

AUTOMATED MOSAICKING OF SUB-CANOPY VIDEO INCORPORATING ANCILLARY DATA

Eric Kee, Graduate Research Assistant
Bradley Department of Electrical and Computer Engineering
Virginia Polytechnic Institute and State University
Durham Hall, Blacksburg, Virginia 24061
ekee@vt.edu

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

A. Lynn Abbott, Associate Professor
Bradley Department of Electrical and Computer Engineering
Virginia Polytechnic Institute and State University
Durham Hall, Blacksburg, Virginia 24061
abbott@vt.edu

ABSTRACT

This work investigates the process of mosaicking overlapping video frames of individual tree stems in sub-canopy scenes captured with a portable multisensor instrument. The robust commercial computer vision systems that are in use today typically rely on precisely controlled conditions. Inconsistent lighting as well as image distortion caused by varying interior and exterior orientation parameters can complicate image mosaicking in a sub-canopy environment. This paper presents how the image, range, and orientation data are used to guide the mosaicking algorithm used in the Tree Measurement System (TMS). The mosaicked images will be an important step in data reduction leading to the continued development of a portable tree-stem sampling instrument.

INTRODUCTION

Mosaicking is a well-studied topic in image processing and has led to the creation of many different types of image mosaics. Mosaics can be generated from data collected of cylindrical and spherical pans and from data containing image scale changes (Rousso et al., 1998, Peleg and Herman, 1997, McMillan and Bishop, 1995). Many different approaches exist to mosaic images; primary mosaicking approaches entail block matching techniques or gradient descent (Shum and Szeliski, 1997, Haralick and Shapiro, 1993). Mosaics are often intended for visual consumption rather than precise object measurement. Systems that generate mosaics for visual appeal often perform destructive image operations such as blurring and warping to eliminate visual blemishes in the resulting mosaic (Rousso et al., 1997). Destructive image operations do not preserve detailed information about objects within a scene. This paper presents a mosaicking system that preserves information integrity in mosaics to support digital measurement of trees. Using a video camera equipped with laser range finder and triple-axis inclinometer, the presented mosaicking system allows for precise analysis of a tree's size and volume.

TMS INSTRUMENT AND FIELD PROCEDURE

The TMS instrument consists of a video camera, pulse laser-rangefinder, and internally mounted inclinometers that measure the instruments orientation in three axes. The video camera has a CCD array (charge-coupled device) type sensor producing a 720 x 480 pixel output image. The instrument has a custom lens system that has an effective focal length of 250 mm for a standard 35 mm film format camera. This results in a field of view of 5.5 degrees high by 8.24 degrees wide. Video and range-orientation (RO) data are captured into two separate sequential data streams onto a videocassette and memory card, respectively, and are later synchronized by TMS. Video data are collected at 30 frames per second and RO data are subsampled to 10 readings per second from 238 raw measurements per second. Additional details of the TMS can be found in (Clark et al. 2001).

The TMS instrument is mounted on a monopod for ease of carrying and stability during data collection. The monopod also restricts instrument rotation about the camera axis, which would increase the difficulty of the automated frame mosaicking process. To collect data, the unit is positioned where tree stem visibility is least impaired and at approximately the same distance from the tree stem as the greatest height being collected. Data collection begins with the instrument aimed at a point on the stem where the range is not obstructed, as this range will be used to validate subsequent range measurements. The instrument is then aligned with the bottom point of the stem (to create a base height for referencing subsequent height measurements) and slowly inclined up the stem until the highest desired point is reached. This defined collection procedure is required for efficient automated data processing by TMS.

The video and RO data streams must be synchronized by sequential correlation. Both data sources are sequentially recorded through time: RO data are stored in incrementally named data files (e.g., file001, file002, etc.) and video data are recorded on tape, which has an intrinsically defined recording rate and sequential order. TMS automates synchronization by sorting through the RO data and computing the times corresponding to desired overlapping video frames determined by orientation values. The frame times are then used to trigger a frame capturing function to extract corresponding frames from the video sequence.

THE MOSAICKING SYSTEM

TMS uses video recorded with the multisensor instrument to create mosaics of an entire tree stem from a single vantage point. To reduce the amount of data processing needed to create a mosaic, a **frame set** of pair wise overlapping frames is captured from the video sequence based on the imprecise orientation estimates from the inclinometers. Errors in the video/RO data synchronization procedure can compound the already noisy inclination measurements. TMS must assemble image pairs in a frame set into a coherent mosaic of overlapping images.

Most current mosaicking algorithms allow for a camera model with short focal length and large differences in object space distances. Mosaicking systems typically assume the mosaic is to be applied as the best match over the entire image; however, TMS need only match the tree stem located in the horizontal and vertical center of the frame. Perspective effects increase as focal length and scene depth decrease. (Scene depth is measured as the distance from the nearest to the farthest image element.) However, given that stem or crown structure depth varies little between frames compared to stem-camera distance, no significant image displacements occur between overlapping frames. Subtle image displacement between overlapping frames tremendously expedites mosaicking, as only translation is needed to find the correspondence between adjacent video frames. Typically, image transformations (e.g., affine, polynomial, perspective transformations) must be applied prior to correspondence to compensate for lens or orientation effects. Because transformation parameters are usually unknown, finding the best transformation may require much iteration—this will impose unacceptable computation time for TMS. While subtle inter-frame perspective changes justify translation assumptions, translation assumptions also support object measurement. Because TMS couples frames with orientation data, it is advantageous to preserve the integrity of image geometries.

TMS uses orientation data from the camera's sensors to determine a mosaic point, or **m-point**, that specifies a translation between images in a **mosaic pair**. If F_i and F_{i+1} , are two frames in a frame set $S = \{F_1, F_2, F_3, \dots, F_n\}$ extracted from a video sequence, an m-point— (x,y) —is a coordinate in F_i that achieves an approximate frame correspondence when the lower-left corner of frame F_{i+1} is positioned at location (x, y) in frame F_i . Figure 1 demonstrates a possible m-point for two sample images. Note that the m-point in Figure 1 does not properly align the depicted trees—we use the term m-point to refer to any possible alignment of a mosaic pair.

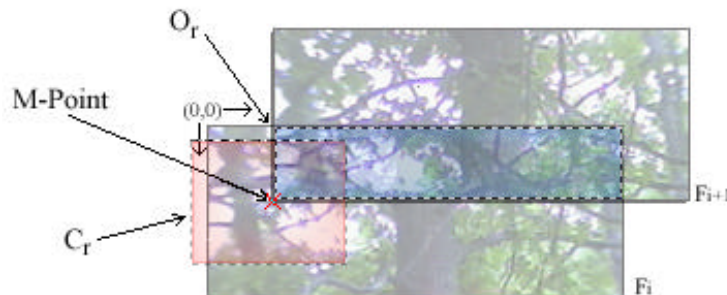


Figure 1: A possible m-point location within a mosaic pair. The red “x” marks a possible m-point. An m-point is determined by the location of the lower left corner of F_{i+1} in frame F_i . Region O_r (the overlap region) is shown in blue; region C_r (the confusion region) is shown in red.

To create an image mosaic, TMS must determine the **actual m-point**—the m-point that creates minimum visual discontinuities when two successive frames are mosaicked. Using the camera’s orientation sensors, a **guess m-point** is calculated which approximates the actual m-point. Because sensor measurements alone do not provide the precision necessary to create a coherent mosaic, the guess m-point must be refined.

The guess m-point between frames F_i and F_{i+1} is calculated using the change in azimuth and inclination between F_i and F_{i+1} from the instrument data using Equation 1.

$$\begin{bmatrix} x \\ \text{frameheight} - y \end{bmatrix} = 4994 \text{Tan} \begin{bmatrix} \Delta \text{Azim} \\ \Delta \text{Incl} \end{bmatrix} \quad (1)$$

In Equation 1, ΔAzim and ΔIncl are the changes in azimuth and inclination between two successive frames, 4994 is the focal length of the camera (in pixels), and $\text{frameheight} - y$ maps inclination offsets to image coordinates (see the image coordinate assumptions in Figure 1). Azimuth and inclination data are subject to instrument and system error. Let $[\Delta I, \Delta A]^T$ represent the inclination and azimuth error (in pixels) inherent in the camera and frame extraction system. The actual m-point lies within $[\pm \Delta I, \pm \Delta A]^T$ pixels of the guess m-point—we refer to the $2[\Delta I, \Delta A]^T$ pixel window that surrounds the guess m-point as the guess m-point’s **confusion region**, C_r (see Figure 1). Assuming no external errors are imposed upon the orientation data, the actual m-point must lie within the confusion region. The mosaicking system uses $[\Delta I, \Delta A]^T = [\pm 80, \pm 80]^T$ pixels to provide some fault tolerance; although, calculations using the manufacturer reported azimuth and inclination accuracies ($\pm 1^\circ$ and $\pm 0.4^\circ$, respectively) suggest that the confusion region should be $[\pm 34, \pm 87]^T$ pixels. Extra vertical fault tolerance is needed, as this is the primary motion direction increasing the probability for orientation-image frame correlation errors along this vector.

Once the guess m-point and confusion region have been computed, TMS must determine the location of the actual m-point within the confusion region. To find the actual m-point, TMS defines an energy function, $E(i,j)$, as a metric for the “goodness” a particular m-point. $E(i,j)$ defines the **energy space** of a mosaic pair—the space of all possible m-points and their energy values. If $E(i,j)$ is a perfect metric, the extreme value in the energy space corresponds to the actual m-point. In this work, we experiment with two energy functions:

$$E(i, j) = \sum_{\forall \text{ColorPlanes}} \frac{\sum_{\forall (x,y) \in C_r} |F_n(x', y') - F_{n+1}(x'', y'')|}{\|O_r\|} \quad (2)$$

$$E(i, j) = \sum_{\forall \text{ColorPlanes}} \frac{\sum_{\forall (x,y) \in C_r} [F_n(x', y') - \mathbf{m}_{F_n}] [F_{n+1}(x'', y'') - \mathbf{m}_{F_{n+1}}]}{\sqrt{\sum_{\forall (x,y) \in C_r} [F_n(x', y') - \mathbf{m}_{F_n}]^2} \sqrt{\sum_{\forall (x,y) \in C_r} [F_{n+1}(x'', y'') - \mathbf{m}_{F_{n+1}}]^2}} \quad (3)$$

Equation 2 is simple normalized SSD using absolute difference rather than squared difference to reduce computation time. We normalize the sum of absolute differences by the number of pixels in the overlap region (O_r) for a given m-point and sum over each color plane in F_n and F_{n+1} (Figure 1 illustrates an overlap region). Without normalization, m-points with large O_r inherently return higher energy values than m-points with small O_r . The normalized sum of absolute differences (NSAD) is zero for an optimum m-point; therefore, we minimize NSAD energy.

Equation 3 is the sum of normalized cross-covariance (NCCV) across all color planes. The extreme point in NCCV energy space corresponds to the maximum value of the NCCV; therefore, we maximize NCCV energy. In equations 2 and 3, the sum is computed for all pixels in the confusion region, C_r . Points, (x,y) , in the confusion region are translated into frame coordinates (x', y') and (x'', y'') for frames F_n and F_{n+1} , respectively. For a given m-point (x,y) , (x', y') and (x'', y'') are assumed to be corresponding pixels—pixels that lie “on top” of each other when F_i is mosaicked with F_{i+1} at m-point (x,y) .

Restricting computation of Equation 1 and 2 within the confusion region reduces data processing and eliminates erroneous values of Equations 1 and 2. Erroneous values of $E(i,j)$ occur when information content is low between a mosaic pair at a given m -point. Low information content for an m -point occurs when a mosaic pair does not contain significant image features, or when the overlap-region between frames is small. Cases of small overlap necessitate the use of orientation data to formulate a guess m -point. Without a guess m -point and confusion region, energy minimization functions must be sensitive to the size of the corresponding overlap region.

Energy computation is used in a 5-level coarse-to-fine refinement scheme, using a Gaussian image pyramid. Coarse-to-fine refinement provides two advantages in TMS: first, computational burden is reduced with smaller data sets; second, coarse-to-fine refinement allows the mosaicking system to discover actual m -points that may lie just beyond the confusion region. At the smallest level of a 5-level Gaussian image pyramid, the confusion region is reduced by a factor of 2^5 , or $[\pm 3, \pm 3]^T$ for our confusion region $([\pm 80, \pm 80]^T)$.

Figure 2 illustrates coarse-to-fine mosaicking. At each successive level of the pyramid, image resolution is one quarter the resolution of the preceding level (image width and height are halved as the pyramid level increases). The coarse-to-fine process begins at level 4 (P_4) of the Gaussian pyramid and ends at level 0 (P_0), the full resolution image. Once an m point, x , has been computed for some P_n , x is used to specify the guess m -point in P_{n-1} . Additionally, computation of x reduces the size of the confusion region in P_{n-1} . Because each pixel in P_n represents four pixels in P_{n-1} , the pyramid confusion region (PC_r) is ideally $[\pm 1, \pm 1]^T$ pixels in P_{n-1} . However, mosaicking errors in P_n necessitate fault tolerance—TMS uses $PC_r = [\pm 4, \pm 4]^T$ for P_{n-1} in coarse-to-fine mosaicking refinement. Experimentation with different size confusion regions suggests that $C_r = [\pm 80, \pm 80]^T$ and $PC_r = [\pm 4, \pm 4]^T$ offers the greatest reliability.

As coarse-to-fine mosaicking processes each P_i , m -point error is reduced. At each successive level of the pyramid, m -point precision increases by a factor of two. Precision optimally reaches ± 1 pixel after coarse-to-fine refinement; at ± 1 pixel precision, azimuth and inclination measures are precise to $\pm 0.023^\circ$.

Finally, TMS computes an m -point for every mosaic pair in a frame set and assembles a mosaic by taking unique strips of data from each frame. From the computed m -points, azimuth and inclination changes are associated with each section of the mosaic—this allows for tree measurement at any point within the mosaic image.

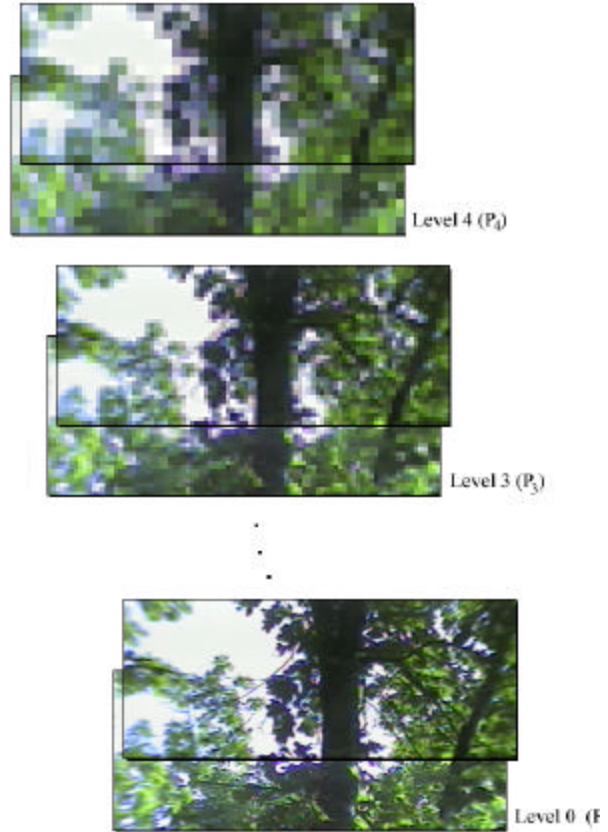


Figure 2: Coarse-to-fine mosaicking using a Gaussian image pyramid.

RESULTS

TMS mosaicking was tested on eight unique tree sequences (data set mt68). For each tree, two unique frame sets were extracted from the original multisensor data. Sixteen mosaics sampled at approximate 2° inclination intervals provided 571 mosaic pairs for TMS to mosaic. Experimentation found NCCV energy to be prohibitively time consuming. Because NCCV and NSAD energy produce comparable results across mt68, NSAD energy is optimal. The results presented below are produced using NSAD energy.

Figure 3 illustrates a typical mosaic generated by TMS using NSAD energy. Because images are not warped to account for scale changes, the upper portion of the tree is noticeably foreshortened. Figure 3 also illustrates TMS's performance on mosaic pairs whose frames differ in contrast. The mutisensor instrument adjusts camera light sensitivity as lighting levels change. Many frame sets in data set mt68 contain significant contrast changes; however, energy minimization was found to be insensitive to contrast effects. Contrast effects generated no error



Figure 3: Mosaic from frame set mt68-252.

conditions in mt68.

Table 1 demonstrates system reliability and provides quantitative analysis of mosaicking results. The actual m-points listed in Table 1 were derived by mosaicking images by hand. Single or double pixel errors found in Table 1 are considered insignificant. (When only small errors are present, it can be argued that the "actual" m-point is no more accurate than the computed m-point.)

Mosaic	Actual M-Point		TMS M-Point		Guess M-Point		Error Dist		Guess Error	
	X	Y	X	Y	X	Y	X	Y	X	Y
0-1	15	212	15	212	-34	210	0	0	49	2
2-3	-11	254	-11	254	17	210	0	0	28	44
4-5	7	252	7	252	34	210	0	0	27	42
6-7	-5	247	-5	247	8	219	0	0	13	28
8-9	-12	212	-12	212	-8	227	0	0	4	15
10-11	-6	249	-6	249	-8	245	0	0	2	4
12-13	2	258	2	258	17	227	0	0	15	31
14-15	-10	239	-10	239	26	219	0	0	36	20
16-17	-6	254	-6	254	0	210	0	0	6	44
18-19	-5	235	-5	235	-17	236	0	0	12	1
20-21	5	254	5	254	-17	245	0	0	22	9
22-23	0	222	0	221	0	236	0	1	0	14
24-25	2	227	2	227	26	227	0	0	24	0
26-27	6	255	7	256	26	219	-1	-1	20	36
28-29	8	256	8	256	8	227	0	0	0	29
30-31	8	242	8	242	-8	219	0	0	16	23
32-33	13	242	13	242	8	227	0	0	5	15
34-35	0	194	0	194	0	219	0	0	0	25
36-37	11	266	11	266	0	210	0	0	11	56
38-39	1	261	1	260	8	227	0	1	7	34
40-41	0	265	0	265	8	236	0	0	8	29
42-43	2	206	2	206	26	227	0	0	24	21
44-45	-4	266	-4	265	17	227	0	1	21	39
46-47	0	279	0	278	-8	227	0	1	8	52
48-49	2	227	2	227	8	227	0	0	6	0
50-51	7	173	7	173	34	166	0	0	27	7
52-53	7	277	7	277	8	227	0	0	1	50
54-55	6	173	6	172	0	245	0	1	6	72
56-57	-19	206	-19	206	17	236	0	0	36	30
58-59	-13	180	-13	180	8	210	0	0	21	30
60-61	19	203	19	204	26	245	0	-1	7	42
62-63	-5	239	-5	239	26	184	0	0	31	55
64-65	-1	199	-1	198	26	210	0	1	27	11
66-67	14	244	14	244	34	236	0	0	20	8
68-69	16	177	16	177	8	227	0	0	8	50
70-71	1	237	1	236	8	245	0	1	7	8
72-73	15	132	15	133	8	166	0	-1	7	34
74-75	2	288	2	288	34	219	0	0	32	69
76-77	3	287	3	286	17	245	0	1	14	42
78-79	-1	218	-1	218	-26	227	0	0	25	9
80-81	77	222	77	222	-52	227	0	0	129	5
82-83	47	207	47	208	-17	219	0	-1	64	12
84-85	4	265	4	266	17	219	0	-1	13	46
86-87	31	195	-108	251	-8	210	139	-56	39	15

Table 1: Quantitative mosaicking results for frameset mt68-0000. All values are in pixels.

After mosaicking the individual frames, inclination and azimuth data are significantly more accurate—TMS is typically accurate to ± 2 pixels of the actual m-point. The error distance column in Table 1 demonstrates system reliability; across 571 mosaics, there were nine error cases (Table 1 is a single tree within dataset mt68). However, system reliability is higher than 98.4% -- only 6 error cases were detrimental to image measurement. Table 1 contains a detrimental error condition; the computed m-point is 150 pixels disparate. Figure 4 demonstrates non-detrimental error conditions. Non-detrimental error conditions will not affect tree measurement precision significantly.



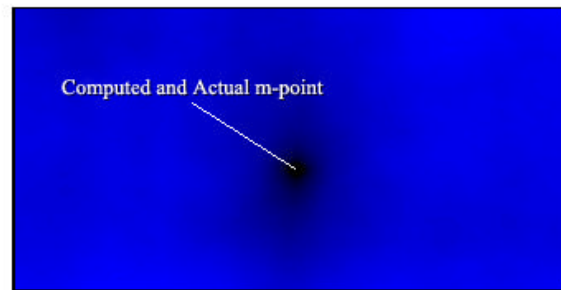
4: A mosaic from frame set mt68-172 with a typical mosaicking error.

Errors in the mosaicking system result from bad orientation data or information ambiguities within a mosaic pair. If orientation data is poorly correlated to the frame set, actual mosaic points will lie outside of the confusion region and will not be examined as a possible solution. When orientation data is poorly correlated, computed m-points assume the best value within the confusion region—this value is often far from correct.

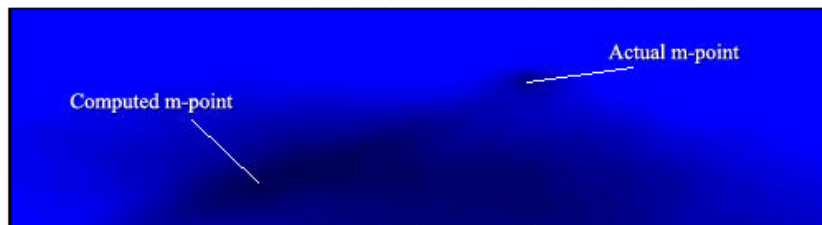
If ambiguities exist in the mosaic pair, the information contained in the two images does not provide a definite actual m-point. Some mosaic pairs are ambiguous even when inspected by hand—these cases will generate errors in any computer vision system. Figure 5 shows two energy spaces: the first energy space contains a well-defined minimum that specifies an accurate actual m-point; the second confusion region contains an ambiguous trough that is associated with diagonal ambiguity found in mosaic pair 86-87 in Table 1. Bright blue indicates areas of high NSAD energy; dark blue indicates areas of low NSAD energy.

Examination of mosaic pair mt68-086-087 reveals the source of information ambiguity—a large diagonal branch is the only significant source of information. Figure 6 shows mosaic pair mt68-086-087 and the source of diagonal ambiguity.

Examination of mosaic pair mt68-086-087 reveals the source of information ambiguity—a large diagonal branch is the only significant source of information. Figure 6 shows mosaic pair mt68-086-087 and the source of diagonal ambiguity.



Energy Space #1:
Well-defined energy space (mosaic pair 00-01)



Energy space #2:
Ambiguous energy space (mosaic pair 86-87)

Figure 5: Energy spaces generated from mosaic pairs in data set mt68. Dark blue indicates areas of low NSAD energy.

Step 2 in Figure 6 illustrates information ambiguity—step 2 shows the alignment of mosaic pair mt68-86-87 at the computed mosaic point. Note that the large diagonal branch is properly aligned between F_{86} and F_{87} . Closer inspection of the mosaic in step 2 reveals subtle clues that the computed mosaic point is incorrect. However, the features that suggest that F_{86} and F_{87} have been improperly

mosaicked are not strong enough to overcome the noise generated by the ambiguous canopy in the mosaic pair.

Figure 7 demonstrates proper alignment of F_{86} and F_{87} at the actual mosaic point in NSAD space. Ambiguous mosaic pairs are typically a product of the energy function (few, if any, mosaic pairs exist that cannot be properly aligned by the human visual system).

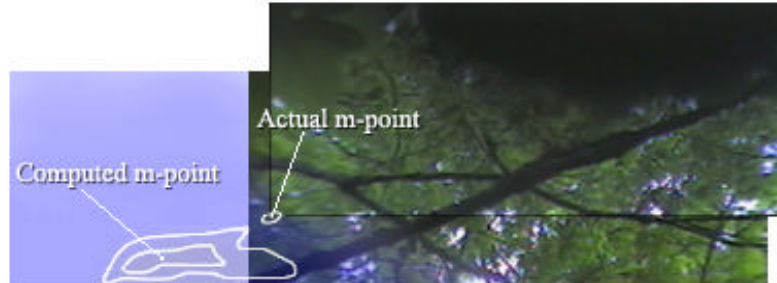


Figure 7: Proper alignment of mosaic pair mt68-86-87. Contour lines are shown in white to demonstrate the shape of the energy space computed by TMS.

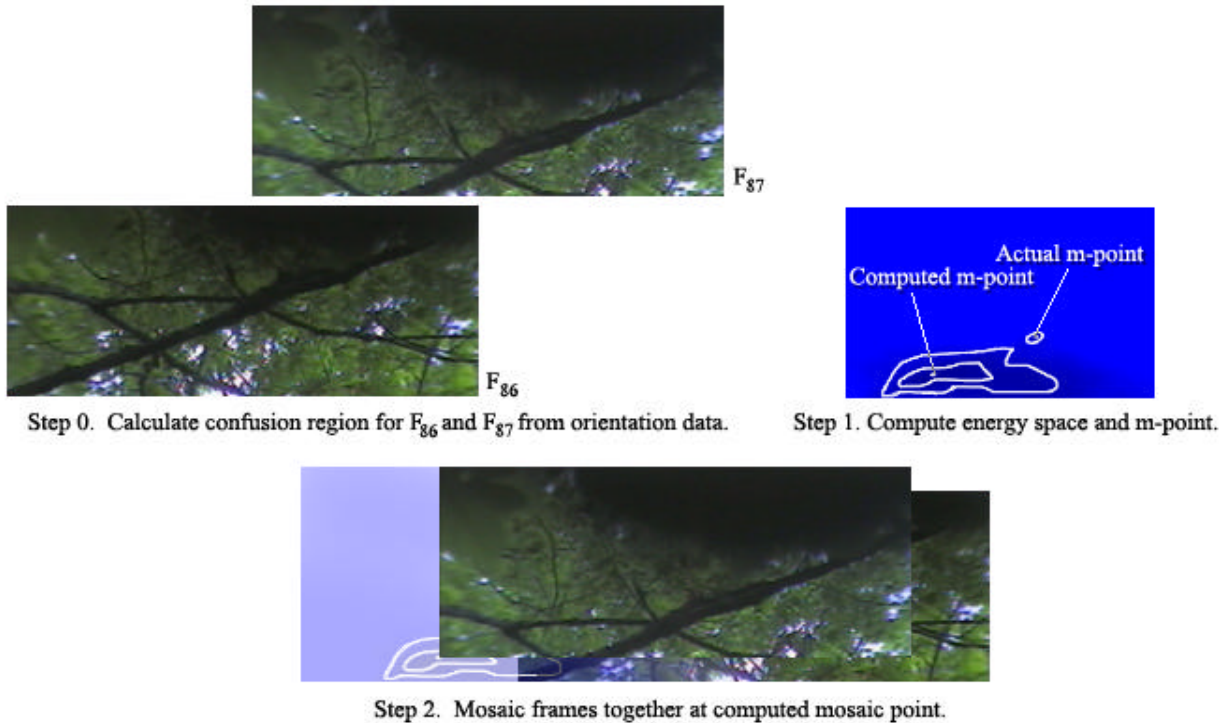


Figure 6: The steps in computation of m-point for mosaic pair mt68-086-087. In Step 0, confusion region for frames 86 and 87 is computed. In Step 1, the energy space within the confusion region is computed (white contour lines are drawn to illustrate the shape of the energy space). Finally, in Step 2, the mosaic pair is aligned at the computed m-point (contour lines are shown to demonstrate the correspondence between the energy space and the mosaic point).

Figures 8-11 present a sampling of other mosaicked trees produced by TMS. Cases of m-point ambiguity are rare within the mt68 data set. The most significant errors are a result of poorly correlated orientation and frame data. If an orientation frame is incorrectly correlated with an image frame, the guess m-point may not place the actual m-point within the confusion region—in these cases, TMS typically cannot recover from the error.



3: Mosaic from frame set mt68-500-588.



Figure 9: Mosaic from frame set mt68-091-171.



Figure 10: Mosaic from frame set mt68-590-627.



Figure 11: Mosaic from set mt68-328

FUTURE WORK

Future work is needed to increase fault tolerance in the mosaicking system. Errors produced by ambiguous energy spaces and erroneous orientation data must be resolved if they create highly discontinuous mpoints. Ambiguous energy spaces may call for more robust energy metrics or alternate mosaicking techniques such as block matching; however, it is primarily difficult to determine that a mosaic point is ambiguous. Future fault tolerance work must also design confidence metrics for mosaic pairs so that TMS may decide when a mosaic pair is suspect of error.

CONCLUSIONS

This paper presents a mosaicking system that uses sensor-based orientation estimates to constrain an energy minimization mosaicking strategy. Through coarse-to-fine processing, TMS refines orientation estimates that facilitate image-based tree measurement. Two common image matching techniques, NSAD and NCCV, are investigated for their mosaicking utility. NSAD energy minimization is shown to be optimum for both its integrity and its efficiency; NCCV energy minimization is more complex and less reliable than NSAD energy minimization. Failures regarding image ambiguity demonstrate faults in the energy minimization technique for image mosaicking.

Image mosaicks can reduce hundreds of video frames in to a single still image that supports accurate tree-stem measurement.

REFERENCES

- Clark, N., S. Zarnoch, A. Clark, and G. Reams. 2001. Comparison of Standing Volume Estimates Using Optical Dendrometers. In: Proceedings of the second annual Forest Inventory and Analysis symposium; Reams, Gregory A.; McRoberts, Ronald E.; Van Deusen, Paul C., eds.; 2000 October 17-18; Salt Lake City, UT. Gen. Tech. Rep. SRS-47. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 123-128.
- Haralick, R.M. and L. G. Shapiro. 1993. *Computer and Robot Vision, Vol. II*. Addison-Wesley Publishing Co. 1993, pp. 289-367.
- McMillan, L. and G. Bishop. 1995. Plenoptic modeling: An image based rendering system. In: SIGGRAPH, Los Angeles, California, August 1995. ACM.
- Peleg, S. and J. Herman. 1997. Panoramic mosaics by manifold projection. In: IEEE Conf. On Computer Vision and Pattern Recognition, pp. 338-343, June 1997.
- Rousso, B., S. Peleg, and I. Finci. 1997. Mosaicing with generalized strips. In: DARPA Image Understanding Workshop, pp. 255-260, May 1997.
- Rousso, B., S. Peleg, I. Finci, and A. Rav-Acha. 1998. Universal mosaicing using pipe projection. In: Int. Conf. On Computer Vision, pp. 945-952, January 1998.
- Shum, H. and R. Szeliski. 1997. Panoramic Image Mosaics. Microsoft Research Technical Report MSR-TR-97-23. 50 pp. [http://www.research.microsoft.com/scripts/pubs/view.asp?TR_ID=MSR-TR-97-23].