January 2001

Review: Extending Visible Band Computer Vision Techniques to Infrared Band Images

Shih-Schon Lin
University of Pennsylvania, shschon@seas.upenn.edu

University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-01-04.

Review: Extending Visible Band Computer Vision Techniques to Infrared Band Images

Abstract

Infared imaging process bears many similarities to the visible imaging process. If visible band computer vision techniques can be used on infrared images with no or small adjustments it would save us the trouble of redeveloping a whole new set of techniques. However, there are important differences in the practical environmental parameters between visible and infrared bands that invalidates many convenient background assumptions inherent to visible-band computer vision techniques. We review here the underlying reasons why some computer vision techniques can while some cannot be applied directly to infrared images. We also examine a few attempts to extend computer vision to infrared images and discuss their relative merits.

Comments

University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-01-04.

Review:
Extending Visible Band Computer Vision Techniques to Infrared Band Images

Shih-Schön Lin
(shschon@grasp.cis.upenn.edu)
Technical Report MS-CIS-01-04
GRASP Laboratory, Computer and Information Science Department
University of Pennsylvania

Abstract

Infrared imaging process bears many similarities to the visible imaging process. If visible band computer vision techniques can be used on infrared images with no or small adjustments it would save us the trouble of redeveloping a whole new set of techniques. However, there are important differences in the practical environmental parameters between visible and infrared bands that invalidates many convenient background assumptions inherent to visible-band computer vision techniques. We review here the underlying reasons why some computer vision techniques can while some cannot be applied directly to infrared images. We also examine a few attempts to extend computer vision to infrared images and discuss their relative merits.

1. Introduction

Computer Vision, especially the part dealing with human vision like imaging modality, evolves mostly around images taken within the "visible band" of electro-magnetic wave spectrum. The environment we live in play a very important role in the selection of this spectral band as "visible" to humans. The main energy source, our sun, emits all frequencies of radiation with peak around the visible band. The temperature of our earth and the composition of our atmosphere results in an environment that radiation in the visible band can travel long distances with relatively low attenuation. Under the level of temperature around the surface of the earth, most material emits little visible band energy of its own. Thus using visible band we have a good passive way of detecting far away events with few environment related interference.

The downside of using visible band, however, is that half of the day the sun is not shining overhead and visible band signal drops below the level of reliable detection.

For the average temperature range of the earth's surface, however, most common material spontaneously emits considerable amount of radiation energy in the band loosely termed as infrared band. Since the photon energy in the infrared region falls within the range of many molecular vibration quantum energy level differences, the gas molecules of earth's atmosphere can easily absorb some bands within the infrared region. There remains, however, several "windows" exist in the infrared band that are not strongly absorbed by the earth's atmosphere and thus can be used for long-range imaging.

Infrared band radiation is first discovered in an experiment by Sir William Herschel [B5]. The radiation is detected indirectly by the heating associated with the absorption of infrared radiation energy. This principle is still in use today in the latest "uncooled" infrared detectors. Although the detection of infrared energy is done almost two centuries ago, the precise quantitative measurement of infrared radiation is difficult and its development lags behind the visible light detectors. One major problem is noise. In visible band the main energy source is the sun or human controlled light source, while in infrared everything around us is a potential light source. Another bottleneck in the development of infrared imaging camera is the material needed for infrared lenses. Except for bands very close to visible bands, ordinary optical glasses are opaque in most infrared band.

Semi-conductor and micro-machining technology greatly improve the properties and performances of infrared detectors, as well as lowering the costs. The lens material, although still few compared to the visible band lens material, already have some commercial products available. With the rapid advances of infrared imaging cameras, the demand of computer vision algorithm to do automatic analysis of infrared images is growing.

Low level computer vision techniques, sometimes classified as image processing techniques, make little assumption on the underlying imaging modality and can thus be applied to infrared images as well with little or no modifications. The relative performance, however, can differ because most infrared images are generally lower resolution and contain more noise than visible-band images. This may improve with time but for now we must deal with it in practical applications.

For higher level computer vision that extracts more abstract or detailed object properties from objects being imaged, e.g. shape from shading, the algorithms are developed in connection with the particular physical properties of the visible band and thus are not directly applicable to other spectral bands.

Since many visible-band computer vision algorithms are well understood and field tested, we would like to apply as many as possible to the infrared band images. We first examine the complete process of visible and infrared imaging. Then we look at several attempts to extend computer vision into infrared images and how far they have gone to expanding the limits.

2. Physical Similarity and Differences in Imaging Process between IR and Visible

The most general imaging process involves the generation of radiation, and altering of the radiation by reflection, refraction, absorption, and scattering, and finally collected by the optical system and captured by the detector. More structured materials, like crystals, can have more peculiar optical effects like rotation of the polarization direction, but since this effect is hardly detectable in usual outdoor or indoor scenes in computer vision applications, we do not pursue them further here.

2.1 Wavelength of Radiation

The main difference between visible and IR radiation is their wavelength (and frequency, since the speed of light in vacuum is the same for all wavelengths). The visible band is defined loosely between 350 nm ~ 780 nm and the band between 780 nm to 1 mm are called IR band. These definitions are rather loosely defined, as human vision has individual variations and IR is not a strictly defined term. Within IR band people often subdivide it into several sub band for convenience. But since different professions work for different ranges of IR the same name of an IR sub band might have different definitions. For example, scientific researcher who work with the whole IR spectrum, defines large sub bands with the "long wave IR" or "Extra Long Wave IR" extending up to the boundary with microwave. In engineering applications like computer vision, the bands of IR that contains more interesting information is narrower, only between the limit of visible band up to about 15000~20000 nm, thus the term "long wave length IR" in computer vision literature is often limited to this range.

2.2 Passive and Active Light Source

On earth surface under normal room temperature, most material surfaces emit little visible light but appreciable radiation within IR band. More specifically, at 300K(Kelvin, absolute temperature, 0C=273.16K) the peak emission occurs at around 10000 nm. This prediction is made by the concept of "black body radiation". The "black body radiation" concept eventually sparked the all-important quantum physics. We show only a few results related to our discussion.

There are several factors involved in the surface thermal emission. The basic concept involved is the concept of energy conservation. However, the energy can be distributed differently among viewing angles and also among different wavelength of radiation. The "black body" is a conceptual ideal surface that absorbs completely any incoming radiation regardless of angle of incidence. This inherently omni-directional definition coupled with an imaginary thermal equilibrium condition eventually leads to the conclusion that a black body is also a perfect emitter that emits the same intensity of radiation in all direction. Thus the directional radiance of a black body in any direction is proportional to the total amount of energy emitted per unit surface area per unit time. Hence we only need to specify the spectral distribution of the black body. Max Planck found the closed form formula for the distribution to be[B4].

$$\frac{dR(\lambda,T)}{d\lambda} = \frac{2\pi hc^2\lambda^{-5}}{\exp(hc/\lambda kT)-1}$$

where

h=6.6256E-34 (Js) (Planck's constant)
c=2.998E8 (m/s) (speed of light in vacuum)
k=1.38054E-23 (J/K) (Boltzmann's constant)
R is the total energy flux emitted by a unit surface patch in thermal equilibrium at temperature T(in Kelvin) per unit wavelength

The form of the formula can be slightly different if we use frequency instead for the spectral unit, but the general properties are the same. This function form has a peak value that occurs at

$$\lambda_{max}(T) = C_1 / T$$

where C1=2897.6 (μm K)

This is Wien's Displacement Law[B5]

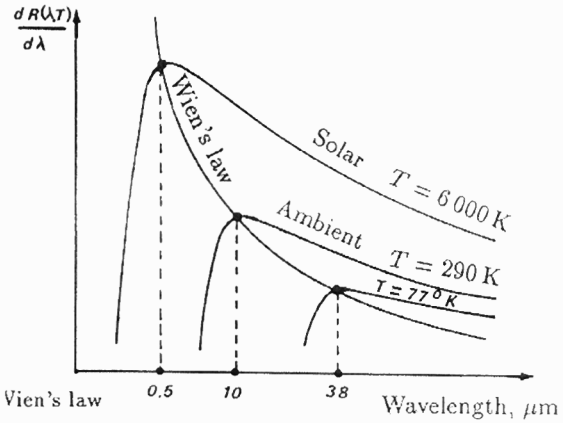

Figure 1 Wien's Displacement Law

The total energy, i.e. light energy including all possible wavelengths, is given by integrating the Planck's Blackbody formula,

$$R(T) = \sigma \cdot T^4$$

This is Stefan-Boltzman Law, where the constant σ = 5.670e-8 (K^4W/m^2). Historically these laws were discovered first experimentally, thus are named after the discoverers.

The blackbody serves as a standard for all surface emission of radiation in the sense that it is the best emitter in the temperature specified. Some real surface like the Sun and some blackened paints have surface emission properties very close to that of an ideal blackbody. Other surfaces are like blackbody only in certain wavelength ranges. To describe this variation from the 'ideal' black body the ratio of the actual energy emitted compared to the ideal quantity is defined to be 'emissivity' ε. Note that real surface emission properties not only differ from the ideal blackbody in terms of wavelength dependence but also on the directional distribution pattern. This leads to several different types of emissivity. The directional spectral emissivity of a surface would have different values at different direction and wavelength. This leads to a large table for only one material. Such table is very difficult both to produce and to use. Thus in practice, only the 'hemispherical' or 'normal' emissivity is listed in most material handbooks. The 'hemispherical' emissivity is the ratio of actual to ideal in all energy summed over all possible directions(a hemisphere). The 'normal' emissivity is the radiance ratio measured along the direction of the surface normal. 'Normal' emissicity is much easier to measure experimentally and for many surfaces the value is roughly proportional to the 'hemispherical' emissivity.

A typical surface spectral emission property is shown in Figure 2[B8].

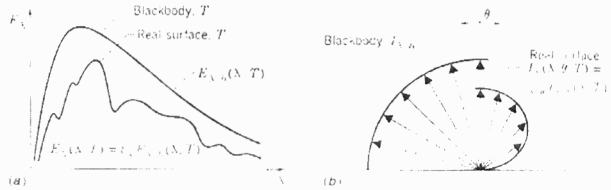

Figure 2 Real surface vs ideal blackbody. (a) Spectral (b) Directional differences

To do a qualitative estimation, we can start from calculating the peak emission band for black body for typical temperatures. As shown in Figure 1, the Sun resembles a 6000K blackbody and the peak emission is right in the middle of the visible band. A typical surface on Earth with ambient temperature around 290K(~17C) has peak at 10 μm, inside the IR band. Another thing to notice is that for an object under room temperature there is very little emission in the visible band. Which explains why most objects in room temperature do not glow (for human eyes) on their own.

This has several impacts on the construction of computer vision algorithms. First, in computer vision algorithms developed for visible band, the self-emission part can be safely ignored. For example, the classical 'shape from shading' technique[B7] is based entirely on surface reflectance. Such technique, while valid in visible band, can not be used in IR band without significant modification because in IR the self-emission contribution can not be ignored. A second implication is that, since in IR image many surface will be light source themselves, the brightness contrast may become larger. When this contrast exceeds the dynamic range of the camera, we get a saturation or decimation effect. This results in loss of features inside a very bright or very dark area. This will cause significant problem for pattern matching vision algorithm designed for visible band.

2.3 Surface interaction with incident radiation

Except for some special phenomenon, all surfaces interact with incident radiation in the following ways: reflection, absorption, and transmission. When these are the only interactions taking place, from conservation of energy we know the incident energy must go into one of the interaction. We can thus define the ratio of energy going into each interaction compared with the total incoming energy as the Absorptivity α, Reflectivity ρ, and Transmissivity τ.

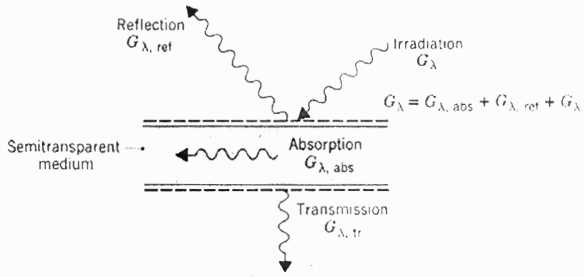

Figure 3 Surface interaction with incident radiation

Again we are faced with the fact that both direction and spectral dependence exists in these concepts. In the reflectivity it is further complicated by the fact that we have two directions (incident and reflected) directions to consider. Like in the case of emissivity, we often simplify the matter by using the ratio of the quantity that is summed over all direction and/or over all wavelengths. From conservation of energy, we have the following basic equations:

$$\rho_\lambda + \alpha_\lambda + \tau_\lambda = 1$$
$$\rho + \alpha + \tau = 1$$

for semi-transparent surfaces and

$$\rho_\lambda + \alpha_\lambda = 1$$
$$\rho + \alpha = 1$$

for opaque surfaces.

Notice that the definitions of these quantities are very different from that of emissivity. Emissitivity definition involves a standard reference surface while the other quantities we see here involve no standard reference. Under certain conditions, however, the numerical values of some emissivity and absorptivity can be the same. The required condition often involves some uniformity in spectral or directional distribution or the proportionality to blackbody properties. We shall check this when we see the use of this convenient equality in some of the algorithm we review.

One of the difficulties involved in IR computer vision is that these passive surface phenomena are mixed up with the self-emission phenomenon. The reflection process is particularly troublesome because the energy can reenter the same surface after multiple reflections. In visible band since we have more control over the light source the problem can often be dealt with by ignoring the weak multi-reflection components. In IR images where light source is scattered all over the scene the problem is much more complicated.

2.4 Tabulated Material Properties

In general, all the radiation related properties discussed so far are functions of temperature, direction(s), and wavelength. However, to tabulate all the dependencies means that for each material there must be a high dimensional grid of data points in order to represent the full functional dependency. It takes a lot of measurement work to build one such table and it is very cumbersome to use if we do complete such a table. In many engineering applications it is usually sufficient to have some average property values to be used with simplified models. Thus in most data tables published, the values listed are only for "spectral normal", or "total normal", which are the most easily measured quantities. For many common materials, this "normal" value is roughly proportional to the "hemispherical" value, see Figure 4 [B8]. Only when there are special needs will the complete functional dependence of a particular material be measured experimentally.

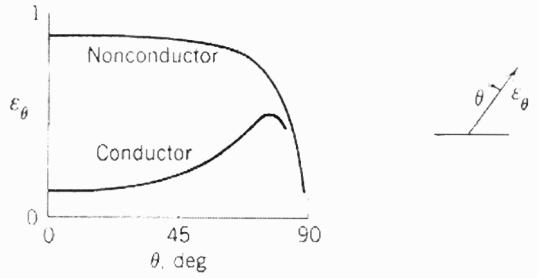

Figure 4 Representative directional dependence of total directional emissivity. The normal (θ=0) value is not far away from the mean (hemispherical) value.

2.5 Atmospheric Effects

Between the object and the observer it is always filled with particles of earth's atmosphere. The earth's atmosphere composes of about 4/5 nitrogen and 1/5 oxygen. Carbon dioxide takes up about 1% and water vapor level varies greatly from area to area and from time to time. Gas molecules and the much larger aerosol particles reflect, refract, absorb, and scatter light with the result of changing the spectral and spatial distribution of light energy.

For indoors visible band vision, the atmospheric effects are mostly insignificant because of the short distances involved and because the air is often stabilized by air conditioning. Thus the air effects are often ignored completely in many computer vision algorithm. In fact, most visible band optics design also ignores the effects of atmosphere partly because they can not be controlled.

In the visible band, the atomic absorption of atmospheric gas molecules is very weak and roughly uniform over the entire visible band. However, in the infrared region, there are several bands that are strongly absorbed by the atmosphere while some there exists some "window" bands that are not absorbed by the atmosphere. So while the radiation of the Sun is very close to that of a 5800K black body, the spectral distribution changes considerably when it reaches the surface of the earth because it passes long distances through the atmosphere, see Figure 5[B8].

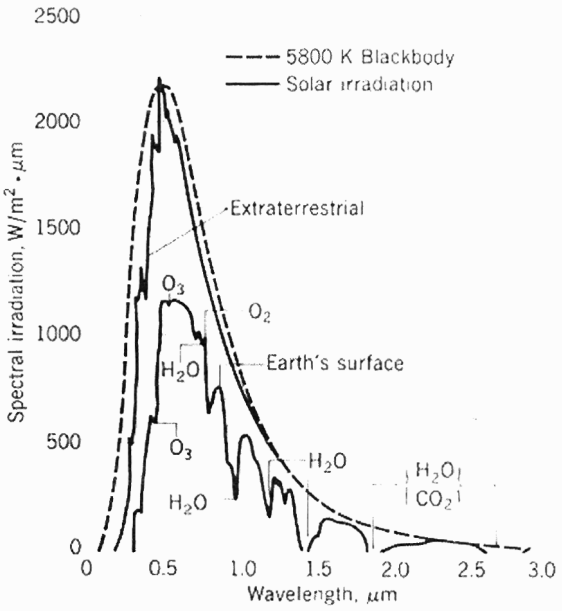

Figure 5 Spectral distribution of the Sun's radiation before and after it enters the earth's atmosphere

The refraction effect is most pronounced when there are significant temperature and/or density differences between air layers. Since air with different temperature and density have different index of refraction, the interface between such two layers acts just like the interface between the lens and air. This is why in hot desert area one can sometimes "see" images of distant city floating in the air. This refractive phenomenon is important and interesting but it is difficult to control because the exact condition of inhomogeneous air is difficult to measure and is changing all the time. Thus for most part it is considered separately.

Scattering effect, like atomic absorption, is easier to model because it can be modeled in a homogeneous atmospheric condition. Depending on the relative size of the particle compared to the wavelength of the incident radiation, there are two types of scattering: Rayleigh scattering which scatters almost uniformly in all direction, and Mie scattering which scatters mostly in directions close to the original incident direction., see Figure 6.[B8] The magnitude of the resulting effects are dependent both on the wavelength of the incident radiation and the thickness of the atmosphere it passes through. For example, the sky looks blue because the scatter cross section is greater for shorter wavelength components. The setting Sun looks red because the Sun light must pass through a thicker layer of atmosphere than in the day time and most short wavelength component are lost due to scattering.

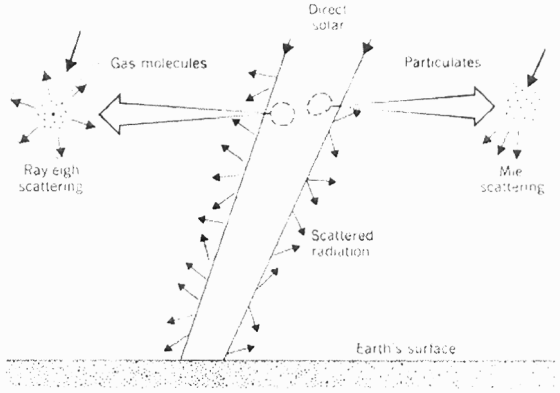

Figure 6 Atmospheric scattering for Sun light

The net result of atmospheric scattering is most often modeled as diffuse lighting, although in fact it is not that uniform as light coming in parallel to the ground is often weak in reality. See Figure 7[B8].

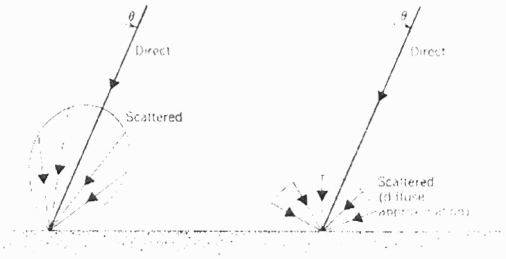

Figure 7 Spatial distribution of day light on Earth's surface. Left: Actual distribution, Right: Diffuse model

2.6 Focusing Devices

Imaging device is often composed of a focusing element and a 2D array of detecting elements. The focusing is necessary because light emanating from a point in space becomes weaker and weaker away from the point because the same amount of energy is distributed to

an increasingly larger space. The focusing element collects all the light energy covered by the aperture and focus them all to the same spot on the 2D array, thus greatly enhance the signal strength for the detector. The design of such focusing elements is often done under the assumption that there is no distortion of scene light by the atmosphere. The focusing element can be composed of refracting elements or reflecting elements alone or both. The index of refraction of a material varies with wavelength of the electro-magnetic wave. Thus a focusing device designed to work under one wavelength range may not work properly under other wavelength. For example, most material that is transparent in visible band can become opaque in longer wavelength infrared bands. Even if the material of the lens remain transparent in other wavelength bands, the focusing power and chromatic aberration characteristic may be different due to differences in index of refraction. The formulae involved (Equation 1 and Equation 2) clearly indicates the dependence on the indices of refraction of the lens materials[B6]:

$$\frac{1}{s_o}+\frac{1}{s_i}=(n_l-1)(\frac{1}{R_1}-\frac{1}{R_2})=\frac{1}{f}$$

Equation 1 Lensmaker's Formula

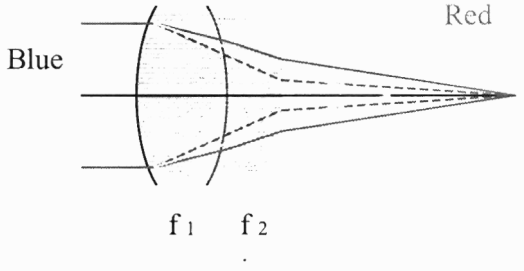

Figure 8 Acromatic doublets composed of one converging lens with focal length f 1 and one diverging lens with focal length f 2

$$\frac{f_{2Y}}{f_{1Y}}=-\frac{(n_{2B}-n_{2R})/(n_{2Y}-1)}{(n_{1B}-n_{1R})/(n_{1Y}-1)}$$

Equation 2 Achromatic doublets formula

Another wavelength related effect is the diffraction limit. The wave property of visible light is not pronounced in many situations because of its relatively short wavelength. However, thermal infrared wavelength is 10 to 100 times longer than that of the visible band so the iris lower bound is higher than that of the visible band lens in order to avoid strong diffraction effects.

2.7 Detectors

The most commonly used digital image detector in computer vision today is CCD (Charge Coupled Device). These device are sensitive to visible band as well as IR that is very close to visible band (about up to 1300 nm). But for other longer wavelength IR band, especially the band that a room temperature black body radiates most CCDs can no longer be used because thermal IR photon energy is much lower than that of the visible band photons. Sensors for the so called "thermal IR" band has been devised for about 200 years, but only until recently can these thermal IR detectors be miniaturized enough to be packaged as a small chip FPA (Focal Plane Array). Even so, in general these thermal IR FPAs are still larger and possesses less pixel density per unit area than ordinary CCD chips due to the special materials and complex structures involved with these thermal IR detector unit.

The IR detectors most widely used today can be roughly divided into 2 groups[B5;B2;B9]. The first group measures the IR indirectly by detecting the changes caused by the heat introduced when absorbing IR radiation. For example, Bolometers measure the heat induced electrical resistance change, while pyroelectric detectors measure the heat induced electrical capacitance change for certain crystals. Pneumatic IR detectors measure the pressure differences induced by heated expansion and thermopiles measure the differences in heat expansion rates. The most significant shortcomings of these types are relatively slow response time but are improving with micro-machining technology (smaller things heat up faster). The main advantage as compared to the other group, the quantum detectors, is that these devices operates at relatively higher temperature and thus do not need expensive and cumbersome cryogenic cooling.

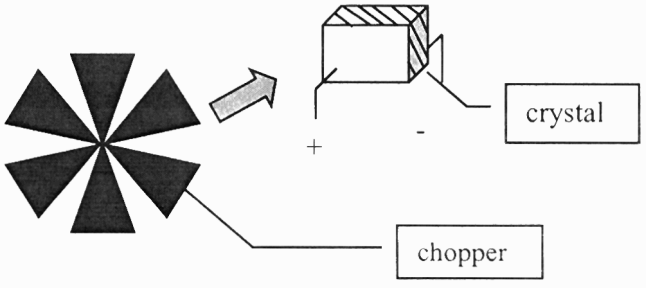

Figure 9 Pyroelectric IR detectors measure the transient current induced by the heated up crystal

surfaces, which in effect changes the capacitance. Since only changes in temperature are detected, a chopper is needed to produce continuous heat changes.

The second group is called "quantum detectors" because they utilize the quantum phenomenon of "photo-electric" effects to detect the IR photon directly. The main advantage is fast response time and ultra-high sensitivity that approaches the theoretical upper limit. The trade-off, however, is that the entire detector must be cooled down to very low temperatures in order to suppress background noise. If we look at Figure 1, we can see that if we cool the detector down to 77K(the temperature of liquid nitrogen), the background radiation peak will be offset to 380000 nm, far away from the signal we want (around 10000 nm). But such low temperature can not be maintained by thermal electric coolers, only specialized cryogenic cooling can maintain such low temperature. This makes the detector as a whole very expensive not only in purchase price but in operation costs as well. The cooling devices also make the whole system bulky, heavy, and power hungry. There is also this inconvenience of having to wait for the pre-cooling every time before use.

If we can get a IR detector that combines the high performance of the quantum detector and the low cost and ease of operation of heat sensing IR detectors many difficulties encountered in IR computer vision can be solved instantly. Recently there are some promising advances to improve the quality of heat sensing IR detector[B9], which may have great impact in the next 10 years. For now, for real time applications, even for limited task like ditch avoidance using IR stereo, the quality of heat sensing IR detectors is still not enough[B12].

3. Effects of IR Properties on Visible Computer Vision Algorithms

Computer vision is very diverse field. Anything that deduce useful information from one or more digital images using computer algorithms counts. For our discussion it is convenient to loosely categorize computer vision algorithms in the following criteria:

- **2D and 3D**: 2D computer vision works with the 2D image data itself and makes little or no assumptions on how the 2D data is acquired from a 3D world. As a result their method can be applied to many different modalities of images but they can not extract high-level information specific to applications. 3D computer vision, however, is based mostly on the projective projection principle and makes use of many reflective properties common in the visible band.
- **Single or Multiple Frames**: Computer vision algorithms start by trying to extract data from a single picture. Since in many imaging process 2D image reveals only one aspect of the object of interests and there can be many ambiguities or unknown information so more sophisticated computer vision algorithms uses multiple images. Multiple images gives more information in either time (video) or space(stereo) or both(moving camera video). The problem is how to register the same object or feature point in multiple images. Many resort to human aid but full automation is the ultimate goal. For automatic correspondence or tracking, there are many intuitive assumptions based on reflective visible band properties.
- **Single Band (Gray Scale) and Multi Band (Color)**: Many computer vision algorithm works with gray-scale images, i.e. each pixel has one value associated with it. With the price of color digital camera dropping every year, 3 band color (RGB) in visible band images are being used more often. Ideally, for each pixel one makes observations in more than one spectral band which yields more information.

Among the differences of IR and visible band images we can distinguish them between technical problems that may improve as we get better IR camera technologies and the problems that are inherent to the IR properties and that can never be removed by using better IR cameras.

- **Technical Problems**: High noise, low spatial resolution. In part this still has to do with the IR properties but we have seen improvement over time with newer IR cameras. The problem of the camera emitting thermal radiation itself, though can not be completely eliminated, can be effectively reduced significantly by cooling down the camera substantially. In Figure 1 we see by Wien's displacement law the camera peak emission band can be shifted away from the ordinary room temperature thermal emission band. The low resolution of the pixel can be improved by micromachining technology to produce smaller pixels.
- **Inherent Problems**:
  - **History effects**: In visible band the brightness and color of one point reflects the (practically) instantaneous lighting and geometry conditions. In thermal IR the self-emission effects are important. Since the self emission depends on the temperature of the object surface and temperature change takes time (noticeably in human time frame), the strength of thermal IR radiation depends not only on the instantaneous states of the object and the environment, but also on the combined effects of the history of state

changes. This effect is inherent to thermal IR and cannot be "removed" by using better IR cameras.

- **Emission and reflection**: The importance of emission component in thermal IR radiation means that the fundamental formula of the reflectance photometry can not account for the whole scene. This is also a material property, not a camera property.
- **High Dynamic Range Differences**: This is not to say we get the benefit of high dynamic range, instead it is a property not easily captured by a camera. This is also a result of the importance of self-emission in the IR radiation. For most object surface except unoxidized metal the reflectance coefficients are low. So an object that only reflects light is much dimmer compared to a light source that emits light itself. In visible band most objects only reflects light so it is not uncommon to see pictures that contains no active light source. Which in turn means all objects brightness are roughly in the same level so we can find one exposure/gain level that spread all the brightness variations nicely to the full dynamic range of the camera. In thermal IR image all objects are emitting radiation and since the total radiation strength is proportional to the 4^{th} degree of object temperature, we expect to see very bright and very dark objects at the same picture almost every time. In this case either the bright object is over exposed or the dark object is under exposed, both of which means losing local texture information.

How exactly does these problems influence the performance of computer vision algorithms if we use the visible band version directly to IR images?

The noise and resolution problem will decrease the performance of 2D, single frame algorithms. Since other more complex algorithms are based more or less to the performance of the basic 2D, single frame algorithms, most of them will suffer indirectly. The reason is that most 2D, single frame algorithm are developed first under the simplified model of no noise and a smooth, continuous 2D surface model (infinitely high resolution). For example, the edge detection algorithms are based on differentiation gradient of a smooth 2D surface. High noise invalidates the smoothness model. Low spatial resolution itself means losing data, especially high spatial frequency data. When the raw data does not contain high frequency information, no algorithm can reconstruct them except guessing with prior knowledge. Low resolution also hurts the statistical assumption of sufficiently large amount of data. In 3D, multi-frame, and multi-band computer vision, when there is a need for correspondence, the most common automatic correspondence finder depends on local statistics of windows of textures. When resolution is low, the number of pixels representing each object of interests are low which makes statistics based method unstable.

The history effect invalidates the basic assumption of brightness constancy constraint of optical flow, which is the basis of multi-frame vision algorithms. The basic optical flow formula of

$$E_x u + E_y v + E_t = 0$$

where E, a function of 2D positions x and y and time t, is the brightness of one object point and Ex Ey Et represents the partial derivatives with respect to x, y and t. Also, u is the 2D apparent object velocity in the x direction and v is the 2D apparent object velocity in the y direction. This equation is valid in the assumption that brightness of an object is constant over time, i.e. dE/dt =0. This is never really true in visible band but without history effect it is not a bad assumption when dt is small The history effects invalidate the brightness constancy assumption in two ways. Because of the temperature dependence of thermal IR radiation, and it is common to see both extremely fast temperature change, like explosion or engine combustion, and very slow temperature change like the natural dissipation of heat. The extremely fast temperature change means the brightness can change significantly even between two consecutive video frames (usually 1/25 ~ 1/30 sec). The very slow dissipation of heat means there can be "ghost image" left behind after a hot or cold object moves, which by applying optical flow blindly can lead to ghost object detection.

The importance of emission in the contribution of brightness invalidates a whole family of shape from shading formula developed for the reflection dominated visible band images. In the reflection case the observed brightness is related to two angles, one is the angle between the surface normal and the direction of the light source, the other is the angle between the surface normal and the direction of observer, thus the term "BRDF"(Bi-directional reflectance distribution function). The thermal emission brightness, however, depends only on the angle between the surface normal to the direction of observer In addition, the emission is strongly dependent on surface temperature while BRDF is relatively insensitive to surface temperature, see Figure 10.

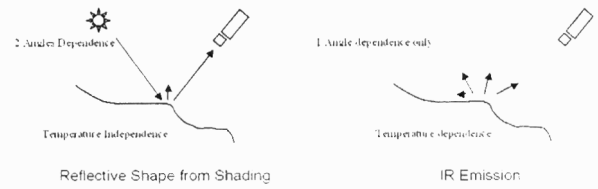

Figure 10 Differences in brightness formulas between reflection and self emission

Even more complicated is the fact that IR radiation is also reflected by surfaces and in situations where both reflection and self-emission are both important contributors to brightness. This can happen in the IR bands that are close to visible bands. In these cases entirely new equations must be used to interpret the observed brightness. We shall see such attempt in one of the paper we review.

For the high dynamic range problem, there is always a tradeoff between linear brightness value and revealing details of every part of the image. When one chooses to have linear pixel values (as required for photometry related algorithms), one can either stretch the brightness resolution to the entire dynamic range and losing the fine resolution in the local variations or one can preserve the local variations and leave the some of the region details saturated or decimated.

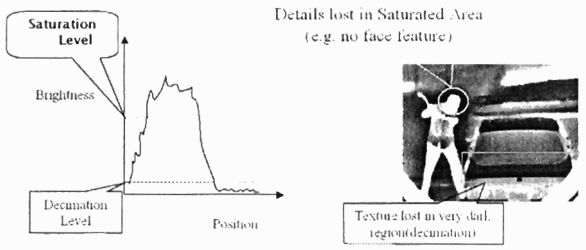

Figure 11 Saturation and decimation caused by high scene dynamic range can be good or bad depending on the task at hand.

The saturation and decimation is bad for texture based matching because the texture pattern inside saturated region is missing, but for some segmentation task saturation and decimation is good because they make the task easier or even saves the trouble altogether, see Figure 11.

4. IR image as single band image(s)

The first paper[B14] we review here try to do the tracking and object pose estimation in a video sequence using slightly modified algorithms developed originally for visible band image(s). The main reason for doing this is the capability to do the same target-tracking task at night or through smoke when visible band cameras are blinded.

The first difficulty they encountered is the low resolution and high noise of the images they had. The resolution is so low that there are only ¼ pixels in both x and y image axis with the resultant pixel count only 1/16 of that of ordinary visible band images they used to process. There are added difficulties that they are working on FLIR(Forward Looking Infra Red) image sequences, which is a military sensor installed on Army helicopters. The rough operation environment on a battlefield causes further image defects like dirty lens and broken pixel elements. Thus it is imperative for them to decrease the noise level and enhance the images before further processing.

The noise remover they use is median filter instead of mean filter. The main reason behind this is to fit the nature of the noise, however, there is also the added bonus of preserving high spatial frequency details in the images for pattern based ego-motion removal. Had they used mean filter, which has the side effect of suppressing high frequency image features, the image based ego motion removal might have failed or performed much worse.

The inherent high dynamic range nature of the scene brightness cause the "cold background" to appear very dark. They use the method "histogram equalization" to make the contrast more uniformly distributed. Note that doing this contrast adjustment destroys the original absolute brightness relationships. For example, if there are two points, one has twice the brightness of the other, this relationship will in general not hold after histogram equalization. The only relationship preserved by histogram equalization is the strength order, i.e. if one pixel is brighter than the other, it will still be brighter than the other pixel after histogram equalization. This is fine in this work because they are not using any photometry information, only the pattern to extract geometric information. On the other hand, histogram equalization may not work if there are more decimation in the "cold background" because histogram equalization only do remapping of existing brightness values and never create new brightness levels. If the brightness variation in the "cold background" is so weak that all background pixels have the same brightness values (or too few brightness values) then the approach of this paper[B14] will fail because they have not feature to do ego motion removal in the second stage.

Having reduced noise and equalized contrast, the images in the video sequence now have more visible band like image properties. However, there are still differences. As we can see clearly in the sample images they showed in Figure 12, even after enhancement the interior details of the vehicles like door, windows, ...etc are not distinguishable. Low spatial resolution, saturation due to strong emission, and different radiation models behind

emission and reflection all play a part in this phenomenon. The algorithm of this paper[B14] still managed to extract the moving object and their direction of movement due to the fact that the contours of the objects are still well preserved. The objects of interests are all equipped with hot running engines so that they stand out much brighter than the background. In this case the saturation both helped and causes trouble. It helped in the sense that a simple threshold in the brightness can segment the object region from the background. It causes trouble in that since the detailed pattern inside the object area are lost, it is difficult to identify the object type, like telling apart trucks from tanks or even distinguish between different brand of trucks.

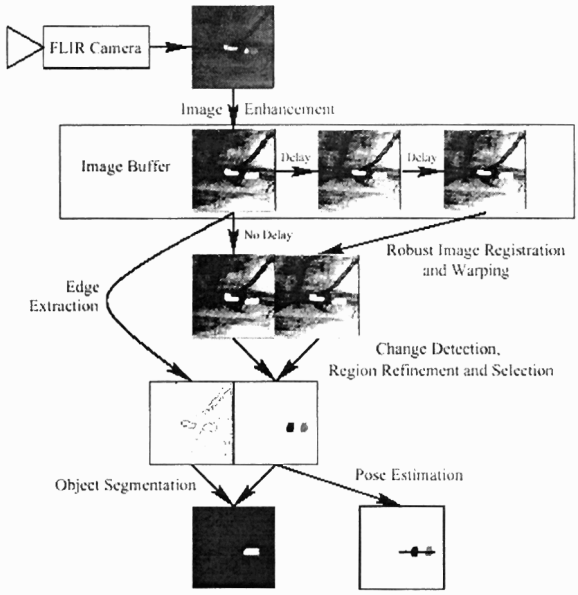

Figure 12 Overall process flow of IMO detection from airborne FLIR video

Since the video sequence is taken by a moving camera on a helicopter the apparent motion in 2D video sequence can be a combined result of object movement and the camera ego motion. The most accurate way to remove the ego motion of the camera is by using motion sensors but this is usually not practical in a battlefield or any uncontrolled field for that matter. Thus the algorithm incorporates an image based ego motion removal method that has been used successfully in visible band images. The core assumptions behind this image based ego motion removal are:

- IMOs(Independently Moving Objects) occupies only a very small portion of each image in the video sequences. This is why they can compensate for ego motion without separating the IMOs first. They just pretend there are no IMOs in the images when they are removing the ego motions.
- The background is stationary.
- There are distinguishable and non-ambiguous feature patterns all over the background so that correspondences of the same points on the background can be found between frames provided that the point is visible in the frames under consideration.
- The camera motion is smooth and the frame rate is high enough so that most scene points are visible in consecutive frames.
- The scene is practically flat so the apparent 2D image captured by the camera can be seen as an affine transformation of the planar scene. Obviously this is an a approximation but a pretty good one for most airborne images as long as the airborne camera is not too close to the scene.

All of the above assumptions are not foolproof, but they hold often enough to be useful. Although these assumptions were devised originally for visible cameras, we can see that most of them are not attached to properties unique to the visible band. Thermal IR band camera only makes the "distinguishable pattern" assumption more difficult to hold. The pattern can be buried in saturation region or decimation region. The pattern can also change with rapid thermal disturbances or fail to move with the object because of the history effect. Since in the sample images provided here this assumption still holds after contrast enhancement, the whole ego motion removal framework can be applied quite well without any modification. In other thermal IR sequences where you see these undesirable effects, the method can break down. For example, if the tank fires its main gun or the truck is hit by a bomb during the sequence, the tracking may be disrupted and never recover.

For the case that the algorithm works, the ego motion is removed by calculating optical flow field between two frames. Because the background is assumed to be stationary and the IMOs are assumed to occupy only a small portion of pixels, we expect to see a dominating displacement in the optical flow and this it can only be caused by the ego motion. Now that we have an initial estimate of the ego motion, we can use this knowledge to exclude flows that deviates a lot from this ego motion. These flows are likely caused by the IMOs. We can recalculate ego motion estimation using only the flows that are likely to be background flows to get a better estimate of the ego motion. The number of iteration can be increased if we have more time or computing power, this is where they claim to have scalability.

They also mention the use of multi-resolution (pyramid). Image Pyramid is a term created in the digital image processing community by [B3]. The idea is to build smaller images that is ½, ¼, … in each dimension of the original images, each of which contains only a band of

spatial frequency information. Since information in lower frequencies can be represented without loss in lower resolution, we get smaller and smaller replica of the original image as well. This aids in the application of the optical flow formula. Recall that optical flow formula is an equation concerning local image gradients and gradients are good approximations only for small displacement. Thus if a point is displaced several pixels away in the next frame, the optical flow formula does not work well. Having increasingly smaller replica of the original image solves the problem, provided that you can still find correspondent patterns in low frequency[B10], because any large displacement will eventually become a one-pixel displacement if you shrink the image enough. If you put all the different sized version of the same image one on top of each other, smaller ones on top of bigger ones, you have a pyramid, thus the name. Pyramid takes time to construct but once constructed the optical flow in all the levels can be computed in parallel and becomes very fast. Thus this method has become the core of many real time applications. Here the algorithm in this paper[B14] claims the potential to become real time because they use image pyramid.

Once the apparent ego motion is determined, they 'warp' one of the frame by affine transformation. They can do this without explicitly recovering 3D information because of the planar scene assumption. This is the main reason why they can use 2D affine transformation. The so called "2D is more robust than 3D" is not the main point, just a side effect. The real problem is that they do not have enough information to recover the 3D structure without the planar scene assumption. After the 'warp', ego motion is removed, and a simple pixel by pixel subtraction would reveal the IMO region, provided that the IMO has very different brightness than the stationary background. In this case thermal IR images actually work better than visible images in the subtraction because the IMOs are all much brighter than the background. Although the threshold value of how much difference in brightness count as an IMO region should not be too hard, they should have mentioned how the threshold value is determined. After initial thresholding, we may get holes in a big IMO region or small fragments of IMO region inside background region. Through prior knowledge of the types of video sequence they are likely to encounter (military FLIRs usually focus its field of view on only a few vehicles at a time), they use the morphological operation opening and closing to weed out fragments. The mask they use is 3 by 3 but no explanation of why this size is used are given.

At this stage we have several blobs of possible IMO candidates. They then use the shape and positional statistics(mean, variance, skewness and kurtosis of each of the coordinates) of each of the region. We can see that these descriptors only account for the silhouettes but not about the interior features like the shape of the doors, windows, ..etc. It is clear in even their sample images that we human observer can not tell which blob is a tank and which blob is a truck unless we read its captions. Thus the method they describe only serve to further eliminate unlikely IMO regions, not to distinguish target types like telling a truck from a tank. In this process they used many prior knowledge that are specific to their test images, like having many bad pixels and bad qualities around the bottom and right edges(other cameras may do better in these area). This may restrict their method to the particular camera they were working with.

Finally, they use the thermal IR image property of saturation to their advantage. Through observation they found the following phenomenon that can be used to locate the head and tail area of a moving vehicle:

Table 1 Head-Tail conditions

object is	in front of object	behind object
appearing	becomes brighter	not observable
moving visible	becomes brighter	becomes darker
disappearing	not observable	becomes darker
moving occluded	not observable	not observable

Notice that those 'not observable' parts are really not very useful. This means that when the IMO is partly outside the image or several IMOs overlap each other, their method can break down.

After detecting several possible heads and tails. We need to pair them in order to find IMO pose and movement directions. The heuristics used here is the shortest distance pairing, which can fail when multiple IMOs are close to each other. This is probably why they use edge information as well to reduce the chance of false pairing. The extraction of pose not only gives more information about IMOs, but they also serves to eliminate some spurious IMO candidates.

All we can reasonably get from the video sequence alone are the position of IMOs and their apparent 2D pose. However, in the paper[B14] they discussed methods to translate the apparent 2D pose into 3D pose by introducing external information, the height of the camera and the range to the target. This is possible in their specific application because the FLIR camera is often mounted on a helicopter or aircraft, on which there is always an altimeter. The problem is that the altimeter reading gives the pressure height, not height relative to the ground. Furthermore, the formula they are using is based on the assumption that the ground is level, not sloped. The distance information may be available because these army helicopters often has laser range finder on board. Overall, the 2D to 3D transformation is not accurate, but is better than nothing. These are only suggestions and no experiments are done.

The results shown in the paper[B14] are pretty good. On the other hand, all the sample images show good natures that may not be true in many situations. All the images contain no more than 2 IMOs and the 2 IMOs seldom overlap each other. There are no rapid temperature changes like firing weapons, starting engines, or being hit by enemy fire. There are no history effects like a hot vehicle start moving after staying at the same position for a while and leaving a hot print on the ground. There is, however, a interesting example showing that the method can some times detect motion better than human observer, as in Figure 13.

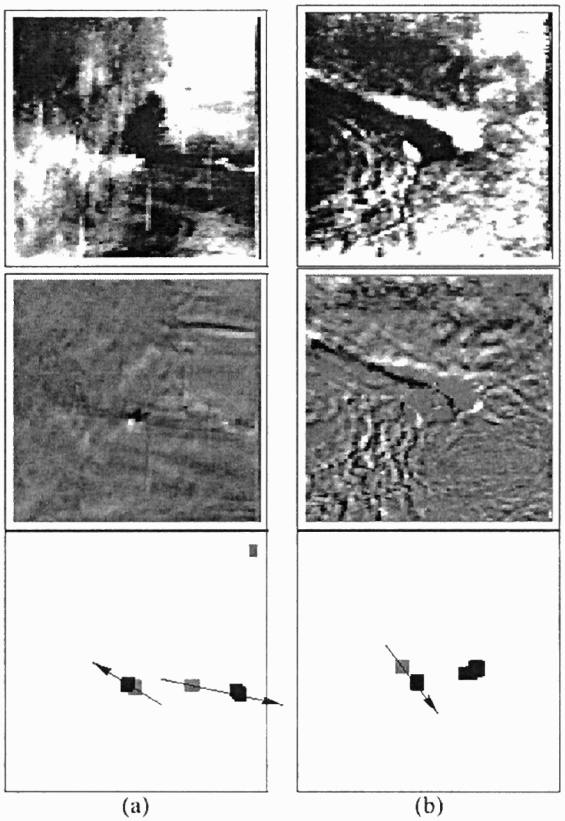

Figure 13 Example showing that the algorithm can sometimes outperform human observer

5. Single IR band multiple views with thermal energy considerations

While in visible band images the pixel values represent the combined effects of surface reflectance properties and the viewing geometry, in the thermal IR band images the pixel values represent predominantly the combined effects of surface temperature, surface emission properties and viewing geometry (though reflectance is still important in some cases). The methods that use only relative pixel values to extract edges ignore this underlying physics of pixel value. To make use of these additional information inferred by the absolute pixel values, the physical formula involving reflectance and heat transfer must be used. In doing so, however, the methods developed would be tied closely to the imaging modality and lose some of the generality for methods based only on relative pixel values.

The first problem encountered when using the physical formula for reflectance and/or heat transfer is that there are a lot of parameters that can not be extracted from the image pixel values alone. Further more, these parameters are very difficult or labor intensive to measure in many applications. Thus simplifying assumptions are invariably made in these methods. The most notable assumption in the visible band is that of Lambertian surface, which states that the surface reflectance values are the same in all direction. This assumption works very well in many applications, but on the other hand it fails often enough in other applications that remedies are needed

This problem is even more complex when we try to use the absolute values of IR images. In addition to the reflectance properties we have emission and heat transfer properties to worry about and they are all mixed up together. However, because of the potential reward of getting a lot more information about the objects in the image, more and more methods are developed using the absolute pixel values in the IR images.

The paper[B11] we review here represents one of the attempts. The main goal is to be able to distinguish object type and even models using only a few thermal IR (8~14 μm) images of the same object taken under different conditions. The idea is to extract some 'invariants' that are the same for a particular object type or model in all these different views. The prime candidates are the physical parameters like heat capacitance, emissitivities, ...etc that in principle should remain the same under normal environmental change and view change. However, it is really difficult to recover these values from a set of images alone, so an alternative is to extract some quantities that are related to these physical parameters.

To start, the principle of energy conservation provides the best starting point to write equalities. The added advantage is that energy is an additive physical quantity, i.e. the total energy is simply the sum of all the energy from each of its contributors. This implies that we get an equality that is linear in form. This will lead to linear differential equations or even linear equations that we can manipulate with linear algebra.

The formulation is based on the energy conservation of a passive (non-heat generating) surface element:

$$W_{abs} = W_{lost}$$

Equation 3 Heat absorbed equals heat lost

Here the Sun is assumed to be the only heat generator, thus: $W_{abs} = W_I \cos\theta_I \alpha_s$

Where

W_I : Solar irradiation when incident normal to the surface

θ_I : Angle between the Sun and surface normal

α_s : surface absorptivity

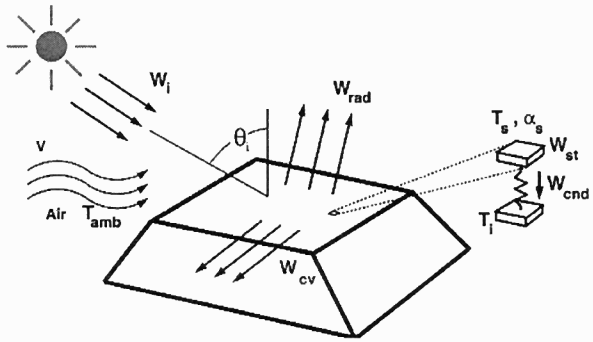

Figure 14 Energy exchange model at the surface of imaged object

There is an unstated simplifying assumption made here. The surface absorptivity should be a function of incident radiation direction, radiation wavelength, and to a lesser extent, surface temperature. Here it is used as if there is a single value for a surface under all direction, wavelength, and temperature.

Since they continue with the statement that the absorptivity is estimated from visible band reflectivity:

$$\alpha_s = 1 - \rho_s$$

Equation 4 Total or band reflectance and absorptivity for opaque object

and the 'fact' that 90% percent of solar radiation energy on the surface of the earth is in the visible band, we can infer that the 'absorptivity' here means 'total directional' absorptivity and the absorptivity and reflectivity properties are assumed 'Lambertian' or 'diffuse'.

Note that this also implicitly restrict the use of this formulation to sunny day solar radiation because surface reflectivity, even in the visible band only, can vary wildly with wavelength (thus we see a colorful world), the reflectivity for different incoming light (with different spectral distribution) should be different. Here the light source is fixed as solar radiation on earth's surface so they can get a single value for one surface. Also, the cos θ factor comes from modeling the Sun as a distant point source, and becomes meaningless in a cloudy day where dominant solar radiation is scattered light from the atmosphere and the angle θ loses its meaning.

The 90% statement is questionable as we can see from Figure 5, which comes from [B8], the solar radiation, even after atmospheric absorption, still has considerable energy in the IR region.

The absorbed energy must be either stored or dissipated. Thus four different ways where the energy absorbed by the surface might go are modeled:

$$W_{lost} = W_{cnd} + W_{st} + W_{cv} + W_{rad}$$

Equation 5 Four 'energy sinks'

W_{cnd} means power lost to heat conduction inside the surface, W_{st} means power stored inside the volume to raise temerature, W_{cv} means energy carried away by convection of the air, and W_{rad} means energy radiated back by the surface.

Each of these phenomena has well known formulas (under certain simplifying assumptions) describing it:

$$W_{cnd} = -k\frac{(T_s - T_{\text{int}})}{\Delta x}$$

Equation 6 Rate of heat flow conducted inward

Heat conduction is assumed to occur only from the surface toward the inner layer. The lateral conduction is assumed negligible and not modeled. In the formula k is thermal conductivity of the material (assumes uniform material type within unit surface and volume); T is the surface temperature; T_{int} is the interior temperature and dx is the distance below the surface. This is a reasonable assumption if

- The surface is smooth and free from shadow
- The surface material is uniform, not mosaic of very different materials.

For energy stored inside the elemental volume to raise surface temperature:

$$W_{st} = C_T \frac{dT_s}{dt}$$

Equation 7 Energy stored to raise surface temperature

Within unit volume the temperature is assumed uniform. C_T is the thermal capacitance of the material comprising the elemental volume, while dt is the unit time.

Heat convection is quite complicated phenomena but here a simplified version is used:

$$W_{cv} = h(T_s - T_{amb})$$

Equation 8 Convected heat transfer

T_{amb} stands for ambient temperature and h alone stands for all the combined effects of wind speed, thermophysical properties of the air, and surface geometry. A very crude formula, but suits the task here because we are practically unable to measure most of the parameter needed, e.g. wind speed and air temperature distribution, so a detailed formula is useless anyway.

Lastly, the surface radiates heat back into the environment:

$$W_{rad} = \varepsilon\sigma(T_s^4 - T_{amb}^4)$$

Equation 9 Energy lost by surface radiation

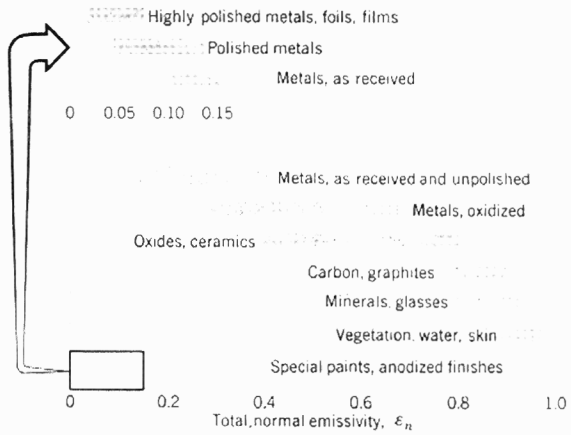

Figure 15 Representative values of the total normal emissivity

Here the formula left many questions unanswered. The Stefan-Boltzman relation for total radiated energy only holds for blackbody radiation. It is true that one can always use a total emissitivity as a multiplicative factor to fix the difference between a black body and a real body, but then the 'total emissitivity' value then changes with temperature. The example data found from one of the reference in this paper [B8] suggest against the approximation that the total-hemispherical emissitivity can all be approximated to be 0.9. See Figure 15, the oxidized metal has emissivities ranging from 0.25 to 0.7and oxides, ceramics has values from 0.4 to 0.8.

Apart from the 'not so good' assumption, the formula also implies that the surface absorptance for the ambient radiation is 100%, if the ambient temperature is the actual ambient temperature. It is true that you can see ambient radiation be written in this same form as black body radiation to make the equation look neat. But in that situation the 'ambient temperature' is NOT the ambient temperature measured with a real thermometer, rather, it is used, like the color temperature, a variable that is adjusted to fit in the formula. The color temperature of the sky is even higher than that of the sun but that does not mean the actual temperature of the air in the sky is that hot.

The distinction is not important if the symbol T_{amb} appears only once here. But now there is another T_{amb} in the convection formula and that T_{amb} is clearly the actual measurable ambient temperature, then this causes a confusion of symbols.

These problems may contribute to the 'not so good' results in the experiment section.

An analogy is made to the RC circuits, but the analogy is not exploited further in this work. The differential terms are treated as just one value and no differential equations are solved. To get the invariant the equality is rewritten into a linear form:

$$a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + a_5x_5 = \bar{\mathbf{a}}^T\bar{\mathbf{x}} = 0$$

Equation 10 Linear form for extracting invariants

Where:

$$\begin{aligned} a_1 &= C_T & x_1 &= -\frac{dT_s}{dt} \\ a_2 &= k & x_2 &= \frac{dT_s}{dx} \\ a_3 &= -(T_s - T_{amb}) & x_3 &= h \\ a_4 &= -\sigma\left(T_s^4 - T_{amb}^4\right) & x_4 &= \epsilon \\ a_5 &= \cos\theta_I & x_5 &= W_I\alpha_s \end{aligned}$$

Equation 11 Separation of "known" and "unknown".

The first question one asks is why divide parameters this way. The answer provided by the authors is that all quantities in the 'a' parts can be guessed with prior knowledge of the object and or derived from image pixel values with the aid of simplifying assumptions. It is really odd, however, that ε is given a guess of 0.9 and still listed in the 'x' unknown side. This raises the question that the acclaimed 5D thermophysical space may actually has only 4 degrees of freedom in their own logic system. (In reality the value ε varies with many parameters as we discussed earlier. This may have saved their experimental data from degenerating.)

Each point on an object imaged at a particular time and place yields one measurement vector

$$\bar{\mathbf{a}} = (a_1, a_2, a_3, a_4, a_5)^T$$

which is measured/guessed, and corresponding vector

$$\bar{\mathbf{x}} = (x_1, x_2, x_3, x_4, x_5)^T$$

which is never used nor measured.

Then comes the introduction of invariants. This is actually misleading because the "invariants" introduced are invariant only under a specific group of transformation, the linear transformation. The authors learned this from the works done in geometric invariants for computer vision [B1], in which the "invariant" are associated with a fixed shape and is unchanged under different view. In the geometric case the "points" or "point sets" physically retain their geometric relationship

in 3D space. However, here the abstract "thermophysical points" do not undergo simple linear transformation in the different pictures. In fact, if the two pictures are taken at the same time, e.g. two views from a stereo rig, the 'measurement vector' would remain the same provided the two thermal cameras are calibrated. This is because the equality derived from 'conservation of energy' involved no observer at all. The only angle θ_I is the angle between the surface normal and the direction of the sun. This angle has nothing to do with the angle of observation, maybe with the exception that the observer may block the sun.

As we discussed in the overview, one of the special properties of thermal images is that they have 'history effect'. This is in complete contrast to affine transformation which has no history effect at all. Put it in another way, the affine transformation operators are commutable, which means the order of application does not effect the final outcome. For thermal process, however, the order of 'transformation' is important as each 'path' incurs different energy and entropy changes. The 'shape' formed by the N points chosen from the an object in this abstract '5D' space may change in each image and the 'affine invariant' may not exist at all. Thus in general this is not the right way to derive 'invariant'.

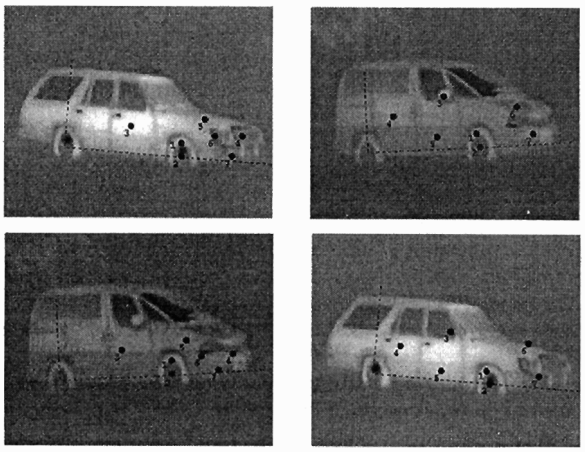

Figure 16 Top row: the car and van object types with points selected on the surface with different material properties and/or surface normals. Bottom row: assignment of point labels under erroneous hypothesis.

Furthermore, the '5D' space is neither a subset of the all possible 'thermophysical states' nor a superset of them. The set of all '5D' point that corresponds to legitimate physical states does not even form a linear space. The basic properties of a vector space require the existence of zero vector and inverse for each vector. However, the zero vector here is not legitimate physical state because it involves a material with zero heat capacitance, which can assume any temperature without heat input. With no zero element the inverse is ill-defined. In fact, a physical state can never have negative values for a_1, a_2, and a_5. Also, there are legitimate physical states that can NOT be represented by any point in the '5D' space. Any time the Sun is not directly visible in front of the surface point, the value of $\cos\theta_I$ and thus a_5 becomes undefined. This is not a uncommon situation as every night the Sun is not visible for any surface point. The experiment data listed in the paper[B11], though spans two days, have no night data. All data were taken between 9AM to 4PM. (See Figure 17, Figure 19, Figure 20) This signifies that they have discovered the problem in experiment but failed to recognize the structural failure of the whole algorithm. Even in day time, if the Sun is blocked by clouds, the $\cos\theta_I$ is ill-defined because the whole solar radiation model should be changed from a distant point source to that of diffuse illumination and very different spectral distribution. This is NOT captured by the '5D' space that is associated only with the point source lighting model pictured in Figure 14. The real physical invariant should be invariant with respect to the transformation between the set of all possible physical states, not the linear transformation on the artificial '5D' space. The linear transformation in the '5D' space can transform legitimate physical state into an illegitimate state, or vice versa, and some physical states can never be the output of the linear transformation. The real physical transformation relation can output any legitimate physical state so it is clearly different from the linear transformation discussed in this paper[B11].

VALUES OF THE I1-TYPE FEATURE USED TO IDENTIFY THE VEHICLE CLASS, TRUCK 1

Hypothesis: Data From:	Truck 1 Truck 1	Truck 1 Tank	Truck 1 Van	Truck 1 Car	Truck 1 Truck 2
11 am	-0.70	27.28	0.33	-	0.68
12 pm	-0.71	0.09	4.83	15.58	4.15
1 pm	-0.45	0.68	0.00	11.73	4.6e12
2 pm	-0.66	-1.00	-∞	71.23	-1.00
3 pm	-0.40	-1.00	-∞	-1.00	-1.00
4 pm	-0.54	-∞	∞	5.42	22.29
9 am	-0.68	1.38	-1.00	-6.66e14	-7.03
10 am	-0.45	-1.00	-∞	6.50	-

The feature consisted of point set {4, 7, 8, 10} corresponding to the points labeled in Fig. 6. The feature value is formed using the thermophysical model of truck 1 and the data from the respective other vehicles. When the feature is applied to the correctly hypothesized data of the tank it has a mean value of –0.57 and a standard deviation of 0.13. This I1-type feature produces a good stability measure of 4.5, and good separability between correct and incorrect hypotheses. The feature values for incorrect hypotheses are at least 3.2 standard deviations away from the mean value for the correct hypothesis.

Figure 17 TABLE 1 of the original paper

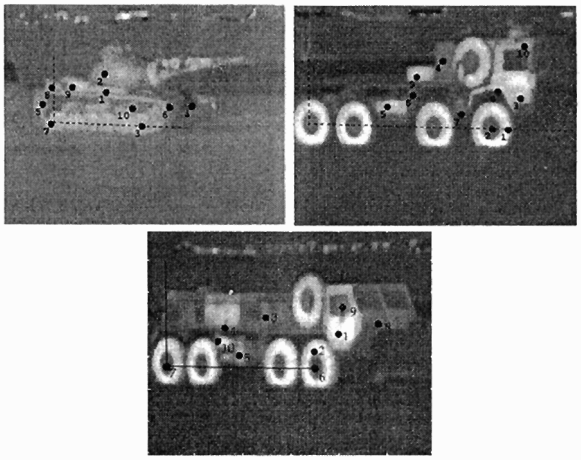

Figure 18 Three of the vehicles used to test the object recognition approach (clockwise from top left) tank, truck 1, and truck 2.

VALUES OF THE I2-TYPE FEATURE USED TO IDENTIFY TRUCK 1

Hypothesis: Data From:	Truck 1 Truck 1	Truck 1 Van	Truck 1 Car	Truck 1 Tank	Truck 1 Truck 2
11 am	-0.16	-193.47	-60.39	59.78	∞
12 pm	-0.28	-387.66	-143.09	20.20	∞
1 pm	-0.09	-525.77	-150.70	-11.23	-1.01e5
2 pm	-0.48	∞	-39.01	-29.38	1.02e5
3 pm	-0.96	-79.45	-1.7e5	-80.83	5.2e5
4 pm	-1.42	-498.51	50.76	∞	∞
9 am	-0.31	-454.87	-252.78	-9.38	29.9e5
10 am	-0.20	-13.90	-246.88	-7.78	∞

The feature consisted of point set {2, 3, 5, 9}, corresponding to the points labeled in Fig. 4. The feature value is formed using the thermophysical model of truck 1 and the data from the respective other vehicles. When this feature is applied to the correctly hypothesized data of truck 1 it has a mean value of –0.49 and a standard deviation of 0.46. The feature values under an incorrect hypothesis are at least 15.2 standard deviations away from the mean value for the correct hypothesis.

Figure 19 TABLE 2 of the original paper

With the '5D' linear space model fundamentally deviate from physical reality, any further derivation based the non-physical '5D linear space' can only be classified as heuristic method, not physics based. In the very limited experiment setup, there is, however, a possibility that in the limit of very short period of time and small environmental change, the result of linear transformation may not differ too much from the result computed by real physical formula. This would impose serious limitation to the application to the whole method, like the weather must be stable with no strong wind and the either the time between views are short or the area must be a remote, undisturbed area. For otherwise even the energy conservation equation would not hold. The authors never mention these fundamental shortcomings in the paper[B11]. Their experiments are all performed under most favorable conditions. The data were acquired over one good weather day(if there were cloudy or rainy period the angle of the sun loses meaning and the measurement matrix could not be taken) in an undisturbed test ground. The vehicles are all parked, not running their own engine and no heavy traffic near by during the testing. Still, the "invariant" values can vary quite a bit, one value may be 7 times greater than the other. This is possibly caused by the fact that the big temperature difference between day and night in a clear day is too much for the linear approximation.

VALUES OF A FEATURE OF TYPE I2 USED TO IDENTIFY THE TANK, FOR CORRECT AND MISTAKEN HYPOTHESES

Hypothesis: Data From:	Tank Tank	Tank Van	Tank Car	Tank Truck 1	Tank Truck 2
11 am	0.88	-648.99	∞	-	25.37
12 pm	0.88	198.98	∞	-38.19	-9.47
1 pm	1.29	-339.60	∞	-	7.68
2 pm	3.25	-154.42	-6.55	-	-7.31
3 pm	4.12	-290.50	-12.48	-	-7.84
4 pm	2.60	-339.92	-20.41	∞	∞
9 am	1.07	-2.42e5	∞	-13.19	-5.31
10 am	0.70	∞	-	-4.64	-3.57

The feature consisted of point set {3, 5, 7, 9}, corresponding to the points labeled in Fig. 4. The feature value is formed using the thermophysical model of tank and the data from the respective other vehicles. When this feature is applied to the correctly hypothesized data of the tank it has a mean value of 2.52 and a standard deviation of 1.82. The feature value under a mistaken hypothesis is at least 3.3 standard deviations away from the average value under the correct hypothesis.

Figure 20 TABLE 3 of the original paper

Coming back to the ideal situation where the linear approximation works, the method is still strongly dependent on hand picking the points and a rather detailed prior knowledge of the scene. Out of the five components in the 'measurement vector', three of them are purely hypothesized with detailed prior knowledge, with the rest still depend quite heavily on simplifying assumptions. With only 4 points used, at most 4 different types of material can influence the value of the 'invariant'. Thus the same object can have quite different 'invariant' values for the same image if the other 4 points are chosen. Further, it may be situations that no points can be chosen for this method because every visible point on the body of the object may all have the same temperature with the environment and a_3 and a_4 are zero for all points.

The experiment data shows good 'inter-class' separation for the 'invariants' but this is expected because the way they do 'inter-class' is to put wrong estimated parameters into the system. With wrong values for the parameters even the equality derived from conservation of energy would not hold in general and all the conditions for ranks fall apart. It is evident in that many determinants are degenerate and the computed 'invariant' becomes infinity.

This method suggested here, even without the fundamental theoretical flaws, would be very impractical for actual applications. For the number one reason people

use thermal IR camera is to see hot things in the dark, at night. This method fails in exactly these situations, it can not see hot things at any time, fails at night, even in day time if it is cloudy. It is even more restricted than a visible band camera.

6. IR as part of multiple band image extending visible band color methods

The more spectral band we can observe an object, the more know about the object. Even if the intensity images of two or more different bands of some object are exactly the same, it still gives provides us a specific signature of the object: maybe most other object we want to distinguish always have different image intensity patterns in different bands. Although it is agreed that more information can potentially improve object recognition, more information also means more data to process for each object. How to efficiently incorporate multi-band image data to aid our computer vision task is an still evolving research area called sensor fusion.

As we have seen from two examples that uses only IR image or images, we get new information about surface thermal properties but the rich reflectance properties in the visible band is not available. This leads to many trade-offs. If we have simultaneously many spectral band images covering a wide range of EM bands then that will give us more information than individual bands. In order to use many bands of information in a cohesive way, the first natural place to look for method is again the visible band. Inside the visible band, a 'color' camera often captures 3 different band information, which when viewed by human eyes would be perceived as Red, Green, and Blue. Although visible band is only a very narrow band compared to the whole EM spectral band, there still exists great variations of spectral properties inside the visible band that the crude division of Red, Green, and Blue bands does not capture all the details. However, since the human eye color vision are approximately based on detecting the RGB bands, it suffices for 'color cameras' to capture the 3 band information and then reproduce them using a mixture of these colors(in the case of printing, the mixture of 3 complementary colors paints Yellow, Cyan, and Magenta). In order to classify 'color' information, a 3D coordinate is established for RGB and later more intuitive alternative, HSV, YUV, ...etc. These are the most studied ways of processing multiple bands in a cohesive manner, thus if we can apply this method beyond the visible band then we can save the effort of developing new band fusion scheme for each different bands.

Depending on the application at hand, the most important advantage derived from using multiple band images may be different. For displaying images for human viewers, the combination of IR and visible may supplement each other because one mode works better in daylight while the other better at night If a Low-Light-Visible camera provides the visible band, then the visible band can supplement the normal reflective texture information while the IR band provides emissive properties of the objects in the image. The representation that best suits human viewer is not a trivial issue By displaying a single monochrome image, it is always necessary to throw out some information because we are displaying two pixel values with only one pixel value If we display a pseudo color image, there is room for more information but the choice of how to map pseudo color to image information is tricky. A color scheme that makes perfect sense for one human operator may appear very confusing for another human operator.

We have mentioned the fact that the names of sub-bands inside the IR region differs greatly between fields of study, even between individuals. In the following we will use the terminology used in the paper [B13] for convenience of discussion. See Table 2

Table 2 IR sub-band definition in the paper

Spectral Band	Spectral Range
Visible (VIS)	0.4-0.75
Near Infrared (NIR)	0.78-1.0
Short-wave Infrared (SWIR)	1-3
Mid-wave Infrared (MWIR)	3-5
Long-wave Infrared (LWIR)	8-12

The paper [B13] provides a very interesting table for approximate flux levels incident on Earth's surface for bands in the visible and IR and during different time and moon conditions:

Table 3 Approximate flux levels on Earth's surface

	VIS	NIR	MWIR	LWIR
Daytime flux clear sky	1.5×10^{17}	1×10^{17}	4×10^{15}	2×10^{17}
Nighttime flux full moon	1.5×10^{11}	-	2×10^{15}	2×10^{17}
Nighttime flux starlight	1.5×10^{9}	9×10^{9}	2×10^{15}	2×10^{17}

Since in the two extremes reflected component and emitted component dominates respectively, in the middle there are transitions between the two mode of dominance and exhibits diurnal variation of texture contrast

There is also a list of bands and pair-wise correlation charts from a satellite ERIM M-7 The trends are high correlation between visible bands, mild correlation between visible and SWIR/MWIR, and mild anti-correlation between visible and LWIR. Since these are satellite images, the imaging condition is somewhat different from that on Earth's surface e.g. much thicker atmosphere between the imaged object and the camera,

different angle of views, ...etc. The results may not be directly applicable to computer vision tasks on Earth's surface.

Table 4 Spectral response of ERIM M-7 sensor

Band	Spectral Response (μm)
1	.36 - .38
2	.42 - .44
3	.44 - .46
4	.46 - .49
5	.48 - .53
6	.52 - .57
7	.57 - .62
8	.64 - .70
9	.71 - .87
10	.90 - 1.03
11	1.23 - 1.30
12	1.54 - 1.64
13	2.04 - 2.30
14	3.38 - 5.38
15	9.38 - 11.59
16	7.17 -12.11

Table 5 Correlation coefficients of the 16 bands for ERIM M-7 data

Band																
#1	1.00															
#2	0.97	1.00														
#3	0.97	0.99	1.00													
#4	0.96	0.99	1.00	1.00												
#5	0.94	0.97	0.98	0.98	1.00											
#6	0.92	0.96	0.98	0.98	0.99	1.00										
#7	0.86	0.90	0.91	0.92	0.96	0.95	1.00									
#8	0.86	0.91	0.92	0.93	0.96	0.96	0.99	1.00								
#9	0.74	0.78	0.81	0.81	0.83	0.88	0.84	0.85	1.00							
#10	0.70	0.74	0.76	0.77	0.78	0.82	0.78	0.80	0.98	1.00						
#11	0.77	0.82	0.83	0.84	0.87	0.87	0.89	0.90	0.88	0.90	1.00					
#12	0.83	0.88	0.89	0.90	0.92	0.90	0.93	0.94	0.77	0.76	0.95	1.00				
#13	0.87	0.91	0.93	0.93	0.95	0.94	0.94	0.95	0.78	0.74	0.91	0.98	1.00			
#14	0.17	0.22	0.23	0.24	0.26	0.21	0.26	0.27	0.24	0.32	0.44	0.44	0.36	1.00		
#15	-0.49	-0.49	-0.50	-0.48	-0.48	-0.54	-0.48	-0.48	-0.40	-0.26	-0.20	-0.26	-0.37	0.59	1.00	
#16	-0.45	-0.44	-0.45	-0.43	-0.43	-0.49	-0.43	-0.43	-0.36	-0.22	-0.15	-0.21	-0.32	0.59	0.96	1.00
	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	#15	#16

About the spectral reflectivity and emissitivity, the function form presented in the paper[B13] is $\rho(\lambda)$ and $\varepsilon(\lambda)$. This form dose not include the dependence on direction(s). Although BRDF (2 directions) and DHR (one direction) are mentioned, the graph presented still does not contain any directional dependence of the reflectivity. We can thus assume that a diffuse assumption for reflectivity and emissitivity is made and the data plotted might be Normal Hemispherical or Normal Normal Reflectance values (under diffuse assumption these values are related by a constant multiplicative factor so the shape of spectral plots are similar). See Figure 21.

There is also another evidence that they used diffuse assumption because of the equation:

$$\varepsilon(\lambda) = 1 - \rho(\lambda)$$

Equation 12 This equation holds only under diffuse light or surface

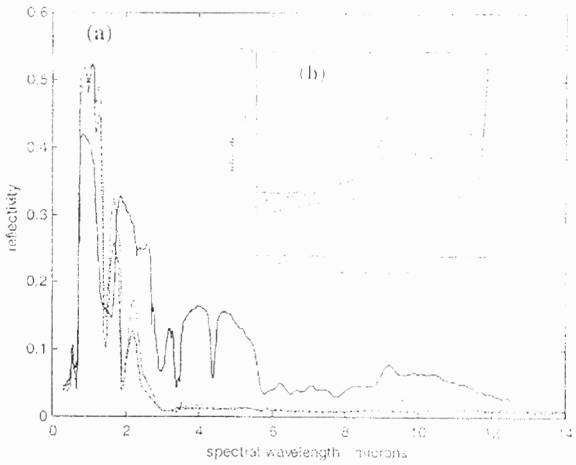

Figure 21 Comparison of measured reflectivity coefficients of green paint (solid), conifers(dashed-dot), and grass(dotted). Inset (b) is a detailed plot of the visible region

This equation actually comes from the conservation of energy of an opaque surface plus the equation that equates $\alpha(\lambda)$ and $\varepsilon(\lambda)$. As we have discussed, emissivities are defined differently than absorptivities and reflectivities. The equality is only for the numerical value. The only assumed true relation is:

$$\alpha_{\lambda,\theta}(\lambda,\theta,\phi) = \varepsilon_{\lambda,\theta}(\lambda,\theta,\phi)$$

Equation 13 The only equality that is always true between α and ε

Here the subscripts are used to denote the value is per wavelength and per unit solid angle based, because the values may be angle based but with no angle dependence (diffuse) or wavelength based but no wavelength dependence (gray). To get the coefficients that are not angle based we need to do the following integration:

$$\alpha_{\lambda}(\lambda) = \frac{\int_{\cap} \alpha_{\lambda,\theta}(\lambda,\theta,\phi) I_{i;\lambda,\theta}(\lambda,\theta,\phi) d\omega}{\int_{\cap} I_{i;\lambda,\theta}(\lambda,\theta,\phi) d\omega}$$

$$\varepsilon_\lambda(\lambda) = \frac{\int_{\cap} \varepsilon_{\lambda,\theta}(\lambda,\theta,\phi) I_{b;\lambda,\theta}(\lambda,T) d\omega}{\int_{\cap} I_{b;\lambda,\theta}(\lambda,T) d\omega}$$

$$= 2\int_{\cap} \varepsilon_{\lambda,\theta}(\lambda,\theta,\phi) d\omega$$

Equation 14 The derivation of hemispherical coefficients

Here the integration is done over the hemisphere about dω, the differential solid angle. The subscript i under I indicates incident irradiation and the subscript b under I indicates black body radiation. Since black body radiation by definition is diffuse(angle independent), thus it can be taken out of the integration both in the numerator and the denominator. The only two conditions that the equation:

$$\alpha_\lambda(\lambda) = \varepsilon_\lambda(\lambda)$$

would hold is that either the incident irradiation is diffuse or both the surface absorptivity and emissivity are diffuse. Since the incident irradiation can take any form, the surface must be diffuse.

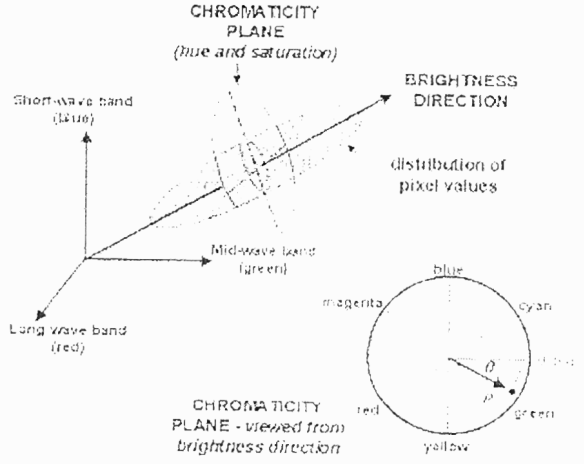

Figure 22 Simple vector representation of an RGB color image, showing the transform coordinates, i.e., the principle component direction and the two chrominant axes in the chromaticity plane

Because of the RGB bands chosen to collect 'color' information in the visible band, mathematically the image signal data can be represented as an array of three-values for each point in the image. This is crude but is quite sufficient for human to discern quite a lot of information.

The most convenient way to integrate the information in all 3 bands is to do a coordinate transform that separates 'brightness' from 'color hue and saturation'. As in Figure 22, often the distribution of pixel values in a typical image will cluster around a prolate spheroid extending from the origin. This direction of pixel concentration is taken to be the brightness direction and the plane orthogonal to this direction is used as the chromaticity plane. On this plane either a polar coordinate of hue(0 to 360 degrees) and saturation(a non-negative value) or in rectangular coordinates.

The idea here is to emulate this representation by substituting the 3 visible bands RGB with some other set of 3 bands. This is quite natural because R, G, and B represents long, middle and short wave length bands inside the visible band. This extension is just to stretch the overall band coverage. Many of the convenience this formulation brings in the visible band can be enjoyed by the IR counterparts, however, some properties may not be the same as the authors expected.

The model starts with n-dimensional coordinate system representing n-bands of sensor outputs. Each pixel has n values that can be expressed as:

$$v_k = \int_{\lambda_{lower}}^{\lambda_{upper}} I(\lambda)\rho(\lambda)\eta_k(\lambda) d\lambda$$

Equation 15 Sensor outputs for reflectance dominated bands

and

$$v_k = \int_{\lambda_{lower}}^{\lambda_{upper}} R_T(\lambda)\varepsilon(\lambda)\eta_k(\lambda) d\lambda$$

Equation 16 Sensor outputs for emission dominated bands

where

- v_k : signal out of a detector for band k
- $I(\lambda)$: spectral distribution of illuminant
- $R_T(\lambda)$: spectral distribution of black body at temperature T.
- $\rho(\lambda)$: spectral reflectivity of the surface
- $\varepsilon(\lambda)$: spectral emissivity of the surface
- $\eta_k(\lambda)$: detector spectral response
- λ_{lower} λ_{upper} : lower and upper limit of a band

This representation already incorporates several big simplifying assumptions:

- **Diffuse surfaces:** ε and ρ has no angle dependence.
- **Atmospheric effects ignored:** no term related to atmospheric effects(scattering, air light, ...etc) is involved.

Still, more simplifications are needed to get the often-used linear form. The simplifying assumption is:

- **Spectral dependence functions of surface reflectivity can be approximated by a linear sum of finite number (n) of basis functions:**

$$\rho(\lambda) = \sum_{j=1}^{n} \sigma_j S_j(\lambda)$$

Equation 17 Approximated reflectivity spectral dependence

We know this simplification will introduce errors in general, but it does introduce big simplification in computation and algorithm design, and many applications do not require high precision in color values. In this simple system we can have

$\vec{v} = \Lambda\vec{\sigma}$ and

$\vec{\sigma} = \Lambda^{-1}\vec{v}$

where the matrix Λ can be derived by substituting Equation 17 into Equation 15 and arrange terms:

$$\Lambda = \begin{pmatrix} \int I(\lambda)S_1(\lambda)\eta_1(\lambda)d\lambda & \int I(\lambda)S_2(\lambda)\eta_1(\lambda)d\lambda & \int I(\lambda)S_3(\lambda)\eta_1(\lambda)d\lambda \\ \int I(\lambda)S_1(\lambda)\eta_2(\lambda)d\lambda & \int I(\lambda)S_2(\lambda)\eta_2(\lambda)d\lambda & \int I(\lambda)S_3(\lambda)\eta_2(\lambda)d\lambda \\ \int I(\lambda)S_1(\lambda)\eta_3(\lambda)d\lambda & \int I(\lambda)S_2(\lambda)\eta_3(\lambda)d\lambda & \int I(\lambda)S_3(\lambda)\eta_3(\lambda)d\lambda \end{pmatrix}$$

Equation 18 Linear Transformation matrix between sensor value and reflectance coordinates

So far this is for one particular pixel, one particular surface point being observed. To really get convenience we would like to have only one matrix Λ that can be used for all pixels/(scene points) in the same image. The problem comes from the term $I(\lambda)$. In a typical scene there are always shadows. Further, there are inter-reflections between surfaces. These all cause $I(\lambda)$ to vary from point to point and as a result the matrix Λ should vary from point to point. In many applications obtaining individual matrix Λ is practically impossible. So further simplifying assumptions are introduced:

- **Ignore inter-reflection:** or assume the effects are much weaker than that of the main illuminant.
- **Ignore shadow:** or assume the number of pixels in shadow is relatively small.

To this point, after so many simplifying assumptions, the model already has 'color constancy' built in. If we know the exact function form of $I(\lambda)$ and all the spectral dependent terms to compute the matrix Λ, then from the observed value vector v we can always get the intrinsic reflectance parameter vector σ, which by assumption of this model is invariant.

In actual applications, none of the spectral distributions in the model are known! Thus the practical color constancy reduces to only the situation that the strength of illuminant is scaled up or down(while the spectral distribution of the illuminant remains unchanged). Then by one further assumption:

- **Gray World:** this is not to say that everything in the scene looks gray, which is not a very useful assumption. Rather, it is to assert that, because of the diversity of material reflectance, it is possible that in a scene the reflectance values are scattered almost uniformly in all possible values. In other words, there are pixels that reflect strongly on red, on green, on blue but none of them is dominating. This assumption can fail, for example, in scenes rich with green vegetation.

The 'color' of the global illuminant is recovered using principle component analysis mentioned in the paper[B13], that statistically treat all points as a whole as a gray reflecting surface and recover the 'color direction' in the color space, which is often the RGB space. HSV is just a change of coordinate that defines two components H and S on a plane orthogonal to the brightness and normalized them so it is more convenient to exploit the 'color constancy'.

The thermal IR band, where emission strength dominates, can be fit into similar model but there are advantages and problems unique to it. Substituting emission counter parts in Equation 17 and Equation 18, we again encounter the problem of getting a global version of $R_T(\lambda)$ that can be used for the whole image. The good thing is that there is only one family of spectral distribution, the Planck's black body radiation, and the general form is known. However, to get the exact spectral distribution we need to know the surface temperature T. This is tricky. Because for scenes that we want to use thermal IR camera, there is usually some interesting temperature differences between scene points. So if we use the simplifying assumption that all surface has the same temperature, we get to continue the argument of 'color constancy' but we can not use the results on most of the interesting scene, like inspecting high temperature factory machinery. In the case we do proceed with color constancy argument, the assumptions change to:

- **Thermal Equilibrium:** surface temperature is constant across the scene. At first, this may seems to be a good approximation for many situations. It is not. The reason is that earth's surface is regularly heated up during the day and cools down during the night. There are always some materials that heat up and cool down faster or slower than others so there are almost always temperature gradient differences. Furthermore, there are a lot of chemical reactions and physical movements taking place all around the world. Chemical reactions generate or absorb heat, physical movement creates heat by friction or from gravitational energy (like rock or water falling down). In fact only in a man-made closed system that thermal equilibrium can be maintained for a long period of time.
- **Ignore any reflection:** this is a restatement of what we start with, but more emphasis on inter-reflection.

Note that since reflection is assumed to insignificant, shadow never arises as a separate problem. The thermal equilibrium assumption can be somewhat relaxed, not by assuming constant temperature, but by arguing that when the temperature differences are small enough, the 'functional shape' of $R_T(\lambda)$ does not change too much.

This argument has the added advantage in that it also enables the statistical recovery of $R_T(\lambda)$ because unlike $I(\lambda)$, $R_T(\lambda)$ never simply scale up and down, it only varies with surface temperature with the result of changing both intensity and shape of function. This is evident in that we have Wien's displacement law about the shifting of function peak with temperature change, see Figure 1. In our argument it is possible then to have simple scaling of $R_T(\lambda)$ just like that of $I(\lambda)$, and thus the statistical recovery of the 'color direction of radiation' is possible under the assumptions:

- **Gray World:** Similar to that of reflectance case, just substitute ρ by ε. However, it is similarly prone to fail, maybe even more so than the reflectance counterpart.
- **Small surface temperature variation between the scenes:** as discussed this is to avoid big 'color change' associated with temperature change.

As a side note, the scheme here is not directly applicable to the usual one-band mono thermal IR images. There must be at least 3 bands or more in the emission dominated bands in order to have meaningful 'color constancy' problem. For mixing up reflectance dominated bands and emission dominated bands or even bands that both phenomena are important, the coordinate scheme extends naturally, but the 'combined color constancy' is infeasible because now the illuminant and radiation varies independently, structurally changing the transformation matrix Λ between scene to scene. This change involves more than one parameter and can not be recovered by a simple principle component analysis. Also, the model becomes more artificial because now both the assumption groups necessary for reflection and emission must be instated, furthermore, as the bands involved grows wider, the 'linear sum of basis function' deviates more from reality. This is also this question of how many reflectance basis functions should be used, but this is not discussed further in the paper[B13].

At the last paragraph the author actually talked about some thing not related to color constancy. With only 2 bands, there is only one plane and it is not possible to get a chromaticity plane, only a line is possible. The subject discussed is actually that about Equation 12, i.e. since emissivity has this tendency to vary in the negative direction of reflectivity, use 'black is hot' display looks more like a visible band image and makes pilot(a human viewer who is used to see visible band reflectance images) feel more comfortable. For 'color constancy' computation, this only amount to a sign change for some of the transformation component and is not essential.

Up to now it is taken for granted that all the pixels from each band are registered, i.e. they are collecting light from the same points for corresponding pixels. This is easier in the case of all visible band image because all bands can share the same optics and even the same detector array, like many color visible band cameras. However, when combining bands with great wavelength differences, it is usually the case the optics useful in one band becomes opaque for the other. Furthermore, the sensor arrays are not sharable. It is still possible to design optically registered device, like we have done in MOOSE project, but the author is right about one thing, for hand adjustment it is difficult to do the alignment to high precision. However, this can be done in factory and fixed. When the demands for multi-spectral camera grows and the IR components get cheaper, such product will come to consumer market in no time.

The 'software solution' of 'rubber sheeting works fine for faraway scene. Since this paper[B13] comes from a Naval research lab, the camera they were working with are probably mounted on airplane or warships and looking for objects at least miles away. In that case the 'parallax' effect caused by bore-sight arrangement is negligible and the scene can often be approximated as planar scene. For indoor close range view that objects are only a few meters away, parallax effects are important(we get stereo vision out of it) and can not be 'corrected' by software.

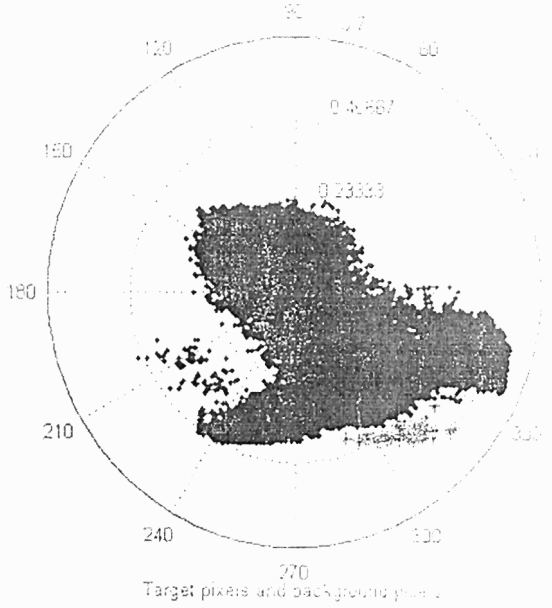

Figure 23 Chromaticity plane scatterplot of vehicle pixels (asterisks) and natural vegetation background pixels (diamonds)

The last application mentioned in this paper[B13] is improving performance for object-background separation. By mapping pixel values from VIS, MWIR and LWIR of a scene to a 3D coordinate treat them as if they are RGB image. On the chromaticity plane(Figure 23), it is clear that the man-made object stands out as they all cluster in the lower right corner of the plot.(this should be the ground truth produced by human observer.) They did not provide the original 3 band images so we do not know how representative it may be. However, we can see this separation happening because the vehicle most probably contains metal, or plastics that has very different thermal properties than the vegetation.

The performance of an algorithm that classify a pixel into two categories can be described by a graph called 'ROC curve', receiver-operator-characteristics curve, see Figure 24. The axes are respectively, false alarm rate and missed detection rate. False alarm means declaring a background pixel as object pixel. Missed detection means declaring a object pixel as background pixel. In a scheme of simple thresholding, setting a high standard for a pixel to be classified as an object will increase the missed detection rate while lowering the false alarm rate and vice versa. For one particular algorithm, the curve of a series of thersholding values will most likely a curve going from upper left to lower right. If a new algorithm that performs better, then both the missed detection and false alarm goes down and the curve as a whole will be closer to the origin, which represents perfect case.

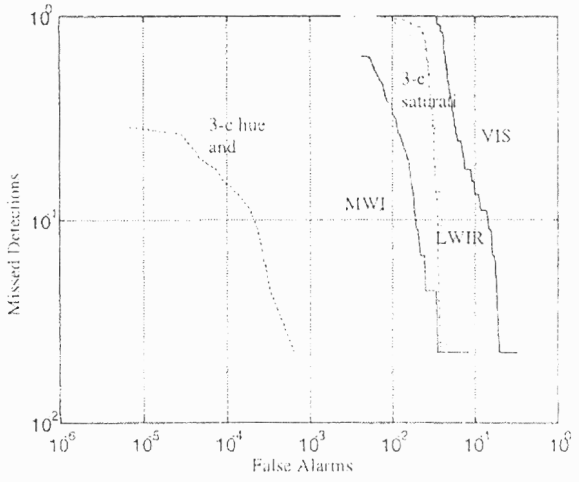

Figure 24 ROC curves for detection of a vehicle in a background of natural vegetation. Single band results are for individual pixel intensities and three band color results are for pixel values in the chromaticity plane with prescribed hue and saturation intensities.

Since here the authors are only demonstrating the advantage introduced with 3 band colors, the methods of separating background and object are all very simple thresholding. The methods that select pixels only with hue between 300 and 330 and then making a thresholding on saturation can be expected to perform well from Figure 23 alone and it does shows order of magnitude improvement in Figure 24 as it is much closer to the origin.

The results shown in this particular picture are promising but we can not draw a conclusion from only one example. However, qualitatively we can expect this scheme to work well in the specific task mentioned, to separate vehicles from background using general argument in physical properties. Vehicles are often composed of materials that are quite different in thermal properties to vegetation. Further, vehicles have engines and move around often so are heated by frictions, too. Further, the paint used by people only look good in visible band only because human eyes can not tell differences beyond the visible range so no efforts are put into making the paints look like vegetation in the IR bands. All these can be well exploited by the 'color scheme' proposed here. However, it should be noted that this scheme is not connected to 'color constancy'. If another set of pictures were taken of the same scene, but with different lighting condition, e.g. at night, then we may still see good separation, but all the 'colors' will change because the thermal temperature do not change linearly with visible light level.

To sum up, the paper[B13] presents some promising ideas of extending visible band color into IR bands. Some of them work as described but there are also some misconceptions involving color constancy. Using the color scheme in one shot to do segmentation is fine, but tracking 'color constancy' is impractical. There are simply too many unknowns involved and the authors failed to recognize that. It is also interesting to see the authors providing numerical statistics of light flux level in day and night and different moon phase, but they did not say how the data is collected or where it comes from, which leaves a lot of questions.

7. Summary

Extending existing visible band image computer vision techniques to infrared band images with no or minor modifications potentially saves a lot of redevelopment time. However, we must be very careful in examining the assumptions, simplifications behind each of the methods. In some cases the problem disappears in infrared, in some cases the problem is worse in infrared. One of the most important cause in these differences is the different dominating modes of brightness generation. In visible band the brightness strength of a scene point comes primarily from reflection. The difficulties involved in this mode are that there are at least two angles involved and for man-made light source the spectral distribution can vary wildly. For infrared band there are only one

angle and one family of reference radiation function (the Planck distribution). However, the close connection with surface temperature introduces a lot of complexities, like history effects. We must examine the functional forms as well as the typical range of parameter variations. Sometimes a reasonable simplification exists within small parameter ranges. Finally, it should be noted that terms like 'constancy', 'invariants', ...etc often comes with a long list of assumptions and may be applicable inside only a specific domain. Misuse of the concept can lead to unpredictable conclusions.

References

[1] *Geometric Invariance in Computer Vision* Cambridge, MA: The MIT Press, 1992.

[2] *Handbook of Computer Vision and Applications (Sensors and Imaging)* San Diego, CA: Academic Press, 1999, pp. 1-623.

[3] Adelson, E. H., Anderson, C. H., Bergen, J. R., Burt, P. J., and Ogden, J. M. Pyramid methods in image processing. RCA Engineer 29[6]. 1984. Princeton, NJ, RCA Corporation.
Ref Type: Magazine Article

[4] Eisberg, R. and Resnick, R., *Quantum Physics of Atoms, Molecules, Solids, Nuclei, and Particles*, 2 ed. John Wiley & Sons, Inc., 1985, pp. 1-713.

[5] Gaussorgues, G., *Infrared Thermography*, 3 ed. London, England: Chapman & Hall, 1994, pp. 1-508.

[6] Hecht, E., *Optics*, 3 ed. Reading, MA, USA: Addison Wesley Longman, Inc., 1998, pp. 1-694.

[7] Horn, B. K. P., *Robot vision* New York USA: McGraw-Hill, 1986, pp. 1-509.

[8] Incropera, F. P. and DeWitt, D. P., *Fundamentals of heat and mass transfer*, 2 ed. New York : John Wiley & Sons, 1985, pp. 1-802.

[9] Lerner, E. J. Uncooled IR detectors move into the mainstream. Laser Focus World 37[5], 201-204. 2001. Tulsa, OK, PennWell.
Ref Type: Magazine Article

[10] Mandelbaum, R., Salgian, G., and Sawhney, H. Correlation-based estimation of ego-motion and structure from motion and stereo. 1998. Kerkyra, Greece, IEEE Computer Society Press. Proceedings of the International Conference on Computer Vision. 1998.
Ref Type: Conference Proceeding

[11] Michel, J., Nandhakumar, N., and Velten, V., "Thermalphysical Algebraic Invariants from Infrared Imagery for Object Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 1, pp. 41-51, Jan.1997.

[12] Owens, K. and Matthies, L. Passive Night Vision Sensor Comparison for Unmanned Ground Vehicle Stereo Vision Navigation. 6-21-1999. Institute of Electrical and Electronics Engineers, Inc. Proceedings of the IEEE Workshop on Computer Vision Beyond the Visible Spectrum. 6-21-1999.
Ref Type: Conference Proceeding

[13] Scribner, D., Warren, P., and Schuler, J. Extending Color Vision Methods to Bands Beyond the Visible. 6-21-1999 Institute of Electrical and Electronics Engineers, Inc Proceedings of the IEEE Workshop on Computer Vision Beyond the Visible Spectrum. 6-21-1999.
Ref Type: Conference Proceeding

[14] Strehl, A. and Aggarwal, J. K. Detecting moving objects in airborne forward looking infra-red sequences. 1-10. 6-21-1999. Fort Collins, CO, USA, IEEE Computer Society Press. IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications. 6-21-1999.
Ref Type: Conference Proceeding