Capabilities

4.1.1. Video Loading

The video loading functional capability allows for videos to be inputted into the system, either through a user loading a video file or an automatic recognition if a new surveillance video is available at a given location. Depending on the hardware capabilities, the video data may be simultaneously transferred during capturing or may be physically loaded by a human into the system for data analysis. This function is considered to be the input stage and is important in order for the system to gain access to the footage at hand.

4.1.2. Before-Image Retrieval for Reference Comparison

Certain imaging algorithms require a before-image to compare to the collected video in order to identify areas of damage. In this case, the system must be able to gather the before-imagery required. We are assuming that this imagery exists and that the system shall have access to these sources. Therefore an additional functional capability is needed to aggregate the require before-images required for the algorithms. This function is also considered as part of the input stage.

The before-image can be categorized into two types. The first type is the actual image, or video footage, of the same geographical location prior to the natural disaster has occurred. This is required for those particular algorithms that compares the current and prior images to determine any differences in the geographical terrain. Significant differences may suggest an area-of-interest, or structural damages. The second type is a repository of images depicting numerous examples of how damages would appear. These images were initially identified by a human and classified accordingly with respect to the features that resembles a collapsed bridge or a destroyed building, for example. The algorithms requiring this second type of before-image would *learn* from the repository and used that as a guideline to identify potential damages in the video.

4.1.3. Damage Detection

Damage detection is the core function of this system. This function deals with given all the data needed, the system must be able to process the data and identify areas where damage has occurred. The system needs to be able to recognize anomalies in the video. This function is part of the processing stage employs the algorithms identified in section 3.6.

4.1.4. Damage Alerting

Once the damage detection portion of the system has identified a region of the video or imagery that could be of interest, the system needs to be able generate an alert around the portion of the video that the analyst should look at. This alert could be a box drawn around the segment of the image or highlighting the image region where the anomaly is detected. It should also identify the time range where this anomaly can be noted.

4.1.5. Detection Alert Verification

When an anomaly is noted and an alert is generated for the first time, the system should have a user verify whether the alert actually identifies disaster damage. The system should then catalog the anomaly as a false alarm so the damage detection algorithms can be improved. For instance, an alert with an image that is verified as an actual damage would be add into the pre-image repository so that the system can continue to learn.

4.1.6. Detection Alert Management

The system should manage all damage alerts that have been created. It should manage the alerts in three categories: alerts that an analyst has not yet looked at or verified, alerts that have been identified as accurate, and alerts that were false alarms. Users using the alert management system should be able to determine which video an alert came from, when and where that video is from, and also when the alert was verified as true or a false positive and who the analyst was who reviewed the alert. The alert management function should also take users directly to the alert that was generated. This function the pairing meta-data stage and alert stage.

5. Feasibility Study

The main goal of this study is to determine the feasibility of various conceptual designs that could potentially be used for different types of post-natural disaster damage assessments with the ultimate objective of replacing current modus operandi This section will cover the following information: what the conceptual designs are, what the evaluation criteria for the conceptual designs are, how well each design meets the criteria defined, the operational scenarios that the system will encounter, design recommendations for each scenario, and future work to be done. This pilot study will examine a couple operational scenarios that will test the effectiveness of these proposed designs.

5.1. Evaluation Criteria

There are three key criteria used to evaluate the feasibility of a system to process aerial video for damage assessment. They are: cost, accuracy, and speed. In order for a solution to be considered feasible, it must balance the three criteria in a way that each criteria suits the needs of the customer.

5.1.1. Cost

The cost of building, deploying and maintaining a system to process aerial videos should be less than the cost of hiring an analyst to do the work. The system will be evaluated on how much it costs to develop and operate the system over the system lifespan compared to how much it would cost to pay a group of analysts to do the same work over the lifespan of the system.

5.1.2. Speed

Speed represents the time the system takes to process a video and return the alerts that have been generated. The current baseline is it takes twelve hours for an analyst to watch a twelve hour video. All conceptual system designs will be evaluated relative to this twelve hour benchmark.

5.1.3. Accuracy

Accuracy represents the number of damage events that the system properly identifies over the total number of damage events that are present in the video. The system should at the very least hit the accuracy of a human analyst watching 12 hours of video. This accuracy will need to be balanced with the number of false alarms, or incidents where the system falsely notes a damage event and none exists. The system must have fewer false alarms than 50% of its total alarms. The system should reduce the overall time required to process aerial video.

5.1.4. System Impacts

The USAF will endure a financial impact if the decision to implement an automated solution is chosen. A hardware analysis will be needed to analyze the necessary computing power of the automated algorithm as well as environmental studies such as server storage and cooling. The overall solution that encompasses the hardware and software must reside in a secure location and in a cooled environment to prevent overheating of the hardware elements. This has the potential to impact facility storage and/or require additional facility space to accommodate the new system. Further impacts will be anticipated to affect the personnel level on staff if an automated solution is chosen to replace human analysts. A new staffing plan will be needed to eliminate redundancy of tasks performed and new positions will be needed to manage the system. A retraining program may be necessary to train new personnel on how to operate the system as well as how to maintain it.

5.2. Scenarios

5.2.1. Mitigatable Disaster Damage

Types: Hurricanes, Monsoons, Floods, Solar Flare, Snowstorm, Human-inflicted

5.2.2. Recommended System Design

Natural disasters such as hurricanes, monsoons, floods, and snowstorms all present the danger of liquid water to an otherwise dry human designed landscape. Cool water from such storms presents an opportunity to use Infrared cameras to easily detect temperature changes on warmer or colder backgrounds. Infrared or IR cameras offer strong and fast image recognition when dealing with rising water levels indicating a flood or snowstorm. IR cameras can easily spot and alert to water areas and humans in flooded water because of the strong contrast of body temperature to the surrounding water. The outlines of flooded areas are also easily identifiable because of water's chemical characteristic to rest at a different temperature in comparison to surrounding landscapes. For these types of disasters, automated algorithms with the ability to use IR sensors are best suited for disaster relief efforts.

5.2.3. Unmitigatable Disaster Damage

Types: Earthquakes, Wildfires, Drought, Tsunami, Tornadoes, Sinkholes, Meteorite

5.2.4. Recommended System Design

Natural or man-made disasters involving fires can use the same principles as floods and water related disasters because IR images will easily detect fire's strong temperature and light difference in comparison to its surrounding areas. However, IR cameras were designed to detect heat changes and the sharp impulse of light emitted from fires can render the IR sensor incapable of performing its job because of the glare on the sensor's lens. Other black and white monochromatic sensors of electro-optical sensors will work well when IR sensors are rendered less than fully capable. The type of algorithm employed will be based upon the disaster classification and whether or not prior video was able to be obtained. Certain disasters like earthquakes give little to no warning they are coming so obtaining previous images of a particular disaster area will be unlikely. In this situation, algorithms that do not require previous images will work best. Meteorites, however, are generally tracked by NASA and their collision course with Earth will be known well enough in advance to plot the impact zone and obtain previous images. These algorithms that employ previous images are more accurate because of image comparison process so they are high recommended to be used when previous images can be captured. The ultimate decision on which algorithm to use when the time comes will depend on a number of factors. Further research is needed to advance the reliability of the technology and investments will need to be made to develop a robust USAF solution that can meet their mission requirements while simultaneously remaining within their operating budget.

5.3. Recommendation for Further Research

The USAF needs to invest in a feasibility study of their own to determine if the available technology meets their mission requirements. One possibility is to extend their research into the specific algorithms described in this project. Another, more general approach, is to focus their research on software and hardware capabilities to process *big data*.

It is recommended that additional research continues with damage detection algorithms. In particular, and to the contrary, focus on algorithms that can identify non-areas-of-interest. Such capabilities will allow the system to better identify the damaged areas by removing background information such as vegetation or large water mass. This type of logic working in conjunction with damage detection algorithms may serve as a check-and-balance to increase the accuracy of damage detection.

Due to the scope of this project, hardware capabilities were not considered in detail for this damage detection system. Therefore, a recommendation for future research would be in hardware designs to speed up processing power. A potential approach would be to employ parallel computing using redundant hardware for simultaneous video analysis. For example, twelve systems could be running simultaneously and each processing a 1-hour segment of the 12-hour video. This would greatly reduce processing time but may drive up costs.

Crowdsourcing is another approach recommended for future research. A high-level logic may start with the system ingesting the video data and splitting the video into 10-second video segments. After that, the system would distribute the segments to millions of registered users. The users would watch the short video clip and ping back a message of what they see. If multiple messages of a particular video segment includes keywords such as "damaged", "collapsed", "broken" and so on, there is a strong indication of a positive damage detection. This entire process may be time consuming, but is highly accurate based on human feedbacks. Needless to say, there will be issues regarding the legality of distributing aerial video to users across the country.

5.4. Conclusion

This project began as a research study into the various alternatives for automatically processing aerial video images as a mean for detecting damages. The algorithms and operational process outlined in the CONOPS fulfill the project requirements. That is, we have identified approaches as to how the system ingest video data, process the data, and generate alerts for an analyst to review. Conceptually, the approaches make sense, but in reality, we must consider how the system will perform when implemented with real-world limitations - accuracy, speed and cost.

There is a need for the system to have some form of a reference image to compare and determine whether the video frame being analyzed contains any anomalies. The accuracy of the system is reliant on the these reference images. To have a repertoire of pre-identified images for the system to learn from may be feasible, but to capture and store every types of damage scenario may not be. The other solution is to obtain pre-disaster videos of the geographical area, but doing so would be costly since that would require a massive amount of resources to fly out and capture videos of all areas. Further, some disasters are impossible to predict and they could occur anywhere. The algorithms as identified are sufficient in detecting damaged areas with acceptable levels of accuracy as high as 86%, but the threshold of the acceptable accuracy level relies on the severity of the damage and what is at stake.

The algorithms are, however, not fast enough to surpass human processing. Currently, the ideal standard is for the system to at least process as fast and as accurate as a human analyst. The current system is not feasible for damage assessment that is reliant on the urgency of time. Such situations include assessment for first responder or for emergency rescue. The system is, however, feasible for research studies such as tracking the changes of coastal lines after a tsunami or determining damages for insurance purposes, for example.

The cost of the system is driven by both software and hardware capabilities. Ideally, the cost to build and maintain the system should be less than that of the manpower required to watch the video in a given amount of time. As such, it is not feasible with the current computing technology to build a system and expect it to be within a reasonable price to meet the ideal requirements.

There are many outlets for improvement to the aerial video processing system. More research into sophisticated approach is within reach that can solve the biggest challenge for this project - processing a large amount of data within a reasonable amount of time with an acceptable accuracy. With respect to accuracy, speed and cost, it is inevitable that at least one element is compromised in exchange for superior performances in the other two.

Acronyms and Abbreviations

Acronym	Meaning
USAF	United States Air Force
DoD	Department of Defense
UAV	Unmanned Aerial Vehicle
COTS	Commercial Off The Shelf

A.1. References

1. Uprenda P Poudel, Gongkang Fu, Kian Ye, Structural Damage Detection Using Digital Video Imaging and Wavelet Transformation, Center for Advanced Bridge Engineering, Civil, and Environmental Engineering Department, Wayne State University
2. Daozheng Chen, Mustafa Bilgic, Lise Getoor, David Jacobs, Dynamic Processing Allocation in Video, IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol 33, No 11, Novinger 2011
3. Harold Trease, Daniel Fraser, Rob Farber, Stephen Elbert, Using Transaction Based Parallel Computing to Solve Image Processing and Computational Physics Problems, www.cca08.org
4. Motion Imagery Standards Board, Inserting Time Stamps and Metadata in High Definition Uncompressed Video, MISB STD 0605.3, June 9th 2011
5. Motion Imagery Standards Board, Compression for Infrared Motion Imagery, MISB Standard 0404, December 3rd 2009
6. Motion Imagery Standards Board, MISB Profile for Aerial Surveillance and Photogrammetry Applications (ASPA), MISB Standard 0301.5 Version 1.5, July 26th 2010
7. Ryan Hruska, Greg Lancaster, Jerry Harbour, Shane Cherry, Small UAV-Acquired High Resolution Georeferenced Still Imagery, Idaho National Laboratory, September 2005
8. N. Voisine, S. Dasiopoulou, V. Mezaris, E. Spyroum T. Athanasiadis, I. Kompatsiaris, Y. Avrithis, M. G. Strintzis, Knowledge Assisted Video Analysis Using A Genetic Algorithm, Information Processing Laboratory, Electrical and Computer Engineering Department, Aristotle University of Thessaloniki, Greece.
9. Yuko Roodt, Willem Visser, Willem Clarke, Image Processing on the GPU: Implementing the Canny Edge Detection Algorithm, University of Johannesburg, http://www.prasa.org/proceedings/2007/prasa07-26.pdf

10. Ozisik, D. *Post-earthquake Damage Assessment Using Satellite and Aerial Video Imagery.* Enschede, The Netherlands: International Institute for Geo-information Science and Earth Observation, 2004.

11. Kim, K., Chalidabhongse, T., Harwood, D., & Davis, L. (2005). *Real-Time Imaging*. Retrieved from SciVerse - ScienceDirect: http://www.sciencedirect.com/science/article/pii/S1077201405000057

12. Voisine, N., Dasiopoulou, S., Mezaris, V., Spyrou, E., Athanasiadis, T., Kompatsiaris, I., ... & Strintzis, M. G. (2005, April). Knowledge-assisted video analysis using a genetic algorithm. In *Proc. of 6th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 2005)*.

13. Montoya, L. (2002). *Urban Disaster Management A Case Study of Earthquake Risk Assessment in Cartago, Costa Rica*. PhD Thesis, Utrecht University, Enschede.

14. Dosi, Cesare. (June 2001) *Environmental values, valuation methods, and natural disaster damage assessment*. Santiago, Chile.