TNO Human Factors
Postbus 23, 3769 ZG Soesterberg

Memo TNO-TM 2002-M040

From:	Dr. A. Toet
To:	Lt.Col. Carl Kutsche, Chief Human Life Sciences, London, UK
Subject:	Final Report Contract No. F61775-01-WE026 on "Perceptual Optimisation of Fused 3D Color Night Vision"
Date:	5 July 2002

1. AIM OF THE PROJECT

The aim of the project was to develop new optimal multimodal image visualisation techniques, based on improved image fusion schemes that optimally exploit and combine the perceptually relevant information from each of the individual nighttime image modalities into a single (stereo) image. Ergonomically combined (or fused) multimodal image representations are of great value for military surveillance and recognition systems.

2. BACKGROUND

Modern electro-optical imaging sensor suites are designed to expand the conditions under which humans can operate. The sensors that constitute a suite typically provide complementary information. A functional piece of equipment should provide a single (fused) image that leads to optimal perceptual awareness in most environmental and operational conditions (in order to "Own the weather " or "Own the night"). A human operator using a suitably combined or fused representation of multisensor imagery should be able to acquire a complete mental representation of the perceived scene, which should result in full situational awareness. For a given observation task, performance with fused imagery should at least be as good (and preferably better) as performance with the individual image modality that yields the optimal performance for that task. Multisensor image fusion should therefore combine and preserve in a single output image all the perceptually important signal information that is present in the individual input images. Knowledge of the nature of the features in each of the input images that determine observer performance is required to develop new multimodal image visualisation techniques that are optimally tuned to human visual perception.

3. APPROACH

Systematic psychophysical experiments were formed in the laboratory to assess the perceptual advantages of fusing the individual image modalities. The information obtained from these experiments was used to devise an optimal image fusion scheme that combines the relevant features in a single image.

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From – To)*
16-07-2002	Final Report	15 June 2001 - 18-Oct-02

4. TITLE AND SUBTITLE

Perceptual Optimization of Fused 3D Color Night Vision

5a. CONTRACT NUMBER
F61775-01-WE026

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Dr. Alex Toet

5d. PROJECT NUMBER

5d. TASK NUMBER

5e. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
TNO Human Factors Institute
Kampweg 5
NL 3769 DE Soesterberg
The Netherlands

8. PERFORMING ORGANIZATION REPORT NUMBER
N/A

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

EOARD
PSC 802 BOX 14
FPO 09499-0014

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
SPC 01-4026

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This report results from a contract tasking TNO Human Factors Institute as follows: The contractor will analyze low-light visible, uncooled long-wave infrared, and laser radar imagery showing vehicles and men deployed in military relevant nighttime scenarios, both in the open and in the trees, during nighttime conditions. The contractor will perform assess the relevant features in multiple image modalities. An optimal image fusion scheme will then be developed which combines the relevant features in a single stereo image. Observer studies will be designed to reveal the distinguishing characteristics of the targets of interest in each of the following image dimensions: 2D Individual image modalities (visible vs. thermal vs. 2D fused); 3D viewing of geometry only vs. 3D viewing with image/fused textures; 3D viewing of point data vs. surface rendering; static vs dynamic 3D scenes (motion pathways, structure from motion); advantage of viewing in stereo with/without 3D manipulation; resolution (pixels on target) reduction; obscuration (natural foliage or systematic deletion of data).

15. SUBJECT TERMS
EOARD, Human Factors, Image Processing, night vision,

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18, NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Valerie Martindale, Maj, USAF
UNCLAS	UNCLAS	UNCLAS	UL	102	19b. TELEPHONE NUMBER *(Include area code)* +44 (0)20 7514 4437

4. PROBLEMS ENCOUNTERED

MIT Lincoln Labs (MIT-LL) has previously demonstrated real-time Short Wave Infra Red (SWIR)/ Long Wave Infra Red (LWIR)/ Laser Radar (LADAR) fusion, and has recently demonstrated low-light visible/uncooled LWIR/ LADAR fusion. MIT-LL has proposed to investigate how this fusion can be further improved by 3D stereo visualization of views from multiple aspects obtained as a target moves around, as the sensor is displaced, or a combination of both. TNO-HF has acquired and investigated most of these techniques. Together with new sensing modalities like LADAR and high-speed image processing techniques, these displays can be used to present complex natural scenes in 3D and in real-time.

At the start of the project we expected to receive a visual + ladar database from MIT Lincoln Labs. Due to personnel changes (Prof Waxman, our contact person, has left MIT-LL) the images could not be delivered, except for one set. We have therefore based our analysis of the perceptual presentation of Ladar information on the following sources:

- A single MIT-LL data set
- Technical knowledge of Ladar
- Knowledge of viewing 3D images
- Knowledge of color coding and color fusion.

We restricted the rest of our studies to nightvision (intensified and thermal) imagery that was readily available. This restriction is not serious, since most of the methods developed in this project will also be applicable to other image modalities.

5. STUDIES PERFORMED

5.1 Effective use of stereoscopic disparity for display of non-literal imagery from multiple sensors (Appendix A)

5.1.1 Rationale

Modern electro-optical imaging sensor suites are designed to expand the conditions under which humans can operate. The sensors that constitute a suite typically provide non-literal images (i.e. images that do not resemble a photograph taken of the same scene). Examples are long- and short-wave infrared, laser radar, and synthetic aperture radar systems. The different sensors typically produce complementary information. A functional piece of equipment should combine and display the perceptually relevant details from each of the individual image modalities in an intuitive way, thereby leading to optimal perceptual awareness in most environmental and operational conditions. Adding stereoscopic disparity to produce a 3D representation of fused imagery provides the user with a better understanding of the spatial layout of the scene. Moreover, it enables the detection of targets which would otherwise remain camouflaged in 2D imagery. This applies particularly to non-literal images, since their non-intuitive nature makes them harder to interpret in the first place

5.1.2 Approach

In recent years an increasing number of 3D visualization displays have become available. In the course of this project TNO Human Factors has acquired and investigated most of these techniques.

5.1.3 Results

This study resulted in a set of guidelines we think provide a good basis for the effective use of stereoscopic disparity for the display of non-literal imagery.

These guidelines were published as a conference proceedings, which is added to this final report as Appendix A (Kooi & Toet, 2001).

5.1.4 Conclusions
Stereoscopic depth gives non-literal images a more intuitive appearance. 3D visualization of non-literal images should therefore be further investigated.

5.2 Detection of dim point targets in cluttered maritime backgrounds through multisensor image fusion (Appendix B)

5.2.1 Rationale
We investigated the problem of small low observable (dim) target detection in cluttered maritime infrared images. This problem is of interest in many applications such as ocean surveillance (e.g. oil spills), search and detection (e.g. speedboats, dinghies, both in civil and in military contexts) and rescue (e.g. swimmers), remote sensing, floating mine detection, etc.

Multispectral IR imaging techniques are frequently deployed in maritime operations, for instance to detect floating mines or to find small dinghies and swimmers during search and rescue operations. However, maritime backgrounds usually contain a large amount of clutter that severely hampers the detection of dim point targets.

5.2.2 Approach
We developed a simple algorithm that deploys the correlation between target signatures in two different (3-5 and 8-12 μm) IR frequency bands to reduce the amount of clutter. First, both individual IR bands are filtered with a morphological opening top-hat transform to extract small details. Second, the resulting detail images are thresholded to produce binary detail images, representing potential target areas. Third, a fused detail image is obtained by taking the intersection (logical AND) of both binary IR detail images. Details that appear in both IR bands remain in this fused detail image, whereas a large fraction of uncorrelated noise details is filtered out. Remaining noise details can be removed by taking into account the temporal characteristics of the target signatures and by using a priori knowledge of structure of the scene and the size of potential targets. The method was tested on two image sequences showing a maritime scene with three kayaks approaching from far away. The scenario was registered in the 3-5 μm and 8-12 μm IR frequency bands, and in the visual range.

5.2.3 Results
The results show that the proposed multispectral image fusion technique improves the detection of dim point targets and significantly reduces the number of false alarms in cluttered maritime backgrounds.

The results were published as a conference proceedings, which is added to this final report as Appendix B (Toet, 2002a).

5.2.4 Conclusions
Fusing the result of a morphological tophat transform of two different (3-5 and 8-12 μm) IR frequency bands enhances the detection of small point targets in cluttered maritime backgrounds.

5.3 Using depth to indicate potential targets (Appendix C)

5.3.1 Rationale
The retionale of this study is the same as that of the previous one, namely the fact that small point targets are extremely hard to detect in cluttered maritime backgrounds.

5.3.2 Approach
We have developed and evaluated a method for highlighting potential targets while keeping them in their context (Appendix C: Hogervorst & Toet, 2002). Potential targets are put at a different depth plane from the rest. This allows the observer to restrict search to a limited set of potential targets, and at the same time keeps the targets within their surrounding contexts (to keep track of the global position within the surroundings). We have evaluated this method with an experiment in which observers had to search for a ring within a large number of C's (with different orientations) acting as distractors.

5.3.3 Results
The results show that when half the distractors are placed in a different depth plane, the search times are (nearly) as fast as when the number of distractors is halved. This shows that observers can restrict search to a single plane.

It shows that the method of putting potential targets in a different plane can speed up the search process while keeping the targets within their context.

5.3.4 Conclusions
The use of differences in depth to indicate potential targets significantly reduces search time.

This method is especially suited for highlighting potential elements in a full color image.

5.4 Perceptual evaluation of different nighttime imaging modalities (Appendix D)

5.4.1 Rationale
The two most common nighttime imaging systems either display emitted infrared (IR) radiation or reflected light, and thus provide complimentary information of the inspected scene. A human operator using a suitably combined or fused representation of IR and (intensified) visual imagery may therefore be able to construct a more complete mental representation of the perceived scene, resulting in a larger degree of situational awareness

5.4.2 Approach
Human scene recognition performance was tested with images of nighttime outdoor scenes. The scenes were registered both with a dual band (visual and near infrared) image intensified low-light CCD camera (DII) and with a thermal middle wavelength band (3-5 μm) infrared (IR) camera. Fused imagery was produced through a grayscale pyramid image merging scheme, in combination with two different color mappings. Observer performance in visual detection and recognition tasks was tested for each of the (individual and fused) image modalities.

5.4.3 Results
The results show that DII imagery contributes most to global scene recognition (situational awareness), whereas IR imagery serves best for the detection and recognition of targets like humans and vehicles. Grayscale fused imagery yields appreciable performance levels in most conditions. With an appropriate color mapping, color fused imagery yields the best overall scene recognition performance. However, an inappropriate color mapping significantly decreases observer performance compared to grayscale image fusion.

The results were described in a paper that has been submitted for publication to the journal "Displays", and which is added to this final report as Appendix D (Toet & Franken, 2002).

5.4.4 Conclusions
The deployment of a DII system in addition to a 3-5 μm IR system through image fusion can increase the performance of human observers when the color mapping relates to the nature of the visual task and the conditions (scene content) at hand.

5.5 Paint the night: applying daytime colors to nighttime imagery (Appendix E)

5.5.1 Rationale
Modern nightvision systems are increasingly deployed in military operations and for surveillance applications. The two most common nighttime imaging systems either display emitted infrared (IR) radiation or reflected light, and thus provide complementary information of the inspected scene. A combined or fused display of these camera signals can therefore provide a more complete representation of the scene.

False color mappings are frequently used to visualise multiband medical and satellite imagery. These mappings are not very useful for the application in multiband nightvision systems, since the resulting images are often hard to interpret because of their unnatural color representation. In this project we developed a simple scheme to give multiband nighttime images a natural daytime color appearance.

5.5.2 Approach
The method employs a transformation to a principal component space that has been derived from a large ensemble of hyperspectral images of natural scenes. In this decorrelated color space the first order statistics of natural color images (target scenes) are transferred to the multispectral nightvision images (source scenes). The

only requirement of the method is that the composition of the source and target scenes are similar to some extent. Hence, the depicted scenes need not be identical, they merely have to resemble each other.

We applied the method to RGB false color nighttime imagery recorded both with a dual band (visual and near infrared) image intensified low-light CCD camera (DII) and with a thermal middle wavelength band (3-5 μm) infrared (IR) camera.

5.5.3 Results
The results convincingly show that the method effectively gives nighttime imagery a daytime appearance.

The results were described in a paper that has been accepted for publication to the journal "Displays", and which is added to this final report as Appendix E (Toet, 2002b).

5.5.4 Conclusions
The simple color mapping presented here gives fused nighttime imagery an appearance that resembles normal color daytime images. The resulting full color representation of nighttime scenes may be of great ergonomic value by making the interpretation (segmentation) of the displayed scene easier (more intuitive) for the observer.

The method is simple and can be applied in real-time.

5.6 Enhanced visualisation of nightvision imagery through fusion with Ladar depth maps (Appendix F)

5.6.1 Rationale
The fusion of range data from a ladar with visual and thermal imagery may help to segment targets from their local background.

5.6.2 Approach
Since we did not obtain the visual and ladar database, we based our analysis of the perceptual benefits of the inclusion of Ladar information in nightvision imagery on the following sources:
- A single MIT data set
- Technical knowledge of Ladar
- Knowledge of viewing 3D images
- Knowledge of color coding and color fusion.

5.6.3 Results
The results of our analysis indicate that mismatches between a (Ladar) depth map and the structural image content will inevitably result in poor viewing comfort and an unpredictable visibility of depth differences.

The results were described in a memo which is added to this final report as Appendix F (Kooi & Toet, 2002).

5.6.4 Conclusions
We conclude that color coding of Ladar depth information into other image modalities appears the most promising approach.

6. RESOURCES USED

6.1 Subjects
A total of 14 subjects participated in the studies that were performed under Contract No. F61775-01-WE026.

6.2 Summary of resources expended for the study
The total costs of the project amount to US $ 87,400.00, from which US $ 24,900.00 was paid for by contract No. F61775-01-WE026 from the European Office of Aerospace Research and Development, Air Force Office of Scientific Research, Air Force Research Laboratory, and the remaining part was largely paid for by Senter, Agency of the Ministry of Economic Affairs of the Netherlands, and some other funding sources.

7. PROJECT SUMMARY

7.1 Findings

In Section 5.1 we derived a set of guidelines for the effective use of stereoscopic disparity for the display of non-literal (i.e. out of band or non-visual) imagery.

In Section 5.2 we developed a simple filter that effectively extracts dim point targets from multiband IR maritime images.

In Section 5.3 we showed how the use of different (stereo) depth planes for the visual presentation minimises the search time required to detect these targets in their cluttered maritime backgrounds.

In Section 5.4 we found that the fused representation of double band intensified imagery with 3-5 μm thermal imagery increases the performance of human observers in visual detection and recognition (situational awareness) tasks when the color mapping is intuitively easy to interpret (relates to common practice).

In Section 5.5 we developed a simple color mapping that gives fused nighttime imagery a normal color daytime appearance.

In Section 5.6 we conclude that color coding of Ladar depth information into other image modalities appears the most promising approach.

7.2 Problems experienced

The only problem encountered in the course of this project was the fact that we were not able to obtain any ladar imagery.

8. CONCLUSIONS

8.1 Results in relation to objectives

As stated in Section 1, the aim of the project was to combine the perceptually relevant information from each of the individual nighttime image modalities into a single (stereo) image that is intuitively easy to interpret. This goal has been achieved for the most part, as we will argue in the following.

In the study described in Section 5.1 we derived a set of guidelines for the effective use of stereoscopic disparity for the display of non-literal (i.e. out of band or non-visual) imagery. Therefore, we know know which factors are critical in the design of stereo presentations of non-literal imagery, and we can therefore optimise these factors in the design process.

In Section 5.2 we developed a simple filter that effectively extracts dim point targets from multiband IR maritime images. Therefore, we can produce fused image representations (mono or stereo) in which these details are optimally enhanced for visual perception.

In Section 5.3 we showed how the use of different (stereo) depth planes for the visual presentation minimises the search time required to detect these targets in their cluttered maritime backgrounds. This demonstrates the effetiveness of stereo presentation of perceptually relevant information (the targets were the perceptually relevant features that were extracted with the method developed in Section 5.2).

In Section 5.4 we found that a color fused representation of double band intensified imagery with 3-5 μm thermal imagery increases the performance of human observers in visual detection and recognition (or situational awareness) tasks when the applied color mapping is intuitively easy to interpret (relates to common practice). We also showed that the application of counter-intuitive color representations can seriously degrade observer performance. Therefore, color mappings that provide natural image representations are of great ergonomic value.

In Section 5.5 we developed a simple color mapping that gives fused nighttime imagery a normal color daytime appearance. The resulting full color representation of nighttime scenes may be of great ergonomic value by

making the interpretation (segmentation) of the displayed scene easier (and therefore probably less time consuming or faster) for the observer. The method is computationally simple and can be applied in real-time.

In this project we did not have the opportunity to evaluate the natural color mapping scheme that was designed in the study described in Section 5.5. Although it is very likely that this mapping will significantly increase observer performance, subject studies need to be performed to test this assumption. We propose to perform such tests in a future study.

In Section 5.6 we analysed the properties of Ladar imagery in relation to other image modalities, and we concluded that conclude that color coding of Ladar depth information into other image modalities appears the most promising approach to the fusion of both data types.

In summary, we achieved most of our objectives by
- designing guidelines for the effective use of stereoscopic disparity for the display of non-literal imagery,
- showing that stereoscopic display of targets in cluttered maritime imagery reduces search time,
- showing that intuitive color representations enhance observer performance in detection and recognition tasks, and
- designing a simple color mapping that gives fused nighttime imagery a normal color daytime appearance.

8.2 Relevance for the US Air Force

The results are relevant for:

airborne systems (like an Air Force Special Ops Command):
- the findings of this study can for instance be used to enhance the display of night vision imagery and flight symbology on HUD's
- a natural color representation of nighttime imagery may increase situational awareness for pilots (it will be easier to distinguisk sky from ground, and different types of terrain, water, or urban areas can more easily be recognized)

outward looking ground systems: a natural color representation of nighttime imagery makes it easier to detect targets and to distinguish different types of terrain or man-made objects therein

maritime surveillance systems: small targets like approaching rubber dinghies can be perceived more quickly and with less effort in cluttered maritime thermal imagery when using an enhanced stereo presentation that incorporates the tophat filter and fusion technique developed in this study.

9. REFERENCES

Hogervorst, M.A. & Toet, A. (2002). Using depth to indicate potential targets.*In Preparation.* (Appendix C).

Kooi, F.L. & Toet, A. (2001). Effective use of stereoscopic disparity for display of non-literal imagery from multiple sensors.*Proceedings of the 4th International Conference on Information Fusion (Fusion2001)* (pp. WeC2-25-WeC2-30). International Society of Information Fusion. (Appendix A).

Kooi, F.L. and Toet, A. (2002). *Enhanced visualisation of nightvision imagery through fusion with Ladar depth maps*Soesterberg, The Netherlands: TNO Human Factors. (Appendix F).

Toet, A. (2002a). Detection of dim point targets in cluttered maritime backgrounds through multisensor image fusion. In W.R. Watkins, D. Clement & W.R. Reynolds (Ed.), *Targets and Backgrounds: Characterization and Representation VIII.* Bellingham, WA: The International Society for Optical Engineering. (Appendix B).

Toet, A. (2002b). Paint the night: applying daylight colors to nighttime imagery. *Displays, Submitted.* (Appendix E).

Toet, A. & Franken, E.M. (2002). Perceptual evaluation of different image fusion schemes. *Displays, In Press.* (Appendix D).

This material is based upon work supported by the European Office of Aerospace Research and Development, Air Force Office of Scientific Research, Air Force Research Laboratory, under contract No. F61775-01-WE026.

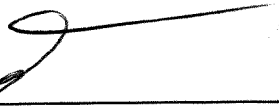
Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the European Office of Aerospace Research and Development, Air Force Offcice of Scientific Research, Air Force Research Laboratory.

The Contractor, TNO Human Factors, hereby declares that, to the best of its knowledge and belief, the technical data delivered herewith under Contract No. F61775-01-WE026 is complete, accurate, and complies with all requirements of the contract.

Date 9-07-2002 Authorized Official: 

Dr. A. Toet

I certify that there were no subject inventions to declare as defined in FAR 52.227-13, during the performance of this contract.

Date 9-07-2002 Authorized Official: 

Dr. A. Toet

Appendix A

Effective use of stereoscopic disparity for display of non-literal imagery from multiple sensors

F.L. Kooi A. Toet

TNO Human Factors

Kampweg 5

3769 DE Soesterberg, The Netherlands

{kooi,toet}@tm.tno.nl

Abstract – Modern electro-optical imaging sensor suites are designed to expand the conditions under which humans can operate. The sensors that constitute a suite typically provide non-literal images (i.e. images that do not resemble a photograph taken of the same scene). Examples are long- and short-wave infrared, laser radar, and synthetic aperture radar systems. The different sensors typically produce complementary information. A functional piece of equipment should combine and display the perceptually relevant details from each of the individual image modalities in an intuitive way, thereby leading to optimal perceptual awareness in most environmental and operational conditions. Adding stereoscopic disparity to produce a 3D representation of fused imagery will provide the user with a better understanding of the spatial layout of the scene. Moreover, it may enable the detection of targets which would otherwise remain camouflaged in 2D imagery. This applies particularly to non-literal images, since their non-intuitive nature makes them harder to interpret in the first place. In recent years an increasing number of 3D visualization displays have become available. TNO Human Factors has acquired and investigated most of these techniques. This paper presents the guidelines we think provide a good basis for the effective use of stereoscopic disparity for the display of non-literal imagery.

Keywords: stereoscopic depth, disparity, image fusion, visual comfort, situational awareness.

1 Introduction

1.1 Non-literal imagery

Modern electro-optical imaging sensor suites are designed to expand the conditions under which humans can operate ("all-time, all weather"). The sensors that constitute a suite typically provide non-literal images (i.e. images that do not resemble a photograph taken of the same scene). Examples are long- and short-wave infrared (LWIR and SWIR), laser radar (LADAR), and synthetic aperture radar (SAR) systems.

The number of different imaging modalities that become available is rapidly growing. Each of these image modalities has its own characteristics, which are usually not intuitive to a human observer. Operators therefore usually require extensive training before they are able to recognize and interpret the information provided by these images [1]. However, the techniques to visualize new types of imagery such that human operators can easily absorb the information therein are developing at a much smaller pace. The human factor therefore becomes increasingly the bottleneck in image processing and interpretation.

Different sensor modalities typically produce imagery containing complementary information. Modern sensor suites therefore combine a set of complementary sensors to cover most of the relevant part of the electromagnetic spectrum. A functional piece of equipment should combine and display the perceptually relevant information (details) from each of the individual image modalities in an intuitive way. This will allow the human operator to easily absorb all the information presented by the different sensors and to construct an internal representation (a mental model) of the overall state of the scene. Only then can the presented imagery lead to optimal perceptual awareness in most environmental and operational conditions.

An additional challenge is the non-literal or unnatural nature of many of the new sensor modalities. Given that the visible wavelengths form only a small part of the electromagnetic spectrum, it is reasonable to expect that natural (visible) images will soon form only a small subset of all image information available. It is also reasonable to expect that non-literal imagery is usually harder to interpret and recognize than standard (visible spectrum) imagery. A good example is SAR imagery. SAR images are typically hard to interpret and recognize. SAR images

contain detailed (highly resolved) structure. However, they lack the pattern of depth cues found in natural images [2].

We are therefore faced with the double challenge of *(1)* visualising different types of non-literal imagery and of *(2)* fusing the complementary information from different sensor modalities into a single coherent and easily interpretable display.

Adding color, motion, or depth to the imagery may in principle aid the human visual system in its detection, recognition, and interpretation tasks.

Previously we have explored the use of color as a visualization tool. We found that when color is used appropriately to code the sensor origin of image details multisensor image fusion schemes lead to enhanced situational awareness and improved target detection [3-6]. However, the inappropriate use of color in image fusion can lead to severe degradations of observer performance [7, 8]. Also, for non-literal images there generally exists no one-to-one mapping between contrast and the spectral reflectance of a material. The goal of producing non-literal images with an appearance similar to natural daytime color images can therefore never be fully achieved. The options are *(1)* to settle for a single mapping that works satisfactory in a large number of conditions, or *(2)* to adapt (optimize) the colour mapping to the situation at hand. However, the last option is not very attractive since a different colour mapping for each task and situation tends to confuse observers [5, 9].

In this paper we focus on the depth dimension, since stereoscopic depth holds the potential to make non-literal images appear more intuitive.

2 3D image presentation

The number of display types is rapidly expanding and now includes several methods to display 3D (stereoscopic) images. It is therefore of interest to explore to what extent 3D presentation can aid the interpretation of fused multisensor images. Recent work shows that it is not trivial that this will be the case. Watkins *et al.* [10] for example show that the benefit of 3D for target detection in natural scenes is only very limited. A second reason to carefully examine the true added value of stereoscopic 3D is that the ability to see stereoscopic depth varies between subjects. It has been reported that approximately 4% of the population does not see stereoscopic depth at all, usually as the result of a lazy (amblyopic) eye. Another 10% experiences difficulties seeing stereo [11]. Here we first present some basic facts about depth vision under normal viewing conditions followed by the restrictions placed by four types of sensors.

2.1 Depth cues

It is generally accepted that the human visual system uses approximately ten different "depth cues" to interpret the 3D layout of flat (2D) images. These cues include:

1. Stereopsis
2. Motion parallax
3. Monocular depth cues
    a. perspective
    b. occlusion
    c. object size familiarity
    d. horizon & relative height in the scene
    e. shading
    f. texture
    g. atmospheric contrast reduction
    h. color

In the normal world all cues are consistent, re-enforcing a consistent depth interpretation. If the cues do not agree, several types of interactions may occur [12]. The important thing is that the viewer looses his sense of confidence, the image is harder to interpret, and occasionally the 3D interpretation can be erroneous, leading to visual illusions. The correct depth percept may be enforced by presenting a movie rather than a picture (motion parallax) or by adding stereoscopic depth. In the next section we work this argument out for specific sensors.

2.2 Depth cues in raw sensor images

Each type of sensor limits the depth cues in its own particular way. We give some examples.

2.2.1 CCD camera

A standard CCD camera provides all monocular depth cues but no stereopsis. A video camera also provides motion parallax if either the camera or the object is moving.

2.2.2 Intensified night vision goggle

A night vision goggle does not provide color. Moreover, it plays tricks with the shading, texture, and atmospheric contrast reduction cues [13]. As a result, the depth percept provided by a night vision goggle is substantially poorer than that provided by a CCD camera. Shading and texture are less visible due to the relatively poor MTF of the intensifier tube. The perception of atmospheric contrast reduction is regularly invoked by the low contrast appearance of the image. This may lead the user to think to be farther away than is the case (an illusion). The fixed angular size of halo's (local overexposure of light) also acts as an erroneous depth cue.

2.2.3 Thermal camera

A thermal camera does not provide color and shading. Texture is impoverished due to the generally limited resolution and distance may be overestimated as a result of the generally high contrast imagery (an illusion). A stationary thermal camera therefore only reliably provides four of the ten depth cues.

2.2.4 SAR image

SAR images also have few depth cues available: only perspective, occlusion, familiarity, and horizon provide natural depth information. SAR shading is unnatural because it is created by the image process (which is an advantage as well as a disadvantage). Texture is often unpredictable due to its dependence on retro-reflectivity.

We therefore hypothesize that a 3D representation will contribute more to the interpretation of non-literal images than to the interpretation of natural images, because non-literal images tend to lack the natural pattern of depth cues. The addition of the stereopsis depth cue to these "depth impoverished sensors" may in principle therefore be very useful.

2.3 Stereoscopic display hardware

Stereoscopic vision exploits the difference between the viewpoints of the left and right eyes. The perception of stereoscopic depth therefore requires a separate image for each of the two eyes. This difference, which we call the disparity, is typically measured in degrees visual angle or arcminutes visual angle. Commonly used methods to present the left and right eye with their own corresponding image include

- anaglyphs (red and green glasses: one eye only sees red while the other perceives only green)
- shutter glasses (usually 60 Hz alternating: the two eyes see alternating frames)
- polarized glasses (requires two projectors, the one polarized perpendicular to the other, and a projection screen that maintains the polarization)
- a head-mounted display (HMD; each eye has its own miniature display)
- Fresnel or lenticular screen which sends the light from alternate columns of pixels to the left and right eye (does not require eyewear)

Here we are primarily concerned with the requirements that a stereoscopic image needs to fulfill rather than the requirements for the hardware.

3 Guidelines for the effective use of stereoscopic disparity

TNO is currently starting to work on the challenge to improve the interpretability of unnatural images by adding stereoscopic disparity. Here we describe the guidelines that must be considered to achieve sufficient visual comfort and sufficient depth resolution. Our starting point is recent work on the visual comfort of binocular images [14] and on stereoscopic thresholds with pixellated displays [15].

3.1 Resolution

How much resolution is required to support a useful 3D percept? Padmos et al. [15] have shown that in the case of HMD's the stereo-threshold on average equals 1/6 pixel. This means that a horizontal image shift of 1/6 pixel in one eye is just noticeable. A general rule of thumb states that text can be read at an optimal reading rate when its size is 3 or more times larger than threshold (the visual acuity). Assuming the same rule of thumb holds in the stereo domain, the disparity needs to be ½ pixel to be useful. Assuming a typical camera of 15 deg field-of-view and 600 pixels resolution, the useful depth difference equals 1.5 arc-minutes. At a 10 m distance this corresponds to a 60 cm depth difference. By comparison, the average stereo-threshold with the naked eye is approximately 0.5 arc-minutes [16] which corresponds to a 20 cm depth difference at 10 m distance. This calculation shows that stereoscopic imagery is useful if the sensors have at least a reasonable resolution, on the order of 1.5 arc-minute per pixel.

3.2 Field of view

There is no inherent reason why a stereo image should have a particular field of view. One does need to be careful however that the image does not contain crossed disparities at the edges. (A crossed disparity makes the image appear to be in front; the edge of an image acts as a window-frame, causing the image to appear to be behind. This conflict can cause eye strain.)

3.3 Contrast

Image contrast has little effect on stereopsis [16]. Stereo with low contrast sensors (like image intensifiers) should therefore work fine.

3.4 Allowable differences between the left and right images

Recently Kooi & Lucassen [14] have measured the impact of nearly all possible binocular asymmetries on visual comfort. The most relevant parameters are presented here.

3.4.1 A luminance difference

A luminance difference can for example result if a bright light or object enters one image but not the other and the luminance gain triggers on the entire image. A difference in luminance between the left and right images has little impact on visual comfort. It can cause the so-called Pullfrich effect: objects moving from left to right appear to be closer than they are if the left image is lighter and farther if the right image is lighter [12]. To our knowledge this illusion has never been reported to be the cause of a serious problem, however.

3.4.2 Contrast difference

A difference in contrast between the left and right images has little impact on visual comfort and stereopsis [17].

3.4.3 Binocular luster

A very important effect is binocular luster[1] because it occurs frequently in (artificial) stereoscopic images and because it can seriously degrade visual comfort. Binocular luster is the perceptual result of local asymmetries in luminance, in particular left/right contrast reversals. The amount of luster will depend on the type of sensor:

- thermal sensors are expected to produce little luster because they primarily detect emitted electro-magnetic radiation
- SAR sensors are expected to produce lots of luster because they detect reflected radiation with a strong retro-reflective component
- Intensified images are expected to produce roughly the same amount of luster as do CCD images
- Fixed pattern noise will produce luster if the noise in the two images is uncorrelated. Correlated noise will appear at a clearly defined depth plane, its distance determined by the camera alignment

Luster can be experienced while viewing metal objects: the metal reflective 'sheen' is located at slightly different locations in the left and right images. Metal sheen does not cause discomfort, but if the displacement (stereoscopic disparity) or shape difference of such luminance differences is greater it will. Because SAR images inherently contain much retro-reflection, we expect that eyestrain caused by luster will be a major limiting human factor for stereoscopic SAR imagery.

[1] Luster refers to *(1)* the appearance of two different surface colors viewed haploscopically and superimposed, the resulting percept being characteristically unstable and aptly described as one surface being seen through the other, and *(2)* a glossiness or sheen associated with metallic surfaces, sometimes called metallic luster.

3.4.4 Inter-camera separation

A small camera spacing will result in small stereoscopic disparities, thereby limiting its added perceptual value. A camera spacing that is too large on the other hand will result in uncomfortably large disparities and possibly in the appearance of (uncomfortable) luster. It is therefore important to choose the camera spacing "right". Much has been written about this topic. As mentioned above, disparities over 1.5 arc-minutes should be well perceived; this value may serve as a lower limit. The literature gives 1 degrees (or 50 to 70 arc-minutes) as the upper limit for comfortable viewing [16, 18]. The upper limit does not only depend on the total amount of disparity but also on the disparity gradient: the rate of change in depth across the image. A steep gradient is less comfortable to view than a gradual change in depth. The same holds true for diplopia (double vision) which is dependent both on the amount of depth difference and on the depth gradient [19]. Ware et al. [20] recently developed an algorithm that dynamically enhances stereo depth cues for moving 3D computer generated images. The algorithm both optimizes stereo disparities and reduces vergence focus conflicts. This improves stereo viewing, while minimizing the occurrence of double images and reducing eyestrain due to the vergence-focus disrcrepancy. This method may prove a useful tool in the development of stereo displays for non-literal imagery.

3.4.5 Horizontal camera alignment

The horizontal alignment of the two camera's determines which distance is imaged at zero disparity. Generally the plane of zero disparity is placed at the distance of greatest interest. The reason to do this is that the viewing comfort is best when the stereoscopic disparity is close to natural. (Otherwise an accommodation-convergence mismatch is introduced.) Problems arise if this distance is not known ahead of time or if it changes greatly during viewing the stereoscopic imagery. In those cases the disparity should either be kept small (by limiting the inter-camera separation) or the viewer should be given control over the distance of zero disparity. This implies that the viewer has manual control over the horizontal camera alignment or can shift the two images in software. The latter option sacrifices part of the horizontal resolution since part of the image will not be visible to both eyes.

3.4.6 Vertical camera alignment

Small amounts of vertical disparity lead to severe eye strain [14]. The vertical alignment therefore needs to be accurately controlled, preferably to within 15 arc-minutes. The same holds true for image rotation which preferably is set to within 1 degree.

3.4.7 Spatial distortion

The left and right eye images should contain minimal distortion with respect to each other at best. We refer to Kooi & Lucassen [14] for details on this topic.

3.4.8 Sharpness and contrast

A difference in (camera) focus is perceived as straining and should be avoided. This is a point to remember when using autofocus camera's. A difference in image contrast is less troublesome but can better be avoided if possible.

3.5 Fused imagery

Above arguments also apply to fused images. Care needs to be taken that the two images are in all respects as equal as possible, except of course the stereoscopic disparity. When the images constituting a stereo pair originate from sensor systems that apply independent and automatic gain regulations their contrast can differ. Moreover, several image fusion schemes inherently apply locally adaptive contrast stretching. For instance, multiresolution greyscale image fusion schemes usually select the perceptually most salient contrast details from both of the individual input image modalities, and fluently combines these pattern elements into a resulting (fused) image. As a side effect of this approach, details in the resulting fused images can be displayed at higher contrast than they appear in the images from which they originate, i.e. their contrast may be enhanced. The degree of enhancement varies with the location in the image and the resolution of the details. This effect may disturb the perception of stereo pairs, because corresponding details may be depicted at different contrasts in both images. This problem can easily be solved by coupling the algorithms alter local image contrast in both images such that they no longer operate independently.

4 Conclusion

There is a need for a more effective visualization of multisensor data. 3D visualization of non-literal images is well worth to investigate experimentally. Stereoscopic depth holds the potential to make non-literal images appear more intuitive.

Acknowledgement

This material is based upon work supported by the European Office of Aerospace Research and Development, Air Force Office of Scientific Research, Air Force Research Laboratory, under contract No. F61775-01-WE026.

References

[1] O'Kane, B.L., Bonzo, D. & Hoffman, J.E. (2001). Perception studies. *Optical Engineering, 40(9).*

[2] Oliver, C. & Quegan, S. (1998). *Understanding synthetic aperture radar images.* Boston: Artech House.

[3] Essock, E.A., Sinai, M.J., McCarley, J.S., Krebs, W.K. & DeFord, J.K. (1999). Perceptual ability with real-world nighttime scenes: image-intensified, infrared, and fused-color imagery. *Human Factors, 41(3),* 438-452.

[4] Sinai, M.J., McCarley, J.S., Krebs, W.K. & Essock, E.A. (1999). Psychophysical comparisons of single- and dual-band fused imagery. In J.G. Verly (Ed.), *Enhanced and Synthetic Vision 1999* (pp. 176-183). Bellingham, WA, USA: International Society for Optical Engineering.

[5] Steele, P.M. & Perconti, P. (1997). Part task investigation of multispectral image fusion using gray scale and synthetic color night vision sensor imagery for helicopter pilotage. In W. Watkins & D. Clement (Ed.), *Proceedings of the SPIE Conference on Targets and Backgrounds, Characterization and Representation III* (pp. 88-100). Bellingham, WA, USA: International Society for Optical Engineering.

[6] Toet, A., IJspeert, J.K., Waxman, A.M. & Aguilar, M. (1997). Fusion of visible and thermal imagery improves situational awareness. In J.G. Verly (Ed.), *Enhanced and Synthetic Vision 1997* (pp. 177-188). Bellingham, WA, USA: International Society for Optical Engineering.

[7] McCarley, J.S. & Krebs, W.K. (2000). Visibility of road hazards in thermal, visible, and sensor-fused night-time imagery. *Applied Ergonomics, 31(5),* 523-530.

[8] Toet, A., Schoumans, N. & IJspeert.J.K. (2000). Perceptual Evaluation of Different Nighttime Imaging Modalities.*Proceedings of the 3rd International Conference on Information Fusion* (pp. TuD3-17-TuD3-23). Paris, France: ONERA.

[9] Krebs, W.K., Scribner, D.A., Miller, G.M., Ogawa, J.S. & Schuler, J. (1998). Beyond third generation: a sensor-fusion targeting FLIR pod for the F/A-18. In B.V. Dasarathy (Ed.), *Sensor Fusion: Architectures, Algorithms, and Applications II* (pp. 129-140). Bellingham, WA, USA: International Society for Optical Engineering.

[10] Watkins, W.R., Heath, G., Phillips, M.D., Valeton, J.M. & Toet, A. (2001). Search and target acquisition: single line of sight versus wide baseline stereo. *Optical Engineering, 40(9),* In Press.

[11] Richards, W. (1970). Stereopsis and stereoblindness. *Experimental Brain Research, 10,* 380-388.

[12] Howard, I.P. & Rogers, B.J. (1995). *Binocular vision and stereopsis.* Oxford, UK: Oxford University Press.

[13] Berkley, W.E. (1992). Night vision goggle illusions and visual training.*Visual problems in night operations* (pp. 9-1-9-6). Neuilly-sur-Seine Cedex, France: North Atlantic Treaty Organization.

[14] Kooi, F.L. & Lucassen, M. (2001). Visual comfort of binocular and 3-D displays. In B.E. Rogowitz & T.N. Pappas (Ed.), *Human Vision and Electronic Imaging VI* Bellingham, WA: The International Society for Optical Engineering.

[15] Padmos, P., Kooi, F.L. & Bijl, P. (2001). Stereo Acuity and Visual Acuity in Head Mounted Displays. *Human Factors,* In Press.

[16] Schor, C.M. & Wood, I. (1983). Disparity range for local stereopsis as a function of luminance spatial frequency. *Vision Research, 23,* 1649-1654.

[17] Legge, G.E. & Gu, Y. (1989). Stereopsis and contrast. *Vision Research, 29,* 989-1004.

[18] Pastoor, S. (1993). Human factors of 3D displays. *Displays, 14.*

[19] Tyler,C.W. (1991). Cyclopean vision. In D.Regan (Ed.), *Vision and visual dysfunction. Vol. 9: Binocular Vision.* (pp. 38-74). London, UK: Macmillan.

[20] Ware, C., Gobrecht, C. & Paton, M.A. (1998). Dynamic adjustment of stereo display parameters. *IEEE Transactions on Systems, Man and Cybernetics - Part A: Systems and Humans, 28(1),* 56-65.

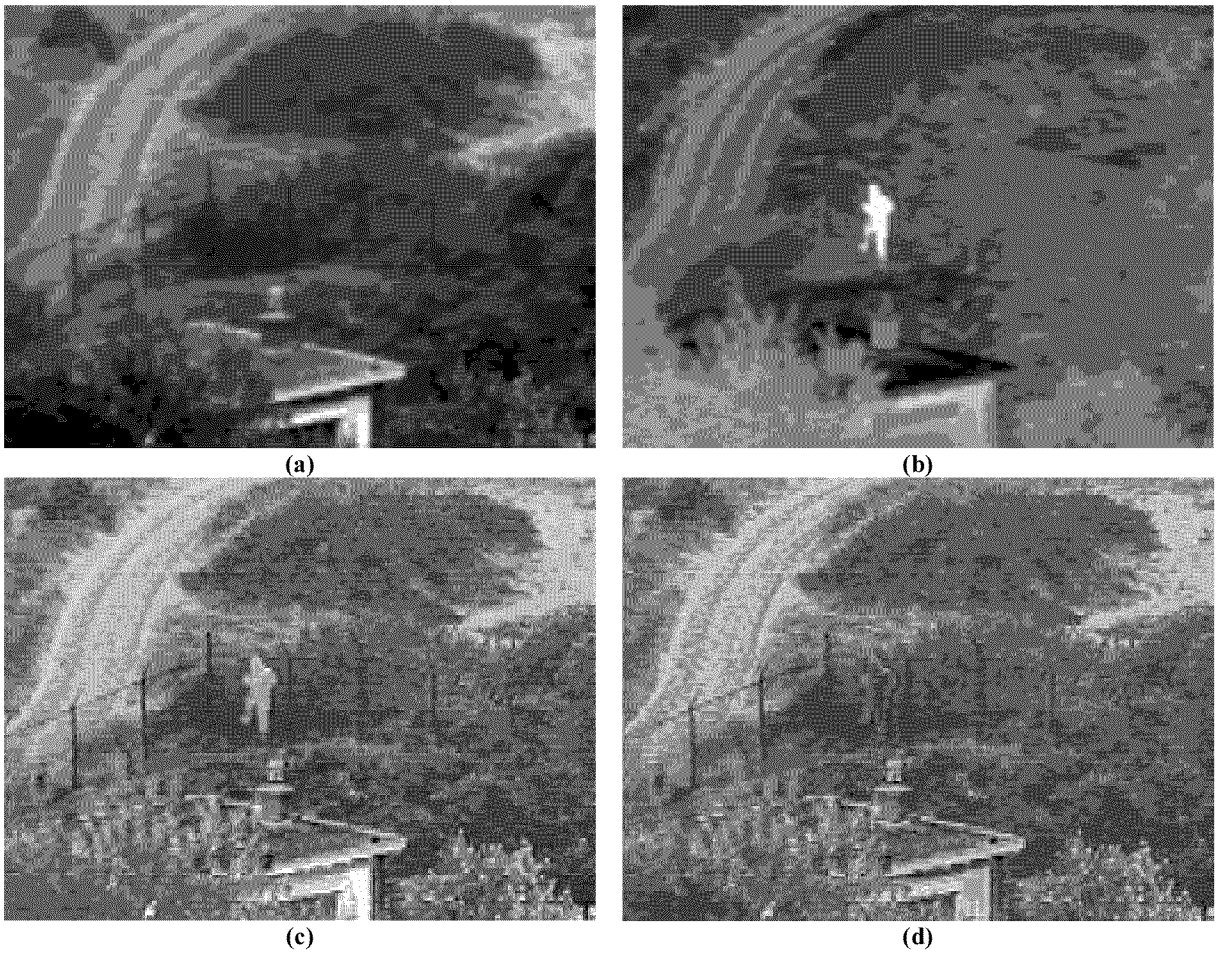


Figure 1 Scene from a nighttime movie sequence. (a) CCD image, (b) thermal (3-5 μm) image, (c) grayscale fused image, and (d) color fused image of a scene representing a man walking behind a fence along a building. The man is not visible in the CCD image. Although he is clearly visible in the thermal image, it is not easy to perceive on which side of the fence he is walking. In the moving grayscale sequence one can easily perceive on which side of the fence the man is walking. Color fused images provide a better sense of depth because the scene is easier to segment visually. The position of the man with respect to the fence is obvious in a stereo presentation of this scene.

Appendix B

Detection of dim point targets in cluttered maritime backgrounds through multisensor image fusion

Alexander Toet*
TNO Human Factors, Kampweg 5, 3769 DE Soesterberg, The Netherlands

ABSTRACT

Multispectral IR imaging techniques are frequently deployed in maritime operations, for instance to detect floating mines or to find small dinghies and swimmers during search and rescue operations. However, maritime backgrounds usually contain a large amount of clutter that severely hampers the detection of dim point targets. Here we present a simple algorithm that deploys the correlation between target signatures in two different (3-5 and 8-12 μm) IR frequency bands to reduce the amount of clutter. First, both individual IR bands are filtered with a morphological opening top-hat transform to extract small details. Second, the resulting detail images are thresholded to produce binary detail images, representing potential target areas. Third, a fused detail image is obtained by taking the intersection (logical AND) of both binary IR detail images. Details that appear in both IR bands remain in this fused detail image, whereas a large fraction of uncorrelated noise details is filtered out. Remaining noise details can be removed by taking into account the temporal characteristics of the target signatures and by using a priori knowledge of structure of the scene and the size of potential targets. The method is tested on two image sequences showing a maritime scene with three kayaks approaching from far away. The scenario was registered in the 3-5 μm and 8-12 μm IR frequency bands, and in the visual range. The results show that the proposed multispectral processing technique has the potential to improve the detection of dim point targets in cluttered maritime backgrounds.

Keywords: Image fusion, infrared, detection, mathematical morphology

1. INTRODUCTION

This paper discusses the problem of small low observable (dim) target detection in cluttered maritime infrared images. This problem is of interest in many applications such as ocean surveillance (e.g. oil spills), search (e.g. speedboats, dinghies) and rescue (e.g. swimmers), remote sensing, floating mine detection, etc.

In the literature there are many useful algorithms for the detection and tracking of targets of significant size. However, in maritime infrared images, targets of interest are usually small and rather dim in a relative dark and cluttered sea surface background. Sea surface structure, reflection and emission changes related to incident angle variations and surface effects are standard features governing the clutter behavior[10]. Sun glint and horizon effects also contribute to clutter. The existence of scanline disturbance and noise further increases the difficulty in proper detection.

Non-linear filtering is a powerful technique to detect small dim targets in cluttered backgrounds[2,4,8,9]. It has also been observed that co-registered multispectral imagery may result in improved target detection[1,8,12]. As result, multispectral IR imaging techniques for target detection have recently received increased attention. Maritime surveillance systems consisting of combined visual and IR sensor systems have already been developed[13].

The aim of the present study is to reduce the false alarm rate of a combined visual and IR surveillance camera system to an acceptable level. The camera system used here registers the 3-5 and 8-12 μm IR bands and the visual band. The combination of a morphological top-hat transform and a thresholding operation is used to detect potential target areas in the raw individual IR band images. The amount of clutter is reduced by taking the intersection of the potential target areas thus found. Finally, the outlines of the remaining potential target areas are projected onto the visual display. Experiments on a large set of sea going kayak images prove the effectiveness of the method presented in this paper.

* E-mail: toet@tm.tno.nl

2. IMAGE REGISTRATION

This section describes the experimental equipment, the location, the scenario, and the registration procedures that were deployed to collect the imagery that is used in the rest of this study.

2.1. Equipment

The following cameras were used to register the imagery that is used in the rest of this study:

- A Radiance HS IR camera (Raytheon), sensitive for 3-5 μm.
- An AIM 256 μLW camera (AEG), sensitive for 8-10 μm.
- A Philips LTC500 CCD camera (f/1.2 50dB s/n at 0.4 Lux) sensitive from visual to near IR (400-900 nm).

The Radiance HS IR focal plane array camera was used to register the 3-5 μm middle wavelength band of the infrared spectrum. The Radiance HS produces an image consisting of 256×256 pixels, with a FOV of 4.4×4.4 degrees^2, and an IFOV of 0.3 mrad. A lens with a focal length of 100 mm was used. The experimentally determined noise equivalent temperature difference (NETD) was 0.045 K.

The AIM 256 μLW focal plane array camera was used to register the 8-10 μm long wavelength band of the infrared spectrum. This camera produces an image consisting of 256×256 pixels, with a FOV of 5.9×5.9 degrees^2, and an IFOV of 0.4 mrad. A lens with a focal length of 100 mm was used. The experimentally determined noise equivalent differential temperature (NEDT) of 0.033 K was little above specification (<0.025 K).

The Philips LTC500 CCD camera was used to register the visual part of the electromagnetic spectrum. This camera is equipped with a zoom lens and produces an image consisting of 752 x 582 pixels. In the narrow angle (tele) zoom setting the FOV is 2×1 degrees^2, and in the wide angle zoom setting the FOV is 14.7×7.4 degrees^2.

The field of view (FOV) of the cameras roughly matched, and was about 5×5 degrees^2 (varying from 4.5 to 6 degrees).

The cameras were rigidly mounted on a common base plate, that was placed on a pan and tilt unit, which was in turn supported by a tripod. Their optical axes were aligned and placed within 60 cm from each other. Since the viewing distances were relatively large, parallax can be neglected for all further purposes.

At the start of each recording session, the non-uniformity corrections (NUC) were determined for the IR cameras.

The camera outputs were continuously recorded using Panasonic AG-7700 SVHS video recorders. The resolution of the stored images is bandwidth limited by these videorecorders. However, video recording allows a full back-up and enables the comparison of dynamic scenes. The videorecorders were set-up in a master-slave configuration, providing each frame with a common time stamp. This enables digitisation of the videotapes with a temporal resolution of less than one frame length.

2.2. Location

The image recording session was held at the fourth floor of a hotel in Den Helder, The Netherlands. The hotel was about 20 m from the coastline, facing the sea. The cameras were placed outside on a balcony overlooking the sea. This condition was chosen to simulate the view from the bridge of a tall ship.

2.3. Environmental conditions

During the image registration period the sky was partly overcast (about 50% cloud cover). The temperature was 7.1 degrees Celsius (Fahrenheit). The wind was blowing from the south, with a windspeed of 9 m/s. Meteorological visibility was about 7 km. Air pressure was 1020 hPa. These conditions resulted in a moderate amount of clutter (breaking waves) in the maritime environment.

2.4. Scenario

The scenario represented a part of the North Sea coastline, with three kayaks approaching the viewing location from far away. As a result, their corresponding image size varies from less than 1 pixel to almost the entire field of view. In this study we will only use images in which the kayaks are not larger than a few pixels.

2.5. Image warping

As mentioned before, the optical axes of all cameras were aligned such that the common area of their FOV was maximised. However, due to imperfections in the alignment procedure and differences in the optics of the cameras, an affine geometric transform is usually required to achieve one-to-one (pixel wise after digitisation) correspondence between the different image modalities. The parameters of such a transform can be determined from a set of fiducial points that are *(a)* well defined and distinctly represented in all image modalities, and *(b)* evenly distributed over the area of the common FOV. At the beginning of each recording session a well defined fiducial point was created by placing a large plastic jerry can on the beach near the coastline. The can was bright (white) and filled with hot water. As a result it was clearly visible in both the visual and infrared image modalities. It was located at such a distance in the scene that its image size was about 2x2 pixels. Images were registered with the jerry can at 9 different positions, evenly distributed over the common FOV of the cameras. This procedure resulted in 9 fiducial points that were later used to compute the parameters of an affine warping transform that maps the images to a common underlying reference grid.

3. IMAGE PROCESSING

This section first presents a brief introduction to mathematical morphology in general and the top-hat transform in particular. Then we describe how the top-hat transform can be employed to detect dim point targets in the individual 3-5 and 8-12 μm IR image bands. Finally, it is argued that uncorrelated noise details can be filtered out by taking the intersection of the potential target areas detected in each of the individual IR bands.

3.1. Mathematical morphology

In this section we briefly define the basic morphological transformations for discrete (e.g. sampled) functions. For an extensive introduction to mathematical morphology we refer to the literature[11].

In the sequel $\mathbb{Z}$ will denote the set of integers and $\mathbb{R}$ the set of real numbers. Discrete (i.e. sampled) images and structuring elements will be denoted by capital letters $A,B,C,\ldots$, their domains and ranges by script letters $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C},\ldots$ Let $F(x)$ be a function defined on a finite discrete d-dimensional domain $\mathcal{F}\subset\mathbb{Z}^d$ (d=1,2,…) and with a discrete and continuous amplitude range $\mathcal{R}\subset\mathbb{R}$ or $\mathbb{Z}$. A *structuring element* is any function $G(x)$ whose support $\mathcal{G}$ is a compact and connected subset of $\mathbb{Z}^d$. In the sequel we will only use flat (binary or bivalued structuring elements $B(x)$, with $B(x)=0$ for $x\in\mathcal{B}$. The extension of our results to multivalued (grayscale) structuring elements is straightforward[3,5,11]. The basic morphological operations on discrete functions with binary structuring elements are defined in Table I. To implement these operations we assume that $F(x)=-\infty$ for $x\notin\mathcal{B}$.

3.2. Top-hat transform

The opening transform removes bright details from an image that are smaller than the size of the structuring element. The residual image, obtained by subtracting the opened image from the original, comprises of only those image features that have been removed by the structuring element in the opening operation. This filter operation is called the *top-hat transform*[6] and is defined as

$$(F-(F\circ B))(x)=(F-(F\ominus\tilde{B})\oplus B)(x)$$

It provides an excellent tool for extracting bright features smaller than a given size from an uneven background.

Table I. Basic morphological transformations for discrete functions.

Name	**Definition** ($x, y, p \in \mathbb{Z}^d$, $\mathcal{B} \subset \mathbb{Z}^d$, d=1,2,…)
set translation of $\mathcal{B}$ by p	$\mathcal{B}_p = \{b+p : b \in \mathcal{B}\}$
symmetric set of $\mathcal{B}$	$\tilde{\mathcal{B}} = \{-b : b \in \mathcal{B}\}$
symmetric function of B	$\tilde{B}(x) = B(-x)$
Minkowski addition of F and B	$(F \oplus B)(x) = \max\{F(y) : y \in \check{\mathcal{B}}_x\}$
Minkowski subtraction of B from F	$(F \ominus B)(x) = \min\{F(y) : y \in \tilde{\mathcal{B}}_x\}$
dilation of F by B	$(F \oplus \tilde{B})(x) = \max\{F(y) : y \in \mathcal{B}_x\}$
erosion of F by B	$(F \ominus \tilde{B})(x) = \min\{F(y) : y \in \mathcal{B}_x\}$
closing of F by B	$F \bullet B(x) = (F \oplus \tilde{B}) \ominus B(x)$
opening of F by B	$F \circ B(x) = (F \ominus \tilde{B}) \oplus B(x)$

3.3. Multiband detection

The top-hat transform can be employed to detect dim point targets in the individual 3-5 and 8-12 μm IR image bands. However, in maritime environments with a high degree of clutter the top-hat transform will yield many false alarms. It is a priori likely that small targets like boats, mines and persons, will be represented in both bands. In contrast, noise details will mostly be resticted to one particular band. Therefore, uncorrelated noise details may be filtered out by taking the intersection of the potential target areas detected in each of the individual IR bands.

4. RESULTS

Figures 1-7 show a set of typical maritime images, representing three kayaks (approximately in the middle of the scene) at high sea. The composition of each of these figures is identical. The upper left and right images correspond respectively to the original Radiance 3-5 μm and AIM 8-12 μm IR images. The middle row in each figure shows the potential target areas detected in the individual IR bands. These areas are obtained by thresholding the top-hat transform of the IR bands. The potential target areas are enlarged by dilation with a disk shaped structuring element of size 3 (radius 1). This is done to compensate for possible offsets in the registration between the individual IR bands. Dilation increases the chances that the potential target areas in the individual IR bands will have at least some degree of physical overlap. By taking the intersection (implemented as a logical AND on both binary IR target area images), a large amount of irrelevant noise details are eliminated and the potential target areas remain. The result of this operation is shown in the lower left image in Figures 1-7. Note that *(a)* the number of false alarms (see middle row of Figures 1-9) is significantly reduced, and *(b)* the correct targets (three kayaks in the middle of the scene) are detected each time. The lower right image in Figures 1-7 shows the enlarged outlines of the potential target areas projected over the corresponding visual CCD image.

5. CONCLUSIONS

The morphological top-hat transform can be used to detect dim point targets in cluttered maritime backgrounds registered in the 3-5 and 8-12 μm IR bands. The number of false alarms (noise details) can be reduced significantly by taking the intersection of the potential target areas (alarms) in both bands. Thus, the proposed multispectral processing technique can improve the detection of dim point targets in cluttered maritime backgrounds.

In the scenario used in this study, the dim target moves very slowly ($v < 0.5$ pixel/frame). The signal to noise ratio can therefore be improved (the definition of the target enhanced) by summation of successive frames. This will result in an increase in target energy and a decrease in clutter energy. Considering the low speed of the target, the summing operation may be implemented by directly adding N consecutive frames under the assumption that the point target stays at a pixel for at least N frames.

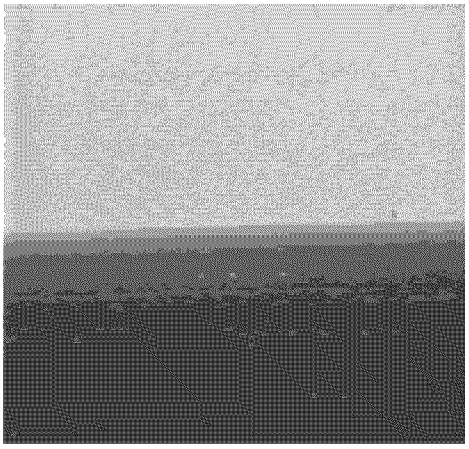
The detection probability can be improved even further by the use of a priori knowledge. Tuning the structuring element used in the top-hat transform to the size and shape of the targets will significantly reduce the amount of false alarms[7]. Information on the distance and possible target locations may also increase the detection probability.

ACKNOWLEDGEMENTS

This material is based upon work supported by the European Office of Aerospace Research and Development, Air Force Office of Scientific Research, Air Force Research Laboratory, under contract No. F61775-01-WE026, and by Senter, Agency of the Ministry of Economic Affairs of the Netherlands.

REFERENCES

1. Abu-Tahnat, M. and Thompson, M.W., Frequency-band comparison using adaptive filter for multispectral imagery, In: A.C. Dubey, R.L. Barnard, C.J. Loe & J.E. McFee (Ed.), *Detection and remediation technologies for mines and minelike targets*, pp. 36-45, The International Society for Optical Engineering, Bellingham,WA, 1996.
2. Deshpande, S.D., Er, M.H., Ronda, V. and Chan, P., Max-mean and max-median filters for detection of small targets, In: O.E. Drummond (Ed.), *Signal and Data Processing of Small Targets 1999*, pp. 74-83, The International Society for Optical Engineering, Bellingham,WA, 1999.
3. Haralick, R.M., Sternberg, S.R. and Zhuang, X., Image analysis using mathematical morphology, *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI, 9(4)* ,pp. 532-550, 1987.
4. Kemper, P.J., Mathematical morphology enhancement of maximum entropy thresholding for small targets, In: D.P. Casasent & A.G. Tescher (Ed.), *Hybrid Image and Signal Processing VI*, pp. 84-91, The International Society for Optical Engineering, Bellingham,WA, 1998.
5. Maragos, P., Pattern spectrum of images and morphological shape-size complexity, In: *IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP 87*, pp. 241-244, IEEE Computer Society Press, Washington, USA, 1987.
6. Meyer, F., Contrast feature extraction, In: J.L. Chermant (Ed.), *Quantitative analysis of microstructure in material sciences, biology and medicine*, Riederer Verlag, Stuttgart, GE, 1978.
7. Moon, V.-S., Zhang, T., Zuo, Z. and Zuo, Z., Detection of sea surface small targets in infrared images based on multilevel filter and minimum risk Bayes test, *International Journal of Pattern Recognition and Artificial Intelligence,* 14*(7)* ,pp. 907-918, 2000.
8. Perez-Jacome, J.E. and Madisetti, V.K., Target detection via combination of feature-based target-measure images, In: I. Kadar & V. Libby (Ed.), *Signal processing, sensor fusion, and target recognition VIII*, pp. 345-356, The International Society for Optical Engineering, Bellingham,WA, 1999.
9. Sang, N., Zhang, T. and Shi, W., Characteristics of contrast and application for small-target detection, In: O.E. Drummond (Ed.), *Signal and Data Processing of Small Targets 1998*, pp. 123-129, The International Society for Optical Engineering, Bellingham,WA, 1998.
10. Schwering, P.B., Maritime infrared background clutter, In: W.R. Watkins & D. Clement (Ed.), *Targets and backgrounds: characterization and representation II*, pp. 255-266, The International Society for Optical Engineering, Bellingham,WA, 1996.
11. Serra, J., *Image analysis and mathematical morphology* , Academic Press, New York, USA, 1982.
12. Singer, P.F. and Sasaki, D.M., Multispectral detection of dim slightly extended targets in heavy clutter, In: O.E. Drummond (Ed.), *Signal and Data Processing of Small Targets 2000*, pp. 96-103, The International Society for Optical Engineering, Bellingham,WA, 2000.
13. Yamamoto, K., Yamada, K. and Kiriya, N., System for maritime surveillance aid, In: S.G. Ackleson (Ed.), *Ocean Optics XIII*, pp. 815-820, The International Society for Optical Engineering, Bellingham,WA, 1997.

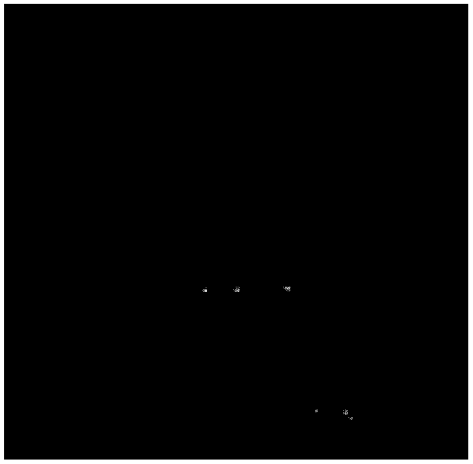


Radiance 3-5 μm

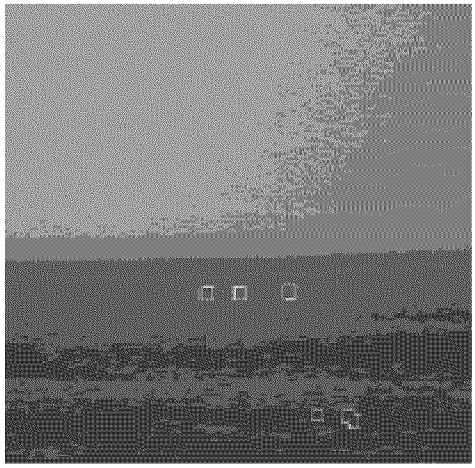
AIM 8-12 μm

Radiance Alarms

AIM Alarms

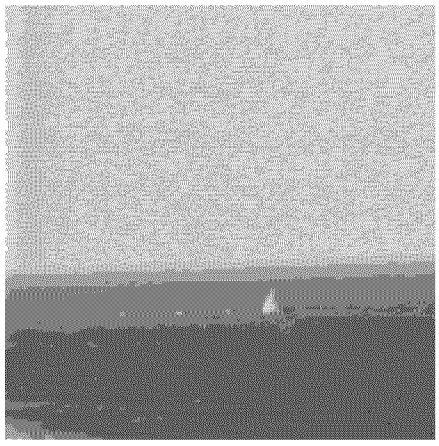


Multiband Alarms

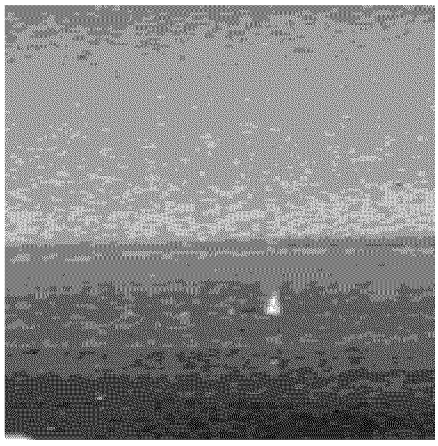


CCD image with potential target areas

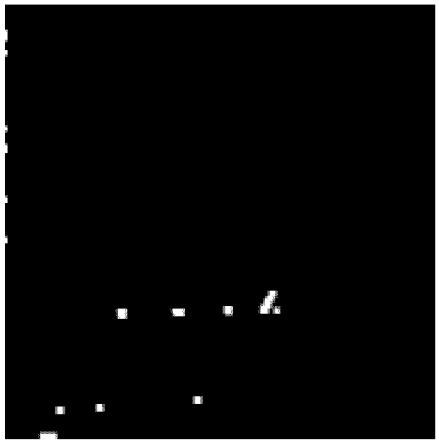
Figure 1 Upper row: corresponding original Radiance 3-5 μm and AIM 8-12 μm IR images. Middle row: potential targets detected in the individual IR bands. Lower row: potential targets detected in both IR bands (left) and target areas projected over the corresponding visual CCD image (right).



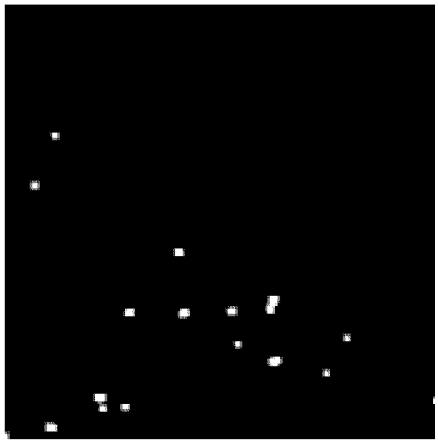
Radiance 3-5 μm



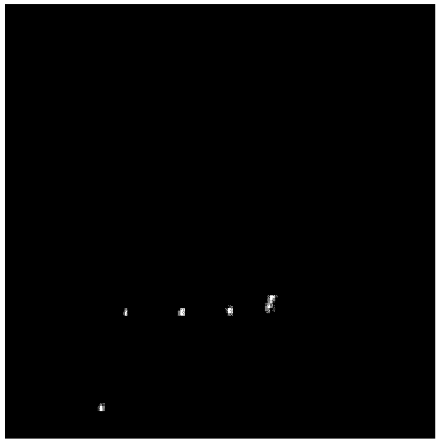
AIM 8-12 μm



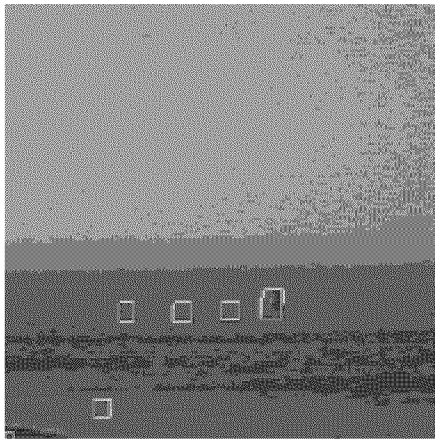
Radiance Alarms



AIM Alarms

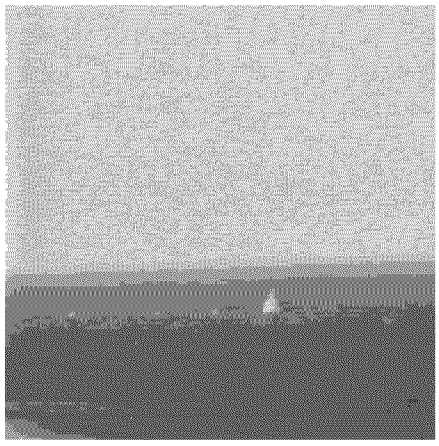


Multiband Alarms

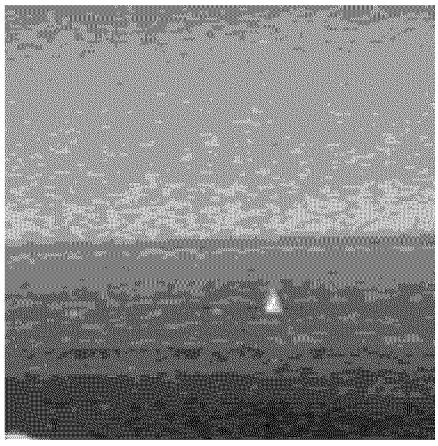


CCD image with potential target areas

Figure 2 Upper row: corresponding original Radiance 3-5 μm and AIM 8-12 μm IR images. Middle row: potential targets detected in the individual IR bands. Lower row: potential targets detected in both IR bands (left) and target areas projected over the corresponding visual CCD image (right). The large triangular object on the right is s buoy.

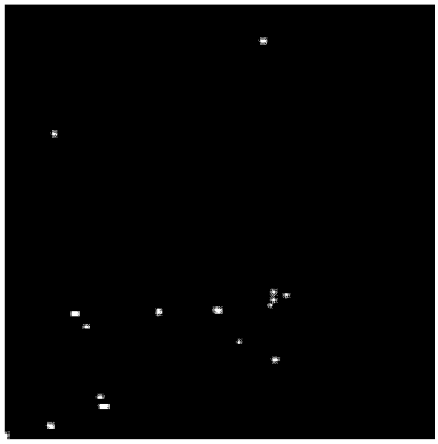


Radiance 3-5 μm

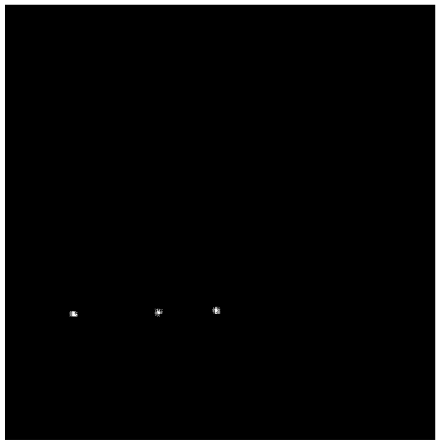


AIM 8-12 μm

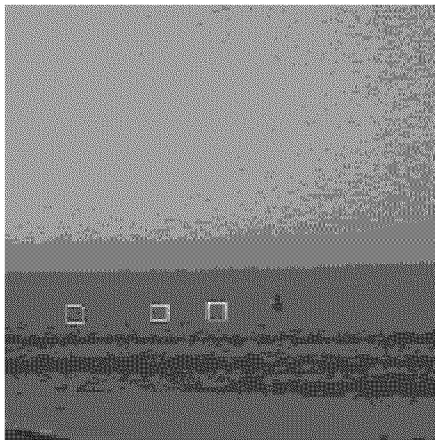
Radiance Alarms



AIM Alarms

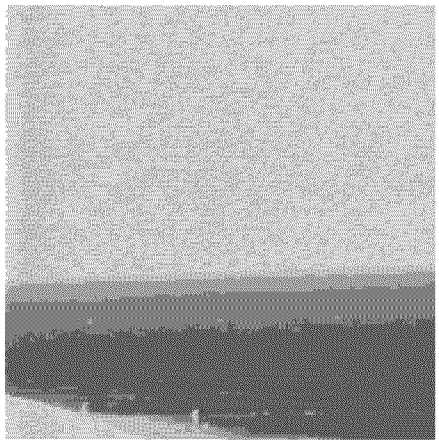


Multiband Alarms

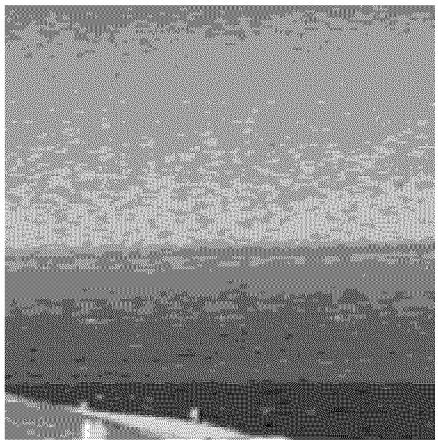


CCD image with potential target areas

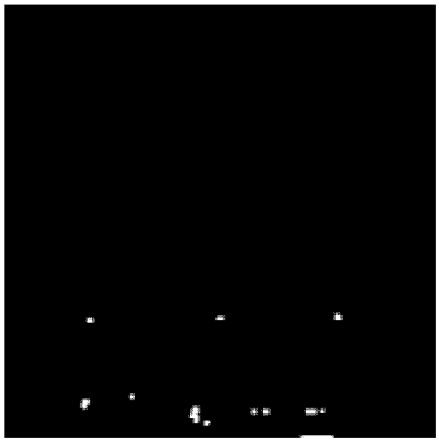
Figure 3 Upper row: corresponding original Radiance 3-5 μm and AIM 8-12 μm IR images. Middle row: potential targets detected in the individual IR bands. Lower row: potential targets detected in both IR bands (left) and target areas projected over the corresponding visual CCD image (right).



Radiance 3-5 μm

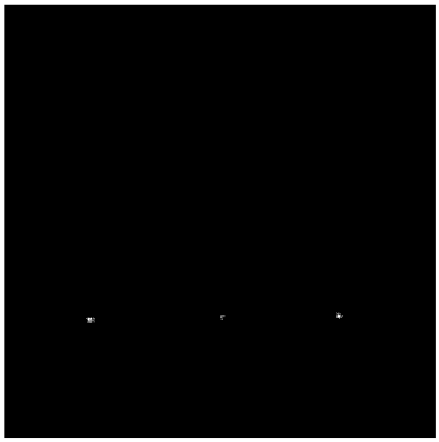


AIM 8-12 μm

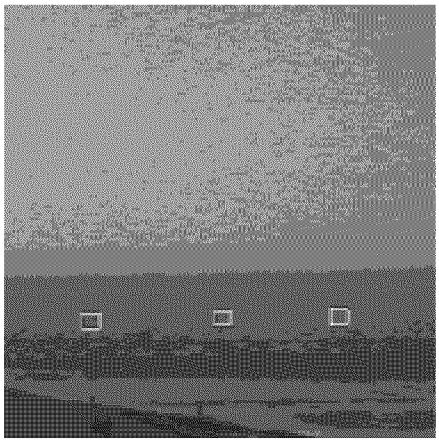


Radiance Alarms

AIM Alarms

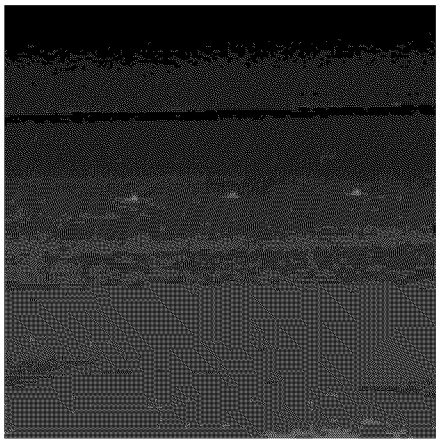


Multiband Alarms

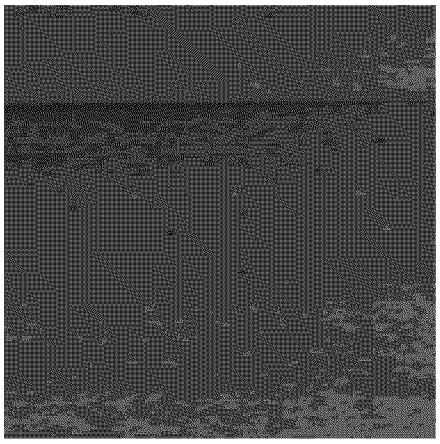


CCD image with potential target areas

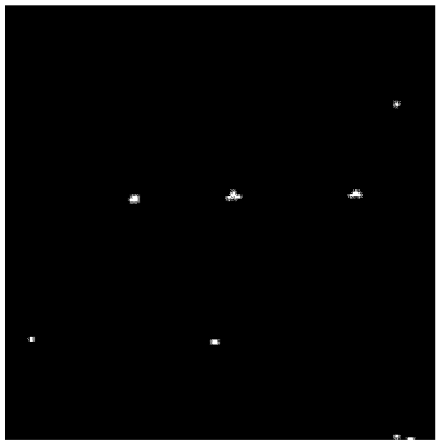
Figure 4 Upper row: corresponding original Radiance 3-5 μm and AIM 8-12 μm IR images. Middle row: potential targets detected in the individual IR bands. Lower row: potential targets detected in both IR bands (left) and target areas projected over the corresponding visual CCD image (right).



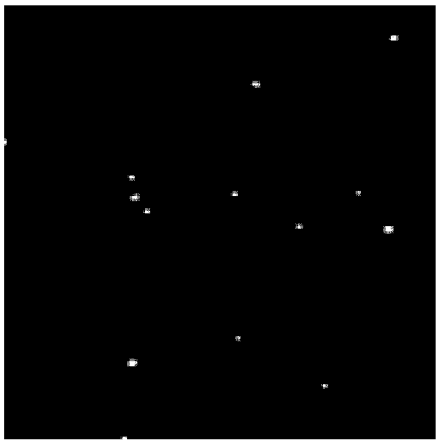
Radiance 3-5 μm



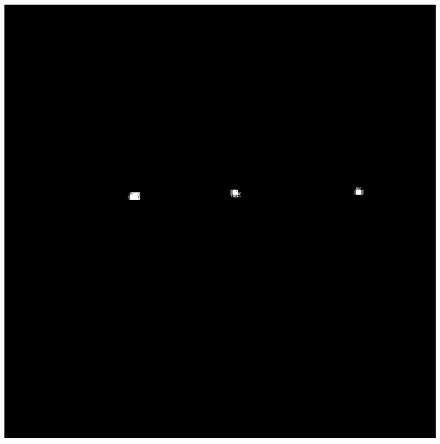
AIM 8-12 μm



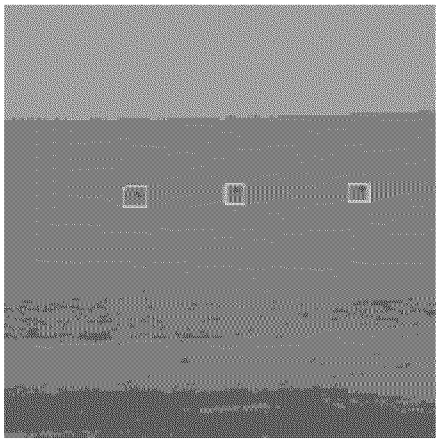
Radiance Alarms



AIM Alarms



Multiband Alarms

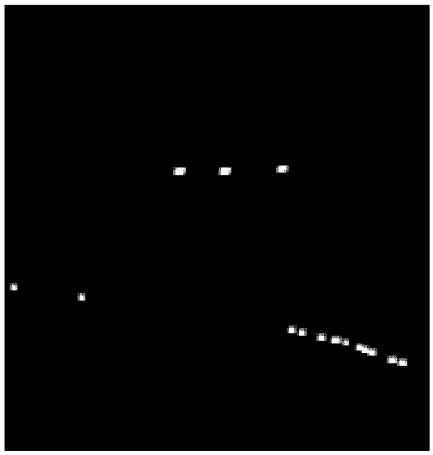


CCD image with potential target areas

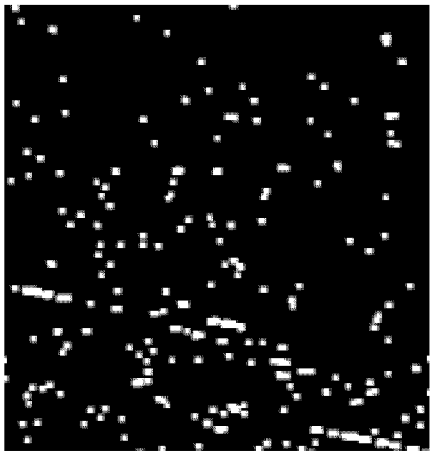
Figure 5 Upper row: corresponding original Radiance 3-5 μm and AIM 8-12 μm IR images. Middle row: potential targets detected in the individual IR bands. Lower row: potential targets detected in both IR bands (left) and target areas projected over the corresponding visual CCD image (right).

Radiance 3-5 μm

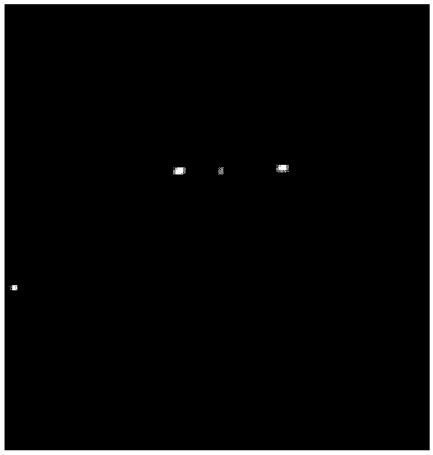
AIM 8-12 μm



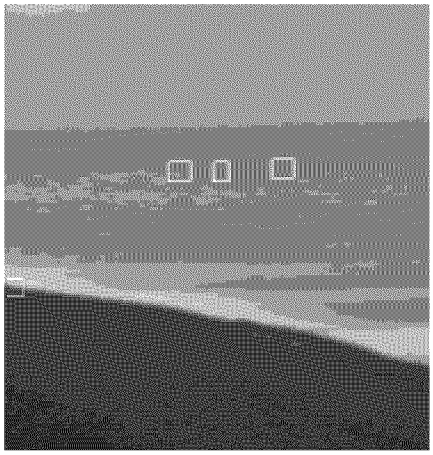
Radiance Alarms



AIM Alarms

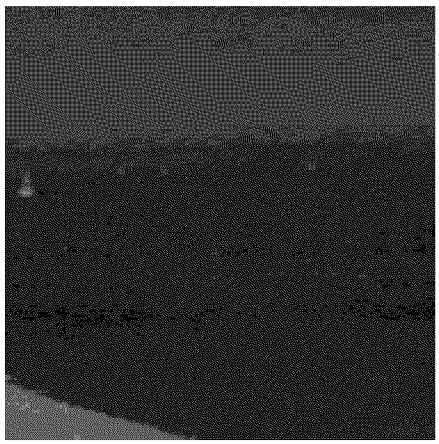


Multiband Alarms

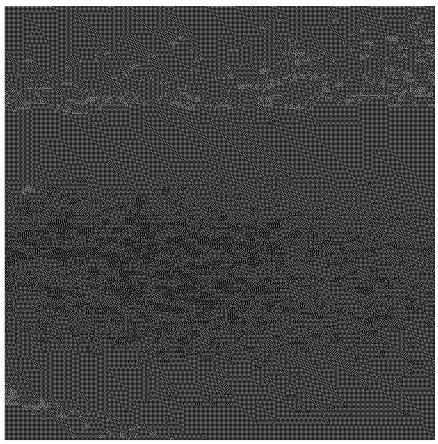


CCD image with potential target areas

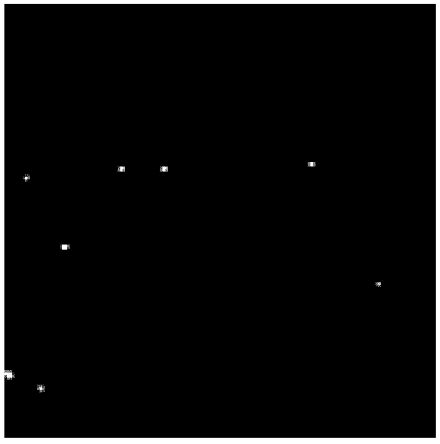
Figure 6 Upper row: corresponding original Radiance 3-5 μm and AIM 8-12 μm IR images. Middle row: potential targets detected in the individual IR bands. Lower row: potential targets detected in both IR bands (left) and target areas projected over the corresponding visual CCD image (right).



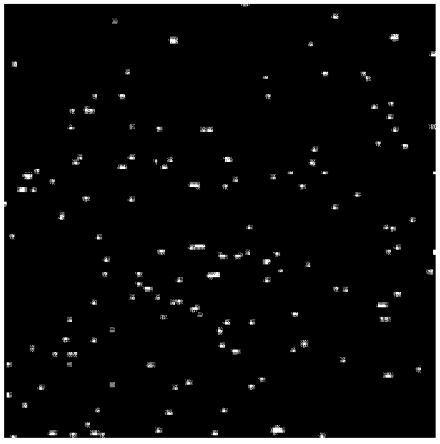
Radiance 3-5 μm



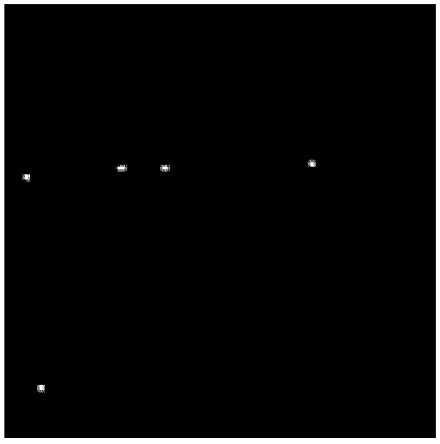
AIM 8-12 μm



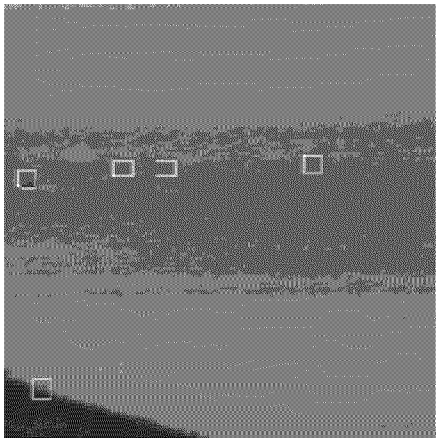
Radiance Alarms



AIM Alarms



Multiband Alarms



CCD image with potential target areas

Figure 7 Upper row: corresponding original Radiance 3-5 μm and AIM 8-12 μm IR images. Middle row: potential targets detected in the individual IR bands. Lower row: potential targets detected in both IR bands (left) and target areas projected over the corresponding visual CCD image (right).

Appendix C

Using depth to indicate potential targets

M.A. Hogervorst & A. Toet

TNO Human Factors, Kampweg 5, 3769 DE Soesterberg, The Netherlands

ABSTRACT

We have developed and evaluated a method for highlighting potential targets while keeping them in their context. Potential targets are put at a different depth plane from the rest. This allows the observer to restrict search to a limited set of potential targets, and at the same time keeps the targets within their surrounding contexts (to keep track of the global position within the surroundings). We have evaluated this method with an experiment in which observers had to search for a ring within a large number of C's (with different orientations) acting as distractors. The results show that when half the distractors are placed in a different depth plane, the search times are (nearly) as fast as when the number of distractors is halved. This shows that observers can restrict search to a single plane. It shows that the method of putting potential targets in a different plane can speed up the search process while keeping the targets within their context. This method is especially suited for highlighting potential elements in a full colour image.

INTRODUCTION

There are several ways in which potential targets can be indicated. Some methods are more suited than others in situations in which it is important that the observer gets a good impression of the location of the target within the surrounding. Another requirement is that search performance (search times, detection rates) are improved due to the image transformation. To keep track of the global location of the target it is important that the spatial relations are not distrurbed by the image transformation and that the surrounding image remains clearly visible. One method to indicate potential targets is by making the colour different from the rest Figure 1, left hand side). Object recognition will be somewhat worsened relative to the a fully black and white image (the luminance range is restricted). Still, it can be expected that this will be a suitable method for black and white images. When the image is in full colour it will be somewhat more difficult to use this method. In that case one may draw black lines in the image indicating the borders of the regions of interest. However, we expect that search times will not drop dramatically using this method, since the observer will still have to search all over the image to detect the borders. To overcome such problems we have tried a third method: placing the potential targets in a different depth plane (see e.g. Figure 1, right hand side). The advantages of this method are:

1) potential targets can be seen within their surrounding context
2) the colours remain the same. This means that the luminance and colour contrasts remain the same, i.e. object recognition performance remains the same.
3) search can be restricted to the regions of interest, as shown in this study. This means that search performance is high.

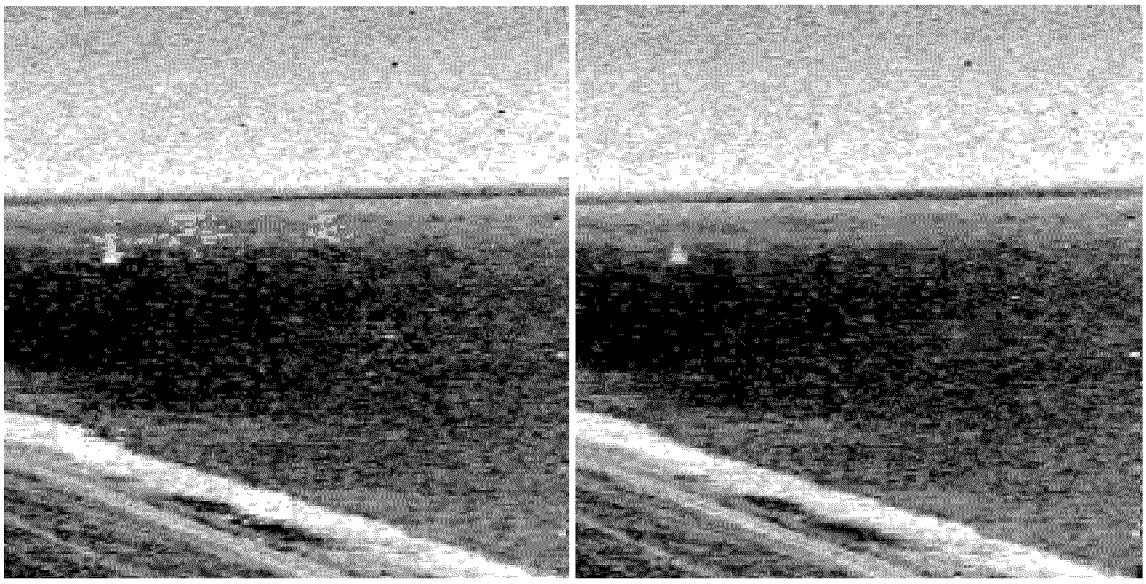


Figure 1. Methods of indicating potential targets. On the left: targets are indicated by using different colours. On the right: potential targets are indicated by putting them in different depth planes (using binocular disparities). When viewed with red-green anaglyphs (glasses with red and green filters), the blue regions in the left image will appear in front of the rest of the image.

METHOD

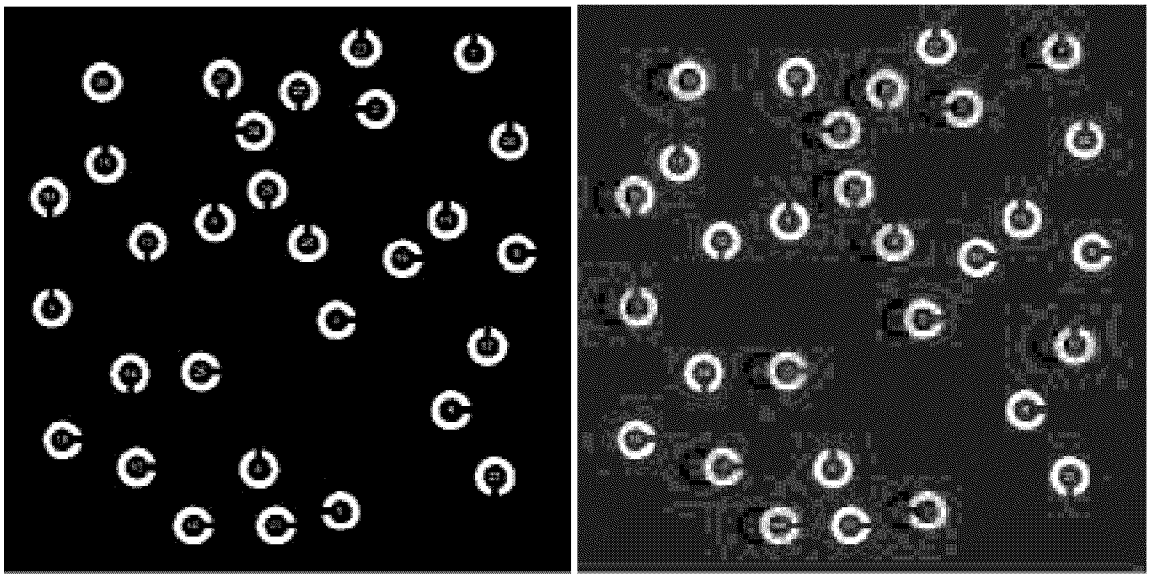


Figure 2. Example stimuli used in the experiment. The images consist of (Landolt-)C symbols of different orientations (the distractors) and a single full ring (the target) with numbers. On the left: a black and white image of an image in which all the objects lie in the same plane. On the right: an anaglyph image (for red-green glasses) in which half of the distacators and the target lie in front of the rest of the image.

We presented images consisting of LandoltC rings (a ring with a gap, see Figure 2) and a single full ring (the target) to two observers (MAH and PB) on a PC monitor. Each object also contained a number. In each trial the observer started the presentation by pressing the <enter> key on a keyboard. This started a timer. As soon as the observer found the target (the full ring) he hit the <enter> key again, which stopped the timer. The observer then entered the number of the target via the keyboard. We used this to check whether the target was detected correctly. The response time and the number of the target were recorded. As soon as the PC was ready to show the next image a sound was played. In each condition 50 images were shown. We compared performance for a situation in which all objects appeared in the same plane with a condition in which half of the distactors appeared in a different plane. In the latter case we used stereo images suited for red-green anaglyph glasses (see Figure 2, right hand side). In this condition, the observer wore red-green anaglyphs (a red filter for the left eye and a green filter for the right eye), which make the green objects are only visible to the right eye and the red objects only visible to the left eye[1]. The target always appeared in a plane behind the image plane, and the observer was aware of this.
Three conditions were run:

1) A condition with black and white images with 1 target and 29 distractors
2) A condition with stereo images with 15 distractors in one plane, and 1 target & 14 disttactors in a plane behind the image plane.
3) A condition with black and white images with 1 target and 14 distractors

By comparing conditions 1 and 2 one can see how much search performance improves when half of the distractors is placed in a different plane. By comparing conditions 2 and 3 one can determine whether search performance is as good with two depth planes as with a single depth plane.

[1] The colour of the background was chosen such that cross-talk was minimised.

RESULTS

The distribution of detection times are shown in Figure 4. The top shows the results from observer MAH and the bottom figure shows the results from observer PB. Only the trials are used in which the target was indicated correctly (almost all of the trials). The figure shows the number of trials for which a response time fell within a certain range (they are binned). This shows the distribution. Shown are the results for conditions 2 ("two planes"), 1 ("single plane") and 3 ("single plane, N=15"), see method section.

Table 1 shows the summarised data. Figure 4 shows that the detection times are well described by a gaussian distribution on a logarithmic scale. Therefore, table 1 shows the *geometric* means (averaged on a logarithmic scale) with their standard error. Also shown are the medians and the 1st and 3rd quartiles, which is independent of assumptions of the distribution.

The (geometric) mean detection times are plotted in Figure 3 for the three conditions observers MAH and PB. The mean detection time is approximately halved when half of the distractors is placed in another depth plane (condition 2, "two planes, N=30" vs. condition 1, "single plane, N=30"). The detection time is roughly the same (but slightly lower) when the second plane is omitted altogether.

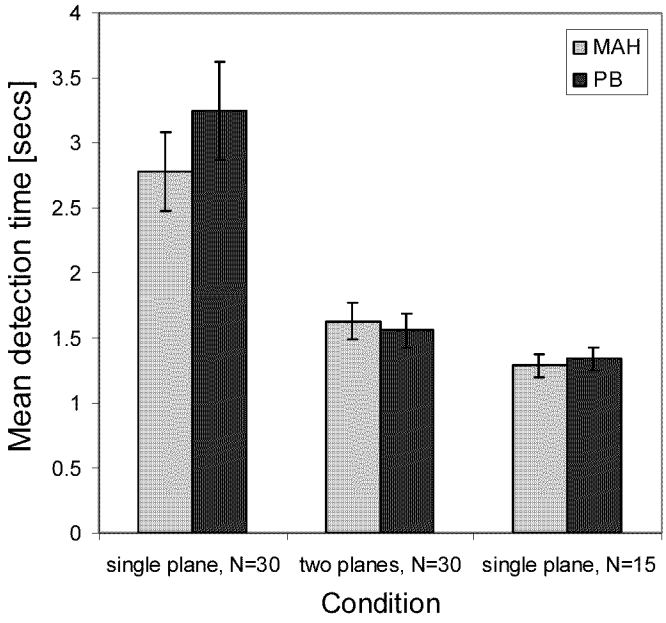


Figure 3. The (geometric) mean search time for the different condition for both observers.

CONCLUSION

The results show that search can be (largely) restricted to the regions of interest (potential targets). Highlighting potential targets in this way ensures high search performance and facilitates good localisation of the target within its context.

TABLE 1. Summary of the results. From left to right: observer, condition, total number of objects N, the geometric average of the detection times, the standard error in the mean detection time, the standard deviation of the detection times (in %), the median and the 1st and 3rd quartiles of the detection times.

observer	condition	N	geom.mean	SE	SD (%)	median	1st quartile	3rd quartile
MAH	single plane	30	2.78	0.3	75	2.86	1.53	4.33
	two planes	30	1.63	0.14	62	1.71	1.22	2.47
	single plane	15	1.29	0.09	51	1.22	0.88	1.79
PB	single plane	30	3.25	0.37	81	3.19	1.69	6.24
	two planes	30	1.56	0.13	61	1.53	0.94	2.42
	single plane	15	1.34	0.09	48	1.31	0.92	1.63

ACKNOWLEDGEMENTS

This material is based upon work supported by the European Office of Aerospace Research and Development, Air Force Office of Scientific Research, Air Force Research Laboratory, under contract No. F61775-01-WE026, and by Senter, Agency of the Ministry of Economic Affairs of the Netherlands.

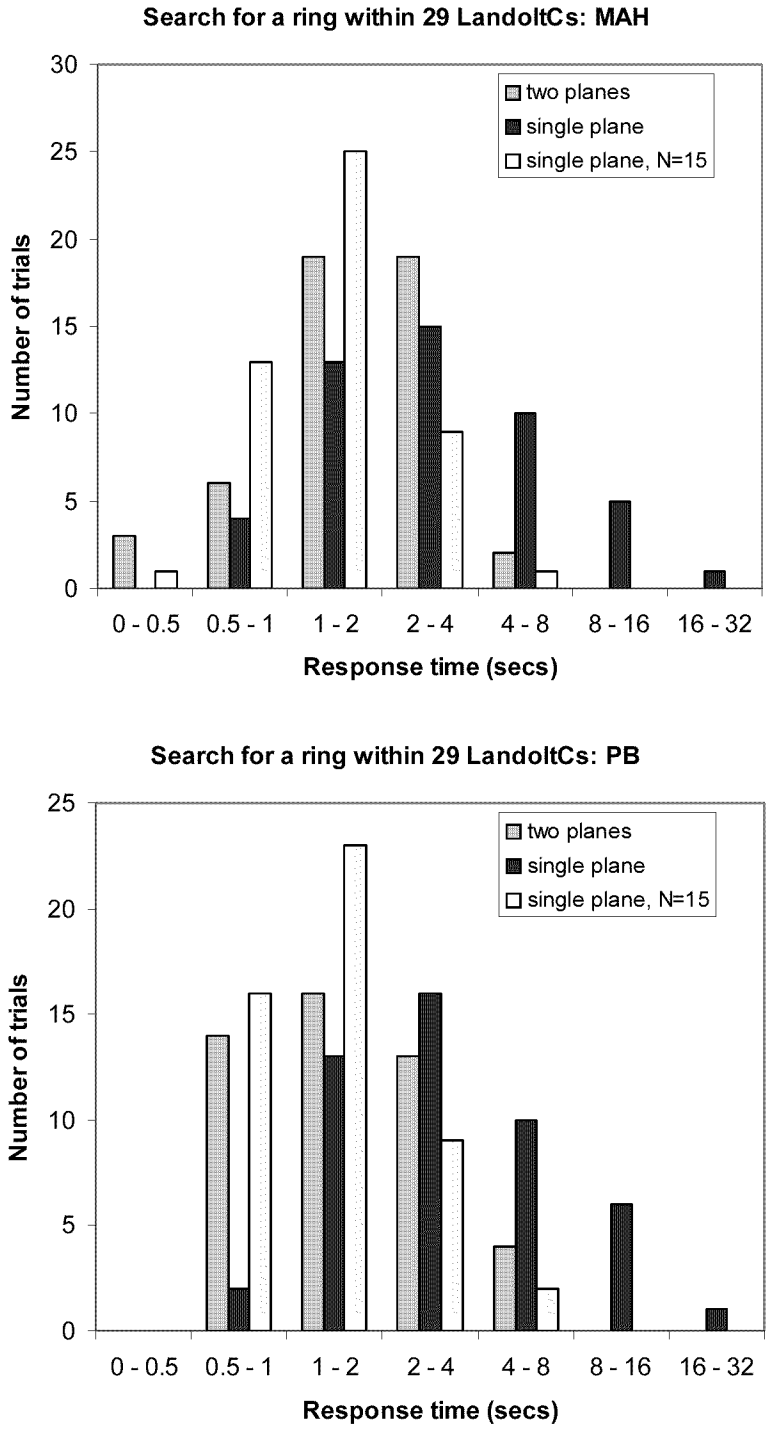


Figure 4. The distributions of search times for observers MAH (top) and PB (bottom) for the three conditions: two plane (stereo images, 15 distactors in one plane, 14 distactors plus the target in another), single plane (black and white images with 1 target and 29 distractors), and single plane N=15 (black and white images with 1 target and 14 distractors). Note that the response time is plotted on a logarithmic scale. Trials in which the wrong target was indicated are omitted.

Appendix D

Perceptual Evaluation of Different Image Fusion Schemes

Alexander Toet Eric M. Franken
TNO Human Factors
Kampweg 5
3769 DE Soesterberg
The Netherlands
Phone +31-3463-56237
Fax +31-3463-53977
Email toet@tm.tno.nl

ABSTRACT

Human scene recognition performance was tested with images of nighttime outdoor scenes. The scenes were registered both with a dual band (visual and near infrared) image intensified low-light CCD camera (DII) and with a thermal middle wavelength band (3-5 μm) infrared (IR) camera. Fused imagery was produced through a grayscale pyramid image merging scheme, in combination with two different colour mappings. Observer performance was tested for each of the (individual and fused) image modalities. The results show that DII imagery contributes most to global scene recognition (situational awareness), whereas IR imagery serves best for the detection and recognition of targets like humans and vehicles. Grayscale fused imagery yields appreciable performance levels in most conditions. With an appropriate colour mapping, colour fused imagery yields the best overall scene recognition performance. However, an inappropriate colour mapping significantly decreases observer performance compared to grayscale image fusion. The deployment of a DII system in addition to a 3-5 μm IR system through image fusion can increase the performance of human observers when the colour mapping relates to the nature of the visual task and the conditions (scene content) at hand.

Keywords: Image fusion, infrared, intensified imagery, scene recognition, situational awareness

1 INTRODUCTION

Modern night-time camera's are designed to expand the conditions under which humans can operate. A functional piece of equipment must therefore provide an image that leads to good perceptual awareness in most environmental and operational conditions (to "Own the weather " or "Own the night"). The two most common nighttime imaging systems either display emitted infrared (IR) radiation or reflected light, and thus provide complimentary information of the inspected scene. IR cameras have a history of decades of development. Although modern IR cameras function very well under most circumstances, they still have some inherent limitations. For instance, after a period of extensive cooling (e.g. after a long period of rain) the infrared bands provide less detailed information due to low thermal contrast in the scene, whereas the visual bands may represent the background in great detail (vegetation or soil areas, texture). In this situation it can be hard or even impossible to distinguish the background of a target in the scene, using only the infrared bands, whereas at the same time, the target itself may be highly detectable (when its temperature differs sufficiently from the mean temperature of its local background). On the other hand, a target that is well camouflaged for visual detection will be hard (or even impossible) to detect in the visual bands, whereas it can still be detectable in the thermal bands. A combination of visible and thermal imagery may then allow both the detection and the unambiguous localisation of the target (represented in the thermal image) with respect to its background (represented in the visual image). A human operator using a suitably combined or fused representation of IR and (intensified) visual imagery may therefore be able to construct a more complete mental representation of the perceived scene, resulting in a larger degree of situational awareness [12].

This study was performed to test the complementarity of information, obtained from different types of night vision systems (IR and image intensifiers), and the capability of several greylevel and colour image fusion schemes applied to these images to combine and convey information, about both the global structure and the fine detail of scenes, to human observers. The image modalities used were conventional single band as well as dual band (visual and near infrared) intensified low-light CCD images (II and DII, respectively) and thermal middle wavelength band (3-5 μm) infrared images. Three different image fusion schemes were investigated. Colour and grayscale fused imagery was produced through a conventional pyramid image merging scheme, in combination with two different colour mappings. This fusion approach is representative for a number of methods that have been suggested in the literature [1, 7, 10, 13-15], and may serve as a starting point for further developments. Observer performance with the individual image modalities serves as a baseline for the performance that should be obtained with fused imagery. The results of these tests indicates to what extent DII and IR images are complementary, and can be used to identify the characteristic features of each image modality that determine human visual performance. The goal of image fusion is to combine and preserve in a single output image all the perceptually important signal information that is present in the individual input images. Hence, for a given observation task, performance with fused imagery should at least be as good (and preferably better) as performance with the individual image modality that yields the optimal performance for that task. Knowledge of the nature of the features in each of the input images that determine observer performance can be used to develop new multimodal image visualisation techniques, based on improved image fusion schemes that optimally exploit and combine the perceptually relevant information from each of the individual nighttime image modalities.

2 METHODS

2.1 Scene recording

A variety of outdoor scenes, displaying several kinds of vegetation (grass, heather, semi shrubs, trees), sky, water, sand, vehicles, roads, and persons, were registered at night with a recently developed dual-band visual intensified (DII) camera (see below), and with a state-of-the-art thermal middle wavelength band (3-5 μm) infrared (IR) camera (Radiance HS). Both cameras had a field of view (FOV) of about 6x6 degrees. Some image examples are shown in Figures 1-5.

The DII camera was developed by Thales Optronics and facilitated a two-colour registration of the scene, applying two overlapping bands covering the part of the electromagnetic spectrum ranging from visual to near infrared (400-900 nm). The short (visual) wavelength part of the incoming spectrum is mapped to the R channel of an RGB colour composite image. The long (near infrared) wavelength band corresponds primarily to the spectral reflection characteristics of vegetation, and is therefore mapped to the G channel of an RGB colour composite image. This approach utilises the fact that the spectral reflection characteristics of plants are distinctly different from other (natural and artificial) materials in the visual and near infrared range [5]. The spectral response of the long-wavelength channel ('G') roughly matches that of a Generation III image intensifier system. This channel is stored separately and used as an individual image modality (II).

Images were recorded at various times of the diurnal cycle under various atmospheric conditions (clear, rain, fog, …) and for various illumination levels (1 lux – 0.1 mlux). Object ranges up to several hundreds of meters were applied. The images were digitized on-site (using a Matrox Genesis frame grabber, using at least 1.8 times oversampling).

2.2 Stimuli

First, the recorded images were registered through an affine warping procedure, using fiducial registration points that were recorded at the beginning of each session. After warping, corresponding pixels in images taken with different cameras represent the same location in the recorded scene. Then, patches displaying different types of scenic elements were selected and cut out from corresponding images (i.e. images representing the same scene at the same instant in time, but taken with different cameras). These patches were deployed as stimuli in the pyschophysical tests. The signature of the target items (i.e. buildings, persons, vehicles etc.) in the image test sets varied from highly distinct to hardly visible.

To test the *perception of detail*, patches were selected that display either buildings, vehicles, water, roads, or humans. These patches are 280×280 pixels, corresponding to a FOV of 1.95×1.95 deg.
To investigate the *perception of global scene structure*, larger patches were selected, that represent either the horizon (to perform a horizon perception task), or a large amount of different terrain features (to enable the distinction between an image that is presented upright and one that is shown upside down). These patches are 575×475 pixels, corresponding to a FOV of 4.0×3.3 deg.

To test if the combined display of information from the individual image modalities may enhance the perception of detail (target recognition) and situational awareness, corresponding stimulus pairs (i.e. patches representing the same part of a scene at the same instant in time, but taken with different cameras) were fused.

Grayscale fused (GF) images were produced by combining the IR and II images though a pyramidal image fusion scheme [1, 10, 13]. A 7-level Laplacian pyramid [1] was used, in combination with a maximum absolute contrast node (i.e. pattern element) selection rule.

Colour fused imagery was produced by the following two methods.

- *Colour Fusion Method 1* (CF1): The short and long wavelength bands of the DII camera were respectively mapped to the R and G channels of an RGB colour image. The resulting RGB colour image was then converted to the YIQ (NTSC) colour space. The luminance (Y) component was replaced by the corresponding aforementioned grayscale (II and IR) fused image, and the result was transformed back to the RGB colour space (note that the input Y from combining the R and G channel is replaced by a Y which is created by fusing the G channel with the IR image). This colour fusion method results in images in which grass, trees and persons are displayed as greenish, and roads, buildings, and vehicles are brownish.
- *Colour Fusion Method 2* (CF2): First, an RGB colour image was produced by assigning the IR image to the R channel, the long wavelength band of the DII image to the green channel (as in Method 1), and the short wavelength band of the DII image to the blue channel (instead of the red channel, as in Method 1). This colour fusion method results in images in which vegetation is displayed as greenish, persons are reddish, buildings are red-brownish, vehicles are whitish/bluish, and the sky and roads are most often bluish.

The multiresolution grayscale image fusion scheme employed here, selects the perceptually most salient contrast details from both of the individual input image modalities, and fluently combines these pattern elements into a resulting (fused) image. As a side effect of this method, details in the resulting fused images can be displayed at higher contrast than they appear in the images from which they originate, i.e. their contrast may be enhanced [9, 11]. To distinguish the perceptual effects from contrast enhancement from those of the fusion process, observer performance was also tested with contrast enhanced versions of the individual image modalities. The contrast in these images was enhanced by a multiresolution local contrast enhancement scheme. This scheme enhances the contrast of perceptually relevant details for a range of spatial scales, in a way that is similar to the approach used in the hierarchical fusion scheme. A detailed description of this enhancement method is given elsewhere [9, 11].

2.3 Apparatus used for stimuli display

A Pentium II 400 MHz computer was used to present the stimuli, measure the response times and collect the observer responses. The stimuli were presented on a 17 inch Vision Master 400 (Iiyama Electric Co., Ltd) colour monitor, using the 1152×864 true colour (32 bit) mode (corresponding to a resolution of 36.2 pixels/cm), with a colour temperature of 6500 K, and a 100 Hz refresh rate.

2.4 Tasks

The perception of the global structure of a depicted scene was tested in two different ways. In the first test, scenes were presented that had been randomly mirrored along the horizontal, and the subjects were asked to distinguish the orientation of the displayed scenes (i.e. whether a scene was displayed right side up or upside down). In this test, each scene was presented twice: once upright and once upside down. In the second test, horizon views were presented together with short markers (55×4 pixels) on the left and right side of the image and on a virtual horizontal line. In this test, each scene was presented twice: once with the markers located at the true position (height) of the horizon, and once when the markers coincided with a horizontal structure that was opportunistically available (like a band of clouds) and that may be mistaken for the horizon. The task of the subjects was to judge whether the markers indicated the true

position of the horizon. The perception of the global structure of a scene is likely to determine situational awareness.

The capability to discriminate fine detail was tested by asking the subjects to judge whether or not a presented scene contained an exemplar of a particular category of objects. The following categories were investigated: buildings, vehicles, water, roads, and humans. The perception of detail is relevant for tasks involving visual search, detection and recognition.
The tests were blocked with respect to both (1) the imaging modality and (2) the task. This was done to minimise observer uncertainty, both with respect to the characteristics of the different image modalities, and with respect to the type of target.

Blocking by image modality yielded the following six classes of stimuli:

1. Grayscale images representing the thermal 3-5 μm IR camera signal.
2. Grayscale images representing the long-wavelength band (G-channel) of the DII images.
3. Colour (R and G) images representing the two channels of the DII.
4. Grayscale images representing the IR and II signals fused by GF.
5. Colour images representing the IR and DII signals fused by CF1.
6. Colour images representing the IR and DII signals fused by CF2.

Blocking by task resulted in trial runs that tested the perception of global scene structure by asking the observers to judge whether

- the horizon was veridically indicated
- the image was presented right side up

and the recognition of detail by asking the observers to judge whether the image contained an exemplar of one of the following categories:

- building
- person
- road or path
- fluid water (e.g. a ditch, a lake, a pond, or a puddle)
- vehicle (e.g. a truck, car or van)

The entire experiment consisted of 42 different trial runs (6 different image modalities $\times$ 7 different tasks). Each task was tested on 18 different scenes. The experiment therefore involved the presentation of 756 images in total. The order in which the image modalities and the tasks were tested was randomly distributed over the observers.

2.5 Procedure

Before starting the actual experiment, the observers were shown examples of the different image modalities that were tested. They received verbal information, describing the characteristics of the particular image modality. It was explained how different types of targets are displayed in the different image modalities. This was done to familiarise the observers with the appearance of the scene content in the different images modalities, thereby minimising their uncertainty.

Next, subjects were instructed that they were going to watch a sequence of briefly flashed images, and that they had to judge each image with respect to the task at hand. For a block of trials, testing the perception of detail, the task was to judge whether or not the image showed an exemplar of a particular category of targets (e.g. a building). For a block of trials, testing the perception of the overall structure of the scene, the task was to judge whether the scene was presented right side up, or whether the position of the horizon was indicated correctly. The subjects were instructed to respond as quickly as possible after the onset of a stimulus presentation, by pressing the appropriate one of two response keys.

Each stimulus was presented for 400 ms. This brief presentation duration, in combination with the small stimulus size, served to prevent scanning eye movements (which may differ among image modalities and target types), and to force subjects to make a decision based solely on the instantaneous percept aroused by the stimulus presentation. Immediately after the stimulus presentation interval, a random noise image was shown. This noise image remained visible for at least 500 ms. It served to erase any possible afterimages (reversed contrast images induced by, and lingering on after, the presentation of the stimulus, that may differ in quality for different image modalities and target types), thereby equating the processing time subjects can use to make their judgement. Upon each presentation, the random noise image was randomly left/right and up/down reversed. The noise images had the same dimensions as the preceding stimulus image, and consisted of randomly distributed sub-blocks of 5×5 pixels. For trial blocks testing the monochrome IR and II imaging modalities and grayscale fused imagery, the noise image sub-blocks were either black or mean grey. For trial blocks testing DII and colour fused imagery, the noise image sub-blocks were randomly coloured, using a color palette similar to that of the modality being tested. In all tests, subjects were asked to quickly indicate their visual judgement by pressing one of two response keys (corresponding to a YES/NO response), immediately after the onset of a stimulus image presentation. Both the accuracy and the reaction time were registered.

2.6 Subjects

A total of 12 subjects, aged between 20 and 55 years, served in the experiments reported below. All subjects have corrected to normal vision, and reported to have no colour deficiencies.

2.7 Viewing conditions

The experiments were performed in a dimly lit room. The images are projected onto the screen of the CRT display. Viewing was binocular, at a distance of 60 cm. At this distance, the images subtended a viewing angle of either 14.8×12.3 or 7.3×7.3 degrees, corresponding to a scene magnification of 3.8.

3 RESULTS

This section reports the results of the observer experiments for the different tasks and for each of the aforementioned image modalities. The first two tasks measure the degree to which the scene structure is correctly perceived. The remaining 5 tasks measure the perception of detail.

For each visual discrimination task the numbers of hits (correct detections) and false alarms (fa) were recorded to calculate $d' = Z_{hits}-Z_{fa}$, an unbiased estimate of sensitivity [4].

The effects of contrast enhancement on human visual performance is found to be similar for all tasks. Figure 6 shows that contrast enhancement significantly improves the sensitivity of human observers performing with II and DII imagery. However, for IR imagery, the average sensitivity decreases as a result of contrast enhancement. This is probably a result of the fact that the contrast enhancement method employed in this study increases the visibility of irrelevant detail and clutter in the scene. Note that this result does *not* indicate that (local) contrast enhancement in general should not be applied to IR images.

Figure 7 shows the results of all scene recognition and target detection tasks investigated here. As stated before, the ultimate goal of image fusion is to produce a combined image that displays more information than either of the original images. Figure 7 shows that this aim is only achieved for the following perceptual tasks and conditions:

- the detection of roads, where CF1 outperforms each of the input image modalities,
- the recognition of water, where CF1 yields the highest observer sensitivity, and
- the detection of vehicles, where three fusion methods tested perform significantly better than the original imagery.

These tasks are also the only ones in which CF1 performs better than CF2. An image fusion method that always performs at least as good as the best of the individual image modalities can be of great ergonomic value, since the observer can perform using only a single image. This result is obtained for the recognition of scene orientation from colour fused imagery produced with CF2, where performance is similar to that with II and DII imagery. For the detection of buildings and humans in a scene, all three fusion methods perform equally well and slightly less than IR. CF1 significantly outperforms grayscale fusion for the detection of the horizon and the recognition of roads and water. CF2 outperforms grayscale fusion for both global scene recognition tasks (orientation and horizon detection). However, for CF2 observer sensitivity approaches zero for the recognition of roads and water.

Rather surprisingly, the response times (not shown here) did not differ significantly between all different image modalities. The shortest reaction times were obtained for the detection of humans (about 650 ms), and the longest response times were found for the detection of the position of the horizon (about 1000 ms).

The following section discusses the results in detail for each of the seven different perception tasks.

3.1 Perception of global structure

The perception of the scene layout was tested by measuring the accuracy with which observers were able to distinguish a scene that was presented right side up from one that was presented upside down, and perceive the position of the horizon.

The first group of bars in Figure 7 (labelled "upright") represents the results for the scene orientation perception task. For the original image modalities, the best results are obtained with the intensified imagery (the II performed slightly better than the DII). The IR imagery performs significantly worse. CF2 performs just as well as II, whereas CF1 performs similar to IR. Graylevel fusion is in between both colour fusion methods. Observers remarked that they based their judgement mainly on the perceived orientation of trees and branches in the scene. CF2 displays trees with a larger colour contrast (red-brown on a light greenish or bluish background) than CF1 (dark green trees on a somewhat lighter green background), resulting in a better orientation detection performance. Also, CF2 produces bright blue skies most of the time, which makes the task more intuitive.

The perception of the true position of the horizon, represented by the second group of bars in Figure 2, is best performed with II imagery, followed by the DII modality. Both intensified visual image modalities perform significantly better than IR or any kind of fused imagery. The low performance with the IR imagery is probably a result of the fact that a tree line and a band of clouds frequently have a similar appearance in this modality. The transposition of these 'false horizons' into the fused image modalities significantly reduces observer performance. For greylevel fused imagery, the observer sensitivity is even reduced to a near-zero level, just as found for IR. Colour fused imagery restores some of the information required to perform the task, especially CF2 that produces blue skies. However, the edges of the cloud bands are so strongly represented in the fused imagery that observer performance never attains the sensitivity level obtained for the intensified visual modalities alone (II and DII).

In both the orientation and horizon perception tasks subjects tend to confuse large bright areas (e.g. snow on the ground) with the sky.

3.2 Perception of detail

The best score for the recognition of *buildings* is found for IR imagery. In this task, IR performs significantly better than II or DII. DII imagery performs significantly better than II, probably because of the colour contrast between the buildings and the surrounding vegetation (red-brown walls on a green background, compared to grey walls on a grey background in case of the II imagery). The performance with fused imagery is slightly less than with IR, and independent of the fusion method.

The detection of *humans* is best performed with IR imagery, in which they are represented as white hot objects on a dark background. II imagery yields a very low sensitivity for this task; i.e. humans are hardly ever noticed in intensified visual imagery. The sensitivity for the detection of humans in DII imagery is somewhat higher, but remains far below that found for IR. In this case, there is almost no additional information in the second wavelength band of the DII modality, and therefore almost no additional colour contrast. As a result, most types of clothing are displayed as greenish, and are therefore hard to distinguish from vegetation. Performance with fused imagery is only slightly below that with IR. There is no significant difference between the different grayscale and colour fusion types.

Roads cannot reliably be recognized from IR imagery (d' becomes even negative, meaning that more false alarms than correct detections are scored). DII performs best of the individual image modalities, and significantly higher than II because of the additional colour contrast (DII displays roads as red-brown, on a green background). Grayscale fused imagery results in a performance that is significantly below that found for DII, and somewhat lower than that obtained for II imagery. This is probably a result of *(1)* the introduction of irrelevant luminance details from the IR imagery, and *(2)* the loss of colour contrast as seen in the DII imagery. CF1 produces colour fused imagery that yields a higher sensitivity than each of the original image modalities, although observer performance is not significantly better than with DII imagery. The additional improvement obtained with this combination scheme is probably caused by the contrast enhancement inherent in the fusion process. The sensitivity obtained for imagery produced by CF2 is near zero. This is probably a result of the fact that this method displays roads with a light blue colour. These can therefore easily be mistaken for water or snow. This result demonstrates that the inappropriate use of colour in image fusion severely degrades observer performance.

Image fusion clearly helps to recognize *vehicles* in a scene. They are best discriminated in colour fused images produced with CF1, that displays vehicles in brown-yellow on a green background. CF2 (that shows vehicles as blue on a brown and green background) and grayscale fusion both result in an equal and somewhat lower observer sensitivity. Fused imagery of all types performs significantly better than each of the original image modalities. The lowest recognition performance is obtained with IR imagery.

Water is best recognised in colour fused imagery produced with CF1. This method displays water sometimes as brown-reddish, and sometimes as greyish. The II, DII and greylevel fusion scheme all yield a similar and slightly lower performance. CF2 results on a near zero observer sensitivity for this task. This method displays water sometimes as purple-reddish, thus giving it a very unnatural appearance, and sometimes as bluish, which may cause confusion with roads, that have the same colour. These results again demonstrate that it is preferable not to use any colour at all (grayscale), than to use an inappropriate colour mapping scheme.

3.3 Summary

Table 1 summarizes the main findings of this study. IR has the lowest overall performance of all modalities tested. This results from a low performance for both large scale orientation tasks, and for the detection and recognition of roads, water, and vehicles. In contrast, intensified visual imagery performs best in both orientation tasks. The perception of the horizon is significantly better with II and DII imagery. IR imagery performs best for the perception and recognition of buildings and humans- DII has the best overall performance of the individual image modalities. Thus, IR on one hand and (D)II images on the other hand contain *complementary* information, which makes each of these image modalities suited for performing different perception tasks.

CF1 has the best overall performance of the image fusion schemes tested here. The application of an appropriate colour mapping scheme in the image fusion process can indeed significantly improve observer performance compared to grayscale fusion. In contrast, the use of an inappropriate colour scheme can severely degrade observer sensitivity. Although the performance of CF1 for specific observation tasks is below that of the optimal individual sensor, for a combination of observation tasks (as will often be the case in operational scenarios) the CF1 fused images can be of great ergonomic value, since the observer can perform using only a single image.

Table 1 The relative performance of the different image modalities for the seven perceptual recognition tasks. Rank orders -1,1, and 2 indicate respectively the worst, second best, and best performing image modality for a given task. The tasks involve the perception of the global layout (orientation and horizon) of a scene, and the recognition of local detail (buildings, humans, roads, vehicles, and water). The different image modalities are: infrared (IR), greyscale (II) and dual band false-colour (DII) intensified visual, grayscale fused images (GF) and two different colour fusion (CF1, CF2) schemes. The sum of the rank orders indicates the overall performance of the modalities.

	IR	II	DII	GF	CF1	CF2
Upright	-1	2	1			2
Horizon	-1	2	1			
Building	2	-1		1	1	1
Human	2	-1		1	1	1
Road	-1		1		2	
Vehicle	-1			2	2	1
Water	-1				2	
Overall	-1	2	3	4	8	5

4 CONCLUSIONS

Nightime images recorded using an image intensified low-light CCD camera and a thermal middle wavelength band (3-5 μm) infrared camera contain *complementary* information. This makes each of the individual image modalities only suited for specific observation task. However, the complementarity of the information of the image modalities can be exploited using image fusion, which would enable multiple observation tasks using a single nighttime image representation.

Since there evidently exists no one-to-one mapping between the temperature contrast and the spectral reflectance of a material, the goal of producing a nighttime image, incorporating information from IR imagery, with an appearance similar to a colour daytime image can never be fully achieved. The options are therefore *(1)* to settle for a single mapping that works satisfactory in a large number of conditions, or *(2)* to adapt (optimize) the colour mapping to the situation at hand. However, the last option is not very attractive since a different colour mapping for each task and situation tends to confuse observers [3, 8].

Multimodal image fusion schemes based on local contrast decomposition do not distinguish between material edges and temperature edges. For many tasks, material edges are the most important ones. Fused images frequently contain an abundance of contours that are irrelevant for the task that is to be performed. Fusion schemes incorporating some kind of contrast stretching enhance the visibility of all details in the scene, irrespective of their visual significance. The introduction of spurious or irrelevant contrast elements in a fused image may clutter the scene, distract the observer and lead to misinterpretation of perceived detail. As a result, observer performance may degrade significantly. A useful image fusion scheme should therefore take into account the visual information content (meaning) of the edges in each of the individual image modalities, and combine them accordingly in the resulting image.

For most perceptual tasks investigated here (except for horizon and road detection), grayscale image fusion yields appreciable performance levels. When an appropriate colour mapping

scheme is applied, the addition of colour to grayscale fused imagery can significantly increase observer sensitivity for a given condition and a certain task (e.g. Colour Fusion Method 2 for orientation detection, both colour fusion methods for horizon detection, Colour Fusion Method 1 for road and water detection). However, inappropriate use of colour can significantly decrease observer performance compared to straightforward grayscale image fusion (e.g. Colour Fusion Method 2 for the detection of roads and water).

For the observation tasks and image examples tested here, optimal overall performance was obtained for images fused using Color Fusion Method 1. The overall performance was higher than for either of the individual image modalities. Note that in this fusion method, no colour mapping is applied to the IR information. Instead, the IR information is blended into the image without changing the colour.

The present findings agree with those from previous studies [2, 3, 6, 8]. The present results will be analysed further

1. to distinguish perceptually relevant features from noise and distracting elements, and
2. to find out if there are features that are consistently mistaken by subjects for another type of scenic detail.
3. to further optimise image fusion techniques (with or without using colour-mapping)

Acknowledgements

This material is partly based upon work supported by the European Office of Aerospace Research and Development, Air Force Office of Scientific Research, Air Force Research Laboratory, under contract No. F61775-01-WE026, and by Senter, Agency of the Ministry of Economic Affairs of the Netherlands.

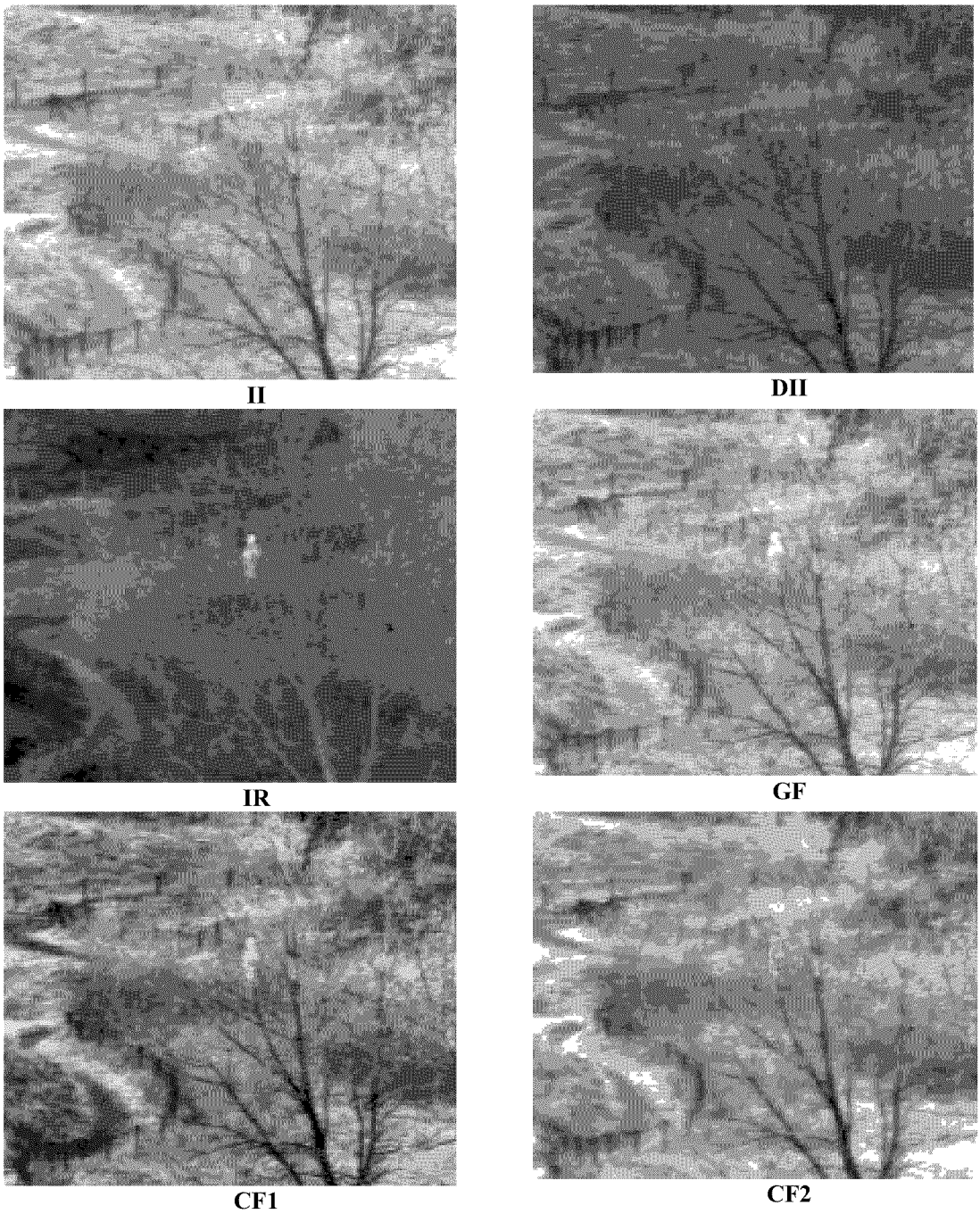


Figure 1 The different image modalities used in this study. II and DII: the long wavelength band and both bands of the false colour intensified CCD image. IR: the thermal 3-5 μm IR image. GF: the greylevel fused image and CF1(2) and colour fused images produced with Method 1(2). This image shows a scene of a person in terrain, behind a tree. .

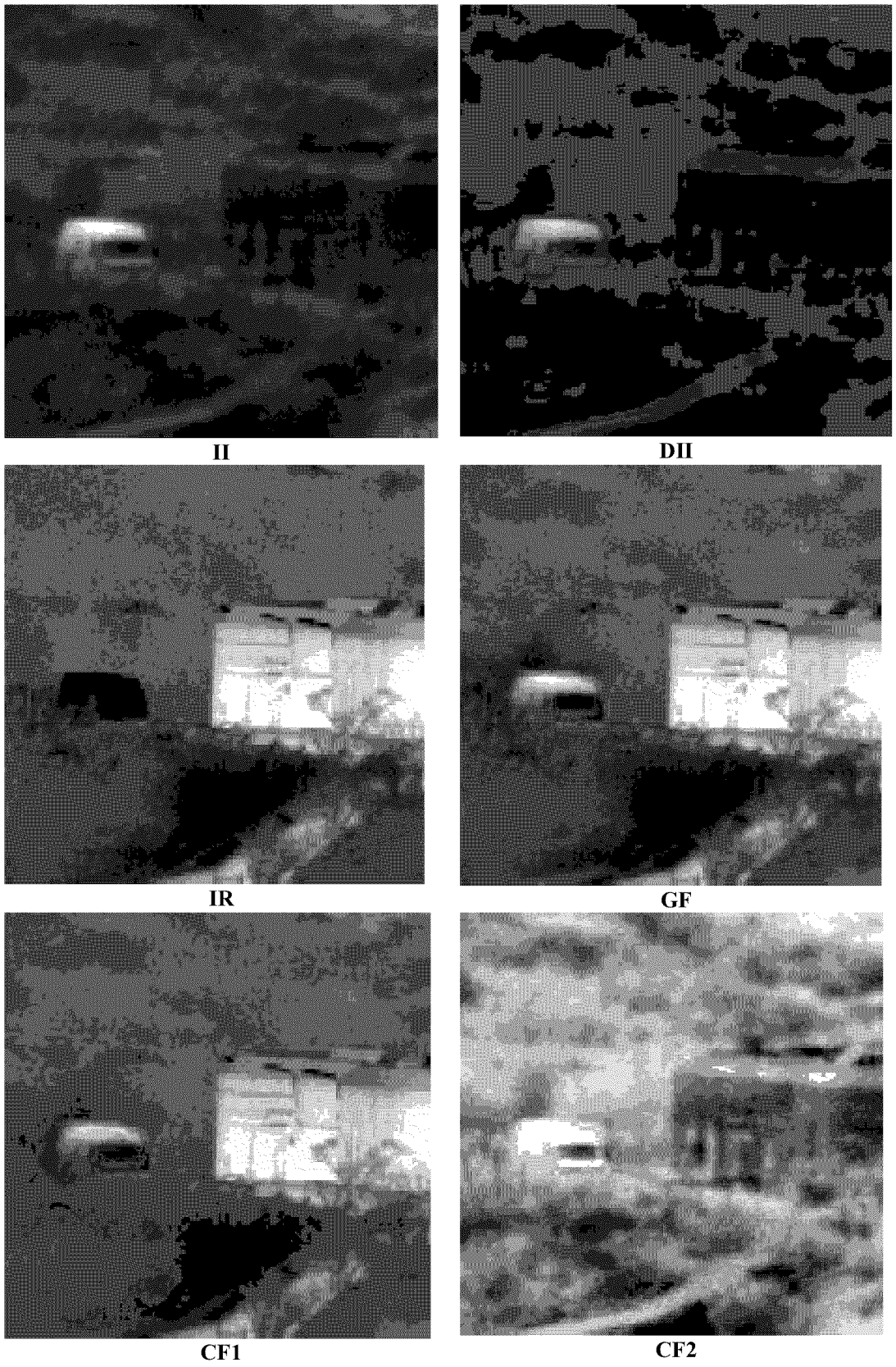


Figure 2 As Figure 1, for a scene displaying a road, a house, and a vehicle.

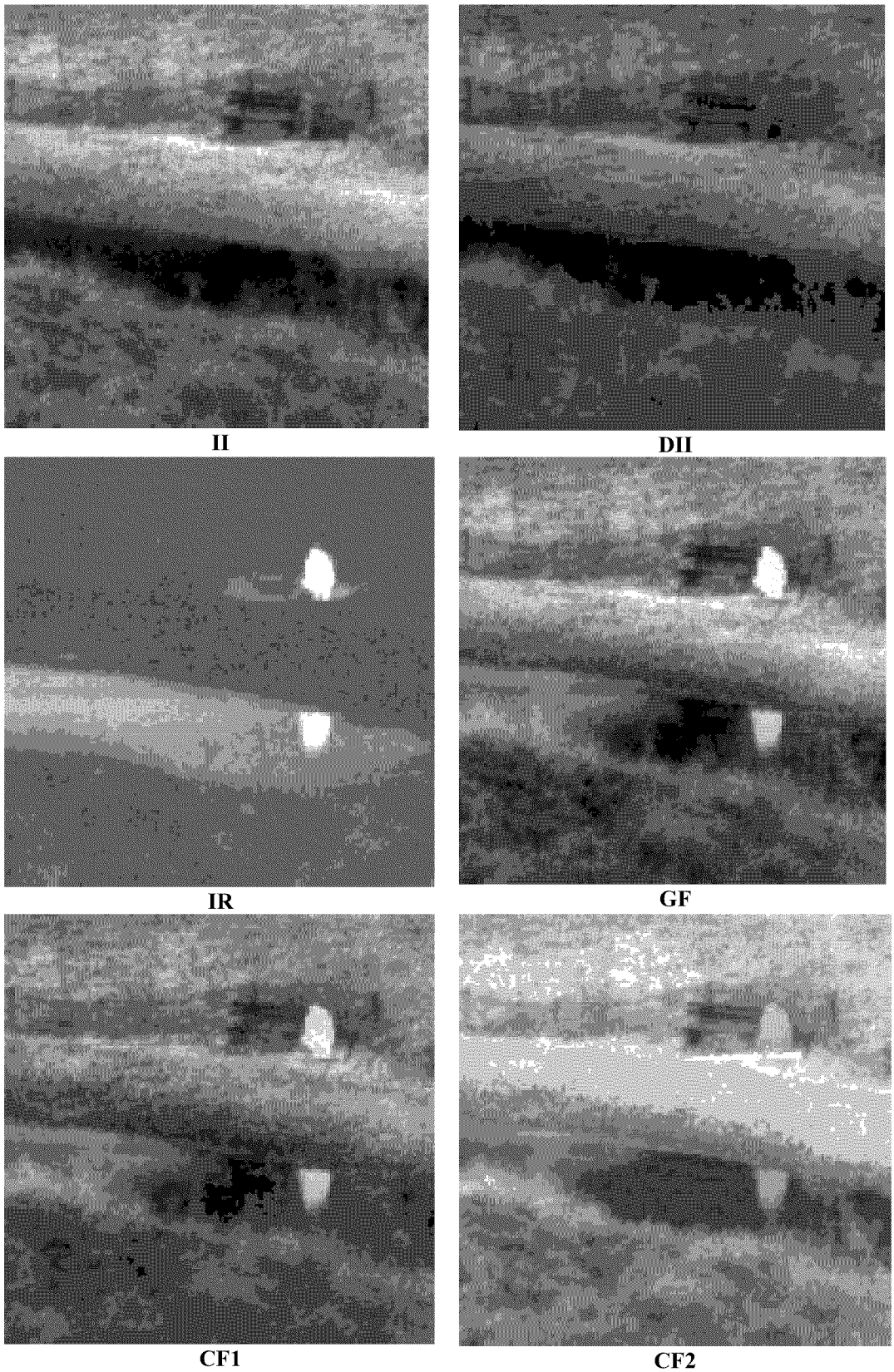


Figure 3 As Figure 1, for a scene displaying a person along a riverside. Notice the reflection of the person's silhouette on the water surface in the thermal image.

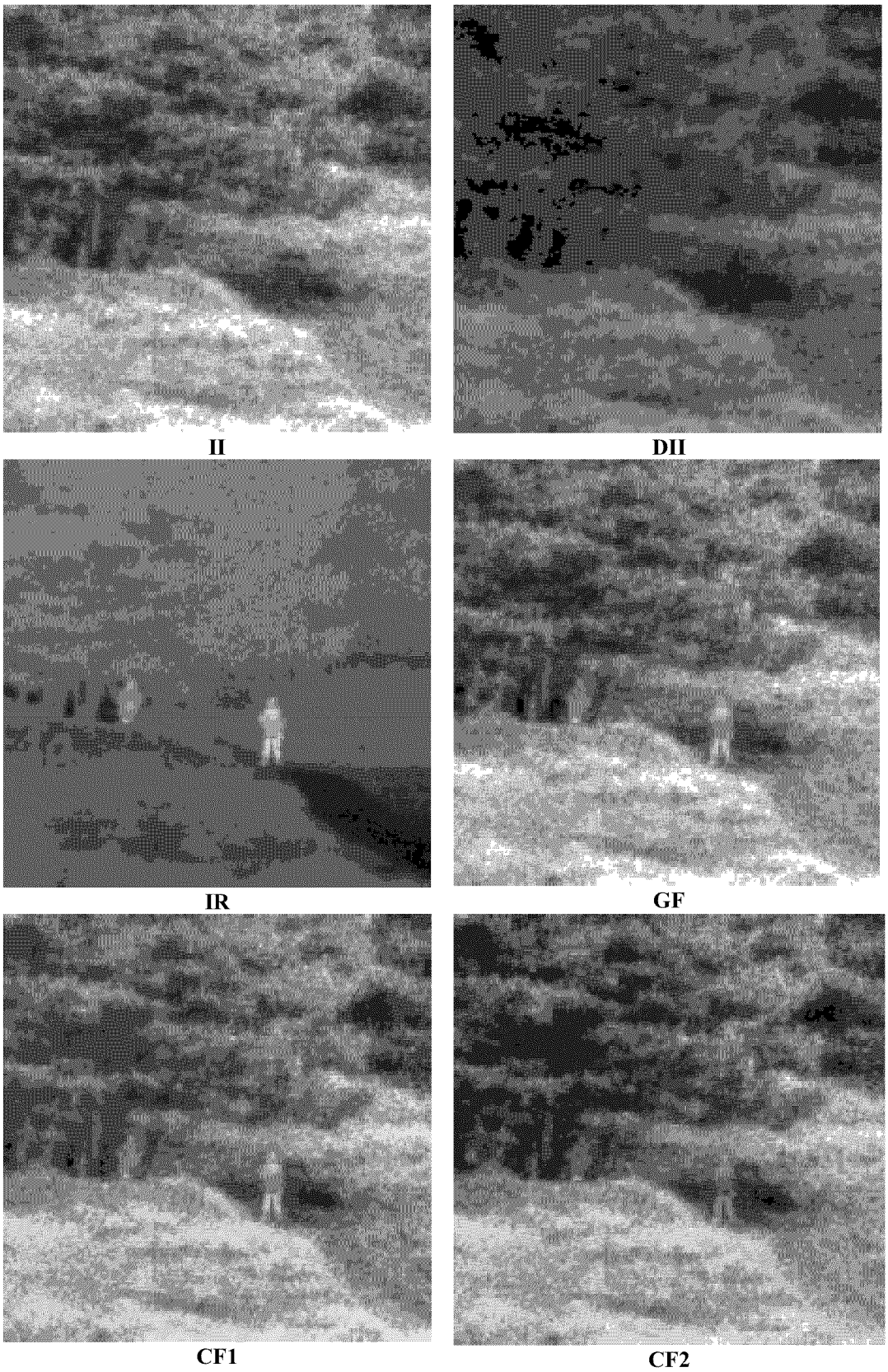


Figure 4 As Figure 1, for a scene displaying people on a road through the woods.

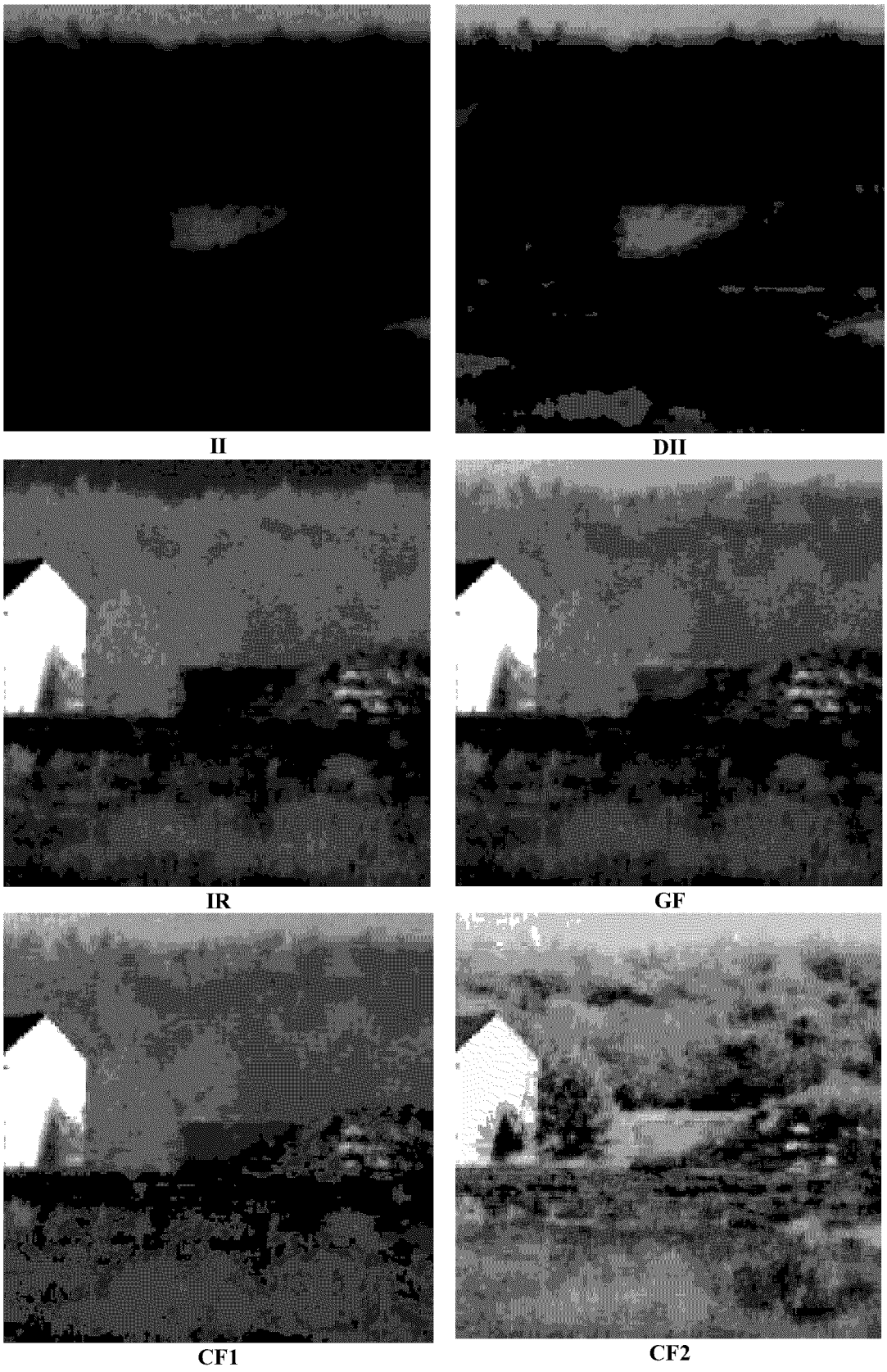


Figure 5 As Figure 1, for a scene displaying a house and trees.

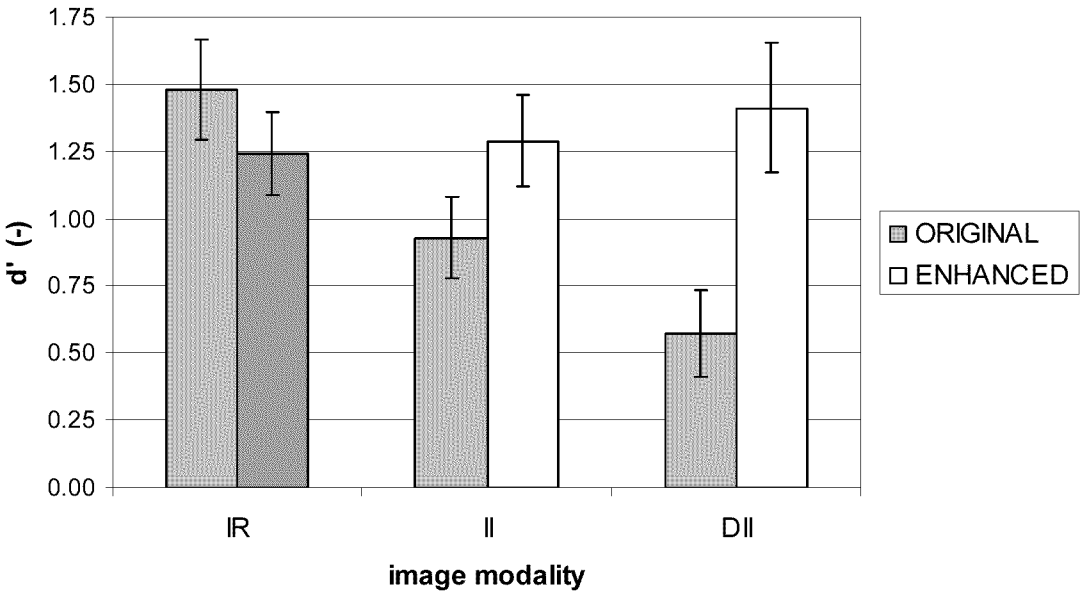


Figure 6 The effect of contrast enhancement on observer sensitivity d'.

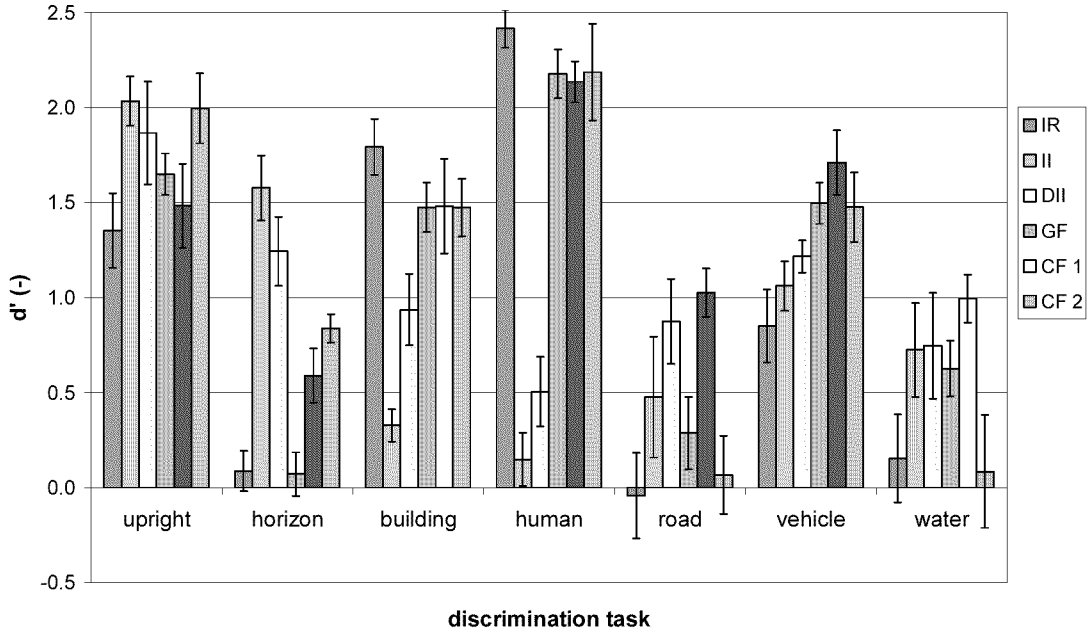


Figure 7 Observer sensitivity d' for discrimination of global layout (orientation and horizon) and local detail (buildings, humans, roads, vehicles, and water), for six different image modalities. These modalities are (in the order in which they appear in the labeled clusters above): infrared (IR), single-band or grayscale (II) and double-band or colour (DII) intensified visual, grayscale (GF) and colour fused (CF1, CF2) imagery.

REFERENCES

[1] Burt, P.J. & Adelson, E.H. (1985). Merging images through pattern decomposition. In A.G. Tescher (Ed.), *Applications of Digital Image Processing VIII* (pp. 173-181). Bellingham, WA: The International Society for Optical Engineering.

[2] Essock, E.A., Sinai, M.J., McCarley, J.S., Krebs, W.K. & DeFord, J.K. (1999). Perceptual ability with real-world nighttime scenes: image-intensified, infrared, and fused-color imagery. *Human Factors, 41(3),* 438-452.

[3] Krebs, W.K., Scribner, D.A., Miller, G.M., Ogawa, J.S. & Schuler, J. (1998). Beyond third generation: a sensor-fusion targeting FLIR pod for the F/A-18. In B.V. Dasarathy (Ed.), *Sensor Fusion: Architectures, Algorithms, and Applications II* (pp. 129-140). Bellingham, WA, USA: International Society for Optical Engineering.

[4] Macmillan, N.A. & Creelman, C.D. (1991). *Detection theory: a user's guide.* Cambridge, MA: Cambridge University Press.

[5] Onyango, C.M. & Marchant, J.A. (2001). Physics-based colour image segmentation for scenes containing vegetation and soil. *Image and Vision Computing, 19(8),* 523-538.

[6] Ryan, D. & Tinkler, R. (1995). Night pilotage assessment of image fusion. In R.J. Lewandowski, W. Stephens & L.A. Haworth (Ed.), *SPIE Proceedings on Helmet and Head Mounted Displays and Symbology Design Requirements II* (pp. 50-67). Bellingham, WA: The International Society for Optical Engineering.

[7] Schuler, J., Howard, J.G., Warren, P., Scribner, D.A., Klien, R., Satyshur, M. & Kruer, M.R. (2000). Multiband E/O color fusion with consideration of noise and registration. In W.R. Watkins, D. Clement & W.R. Reynolds (Ed.), *Targets and Backgrounds VI: Characterization, Visualization, and the Detection Process* (pp. 32-40). Bellingham, WA, USA: The International Society for Optical Engineering.

[8] Steele, P.M. & Perconti, P. (1997). Part task investigation of multispectral image fusion using gray scale and synthetic color night vision sensor imagery for helicopter pilotage. In W. Watkins & D. Clement (Ed.), *Proceedings of the SPIE Conference on Targets and Backgrounds, Characterization and Representation III* (pp. 88-100). Bellingham, WA: The International Society for Optical Engineering.

[9] Toet, A. (1990). Adaptive multi-scale contrast enhancement through non-linear pyramid recombination. *Pattern Recognition Letters, 11,* 735-742.

[10] Toet, A. (1990). Hierarchical image fusion. *Machine Vision and Applications, 3,* 1-11.

[11] Toet, A. (1992). Multi-scale contrast enhancement with applications to image fusion. *Optical Engineering, 31(5),* 1026-1031.

[12] Toet, A., IJspeert, J.K., Waxman, A.M. & Aguilar, M. (1998). Fusion of visible and thermal imagery improves situational awareness. *Displays, 18,* 85-95.

[13] Toet, A., Ruyven, J.J. & Valeton, J.M. (1989). Merging thermal and visual images by a contrast pyramid. *Optical Engineering, 28,* 789-792.

[14] Toet, A. & Walraven, J. (1996). New false colour mapping for image fusion. *Optical Engineering, 35(3),* 650-658.

[15] Waxman, A.M., Gove, A.N., Fay, D.A., Racamato, J.P., Carrick, J.E., Seibert, M.C. & Savoye, E.D. (1996). Color night vision: opponent processing in the fusion of visible and IR imagery. *Neural Networks, 9(6).*

Appendix E

Paint the Night:

Applying Daytime Colours to Nighttime Imagery

Alexander Toet

TNO Human Factors
Kampweg 5
3769 DE Soesterberg
The Netherlands
Phone +31-3463-56237
Fax +31-3463-53977
Email toet@tm.tno.nl

EXTENDED ABSTRACT

Reinhard *e.a.* (Reinhard et al., 2001) recently introduced a method to transfer one image's colour characteristics (the target image) to another (the source image). Here we show that this method can also be used to transfer the natural appearance of daylight colour imagery to fused multispectral nightvision images. The method employs a transformation to a principal component space that has been derived from a large ensemble of hyperspectral images of natural scenes (Ruderman et al., 1998). In this decorrelated colour space the first order statistics of natural colour images (target scenes) are transferred to the multispectral nightvision images (source scenes). The only requirement of the method is that the composition of the source and target scenes are similar to some extent. Hence, the depicted scenes need not be identical, they merely have to resemble each other.

Modern nightvision systems are increasingly deployed in military operations and for surveillance applications. The two most common nighttime imaging systems either display emitted infrared (IR) radiation or reflected light, and thus provide complementary information of the inspected scene. A combined or fused display of these camera signals can therefore provide a more complete representation of the scene. A false colour mapping can be used to enable human observers to interpret the displayed information more easily. The simple colour mapping presented here gives fused nighttime imagery an appearance that resembles normal colour daytime images. It is therefore likely that this mapping will help to make the visual interpretation of the nightvision imagery more intuitive.

Since there evidently exists no one-to-one mapping between temperature contrast and spectral reflectance of a material, the goal of producing a nighttime image, incorporating information from IR imagery, with an appearance identical to a colour daytime image can never be fully achieved. The options are therefore *(1)* to settle for a single mapping that works satisfactory in a large number of conditions, or *(2)* to adapt (optimize) the colour mapping to the situation at hand. Reinhard's colour transfer method (Reinhard et al., 2001) provides both of these options. A single mapping can be applied to a large ensemble of fused multispectral images by adopting a generic target (colour daylight) image that has characteristics similar to those of the members of the ensemble. The mapping can be optimised for each individual nightvision scene by selecting a target image with the same structural content. For surveillance systems, that usually register a fixed scene, the optimisation is easily performed using a daylight colour image of the same scene that is being monitored.

We apply the method to RGB false colour nighttime imagery recorded both with a dual band (visual and near infrared) image intensified low-light CCD camera (DII) and with a thermal middle wavelength band (3-5 μm) infrared (IR) camera. The results show that the method can be used effectively to give nighttime imagery a daytime appearance. In addition, we show how the luminance contrast of the resulting images can be enhanced by replacing luminance component of the resulting colour image with a grayscale fused representation of the three input bands.

Summary

We show that a recently introduced method to transfer one image's colour characteristics to another (Reinhard et al., 2001) can be used effectively to give multispectral nighttime imagery a natural daytime colour appearance. The contrast of the resulting colour imagery can be improved by mapping a grayscale fused representation of the individual image bands to the luminance component of the resulting colour images. Only 6 numbers (the three components of respectively the mean and standard deviation of the image components in ***LMS*** cone space) are required to apply a natural daytime colour appearance to multispectral nighttime imagery. The resulting full colour representation of nighttime scenes may be of great ergonomic value by making the interpretation (segmentation) of the displayed scene easier (more intuitive) for the observer.

Keywords: Image fusion, infrared, intensified imagery, scene recognition, situational awareness, false colour, pyramid

1 INTRODUCTION

Modern night-time camera's are designed to expand the conditions under which human obervers can operate. A functional piece of equipment must therefore provide an image that leads to good perceptual awareness in most environmental and operational conditions (to "Own the weather " or "Own the night"). The two most common nighttime imaging systems either display emitted infrared (IR) radiation or reflected light, and thus provide complementary information of the inspected scene. A suitably combined or *fused* representation of IR and (intensified) visual imagery may enable an observer to construct a more complete mental representation of the perceived scene, resulting in a larger degree of situational awareness (Toet et al., 1998). A false colour representation of fused nighttime imagery that closely resembles a natural daylight colour scheme will help the observer by making scene interpretation more intuitive.

The rapid development of multi-band infrared and visual nightvision systems has led to an increased interest in colour fused ergonomic representations of multiple sensor signals (Aguilar et al., 1998; Aguilar et al., 1999; Aguilar & Garret, 2001; Essock et al., 1999; Fay et al., 2000; Schuler et al., 2000; Scribner et al., 1999; Varga, 1999; Waxman et al., 1995; Waxman et al., 1996; Waxman et al., 1997; Waxman et al., 1998; Waxman et al., 1999). Simply mapping multiple spectral bands of imagery into a three dimensional colour space already generates an immediate benefit, since the human eye can discern several thousand colours, while it can only distinguish about 100 shades of gray at any instance. Combining bands in colour space therefore provides a method to increase the dynamic range of a sensor system (Driggers et al., 2001). Experiments have convincingly demonstrated that appropriately designed false colour rendering of nighttime imagery can significantly improve observer performance and reaction times in tasks that involve scene segmentation and classification (Essock et al., 1999; Sinai et al., 1999; Toet et al., 1997; Toet & IJspeert, 2001; Varga, 1999; White, 1998). However, inappropriate colour mappings may hinder situational awareness (Krebs et al., 1998; Toet & IJspeert, 2001; Varga, 1999). One of the main reasons seems to be the counter intuitive appearance of scenes rendered in artificial colour schemes and the lack of colour constancy (Varga, 1999). Hence, an ergonomic colour scheme should produce night vision imagery with a natural appearance and with colours that are invariant for changes in the environmental conditions.

Reinhard *e.a.* (Reinhard et al., 2001) recently introduced a method to transfer one image's colour characteristics to another. Here we show that this method can be applied to transfer the natural colour characteristics of daylight colour imagery to fused multispectral nightvision images. The method employs a transformation to a principal component space that has recently been derived from a large ensemble of hyperspectral images of natural scenes (Ruderman et al., 1998). In this decorrelated colour space the first order statistics of natural colour images (target scenes) are transferred to the multispectral nightvision images (source scenes). The only requirement of the method is that the composition of the source and target scenes are similar to some extent. Hence, the depicted scenes need not be identical, they merely have to resemble each other. For surveillance systems, that usually register a fixed scene, a daylight colour image of the same scene that is being monitored can be used to derive an optimal colour mapping.

Here we apply the method of Reinhard *e.a.* (Reinhard et al., 2001) to transfer the characteristics of natural daylight colour images to false colour fused nighttime imagery. We demonstrate the effectiveness of the method for the combined (fused) display of visual (400-700 nm) and near infrared (700-900 nm) intensified low-light CCD images and thermal middle wavelength band (3-5 μm) infrared images. The results show that the method can be used effectively to give nighttime imagery a daytime appearance. Reinhard's (Reinhard et al., 2001) colour transfer method is in fact a simplification of a more general method that employs a principal component analysis, and applies when the images have a natural origin. The method can also be applied to images representing man made objects by using a full principal component analysis.

2 IMAGERY

A variety of outdoor scenes, displaying several kinds of vegetation (grass, heather, semi shrubs, trees), sky, water, sand, vehicles, roads, and persons, were registered at night with a recently developed dual-band visual intensified (DII) camera (see below), and with a state-of-the-art thermal middle wavelength band (3-5 μm) infrared (IR) camera (Radiance HS). Both cameras had a field of view (FOV) of about 6x6 degrees.

The DII camera was developed by Thales Optronics (Delft, The Netherlands) and facilitated a two-colour registration of the scene, applying two overlapping bands covering the part of the electromagnetic spectrum ranging from visual to near infrared (400-900 nm). The short (visual) wavelength part of the incoming spectrum was mapped to the R channel of an RGB false colour composite image. The long (near infrared) wavelength band corresponds primarily to the spectral reflection characteristics of vegetation, and was therefore mapped to the G channel of the RGB false colour composite image. This approach utilises the fact that the spectral reflection characteristics of plants are distinctly different from other (natural and artificial) materials in the visual and near infrared range (Onyango & Marchant, 2001). The spectral response of the long-wavelength (G) channel roughly matches that of a Generation III image intensifier system.

Images were recorded at various times of the diurnal cycle under various atmospheric conditions (clear, rain, fog, …) and for various illumination levels (1 lux – 0.1 mlux). Object ranges up to several hundreds of meters were applied. The images were digitized on-site (using a Matrox Genesis frame grabber, using at least 1.8 times oversampling).

The recorded DII and IR images were first registered through an affine warping procedure, using fiducial registration points that were recorded at the beginning of each session . After warping, corresponding pixels in images taken with the different cameras represent the same location in the recorded scene. Then, a fused false colour RGB image was produced by assigning the IR image to the (empty) B channel of the false colour DII image. Finally, patches displaying different types of scenic elements were selected and cut out from the resulting false colour fused images. These patches were deployed as test images in the rest of this study. They display either buildings, vehicles, water, roads, trees, heather or humans. These details were selected because their signature varies strongly among the different image modalities. This false colour fusion scheme results in images in which grass, trees and persons are displayed as greenish, and roads, buildings, and vehicles are brownish. Some image examples of the individual image modalities and the fused RGB false colour representation are shown in Figures 1-7.

3 COLOUR TRANSFER

The false colour images resulting from the aforementioned fusion scheme have an unnatural colour appearance. The aim of the present study is to give these images the appearance of normal daylight colour images. In this section we introduce a simple technique to transfer the colour characteristics from natural daylight imagery to false colour nightvision imagery. A similar method was recently introduced to enhance the colour representation of synthetic imagery (Reinhard et al., 2001).

The method is as follows. Let the input multiband nightvision image be the source image, and let a normal daylight colour photograph be the target image. First, the source and target image are both transformed to the ***LMS*** cone response space. The different bands of multisensor signals and daytime colour images are usually correlated. Since we want to be able to transfer the characteristics of daytime colour images to false colour fused nighttime images we first need to transform the input (multiband fused and daytime) colour imagery to a space which minimizes the correlation between channels. Therefore, through principal component analysis, we rotate the axes in the ***LMS*** cone space to achieve maximal decorrelation between the data points. Then, the mean and standard deviation of the source image is set equal to those of the target image. Finally, the source image is transformed back to RGB space for display. The result is a colour representation of the multiband nightvision image that resembles a normal daylight image.

In the following sections we first discuss the RGB to ***LMS*** transform. Then, we present a colour transfer method that employs a principal component transform in ***LMS*** cone space. Finally, we will show that for natural scenes the principal component transform can effectively be replaced by a fixed transform to $l\alpha\beta$ space (Ruderman et al., 1998). This space has recently been derived from a principal component transform of a large ensemble of hyperspectral images that represents a good cross-section of natural scenes. The resulting data representation is compact and symmetrical, and provides automatic decorrelation to higher than second order.

3.1 RGB to ***LMS*** transform

First the RGB tristimulus values are converted to device independent XYZ tristimulus values. This conversion depends on the characteristics of the display on which the image was originally intended to be displayed. Because that information is rarely available, it is common practice to use a device-independent conversion that maps white in the chromaticity diagram to white in RGB space and vice versa (e.g. Fairchild, 1998).

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.5141 & 0.3239 & 0.1604 \\ 0.2651 & 0.6702 & 0.0641 \\ 0.0241 & 0.1228 & 0.8444 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

The device independent XYZ values are then converted to LMS space by

$$\begin{bmatrix} L \\ M \\ S \end{bmatrix} = \begin{bmatrix} 0.3897 & 0.6890 & -0.0787 \\ -0.2298 & 1.1834 & 0.0464 \\ 0.0000 & 0.0000 & 1.0000 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (2)$$

Combination of (1) and (2) results in

$$\begin{bmatrix} L \\ M \\ S \end{bmatrix} = \begin{bmatrix} 0.3811 & 0.5783 & 0.0402 \\ 0.1967 & 0.7244 & 0.0782 \\ 0.0241 & 0.1288 & 0.8444 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{3}$$

The data in this colour space shows a great deal of skew, which is largely eliminated by taking a logarithmic transform:

$$\begin{aligned} \boldsymbol{L} &= \log L \\ \boldsymbol{M} &= \log M \\ \boldsymbol{S} &= \log S \end{aligned} \tag{4}$$

The inverse transform from ***LMS*** cone space back to RGB space is as follows. First, the ***LMS*** pixel values are raised to the power ten to go back to linear LMS space. Then, the data can be converted from LMS to RGB using the inverse transform of Equation (3):

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 4.4679 & -3.5873 & 0.1193 \\ -1.2186 & 2.3809 & -0.1624 \\ 0.0497 & -0.2439 & 1.2045 \end{bmatrix} \begin{bmatrix} L \\ M \\ S \end{bmatrix} \tag{5}$$

3.2 Transfer method I: principal component transform

The principal component transform (e.g. Hall, 1979; Richards, 1986; Taylor, 1999) effectively rotates the ***LMS*** coordinate axes such that the pixel components are maximally decorrelated. The set of normalized eigenvectors of the covariance matrix of the set of pixel values, arranged in order of increasing eigenvalues, constitute the column vectors of the corresponding rotation matrix. Let R_t be the rotation matrix that decorrelates the target pixels. The pixel values of the source and target images in this new coordinate system are then respectively given by

$$\begin{bmatrix} \boldsymbol{L}'_s \\ \boldsymbol{M}'_s \\ \boldsymbol{S}'_s \end{bmatrix} = R_t \begin{bmatrix} \boldsymbol{L}_s \\ \boldsymbol{M}_s \\ \boldsymbol{S}_s \end{bmatrix} \tag{6}$$

and

$$\begin{bmatrix} \boldsymbol{L}'_t \\ \boldsymbol{M}'_t \\ \boldsymbol{S}'_t \end{bmatrix} = R_t \begin{bmatrix} \boldsymbol{L}_t \\ \boldsymbol{M}_t \\ \boldsymbol{S}_t \end{bmatrix} \tag{7}$$

where the indices s and t refer to the source and target images respectively.

First, the mean is subtracted from the data points:

$$\begin{aligned} \boldsymbol{L}^* &= \boldsymbol{L}' - \langle \boldsymbol{L}' \rangle \\ \boldsymbol{M}^* &= \boldsymbol{M}' - \langle \boldsymbol{M}' \rangle \\ \boldsymbol{S}^* &= \boldsymbol{S}' - \langle \boldsymbol{S}' \rangle \end{aligned} \tag{8}$$

Then, the source data points are scaled with the ratio of the standard deviations of the source and target images respectively:

$$
\begin{aligned}
\boldsymbol{L}_s^{+} &= \frac{\sigma_t^{\boldsymbol{L}}}{\sigma_s^{\boldsymbol{L}}} \boldsymbol{L}_s^{*} \\
\boldsymbol{M}_s^{+} &= \frac{\sigma_t^{\boldsymbol{M}}}{\sigma_s^{\boldsymbol{M}}} \boldsymbol{M}_s^{*} \\
\boldsymbol{S}_s^{+} &= \frac{\sigma_t^{\boldsymbol{S}}}{\sigma_s^{\boldsymbol{S}}} \boldsymbol{S}_s^{*}
\end{aligned}
\tag{9}
$$

After this transformation, the resulting data points have standard deviations that correspond to those of the target image. Before reconstructing the RGB representation the averages computed over the target image are added to the source image:

$$
\begin{aligned}
\boldsymbol{L}_s^{\oplus} &= \boldsymbol{L}_s^{+} + \left\langle \boldsymbol{L}_t^{'} \right\rangle \\
\boldsymbol{M}_s^{\oplus} &= \boldsymbol{M}_s^{+} + \left\langle \boldsymbol{M}_t^{'} \right\rangle \\
\boldsymbol{S}_s^{\oplus} &= \boldsymbol{S}_s^{+} + \left\langle \boldsymbol{S}_t^{'} \right\rangle
\end{aligned}
\tag{10}
$$

The result is transformed back to RGB space via the inverse rotation R_t^{-1} , logLMS, LMS, and XYZ colour space using Equation (5).

3.3 Transfer method II: *lαβ* transform

Ruderman *e.a.* (Ruderman et al., 1998) recently derived a colour space, called *lαβ,* which effectively minimises the correlation between the ***LMS*** axes. This result was derived from a principal component transform to the logarithmic ***LMS*** cone space representation of a large ensemble of hyperspectral images that represented a good cross-section of natural scenes. The principal axes encode fluctuations along an achromatic direction (*l*), a yellow-blue opponent direction (*αβ*), and a red-green opponent direction (*αβ*). The resulting data representation is compact and symmetrical, and provides automatic decorrelation to higher than second order.

Ruderman *e.a.* (Ruderman et al., 1998) presented the following simple transform to decorrelate the axes in the ***LMS*** space:

$$
\begin{bmatrix} l \\ \alpha \\ \beta \end{bmatrix} =
\begin{bmatrix} \frac{1}{\sqrt{3}} & 0 & 0 \\ 0 & \frac{1}{\sqrt{6}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} \end{bmatrix}
\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -2 \\ 1 & -1 & 0 \end{bmatrix}
\begin{bmatrix} \boldsymbol{L} \\ \boldsymbol{M} \\ \boldsymbol{S} \end{bmatrix}
\tag{11}
$$

If we think of the **L** channel as red, the **M** as green, and **S** as blue, we see that this is a variant of a colour opponent model:

$$
\begin{aligned}
\text{Achromatic} &\propto r+g+b \\
\text{Yellow-blue} &\propto r+g-b \\
\text{Red-green} &\propto r-g
\end{aligned}
\tag{12}
$$

Thus the *l* axis represents an achromatic channel, while the α and β channels are chromatic yellow-blue and red-green opponent channels.

After processing the colour signals in the $l\alpha\beta$ space the inverse transform of Equation (11) can be used to return to the ***LMS*** space:

$$\begin{bmatrix} \boldsymbol{L} \\ \boldsymbol{M} \\ \boldsymbol{S} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -2 & 0 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{3}}{3} & 0 & 0 \\ 0 & \frac{\sqrt{6}}{6} & 0 \\ 0 & 0 & \frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} l \\ \alpha \\ \beta \end{bmatrix} \tag{13}$$

The processing in the $l\alpha\beta$ space is similar to the processing applied in the previous section, and given by Equations (8)--(10). First, mean is subtracted from the source and target data points:

$$\begin{aligned} l^* &= l - \langle l \rangle \\ \alpha^* &= \alpha - \langle \alpha \rangle \\ \beta^* &= \beta - \langle \beta \rangle \end{aligned} \tag{14}$$

Then, the source data points are scaled with the ratio of the standard deviations of the source and target images respectively:

$$\begin{aligned} l_s' &= \frac{\sigma_t^l}{\sigma_s^l} l_s^* \\ \alpha_s' &= \frac{\sigma_t^\alpha}{\sigma_s^\alpha} \alpha_s^* \\ \beta_s' &= \frac{\sigma_t^\beta}{\sigma_s^\beta} \beta_s^* \end{aligned} \tag{15}$$

After this transformation the pixels comprising the multiband source image have standard deviations that conform to the target daylight colour image. Finally, in reconstructing the $l\alpha\beta$ transform of the multiband source image, instead of adding the previously subtracted averages, the averages computed for the target daylight colour image are added. The result is transformed back to RGB space via logLMS, LMS, and XYZ colour space using Equations (13) and (5).

4 OPTIMISING LUMINANCE CONTRAST

When combining the different bands of a multiband nightvision system into a single colour display it is essential that the relevant contrast details of the individual bands are preserved in the final colour image, and that no spurious pattern elements (that could interfere with subsequent analysis) are introduced by the merging process. The multiband images used in this study were obtained by mapping visual (400-700 nm) and near infrared (700-900 nm) intensified low-light CCD images and thermal middle wavelength band (3-5 μm) infrared images to respectively the Red, Green and Blue channels of an RGB false colour image. Note that the contrast of an image detail may vary strongly among the different bands. In some conditions a detail may even be represented with opposite contrast in different spectral bands. The combination of the individual image bands into a single colour image may therefore significantly reduce the luminance contrast of an image detail. As a result, a detail that is clearly visible in the individual image bands may be much less visible in the final colour representation, due to a lack of luminance contrast.

We applied the following procedure to preserve the luminance contrast (and hence the visibility) of perceptually relevent image details in the final colour representation of the multiband nighttime images.

First we fused the 3 abovementioned individual bands into a single grayscale representation though a pyramidal image fusion scheme (Burt & Adelson, 1985; Toet, 1989; Toet et al., 1989; Toet, 1990). A 7-level Laplacian pyramid (Burt & Adelson, 1985) was used, in combination with a maximum absolute contrast node (i.e. pattern element) selection rule. This procedure ensures that perceptually relevant image details from all individual bands are represented in the final grayscale fused image.

Then the RGB colour images resulting from the application of Method I to the multiband nighttime images was transformed to a HSV (hue, saturation, value) representation (Pratt, 1991). This transformation effectively decouples the color information (H and S) from the luminance information (V). Hue represents the dominant color as seen by an observer, saturation refers to the amount of dilution of the color with white light, and value defines the average brightness. The luminance component may therefore be processed independently of the image's colour information. We used this property to replace the luminance component of the HSV transformed multiband nighttime image by the grayscale fused image. Finally, we transformed the resulting images back to an RGB representation. The result is a false colour representation of the 3 band nightvision imagery, with optimal luminance contrast (resulting from the application of the pyramidal grayscale fusion method), and with colour characteristics similar to a daylight photograph of the depicted scene (resulting from the application of Method I).

5 EXAMPLES

Figures 8-14 show some results of the application of both colour transfer Methods I and II to the set of false colour RGB multispectral nightvision images from Figures 1-7. In each of these examples, the target image (8-14 b) is a daylight colour photograph of the same scene that was also recorded at night in full darkness with both the dual-band image intensifier and the thermal middle wavelength infrared camera, and that is represented as a false colour RGB image in (8-14 a). In some of the target photographs, the colour is washed out somewhat because of atmospheric attenuation, since the viewing distance was about 800 m. Note that the colour of the target photograph (8-14b) returns in the processed false colour nighttime imagery (8-14 c and d). Hence, the colour transfer method is effective in giving nighttime imagery a daytime appearance. The results produced with Methods I (8-14 c) and II (8-14 d) are quite similar. That implies that the costly principal component analysis of Method I can be replaced by the simple matrix transform of Method II when viewing natural imagery.

Figures 15-20 demonstrate the fact that the actual choice of the target image is not critical, as long as the structural content (statistics) and the colour distribution are similar to those of the source scene. In each of these Figures the top image represents the false colour RGB image obtained by mapping respectively the visual (400-700 nm) and the near infrared (700-900 nm) intensified low-light CCD images and the thermal middle wavelength band (3-5 μm) infrared images to the Red, Green and Blue channels of an RGB colour image. The left column represents the different target images used in the colour transfer Method II. The right column represents the result of applying the colour transfer Method II to the image on top with the corresponding image in the left column as the target. The results show that the appearance of the final result resembles that of the target image in each case.

Figure 15 also shows an example (5^{th} row) in which an oil painting called "Old Oak Tree", by the Dutch painter Barend Cornelius Koekkoek (1803-1862), was adopted as the target image. The corresponding result has an appearance which is quite natural. However, the example shown in the 4^{th} row of Figure 15 shows that the method fails when the target and source images are too dissimilar in composition. In this case, the target image displays green grass and green trees with a bright blue sky in the background. The source image in contrast only shows brownish trees and vegetation. As a result, the transfer of statistics fails in this case.

Figures 21-27 illustrate the effect of mapping the grayscale fused nighttime imagery to the luminance component of the result of the colour transfer Method I. The results show that luminance contrast indeed increases in most cases, resulting in images in which the depicted details are easier to perceive (easier to segment from their background).

Figures 28-30 demonstrate the general applicability of the colour transfer method by applying it to arbitrary combinations of normal and intensified grayscale imagery, and 3-5 and 8-12 μm thermal imagery. In all cases, the results show a colour image with natural characteristics.

6 DISCUSSION

We showed that a recently introduced method to transfer one image's colour characteristics to another (Reinhard et al., 2001) can be used effectively to give multispectral nighttime imagery a natural daytime colour appearance. The contrast of the resulting colour imagery can be improved by mapping a grayscale fused representation of the individual image bands to the luminance component of the resulting colour images.

The colour transfer method employs a transformation to a principal component space. In this decorrelated colour space the first order statistics of natural colour images (target scenes) are transferred to the multispectral nightvision images (source scenes). We applied the method to a set of RGB false colour nighttime images recorded both with a dual band (visual and near infrared) image intensified low-light CCD camera (DII) and with a thermal middle wavelength band (3-5 μm) infrared (IR) camera. In each case, the resulting false colour nighttime images adopted the appearance of the daytime colour images of the corresponding scene.

The colour transfer method only uses the first order statistics of natural colour images that are representative of the depicted scene. This implies that only 6 numbers (the three components of respectively the mean and standard deviation of the image components in ***LMS*** cone space) are required to apply a natural daytime colour appearance to multispectral nighttime imagery. Hence, there is no need to actually store the target images from which the colour information (the first order statistics) is derived. A system that is equipped with a look-up table of characteristic numbers for different types of backgrounds is sufficient to enable the observer to adjust the colour mapping to the scene being viewed.

Nighttime images recorded with an intensified low-light CCD camera and a thermal middle wavelength band (3-5 μm) infrared camera contain *complementary* information. This makes each of the individual image modalities only suited for specific observation tasks. In many operational conditions different nightvision systems are used side by side. By using a combined or fused display method the complementarity of the information in the image image modalities can be fully exploited, thus enabling multiple observation tasks to be performed with a single nighttime image representation. A full colour representation of nighttime scenes may be of great ergonomic value by making the interpretation (segmentation) of the displayed scene easier (more intuitive) for the observer.

Since there evidently exists no one-to-one mapping between the temperature contrast and the spectral reflectance of a material, the goal of producing a nighttime image, incorporating information from IR imagery, with an appearance identical to a colour daytime image can never be fully achieved. The method employed here allows one *(1)* to settle for a single mapping that works satisfactory in a large number of conditions (.e.g by selecting the colour statistics of a generic representative scene), or *(2)* to adapt (optimise) the colour mapping to the situation at hand (e.g. by selecting the colour statistics that perfectly match the scene at hand).

Acknowledgements

The authors thank Thales Optronics for providing the DII camera, and Hans Winkel (TNO-FEL), Jan Kees IJspeert and Nicole Schoumans (TNO-HF) for their help with the image registration.

This material is partly based upon work supported by the European Office of Aerospace Research and Development, Air Force Office of Scientific Research, Air Force Research Laboratory, under contract No. F61775-01-WE026, and by Senter, Agency of the Ministry of Economic Affairs of the Netherlands.

REFERENCES

Aguilar, M., Fay, D.A., Ireland, D.B., Racamoto, J.P., Ross, W.D. & Waxman, A.M. (1999). Field evaluations of dual-band fusion for color night vision. In J.G. Verly (Ed.), *Enhanced and Synthetic Vision 1999* (pp. 168-175). Bellingham, WA: The International Society for Optical Engineering.

Aguilar, M., Fay, D.A., Ross, W.D., Waxman, A.M., Ireland, D.B. & Racamoto, J.P. (1998). Real-time fusion of low-light CCD and uncooled IR imagery for color night vision. In J.G. Verly (Ed.), *Enhanced and Synthetic Vision 1998* (pp. 124-135). Bellingham, WA: The International Society for Optical Engineering.

Aguilar, M. & Garret, A.L. (2001). Biologically based sensor fusion for medical imaging. In B.V. Dasarathy (Ed.), *Sensor Fusion: Architectures, Algorithms, and Applications V* (pp. 149-158). Bellingham, WA: The International Society for Optical Engineering.

Burt, P.J. & Adelson, E.H. (1985). Merging images through pattern decomposition. In A.G. Tescher (Ed.), *Applications of Digital Image Processing VIII* (pp. 173-181). Bellingham, WA: The International Society for Optical Engineering.

Driggers, R.G., Krapels, K.A., Vollmerhausen, R.H., Warren, P.R., Scribner, D.A., Howard, J.G., Tsou, B.H. & Krebs, W.K. (2001). Target detection threshold in noisy color imagery. In G.C. Holst (Ed.), *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XII* (pp. 162-169). Bellingham, WA: The International Society for Optical Engineering.

Essock, E.A., Sinai, M.J., McCarley, J.S., Krebs, W.K. & DeFord, J.K. (1999). Perceptual ability with real-world nighttime scenes: image-intensified, infrared, and fused-color imagery. *Human Factors, 41(3),* 438-452.

Fairchild, M.D. (1998). *Color appearance models.* Reading, MA: Addison Wesley Longman, Inc.

Fay, D.A., Waxman, A.M., Aguilar, M., Ireland, D.B., Racamato, J.P., Ross, W.D., Streilein, W. & Braun, M.I. (2000). Fusion of multi-sensor imagery for night vision: color visualization, target learning and search.*Proceedings of the 3rd International Conference on Information Fusion* (pp. TuD3-3-TuD3-10). Paris, France: ONERA.

Hall, E.L. (1979). *Computer Image Processing.* New York, USA: Academic Press.

Krebs, W.K., Scribner, D.A., Miller, G.M., Ogawa, J.S. & Schuler, J. (1998). Beyond third generation: a sensor-fusion targeting FLIR pod for the F/A-18. In B.V. Dasarathy (Ed.), *Sensor Fusion: Architectures, Algorithms, and Applications II* (pp. 129-140). Bellingham, WA, USA: International Society for Optical Engineering.

Onyango, C.M. & Marchant, J.A. (2001). Physics-based colour image segmentation for scenes containing vegetation and soil. *Image and Vision Computing, 19(8),* 523-538.

Pratt, W.K. (1991). *Digital image processing, 2nd edition.* New York, USA: Wiley.

Reinhard, E., Ashikhmin, M., Gooch, B. & Shirley, P. (2001). Color transfer between images. *IEEE Computer Graphics and Applications, 21(5),* 34-41.

Richards, J.A. (1986). *Remote sensing digital image analysis.* Berlin: Springer Verlag.

Ruderman, D.L., Cronin, T.W. & Chiao, C.-C. (1998). Statistics of cone responses to natural images: implications for visual coding. *Journal of the Optical Society of America A, 15(8),* 2036-2045.

Schuler, J., Howard, J.G., Warren, P., Scribner, D.A., Klien, R., Satyshur, M. & Kruer, M.R. (2000). Multiband E/O color fusion with consideration of noise and registration. In W.R. Watkins, D. Clement & W.R. Reynolds (Ed.), *Targets and Backgrounds VI: Characterization, Visualization, and the Detection Process* (pp. 32-40). Bellingham, WA, USA: The International Society for Optical Engineering.

Scribner, D.A., Warren, P. & Schuler, J. (1999). Extending color vision methods to bands beyond the visible.*Proceedings of the IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications* (pp. 33-40).Institute of Electrical and Electronics Engineers.

Sinai, M.J., McCarley, J.S., Krebs, W.K. & Essock, E.A. (1999). Psychophysical comparisons of single- and dual-band fused imagery. In J.G. Verly (Ed.), *Enhanced and Synthetic Vision 1999* (pp. 176-183). Bellingham, WA: The International Society for Optical Engineering.

Taylor,P. (1999). Statistical methods. In M.Berthold & D.J.Hand (Eds.), *Intelligent data analysis.* (pp. 67-127). Berlin, GE: Springer Verlag.

Toet, A. (1989). Image fusion by a ratio of low-pass pyramid. *Pattern Recognition Letters, 9,* 245-253.

Toet, A. (1990). Hierarchical image fusion. *Machine Vision and Applications, 3,* 1-11.

Toet, A. & IJspeert, J.K. (2001). Perceptual evaluation of different image fusion schemes. In I. Kadar (Ed.), *Signal Processing, Sensor Fusion, and Target Recognition X* (pp. 436-441). Bellingham, WA: The International Society for Optical Engineering.

Toet, A., IJspeert, J.K., Waxman, A.M. & Aguilar, M. (1997). Fusion of visible and thermal imagery improves situational awareness. In J.G. Verly (Ed.), *Enhanced and Synthetic Vision 1997* (pp. 177-188). Bellingham, WA, USA: International Society for Optical Engineering.

Toet, A., IJspeert, J.K., Waxman, A.M. & Aguilar, M. (1998). Fusion of visible and thermal imagery improves situational awareness. *Displays, 18,* 85-95.

Toet, A., Ruyven, J.J. & Valeton, J.M. (1989). Merging thermal and visual images by a contrast pyramid. *Optical Engineering, 28,* 789-792.

Varga, J.T. (1999). *Evaluation of operator performance using true color and artificial color in natural scene perception* (Report AD-A363036). Monterey, CA: Naval Postgraduate School.

Waxman, A.M., Aguilar, M., Baxter, R.A., Fay, D.A., Ireland, D.B., Racamoto, J.P. & Ross, W.D. (1998). Opponent-color fusion of multi-sensor imagery: visible, IR and SAR.*Proceedings of the 1998 Conference of the IRIS Specialty Group on Passive Sensors* (pp. 43-61).

Waxman, A.M., et al. (1999). Solid-state color night vision: fusion of low-light visible and thermal infrared imagery. *MIT Lincoln Laboratory Journal, 11,* 41-60.

Waxman, A.M., Carrick, J.E., Fay, D.A., Racamato, J.P., Augilar, M. & Savoye, E.D. (1996). Electronic imaging aids for night driving: low-light CCD, thermal IR, and color fused visible/IR.*Proceedings of the SPIE Conference on Transportation Sensors and Controls* Bellingham, WA: The International Society for Optical Engineering.

Waxman, A.M., Fay, D.A., Gove, A.N., Seibert, M.C., Racamato, J.P., Carrick, J.E. & Savoye, E.D. (1995). Color night vision: fusion of intensified visible and thermal IR imagery. In J.G. Verly (Ed.), *Synthetic Vision for Vehicle Guidance and Control* (pp. 58-68). Bellingham, WA: The International Society for Optical Engineering.

Waxman, A.M., Gove, A.N., Fay, D.A., Racamoto, J.P., Carrick, J.E., Seibert, M.C. & Savoye, E.D. (1997). Color night vision: opponent processing in the fusion of visible and IR imagery. *Neural Networks, 10(1),* 1-6.

White, B.L. (1998). *Evaluation of the impact of multispectral image fusion on human performance in global scene processing* (Report AD-A343639). Monterey, CA: Naval Postgraduate School.

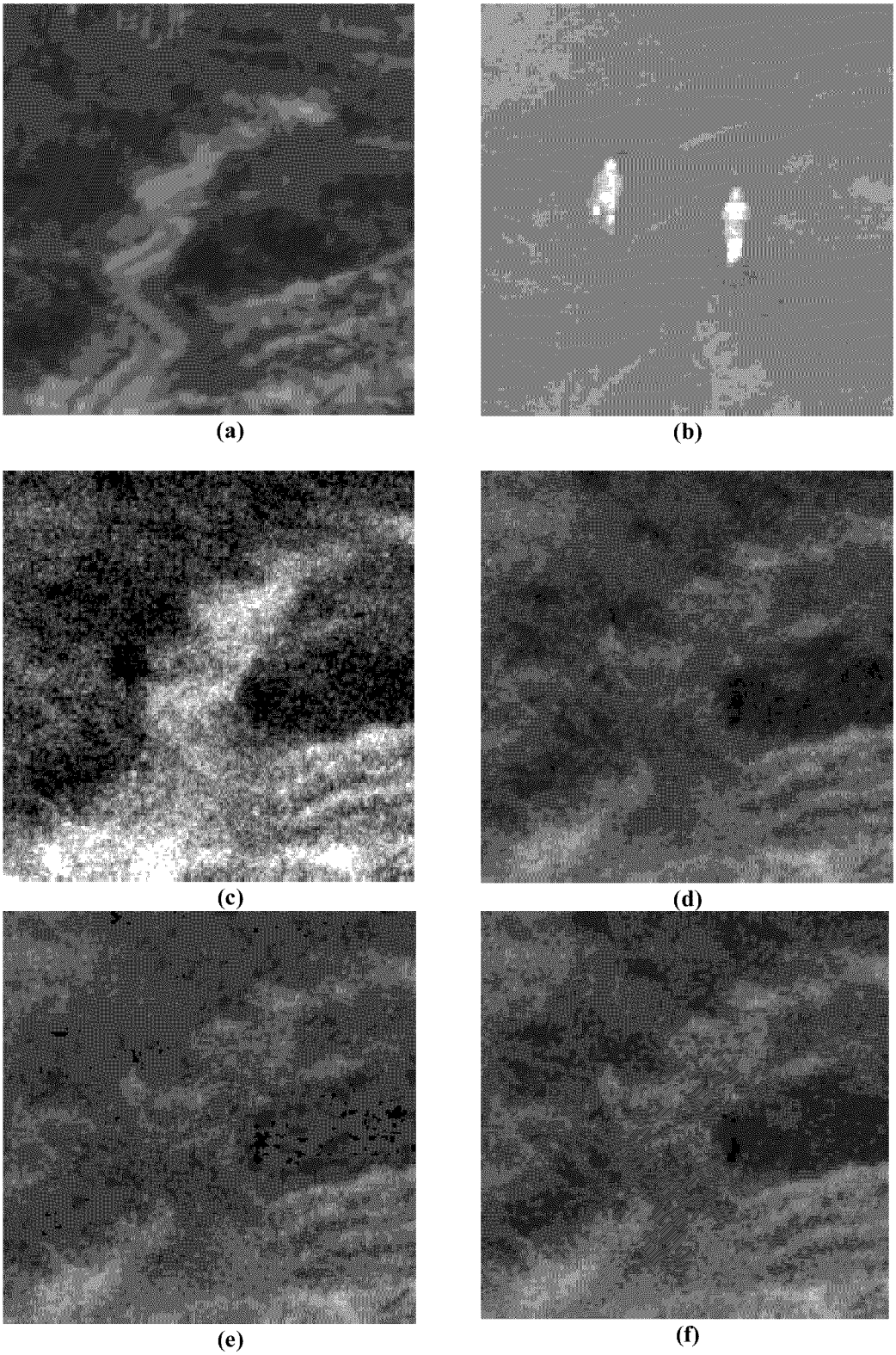


Figure 1 **(a)** Daylight colour photograph of a sandy path through the heath. **(b-d)** Images of the same scene in complete darkness resulting from respectively **(b)** a 3-5 μm infrared camera, **(c)** the visual (400-700 nm) and **(d)** the near infrared (700-900 nm) bands of a double-band image intensifier. **(e)** False colour image obtained by mapping the intensified images (c) and (d) to the R and G channels of an RGB colour image. **(f)** False colour image obtained by mapping the infrared image (b) to the B channel of image (e). Notice the presence of 2 persons in the nighttime images (b-f; but not in a).

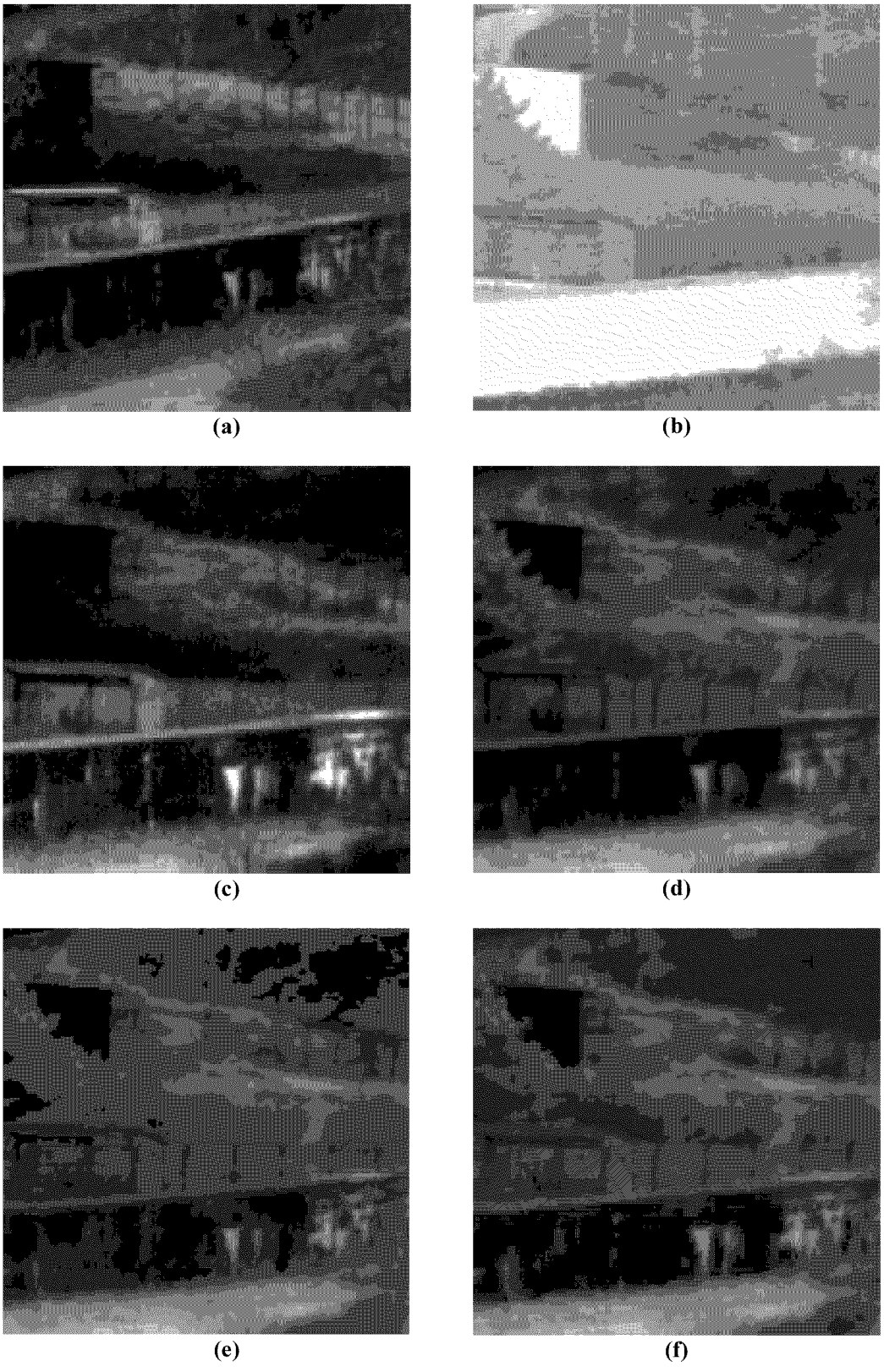


Figure 2 As **Figure 1** for a scene representing a stone wall, with trees, fences and a small building in the background.

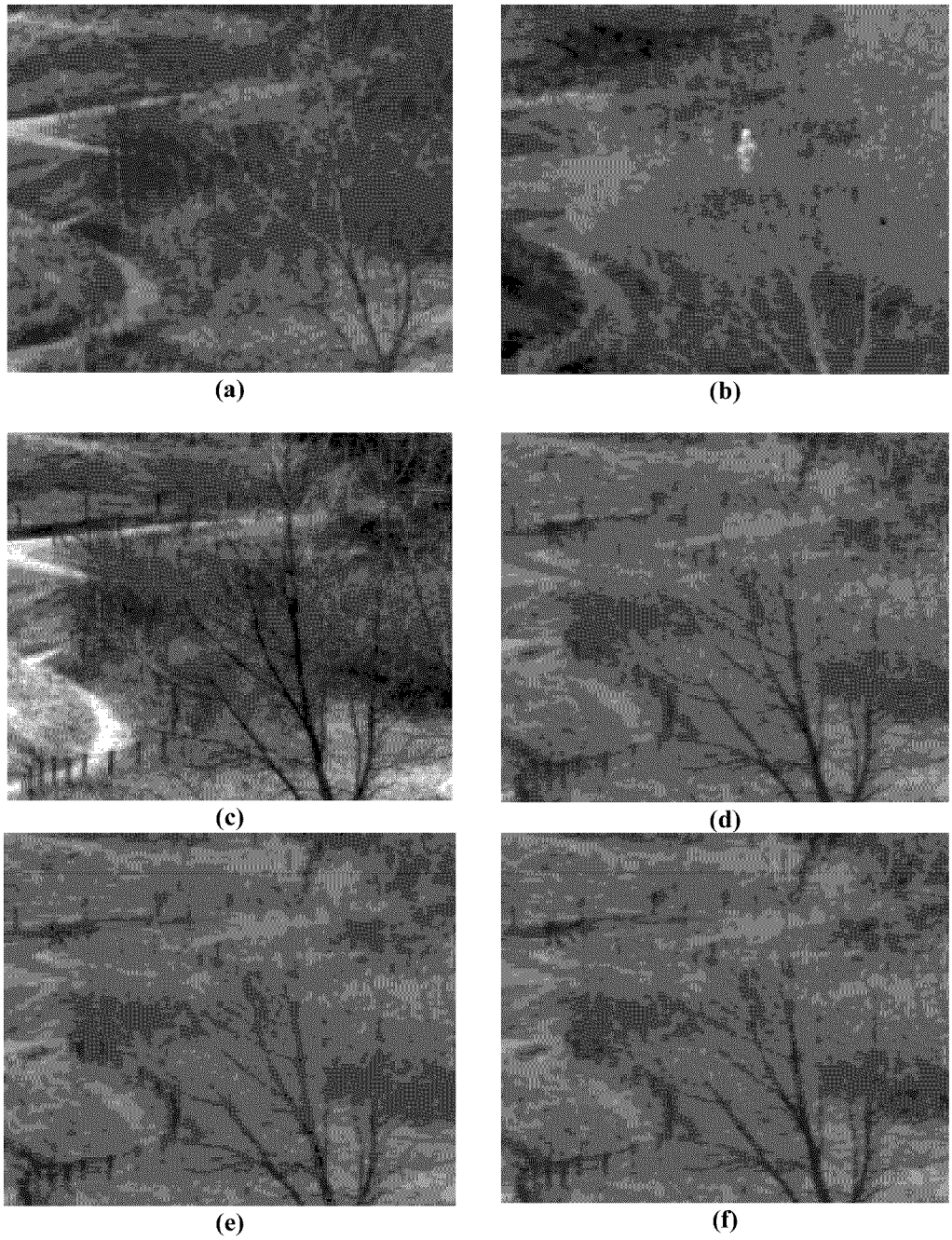


Figure 3 As Figure 1 for a scene representing a sandy path, trees, and fences. Notice the presence of a person in the nighttime images.

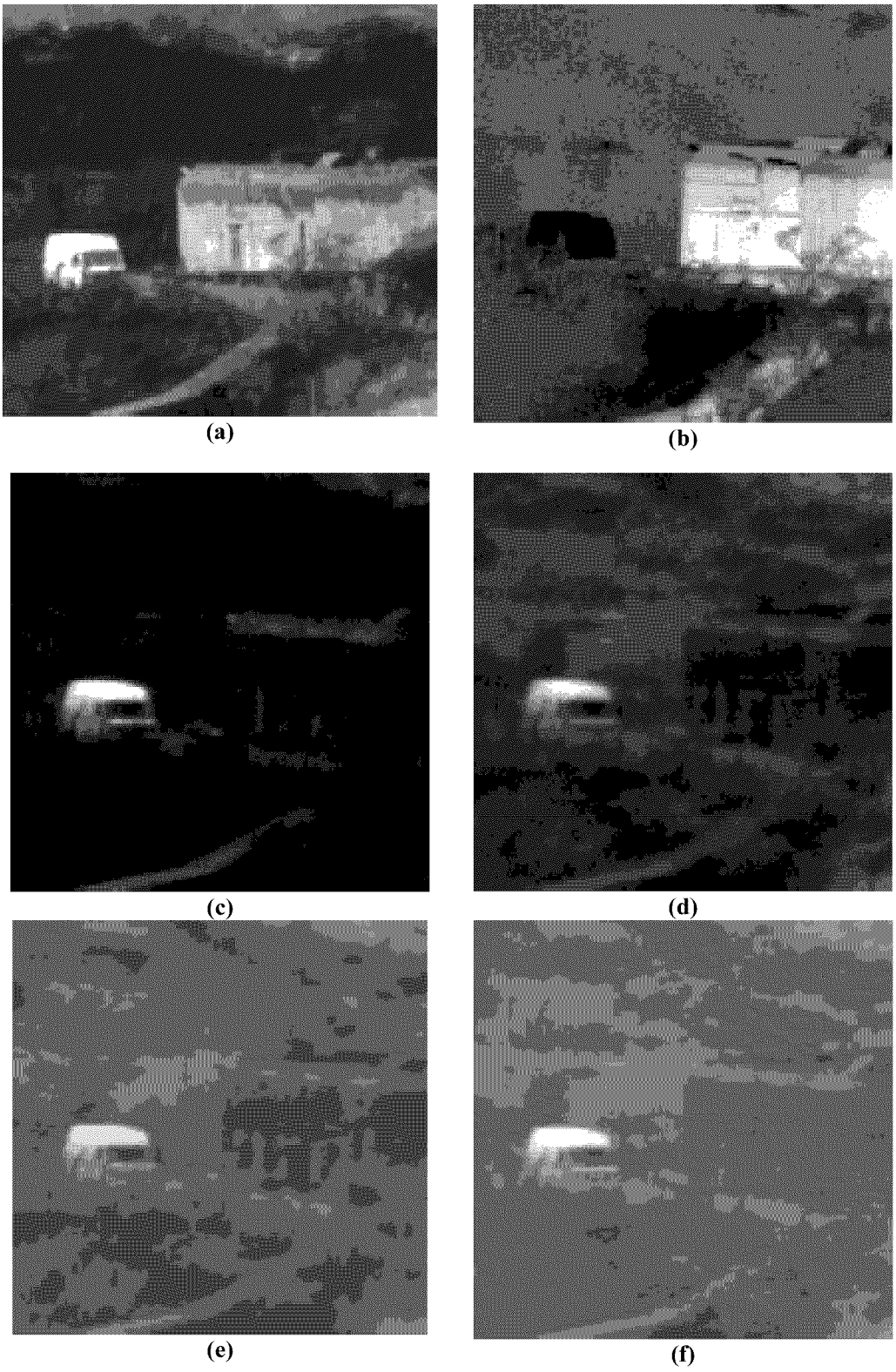


Figure 4 As Figure 1 for a scene representing a stone building, a wall, and a white van.

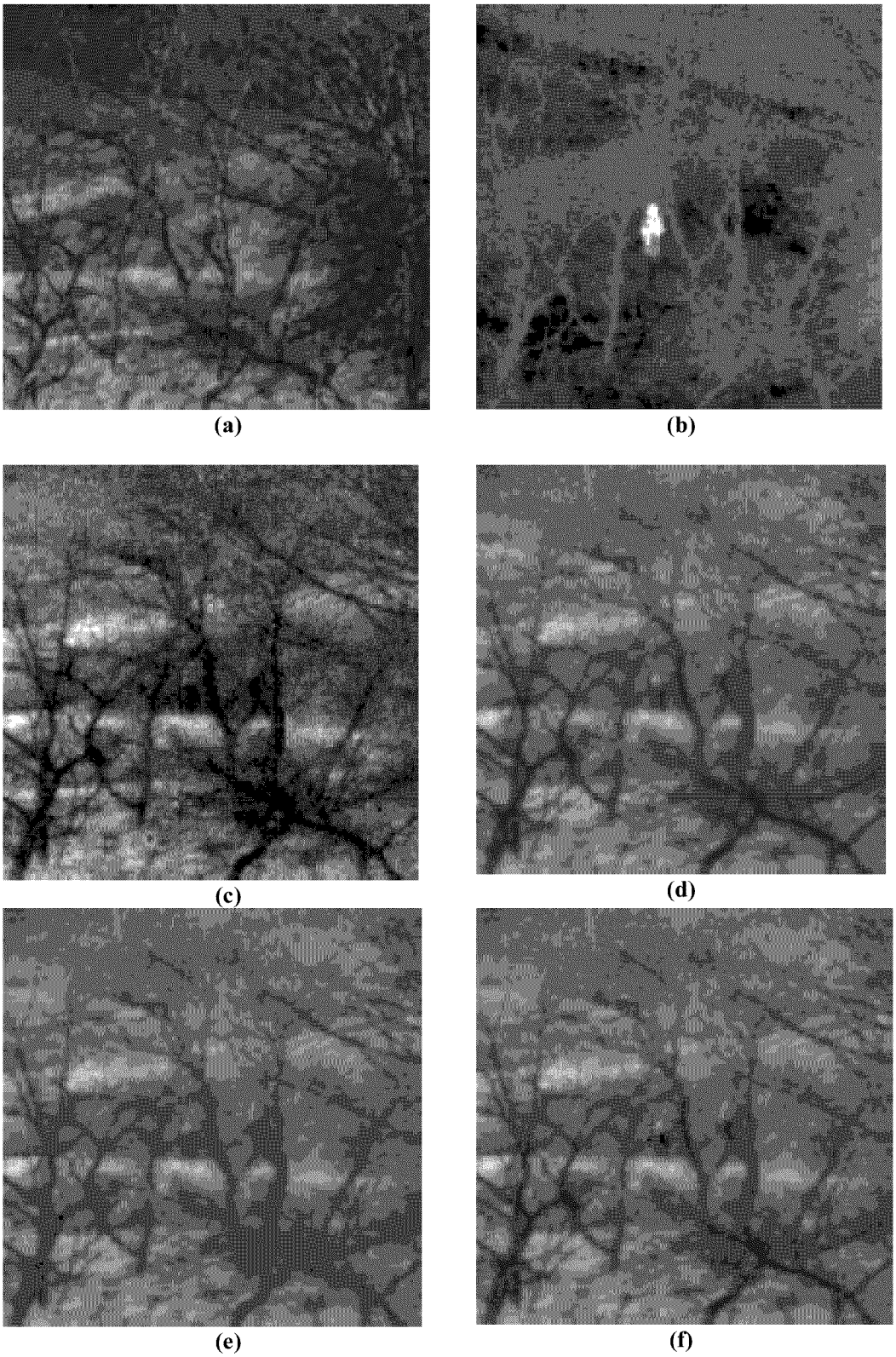


Figure 5 As Figure 1 for a scene representing trees and grassland. Notice the presence of a person in the nighttime images.

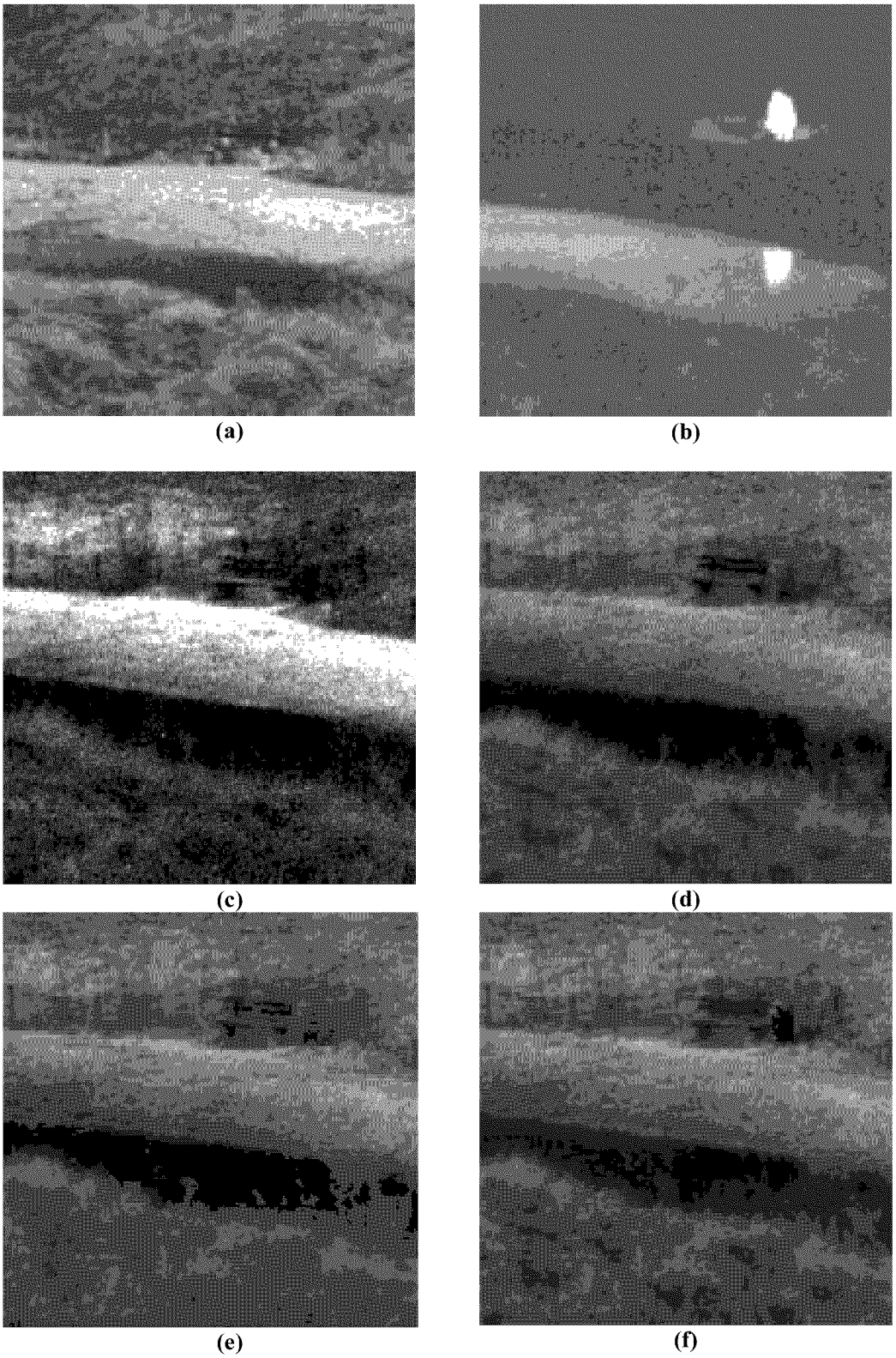


Figure 6 As Figure 1 for a scene representing a bench near the lakeside. Notice the presence of the person in the nighttime images, and his thermal reflection of in the water.

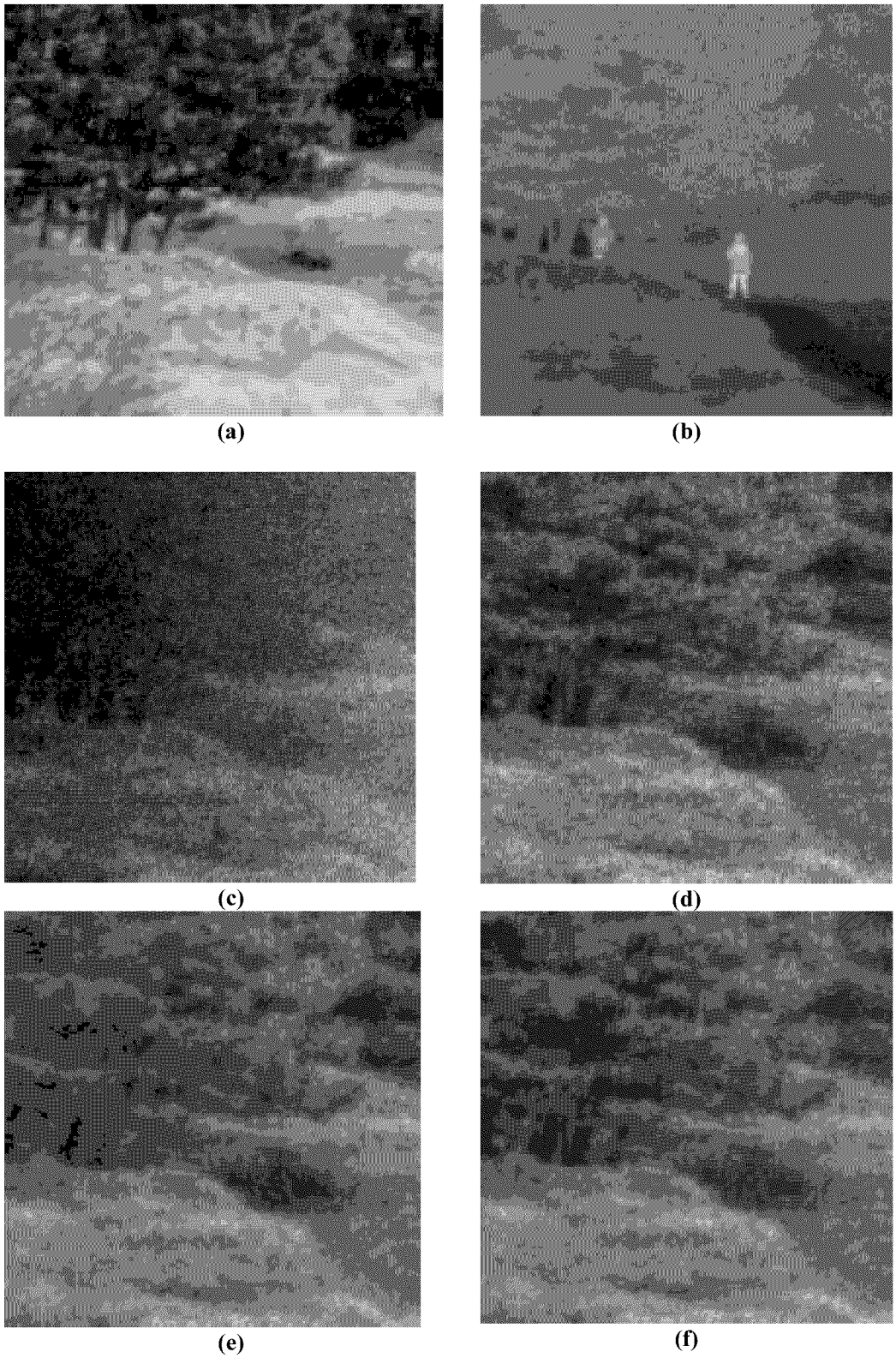
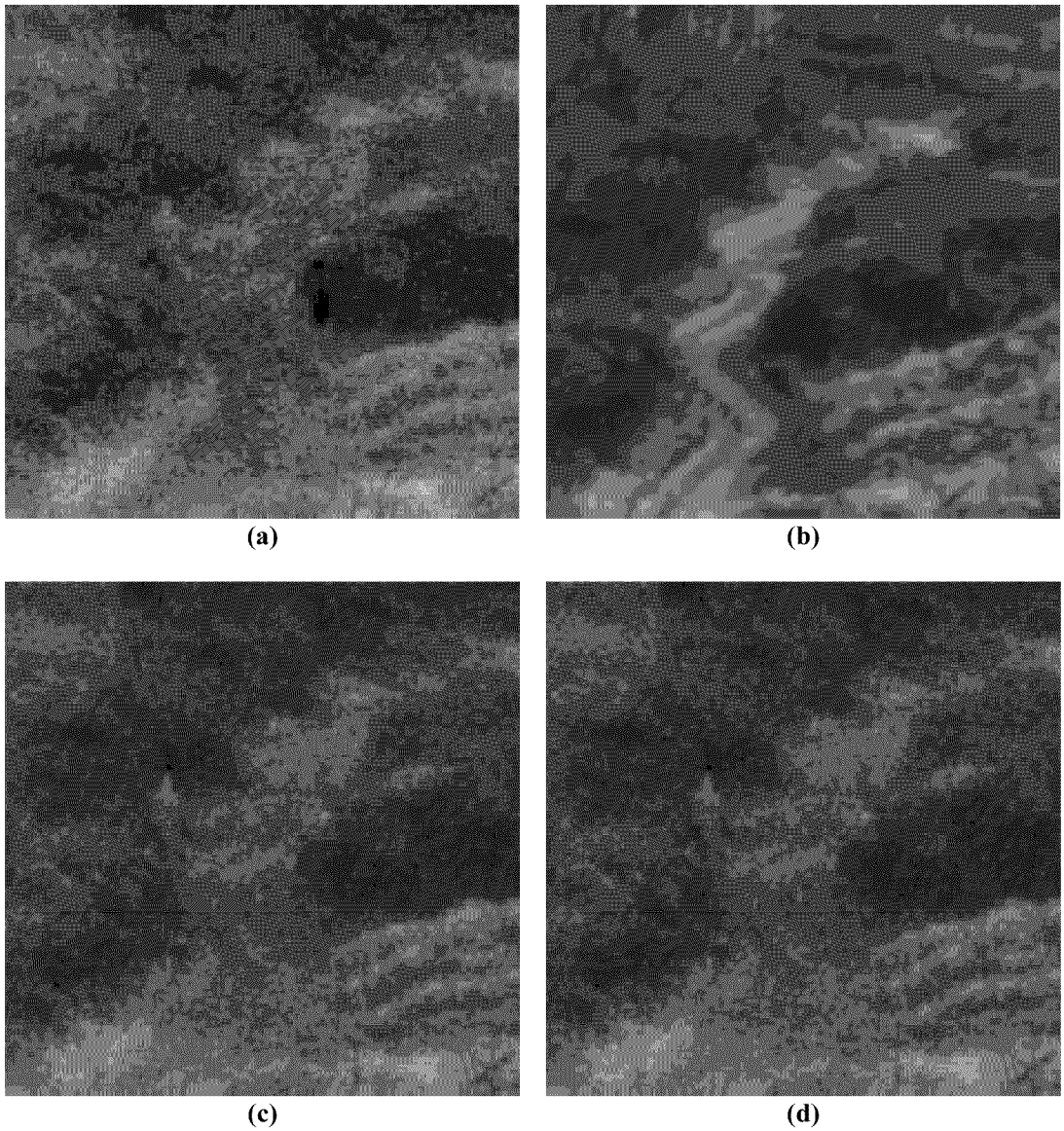


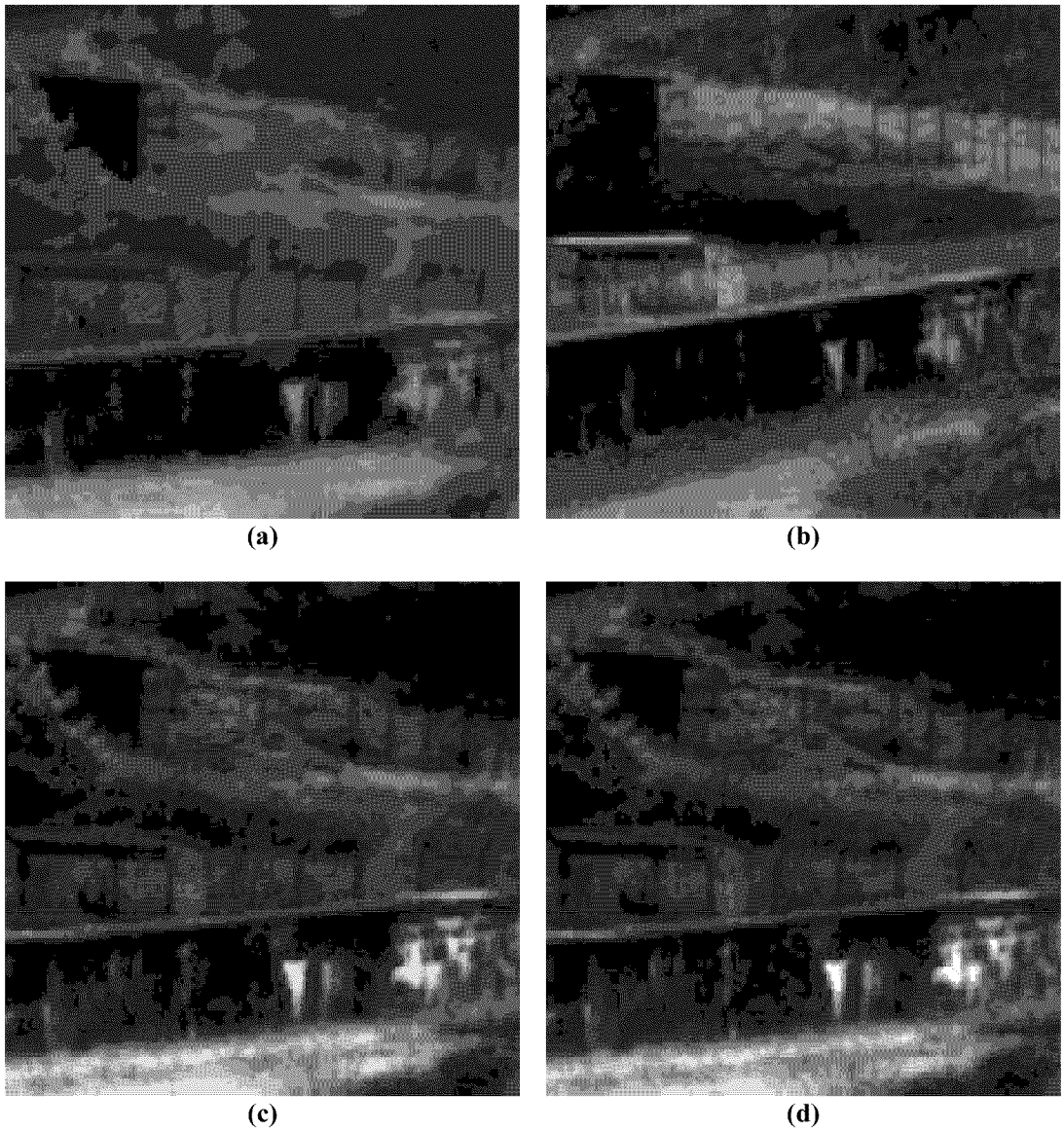
Figure 7 As Figure 1 for a scene representing a sandy path through some trees. Notice the presence of 2 persons in the nighttime images.



(a) **(b)**

(c) **(d)**

Figure 8 This figure shows **(a)** the false colour 3 band multispectral image from Figure 1f, **(b)** the daylight colour photograph of the same part of the scene, **(c)** the result of the transfer of the colours from (b) to (a) by Method I, and **(d)** the result of the transfer of the colours from (b) to (a) by Method II.



(a) **(b)**

(c) **(d)**

Figure 9 This figure shows **(a)** the false colour 3 band multispectral image from Figure 2f, **(b)** a daylight colour photograph of the same part of the scene, **(c)** the result of the transfer of the colours from (b) to (a) by Method I, and **(d)** the result of the transfer of the colours from (b) to (a) by Method II.

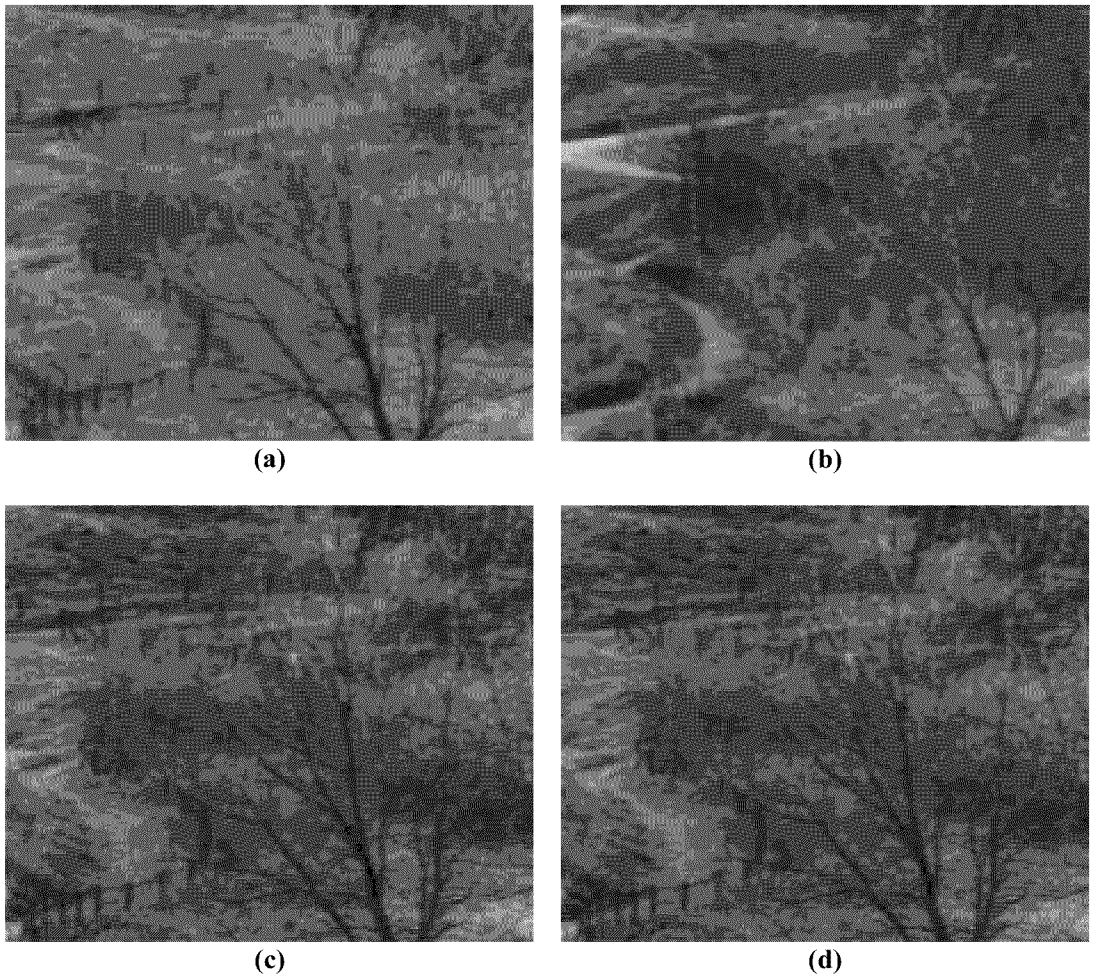
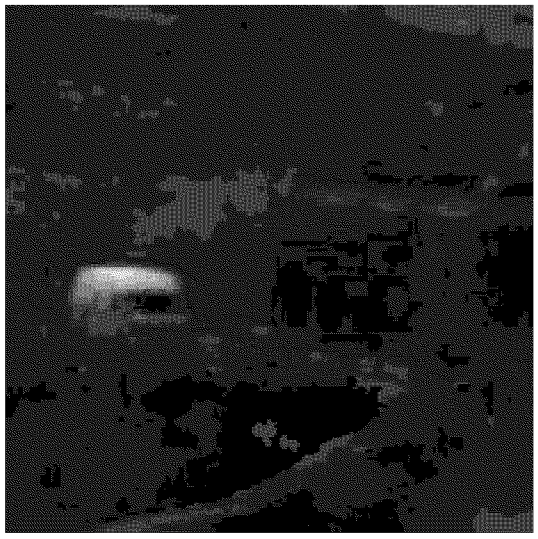
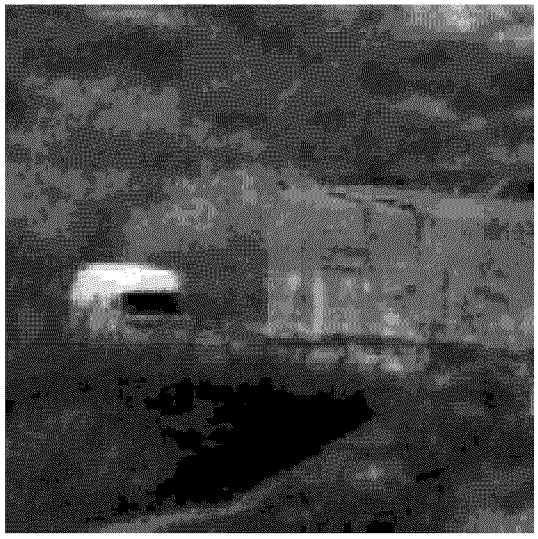


Figure 10 This figure shows **(a)** the false colour multispectral image from Figure 3f, **(b)** a daylight colour photograph of the same part of the scene, **(c)** the result of the transfer of the colours from (b) to (a) by Method I, and **(d)** the result of the transfer of the colours from (b) to (a) by Method II.

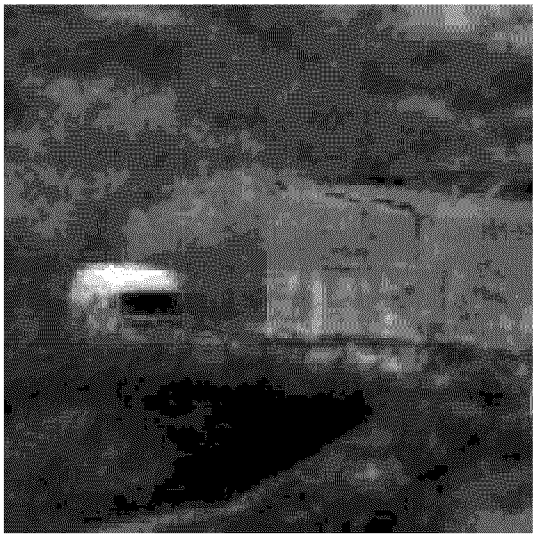


(a)

(b)

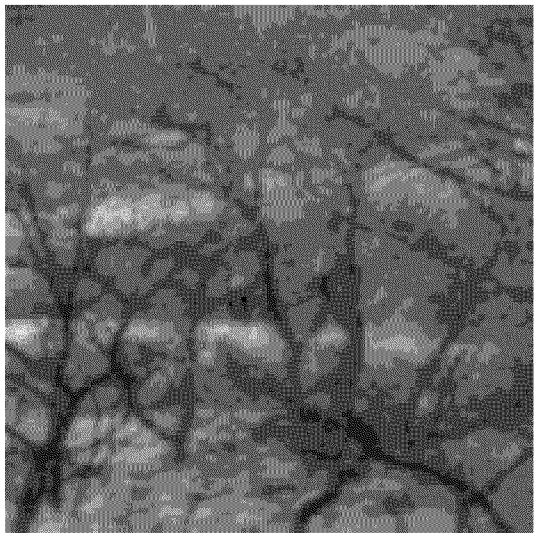


(c)

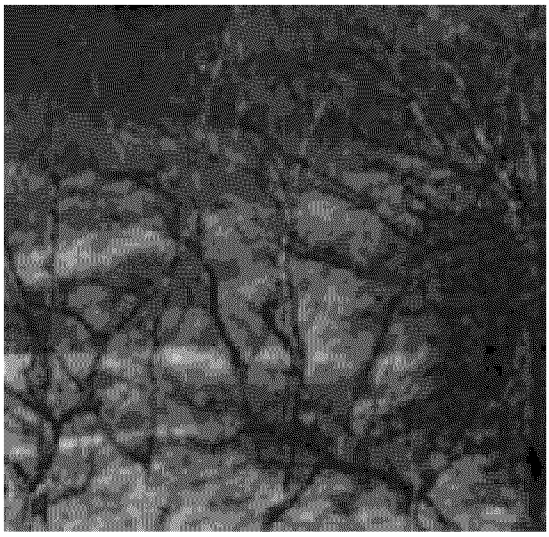


(d)

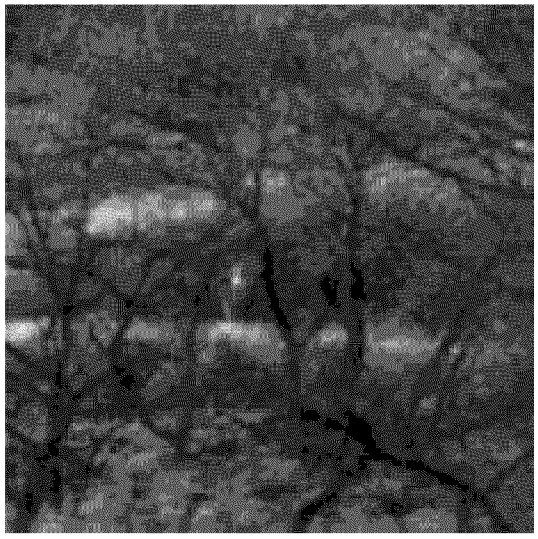
Figure 11 This figure shows **(a)** the false colour multispectral image from Figure 4f, **(b)** a daylight colour photograph of the same part of the scene, **(c)** the result of the transfer of the colours from (b) to (a) by Method I, and **(d)** the result of the transfer of the colours from (b) to (a) by Method II.



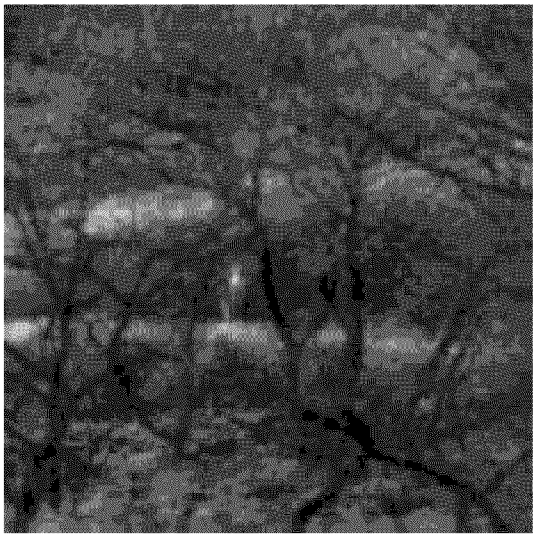
(a)



(b)

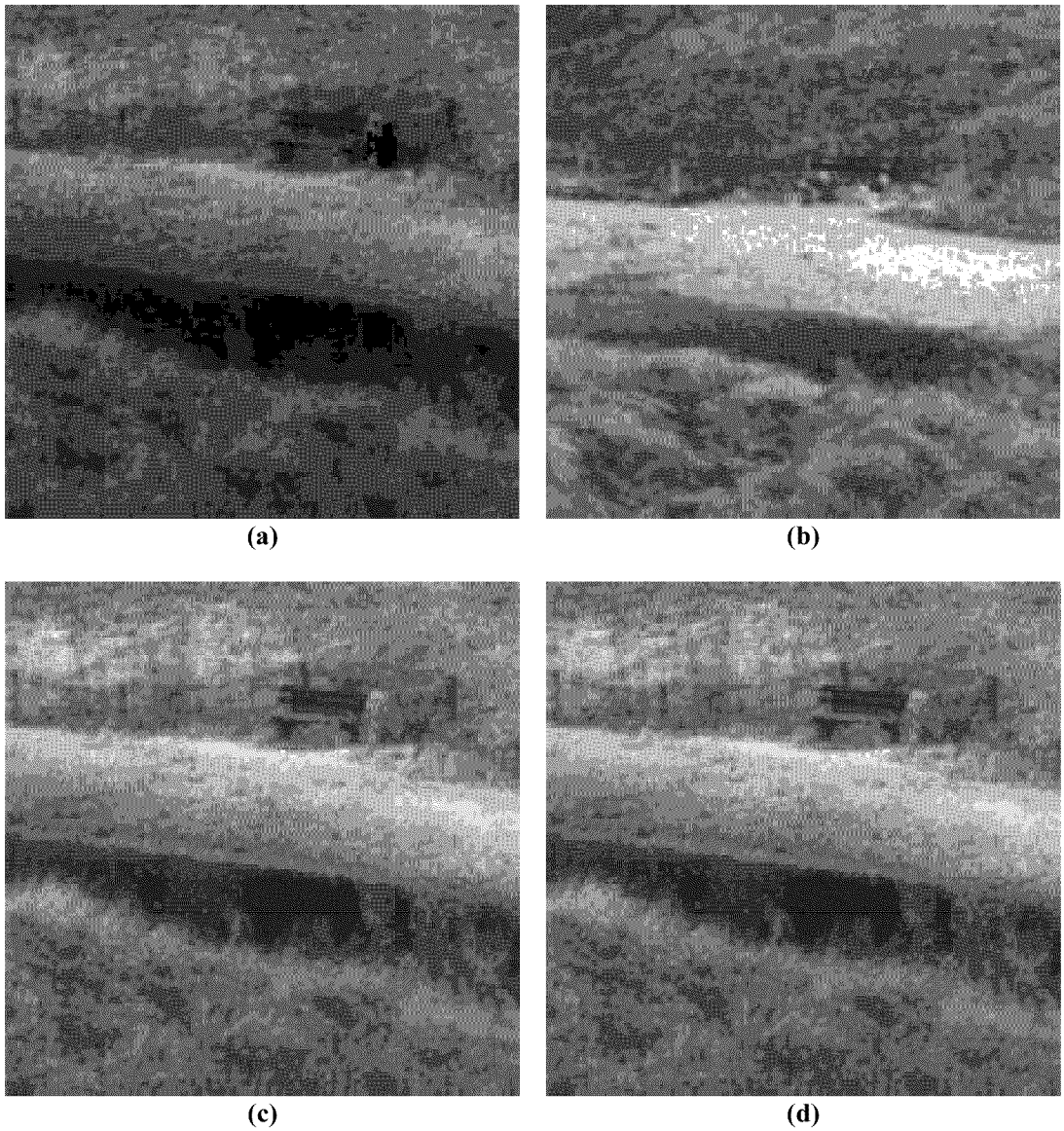


(c)



(d)

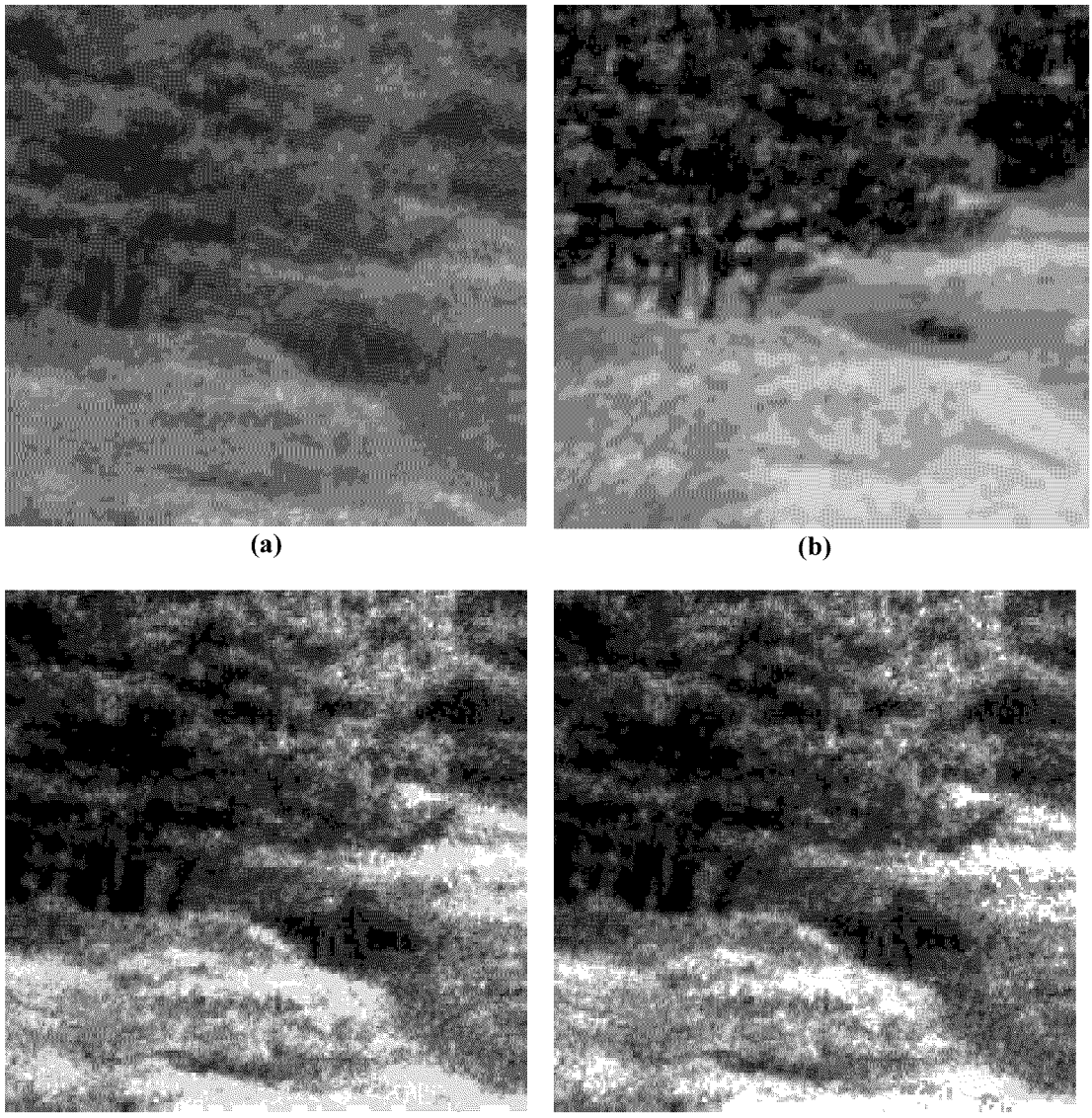
Figure 12 This figure shows **(a)** the false colour multispectral image from Figure 5f, **(b)** a daylight colour photograph of the same part of the scene, **(c)** the result of the transfer of the colours from (b) to (a) by Method I, and **(d)** the result of the transfer of the colours from (b) to (a) by Method II.



(a) **(b)**

(c) **(d)**

Figure 13 This figure shows **(a)** the false colour multispectral image from Figure 6f, **(b)** a daylight colour photograph of the same part of the scene, **(c)** the result of the transfer of the colours from (b) to (a) by Method I, and **(d)** the result of the transfer of the colours from (b) to (a) by Method II.



(a) **(b)**

(c) **(d)**

Figure 14 This figure shows **(a)** the false colour multispectral image from Figure 7f, **(b)** a daylight colour photograph of the same part of the scene, **(c)** the result of the transfer of the colours from (b) to (a) by Method I, and **(d)** the result of the transfer of the colours from (b) to (a) by Method II.

Figure 15 Illustration of the use of different target images in the colour transfer Method II. Top: the 3 band multispectral source image from Figure 5f. Left column: the different target images. Right column: the corresponding results of the colour transfer Method II. The top left column is the daylight photograph of the same scene. The lower left target image is a painting called "Old Oak Tree" by the Dutch master Barend Cornelius Koekkoek (1803-1862).

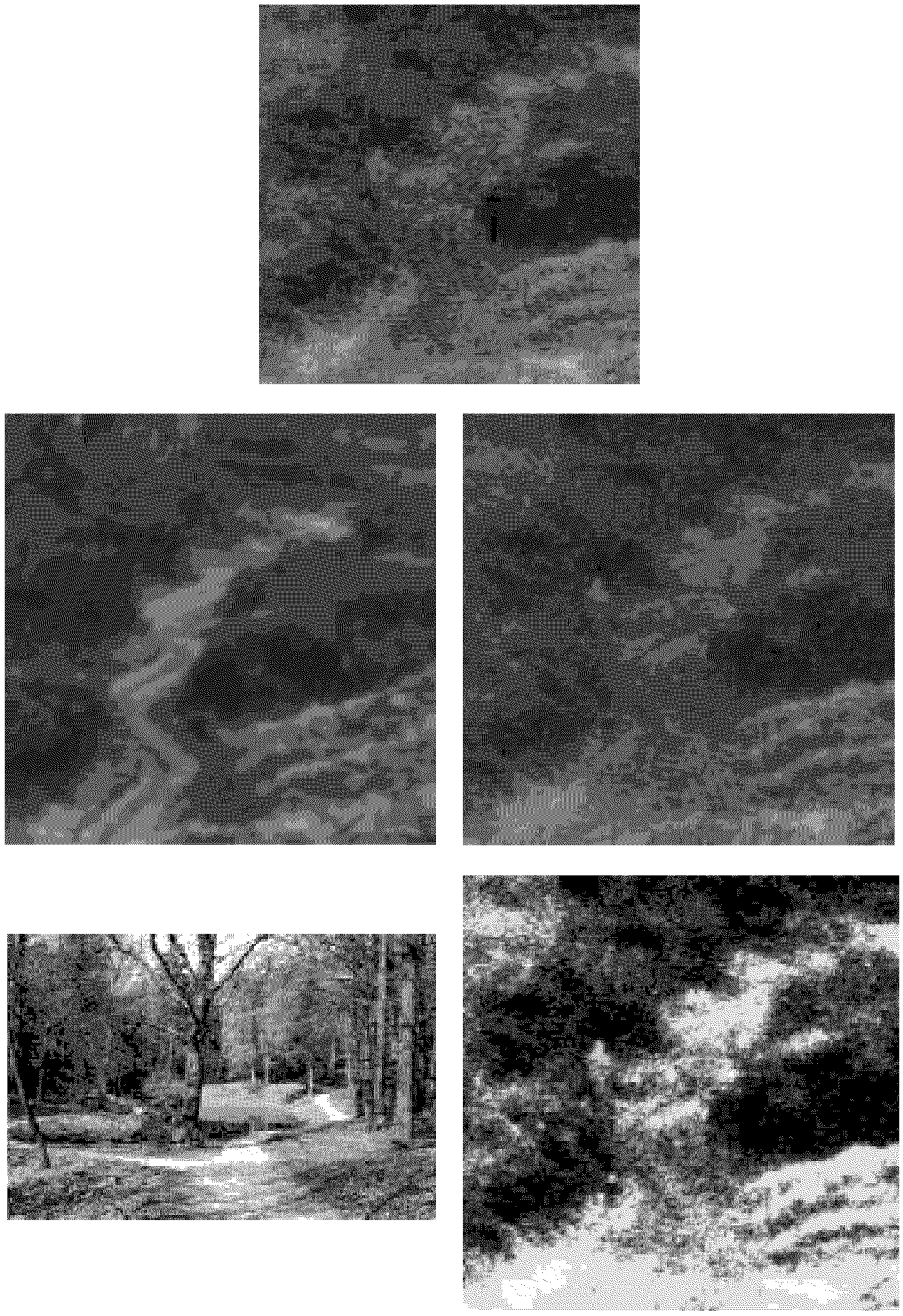


Figure 16 Illustration of the use of different target images in the colour transfer Method II. Top: the 3 band multispectral source image from Figure 3f. Top left column: the daylight photograph of the scene. Bottom left: arbitray scene in the woods. Right column: the corresponding results of the colour transfer Method II.

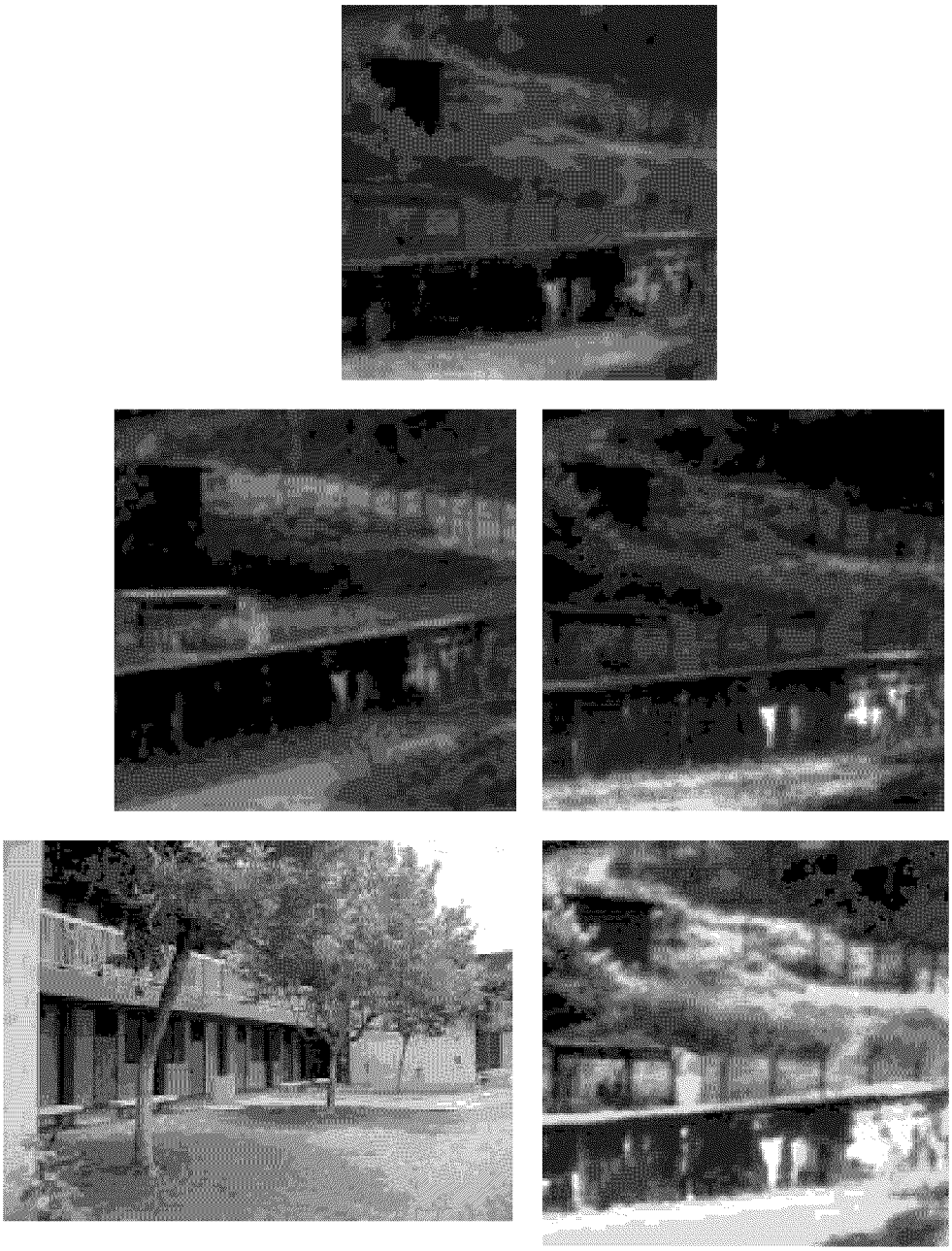


Figure 17 Illustration of the use of different target images in the colour transfer Method II. Top: the 3 band multispectral source image from Figure 2f. Top left column: the daylight photograph of the scene. Bottom left: arbitray image of building with and trees. Right column: the corresponding results of the colour transfer Method II.

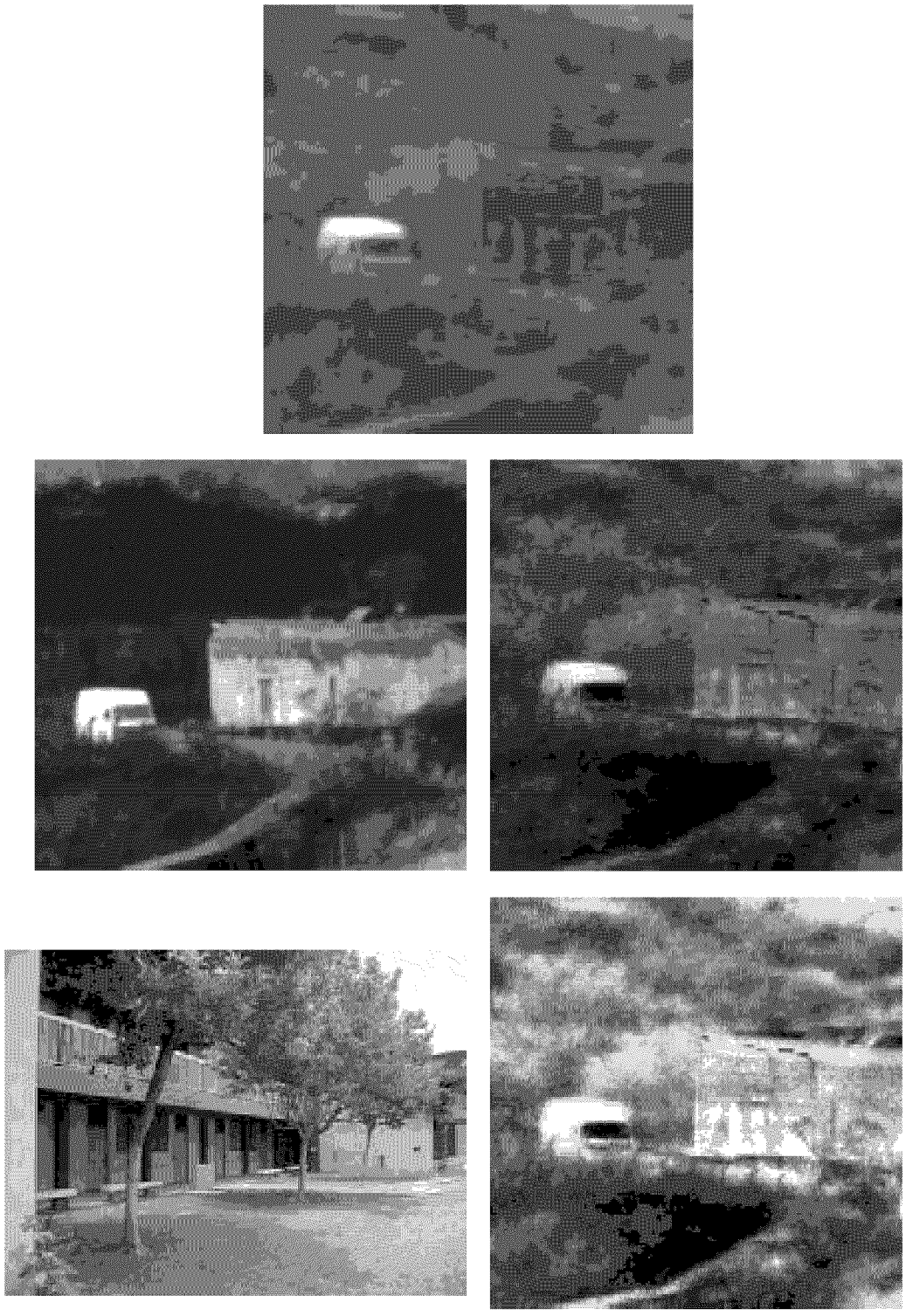


Figure 18 Illustration of the use of different target images in the colour transfer Method II. Top: the 3 band multispectral source image from Figure 4f. Top left column: the daylight photograph of the scene. Bottom left: arbitray image of building with trees. Right column: the corresponding results of the colour transfer Method II.

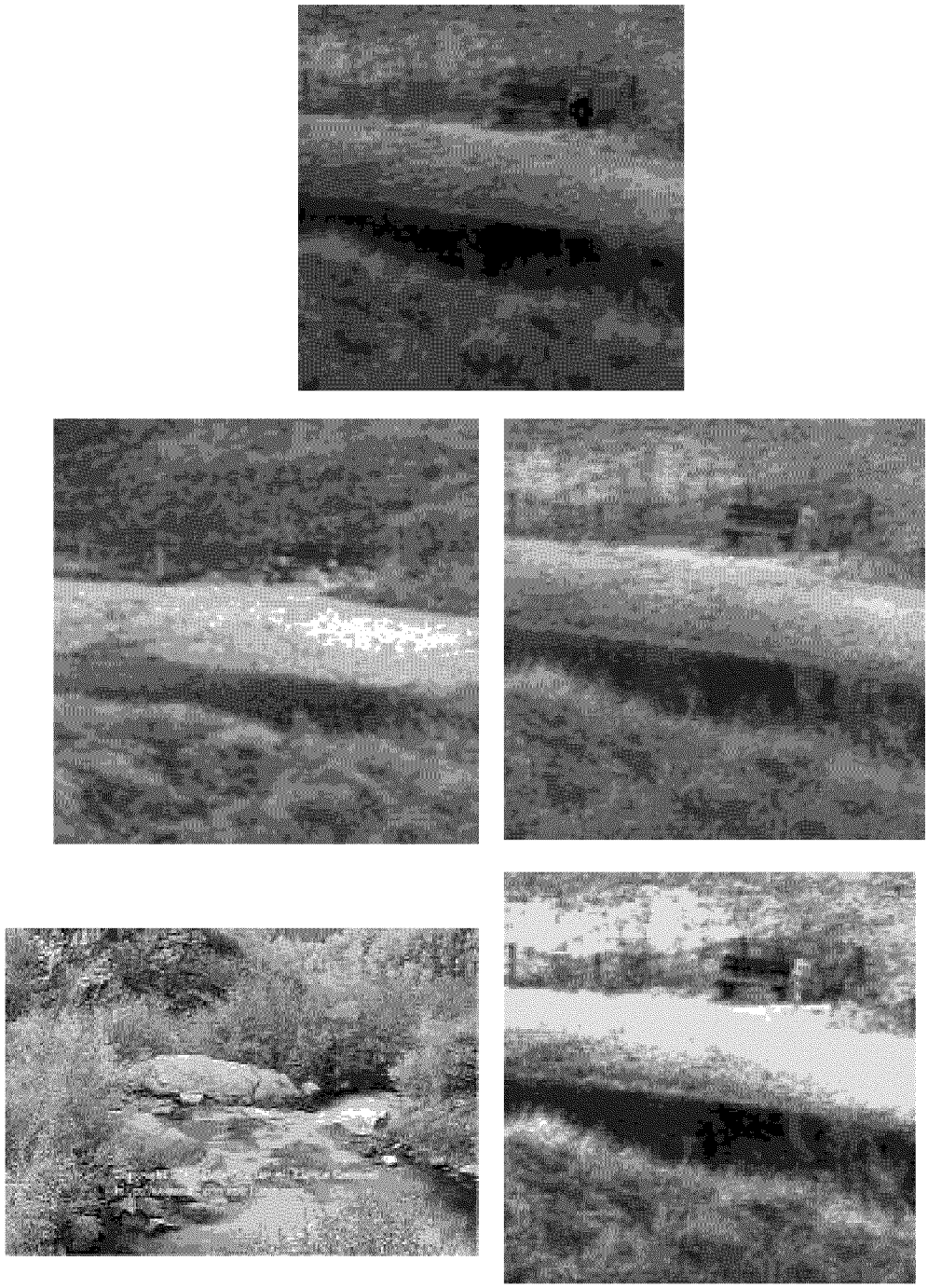


Figure 19 Illustration of the use of different target images in the colour transfer Method II. Top: the multispectral source image from Figure 6f. Top left column: the daylight photograph of the scene. Bottom left: arbitray image of riverside. Right column: the corresponding results of the colour transfer Method II.

Figure 20 Illustration of the use of different target images in the colour transfer Method II. Top: the multispectral source image from Figure 7f. Top left column: the daylight photograph of the scene. Bottom left: arbitray image of a path through trees. Right column: the corresponding results of the colour transfer Method II.

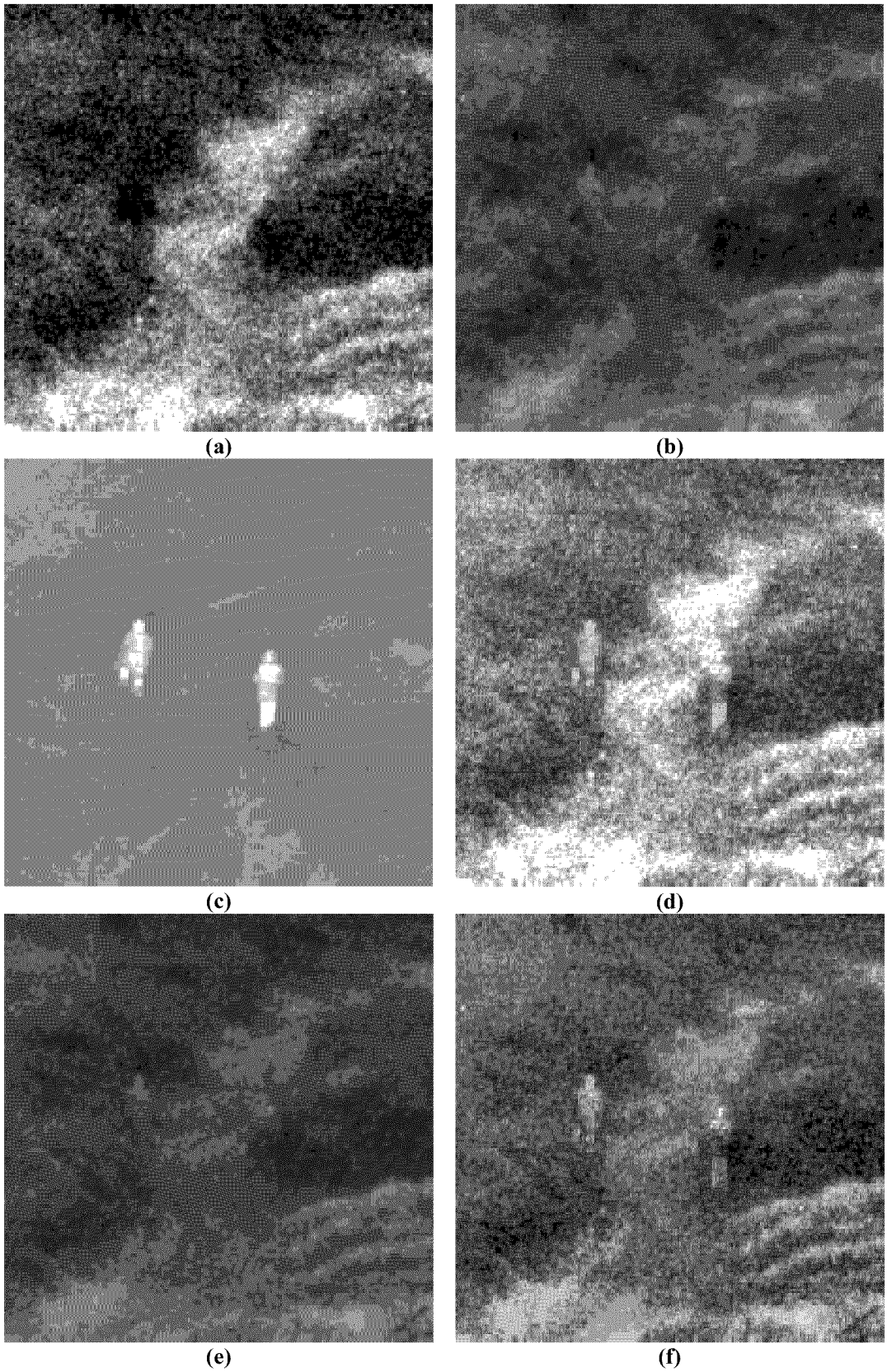


Figure 21 Nighttime images of two persons standing alongside a sandy path through the heath (from Figure 1). **(a)** The visual (400-700 nm) and **(b)** the near infrared (700-900 nm) channels of a double-band image intensifier. **(c)** The corresponding thermal 3-5 μm image of the same scene. (d) The grayscale fused representation of images a,b, and c. **(e)** The multiband colour image from Figure 8**c**. **(f)** Result of replacing the luminance component of (e) with the grayscale fused image (d).

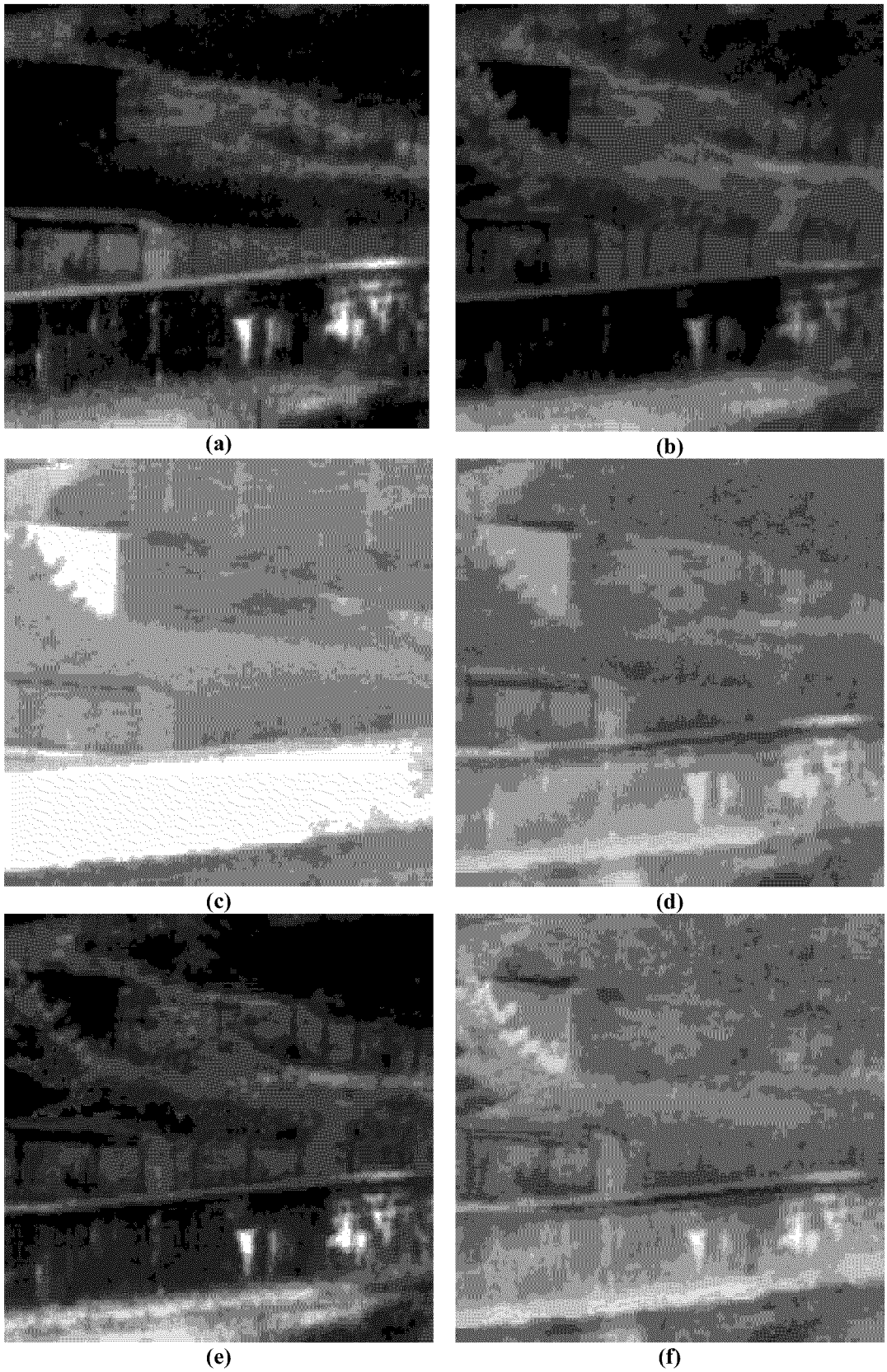


Figure 22 As Figure 21, for the images from Figure 2.

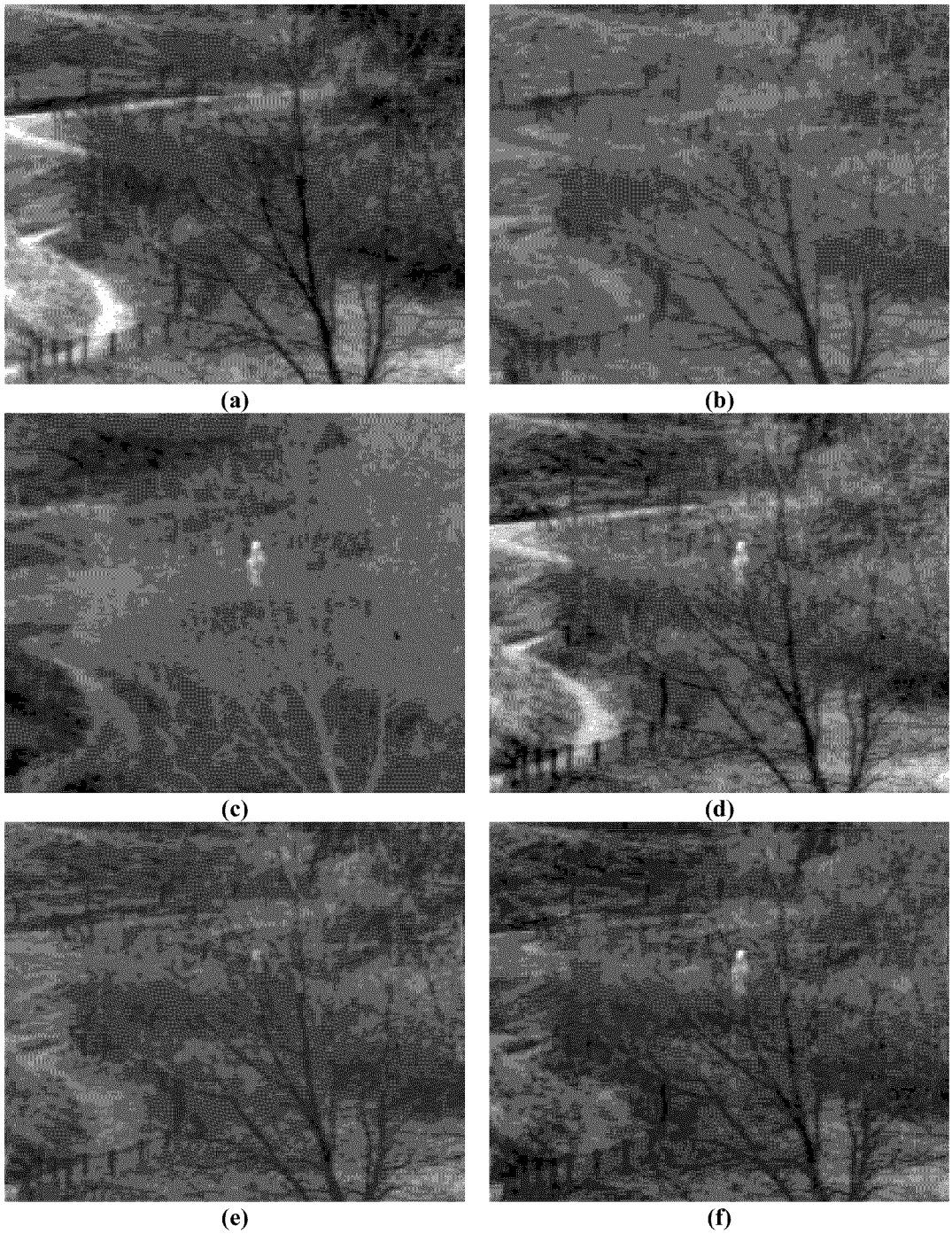


Figure 23 As Figure 21, for the images from Figure 3.

Figure 24 As Figure 21, for the images from Figure 4.

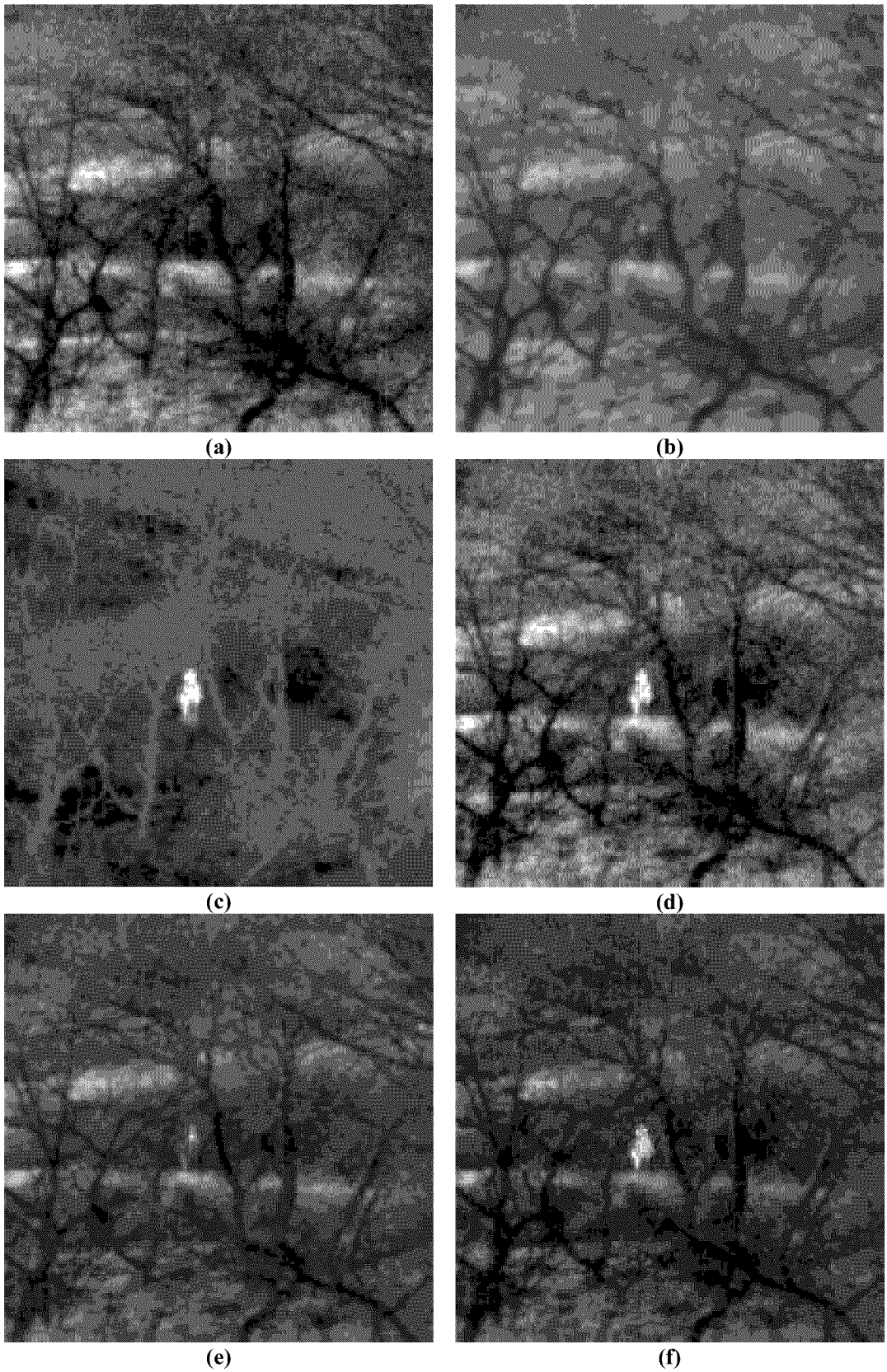


Figure 25 As Figure 21for the images from Figure 5.

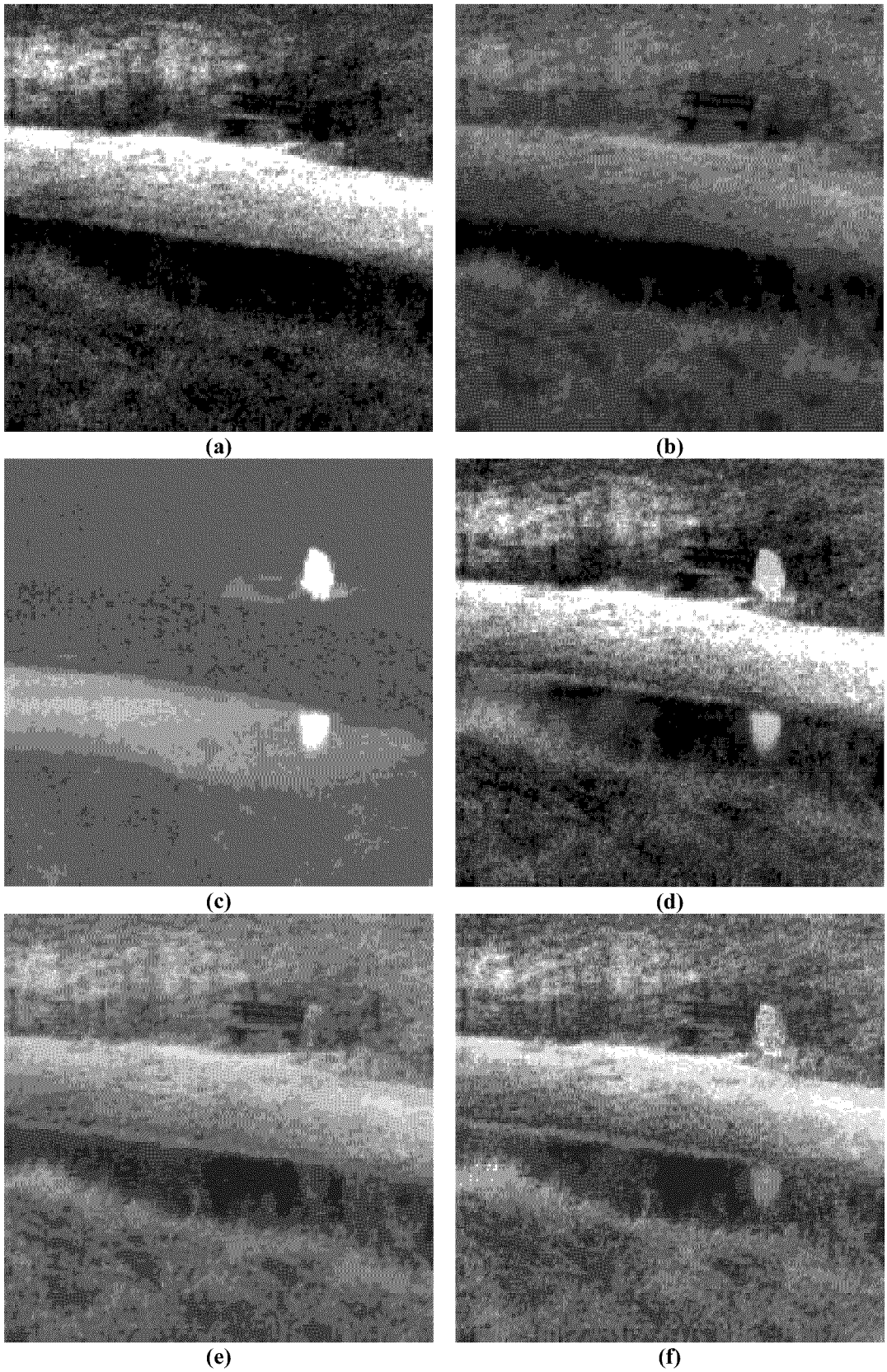


Figure 26 As Figure 21, for the images from Figure 6.

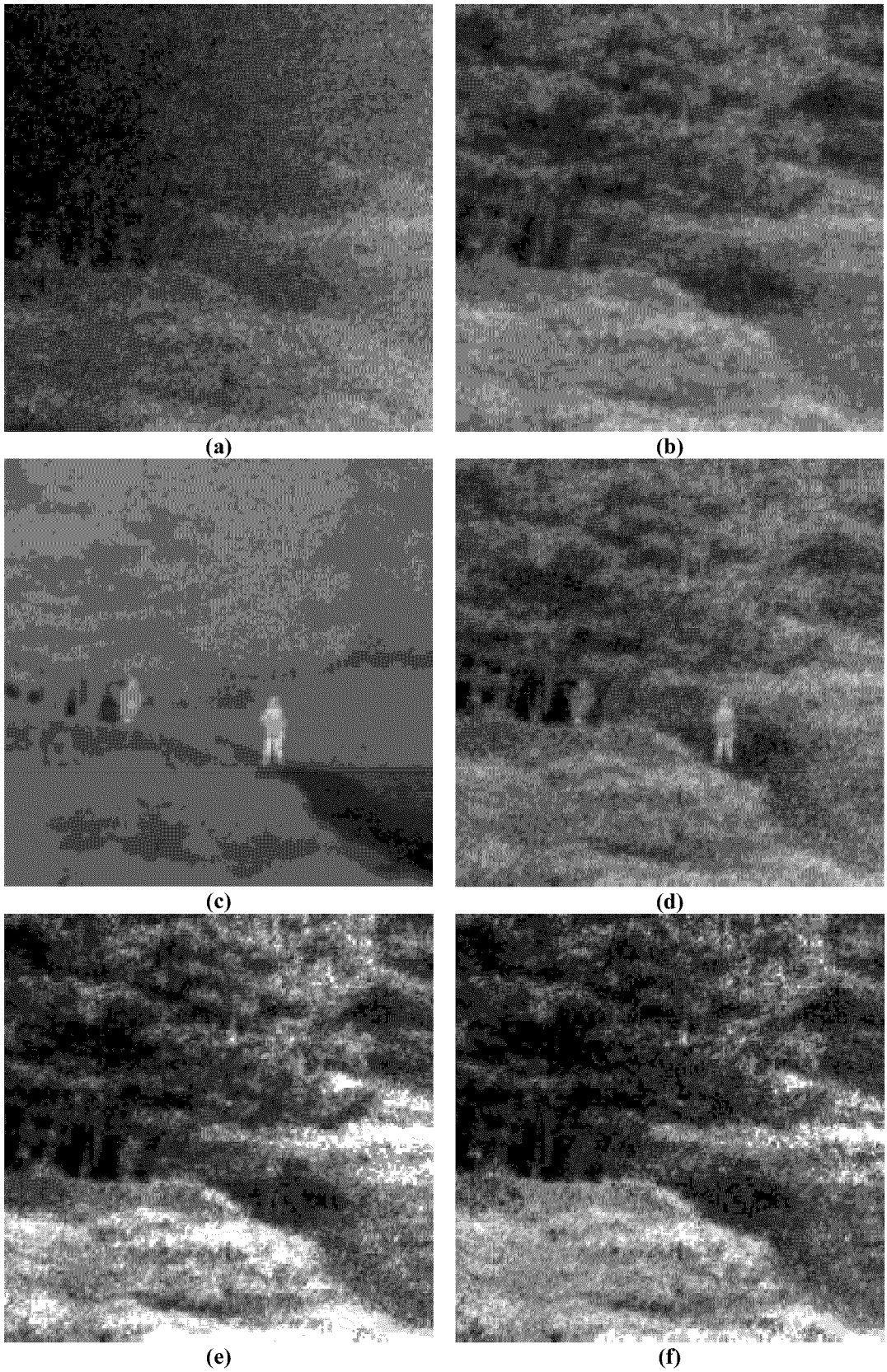


Figure 27 As Figure 21, for the images from Figure 7.

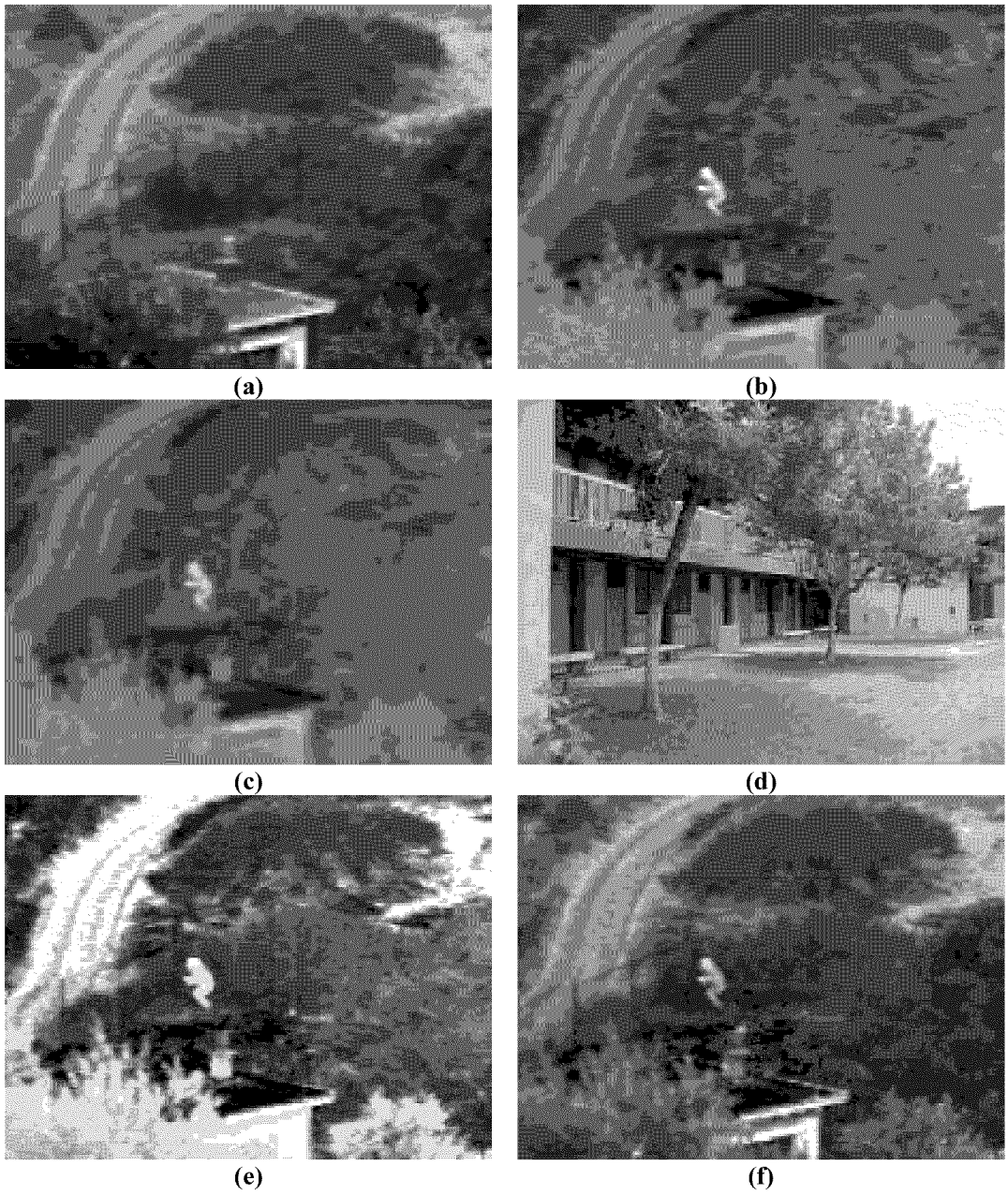


Figure 28 Nighttime images of a person walking along the outside of a fence. **(a)** The visual (400-700 nm) intensified image. **(b)** The corresponding thermal 3-5 μm image of the same scene. **(c)** The false colour image obtained by mapping (a) and (b) to respectively the B and G channels of an RGB image representation. **(d)** Arbitrary colour image with similar content. **(e)** Result of colour transfer Method II applied to (c) with (d) as target image. **(f)** Result of replacing the luminance component of (e) with the grayscale fused image of (a) and (b) (not shown here).

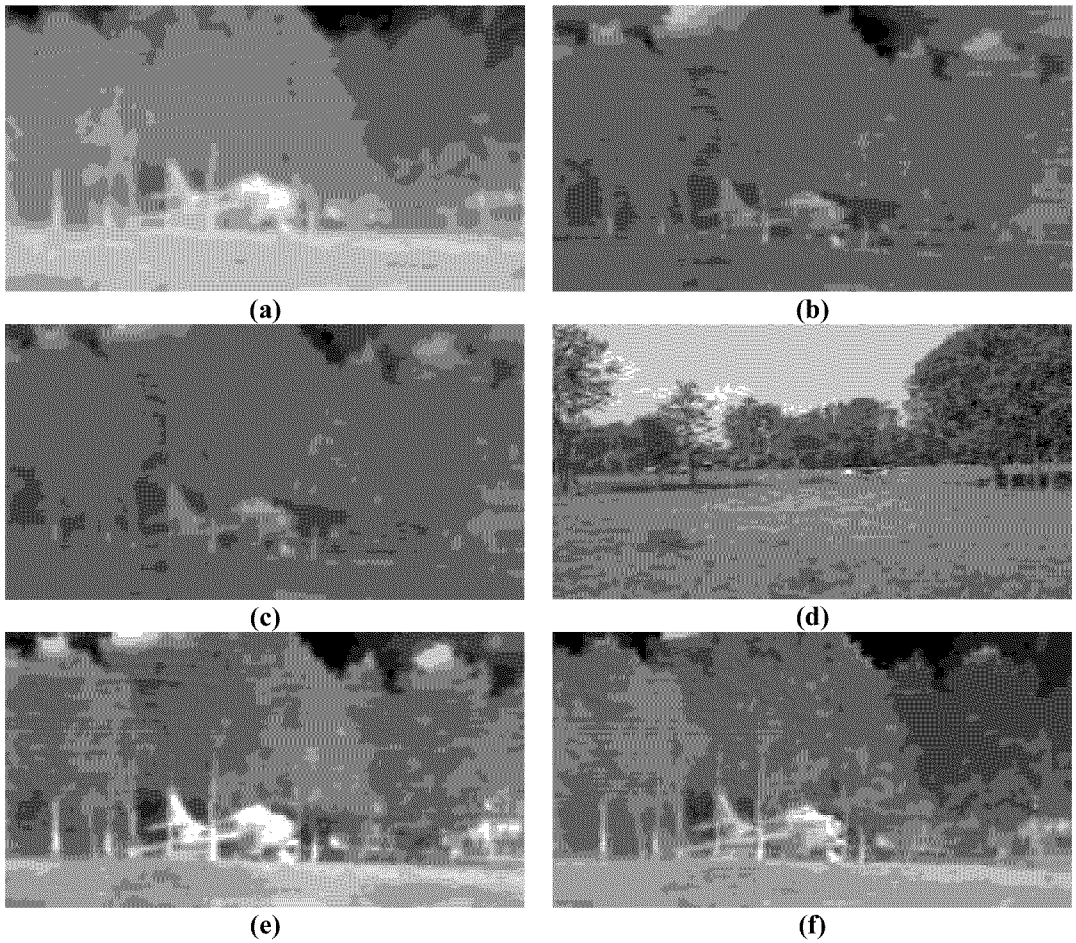


Figure 29 Nighttime images of an airplane hidden in a treeline. **(a)** The thermal 3-5 μm image. **(b)** The corresponding thermal 8-10 μm image of the same scene. **(c)** The false colour image obtained by mapping (a) and (b) to respectively the R and G channels of an RGB image representation. **(d)** Arbitrary colour image of a similar scene. **(e)** Result of colour transfer Method II applied to (c) with (d) as target image. **(f)** Result of replacing the luminance component of (e) with the grayscale fused image of (a) and (b) (not shown here).

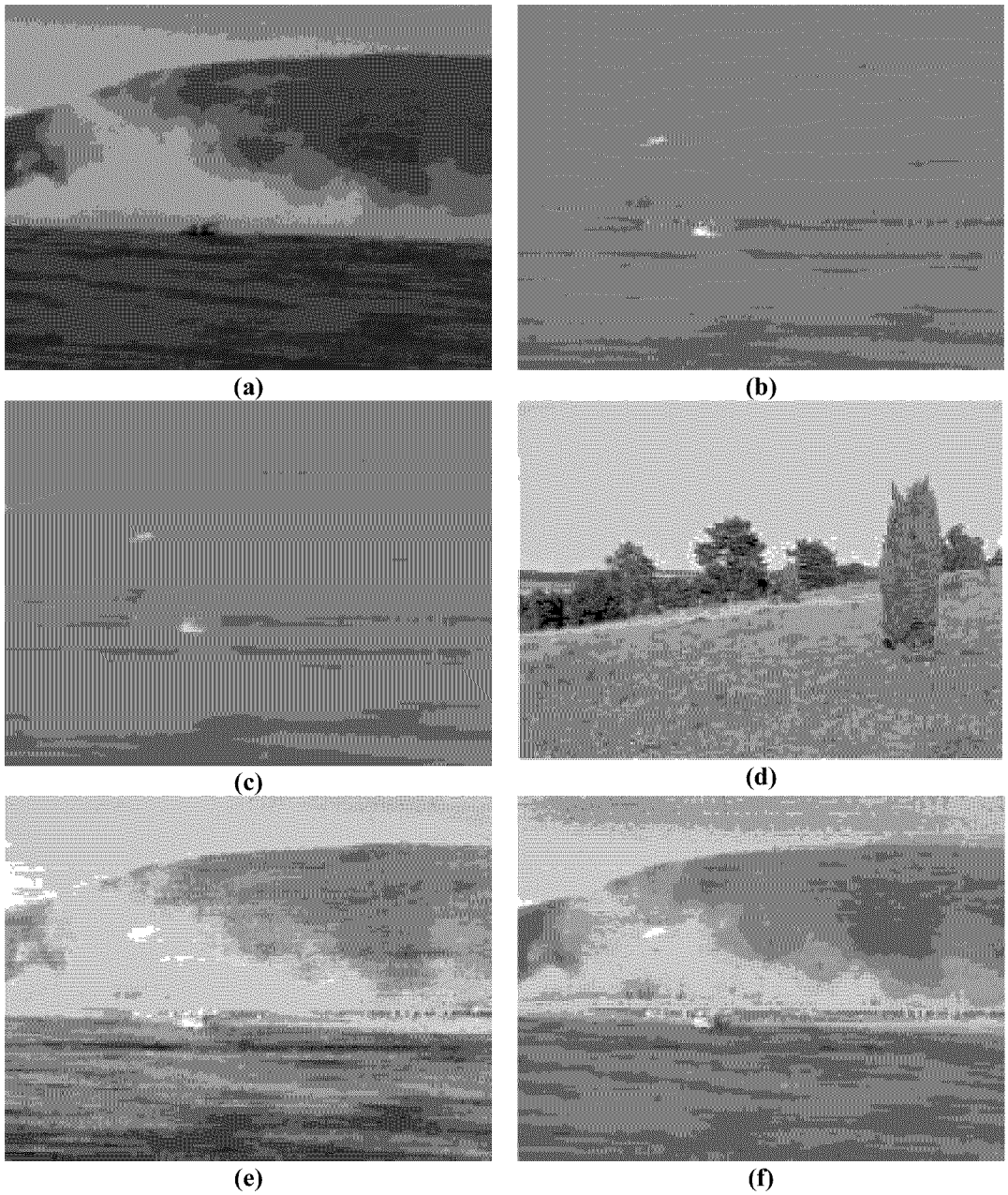


Figure 30 Images of battlefield with trucks, persons and a helicopter behind a smoke screen, and a hill in the background. **(a)** The visual CCD grayscale image. **(b)** The corresponding thermal 8-12 μm image of the same scene. **(c)** The false colour image obtained by mapping (a) and (b) to respectively the B and G channels of an RGB image representation. **(d)** Arbitrary colour image of a similar scene. **(e)** Result of colour transfer Method II applied to (c) with (d) as target image. **(f)** Result of replacing the luminance component of (e) with the grayscale fused image of (a) and (b) (not shown here).

Enhanced visualisation of nightvision imagery through fusion with Ladar depth maps

F. Kooi & A. Toet

1. INTRODUCTION

It has been argued that the fusion of range data from a ladar with visual and thermal imagery may help to discriminate background from target pixels (Aggarwal & Chu, 1992; Hackett & Shah, 1989; Mitiche et al., 1982; Rogers et al., 1989). Many objects of interest, such as vehicles and buildings, are characterized in at least some of their extent by relatively smooth surfaces with slowly varying range data that result in small values of gradients of the range data. A threshold based on the first derivative of the log-transformed range gradient can effectively be used to segment the targets from their background (Rogers et al., 1989).

At the start of the project we expected to receive a visual + ladar database from MIT Lincoln Labs. Due to personnel changes (Prof Waxman has left MIT) the images could not be delivered, except for one set. We have therefore based our analysis of the perceptual benefits of the inclusion of Ladar information in nightvision imagery on the following sources:

- A single MIT data set
- Technical knowledge of Ladar
- Knowledge of viewing 3D images
- Knowledge of color coding and color fusion

Initially we expected that a 3D visualization would be most useful. However, our final conclusion is that color coding is more promising.

In this report we argument the case and derive some "design rules" for the use of Ladar data in both types of visualization.

2. STEREOSCOPIC 3D IMAGE

Mismatches between a (Ladar) depth map and the structural image content will inevitably result in poor viewing comfort and an unpredictable visibility of depth differences. Causes for these effects are:

- Low resolution of the depth map
- Local lack of depth information (lack of reflection)
- Local lack of image information (zero intensity; lack of light)
- The problem of occlusion, resulting in a need to fill in missing image parts
- Can't deal with transparency
- The need to enhance the depth to perceptually detect small depth differences

These effects result in very high depth gradients at many places in the scene which adversely affects visual perception (Kooi & Lucassen, 2001; Tyler, 1991).

The primary advantage of depth coding is that it can convey most of the information contained in the Ladar image. The alternative (color coding) requires pre-processing which implies that part of the depth information is discarded.

3. COLOR LINE CODING

The problems described above (limited viewing comfort and visibility of depth information) may in principle be avoided by color coding the depth information. We propose to emphasize segments with steep depth gradients by outlining them with colored borders, keeping the original image as 'clean' as possible. We expect that the major stumbling block will be the selection criterion or depth threshold: which type of depth gradients will be colored and which won't? Because the MIT Ladar data turned out not to be available after all, we have not been able to experimentally test our ideas on real data. However, we do have some suggestions how to proceed:

- Make the colored lines equiluminant with the local image content, except on those locations where the image is too dark. In these cases increase the line brightness to approximately 3x threshold visibility.
- Only show the lines when they do not directly correspond to structures in the image. If the structure of the depicted scene is clearly visible, depth changes will also be easily perceived.
- Use distinct colors to code different depth gradient magnitudes. Use for example three categories: very steep = red / steep = yellow / shallow = green OR use different line thickness.
- Reduce the amount of clutter by only showing extended lines, not short segments.

4. CONCLUDING REMARKS

Our analysis of the characteristics of Ladar imagery indicates that mismatches between a (Ladar) depth map and the structural content of other image modalities will inevitably result in poor viewing comfort and an unpredictable visibility of depth differences. We therefore conclude that color coding the depth information should be further exploited.

Remaining issues that need further investigation include:

- A situation in which a camouflaged vehicle is positioned at an angle with respect to the viewing point, resulting in a shallow depth gradient across the vehicle. We expect this situation won't be a problem, since we only represent the outline anyway.
- A situation in which a vehicle is covered with a camouflage net in front of a treeline. It is a priori not clear whether there will be sufficient depth nformatio to segment the target from its local background.

REFERENCES

Aggarwal, J.K. & Chu, C.-C. (1992). Image interpretation using multiple sensing modalities. *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI, 14(8),* 840-847.

Hackett, J.K. & Shah, M. (1989). Segmentation using intensity and range data. *Optical Engineering, 28(6),* 667-674.

Kooi, F.L. & Lucassen, M. (2001). Visual comfort of binocular and 3-D displays. In B.E. Rogowitz & T.N. Pappas (Ed.), *Human Vision and Electronic Imaging VI* (pp. 586-592). Bellingham, WA: The International Society for Optical Engineering.

Mitiche, A., Gil, B. & Aggarwal, J.K. (1982). On combining range and intensity data. *Pattern Recognition Letters, 1,* 87-92.

Rogers, S.K., Tong, C.W., Kabrisky, M. & Mills, J.P. (1989). Multisensor fusion of ladar and passive infrared imagery for target segmentation. *Optical Engineering, 28(8),* 881-886.

Tyler,C.W. (1991). Cyclopean vision. In D.Regan (Ed.), *Vision and visual dysfunction. Vol. 9: Binocular Vision.* (pp. 38-74). London, UK: Macmillan.