Linear Adaptive Infrared Image Fusion

Chuong T. Nguyen and Joseph P. Havlicek
School of Electrical and Computer Engineering
University of Oklahoma, OK, USA
chuong@ou.edu, joebob@ou.edu

Abstract—We introduce a new adaptive image fusion algorithm to address the decomposition level problem in the multiresolution fusion technique. The fusion scheme consists of two parallel processes which are linearly combined on a spatially adaptive basis: lowpass fusion to exploit the inherent target-background discrimination capability of infrared sensors and highpass fusion to capture texture and shape details leveraging the visible band imagery. The weighting coefficients for each process are determined by maximizing a no-reference fusion quality metric. We demonstrate the effectiveness of the proposed algorithm against a variety of practical infrared fusion tasks where it delivers tangible performance gains relative to several well-known state of the art fusion techniques.

Keywords-image fusion, SSIM, infrared fusion

I. INTRODUCTION

Image fusion is a technique to combine images captured by one or multiple sensors into a single output image. Such techniques play an important role in applications such as photography, night vision, and remote sensing where one exposure or one sensor can not achieve the desired goal. For example, in landscape photography, it is difficult to shoot a high depth-of-field scene and maintain front-to-back sharpness with one exposure even if a small aperture is used. As the aperture decreases, the image sharpness degrades because of optical diffraction. Nevertheless, an optimal front-to-back sharpness of any particular scene can be achieved by blending multiple shots with larger apertures and different focused regions. In surveillance applications, infrared sensors are often used together with visible sensors to enhance image quality of ambient scenes. Infrared sensors, on the other hand, can not capture object textures as well as visible sensors. Therefore, a fusion of infrared and visible sensors can produce a single image that retains textures while maximizing target-background discriminability. The fused image can then be used to improve tracking algorithms or to assist human observers.

While image fusion is a board subject, we restrict our attention to the fusion task where multiple co-boresighted sensors capture the same scene from identical or nearly identical optical axes. Image fusion algorithms can be divided into two categories: pixel-based approaches and region-based approaches [1]. In the pixel-based approaches, the multi-resolution (MR) fusion scheme is the most popular framework in the literature. Input images are first decomposed into multiple scales using, e.g., a Laplacian pyramid or a wavelet transform [2], [3]. In each scale, the output image is determined from the transformed input images on a pixel by pixel basis. The decision scheme is usually the max filter where the largest transformed coefficient is transferred to the output. The final fused image is then obtained by inverting the multiscale representation. However, the pixel-based fusion algorithms do not integrate spatial information in images and are sensitive to registration errors [1]. The region-based approaches address this problem by accounting for regions or objects through a segmentation of the input images prior to the fusion step [4], [5], [6]. The fusion algorithms then make decisions based on regions instead of pixel-wise. However, region-based algorithms depend on the accuracy of the segmentation and on the region merging steps, both of which are error prone.

Many successful fusion algorithms share a common theme. These algorithms operate in a transform domain and select important features of the source images such as edges and energy to transfer to the output image. Hence, the performance of these algorithms depends on the chosen image transformation. For instance, fused images produced by the translation invariant wavelet transform contain less distortions than those produced by the decimated wavelet transform [3]. Recently, there has been a growing interest in designing fusion algorithms by optimizing image quality metrics [7], [8]. In this paper, we pursue an optimization based linear adaptive fusion (LAF) algorithm for infrared image fusion. The LAF consists of two parallel processes, highpass fusion and lowpass fusion. The fused image is computed as a weighted linear combination of these two processes. The weighting coefficients are determined by maximizing the no-reference fusion quality index proposed by Piella and Heijmans [9]. We evaluate the effectiveness of the proposed algorithm against four major classes of image fusion quality metrics. The simulation results indicate that the LAF consistently delivers performance gains relative to state of the art fusion techniques.

II. BACKGROUND

Most recent fusion algorithms operate in a transform domain, e.g., multiresolution (MR) analysis [1]. In this scheme, each input image is decomposed into multiple subbands. For each subband, a fusion operator is then defined to select important features from the input subbands to transfer to the output subband. The fused image is then obtained by reconstructing from the output subbands.

One of the earliest MR schemes is the Laplacian pyramid (LP) proposed by Burt and Adelson [10]. Li, Manjunath, and Mitra [2] used the maximally decimated discrete wavelet transform (DWT) to improve the fusion performance. To overcome the translation invariance limitations of the DWT, Hill, Canagarajah, and Bull performed image fusion using the complex wavelet transform [3].

Most of the MR fusion methods are pixel-based, i.e., the fusion operator is applied on a pixel-by-pixel basis. These techniques do not use spatially coherent information in images such as edges and local feature statistics. Zhang and Blum [4] used edge detection and region labeling algorithms to find region of interests to create a fusion decision map. They then performed image fusion in the wavelet domain. Region identification is also a common theme

in subsequent region-based image fusion works by Lewis and O'Callaghan [11] and by Piella [5].

III. LINEAR ADAPTIVE FUSION

Infrared sensors can discriminate between targets and backgrounds (clutter) because targets such as humans and machines are usually warmer than the clutter. Visible sensor, however, is capable of capturing textures of the scene. Based on this observation, we preserve characteristics of infrared sensors using a lowpass fusion process. Scene details of the visible sensors such as textures and shapes are retained by a highpass fusion process. However, the decomposition into lowpass and highpass components varies for different image pairs. For example, a certain image pair may contain more textures than others. Hence, a fixed highpass/lowpass decomposition level for every image pairs is not an optimal choice. Here, we address the highpass/lowpass decomposition level selection using an image quality optimization approach.
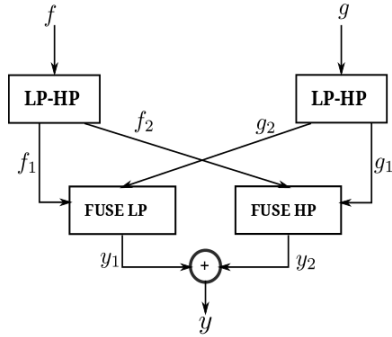

Figure 1.   Block diagram of LAF image fusion.

The proposed algorithm is illustrated in Fig. 1. We assume that the images f and g are captured by a visible sensor and an infrared sensor. First, f and g are decomposed into lowpass and highpass components. Let h be a Gaussian low-pass filter with adaptive bandwidth parameter σ. The spectrum of this lowpass filter is given by

$$\widehat{h}(\boldsymbol{\omega}, \beta) = \exp\left[-\frac{||\boldsymbol{\omega}||^2}{2\beta^2}\right]. \qquad (1)$$

In particular, the lowpass images are then obtained as $f_1 = f * h$ and $g_1 = g * h$. The highpass images are computed as $f_2 = f - f_1$ and $g_2 = g - g_1$.

The lowpass fusion decision is given by

$$y_1(\beta) = \frac{\bar{f}_1}{\bar{f}_1 + \bar{g}_1} f_1 + \frac{\bar{g}_1}{\bar{f}_1 + \bar{g}_1} g_1, \qquad (2)$$

where $\bar{f}_1$ and $\bar{g}_1$ denote the local mean of f_1 and g_1 computed in a local neighborhood, e.g., an 8-by-8 window.

The highpass fusion scheme is computed based on local variance

$$y_2(\beta) = \frac{\sigma_{f_2}^2}{\sigma_{f_2}^2 + \sigma_{g_2}^2} f_2 + \frac{\sigma_{g_2}^2}{\sigma_{f_2}^2 + \sigma_{g_2}^2} g_2, \qquad (3)$$

where $\sigma_{f_2}^2$ and $\sigma_{g_2}^2$ are the local variance of f_2 and g_2 computed in a neighborhood about each pixel, e.g., within an 8-by-8 window.

The fused image y is then obtained as a summation of (2) and (3). After some rearrangement, the fused image can be rewritten as

$$y(\beta) = \frac{h * \bar{f}}{h * (\bar{f} + \bar{g})} [h * (f - g)] +$$
$$\frac{h * \sigma_f^2}{h * (\sigma_f^2 + \sigma_g^2)} [(f - g) - h * (f - g)]. \qquad (4)$$

The fused image y in (4) is a function of β. We find the optimal value of β that maximizes the no-reference fusion quality index

$$\arg\max_{\beta} Q(f, g, y(\beta)), \qquad (5)$$

where $Q(f, g, y(\beta))$ is the quality measure of the fused image y given input images f and g. The quality index $Q(f, g, y(\beta))$ is given in [9] by

$$Q(f, g, y(\beta)) = \lambda Q(f, y(\beta)) + (1 - \lambda) Q(g, y(\beta)). \qquad (6)$$

where $Q(f, y(\beta))$ and $Q(g, y(\beta))$ are the full-reference structural similarity index (SSIM) [12]. The mixing coefficient λ is the weighted salient feature computed from the image pairs (f, y) and (g, y). Here, we use local variance as the salient feature. Once the bandwidth parameter β is obtained, the final fused image y is computed as in (4).

IV. RESULTS AND DISCUSSION

We computed LAF against twelve publicly available sets of images obtained from *www.imagefusion.org* or from our website with Matlab source code[1]. We compare the LAF results with seven other fusion algorithms: DWT (decimated discrete wavelet transform), SIDWT (shift invariant discrete wavelet transform), DT-CWT (dual-tree complex wavelet transform), LP (Laplacian pyramid), RL (ratio pyramid), GL (gradient pyramid), and AVG (average) [20]. The performance of these seven algorithms and LAF were quantified using four classes of quality metrics: information theoretic measures [13], [14], image feature measures [15], [16], structural similarity measures [9], [17], and perceptually motivated quality measures [18], [19]. For any given metric (except the Chen-Varshney (VC) metric [18]), a higher score means better performance. Mathematical descriptions of these quality metrics are found in the review by Liu et al. [20]. The number of decomposition levels for the MR method was set to 4. LAF used the default 8×8 window in the calculation of the SSIM index and the local variance feature.

We tabulate the fusion results for the Bristol set and the Kayak set in Table I and in Table II. Quality metrics are tabulated across rows and fusion algorithms are listed down each column. For each quality metric, we indicate the best fusion method with bold numbers. We note that none of the fusion methods wins in every test case. This observation is consistent with the detail study in [20]. However, LAF performs best with respect to most of the quality measures, especially with respect to the Piella metric [9] which LAF is designed for. In addition, we studied the overall performance of LAF for all of the test images. For each test set, we ranked the eight fusion algorithms from 1 to 8 based on quality scores. We then averaged the ranking scores over twelve test cases. In this ranking study, LAF delivers the best overall fusion

Table I
FUSION PERFORMANCE COMPARISON AGAINST THE BRISTOL SET.

Method	MI [13]	Wang [14]	Xydeas [15]	PWW [16]	Piella [9]	Cvejie [17]	VC [18]	CB [19]
DWT	0.5617	0.8118	0.5634	0.4110	0.5965	0.4582	1075.7109	0.3445
SIDWT	0.4023	0.8058	0.3987	0.3986	0.4737	0.5068	1065.9187	0.3195
DT-CWT	0.2963	0.8039	0.5314	0.4296	0.6351	0.5055	845.6816	0.4436
LP	0.3361	0.8048	0.6251	0.4641	0.6943	0.5284	**671.9634**	**0.4821**
RL	0.3242	0.8035	0.1635	0.2864	0.1018	0.3782	1648.4622	0.4530
GL	0.3832	0.8055	0.5608	0.4739	0.6387	**0.5289**	786.7205	0.4152
AVG	0.4348	0.8066	0.4185	0.3709	0.4564	0.5148	891.8675	0.4653
LAF	**0.5844**	**0.8128**	**0.6490**	**0.5417**	**0.6946**	0.4806	1004.9463	0.4795

Table II
FUSION PERFORMANCE COMPARISON AGAINST THE KAYAK SET.

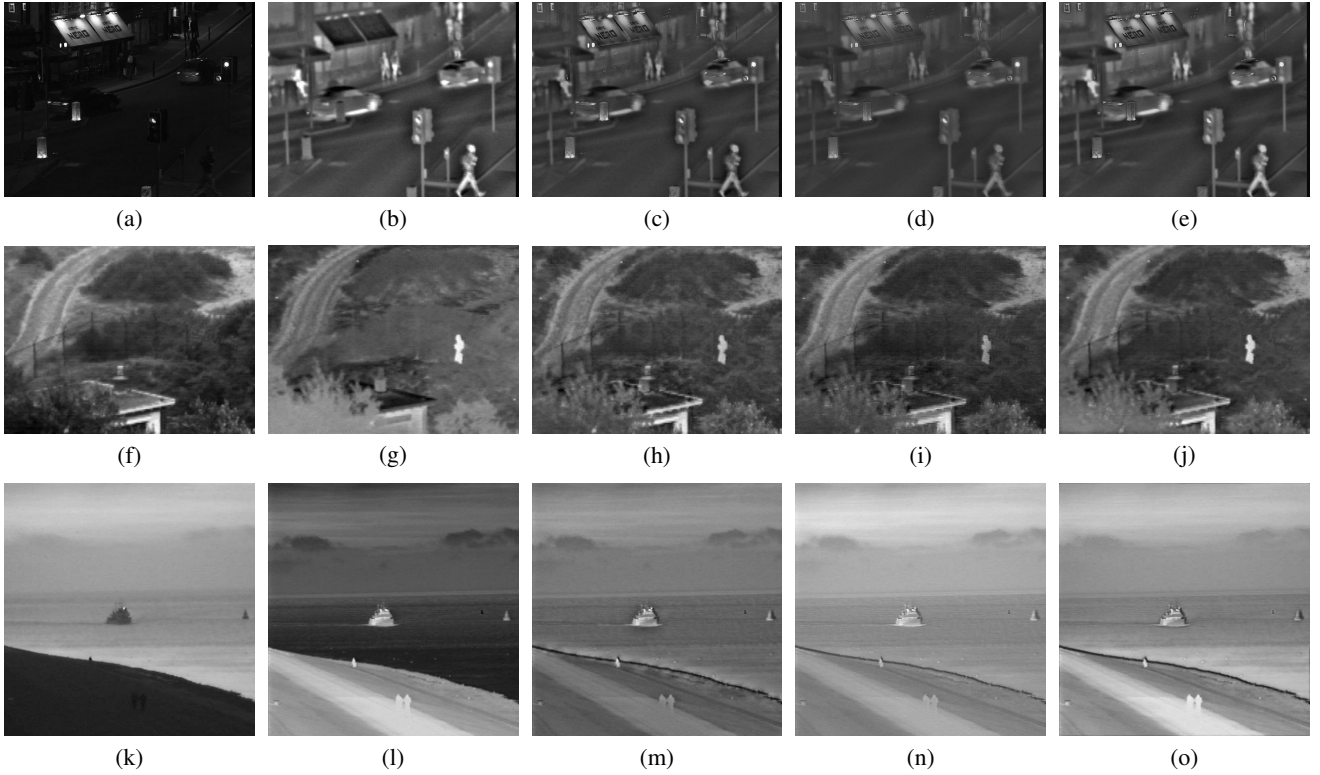| Method	MI [13]	Wang [14]	Xydeas [15]	PWW [16]	Piella [9]	Cvejie [17]	VC [18]	CB [19]
DWT	0.4376	0.8078	0.4753	0.3294	0.1233	0.4508	1681.1400	0.3500
SIDWT	0.5359	0.8099	0.5216	0.4212	0.1500	0.5602	969.4817	0.3507
DT-CWT	0.4225	0.8075	0.5050	0.3851	0.1345	0.5275	923.0406	0.3761
LP	0.4660	0.8083	0.5429	**0.5306**	0.1821	0.5760	1040.5504	0.4358
RL	0.4562	0.8079	0.3753	0.3786	0.1061	0.4841	1164.6833	**0.4815**
GL	0.5005	0.8091	0.5459	0.4877	0.1758	0.5727	988.2112	0.3830
AVG	0.6401	**0.8129**	0.5504	0.4980	0.1763	0.5997	976.6850	0.4216
LAF	**0.6405**	0.8120	**0.5560**	0.3447	**0.1967**	**0.6009**	**921.2385**	0.3667

Figure 2. Image fusion results of LP, SIDWT, and LAF for the Bristol, UNCamp, and Kayak test sets. Visible input images are shown in the first column. Infrared input images are depicted in the second column. LP, SIDWT, and LAF fusion results are shown in the third, fourth, and fifth columns, respectively.

performance. It is interesting to observe that in some test cases the simple average method performs better than other tested methods, including LAF, with respect to the MI measure. However, the MI measure computes image fusion quality using distributions which may not be appropriate for evaluating the quality of visual scenes.

In Fig. 2, we show the fused outputs of LAF and two other regularly benchmarked methods, LP and SIDWT [9], [20]. The first and second columns are input images taken from two different sensors, visible and infrared. The LP fusion outputs are given in the third column. The fourth column depicts the results of SIDWT. The last column contains the LAF fusion results. The test image sets are Bristol, UNCamp, and Kayak [21]. We observe that LAF retains the contrast of the input images better than other two methods, especially in important regions where human body is presented against the ambient clutter background as in Fig. 2(e).

Besides producing superior fusion results, LAF has the advantage of not requiring parameter tuning. However, LAF runs slower than other tested methods. For a 256x256 fusion task, the current Matlab implementation takes an average of 5 seconds on an Intel 2GHz single-core processor with 2GB of RAM.

V. CONCLUSIONS

We introduced a linear adaptive fusion (LAF) algorithm for heterogeneous sensors vision system. The fused image is computed as a weighted linear combination of visible and infrared input images. To take advantage of the unique operating characteristic of infrared sensors, we divide the fusion scheme into two parallel processes: lowpass fusion and highpass fusion. The weighting coefficients of each process are determined by maximizing the SSIM-based no-reference fusion quality measure. We demonstrated the effectiveness of the proposed algorithm quantitatively and qualitatively. The experimental results indicate that LAF consistently delivers superior performance relative to seven competing techniques. We are currently evaluating the performance of LAF under noisy and high distortion scenarios.

REFERENCES

[1] G. Piella, "A general framework for multiresolution image fusion: from pixels to regions," *Information Fusion.*, vol. 4, pp. 259–280, 2003.

[2] H. Li, B.S. Manjunath, and S.K. Mitra, "Multi-sensor image fusion using the wavelet transform," in *Proc. IEEE Int'l. Conf. Image Proc.*, Austin, TX, Nov. 13-16, 1994, vol. 1, pp. 51–55.

[3] P. Hill, N. Canagarajah, and D. Bull, "Image fusion using complex wavelets," in *British Machine Vision Conf.*, Sept. 2-5 2002, pp. 487–496.

[4] Z. Zhang and R. S. Blum, "A region based image fusion scheme for concealed weapon detection," in *Proc. Conf. Info. Sciences and Syst.*, Baltimore, MD, Mar., 1997, pp. 168–173.

[5] G. Piella, "A region-based multiresolution image fusion algorithm," in *Proc. IEEE Int'l. Conf. Info. Fusion*, July 2002, vol. 2, pp. 1557–1564.

[6] V. S. Petrović and C. S. Xydeas, "Gradient-based multiresolution image fusion," *IEEE Trans. Image Proc.*, vol. 13, no. 2, pp. 178–182, Feb. 2004.

[7] Y. Zheng, E.A. Essock, and B.C. Hansen, "Advanced discrete wavelet transform fusion algorithm and its optimization by using the metric of image quality index," *Optical Engineering*, vol. 44, no. 3, pp. 037003–1–037003–12, Mar. 2005.

[8] V. Petrovic and T. Cootes, "Objectively optimised multisensor image fusion," in *Proc. Int'l. Conf. Info. Fusion.*, Jul. 10-13 2006, pp. 1–7.

[9] G. Piella and H. Heijmans, "Image fusion using complex wavelets," in *Proc. IEEE Int'l. Conf. Image Proc.*, Barcelona, Spain, Sept. 14-18 2003, vol. 2, pp. 173–176.

[10] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image coding," *IEEE Trans. Commun.*, vol. COM-31, no. 4, pp. 532–540, Apr. 1983.

[11] J.J. Lewis, R. J. O'Callaghan, S.G. Nikolov, D. R. Bull, and C.N. Canagarajah, "Region-based image fusion using complex wavelets," in *Proc. IEEE Int'l. Conf. Info. Fusion*, Sweden, 2004, pp. 555–562.

[12] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Proc.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[13] M. Hossny, S. Nahavandi, and D. Vreighton, "Comments on 'information measure for performance of image fusion'," *Electronics Letters*, vol. 44, no. 18, pp. 1066–1067, Aug. 2008.

[14] Q. Wang and Y. Shen, "Performances evaluation of image fusion techniques based on nonlinear correlation measurement," in *Instru. Meas. Tech. Conf.*, Italy, May. 18-20, 2004, pp. 472–475.

[15] C. S. Xydeas and V. Petrovic, "Image fusion performance measure," *Elect. Letters.*, vol. 36, no. 4, pp. 308–309, Feb. 2000.

[16] Z. Liu, D.S. Forsyth, and R. Laganière, "A feature-based metric for the quantitative evaluation of pixel-level image fusion," *Computer Vision and Image Understanding*, vol. 109, no. 1, pp. 56–68, Jan. 2008.

[17] N. Cvejic, A. Loza, D. Bul, and N. Canagarajah, "A similarity metric for assessement of image fusion algorithms," *Int'l. J. Signal Proc.*, vol. 2, no. 3, pp. 178–182, 2005.

[18] H. Chen and P.K. Varshney, "A human perception inspired quality metric for image fusion based on regional information," *Information Fusion*, vol. 8, no. 2, pp. 193–207, Apr. 2007.

[19] Y. Chen and R.S. Blum, "A new automated quality assessment algorithm for image fusion," *Image and Vision Comput.*, vol. 27, no. 10, pp. 1421–1432, Sep. 2009.

[20] Z. Liu, E. Blasch, Z. Xue, J. Zhao, R. Laganiére, and W. Wu, "Objective assessement of multiresolution image fusion algorithms for context enhancement in night vision: A comparative study," *IEEE Trans. Pattern. Anal. and Machi. Intell.*, vol. 34, no. 1, pp. 94–109, Jan. 2012.

[21] J. J. Lewis, S. G. Nikolv, A. Toet, D. R. Bull, and C. N. Canagarajah, "Uni-modal versus joint segmentation for region-based image fusion," *Proc. Int'l. Conf. Info. Fusion.*, pp. 1–8, Jul. 2006, please cite me whenever.