# REGISTRATION AND FUSION OF MULTI-SPECTRAL IMAGES USING A NOVEL EDGE DESCRIPTOR

*Nati Ofir, Shai Silberstein, Dani Rozenbaum, Yosi Keller, Sharon Duvdevani Bar*

Elbit Systems, Bar-Ilan University, Computer Vision and Algorithms Ltd.

## ABSTRACT

In this work we propose a fully end-to-end approach for multi-spectral image registration and fusion. Our fusion method combines images from different spectral channels into a single fused image using approaches for low and high frequency signals. A prerequisite of fusion is the geometric alignment between the spectral bands, commonly referred to as registration. Unfortunately, common methods for image registration of a single spectral channel might prove inaccurate on images from different modalities. For that end, we introduce a new algorithm for multi-spectral image registration, based on a novel edge descriptor of feature points. Our method achieves an accurate alignment allowing us to further fuse the images. It is experimentally shown to produce a high quality of multi-spectral image registration and fusion under challenging scenarios.

***Index Terms***— Multi-Spectral Imaging, Image Registration, Image Fusion

## 1. INTRODUCTION

This work addresses the problem of multi-spectral registration and fusion. Different spectral channels capture different scenes, therefore their fusion is interesting and their registration is challenging. In Figure 1 for example, the visible channel in wavelength $0.4 - 0.7\mu m$, captures the colors of the scene on the one hand, but suffers from low visibility due to haze on the other hand. On the right, the Middle-Wave-Infrared (MWIR) channel in wavelength of $3 - 5\mu m$, is complementary to the visible channel. It does not capture colors, but utilizes a good visibility even in the presence of haze. For an example of good visibility, see the objects in the far road, where "hot" cars can be easily seen in the thermal (MWIR) image. The fused image in the middle contains the advantages of both channels: good visibility of far distant objects combined with color information. In this work we introduce a method for performing registration of such images, that manages to align multi-spectral images even though their captured information differs significantly. In addition, we introduce a fast method for fusion of color-visible with IR-thermal images.

Image registration is a fundamental problem in computer



**Fig. 1**. Multi-spectral image registration and fusion. Left: RGB image of the visible channel $0.4 - 0.7\mu m$, contains the information about the color. Right: aligned grayscale MWIR images $3 - 5\mu m$, "sees" behind the haze, captures information about the cars in the far road. Center: fused image with the advantages of both channels, color information combines with good visibility of far objects.

vision with profound research. A classic way to align two images from the same modal [25, 27] is by Scale-Invariant-Feature-Transform (SIFT) [18]. Even though this method is robust, it might prove inefficient for images from different modalities. As the appearance of the same object, under different spectral wavelengths, varies. Thus, we propose to perform registration of multi-spectral images in a similar way, but to use instead a novel edge descriptor. Our method detects keypoints by Harris corner detection [11] and describes them using an edge descriptor. We quantify the keypoint similarities by correlation of descriptors and compute the final transformation by a novel derivation of Random-Sample-Consensus (RANSAC). The proposed scheme produces accurate registration results on multi-spectral images, since their edge information is indeed invariant to different wavelengths. Moreover, since measuring similarities by our edge descriptors is fast, our scheme can be used in real-time vision systems.

The paper is organized as follows. In Section 2 we cover previous work on the topics of registration and fusion. In Section 3 we introduce our method, for registration of multi-spectral images, based on our new edge descriptor. Then, in Section 4, we explain how to fuse two registered images: a color image of the visible channel, and a MWIR gray-scale image. In Section 5, we demonstrate evaluations of the accuracy of our registration algorithm, as well as examples of fusion of images captured from different sensors.

## 2. PREVIOUS WORK

Image registration is a fundamental problem in computer vision which has been studied for decades. Image registration methods [3, 27] are used to fuse images, for matching stereo to recover object shape, or to produce wide panorama images. Early methods rely on primitive characteristics of the images, such as, schemes that find translation by correlation [22] of image pixels, [19] solves registration based on the spatial intensity gradient of the image. Reddy et al. [23] estimate translation, rotation and scale by and FFT-based technique. Other schemes rely on key-points [11, 18] and invariant descriptors to estimate feature-based registration [25].

One group of works tried to solve the problem of image registration on images from different modalities or spectral channels [7, 24]. In [20] registration was carried out by two medical images based on maximizing the mutual information. Since contours and gradients are invariant to different spectral images, [4, 14, 15, 16] offer to utilize them for registration. When trying to solve only translation, measuring correlation on Sobel image [10], or Canny image [6] may help to register multi-modal images. We show experimentally that our method outperforms these approaches. Note that we align images of a wide spectral range, visible to MWIR, while other methods [4, 15] work on a narrower range, visible to Near-Infrared (NIR). When spectral ranges are closer, the captured images are more similar and therefore easier to align. Aguilera et al. [1] solves registration by FAST features [26] and unique descriptors for non linear intensity variations. Aguilera et al. [2] measures cross spectral similarity using a Convolutional-Neural-Network (CNN). They achieve high accuracy in classification of patches as same or different, however, their method cannot be incorporated into a complete registration scheme due to its insensitivity to slight geometrical transformations. Our approach utilizes the fact that edges are invariant to multi-spectral images, and use it within a feature-based registration. Edge descriptors are effective and relevant for our purpose as they are sensitive to geometric transformation and their computation is fast. In Section 3 we detail our novel edge descriptor and how to use it for multi-spectral registration.

The fusion of images retrieved from different sensors is a fundamental image processing problem addressed by many previous works [13, 8]. The most basic group of fusion methods, named $\alpha$ blending, is formulated as a linear combination, $\alpha I_1 + (1 - \alpha) I_2$, where $I_1, I_2$ are the input images. Early methods of image fusion are based on wavelets transform [17], or on Laplacian pyramid of an image [5]. More advanced approaches rely on Principal-Component-Analysis (PCA) transform of the fused images [21], and on Intensity-Hue-Saturation (IHS) technique [12]. In this work, we introduce a new technique named High-Pass-Low-Pass (HPLP). This method applies different fusion techniques to the low frequency signals and the high frequencies, as detailed in Section 4.

## 3. MULTI-SPECTRAL IMAGE REGISTRATION

The proposed scheme for multi-spectral image registration consists of three steps: corners detection by Harris [11], feature matching by a novel edge descriptor and iterative Random-Sample-Consensus (RANSAC) [9] for outliers rejection.

**Corners Detection.** Our multi-spectral registration scheme is based on feature-points, whereas each point is a Harris corner [11]. This method is useful for our purpose, since the groups of corners of the two spectral images have a significant overlap. We apply the following algorithm to detect the corners in each spectral image.

Let $I$ be the input image, $I_x, I_y$ be the horizontal and vertical derivatives of $I$ respectively. Given a Gaussian window $w$ compute the Harris matrix for every pixel:

$$A = \sum_u \sum_v w(u,v) \begin{pmatrix} I_x(u,v)^2 & I_x(u,v)I_y(u,v) \\ I_x(u,v)I_y(u,v) & I_x(u,v)^2 \end{pmatrix} \tag{1}$$

A corner is characterized by a large positive eigenvalues $\lambda_1, \lambda_2$ of the matrix $A$. Since the computation of exact eigenvalues is expensive, we compute the corner score: $S = \lambda_1 \lambda_2 - k(\lambda_1 + \lambda_2)^2 = \det(A) - k \cdot trace^2(A)$, where $k$ is a sensitivity parameter. For every pixel $x$ we compute the score $S$ to obtain a Harris score image $S(x)$. In practice, we apply a non-maximal-suppression using a window of size $w_1 \times w_1$, such that every corner pixel, is a local maximum in its surrounding window in $S(x)$.

**Feature Matching by an Edge Descriptor.** We encode every corner point in one of the multi-spectral images by a novel edge descriptor. The uniqueness of this descriptor is its invariance over different wavelengths. Note that although the same object is acquired differently in the visible and MWIR channels, it has high similarity after applying our edge descriptor.

Denote by $C_V$ the set of corners in the visible image and by $C_{IR}$ the corners in the MWIR image. Every point $p \in C_V$ or $p \in C_{IR}$ is basically described by its center pixel $(x_p, y_p)$. In addition, in a surrounding window of size $w_2 \times w_2$, we compute the Canny [6] edge image $E_p$ and the Gradient directions $G_p$. Note that $E_p$ is a binary image, $G_p$ is an image with indexes of directions. The angular $[0, 2\pi]$ domain is quantized in $G_p$ to $k_1$ bins. Thus, the descriptor $D_p$ of a point contains the following information: $D_p = \{x_p, y_p, E_p, G_p\}$.

Given two points, $p \in C_V$ and $q \in C_{IR}$, they are matched using their descriptors $D_p, D_q$. We first require that the difference in the gradient direction, will be less or equal to a single bin in any pixel $x$: $SameGrad_{p,q}(x) = |G_p(x) - G_q(x)| \mod 16 \leq 1$. Hence, the similarity measure between $p$ and $q$ is the normalized correlation of the pixels that are in the

same Gradient directions:

$$Similarity(p,q) = \frac{\sum_x E_p(x) \wedge E_q(x) \wedge SameGrad_{p,q}(x)}{\sqrt{\sum_x E_q(x)}}. \tag{2}$$

Then, for every corner $p \in C_V$, search the corners $q \in C_{IR}$ for the match with the highest similarity. We normalize $Similarity(p,q)$ by $\sqrt{\sum_x E_q(x)}$ to avoid high scores in cases of many edge pixels in $E_q$.

**Iterative RANSAC.** We compute the final transformation between the images using an iterative version of RANSAC [9]. For that we apply three iterations, the first is applying RANSAC to the regular axis. We compute for each corner $p \in C_V$ its best match $q^* \in C_{IR}$, and get a group of matches $M_1$. This group consists of inlier and outlier matches, therefore we apply the RANSAC outliers rejection approach. We sample small groups of matches $m_1, ..., m_l, ... \in M_1$ and compute by least square the transformation $T_l$ that best corresponds to these matches for every subgroup $m_l$. Each match induces two linear constraints and therefore if our transformation $T$ is characterized by $n$ parameters, $|m_l| = \lceil \frac{n}{2} \rceil$. Finally, we select the best transformation $T^*$ derived from the best group of matches $m^*$, that has the greatest support in the whole group of matches $M_1$. The support of a transformation is the number of other matches that "agree" with it, a match "agree" with a transformation if

$$||T_{2\times3} \begin{pmatrix} x_p \\ y_p \\ 1 \end{pmatrix} - \begin{pmatrix} x_{q^*} \\ y_{q^*} \\ 1 \end{pmatrix}||_2 \leq RansacDistance. \tag{3}$$

The $RansacDistance$ in the first iteration is $rd_1$. We denote by $T_1$ the transformation that is found by RANSAC in the first iteration.

In the second iteration we repeat the same process as in the first, using $T_1$ as an initial guess. We restrict the group of all matches in the second iteration $M_2$ as follows. $p \in C_V$ can be matched to $q \in C_{IR}$ only if their distance under $T_1$ (similar to Eq. (3)), is less than $MatchDistance$. In the second iteration, $MatchDistance = md_1$ and $RansacDistance = rd_1$ such that $md_1 < \infty$. We denote by $T_2$ the transformation found by our iterative RANSAC in the second iteration.

In the third and final iteration, we use $T_2$ as an initial guess and decrease the thresholds, $RansacDistance = rd_2$ and $MatchDistance = md_2$ such that $rd_2 < rd_1, md_2 < md_1$. As these thresholds are relatively low, the final transformation is more accurate. Finally, we denote by $T_3$ the transformation found by RANSAC in the third iteration. This is the final transformation of our multi-spectral registration algorithm. Note that we can search for different types of transformations between the input images.

## 4. FAST IMAGE FUSION

Given two registered images, a visible channel image $V$ and a MWIR image $IR$, we would like to obtain their colored fu-

sion image $F_c$, whose luminance gray-scale image is $F$. Since $V$ is an RGB image, we denote by $Y(V)$ its luminance channel and by $R(V), G(V), B(V)$ its color channels. In this section we introduce a new fusion technique denoted High-Pass-Low-Pass (HPLP). This method fuses $Y(V)$ with $IR$ using a different approach for the low-frequency signals in the image than for the high-frequency signals. The proposed fusion method is fast and can be run in real-time algorithms easily.

Given an image $I$ and a Gaussian kernel $g_\sigma$ with standard deviation of $\sigma$, we divide the image to high and low frequencies by convolution, $LP_\sigma(I) = I * g_\sigma$ for low pass, and then the high-pass is $HP_\sigma(I) = I - LP_\sigma(I)$. We fuse the low-frequencies by alpha-fusion,

$$LP_\sigma(F) = \alpha \cdot LP_\sigma(Y(V)) + (1 - \alpha) \cdot LP_\sigma(IR). \tag{4}$$

The high-frequency image is fused such that in each pixel we take the value of the channel with the highest absolute value. Denote by $b$ the indicator $b = 1_{|HP_\sigma(Y(V))| \geq |HP_\sigma(IR)|}$, then for any pixel,

$$HP_\sigma(F) = b \cdot HP_\sigma(Y(V)) + (1 - b) \cdot HP_\sigma(IR). \tag{5}$$

The fused image is given by, $F_\sigma = LP_\sigma(F) + gain \times HP_\sigma(F)$. We repeat the process for three levels of $\sigma$, $\sigma_1 = 3, \sigma_2 = 5, \sigma_3 = 7$ and fuse all the three with equal weights, $F = \frac{1}{3}(F_{\sigma_1} + F_{\sigma_2} + F_{\sigma_3})$.

We have a gray-scale fusion of $V$ and $IR$, we aim to compute the color fusion $F_c$ and apply color restoration based on $V$. Aiming to preserve the ratio, $\frac{F}{Y(V)} = \frac{R(F_c)}{R(V)} = \frac{G(F_c)}{G(V)} = \frac{B(F_c)}{B(V)}$. The final color fusion is this ratio times the input visible image, $F_c = \frac{F}{Y(V)} \cdot V$.

## 5. EXPERIMENTS

We experimentally verified the proposed scheme using multiple real images captured by multi-spectral imaging systems. The transformation between the channels is unknown in advance and we found it automatically using our method. We evaluated the accuracy of our approach by creating a dataset of manually-aligned images, and then finding a solution to a known simulated transformation. We tested our method on this dataset and compared it to several approaches for solving multi-spectral registrations. The proposed scheme was implemented in C++ and it runs in less than 4 seconds on a CPU. By utilizing parallel computing on a GPU its runtime is significantly reduced to 0.5 seconds. The proposed fusion scheme is implemented in Matlab and its runtime is negligible.

In Table 1 we compare between our accuracy to those of correlation of Canny [6] images, correlation of Sobel [10] images, maximization of mutual-information [20]. We also compared to [1] but this approach failed when applied to our dataset. We simulated a known translation between the channels and tested the various methods when trying to solve

| Algorithm | VIS-SWIR | VIS-MWIR |
|---|---|---|
| Our method | **0.62** | **0.76** |
| Canny | 2.13 | 1.43 |
| Sobel | 3.84 | 3.2 |
| Mutual Information | 1.38 | 2.48 |
| LGHD | 24.1 | 8.13 |

**Table 1**. Accuracy in pixels of multi-spectral registration when searching for translation only. Our method is compared to correlation of Canny, correlation of Sobel, maximization of mutual-information and LGHD. As can be seen (in bold), our approach achieves the highest accuracy and it is the only one which is sub-pixel accurate (score is less than 1).
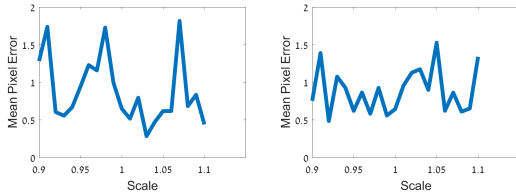


**Fig. 2**. Accuracy in pixels of multi-spectral registration as a function of scale between the images. Left: visible to SWIR registration. Right: visible to MWIR registration. The accuracy is around 1 pixel for all scales between 0.9 to 1.1 in both graphs.

translation only. Eventually, we find the Euclidean distance between the known translation to the output of each method to derive accuracy in pixels. We repeat this test for several images and store the mean of accuracy of all iterations. We test the performance for visible to Short-Wave-Infrared (SWIR, $1.4 - 3\mu m$) registration, and for visible to MWIR registration. It follows that in both wavelengths we achieve the most accurate results among all the compared approaches. Moreover, our algorithm is the only one that produces sub-pixel registration accuracy.

We evaluate the proposed scheme accuracy on different scales between two spectral channels in Figure 2. We report the Euclidean distance between the estimated translation to the simulated one. The same distance in the scale parameter is less than $0.001$ and therefore negligible. Our accuracy is $\sim 1$ pixel, in both SWIR and MWIR. Moreover, we achieve this accuracy level along all simulated scale transformations from 0.9 to 1.1.

Figure 3 shows the registration of visible and MWIR channels. We fuse the aligned images by the HPLP scheme. Where the left column depicts all the information about colors of the captured objects, while the right column depicts all the thermal information of the objects. The image of the "hot" factory building demonstrates how the two channels are naturally different. This object is very hot in the MWIR image and yet very colorful in the visible image. In the fused image



**Fig. 3**. Example of a multi-spectral image registration and fusion. Left: RGB image of the visible channel $0.4 - 0.7\mu m$, contains color information. Right: aligned grayscale MWIR images $3 - 5\mu m$ captures thermal information of objects. Center: fusion image with the advantages of both channels.



**Fig. 4**. Example of HPLP fusion captured by an aerial platform. From left to right: visible image, regular HPLP fusion and MWIR image.

we can see its colors together with its brightness indicating a high temperature.

In Figure 4 we show images captured by an aerial platform. The visible image captures the color of the clouds, but cannot depict far objects, whereas the MWIR image captures the far land. As can be seen, the proposed fusion conveys the interesting information captured in both channels.

## 6. CONCLUSIONS

We introduced a novel method to register images captured by multi-spectral sensors. Our approach is feature based and utilizes the robustness of edge descriptors to different wavelengths. The multi-spectral registration is accurate to allow the fusion of images. To that end we developed a new fusion scheme, named High-Pass-Low-Pass, that fuses the images differently for low and high frequencies. Our method outputs informative fusion results that utilize the benefits of both spectral channels.

## 7. REFERENCES

[1] C. Aguilera, A. D. Sappa, and R. Toledo. Lghd: A feature descriptor for matching across non-linear intensity variations. In *Image Processing (ICIP), 2015 IEEE International Conference on*, page 5. IEEE, Sep 2015.

[2] C. A. Aguilera, F. J. Aguilera, A. D. Sappa, and R. Toledo. Learning cross-spectral similarity measures with deep convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 1–9, 2016.

[3] L. G. Brown. A survey of image registration techniques. *ACM computing surveys (CSUR)*, 24(4):325–376, 1992.

[4] M. Brown and S. Süsstrunk. Multispectral SIFT for scene category recognition. In *Computer Vision and Pattern Recognition (CVPR11)*, pages 177–184, Colorado Springs, June 2011.

[5] P. Burt and E. Adelson. The laplacian pyramid as a compact image code. *IEEE Transactions on communications*, 31(4):532–540, 1983.

[6] J. Canny. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, (6):679–698, 1986.

[7] C. Chen, Y. Li, W. Liu, and J. Huang. Sirf: simultaneous satellite image registration and fusion in a unified framework. *IEEE Transactions on Image Processing*, 24(11):4213–4224, 2015.

[8] L. Dong, Q. Yang, H. Wu, H. Xiao, and M. Xu. High quality multi-spectral and panchromatic image fusion technologies based on curvelet transform. *Neurocomputing*, 159:268–274, 2015.

[9] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[10] W. Gao, X. Zhang, L. Yang, and H. Liu. An improved sobel edge detection. In *Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on*, volume 5, pages 67–71. IEEE, 2010.

[11] C. Harris and M. Stephens. A combined corner and edge detector. In *Alvey vision conference*, volume 15, pages 10–5244. Manchester, UK, 1988.

[12] C. He, Q. Liu, H. Li, and H. Wang. Multimodal medical image fusion based on ihs and pca. *Procedia Engineering*, 7:280–285, 2010.

[13] B. Huang, H. Song, H. Cui, J. Peng, and Z. Xu. Spatial and spectral image fusion using sparse matrix factorization. *IEEE Transactions on Geoscience and Remote Sensing*, 52(3):1693–1704, 2014.

[14] M. Irani and P. Anandan. Robust multi-sensor image alignment. In *Computer Vision, 1998. Sixth International Conference on*, pages 959–966. IEEE, 1998.

[15] Y. Keller and A. Averbuch. Multisensor image registration via implicit similarity. *IEEE transactions on pattern analysis and machine intelligence*, 28(5):794–801, 2006.

[16] H. Li, B. Manjunath, and S. K. Mitra. A contour-based approach to multisensor image registration. *IEEE transactions on image processing*, 4(3):320–334, 1995.

[17] H. Li, B. Manjunath, and S. K. Mitra. Multisensor image fusion using the wavelet transform. *Graphical models and image processing*, 57(3):235–245, 1995.

[18] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60:91–110, 2004.

[19] B. D. Lucas, T. Kanade, et al. An iterative image registration technique with an application to stereo vision. 1981.

[20] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE transactions on medical imaging*, 16(2):187–198, 1997.

[21] U. Patil and U. Mudengudi. Image fusion using hierarchical pca. In *image Information Processing (ICIIP), 2011 International Conference on*, pages 1–6. IEEE, 2011.

[22] W. K. Pratt. Correlation techniques of image registration. *IEEE transactions on Aerospace and Electronic Systems*, (3):353–358, 1974.

[23] B. S. Reddy and B. N. Chatterji. An fft-based technique for translation, rotation, and scale-invariant image registration. *IEEE transactions on image processing*, 5(8):1266–1271, 1996.

[24] X. Shen, L. Xu, Q. Zhang, and J. Jia. Multi-modal and multi-spectral registration for natural images. In *European Conference on Computer Vision*, pages 309–324. Springer, 2014.

[25] M. Subramanyam et al. Automatic feature based image registration using sift algorithm. In *Computing Communication & Networking Technologies (ICCCNT), 2012 Third International Conference on*, pages 1–5. IEEE, 2012.

[26] G. Takacs, V. Chandrasekhar, S. Tsai, D. Chen, R. Grzeszczuk, and B. Girod. Unified real-time tracking and recognition with rotation-invariant fast features. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 934–941. IEEE, 2010.

[27] B. Zitova and J. Flusser. Image registration methods: a survey. *Image and vision computing*, 21(11):977–1000, 2003.