

EXPLOITING FOVEATION IN USER-CENTRED IMAGE FUSION

*E. Basaeed**, *M. Al-Mualla*

Department of Computer Engineering
Khalifa University of Science,
Technology and Research
Sharjah, U.A.E
Email: {ebasaeed, almualla}@kustar.ac.ae

D. Vernon

Department of Robotics, Brain,
and Cognitive Sciences
Italian Institute of Technology
Genoa, Italy
Email: david@vernon.eu

ABSTRACT

The selection of image fusion techniques has always been a compromise between effectiveness and efficiency. In this paper, foveation (using log-polar transformation) is introduced to satisfy the real-time requirement of applications where the fused image is eventually presented to a human to interpret. Log-polar transformation achieves a data reduction without a subsequent loss in perceptual information. Thus, it can be used to reduce the execution time of any existing image fusion technique without modifying the technique itself. In this paper, the log-polar transformation is integrated with three widely-used image fusion techniques: image averaging, the Laplacian pyramid, and the Discrete Wavelet Transform. Then, the proposed fusion process is objectively evaluated using a set of established metrics; they are Mutual Information (MI), Xydas & Petrovic ($Q^{AB/F}$), and Piella & Heijmans (Q_w).

Index Terms— Data Fusion, Image Fusion, Pixel-level Image Fusion, Foveation, Log-polar Mapping.

1. INTRODUCTION

In recent years, research interest in data fusion has increased greatly. Within the field of data fusion, image fusion evolved as a distinct sub-area. Image fusion is defined as “the process of combining two or more different images into a single composite image with extended information content [1]”. Three conditions must be satisfied in this process [2]: it should 1) preserve all relevant information, 2) eliminate irrelevant information and noise, and 3) minimize artefacts and inconsistencies in the fused image. The fused image should be more suitable for further processing and/or visual interpretation [3]. Image fusion is used in the fields of, amongst others, remote sensing, security and law enforcement, military applications, industrial engineering, and robotics.

*The author would like to acknowledge the support of H.H. General Sheikh Mohammed Bin Zayed Al-Nahyan Program for Postgraduate Scholarships (Buhooh).

In our application, visible and infrared images are fused to facilitate tele-operation of a robot. In such an active vision system, continuous real-time operation is essential [4]. Hence, one should select a fusion system that is effective and efficient. However, fusion techniques available are either computationally simple but with poor quality or effective and computationally expensive. Therefore, the selection of a suitable fusion technique has always been a trade-off between fusion quality and computational complexity. The ultimate goal, though, is to find a way to fuse with a high quality yet with an acceptable delay. A common way of achieving this goal, especially in robotics, is by reducing the size of the input data-set. Data reduction yields also a subsequent reduction in processing power [5], communication bandwidth [6], and storage requirements. The idea, then, would be to reduce the size of input images and fuse them using existing fusion methods.

This paper is organized as follows. The next section investigates a data reduction technique called *foveation*. Section 3 introduces the foveation-based image fusion system. Results of a time analysis are presented in section 4. In the last section, the performance of some representative image fusion methods with/without log-polar transformation is discussed.

2. REDUCING DATA-SET SIZE USING FOVEATION

In applications where the output image is eventually presented to a human (as in our case), human factors can be exploited to reduce the data set size. The retina is known to have a high resolution in the fovea and its resolution smoothly decreases when moving away from the fovea towards the periphery [4]. In other words, it has a space-variant spatial resolution [6]. Hence, by matching the resolution of the input image to the resolution fall-off in the retina, the data set size can be considerably reduced while maintaining its perceptual value [6]. Images with a space-variant resolution are called foveated images. One widely used transformation of modelling the sampling pattern in the retina is the log-polar mapping.

Photoreceptors in the retina are non-uniformly distributed as they are dense in the fovea and become sparser as we

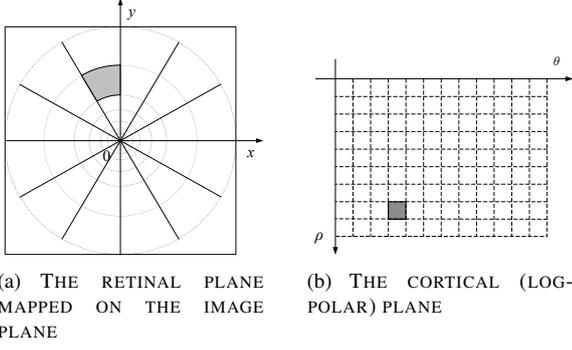


Figure 1. The log-polar mapping (based on [7]).

move away towards the periphery [8]. Moreover, they are located non-linearly in the retina with a radial symmetry [8]. The non-uniform distribution is introduced by the logarithmic scaling whereas the non-linearity is modelled by the polar mapping [9]. As such, log-polar transformation models the mapping of the retina into the visual cortex [5], hence the name the *retino-cortical* transform [9]. Log-polar is an excellent transformation as it results in a foveated image that has a smoothly varying spatial resolution without blocking artefacts [6]. In addition, it depicts in great detail the region where the focus of attention is (given that it is properly indicated) while it facilitates a better situational awareness with its wide field-of-view [10].

The retinal plane is modelled as a set of rings with logarithmically decreasing widths as we approach the center. However, with the logarithmic mapping, an infinite number of rings is required in order to cover the centre (or fovea). Since this is not physically realizable, linear mapping is used in the fovea whereas the logarithmic mapping is used in the periphery. To be more precise, the last ring in the fovea is the first ring in the logarithmic mapping. Each ring consists of a fixed number of sectors which model the receptive fields in the retina. Each of these sectors define the set of pixels in the image plane (see Figure 1(a)) that will determine the value of the corresponding log-polar pixel (refer to Figure 1(b)). Log-polar mapping is a transformation between the Cartesian domain (x, y) (image plane) and the log-polar plane (ρ, θ) (cortical plane). Equations (1) & (2) define the relationship between ρ and (x, y) for the fovea and the periphery, respectively [7]. The relationship between θ and (x, y) is defined in Eq. (3) which is the same for both the fovea and the periphery regions [7]. A conversion from the log-polar representation to the image plane is required, that is converting from (ρ, θ) back to (x, y) domain as defined in the following equations [7]. Equations (4) & (5) define the relationship in the periphery whereas the last two specify the relationship in the fovea.

$$\rho(x, y) = k \cdot \log_{\lambda} \frac{\sqrt{x^2 + y^2}}{r_0} \quad (1)$$

$$\rho(x, y) = k \cdot \sqrt{x^2 + y^2} \quad (2)$$

$$\theta(x, y) = \frac{\Theta}{2\pi} \arctan \frac{y}{x} \quad (3)$$

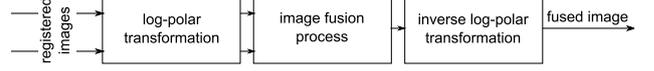


Figure 2. A general block diagram of an image fusion system exploiting log-polar transformation.

$$x(\rho, \theta) = r_0 \cdot \lambda^{\rho/k} \cos \frac{2\pi\theta}{\Theta} \quad (4)$$

$$y(\rho, \theta) = r_0 \cdot \lambda^{\rho/k} \sin \frac{2\pi\theta}{\Theta} \quad (5)$$

$$x(\rho, \theta) = \frac{\rho}{k} \cos \frac{2\pi\theta}{\Theta} \quad (6)$$

$$y(\rho, \theta) = \frac{\rho}{k} \sin \frac{2\pi\theta}{\Theta} \quad (7)$$

where Θ is the total number of sectors per ring and k is used to scale the mapping to cover the desired area. λ is the logarithm index which is the ratio between two adjacent rings in the logarithm mapping and r_0 represents the radius of the innermost circle in the logarithmic part; they are calculated by the following equations, respectively [11]:

$$r_0 = \frac{1}{\lambda^F(\lambda - 1)} \quad (8)$$

$$\lambda = \frac{1 + \sin \frac{\pi}{\Theta}}{1 - \sin \frac{\pi}{\Theta}} \quad (9)$$

where F represents the number of rings in the fovea region, that is defined as [11]:

$$F = \left\lfloor \frac{\lambda}{\lambda - 1} \right\rfloor \quad (10)$$

Foveated images have been proven effective in a number of robotic applications in such processing power and communication bandwidth are limited resources [4]. In fact, studies confirmed that operators perform better with foveated images compared to non-foveated ones in limited bandwidth conditions [6].

3. THE PROPOSED IMAGE FUSION SYSTEM

The log-polar transformation can be integrated with image fusion as depicted in Figure 2. Registered input images are first transformed from the conventional image plane to the log-polar representation. Then, images in the log-polar representation are fused. After fusion, the fused image is transformed to the conventional image plane again for display (a fused image before/after the inverse log-polar mapping are depicted in Figure 3).

Across the years, several image fusion methods have been developed. In general, they fall under three broad categories; they are mono-scale, multi-scale, and biological fusion methods. In mono-scale methods, the fusion algorithm is directly performed on registered raw input images. Multi-scale fusion methods can be identified by the multi-scale transformation process that precedes the fusion. The fusion is then performed on the transformed images rather



(a) THE FUSED IMAGE IN THE LOG-POLAR REPRESENTATION.



(b) THE FUSED IMAGE TRANSFORMED INTO THE IMAGE PLANE.

Figure 3. The fused image before/after applying the inverse log-polar transform.

than the raw images. The last category includes algorithms where input images are presented in a way that facilitates a mental fusion of the input images. In other words, the fused image is only realized in the brain, hence the name, biological fusion methods.

For the purpose of performing a comparative analysis, three well-known techniques are considered, namely, image averaging, Laplacian pyramid, and Discrete Wavelet Transform (DWT). The first is a mono-scale method whereas the other two are multi-scale methods. The selected multi-scale methods follow the general framework of multi-scale methods introduced by [12]. Image averaging is the simplest image fusion algorithm where the fused image, F , is a pixel-by-pixel averaging of the two input images, I_1 & I_2 , as in [13]:

$$F = 0.5I_1 + 0.5I_2 \quad (11)$$

It has the lowest computational complexity out of the three. With the Laplacian pyramid, input images are decomposed into five levels using a 5×5 Gaussian filter (with $a = 0.375$). The saliency of a coefficient is taken as the maximum absolute value in a 3×3 surrounding window. Corresponding coefficients are combined by taking the coefficient of the highest saliency whereas approximation images are combined using averaging. No consistency checking is performed with the Laplacian pyramid fusion. In the third method, the input images are decomposed into five levels using a Daubechies wavelet filter with filter length equals 8 ('db4' in Matlab). Corresponding coefficients in the multi-scale representations of input images are combined by taking the one of the highest absolute value, that is 'sample-based absolute' saliency measure with 'max' combination method. The spatial consistency is assured using a 3×3 consistency window. Approximation images are averaged to get the approximation image of the fused image.

4. TIME ANALYSIS

In our platform, the visible and infrared images, after registration, are with spatial dimensions of 648×488 . It is noteworthy that not all of the input image is mapped into the log-polar domain. The biggest square in the center of the image that can fit completely in it dictates the area that will be mapped. In this particular case, it will be a

Table 1. The improvement in execution time when using log-polar mapping

Method	Ratio of time taken to fuse images without/with log-polar mapping
Image averaging	0.021951
Laplacian	2.494605
DWT	6.730341

488×488 square in the center of the image. This square will be mapped into a log-polar image with spatial dimensions of 252×152 , which are considered to be high resolution for log-polar images. In its second column, Table 1 lists the ratio of the execution time taken by different fusion techniques to fuse input images in their normal representation (648×488) to the time taken to fuse (252×152) log-polar images. The table clearly shows the advantage of using log-polar images instead of the full-resolution image in all fusion techniques apart from the simple image averaging method. The image averaging method is time efficient and the overhead time in mapping to log-polar and from it exceeds the reduction of execution time as a result of reducing the size of the input data set.

5. PERFORMANCE EVALUATION

Subjective and objective evaluation measures can be used to determine the effect of introducing log-polar transformation in image fusion. Because subjective evaluation requires lots of time, effort, and physical & human resources, we confine our comparison to objective evaluation using established objective measures. The measure are Mutual Information (MI) [14], Xydas & Petrovic ($Q^{AB/F}$) [15], and Piella & Heijmans (Q_w) [16] metrics. These measures, although not perfect, provide a good insight into the quality of fused images.

For the purpose of testing, sixteen pairs of images are used. These images resemble different situations in the application domain. Fused images from input images in the normal plane and those which are produced from input images in the log-polar representation are evaluated. The log-polar mapping results in fused images that are blurred in the periphery and sharp in the fovea. Therefore, in order to assure a fair evaluation, only a small area at the center of the scene is considered for objective evaluation. This area has width and height equal one fourth of the width and height of the whole image. This decision is justified in actively controlled systems as in our case. All objective results are averaged and they are summarized in Table 2.

It is noticeable from the table that, based on these measures, the introduction of log-polar transformation degrades the quality of the fused image in all cases. However, these numbers should be accepted with caution as they treat pixels in the fovea and in the periphery the same. Thus, they do not perfectly reflect the fact that log-polar images are perceptually identical to the original images. In this sense, these measures may be used to compare log-polar image

Table 2. Objective evaluation of fusion techniques

Fusion methods	Log-polar transform	Performance measures		
		Q_w	MI	$Q^{AB/F}$
Image averaging	N	0.676	2.908	0.412
	Y	0.411	1.252	0.282
Laplacian	N	0.818	2.591	0.858
	Y	0.552	1.628	0.512
DWT	N	0.807	2.443	0.748
	Y	0.593	1.314	0.497

fusion methods with each other but not with those that do not use log-polar transformation. So, for log-polar fused images, Xydas & Petrovic ($Q^{AB/F}$) and Mutual Information (MI) metrics prefer the Laplacian pyramid fusion whereas Piella & Heijmans (Q_w) metric, with a fraction, favours the Discrete Wavelet Transform fusion.

6. CONCLUSION

The log-polar transformation is potentially a very effective way to reduce execution time in user-centred image fusion. It not only reduces processing power, communication bandwidth, and storage requirements but also maintains perceptual information. It achieves this by reducing input data set size to match the space-variant sampling pattern in the retina. In other words, the resultant fused images have high resolution in the fovea and the resolution smoothly decreases while moving away from the fovea towards the periphery. Furthermore, this solution requires no change either in the hardware used or in the fusion technique. For the quality of the log-polar to be evaluated properly, space-variant objective evaluation measures are needed. These measures should be compared with subjective evaluation results in order to be proven effective.

7. REFERENCES

- [1] A. Toet and J. Walraven, "New false color mapping for image fusion," *Optical Engineering*, vol. 35, no. 3, pp. 650–8, Mar. 1996.
- [2] O. Rockinger, "Pixel level fusion of image sequences using wavelet frames," *Proceedings of the 16th Leeds Annual Statistical Research Workshop*, pp. 149–154, 1996.
- [3] J. J. Lewis, R. J. O'Callaghan, S. G. Nikolov, D. R. Bull, and C. N. Canagarajah, "Region-based image fusion using complex wavelets," *The 7th International Conference on Information Fusion*, vol. 1, pp. 555–562, 2004.
- [4] F. Panerai, C. Capurro, and G. Sandini, "Space variant vision for an active camera mount," *Visual Information Processing IV*, vol. 2488, pp. 284–96, Apr. 1995.
- [5] N. Oshiro, N. Maru, A. Nishikawa, and F. Miyazaki, "Binocular tracking using log polar mapping," *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. vol.2, pp. 791–8, Nov. 1996.
- [6] W. S. Geisler and J. S. Perry, "A real-time foveated multiresolution system for low-bandwidth video communication," *Human Vision and Electronic Imaging III*, vol. 3299, pp. 294–305, 1998.
- [7] F. Berton, G. Sandini, and G. Metta, "Anthropomorphic visual sensors," *Encyclopedia of Sensors*, vol. 10, pp. 1–16, 2006.
- [8] R. Manzotti, A. Gasteratos, G. Metta, and G. Sandini, "Disparity estimation on log-polar images and vergence control," *Computer Vision and Image Understanding*, vol. 83, no. 2, pp. 97–117, 2001.
- [9] S. Zokai and G. Wolberg, "Image registration using log-polar mappings for recovery of large-scale similarity and projective transformations," *IEEE Transactions on Image Processing*, vol. 14, no. 10, pp. 1422–34, Oct. 2005.
- [10] M. Tistarelli and G. Sandini, "On the advantages of polar and log-polar mapping for direct estimation of time-to-impact from optical flow," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 4, pp. 401–10, Apr. 1993.
- [11] F. Berton, "A brief introduction to log-polar mapping," Technical report, LIRA-Lab, University of Genova, Feb. 2006.
- [12] R. S. Blum, Z. Xue, and Z. Zhang, "An overview of image fusion," in *Multi-sensor Image Fusion And Its Applications*, R. S. Blum and Z. Liu, Eds., vol. 26 of *Signal Processing and Communication*, pp. 1–35. CRC Press, 2005.
- [13] J. J. Lewis, R. J. O'Callaghan, S. G. Nikolov, D. R. Bull, and N. Canagarajah, "Pixel- and region-based image fusion with complex wavelets," *Information Fusion*, vol. 8, no. 2 SPEC ISS, pp. 119–130, 2007.
- [14] S. Li, J. T. Kwok, and Y. Wang, "Multifocus image fusion using artificial neural networks," *Pattern Recognition Letters*, vol. 23, no. 8, pp. 985–997, 2002.
- [15] V. Petrovic, T. Cootes, and R. Pavlovic, "Dynamic image fusion performance evaluation," *The 10th Conference of the International Society of Information Fusion*, pp. 1154–60, July 2007.
- [16] N. Cvejic, A. Loza, D. Bull, and N. Canagarajah, "A novel metric for performance evaluation of image fusion algorithms," *Transactions on Engineering, Computing and Technology*, vol. 7, pp. 80–86, 2005.