An analysis of fusion algorithms for LWIR and visual images

Jason de Villiers
Council for Scientific and Industrial Research, Pretoria, South Africa

Robert Jermy
Council for Scientific and Industrial Research, Pretoria, South Africa
Email: rjermy@csir.co.za

***Abstract*—This paper presents a comparison of methods to fuse pre-registered colour visual and long wave infra-red images to create a new image containing both visual and thermal cues. Three methods of creating the artificially coloured fused images are presented. These three methods along with the raw visual and LWIR imagery are then evaluated using the Analytical Hierarchy Process for three different scenarios using a set of 32 observers. The scenarios entail bright, dim and dark conditions which directly affect the amount of visual information available. Both the standard method and a novel voting methodology are used to evaluate the results, the latter providing similar ranking but better discrimination between the voter's preferences. The results show that fused images are preferred for non-dark conditions with the thermal based hue offset algorithm being preferred.**

I. Introduction

This paper creates and evaluates multi-spectral fused images. Specifically visual and Long Wave Infrared (LWIR) imagery are used. The fusion of data is a two-step process consisting of registering the images and then using the registered images to create a false-colour image for display purposes. This paper focusses on the false-colour image creation and the evaluation of such images for the purpose of object of interest detection and subsequent identification.

A. Cross spectral registration

Registration is the first step in multi-spectral fusion, it entails the warping of the two images such that corresponding features in the images coincide. The authors have previously shown [1] that standard feature descriptors yielded poor results when used to generate homographies to register cameras of spectral bands as different as the visual spectrum ($0.3\mu m$ to $0.7\mu m$) and LWIR ($8\mu m$ to $14\mu m$). Ergo this work used the authors' patented robot arm calibration technique [2] to determine each camera's forward and reverse distortion parameters (required to correct the barrel effects), focal length, pixel size, principal point and relative six degree of freedom (DOF) positions. These parameters were then used to project the cameras onto one or more virtual projection geometries [3] to photogrammetrically stitch them in real time.

B. Creating meaningful false coloured fused images

Once the images can be correctly overlaid, the question arises as to how to display this data. A standard computer screen has only 24 bits resolution per pixel, 8 bits for each of red, green and blue channels. This exactly coincides with the data received from the visual camera, however one wants to simultaneously display the thermal data (which is typically of at least 10 bit resolution) together with the visual information. Intuitively one may associate hot items with red and cold items with blue, however these colours may already be present in the visual image. For instance a blue vehicle with its engine running will also be hot, and so must simultaneously be rendered both blue and red.

For this work it is assumed that not only is the detection of objects of interest important but so too is their identification. LWIR imagery could be able to detect a person or vehicle in a restricted area but normally not be able to provide information to apprehend the person (e.g. "Stop the guy wearing the awesome Metallica T-shirt!"). For this reason there is no clear best algorithm or heuristic to judge algorithms. Ergo this work presents a suite of methods to create false colour images as well as an assessment by a panel of evaluators as to which algorithm performs the best in three different lighting situations.

C. Related Work

In their paper Hama et. al [4] make use of three layers of image fusion to create meaningful images. The first layer was pixel-level fusion, where the values of pixels are merged. The second layer was feature-level fusion, where salient features are detected in each image and are highlighted in the fused image. Finally they use decision-level fusion which enhances features in the fused image, while suppressing conflicts. Zheng [5] make use of channel-based colour fusion, he modified the red channel of the input image with the corresponding pixel value from the LWIR image. Li et. al. [6] used a similar channel-based fusion method to Zheng, however they also changed the colour space of the image to YC_BC_R and weighted those values using the LWIR value and then modified the fused image to look similar to a separate sample image. This work uses pixel-level fusion, since the fastest possible fusion method was required. Since the input visual images are colour each colour channel is modified in some way during the fusion process. Image colour space transformations were used specifically transformation from Red-Green-Blue (RGB) to Hue-Saturation-Value (HSV) and the reverse transform. These methods allowed hot and cold pixels to be highlighted with modifications to the red, green, and blue channels and then made more visible with a modification to the value channel in

the (HSV) colour space.

The Analytical Hierarchy Process (AHP) [7] is a widely accepted decision making method which not only determines each evaluator's preference but also provides a measure of their consistency. An overview of AHP is presented in §III.

There is a precedent for using AHP in selection of security-related technologies: Baumbach [8] used a similar process to select optimal camouflage patterns for uniforms.

D. Paper organisation

The rest of this paper is organised as follows: Section II describes the real time fusion algorithms. Section III explains the process used to evaluate the algorithms. Section IV describes the three scenarios used for the evaluation. Section V provides the results of the evaluation and the fusion algorithms. Section VI summarises the results and places them in context.

II. Fusion Algorithms

This section describes the three fusion algorithms that were evaluated. The algorithms make use of both the RGB and HSV colour spaces. The definition and conversion between these colour spaces is widely standardised and can be found in any text book on image processing (e.g. [9]).

Some of the algorithms require a definition of to what extent the item is considered hot or cold. For this the simple clamped linear fuzzy membership functions given in Eq. 1 and Eq. 2 were used.

$$f_H = \begin{cases} 0 & \text{if } \mathfrak{T} \in [0, \alpha_{H1}] \\ \frac{\mathfrak{T}-\alpha_{H1}}{\alpha_{H2}-\alpha_{H1}} & \text{if } \mathfrak{T} \in (\alpha_{H1}, \alpha_{H2}] \\ 1 & \text{if } \mathfrak{T} \in (\alpha_{H2}, 1] \end{cases} \tag{1}$$

$$f_C = \begin{cases} 1 & \text{if } \mathfrak{T} \in [0, \alpha_{C1}] \\ \frac{\alpha_{H2}-\mathfrak{T}}{\alpha_{H2}-\alpha_{H1}} & \text{if } \mathfrak{T} \in (\alpha_{C1}, \alpha_{C2}] \\ 0 & \text{if } \mathfrak{T} \in (\alpha_{C2}, 1] \end{cases} \tag{2}$$

where:

f_H = the membership function for hot elements,
f_C = the membership function for cold elements,
α_{H1} = the lower threshold for the hot element,
α_{H2} = the upper threshold for the hot element,
α_{C1} = the lower threshold for the cold element,
α_{C2} = the upper threshold for the cold element, and
$\mathfrak{T}$ = the normalised intensity value of the LWIR pixel.

A. Thermal Tinged Gray Scale

The thermal tinged gray scale (TTGS) algorithm creates a false coloured image based on a grayscale version of the visual pixel. The visual pixel is converted to grayscale using the formula in Eq. 3. The RGB output pixel is created by amplifying the red and blue channels by f_H and f_C as determined by Eq. 1 and 2 respectively. The green channel is set such that the geometric mean of the three channels is the same as the grayscale value.

$$P'_{RGB} = \begin{bmatrix} \mathfrak{G}\,(1+f_H) \\ \frac{\mathfrak{G}}{(1+f_H)(1+f_C)} \\ \mathfrak{G}\,(1+f_C) \end{bmatrix} \tag{3}$$

where:

P'_{RGB} = the output RGB pixel,
P_{RGB} = the input RGB pixel, and

$$\mathfrak{G} = \begin{bmatrix} 0.2126 \\ 0.7152 \\ 0.0722 \end{bmatrix}^T \cdot P_{RGB}.$$

B. Fuzzy

The fuzzy algorithm creates a colour output pixel where hot areas are highlighted red and cold areas are highlighted blue. The input RGB pixel has its channels weighted as shown in Eq. 4 below, this step makes hot pixels more red, and cold pixels more blue. This modified pixel is then converted to the HSV colour space where the hue value is shifted towards red by the hot LWIR weighting value, and towards blue by the cold LWIR weighting value, the value channel is increased slightly to make the colours brighter. This HSV pixel is then converted back to RGB and used as the final output pixel.

$$P'_{RGB} = f_{HtoR}\left({}^{T}P_{HSV} + \begin{bmatrix} f_C \cdot 5^\circ - f_H \cdot 8^\circ \\ 0 \\ 0.02 \end{bmatrix}\right) \tag{4}$$

where:

${}^{T}P_{HSV}$ = an intermediate pixel value in HSV space,

$$= f_{RtoH}\begin{pmatrix} (1-\kappa_1 f_H)\, P_{RGB}.R + \kappa_1 f_H \mathfrak{T} \\ (1-\kappa_2 f_H f_C)\, P_{RGB}.G + \kappa_2 f_H f_C \mathfrak{T} \\ (1-\kappa_3 f_C)\, P_{RGB}.B + \kappa_3 f_C \mathfrak{T} \end{pmatrix},$$

P'_{RGB} = the output RGB pixel,
f_{HtoR} = the function to transform HSV to RGB,
f_{RtoH} = the function to transform RGB to HSV, and
κ_n = the n$^{\text{th}}$ interpolation factor.

C. Thermal Based Hue Offset

The thermal based hue offset (TBHO) algorithm fuses each channel of the input RGB pixel with the LWIR pixel using Eq. 5 and the weights in Table V, this pixel is then converted to the HSV colour space and one of two modifications is made: if the LWIR pixel is hot then the hue channel of the HSV pixel is shifted towards red, conversely if the LWIR pixel is cold then the hue channel of the HSV pixel is shifted towards blue. The output RGB pixel is then calculated from the redder or bluer HSV values if either of these changes were made; or remains the modified RGB pixel value if neither change was made.

(a) LWIR image

(b) Visual image

(c) TTGS fused image

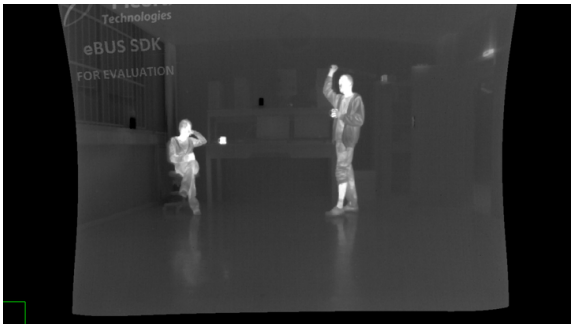
(d) Fuzzy fused image

(e) TBHO fused image

Fig. 1. Bright Scenario

TABLE I
BRIGHT SCENARIO ALGORITHM RANKINGS

Metric	Consistency Ratio (%)	Eigenvector				
		IR	Vis	TTGS	Fuzzy	TBHO
Mean	1.73	0.10	0.20	0.21	0.25	0.23
Mean $(CR < 20)$	1.41	0.14	0.16	0.28	0.21	0.21
Votes		-214.7	53.7	104.6	58.8	-2.4

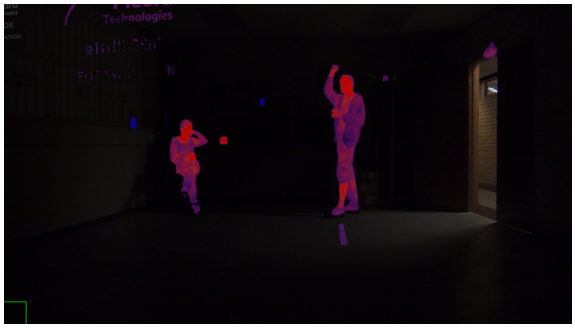


(a) LWIR image

(b) Visual image

(c) TTGS fused image

(d) Fuzzy fused image

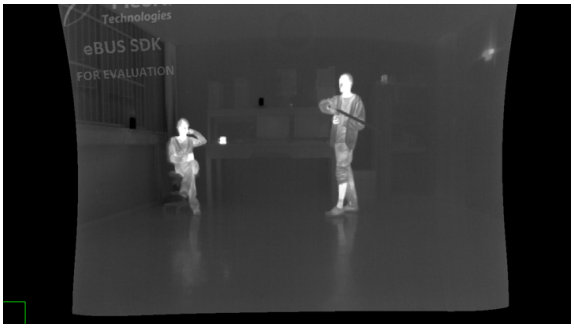


(e) TBHO fused image

Fig. 2. Dim Scenario

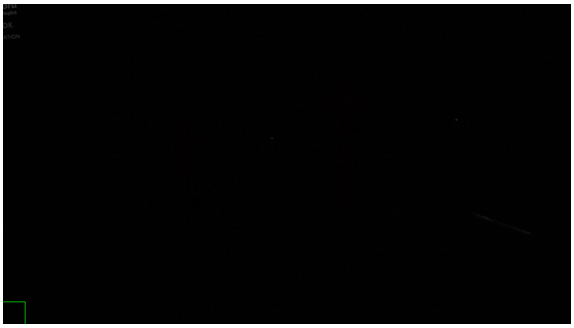
TABLE II
DIM SCENARIO ALGORITHM RANKINGS

Metric	Consistency Ratio (%)	Eigenvector				
		IR	Vis	TTGS	Fuzzy	TBHO
Mean	2.58	0.29	0.05	0.06	0.25	0.36
Mean $(CR < 20)$	2.02	0.27	0.04	0.06	0.24	0.39
Votes		274.1	-378.2	-361.0	82.4	382.6



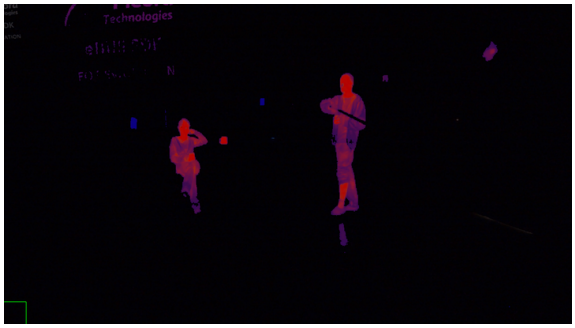
(a) LWIR image

(b) Visual image



(c) TTGS fused image

(d) Fuzzy fused image



(e) RWHS fused image

Fig. 3. Dark Scenario

TABLE III
DARK SCENARIO ALGORITHM RANKINGS

Metric	Consistency Ratio (%)	Eigenvector				
		IR	Vis	TTGS	Fuzzy	TBHO
Mean	6.13	0.44	0.04	0.04	0.17	0.30
Mean $(CR < 20)$	5.14	0.42	0.04	0.04	0.18	0.31
Votes		701.7	-418.5	-421.7	-102.0	240.5

$$P'_{RGB} = \begin{cases} f_{HtoR}\left({}^{T}P_{HSV_C}\right) & \text{if } \mathfrak{T} \in [\alpha_{C1}, \alpha_{C2}) \\ f_{HtoR}\left({}^{T}P_{HSV_H}\right) & \text{if } \mathfrak{T} \in (\alpha_{H1}, \alpha_{H2}] \\ P_{RGB} & \text{otherwise} \end{cases} \quad (5)$$

where:

$${}^{T}P_{HSV_H} = f_{RtoH}\left({}^{T}P_{RGB}\right) - \begin{bmatrix} 10^\circ \\ 0 \\ 0 \end{bmatrix}, \text{ and}$$

$${}^{T}P_{HSV_C} = f_{RtoH}\left({}^{T}P_{RGB}\right) + \begin{bmatrix} 8^\circ \\ 0 \\ 0 \end{bmatrix}$$

P'_{RGB} = the output RGB pixel,

f_{HtoR} = the function to transform HSV to RGB,

f_{RtoH} = the function to transform RGB to HSV, and

$${}^{T}P_{RGB} = \begin{bmatrix} \kappa_1 P_{RGB}.R + (1-\kappa_1)\,\mathfrak{T} \\ \kappa_2 P_{RGB}.G + (1-\kappa_2)\,\mathfrak{T} \\ \kappa_3 P_{RGB}.B + (1-\kappa_3)\,\mathfrak{T} \end{bmatrix}.$$

III. Analytical Hierarchy Process

The AHP [7], developed by Saaty in the 1970s, is a structured method to perform complex decision analysis. It breaks down a complicated problem into simpler sub-problems, which are then assessed by a panel. In this work the decision is which of the fusion methods, or raw video inputs provides the best situational awareness for three given scenarios.

All possible pair-wise comparisons of the five fusion algorithms (three fusions techniques presented in §II plus the raw LWIR and raw visual images) were presented in a randomised order. For each pairing each evaluator selected which algorithm was better and indicated how much better that algorithm performed using a scale from one (the algorithms perform equally well) through nine (the chosen algorithm is extremely superior). Where more than one evaluator was present no discussion was allowed.

This information was used to generate an $N \times N$ matrix for each evaluator with elements as described by Eq. 6.

$$a_{i,j} = \begin{cases} 1 & \text{if } i = j \\ R & \text{if Alg } i \text{ chosen over Alg } j \\ R^{-1} & \text{if Alg } j \text{ chosen over Alg } i \end{cases} \quad (6)$$

where:

N = the number of options being compared,

$i, j \in [1, 2, \ldots N]$, and

R = the evaluator's rating from 1 to 9.

The dominant eigenpair of this matrix is then sought and the eigenvector normalised such that its sum of elements is unity. According to Saaty [7] a Consistency Ratio (CR) from the eigenvalue (λ) and the Random Consistency Index (RCI) can be calculated as per Eq. 7.

TABLE IV
Random Consistency Index

N	1	2	3	4	5
RCI	0.0	0.0	0.58	0.9	1.12
N	6	7	8	9	10
RCI	1.24	1.32	1.41	1.45	1.49

$$CR = \frac{100.0 \times (\lambda - N)}{RCI(N) \times (N-1)} \quad (7)$$

where:

CR = the consistency ratio,

λ = the dominant eigenvalue,

N = the number of items being compared, and

$RCI(N)$ = is as per Table IV.

The ideal CR is 0%, indicating that the panel member had no contradictory selections and rankings, Saaty further states that a CR of 20% or higher may result in anomalous findings if used. The eigenvector meanwhile provides the percentage preference of the panel member for each of the items being considered. AHP thus provides both the preferences of an evaluator as well as the credibility of the assessment.

A. New Voting Methodology

Once the evaluation of each panel member's individual algorithm preferences has been determined as described in §III, the question arises as to how all of these assessments should be combined to yield the panel's verdict. The standard method of doing this is to create the geometric mean decision matrix, and obtain the dominant eigenpair of this matrix to use as the aggregate. In this work the global geometric mean eigenpair, the geometric mean eigenpair of evaluators with a CR less than 20% and a new vote casting method was used.

The vote casting method was created to remedy a flaw in the AHP process which causes people who rate algorithms as being of similar performance (i.e. ratings of 1 through 4) to have higher consistency. Whereby those expressing strong opinions between the presented algorithms tend to have a higher CR and thus may get excluded from the decision making process. This skews the algorithm in favour of the choices of people who do not really have a strong preference for any of the algorithms. So the following vote casting strategy was determined:

$$V_i^j = \begin{cases} 0 & \text{if } CR > 100.0 \\ (100.0 - CR) \times (\vec{v}_i^{\,j} - \frac{1}{N}) & \text{if } CR \leqslant 100.0 \end{cases} \quad (8)$$

where:

V_i^j = the evaluator j's vote for algorithm i,

$\vec{v}_i^{\,j}$ = the i^{th} element of the evaluator j's eigenvector,

N = the number of algorithms, and

CR = the evaluator j's consistency ratio as per Eq. 7.

TABLE V
VALUES OF FACTORS USED IN ALGORITHMS

Scene	Algorithm	α_{H1}	α_{H2}	α_{C1}	α_{C2}	κ_1	κ_2	κ_3
Dark and Dim	TTGS	0.55	1.0	0.0	0.45			
	Fuzzy	0.4	1.0	0.0	0.35	0.7	0.1	0.5
	TBHO	0.55	1.0	0.0	0.45	0.3	0.9	0.5
Outside	TTGS	0.65	1.0	0.0	0.1			
	Fuzzy	0.6	1.0	0.0	0.075	0.7	0.1	0.5
	TBHO	0.65	1.0	0.0	0.1	0.3	0.9	0.5

Eq. 8 causes each person to vote for the algorithms that they prefer and negatively vote for algorithms they dislike. The strength of a vote is proportional to both the evaluator's consistency and the strength of their preference for the algorithm. The total of each person's votes is always zero. The total vote for each algorithm is merely the sum of each evaluator's vote for that algorithm.

IV. SCENARIOS

This section describes the three scenarios presented to a total of 32 evaluators. For each scenario all possible pairs were presented in a random order and with random left/right assignments to help prevent evaluator's consciously preferring a specific algorithm. The same question was asked of the evaluators when considering each scenario: Which of the two current algorithms provides the best combination of potential threat detection and then identification of that threat?

In all the images a small Pleora Software Development Kit (SDK) [10] watermark is evident in the top left corner of the visual images and a larger watermark in the LWIR images. This was due to the laptop used to capture the images only having the free non-licensed version of the SDK.

The visual camera used was a Sony FCB-SE600 Colour Block Camera with a 1920×1080 resolution and maximum horizontal field of view (FOV) of $78.5°$. The SE600 was fitted with a Stemmer GigE Vision backpack. The LWIR camera used was a Xenics Gobi 640 GigE microbolometer with an approximate horizontal FOV of $60°$ and a resolution of 640×480.

A. Bright

The first scenario is a bright, well illuminated outdoor scene. There is a large amount of colour detail in the visual images including both red and blue vehicles. Much of the scene was already hot due to solar heating effects. Figure 1 shows the five images presented.

B. Dim

The second scene was an indoor scene, this was so the lighting could be controlled to ensure it was dim. The amount of colour information throughout the scene varied depending on proximity to the open door. Both hot and cold items were placed in the scene along with two people, one wielding a weapon. Figure 2 provides the 5 images shown pairwise to the evaluators.

TABLE VI
CONSISTENCY RATIO STATISTICS

Scenario Name	Consistency Ratio			
	Min	Mean	Max	St Dev.
Bright	3.71	21.10	84.78	18.74
Dim	3.09	22.20	143.51	27.94
Dark	3.49	15.89	128.71	21.60

C. Dark

The final scenario presented to the evaluator's was a dark scene. Negligible visual/colour information is evident (see Figure 3(b)) due to the lighting conditions. Structurally, the scene is almost identical to that described in Section IV-B. Figure 3 shows the images for the dark scenario.

V. RESULTS

The results of the aggregation of the individual evaluator's votes are given in Tables I, II and III for the bright, dim and dark scenes respectively. In these tables the highest scoring algorithm for each scene and aggregation method is indicated. Table VI provides the evaluator's consistency ratio statics for each scenario. Note that the CR of the geometric means for both the bright and dim datasets is lower than the minimum individual CR for those datasets and the dark dataset's CR is only marginally worse than the minimum. This was why the voting aggregation method was developed. The ranking of algorithms provided by the voting method is almost identical to that of the geometric mean based ranking, only the third and fourth ranked positions are swapped around for the bright dataset. The voting method seems to provide a better indication of the relative preferences of the algorithms. The voting method also indicates whether the algorithm was beneficial (i.e. had a positive total of votes) or detrimental (negative vote tally) for each scenario.

It can be seen that where usable information is provided by the both cameras then a fused false-colour image is preferred. In the dark scenario there was no visual information and evaluators strongly preferred the raw LWIR image, the second most favoured algorithm was TBHO algorithm.

For the bright data there was a strong preference (excluding the global geometric mean) for TTGS, which discards the visual colour information and uses red or blue tinting based on the thermal information. Close second choices were the 'Fuzzy' and TBHO algorithms.

VI. Conclusions

This paper developed three fusion algorithms and evaluated them and the raw imagery in three different scenarios using a panel of evaluators. AHP was used to perform the evaluation and a new method of aggregating the AHP results was suggested which provides an improved discrimination of preferences between algorithms.

The TBHO algorithm is the overall highest ranked fusion algorithm across the thee scenarios and is recommended for situations where only one fusion algorithm may be implemented. Implementations capable of allowing several algorithms should provide the raw LWIR, TBHO and TTGS algorithms.

References

[1] J. Cronje and J. P. de Villiers, "A comparison of image features for registering LWIR and visual images," in *Proceedings of the 23nd Annual Symposium of the Pattern Recognition Society of South Africa*, ser. PRASA2012, 2012.

[2] J. P. de Villiers and J. Cronje, "A method of calibrating a camera and a system therefor," International Patent PCT/IB2012/056 820, 11 30, 2012.

[3] ——, "Improved real-time photogrammetric stitching," in *Proc. SPIE 8744, Automatic Target Recognition XXIII*, vol. 8744, 2013, pp. 874 406–874 406–9.

[4] T. T. Zin, H. Takahashi, T. Toriu, and H. Hama, "Fusion of infrared and visible images for robust person detection," in *Image Fusion*, O. Ukimara, Ed. Intechopen, 2011.

[5] Y. Zheng, "An exploration of color fusion with multispectral images for night vision enhancement," in *Image Fusion and Its Applications*, O. Ukimara, Ed. Intechopen, 2011.

[6] G. Li, S. Xu, and X. Zhao, "Fast color-transfer-based image fusion method for merging infrared and visible images," pp. 77 100S–77 100S–12.

[7] T. L. Saaty, *The Analytical Hierarchy Process (AHP)*. New York, USA: McGraw Hill, 1980.

[8] J. Baumbach, "Psychophysics of human vision and camouflage pattern design," 2010.

[9] R. C. Gonzalez and R. E. Woods, *Digital image processing*. Addison-Wesley Publishing Company, 2002.

[10] P. T. Inc, *eBus SDK*, 2013, http://www.pleora.com/support-center/documentation-and-downloads/79.