FORMULATING AN IMAGE MATCHING STRATEGY FOR TERRESTRIAL STEM DATA COLLECTION USING A MULTISENSOR VIDEO SYSTEM

Neil Clark, research forester
USDA Forest Service
Southern Research Station 4702 – Integrated Life Cycle of Wood
1650 Ramble Road
Blacksburg, Virginia 24061
neclark@vt.edu

ABSTRACT

A multisensor video system has been developed incorporating a CCD video camera, a 3-axis magnetometer, and a laser-rangefinding device, for the purpose of measuring individual tree stems. While preliminary results show promise, some changes are needed to improve the accuracy and efficiency of the system. Image matching is needed to improve the accuracy of length measurements spanning multiple video frames. Formulation of a knowledge base can enhance the speed and robustness of the matching algorithm. This paper will demonstrate how characteristics of sub-canopy video collection and parameters of the video system can be used to formulate a suitable image matching algorithm.

INTRODUCTION

Image matching is a preliminary step to image mosaicking. In the realm of satellite imagery, and some aerial imagery, data is acquired and delivered with a certain amount of geometric registration. Normally with aerial data this is done by registration to ground control points or tie points between images. In practice, these points are usually manually located. Then normally only a simple transformation is required to match the images as the scale, translation, and rotation parameters are controlled. The entire aerial process is simplified due to calibrated geometric parameters of traditional metric aerial cameras and digital satellite arrays, a small parallax to flying height ratio, sufficient overlap, and lack of significant occlusions.

These are the traditional tools and techniques that have been used in the forestry sector over the last half-century. The last decade has introduced affordable and extremely powerful numerical processing ability. Likewise, many traditionally difficult, labor-intensive, analogue methods have been approached in a digital manner. In some cases, there have been major improvements and production gains made. In other cases, programming a computer to perform certain complex analyses has presented some new problems. Incorporating adaptive methods, context, and flexibility into exacting computer systems has proven to be difficult. There is a large amount of image processing research occurring to solve problems from industrial mensuration and quality control (Atkinson 1996) to 3-D video stitching for virtual reality displays (Peleg et al. 2000). To date many useful general algorithms have been developed, but the complexities of each problem often render these general approaches incapable or inefficient solutions.

The problem faced here is less concerned with total image mosaicking as it is with finding a rapid, automated, and robust method for matching video frames captured in a sub-canopy environment. A multisensor video system, incorporating a CCD video camera, a 3-axis magnetometer, and a laser-rangefinding device, has been developed for the purpose of measuring individual tree stems (Clark 2001). The concept being to capture image, range, and orientation data in the woods (Figure 1), then postprocessing these data to obtain certain metrics (e.g., diameters, heights, form) and perhaps model the stem in 3 dimensions.

This paper will present some of the obstacles involved in accomplishing such a task. Methods of tackling these obstacles will be presented with their associated advantages and disadvantages. Image matching is an admittedly time consuming task, yet this time can be reduced by a data driven operation sequence (Sonka et al. 1993). Data

specific to the application of this instrument in real world conditions will be analyzed to determine a favorable matching strategy.

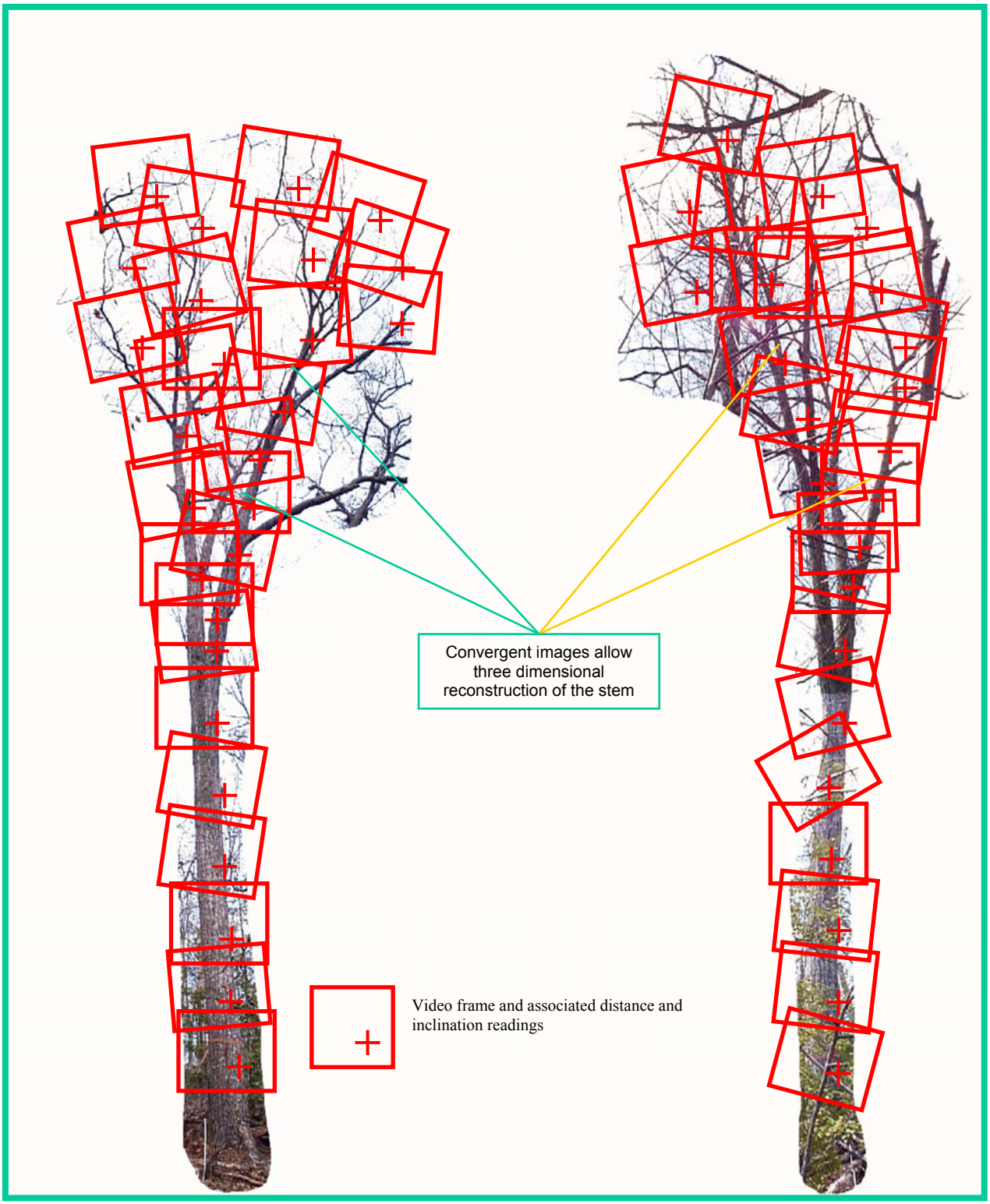

Figure 1. Video frames are collected by scanning the stem using the multisensor instrument.

MATCHING CHALLENGE

There are many factors to consider in formulating an image matching strategy including acceptable precision/error limits, scene characteristics, instrument parameters, and the method of image acquisition. The goal of this project is to obtain a confident image match with minimal overlap at a moderate speed. Finding one definite match on the stem is more important than mosaicking the images, so precise determination of all orientation parameters is unnecessary. Use of a third-party software developed for panoramic image formation demonstrated that matching could be performed rapidly on images of subcanopy forest scenes. However, this software had an approximately 20 percent failure rate on the frames analyzed.

Data reduction

Each video frame (image) from the camera used in this study has an output resolution of 720 x 480 pixels by 3 colors. This means that there are over 1 million values to analyze. Moreover, given the iterative nature of many of these processes, the computational burden and complexity is extreme. To overcome this problem the data must be compressed, aggregated, simplified, omitted, or reduced. However, any method that is chosen should not negate the ability to positively match the images. Concurrently with making the data reduction decision, the geometric primitives used for matching must also be chosen.

Selecting geometric primitives and type of analysis

There are numerous approaches commonly used to match images as evidenced by the immense number of publications. There are literally thousands of publications regarding matching based on a large number of methods applied to any number of geometric primitives, of any dimension within any domain, for many types of spatially related data. The two general classes of geometric primitives used in image matching are area-based or feature-based (Heipke 1996). Area-based methods consider the correlation of gray value windows (matrix of pixels). Usually a small window (*i* rows x *j* columns) also referred to as the template (*t*) window from one image is compared with sample (*s*) windows that are subsets of a larger window (M x N), and sum of squares differences or a correlation function such as,

$$corr = \frac{\sum_{row=1}^{i}\sum_{col=1}^{j}\left(BV_t(row,col)-\mu_t\right)\left(BV_s(row,col)-\mu_s\right)}{\sqrt{\sum_{row=1}^{i}\sum_{col=1}^{j}\left(BV_t(row,col)-\mu_t\right)^2\sum_{row=1}^{i}\sum_{col=1}^{j}\left(BV_s(row,col)-\mu_s\right)}}$$

where BV represents the brightness value of a pixel, and μ represents a mean brightness value of the small windows. This function is applied iteratively within the larger window to find a maximum correlation value. Problems can arise when the template window lacks significant texture or is not unique within the search image, resulting in multiple potential matches. Some work has been done to find an optimum window size (Okutomi 1992, Salama 1999). Occlusion, deformation, or illumination effects may also result in absence of any matches.

Feature-based methods extract features (points, edges, clusters, structures, etc.) from each image prior to matching. Depending on the effectiveness of the algorithm used, this usually minimizes some noise and gain effects. Flexibility must also be a consideration in the search algorithm in regards to occlusion and deformation.

Marapane and Trivedi (1994) list seven important characteristics of primitives: 1) dimensionality, 2) size, 3) contrast, 4) semantic content, 5) density of occurrence, 6) ease of extraction, and 7) uniqueness. Some of these characteristics are interrelated with others, e.g., contrast is related to semantic content and ease of extraction. A feature such as a point, with low dimensionality is less affected by spatial distortions than an edge, region, or surface. Size can be an important factor when using certain strategies such as “coarse-to-fine” where large

primitives will increase the efficiency of the search and small primitives will be overlooked. High semantic content usually will provide confirmation of a correct match. However, *a priori* knowledge of the scene is usually required and the computations to derive the information are sometimes complex and time consuming. As previously mentioned, if the density of occurrence is too great many ambiguous solutions are obtained which must then be pruned if a one-to-one match is required. The greater the complexity of the primitive the more difficult and time consuming it will be to extract. For instance, although an "airplane" may be a great primitive with tremendous semantic content, the amount of image processing and knowledge-base searching may prove to be unnecessary compared to a line matching strategy. Lastly, unique primitives reduce the number of ambiguous solutions resulting in a confident match.

Take the case of matching two views of a red laser pointer mark on a white screen for instance. This task could be approached by multiple means. A grayscale window selected at random using an area-based approach as mentioned previously, would likely not be the best approach. If the window did not contain the point, there would be no contrast (except perhaps noise), no uniqueness and no semantic content. However, if the window was selected to contain the point, there would be sufficient uniqueness and contrast to make a confident match. Either situation would have a huge computational burden. Alternatively, if *a priori* information was known about the search, i.e. searching for a single red point, the strategy could be improved. In this case, searching for pixels with a blue value below a certain threshold may be a more efficient method.

The objective is to find a unique entity (area or feature) that is distinct, visible in both images, and geometrically and radiometrically invariant. Geometric variation should not be extreme, as the viewing angle and range are not subjected to severe changes between frames. Radiometric variation is significant using video in the subcanopy environment. This problem can be addressed by quanitzation or classification. Brightness values at the extremities are more reliable as they are less likely to be lost to gain control adjustments.

Point features are the most geometrically invariant features and the least complex. The disadvantages of points are that they have little semantic content and their small size and dimensionality give them a high probability of redundancy. In the digital image, pixels can represent point values. Some points can have more semantic content than other points. In the laser pointer example, the pixel with the lowest blue or green value would have the greatest semantic content. Unfortunately in the natural scenes unique spectral points are rare and would be impossible to determine *a priori*. Edges can be point features that have more semantic content than a brightness value. Edges can have intensities and orientation. These edges can be further analyzed to find corners, which may be represented by point values having even greater semantic content. These corners or specified points features can further be combined to form a structure which will increase the uniqueness of the feature, but will also introduce the potential of geometric variance.

Lines, edge segments, and regions are all higher-level primitives of increasing complexity, thus increasing semantic content. Lines are not likely to be useful in the analysis of these natural scenes where perfect lines are unlikely to be present. Edge segments and regions are similar techniques focusing on discontinuity and continuity properties, respectively. Both of these techniques can also be computationally intensive and affected by radiometric variance. Because of the size and characteristics of the region based approach it is likely more stable than edge finding.

The approach that will be used from this point is to use preliminary analysis of the scene to drive the matching strategy. The least computationally intense methods will be attempted first. If a confident match is not obtained, the algorithm will progress through the more intensive and thorough methods.

IMAGE CHARACTERISTICS

218 video frames were selected from data collected at Bent Creek Experimental Forest in Asheville, North Carolina in June of 2000. In all there were 54 overlapping image sequences consisting of 117 hardwood and 101 pine frames. Frames were classified into 3 general categories based on stem position. There were 51 lower, 100 middle, and 67 upper stem frames. The background of the lower frames lacked any sky, middle frames contained

some sky but lacked extreme inclination, and the upper stem frames contained the tree crown and extreme inclination. These frames were examined for qualitative characteristics that would assist in determining a basis for an image matching strategy.

Generally, natural forest scenes are highly heterogeneous in regards to texture. The advantage to this is the ease in determination of unique primitives that reduce the probability of ambiguous solutions. The disadvantage is the difficulty in setting thresholds and constraints for the matching methods. The color space is not quite as diverse, which may lead to some confusion when coarse methods are used.

Spectral difference

Spectral characteristics using a video system have the potential of drastic change from one image frame to another since the dynamic range is being continuously adjusted (Figure 2). As a consequence, texture information may be lost between frames. Additionally in the context of sampling stems in the field, illumination is highly variable. The sun is a very intense light source. The amount of incident radiation is also determined by atmospheric conditions, vertical position relative to the horizon, the viewing azimuth in relation to the sun, and the canopy density for any given frame. Inherent with the desire to match frames at varying heights along a stem, the viewing angle in relation to the sun is changing constantly. This has effects on optimizing a matching strategy.

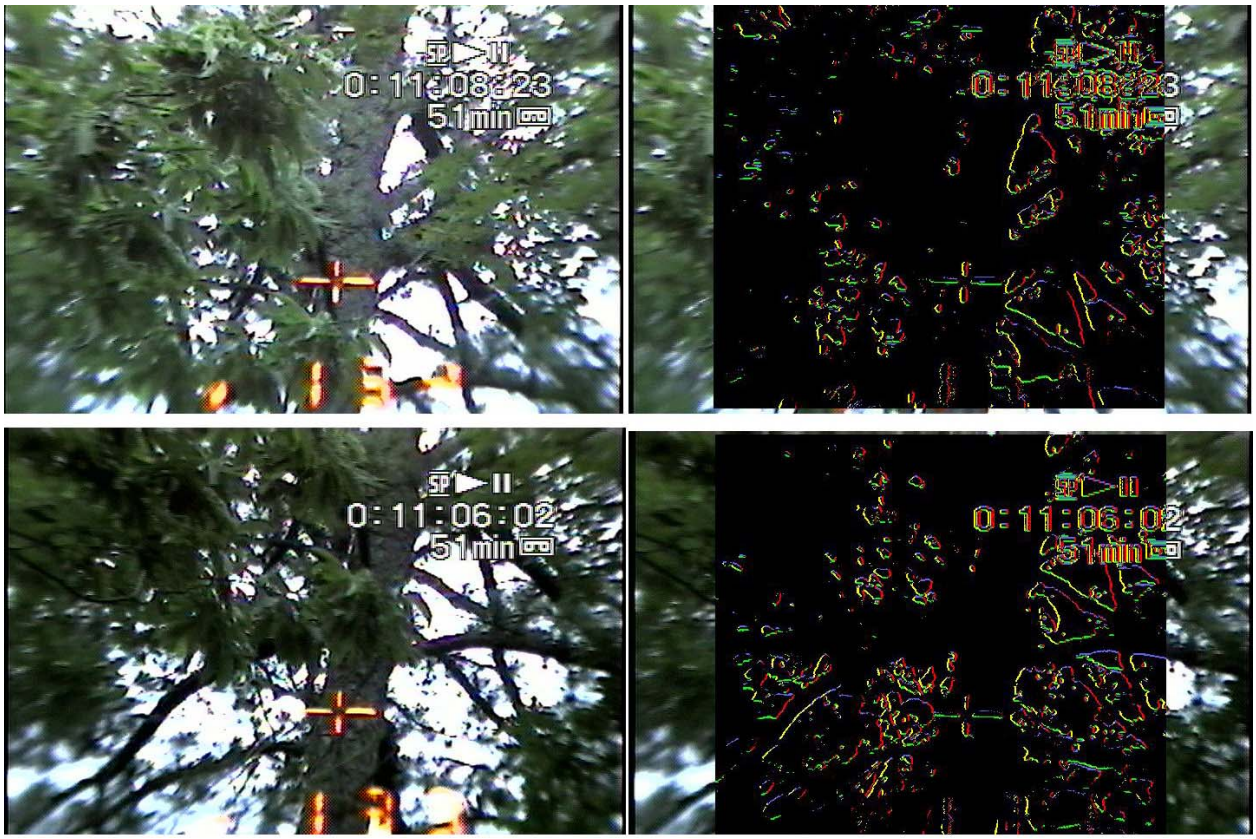

Figure 2. Example of automatic gain control effects on global image brightness

Upper images, such as those in Figure 2, almost always contain some amount of sky as background. Low texture and high brightness characterize sky regions. The automatic gain control adjusts for the bright conditions resulting also in a low textured and dark stem. One advantage in these scenes is a high contrast between the stem and sky simplifying separability of these two entities using a simple gray level threshold. This lack of texture would reduce the effectiveness of fine scale area-based methods. However, a region-based approach using sky patches, or an edge-based approach using branches or stem anomalies may produce a very unique primitive from which to obtain a highly confident match.

Lower images generally are not characterized by extreme step edges as those where the sky is in the background contrasted with the opaque stem. Lower images are also less likely to suffer from extreme changes in dynamic range. Some of the brightest pixels may appear on the stem depending on the reflectance properties of the bark. Some tree species, like white oak (*Quercus alba*) have bark that reflects all wavelengths (white light). Other species (e.g., "platy", *Pinus* bark) may absorb more diffuse radiation, but exhibit specular properties resulting in uncharacteristically high brightness values at certain viewpoints and illumination angles. The fact that a stem may exhibit spectral characteristics on both ends of the range may make background separation more difficult from a gray-level thresholding point of view, but it also provides textural information.

One characteristic that was found to be useful for the case of lower images with the stem in the foreground was a dominant color analysis. It can be seen in Figure 3 that green vegetation makes up a significant proportion of the background. The background is highly textured and has a large dynamic range compared to the stem (in most cases). This background may be used if the stem lacks unique texture or edge features, but only as a last resort as the difference in depth between the background and the stem my have detrimental effects on correct matching. Additionally, since leaves are the source of much of the background (or foreground) texture their spatial stability is not certain under windy conditions and changing view angles may have extreme effects on their spectral characteristics. This simple separation using dominant color is useful for limiting the area considered for matching. Limiting the textural analysis to the dominant color area will improve the results.

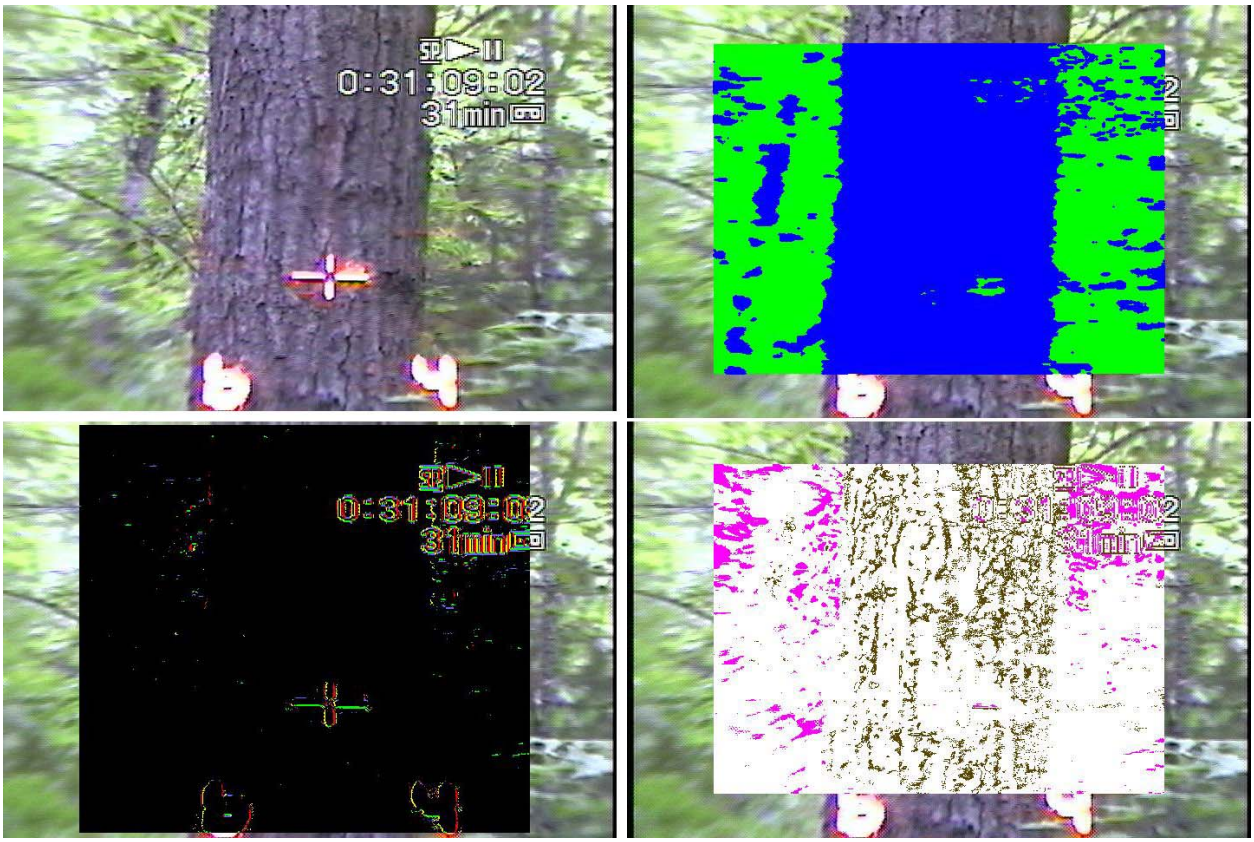

Figure 3. Image of the lower portion of the stem. The lower left portion demonstrates a lack of strong edges. The upper right portion shows how a dominant color analysis can efficiently stratify the image and focus the search. The lower right demonstrates how the local texture can be used to find unique features.

There are occasions where the stem is occluded to a significant extent on the lower images (Figure 4). In this case there may not be significant bark texture to assist in the matching. The dominant color analysis also may not provide any distinction. When this is the case, features in the foreground may have to be used to conduct the

matching, or perhaps creating a binary image based on brightness value threshold, and using a region, edge, or area search.

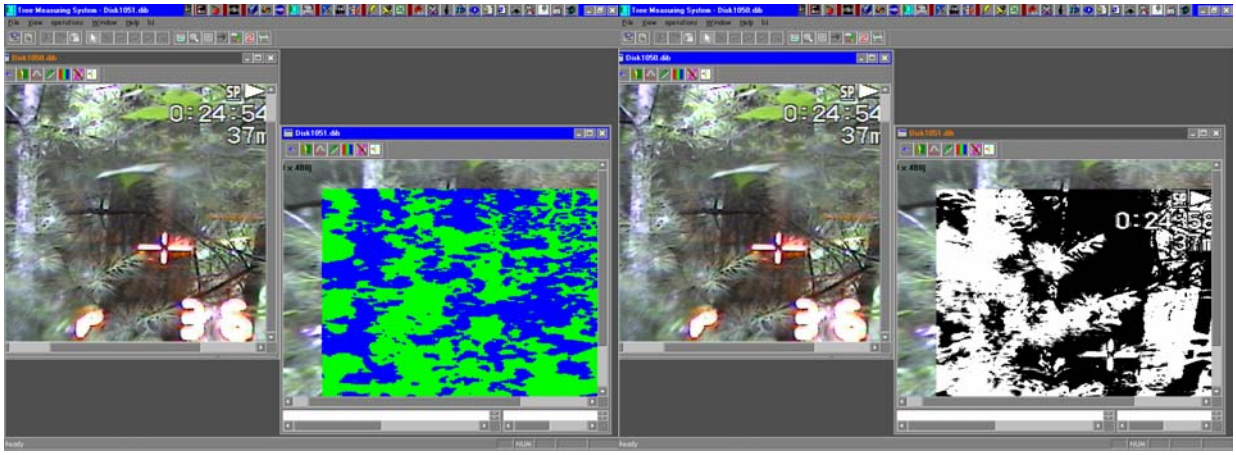

Figure 4. Image where the stem is in the background. In this instance dominant color analysis does not yield promising results and a global binary threshold is applied. Edges can be derived from the binary image and used for matching.

Spatial characteristics

Some spatial constraints are provided by the instrument. Intrinsically the instrument will be collecting data in an upright position resulting in all of the image frames to be oriented such that the bottom will be lower in object space than the top limiting the amount of rotation that will occur. Since the main orientation of a tree bole is vertical, that will be the primary direction of optical flow. Rotations about the x-axis (pitch), y-axis (roll), and z-axis (yaw) are captured by the instrument and used for initial positioning of the frames. These rotation readings made by the magnetometer are somewhat sensitive to camera motion, as such they are only considered for initial estimates and not actual measurements.

The tree should be scanned from bottom to top (or desired point of measurement) from a set position for each measurement. This should limit uncontrolled scale variation beyond that of the perspective projection. This also limits movement around the tree, so there are no great azimuth variations with respect to the tree stem. The pitch information from the magnetometer should be accurate enough to give an initial estimate of direction and magnitude of optical flow between frames and to guarantee overlap.

Due to the morphology of trees, diameter is inversely proportional to height. Intrinsically, following proper collection protocol, the stem should be targeted with the LED crosshairs. These constraints will assist in gleaning information helpful for separating the stem from the background. This will also assist in finding the edges of the stem, which will have to be done at a later stage to obtain diameter measurements. There are certain assumptions that can be made about stem form that will assist in determination of stem and background.

Hierarchical Approach

Hierarchical methods have been used by some (Marapane and Trivedi 1994) to obtain very consistent results. In these instances, higher-level primitives are used for guidance for lower level analysis for precision and confirmation. Since the goal of this study is a simple image matching of a single area rather than stereo analysis, this extra time is unwarranted. Here a hierarchical approach from lower to higher-level primitives is proposed for efficiency.

Primitive selection can be guided by image properties. Figure 5 shows the proposed approach. The terms upper and lower will be used here as previously with reference to their position on the stem and background properties (ground or sky). The terms top and bottom are used to denote the juxtaposition of the overlapping video frames to one another.

First, the top portion of the bottom image will be tested for edges of sufficient intensity. These strong edges will likely be apparent in both images and an edge matching strategy can be used. If a sufficient number of strong edges are not found, a determination will be made between upper and lower stem images. For the upper stem images, sky or leaves regions may be segmented and a region matching approach will be attempted. Lower images will be analyzed to determine whether a dominant color analysis will assist in stratifying the image. If sufficient texture exists among the blue dominated pixels, an area-based template matching approach will be used within this stratum. If texture is lacking among the blue dominated pixels, a binary threshold will be applied to the image and an edge finding strategy will be employed.

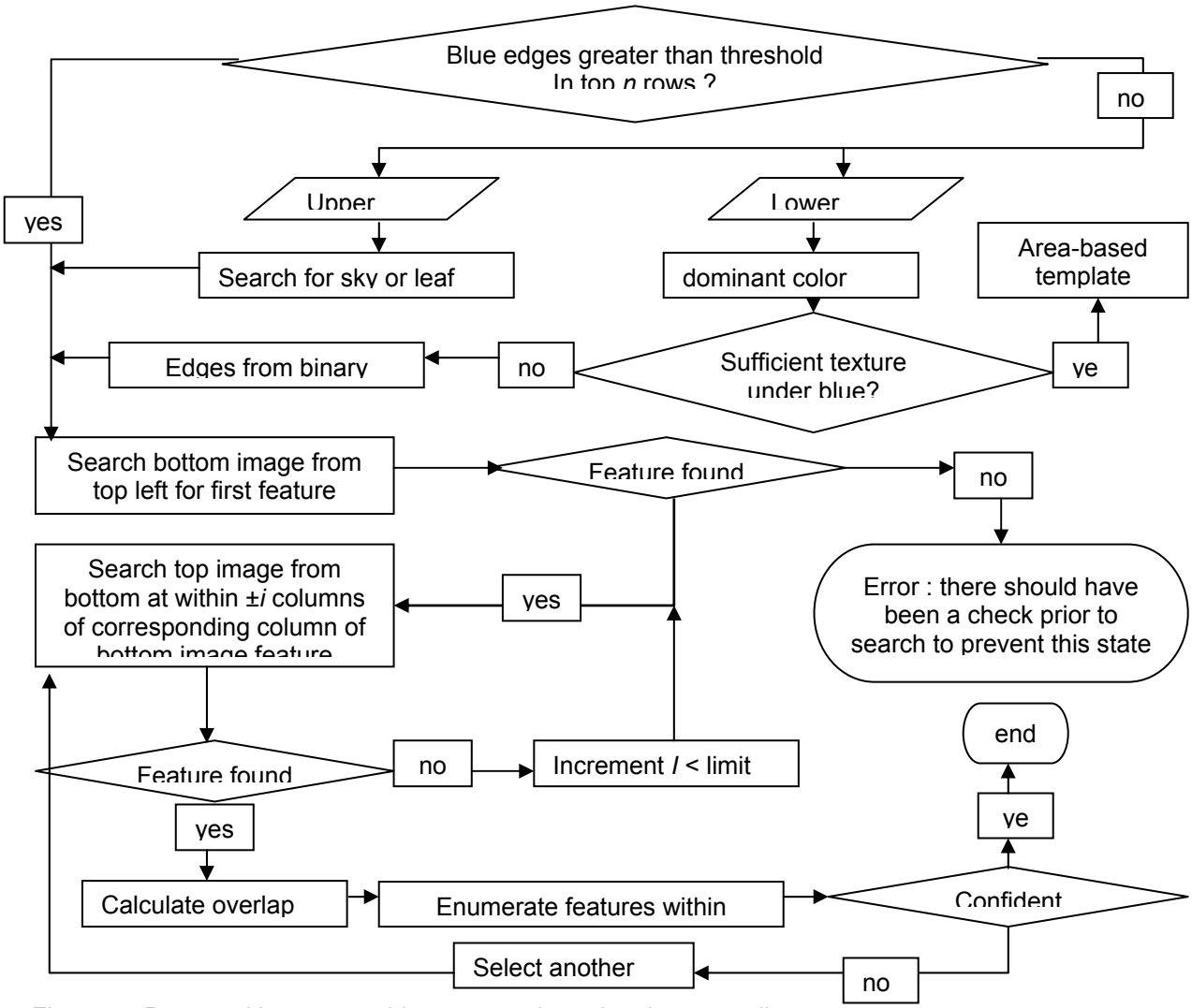

Figure 5. Proposed image matching strategy based on image attributes.

Corner pixels, edges, or regions will be progressively desired primitives. The topmost central feature from the bottom image will be searched first. Since the amount of overlap is unknown, this feature has a higher probability of being contained on the top image. The search space will be constrained initially by columns because horizontal translation should be minimal. The top image will be searched from the bottom within this columnar constraint. If no match is found, the columnar constraint will be eased. As each initial feature is located, the image overlap area

will be calculated and additional features contained within this region will be enumerated. The proportion of matched features to the total features examined will be used to determine if a confident match is obtained. If a match is not obtained, the feature may have escaped detection in the top image. In this case another feature from the bottom image is selected and the process is run again.

CONCLUSIONS

There is a compromise between small simple primitives that have a higher probability of repetition and large, complex primitives that may be time consuming to formulate. Primary interest is to find the most unique primitive so that there is only one singly dominant correspondence. The desire is that this primitive would also be small in dimension to reduce computational burden, and small in size to reduce the effects of geometric distortion. Features are more reliable with increasing size, dimension, and complexity, however the computational burden to derive these features can be large. A strategy is proposed here to start with the simplest, but unique primitives and progressing high-level features only as a last resort.

This paper has presented some of the properties of subcanopy images captured using a multisensor video system. Simple primitives are favored overall, but if no unique simple primitives are apparent additional processing must be done. Images are classified into upper and lower stem regions to guide matching strategy. A region-based approach is chosen as the preferred method for upper stem images due to the common presence of sky regions. Textural information is more available and more informative for matching of the lower images.

Of the many methods and algorithms available for image matching, there is no suitable general purpose solution for all imaging situations. This paper has demonstrated how image discrimination can be used to guide the approach for efficiency and reliability. This method will soon be implemented and evaluations can then be made and means of improvement explored.

LITERATURE CITED

Atkinson, K. 1996. Close Range Photogrammetry and Machine Vision. Whittles Publishing. Scotland, UK. 371 pp.

Clark, N., S. Zarnoch, A. Clark III and G. Reams. 2001. Comparison of standing volume estimates using optical dendrometers. In 2nd Annual FIA Symposium, Salt Lake City, UT November 2000. (In Press)

Heipke, C. 1996. Overview of image matching techniques. OEEPE Workshop on the Application of Digital Photogrammetric Workstations. March 4-6, Lausanne, Switzerland
< http://dgrwww.epfl.ch/PHOT/workshop/wks96/art_3_1.html >

Marapane, S.B., and M.M. Trivedi. 1994. Multi-primitive hierarchical (MPH) stereo analysis. In: IEEE Transactions on Pattern Analysis and Machine Intelligence. 16(3):227-240.

M. Okutomi and T. Kanade, "A Locally Adaptive Window for Signal Matching," International Journal of Computer Vision, vol. 7, no.2, pp.143-162, 1992.

Peleg, S., B. Rousso, A. Rav-acha, and A. Zomet. 2000. Mosaicing on adaptive manifolds. IEEE Transactions on Pattern Analysis and Machine Intelligence. 22(10):1144-1154.

Salama, Gouda Ismail Mohamed. 1999. Monocular and Binocular Visual Tracking. PhD Dissertation. Virginia Polytechnic Institue and State University. 196 pp. http://scholar.lib.vt.edu/theses/available/etd-0104100-232806/

Sonka, M., V. Hlavac, and R. Boyle. 1993. Image Processing, Analysis, and Machine Vision. Chapman & Hall, Inc. New York. P. 178 555 pp.

Proceedings of the

18th Biennial Workshop on Color Photography and Videography in Resource Assessment

Amherst, Massachusetts
May 16–18, 2001