2009-12-07

Fused Visible and Infrared Video for use in Wilderness Search and Rescue

Dennis Eggett

Michael A. Goodrich
mike@cs.byu.edu

Bryan S. Morse
morse@byu.edu

Nathan Rasmussen

Original Publication Citation

N. Rasmussen, B. Morse, M. Goodrich, and D. Eggett, Fused Visible and Infrared Video for Use in Wilderness Search and Rescue, Proceedings of the IEEE Workshop on Applications of Computer Vision (WACV) , 29.

Fused Visible and Infrared Video for use in Wilderness Search and Rescue

Nathan D. Rasmussen[1], Bryan S. Morse[1], Michael A. Goodrich[1], and Dennis Eggett[2]
Brigham Young University
[1]Department of Computer Science
[2]Department of Statistics

Abstract

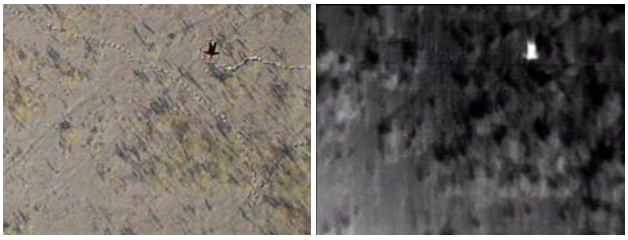
Mini Unmanned Aerial Vehicles (mUAVs) have the potential to assist Wilderness Search and Rescue groups by providing a bird's eye view of the search area. This paper proposes a method for augmenting visible-spectrum searching with infrared sensing in order to make use of thermal search clues. It details a method for combining the color and heat information from these two modalities into a single fused display to reduce needed screen space for remote field use. To align the video frames for fusion, a method for simultaneously pre-calibrating the intrinsic and extrinsic parameters of the cameras and their mount using a single multi-spectral calibration rig is also presented. A user study conducted to validate the proposed image fusion methods showed no reduction in performance when detecting various objects of interest in this single-screen fused display compared to side-by-side viewing. Most significantly, the users' increased performance on a simultaneous auditory task showed that their cognitive load was reduced when using the fused display.

1. Introduction

Search and Rescue teams throughout the Western United States are often called out to assist individuals who find themselves lost or in peril. These teams often utilize pilots to assist them from airplanes or helicopters, giving them a bird's eye view of the area. These aerial searchers speed up the search by covering large areas in a short period of time.

Aerial searching, however, comes with a number of disadvantages. Conventional aircraft are very costly to purchase and operate and pose potential danger to the pilot and crew members. In 2006 there was an incident in Utah where a sheriff's deputy was killed during a search when the helicopter he was flying in hit power lines and crashed [1].

In recent years, there have been great advances in the use of Unmanned Aerial Vehicles (UAVs) to obtain video from the air. This aerial video can replace aerial searchers and reduce the costs and dangers of searching from the air.



(a) Visible (b) Infrared

Figure 1. Aligned and synchronized frames from Visible and IR cameras showing a person lying on the ground.

Mini Unmanned Aerial Vehicles (mUAVs) have also been used because they bring additional advantages. mUAVs can be launched on location in many terrains and cost far less to purchase and operate, making them more attainable for Search and Rescue groups that have limited funding. Throughout the remainder of the paper, when we use the term UAV we are specifically referring to mini UAVs.

UAVs are capable of carrying a variety of sensors. The most common sensor carried by these is a visible-spectrum color camera, which we will refer to as a "Visible camera". These cameras provide views similar to those that pilots can obtain in manned aircraft.

Another sensor that has potential to be very helpful in Wilderness Search and Rescue is an Infrared (IR) camera. This type of camera increases the information that an aerial searcher can obtain by being able to see the heat our bodies produce. IR cameras are most useful during times or in areas when emitted body heat is greater than heat emitted from the surroundings, such as during the nighttime, in the early morning hours, over snow covered ground, or over water. In many of these conditions, IR cameras could be used in conjunction with Visible cameras to maximize the information that the searchers receive.

Searching with both IR and Visible cameras presents a great deal of information to users all at once. Figure 1 shows an example of the information from aligned Visible and IR frames. Each frame, from both cameras, has the potential of containing new information to be analyzed, yet

978-1-4244-5498-3/09/$25.00 ©2009 IEEE

the viewer only has a fraction of a second to do so. Significant screen real estate is needed to display both videos simultaneously, making a small portable system for use in Wilderness Search and Rescue difficult to build and use.

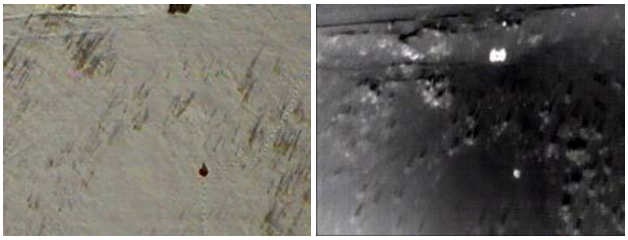
A UAV equipped with IR and Visible cameras has the potential to be an excellent asset to a Search and Rescue team. This paper presents a novel method for overcoming the increased information and screen space needed to use these two cameras. This is done by combining the information from the two cameras into a single view, retaining enough information from both image modalities so that searchers are able to successfully find people and objects of interest. To be capable of combining the information we have also developed a new method for calibrating the intrinsic and extrinsic parameters of our multi-modal cameras and their shared mount using a multi-spectral calibration rig.

2. Related Work

The area of aerial search and surveillance using UAVs has seen an expansion in recent years due to their increasing availability (for example [2–4]). This expansion has opened the way for different areas of research using aerial platforms for obtaining video.

Aligning and interpreting multimodal images or video streams is not a new concept (e.g., [5]). Some have used manually-selected points to align frames from uncalibrated sequences from different modalities [6, 7], but this requires manual intervention. Silhouette extraction [8, 9] has been used for IR and Visible alignment in systems where a person is walking across the scene, but this requires a specific object that can be singled out in both modalities and only aligns the extracted object rather than the entire frame. Mutual information [10–12] has been shown to be an effective method for automatically aligning sequences from uncalibrated cameras of different modalities when images have enough statistical correlation, but it requires significant computation. Another approach is to exploit the temporal correlation between different video modalities rather than spatial correlation [13, for example], but as with mutual information approaches this can be computationally expensive and may not be feasible for real-time application.

The approach presented here uses automatic precalibration of the cameras prior to video acquisition, thus requiring relatively little computational overhead to process continuous real-time video sequences. Automatic calibration of two cameras for alignment of their videos has been done for stereo reconstruction [14,15]; however, these methods don't often work well when working with different image modalities. Multimodal markers have been employed in medical imaging to align imaging with different modalities [16, 17] but can only be used when the markers can be placed in advance. While we cannot set up markers on the ground in all areas where searches will be performed, we can use external multimodal markers to pre-calibrate the camera mount and align the images.



(a) Visible (b) Infrared

Figure 2. An example of the raw video from the cameras. The frames are synchronized, but not aligned.

Once the images have been aligned they can then be interpreted using joint information. Many researchers work with IR and Visible imagery to make decisions on whether there is fire [18], or whether a person is in view [8, 19]. These methods look at each sensor independently and use the correlation of objects in both frames to make decisions. In Wilderness Search and Rescue there are many objects we look for that may not show up simultaneously in both modalities, so a method to present both forms of this information to the user is needed.

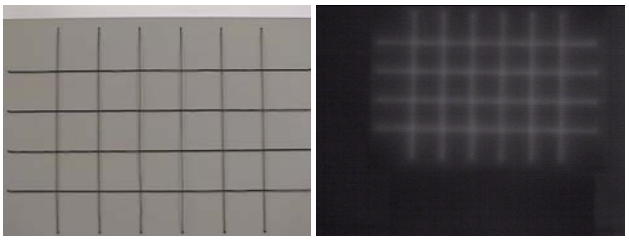
Many have worked in the area of image fusion, visually fusing information from two or more modalities into a single view [20–23, among many others]. Most of this work has used greyscale sensors, which allow them to find features from the different images to include in their output image, blending them to create other greyscale images (e.g., [20]) or to create color visualizations (e.g., [21]). In Wilderness Search and Rescue, the color information obtained from a Visible camera is important and cannot be combined in the same way that greyscale images can. Rather than a single channel of information from each sensor, we have multiple channels from the Visible camera that needs to be looked at jointly to retain the color information.

3. Methods

To combine video from IR and Visible cameras, the individual cameras as well as their shared mount need to be calibrated. Once the calibration has been performed, the frames can be warped into alignment and combined into a single image. Figure 2 shows sample frames before any alignment has been performed.

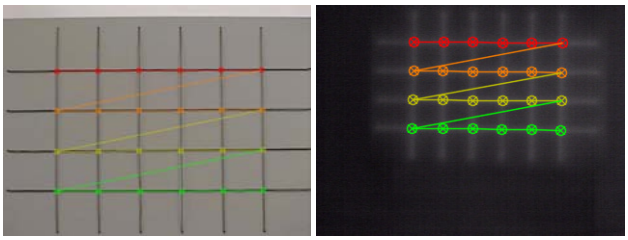
3.1. Image Alignment

To align the IR and Visible imagery, we need to first calibrate each of the cameras, followed by calibrating the mount that holds the cameras together. These could be done separately; however, we have developed a method that can do both of these in a sequential fashion using the same data. To do this, a set of objects need to be chosen that can



(a) Visible (b) Infrared

Figure 3. The multi-spectral calibration rig.



(a) Visible (b) Infrared

Figure 4. Located grid points after applying refinement.

provide both internal camera calibration as well as external mount calibration. To be useful, these objects need to be detectable in both the Visible and Infrared spectra. The camera calibration requires a pattern with known distances, and the rig calibration requires some way of pairing corresponding points from one image to the other. To satisfy all of these constraints simultaneously we use a grid pattern of wires (Figure 3) that have a small electrical current running through them so that they warm up slightly and emit heat. From this grid, corners are extracted where the wires meet, and then these points are used to calibrate the intrinsic and extrinsic parameters of the cameras and their mount.

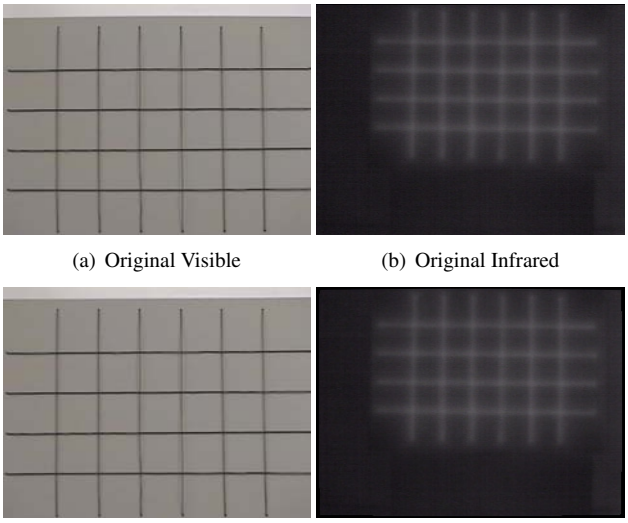
3.1.1 Point Extraction

Extracting the points from the grid of lines (Figure 3) is done by first using the Hough transform [24] to identify possible lines in the image. Intersecting lines are checked to verify that the intersection has at least a 45° angle between them (to remove near parallel intersections), and these lines are intersected to produce possible points. These points are then fit to the grid pattern we are trying to recover, and checks are performed to verify that the recovered pattern is a grid. The user can easily reject any bad set of points that may make it through all of the checks.

Once the grid is located, the points are refined using the lines that define them. A small window is extracted around each point to perform the refinement. We cannot use the entire line, as this would remove any distortions due to the lens and drastically change or skew our calibration results. By using a local area the distortions remain intact and can be correctly discovered and removed in the camera calibration.

To refine the points, we first estimate the equation of each line, then determine the intersection of these lines. Weighted linear regression is used to refine the line parameters for each line to reduce inaccuracies in the corner locations introduced by the Hough transform as well as from the limited resolution and sensitivity of the cameras (as shown in the fuzziness of the lines in Figure 3).

Since two separate lines are present in the small window around the intersection, not all of the points can be used in the refinement. A small area around each line can be used,



(a) Original Visible (b) Original Infrared

(c) Undistorted Visible (d) Undistorted Infrared

Figure 5. The same lines as shown in Figure 3 before and after intrinsic calibration and application of the distortion parameters. It is easiest to notice in the Visible images that the lines are straighter after performing the calibration and removing the lens distortions.

but to do so requires an approximation for the line. Due to the grid nature of the points, an approximation of the line can be found by using the neighboring intersections. An E-M [25] style iteration is then used in which the line approximation is updated and reapplied to the data to remove the biasing due to the original line approximation. The refinement provides much more accurate point locations and produces good results in the calibration. Figure 4 shows the results of finding and refining the point locations.

3.1.2 Intrinsic Camera Calibration

To calibrate each camera, the grid points from each acquired image are used. The Bouguet method for camera calibration is then applied to the image points [26, 27, with an implementation available in OpenCV and for Matlab]. This method gives both the intrinsic camera parameters as well as the distortion coefficients that correct for lens distortions. Figure 5 shows the lines after camera calibration has been performed and the distortions removed.

3.1.3 Extrinsic Camera Calibration

To calibrate the extrinsic parameters of the camera mount, we use the Visible/IR point correspondences to determine the homography between the calibration images, then factor this to determine the translation and rotation between the two cameras. Since we typically fly between 60 and 100 meters above the ground in order to provide for UAV safety while still maintaining sufficient resolution of the potential search area, the translation between the two cameras (only a few inches) becomes negligible, and the homography reduces to only the rotation component.

To find the homography, multiple image pairs are used and the results are averaged. The homography from each image pair is taken separately due to changes in the orientation of the grid plane that modify the plane normal and resulting translational component of the homography. The set of point pairs p_i and q_i from the Visible and IR images respectively are brought into a similar reference frame by applying their respective calibrations matrices K_1 and K_2:

$$p_i' = K_1^{-1} p_i \tag{1}$$

$$q_p' = K_2^{-1} q_i \tag{2}$$

The points p_i' and q_i' are then used to calculate the homography H such that $p_i' = H q_i'$ using the 8 point algorithm [28]. H is then decomposed into rotation R, translation $\frac{1}{d}T$, and plane normal N as in [29]:

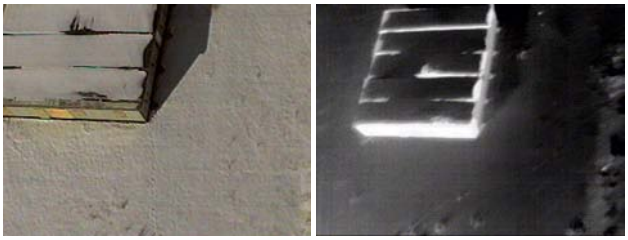
$$H = R + \frac{1}{d} T^T N \tag{3}$$

This decomposition produces four possible solutions, but only one is physically correct. Two of the solutions can be immediately removed since the z component of their plane normals is negative, orienting the plane facing away from the cameras. We remove the final ambiguity by again looking at the z component of the plane normal N and choosing the solution with the largest z component, since the plane on which the points lie is always close to perpendicular to our cameras.

Once each rotation has been isolated, the set of rotations are averaged together to get the final rotation estimate. This rotation then needs to be brought back into the image space, which will add in any scale and possible translational differences due to the cameras. This is done by applying the calibration matrices (K_1 and K_2) from the cameras to get the warp W that aligns the IR frame to the Visible frame:

$$W = K_1 R K_2^{-1} \tag{4}$$

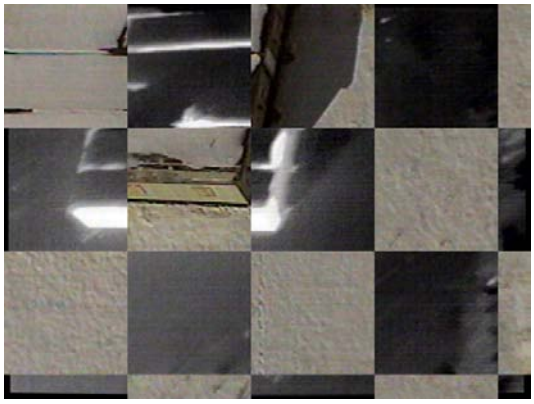
The results of the extrinsic camera calibration are shown in Figure 6 and 7. The misalignments are easy to see in the uncalibrated images of both figures but have been removed through the calibration methods as shown in the calibrated



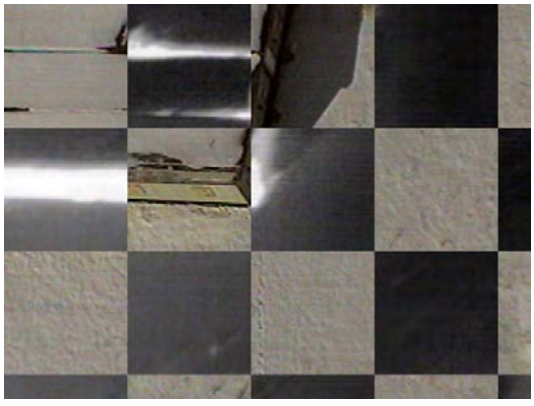
(a) Uncalibrated Visible (b) Uncalibrated IR

(c) Calibrated Visible (d) Calibrated IR

Figure 6. Sample image pair before and after calibration.



(a) Uncalibrated



(b) Calibrated

Figure 7. Checkerboard composite examples of the calibration using the same image frames from Figure 6. These images show both the IR and Visible images displayed in a checkerboard fashion on top of each other to show the alignment. (a) shows the alignment before calibration and (b) shows the alignment after calibration.

(a) Visible (b) IR (c) Fused

Figure 8. Example of IR and Visible image fusion. This result uses 50% transparency.

images. In Figure 7, the Visible and Infrared images are displayed in a checkerboard fashion, showing how well the edges of the building are aligned in the two images.

3.2. Image Fusion

Combining the IR and Visible imagery into a single image has the potential for greatly reducing the load on the user watching the video, as well as saving available screen space for field use. Our method involves highlighting objects (or areas) in the Visible frame where hot objects are found in the Infrared frame by creating a heat overlay to place over the Visible image. This is done by taking advantage of colors that are not common in the wilderness areas we image and also of the temporal dimension of the video.

To perform this highlighting we start with images that are aligned using the calibration methods described in Section 3.1. We create an overlay image $O(x, y)$ containing the heat information to be shown to the user. An HSV value is assigned to our overlay depending on whether the value of the IR image $I(x, y)$ is above a specified threshold t:

$$O(x,y) = \begin{cases} \mathit{HSV}(H, I(x,y), I(x,y)) & \text{if } I(x,y) > t \\ \mathit{HSV}(0,0,0) & \text{otherwise} \end{cases} \quad (5)$$

When the IR value is above the threshold we give the overlay image a pre-selected hue H and the IR value $I(x, y)$ for both its saturation and value. This allows the color to be brighter or darker when the object is hotter or colder. For all of our video we picked a hue of 300° on a 360° color wheel, giving us a magenta color. This is a rare color in wilderness areas and it helps give the user a sense of heat at the same time. Any hue can be used that gives the searcher an understanding of what is hot and is atypical in the target search area. We use a threshold of 150 for segmenting the heat information in the infrared video, though this can be adjusted by the user. The effect of the threshold on detection rates is left as an area of future work.

After the overlay is created we combine it with the Visible image $V(x, y)$ to create our fused image $F(x, y)$. The threshold t from Equation 5 is used again. A user-controllable opacity α is used to overlay $O(x, y)$ on $V(x, y)$ to produce the fused image $F(x, y)$:

$$F(x,y) = \begin{cases} V(x,y)(1-\alpha) + O(x,y)\alpha & \text{if } I(x,y) > t \\ V(x,y) & \text{otherwise} \end{cases} \quad (6)$$

The opacity allows the user to control to what extent they see the original Visible image vs. the thermal overlay, tailoring the output image to the user's preferences.

The colored overlay has the potential of either reducing the understanding of the original Visible image due to the added color or to be barely noticeable. To compensate for this, the temporal nature of video is leveraged by turning the overlay on and off at a user-specified rate (since anecdotal evidence suggests that users find different rates to be effective), thus interleaving a number of original Visible frames with the new fused frames. This allows the original information to be visible for a few frames and then the heat information to also be available (overlaid) for a few frames. This also attracts the visual attention [30] of the user.

Figure 8 shows an example of this method. We can see the original Visible and Infrared frames that were used as well as the resulting fused frame. This image was created using 50% opacity, allowing the original Visible frame to still be seen through the overlay.

4. Results

All of the video and images in this paper were obtained using a KX141 color camera from Black Widow AV [31] and FLIR's Pathfind IR infrared camera [32]. The Visible video was captured at 640×480, and the Infrared video was captured at 320×240. The aerial video was obtained by

(a) Mini UAV (b) Cameras and Mount

Figure 9. The hardware used to obtain the aerial imagery.

Display Method	Least Squares Means (Std. Error) of False Positives
Combined	1.85(0.30)
Side-By-Side	2.44(0.30)

Table 1. False positive results from the user study. The difference has a p value of 0.1516. While this is not extremely significant it was the next closest significant item in all of the statistics run against the primary task.

Display Method	Least Squares Means (Std. Error) of Miscounted Tones
Combined	0.91(0.18)
Side-By-Side	1.35(0.18)

Table 2. Miscounted tones from the user study. The users more accurately reported the number of tones played when watching the Combined video display compared with the Side-By-Side video display. The difference has a p value of 0.0417.

flying the cameras on a five-foot flying-wing remote control plane (Figure 9a) outfitted with the Procerus Technologies Kestrel Autopilot system [33].

A simple camera mount (Figure 9b) was built to retain the camera configuration, and the cameras and mount were calibrated using the methods from Section 3.1.

Figure 10 shows a sequence of frames using our fusion methods including the Visible frame, the aligned Infrared frame, and our fused frame. The frames show the image of a person lying on the ground passing though the frames. Note that even though frame slipping (temporal misalignment) due to independent capture devices for the two streams can cause slight misalignment, the indication of heat near a person is enough to draw a user's attention to them.

To validate our image fusion method, we conducted a user study to compare performance on a detection task given Side-By-Side Visible and Infrared videos compared with the fused video (which was labeled "Combined" in the study). To test the subjects' cognitive load while performing this task, a secondary task was added, in which they were asked to count the number of tones that were played through headphones. This secondary task had a low- and a high-difficulty setting to adjust the cognitive load placed on the user. In the low-difficulty setting, subjects were asked to count how many times a single tone was played. In the high-difficulty setting, they were asked to count two different tones and keep track of how many times each was played. This secondary audio task loosely simulates what a searcher may need to do as they interact with others who control the plane or need information about objects found while still searching through new video. Performance on this secondary task reflects the amount of mental workload required in the primary task; high performance means low workload and vice versa.

The user study involved a set of 32 volunteer subjects comprised of male and female college-age students who were told only that they were helping to evaluate two potential display methods for search and rescue. Eight videos, each approximately one minute in length, were watched by each subject. Collectively these contained 15 objects: eight people lying on the ground and seven red circles. The videos were taken during the winter, and the objects were placed to try to require using information from both the IR and Visible videos for detection. The video order, display methods, and secondary tasks were randomized to minimize ordering effects, and subjects saw each combination of display method and secondary task twice. We analyzed the data using mixed models ANOVA with subjects as a random blocking factor. All other variables were fixed effects and their differences were evaluated. The results are presented as Least Squared Means.

On the primary task, subjects were able to detect 90% of the objects without any significant differences due to the display method or the secondary task being performed. The relative difficulty of the individual videos was the only statistically significant ($p < 0.0001$) factor that affected the performance on this task. There was a slight statistical trend ($p = 0.15$) shown in the number of false positives subjects detected depending on the display method, where there were approximately 25% fewer false positives using the Combined display (Table 1).

On the secondary task, a clear improvement came when using the Combined display. Subjects reported the number of tones played with approximately 33% better accuracy ($p = 0.04$) while viewing the Combined display than while viewing the Side-By-Side display (Table 2). This strongly suggests a reduced cognitive load when working with the Combined display.

The subjective feedback confirmed the findings of the primary and secondary task analysis. On the preference questions asking which display was preferred for the detection task and which display would be preferred for use in a real search and rescue task, the responses were mixed, confirming the primary analysis that showed similar detection performance with either the Combined or Side-By-Side display. On the question asking which display was easier to watch, more subjects felt that the Combined display was easier, confirming our secondary task findings that the cognitive load was less with this display.

5. Conclusions

Using our calibration and fusion methods we are able to create a fused display that allows users to do just as well as they can with a simple side-by-side display, while requiring less cognitive effort on the part of the users and less screen space. This has great potential for assisting searchers when using both of these imaging modalities simultaneously.

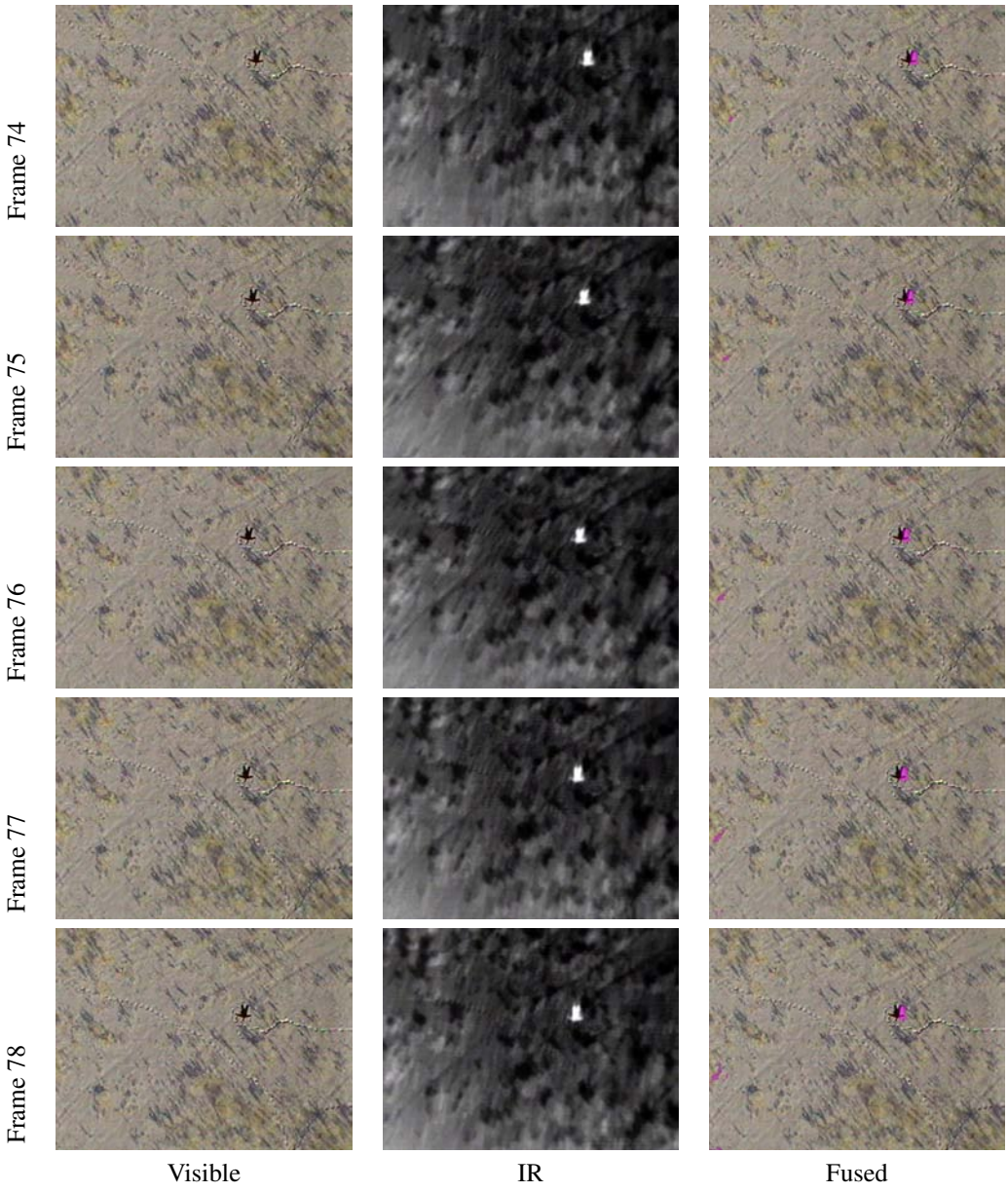


Figure 10. An image sequence showing a person lying on the ground.

The multi-modal video alignment developed here to align the frames for image fusion opens up a number of different areas that can be pursued with both IR and Visible cameras. Infrared video mosaics can easily be created using frame alignment information from the Visible video. With this same video alignment, filtering and super-resolution could be performed on the Infrared imagery to provide better information to the user.

The fusion methods we developed could be enhanced. Due to the thresholding of the infrared video, information is lost that could be detectable when looking at the separate videos. This might be mitigated by using adaptive thresholding or a segmentation method suitable for Infrared imagery. Our fusion method has not been tested in all settings. In our user study and in all of the aerial images presented in this paper we used imagery of winter scenes with snow on the ground. The fusion methods need to be tested during the summer, and some adaptation of the thresholding may need to take place to correctly segment a person's heat vs. the heat of other objects. Fatigue levels would also be interesting to test using our fusion methods compared with using side-by-side videos, though with the decreased cognitive load we could expect that the fatigue levels would be significantly reduced when using the fused display.

The calibration and fusion methods presented here show great potential for assisting users in Wilderness Search and Rescue. They allow heat information to be added to the color imagery obtained from UAVs. The fusion methods decrease the cognitive load on the searcher while maintaining the ability to correctly detect objects of interest.

Acknowledgments

The work was partially supported by the National Science Foundation under grant numbers 0534736 and 0812653. Any opinions, findings, and conclusions or recommendations expressed are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[1] KSL5 News, "Detective dies from helicopter crash injuries," November 2006. [Online]. Available: http://www.ksl.com/?sid=666341&nid=148

[2] D. Gibbins, P. Roberts, and L. Swierkowski, "A video geo-location and image enhancement tool for small unmanned air vehicles (UAVs)," in *Intelligent Sensors, Sensor Networks and Information Processing Conference*, 2004, pp. 469– 473.

[3] F. Rafi, S. M. Khan, K. Shafiq, and M. Shah, "Autonomous target following by unmanned aerial vehicles," in *SPIE Defense and Security Symposium*, Orlando, FL, 2006.

[4] M. A. Goodrich, B. S. Morse, D. Gerhardt, J. L. Cooper, M. Quigley, J. A. Adams, and C. Humphrey, "Supporting wilderness search and rescue using a camera-equipped mini UAV," *Journal of Field Robotics*, vol. 25, no. 1-2, pp. 89–110, 2008.

[5] N. Nandhakumar and J. Aggarwal, "Integrated analysis of thermal and visual images for scene interpretation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 10, no. 4, pp. 469–481, 1988.

[6] C. O'Conaire, N. O'Connor, E. Cooke, and A. Smeaton, "Comparison of fusion methods for thermo-visual surveillance tracking," in *9th International Conference on Information Fusion*, July 2006, pp. 1–7.

[7] F. Bunyak, K. Palaniappan, S. K. Nath, and G. Seetharaman., "Flux tensor constrained geodesic active contours with sensor fusion for persistent object tracking," *Journal of Multimedia*, vol. 2, pp. 20–33, August. 2007.

[8] J. Han and B. Bhanu, "Detecting moving humans using color and infrared video," in *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, Aug 2003, pp. 228–233.

[9] L. Zheng and R. Laganiere, "Registration of IR and EO video sequences based on frame difference," in *Canadian Conference on Computer and Robot Vision*, 2007, pp. 459–464.

[10] J. Pluim, J. Maintz, and M. Viergever, "Mutual-information-based registration of medical images: a survey," *IEEE Transactions on Medical Imaging*, vol. 22, no. 8, pp. 986–1004, Aug. 2003.

[11] P. Viola and W. M. Wells, III, "Alignment by maximization of mutual information," *International Journal of Computer Vision*, vol. 24, no. 2, pp. 137–154, 1997.

[12] Y. Lin and G. Medioni, "Map-enhanced UAV image sequence registration and synchronization of multiple image sequences," in *IEEE Conference on Computer Vision and Pattern Recognition*, June 2007, pp. 1–7.

[13] Y. Caspi and M. Irani, "Alignment of non-overlapping sequences." in *IEEE International Conference on Computer Vision, July 7-14, 2001, Vancouver, British Columbia, Canada*, 2001, pp. 76–83.

[14] J. Knight and I. Reid, "Self-calibration of a stereo rig in a planar scene by data combination," in *15th International Conference on Pattern Recognition*, vol. 1, 2000, pp. 411–414 vol.1.

[15] A. Zisserman, P. Beardsley, and I. Reid, "Metric calibration of a stereo rig," in *IEEE Workshop on Representation of Visual Scenes*, 24 Jun 1995, pp. 93–100.

[16] J. B. A. Maintz and M. A. Viergever, "A survey of medical image registration," *Medical Image Analysis*, vol. 2, no. 1, pp. 1–36, 1998.

[17] P. van den Elsen, L. van 't Zelfde, and M. Viergever, "Near-automatic detection of arrow-shaped markers for CT/MRI fusion," in *11th IAPR International Conference on Pattern Recognition*, vol. I, 1992, pp. 755–759.

[18] L. Merino, F. Caballero, J. M. de Dios, and A. Ollero, "Cooperative fire detection using unmanned aerial vehicles," in *IEEE International Conference on Robotics and Automation*, 2005, pp. 1884–1889.

[19] P. Rudol and P. Doherty, "Human body detection and geolocalization for UAV search and rescue missions using color and thermal imagery," in *IEEE Aerospace Conference*, March 2008, pp. 1–8.

[20] P. Burt and R. Kolczynski, "Enhanced image capture through fusion," in *Fourth International Conference on Computer Vision*, 11-14 May 1993, pp. 173–182.

[21] A. Waxman, A. Gove, M. Siebert, D. Fay, J. Carrick, J. Racamato, E. Savoye, B. Burke, R. Reich, W. McGonagle *et al.*, "Progress on color night vision: Visible/IR fusion, perception and search, and low-light CCD imaging," in *Proceedings of SPIE*, vol. 2736, 1996, p. 96.

[22] T. Wilson, S. Rogers, and M. Kabrisky, "Perceptual-based image fusion for hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, no. 4, pp. 1007–1017, Jul 1997.

[23] P. Simard, N. K. Link, and R. V. Kruk, "Evaluation of algorithms for fusing infrared and synthetic imagery," 2000.

[24] J. Illingworth and J. Kittler, "A survey of the hough transform," *Computer Vision, Graphics, and Image Processing*, vol. 44, no. 1, pp. 87–116, 1988.

[25] A. P. Dempster, N. M. Laird, and D. B. Rdin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society, Series B*, vol. 39, pp. 1–38, 1977.

[26] J.-Y. Bouguet and P. Perona, "Camera calibration from points and lines in dual-space geometry," California Institute of Technology, Tech. Rep., 1997.

[27] ——, "3d photography on your desk," *Sixth International Conference on Computer Vision*, pp. 43–50, Jan 1998.

[28] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.

[29] E. Michaelsen, M. Kirchhof, and U. Stilla, "Sensor pose inference from airborne videos by decomposing homography estimates," in *International Society for Photogrammetry and Remote Sensing Congress*, July 2004.

[30] V. Bruce, M. A. Georgeson, and P. R. Green, *Visual Perception: Physiology, Psychology, and Ecology*. Psychology Press, August 2003.

[31] "Black Widow AV — Wireless Aerial Video Solutions". [Online]. Available: http://www.blackwidowav.com

[32] "FLIR Pathfind IR". [Online]. Available: http://www.corebyindigo.com

[33] "Procerus Technologies". [Online]. Available: http://www.procerusuav.com