1. Introduction

Unmanned aerial vehicles (UAVs) are aircrafts which have the capability of flight without an onboard pilot. These vehicles are remotely controlled, semi-autonomous, autonomous, or have a combination of these capabilities. UAV’s has its applications in a whole lot of domains. Image processing applications with specific importance to surveillance and reconnaissance is of immense interest.

UAVs are equipped with imaging sensor platform, which operates remotely controlled, semi-autonomously or autonomously, without a pilot sitting in the vehicle. The platform may have a small or medium size still-video or video camera, thermal or infrared camera systems, airborne light detection and ranging (LIDAR) system, or a combination thereof. All these different kinds of cameras are an effective sensor tool which is portable, light weight and airborne in a platform on the UAV.

Thermal images have a valuable advantage over the visual images. Thermal images do not depend on the illumination, the output is the projection of thermal sensors of the emissions of heat of the objects. This unique merit gives rise for effective segmentation of objects. Ultimately, surveillance measure using an UAV gets improved.

With the development of new imaging sensors arises the need of a meaningful combination of all employed imaging sources. Image fusion of visual and thermal sensing outputs adds a new dimension in making the target tracking more reliable. Target tracking at instances of smoke, fog, cloudy conditions gets improved. With conditions of same background colour perception of target unnoticed getting eliminated with thermal image inclusion, image fusion gives complementary information. A holistic system which represents the combined fused data is perceived at the control level of the UAV’s.
2. Thermal imaging

Thermography which uses black body radiation law makes it to have information gathering without visible illumination. Thermal imaging cameras detect radiation in the infrared (IR) range of the electromagnetic spectrum (3-6 µm and 8–14 µm). The charge coupled device (CCD) and complementary metal oxide semiconductor (CMOS) sensors are used for visible light cameras. These can detect only the non thermal part of the infrared spectrum called near-infrared (NIR). On the other hand, thermal imaging cameras make use specialized focal plane arrays (FPAs) that respond to longer wavelengths (mid- and long-wavelength infrared).

There is also a difference between how far one can see with a cooled and with an uncooled thermal imaging camera. Cooled camera systems are more expensive, but generally have a longer range than uncooled systems under many conditions. Extremely long range thermal imaging applications are best served by cooled camera systems. This is particularly true in the midwave band in humid atmospheric conditions.

The heat radiation is focused onto special receptors in the camera which convert it into a format which is displayed on a monitor in monochrome which is recognisable by the human eye. The objects emitting the greatest intensity of heat are usually presented as the darkest (black) in the greyscale, i.e. known as ‘black-hot’. Many cameras have a function whereby the functionality can be switched from ‘blackhot’ to ‘white-hot’ and back again at the operator’s wish. Probably the greatest ‘enemy’ of thermal imaging is extended rainfall since that has the effect of cooling all inanimate objects and severely reducing contrast.

Thermal imaging makes it possible of real time target tracking. Detection of targets in dark and low light conditions can be done. All weather operation and dull, dirty and dangerous (DDD) roles are possible.

Thermal imaging cameras produce a clear image in the darkest of nights, in light fog and smoke and in the most diverse weather conditions. There has also been an increased interest in thermal imaging for all kinds of security applications, from long-range surveillance at border crossings, truck and shipping container inspection, to the monitoring of high-security installations such as nuclear power stations, airports, and dams. But thermal imaging has a lot more to offer than just a night vision solution for security applications.

Car manufacturers are integrating night vision modules for driver vision enhancement into cars. By helping drivers to see at night, accidents can be avoided. Boats and yachts are being equipped with thermal imaging cameras for night time navigation and other maritime applications like man overboard searches.

Often the thermal imager is just a small part of the complete system as in an UAS, so it needs to be as small, light and inexpensive as possible. Low-cost thermal imager is used as a pilot’s night vision enhancement. It helps pilots by enhancing the ability to see terrain and other aircraft at long ranges, even in total darkness, light fog, dust and smoke. Thermal imaging is a technology that enables detection of people and objects in total darkness and in very diverse weather conditions.

A typical application for thermal imaging is border security, where most threats occur at night. Thermal imaging allows the aircraft to fly in total darkness and detect targets through smoke. The same aircraft can also be used to detect such things as forest fires. Areas which are hotter than the surroundings can indicate the start of a fire and can clearly be seen on a thermal image.
3. Image segmentation

Segmentation is the key and the first step to automatic target recognition, which will directly affect the accuracy of the following work. As a result, the division methods and its precision degree are essential. Infrared heat wave image is different from the visible light images. It reflects the distribution of the object surface temperature and latent characteristics of material form.

The infrared heat radiation, due to the imperfections of the system, will bring a variety of noise in the imaging process. The noise of complex distribution of infrared images makes the signal to noise ratio lower than visible light images.

3.1 2D OTSU algorithm

The two dimensional Otsu algorithm is given as follows. Suppose an image pixel size is M × N, gray-scale of the image ranges from 0 to L-1. The neighborhood average gray \( g(m,n) \) of the coordinate definition \((m,n)\) pixel point is as follows:

\[
g(m,n) = \frac{1}{k \times k} \sum_{i=-(k-1)/2}^{(k-1)/2} \sum_{j=-(k-1)/2}^{(k-1)/2} f(m+i, n+j)
\]

Calculating the average neighbourhood gray of each pixel point, a gray binary group \((i, j)\) may form. We use \( C_{ij} \) to represent the occurrence frequency of \((i, j)\). Then the probability \( P_{ij} \) of vector \((i, j)\) may be determined by the formula:

\[
P_{ij} = \frac{C_{ij}}{M \times N}
\]

Here, \( 0 \leq i, j < L \), and \( \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} P_{ij} =1 \).

Assuming the existence of two classes \( C_0 \) and \( C_1 \) in Two dimensional form, the histogram represents their respective goals and background, and with two different probability density distribution function. If making use of two-dimensional histogram threshold vector \((s, t)\) to segment the image (of which \( 0 \leq s, t < L \)), then the probability of two classes are respectively: The probability of background occurrence is:

\[
\omega_0 = P(C_0) = \sum_{i=0}^{s} \sum_{j=0}^{r} P_{ij} = \omega_0 (s,t)
\]

The probability of object occurrence is:

\[
\omega_1 = P(C_1) = \sum_{i=s+1}^{L-1} \sum_{j=t+1}^{L-1} P_{ij} = \omega_1 (s,t)
\]

The definition of dispersion matrix:

\[
\sigma_B = \omega_0 \left[(\mu_0-\mu_0)(\mu_0-\mu_0)^T\right] + \omega_1 \left[(\mu_1-\mu_1)(\mu_1-\mu_1)^T\right]
\]

When the track of the above-mentioned dispersion matrix gets the maximum, the corresponding threshold of segmentation is the optimal threshold \((S, T)\), namely:

\[
\text{tr}(\sigma_B(S, T)) = \max_{0 \leq s, t < L} \{\text{tr}(\sigma_B(S, T))\}
\]

We know that 2-D thermal images with noise segmented by Otsu way may get better results compared to one dimensional threshold segmentation methods. However, the computation cost gets huge, which is because the determination of the optimal threshold need to travel all the \( s \) and \( t \), of which \( 0 \leq s, t < L \). That is to say, the more gray scale value of images is, the longer choice time of the threshold is.
The segmentation of various thermal images is illustrated here. The segmentation results can be made more efficient in identification of targets and optimized by a number of methods. These segmentation results are used in fused image target tracking. One such method will be determining optimum threshold using histogram analysis. The method of using chaos based genetic algorithm makes the process time lower. The chaos based genetic algorithm uses Otsu algorithm as fitness function and proceeds for segmentation.

The so-called chaos refers to the uncertainty in the system appearing in seemingly without rules. From the mathematical sense, for determined series initial values, it is possible to predict the long-term behaviour of the system, even know its past behaviour and state by the power system. It is a unique phenomenon of non-linear systems. Chaos has randomicity, ergodicity and regularity. Based on these characteristics of chaos, using chaotic variables to optimize search is no doubt better than random search. Logistic map is the most basic chaotic map. The chaos equation of the Logistic definition can be described as follows:

\[ x_{n+1} = ux_n(1-x_n) \; ; \; \quad 0 < u \leq 4 \]  

(7)

\( x_n \) lies between 0 and 1. \( n \) varies from 0,1,2, etc.

Fig. 1. Thermal Image

Fig. 2. Segmented Image
When \( u = 4 \), the system is the chaos state of the biggest ergodicity. Genetic algorithm is a kind of random search algorithm drawn from natural selection and natural genetic mechanisms of biology, which is particularly well suited to deal with the complex and nonlinear problems that traditional search methods are difficult to resolve.

![Thermal Image](image1.png)

Fig. 3. Thermal Image

![Segmented Image](image2.png)

Fig. 4. Segmented Image

The segmentation results are shown in Fig. 1 to Fig. 4 with the segmented image of the thermal image.

### 4. Image fusion process

Image fusion is the process by which two or more images are combined into a single image retaining the important features from each of the original images. The fusion of images is often required for images acquired from different instrument modalities or capture techniques of the same scene or objects. Important applications of the fusion of images include medical imaging, microscopic imaging, remote sensing, computer vision, and
robotics. Fusion techniques include the simplest method of pixel averaging to more complicated methods such as principal component analysis (PCA) and wavelet transform (WT) based fusion.

Image fusion is the process of relevant information extraction from two or more images. The resulting image will encompass more information than any of the given input images. A multi sensor, multi temporal and multi view information techniques are required which overcome the limitations of using single sensor information. The benefits of image fusion are extended range of operation, extended spatial and temporal resolution and coverage, reduced uncertainty, higher accuracy, reliability and compact representation.

Various kinds of image fusion for visual, Infrared and Synthetic Aperture Radar (SAR) images exist. Some of the primitive fusion process algorithms also have disadvantages. Direct fusion method makes the image blurred. The pixel based image fusion is computationally complex for large and high resolution images. The image averaging method produces a reduced contrast of information. The requirement is thus a novel and efficient information fusion process. Image fusion has a significant role of recognition of targets and objects. Target identification, localisation, filtering and data association forms an important application of the fusion process. Thus an effective surveillance and reconnaissance system can be formed.

There is information which is redundant and at the same time complementary too. The following summarize several approaches to the pixel level fusion of spatially registered input images. A generic categorization of image fusion methods in the following:

- Linear Superposition, Nonlinear Methods, Expectation Maximisation, Image Pyramids, Wavelet Transform, Generic Multiresolution Fusion Scheme, Optimization Approaches, Artificial Neural Networks, Fuzzy Techniques

Some of the prominent applications of image fusion are:


### 4.1 Wavelet transform

A signal analysis method similar to image pyramids is the discrete wavelet transform. The main difference is that while image pyramids lead to an over complete set of transform coefficients, the wavelet transform results in a nonredundant image representation. The discrete two dimensional wavelet transform is computed by the recursive application of low pass and high pass filters in each direction of the input image (i.e. rows and columns) followed by sub sampling.

These basis functions or baby wavelets are obtained from a single prototype wavelet called the mother wavelet, by dilations or contractions (scaling) and translations (shifts). They have advantages over traditional Fourier methods in analyzing physical situations where the signal contains discontinuities and sharp spikes.

Image fusion process is achieved by multiresolution decomposition at fourth level. The multiwavelet decomposition coefficients of the input images are appropriately merged and a new fixed image is obtained by reconstructing the fused multiwavelet coefficients. The theory of multiwavelets is also based on the idea of multiresolution analysis (MRA) as shown in Fig. 5. During a single level of decomposition using a scalar wavelet transform, the two-dimensional (2-D) image data is replaced with four blocks corresponding to the subbands representing either low-pass or high-pass filtering in each direction.
The Haar wavelet is a certain sequence of rescaled "square-shaped" functions which together form a wavelet family or basis. Wavelet analysis is similar to Fourier analysis in that it allows a target function over an interval to be represented in terms of an orthonormal function basis. The Haar wavelet is also the simplest possible wavelet.

**4.2 FIS based image fusion**

Neural Network and Fuzzy Logic approach can be used for sensor fusion. Such a sensor fusion could belong to a class of sensor fusion in which case the features could be input and decision could be output. The system can be trained from the input data obtained from the sensors.

The basic concept is to associate the given sensory inputs with some decision outputs. The following algorithm for pixel level image fusion using Fuzzy Logic illustrate the process of defining membership functions and rules for the image fusion process using FIS (Fuzzy Inference System) editor of Fuzzy Logic toolbox in MATLAB. The process flow of the process is as follows:

1. The visual and thermal image forms the inputs for the fusion system. The inputs must be with the same field of view.
2. With a gray level conversion applied, the two images are transformed to a column form.
3. The number and type of membership functions for both the input images are given to the FIS. The rule base for FIS decides how the output fused image should be.
4. The fused image is converted back to matrix format from column form.

The fuzzy system considered here is Mamdani type and its’ FIS model in MATLAB is shown in Fig. 6. The Mamdani rule base is a crisp model of a system, i.e. it takes crisp inputs and produces crisp outputs. It does this with the use of user-defined fuzzy rules on user-defined fuzzy variables.

Fig. 5. Wavelet Multi Resolution Analysis

The Haar wavelet is a certain sequence of rescaled "square-shaped" functions which together form a wavelet family or basis. Wavelet analysis is similar to Fourier analysis in that it allows a target function over an interval to be represented in terms of an orthonormal function basis. The Haar wavelet is also the simplest possible wavelet.
Fig. 6. FIS Model

Fig. 7. FIS Membership Function
Fig. 8. Rule Base of FIS

Fig. 9. Surface View of FIS
The idea behind using a Mamdani rule base to model crisp system behavior is that the rules for many systems can be easily described in terms of fuzzy variables. Thus we can effectively model a complex non-linear system, with common-sense rules on fuzzy variables. The operation of the Mamdani rule base can be broken down into four parts: 1) mapping each of the crisp inputs into a fuzzy variable (fuzzification), 2) determining the output of each rule given its fuzzy antecedents, 3) determining the aggregate output of all of the fuzzy rules; 4) mapping the fuzzy output to crisp output (defuzzification). The membership function used is Guassian as shown in Fig. 7. The rule base is formed for the fusion process as shown in Fig. 8. The surface view of the two inputs and one output is represented in Fig. 9.

5. Target tracking of objects

Automatic detection and tracking of interested targets from a sequence of images obtained from a reconnaissance platform is an interesting area of research for defence related application. The video images are obtained from an unmanned aerial vehicle (UAV) with on-board guidance and navigation system. The aircraft carries a multispectral camera which acquires images of the territory and sends the information to a ground control station (GCS) in real time. During flight, the pilot in the ground control station may identify a region of interest as a target. This identification can be click and target type or an intelligent perception type. The target which appears on a small window could be tracked by engaging track mode. Optical flow is an approximation of the local image motion based upon local derivatives in a given sequence of images. That is, in 2D it specifies how much each image pixel moves between adjacent images while in 3D it specifies how much each volume voxel moves between adjacent volumes. The 2D image sequences used are formed under perspective projection via the relative motion of a camera and scene objects. The differential methods for determining the optical flow are Lucas-Kanade, Horn-Schunck, Buxton-Buxton, Black-Jepson and variational methods.

**Lucas–Kanade method** is a widely used differential method for optical flow estimation developed by Bruce D. Lucas and Takeo Kanade. The assumption in this method is that the flow is essentially constant in a local neighbourhood of the pixel under consideration, and solves the basic optical flow equations for all the pixels in that neighbourhood, by the least squares criterion. **Horn–Schunck method** of estimating optical flow is a global method which introduces a global constraint of smoothness to solve the aperture problem. The Horn-Schunck algorithm assumes smoothness in the flow over the whole image. Thus, it tries to minimize distortions in flow and prefers solutions which show more smoothness.

5.1 Tracking of objects in fused images

A brief representation of the process used is as represented in Fig. 10. The visual and thermal video is first obtained from an airborne camera. The cameras used must be of acceptable resolution with good preprocessing features. A high resolution, light weight, rugged, portable, stabilized platform sensing element is required to be mounted on the airframe. The images are preprocessed for ego-sensory motion, atmospheric disturbances and inherent noise. Then a fusion process of thermal and visual images is adapted. The algorithms for such a fusion process would involve processing of information separately and then fusing the data. The other way is to fuse and process the data available.
Fig. 10. Process Flow
The transmission of information from on board real time processing and display of all required data are done in the ground station. The control interaction with UAV is done by wireless communication transfer.

The video file obtained from the camera is loaded as an avi file. It is processed in MATLAB. It aids in real time and efficient implementation of the algorithm. The video is processed for detection of targets in a sequence of frames. The target detection and localisation process commences from applying the frames to an optical flow pattern which generates vectors. The target detection is done with respect to feature matching and extraction from the initial frame. The target which is the region of interest is separated out effectively. By using this novel technique, an optimized search process for an effective segmented image with less noise effects when compared to other algorithms is obtained. The targets are thus effectively located and segmented. The output frames are integrated to form a video file. Then effective tracking algorithms are applied for the process. Filtering and the required data association for the output obtained are subsequently done. The inputs needed for processing include that of two images (Visual and thermal) for image fusion. The steps involved are:

1. The airborne video is obtained from the UAV with
   a. A visual and thermal camera onboard the UAV.
   b. Wireless transmission of video to ground control station.
2. The inputs needed are two videos (visual and thermal) which are split into images with respect to frames. The factors to be considered are:
   a. Two videos are to be of same resolution and frame rate.
   b. Two images are to be considered with the same field of view (FOV).
3. Segmentation process is applied to the images and the targets are segmented.
4. For the fusion process:
   a. Consider two images at a time.
   b. Apply wavelet based image fusion with Haar transform with four level decomposition of the image.
   c. An inverse transform is applied to get back the fused image.
5. For tracking algorithm:
   a. The tracking algorithm is implemented in SIMULINK in MATLAB. The fused images are stitched to form a video of the same frame rate.
   b. The fused image is applied to an optical flow technique. The optical flow technique is based on Horn-Schunck method.
   c. Further segmentation accompanied with thresholding principles is applied for the output obtained after computing optical flow.
   d. Blob detection principles accompany the process for spotting the targets in the frame of concern. Thus the targets are given a rectangular boundary of identification.

6. Results

Image fusion results based on wavelet transform are discussed. Two images (Visual and thermal) are taken as shown in Fig. 11 and Fig. 12.

The visual image gives a realistic human sensing view. The thermal image identifies the target with the temperature difference coming into the picture with objects possessing different emissivity values.
The fused image is given as Fig. 13. The fusion process is applied for a concealed weapon detection process. The results for such an application are given below in Fig. 14 to Fig. 15. The visual and thermal images are fused and the fused image gives combined information in one representation.
The thermal images clearly give the information about concealed weapons. The fused image gives both the RGB based information present in visual image and the concealed weapon information present in thermal image. The fused image is thus capable of giving enhanced detection capabilities.

6.1 Target tracking results
For target tracking, the database of thermal and visual images is considered. The video obtained is converted to a series of frames. Each thermal and visual image is fused and at the end the fused images are combined to form a video which is given for target tracking process.

One such visual and thermal image from the sequence of frames is shown in Fig. 16 and Fig. 17. The fused image obtained as a result of wavelet transform using Haar wavelet is shown in Fig. 18.
The target tracking results including segmentation, optical flow and identification of targets in their corresponding frame are shown in Fig. 19, Fig. 20 and Fig. 21. For the four sample frames considered, the segmentation results and tracking results are illustrated. The segmentation shows that of humans being segmented according to the threshold values.
The optical flow which is found out from the images gives the movement of the humans which is being tracked and given a rectangular representation.
The following results correspond to taking the thermal images alone and doing target tracking. The target tracking for aerial aircrafts’ motion is shown in Fig. 22 and Fig. 23. These images correspond to the optical flow and the target detection from a real time video sequence. The target tracking for identification of humans is given in Fig. 24. The targets are tracked with respect to the movement of objects identified with optical flow. In this frame, an airborne flight is being tracked. This aids to the visual perception other than usage of RADAR.

Fig. 22. Optical Flow Detection

Fig. 23. Target Detected Image
Further sample results for tracking of objects in thermal video is shown in Fig. 24. This is an airborne thermal video which is tracked for vehicles and humans if present. The ego sensory motion is present in this kind of a dynamic moving platform. They have to be compensated for a stabilized video and for the tracking process has to be improved for a multiple target tracking environment.

Fig. 24. Target Detection in Aerial Image

7. Conclusion

The airborne images obtained from an UAV are analysed in ground control station. By using the thermal images, all weather and night operation are possible. Visual and thermal image fusion is done and the fused image is given for target tracking. This system has the benefit of enhanced target tracking application wherein only visual or thermal target tracking would not provide sufficient efficiency. Thus the image fusion process augments information leading to an improved system as a whole. The overall system incorporates segmentation, fusion and target tracking principles.
8. Acknowledgements

The authors are grateful to The Vice Chancellor, Anna University and The Dean, Madras Institute of Technology, Anna University for providing organizational facilities for carrying out this research work at Madras Institute of Technology. We would like to express our sincere thanks & gratitude to The Director NAL for providing sufficient funds for successful completion this sponsored research work. We also express special thanks to V.S.Chandrasekhar, B.P.Shasidhara and K.V.Srinivasan, Scientists, Aeronautical Development Establishment(ADE), DRDO, Bangalore. for their inspiration and encouragement. We would like to convey our thanks and wishes to all the Avionics students.

9. References


A well-known statement says that the PID controller is the “bread and butter” of the control engineer. This is indeed true, from a scientific standpoint. However, nowadays, in the era of computer science, when the paper and pencil have been replaced by the keyboard and the display of computers, one may equally say that MATLAB is the “bread” in the above statement. MATLAB has become a de facto tool for the modern system engineer. This book is written for both engineering students, as well as for practicing engineers. The wide range of applications in which MATLAB is the working framework, shows that it is a powerful, comprehensive and easy-to-use environment for performing technical computations. The book includes various excellent applications in which MATLAB is employed: from pure algebraic computations to data acquisition in real-life experiments, from control strategies to image processing algorithms, from graphical user interface design for educational purposes to Simulink embedded systems.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following:
