Merging thermal and visual images by a contrast pyramid

Alexander Toet, van Ruyven, J.J. & Valeton, J.M.

This paper was published in

Optical Engineering, 28(7), pp. 789-792

Merging thermal and visual images by a contrast pyramid

A. Toet, MEMBER SPIE
Institute for Perception TNO
Kampweg 5
3769 DE Soesterberg, The Netherlands

L. J. van Ruyven
Philips Usfa B.V.
Advanced Development Bureau
Meerenakkerweg 1
5600 MD Eindhoven, The Netherlands

J. M. Valeton
Institute for Perception TNO
Kampweg 5
3769 DE Soesterberg, The Netherlands

Abstract. Integration of images from different sensing modalities can produce information that cannot be obtained by viewing the sensor outputs separately and consecutively. This paper introduces a hierarchical image merging scheme based on a multiresolution contrast decomposition (the ratio of a low-pass pyramid). The composite images produced by this scheme preserve those details from the input images that are most relevant to visual perception. The method is tested by merging parallel registered thermal and visual images. The results show that the fused images present a more detailed representation of the depicted scene. Detection, recognition, and search tasks may therefore benefit from this new image representation.

Subject terms: image processing; ratio of low-pass pyramid; image fusion; visual perception; contrast decomposition; sensor fusion.

Optical Engineering 28(7), 789-792 (July 1989).

CONTENTS

1. INTRODUCTION

Integration of visual (CCD) and thermal (FLIR) images can produce information that cannot be obtained by viewing the sensor outputs separately and consecutively. In defense applications, for example, details of targets that are hard to detect in a visual image (that have low visual contrast) can sometimes easily be seen in a thermal image. Incomplete representation of targets in thermal images may result from large temperature gradients within these objects. Also, the exact location of targets in FLIR images may be hard to assess when the background has low thermal contrast. The increased information content of integrated FLIR and CCD images is expected to improve observer performance for a range of different tasks, e.g., the control of remotely piloted vehicles, driving in hostile environments, and surveillance.

The essential problem in merging images for visual display is "pattern conservation": important details of the component images must be preserved in the resulting composite image, while the merging process must not introduce spurious pattern elements that could interfere with subsequent analysis. Simple methods to combine image details (e.g., cutting and pasting, sometimes followed by edge blurring) often create edge artifacts between regions taken from different imges. Burt and Adelson' introduced a hierarchical image merging scheme based on the Laplacian pyramid.[2,3] The Laplacian decomposition scheme is based on local luminance differences. However, the human visual system is sensitive only to local luminance contrasts. Thus, Burt and Adelson's scheme is not necessarily a good choice when the final result is to be judged by human observers.

In this paper we introduce a so-called contrast or ratio of low pass (ROLP) pyramid that fits models of the human visual system .[4,5] In the ROLP image fusion scheme, judgments on the relative importance of pattern segments are based on their local luminance contrast. First the input images are decomposed into sets of light and dark blobs on different levels of resolution. Thereafter, the absolute contrast of blobs at corresponding locations and at corresponding levels of resolution are compared. The actual image fusion is done by selecting the blobs with maximum absolute luminance contrast. The fused image is reconstructed from the set of blobs or pattern primitives thus

---

Invited Paper VI-110 received Nov. 14, 1988; revised manuscript received March 3, 1989; accepted for publication May 2, 1989.

obtained. As a result, perceptually important details with a relatively high local luminance contrast are preserved in the composite image.[7]

2. METHOD

2.1. Contrast pyramid

The construction of the ROLP pyramid is similar to the construction of the popular Laplacian pyramid.[2] First a Gaussian or low-pass pyramid is constructed. This is a sequence of images in which each image is a low-pass-filtered and subsampled copy of its predecessor. Let array G_o contain the original image. This array becomes the bottom or zero level of the pyramid structure. Each node of pyramid level ℓ ($1 \le \ell \le N$, where N is the index of the top level of the pyramid) is obtained as a Gaussian-weighted average of the nodes at level $\ell - 1$ that are positioned within a 5×5 window centered on that node.

Convolving an image with a Gaussian-like weighting function is equivalent to applying a low-pass filter to the image. Gaussian pyramid construction generates a set of low-pass-filtered copies of the input image, each with a bandlimit one octave lower than that of its predecessor. Because of the reduction in spatial frequency content, each image in the sequence can be represented by an array that is half as large as that of its predecessor. The process that generates each image in the sequence from its predecessor is called a REDUCE operation since both the sample density and the resolution are decreased. Thus, for $1 \le \ell \le N$ we have

$$G_\ell = \mathrm{REDUCE}[G_{\ell-1}] , \tag{1}$$

meaning

$$G_\ell(i,j) = \sum\sum_{m,n=-2}^{2} w(m,n) G_{\ell-1}(2i + m, 2j + n) . \tag{2}$$

The weighting function w(m,n) is separable: w(m,n) = w'(m)w'(n), where w'(0) = a, w'(1) = w'(-1) = 0.5, and w'(2) = w'(-2) = a/2. A typical value of a is 0.4.

Because we are primarily interested in merging images for visual display, we demand that visually important details of the component images be preserved in the resulting composite image. It is a well-known fact that the human visual system is sensitive to local luminance contrast. If an image fusion scheme is to preserve visually important details, it must exploit this fact.

We now present an image decomposition scheme that is based on local luminance contrast. This scheme computes the ratio of the low-pass images at successive levels of the Gaussian pyramid. Since these levels differ in sample density, it is necessary to interpolate new values between the given values of the lower frequency image before it can divide the higher frequency image. Interpolation can be achieved simply by defining the EXPAND operation as the inverse of the REDUCE operation. Let $G_{e,k}$ be the image obtained by applying EXPAND to G_e k times. Then,

$$G_{\ell,0} = G_\ell , \tag{3}$$

$$G_{\ell,k} = \mathrm{EXPAND}[G_{\ell,k-1}] , \tag{4}$$

meaning

$$G_{\ell,k}(i,j) = 4 \sum\sum_{m,n=-2}^{2} w(m,n) G_{\ell,k-1}\left(\frac{i+m}{2}, \frac{j+n}{2}\right) , \tag{5}$$

where only integer coordinates contribute to the sum. A sequence of ratio images R_i is defined by

$$R_i = \frac{G_i}{\mathrm{EXPAND}[G_{i+1}]} , \quad \text{for } 0 \le i \le N-1 , \tag{6}$$

$$R_N = G_N . \tag{7}$$

Thus, every level R_i is a ratio of two successive levels in the Gaussian pyramid. Luminance contrast is defined as $C = (L - L_b)/L_b = (L/L_b) - I$, where L denotes the luminance at a certain location in the image plane, L_b represents the luminance of the local background, and $I(i,j) = 1$ for all i,j. When C_i is defined as $C_i = G_i/\mathrm{EXPAND}[G_{i+1}] - I$ we have $R_i = C_i + I$. For these reasons we refer to the sequence R_i as the ROLP or contrast pyramid.

The ROLP pyramid is a complete representation of the original image. G_o can be recovered exactly by reversing the steps used in the construction of the pyramid:

$$G_N = R_N , \tag{8}$$

$$G_i = R_i \, \mathrm{EXPAND}[G_{i+1}] , \quad \text{for } 0 \le i \le N-1 . \tag{9}$$

This property of the ROLP pyramid is essential for the image merging scheme described below.

2.2. Merging scheme

The image merging scheme can be cast into a three-step procedure. First, a ROLP pyramid is constructed for each of the source images. We assume that the different source images are in register and have the same dimensions. The latter restriction is not very serious since it can be shown that the method can also be applied to images with different definition regions as long as there is a common overlap.[1] Second, a ROLP pyramid is constructed for the composite image by selecting values from corresponding nodes in the component pyramids. The actual selection rule depends on the application and may be based on individual node values or on masks or confidence estimates. For example, in the case of the fusion of two input images A and B into a single output image C and maximum absolute contrast as a selection criterion, we have, for all i, j, and l

$$RC_\ell(i,j) = \begin{cases} RA_\ell(i,j) , & \text{if } |RA_\ell(i,j) - 1| > |RB_\ell(i,j) - 1| , \\ RB_\ell(i,j) , & \text{otherwise} , \end{cases} \tag{10}$$

where RA and RB represent the ROLP pyramids for the two source images and RC represents the fused output image. Finally, the composite image is recovered from its ROLP pyramid representation through the EXPAND and multiply reconstruction procedure.

3. EXPERIMENT

In this section we describe an experiment we performed to test the ROLP image fusion scheme. First, we simultaneously recorded spatially registered CCD and FLIR images on video tape. The images were thereafter digitized and brought into register. Finally, we digitally merged corresponding images using the ROLP scheme.

3.1. Image acquisition

Figure 1 is a schematic drawing of the experimental setup used to record the CCD and FLIR images. The CCD and FLIR

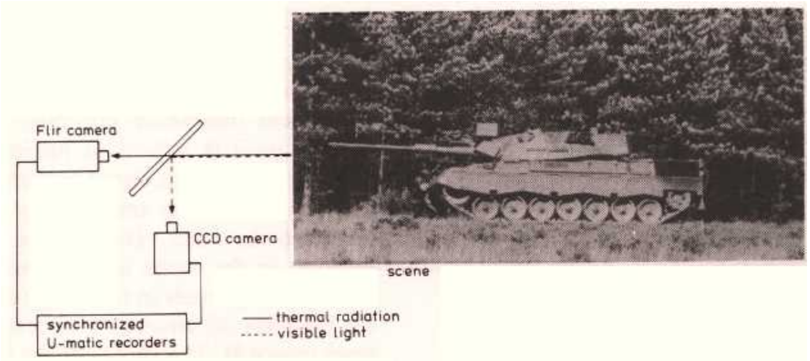


Fig. 1. Schematic representation of the experimental setup used to record the CCD and FLIR images.

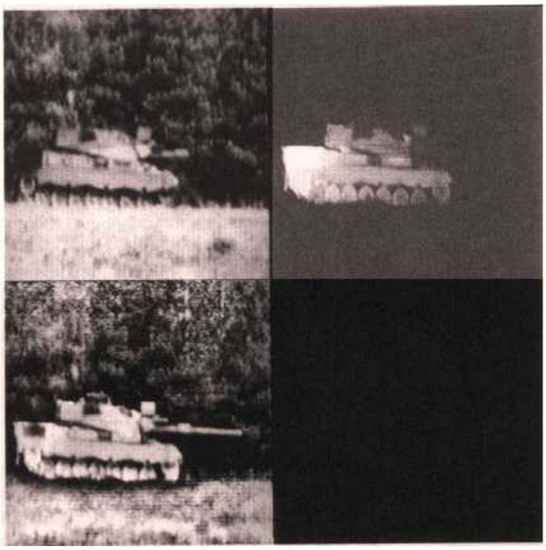


Fig. 2. The CCD (upper left) and FLIR (upper right) images of a tank. The result of the image merging scheme is shown on the lower left.

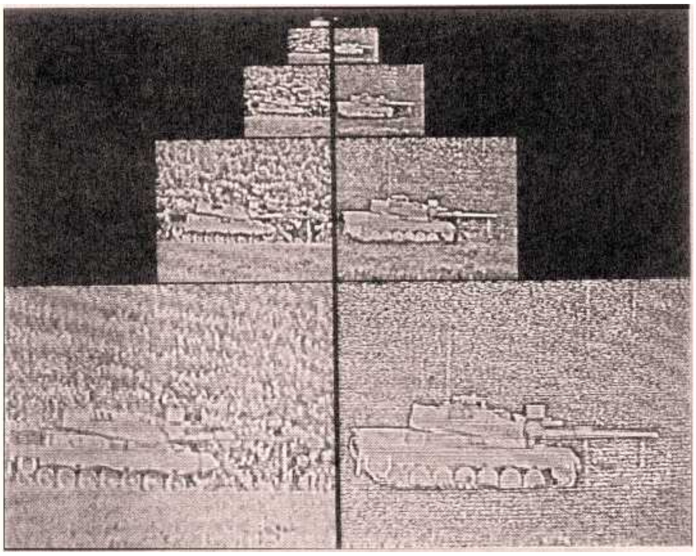


Fig. 3. A five-level ROLP pyramid for the visual image from Fig. 2(a) (left column) and the thermal image from Fig. 2(b) (right column).

cameras are directed along the same optical axis. This is done by using a slanting germanium mirror. Germanium transmits thermal radiation while reflecting visible light. Spatial image registration was obtained by creating a common Cartesian coordinate grid for both image modalities. This was done by placing nine light bulbs in the scene. The bulbs were attached to three vertical, equidistant poles. They were clearly visible in both image modalities and small enough to provide well-defined reference points. For spatial calibration of the recordings, the CCD and FLIR images were displayed on the R and G channels of an RGB monitor. Spatial registration was obtained by superimposing the R and G grid points (i.e., the corresponding images of the light bulbs) by adjusting the magnification, tilt, and direction of the CCD camera.

Prior to each recording session some shots of the reference grid were taken. After digitization of the recordings, these shots were used to correct for small image distortions. This was done by affine warping transforms. The bulbs were removed prior to recording objects in the scene. The signals from both cameras were recorded on U-matic video tape.

3.2. Image merging

Figure 2 shows the CCD and FLIR images of a tank together with the final result of the merging process. In the CCD image it is hard to distinguish the front part of the gun barrel, the back part of the vehicle (containing the engine room), and the man hood located on top of the tank, all of which have very low contrast in the visual image. In the thermal image, however, these parts are clearly visible. The background, which is clearly visible in the CCD image, is nearly indistinguishable in the FLIR image. Notice that all of the aforementioned details from both image modalities are clearly represented in the fused image.

Figure 3 shows the ROLP pyramids of the CCD (left column) and FLIR (right column) images from Fig. 2. Notice that the front part of the gun barrel, the back part of the vehicle, and the man hood on top of the tank are accentuated in the ROLP pyramid of the FLIR image and are only very weakly represented in the ROLP pyramid of the CCD image. However, the background, which is accentuated in the ROLP representation of the CCD image, is very noisy in the FLIR image.

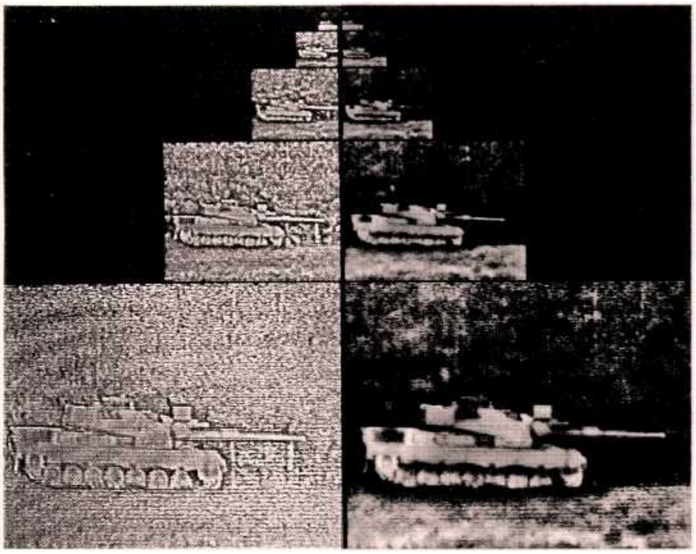


Fig. 4. The composite ROLP pyramid (left column) and the reconstructed composite Gaussian pyramid (right column) constructed from the pyramids in Fig. 3. The lower image in the right column represents the final result.

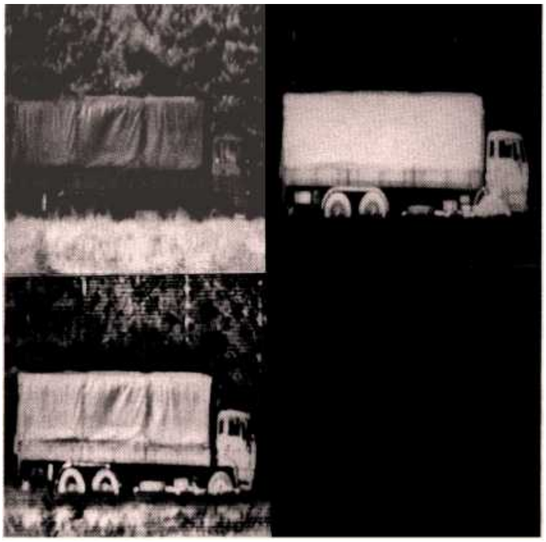


Fig. 5. The CCD (upper left) and FUR (upper right) images of a truck. The result of the image merging scheme is shown on the lower left.

In the examples presented in this paper we used the maximum absolute contrast node selection rule. Figure 4 shows the composite ROLP pyramid (left column) and the reconstructed composite Gaussian pyramid (right column). The bottom level of the composite Gaussian pyramid represents the final result. This fused image contains those details from both input images that have maximum local contrast.

Another example of the image fusion scheme is shown in Fig. 5. In this example the CCD and FLIR images of a truck were fused by the ratio of contrast pyramid. The result shows that the sharp outlines of the FLIR image of the truck enhance the contrast of the truck in the fused image. The outlines of the driver's cabin, the gas tank, and the wheels are not clearly visible in the CCD image but are well represented in the FLIR image. In the fused image all of these features are combined.

4. DISCUSSION

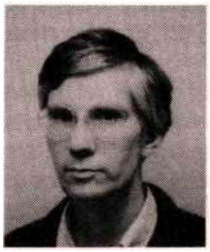
In this paper we have introduced an image fusion method intended for human observation. The method preserves details of high local luminance contrast. The contrast decomposition scheme used is related to luminance processing in the early stages of the human visual system. The method is quite general and can be used to transfer any useful information from one image to the other. This can be done independently for each location in the scene and on every level of resolution. Thus, information present in the image obtained from sensor A can be used to filter information at corresponding locations in the image from sensor B. The choice of the filter operation depends on the application (and can be anything from contrast enhancement to smoothing and thresholding). This method can be used to merge images from a variety of sensing modalities prior to display.

5. REFERENCES

1. P. J. Burt and E. H. Adelson, "Merging images through pattern decomposition," in *Applications of Digital Image Processing* VIII, A. G. Tescher, ed., Proc. SPIE 575, 173-181 (1985).
2. P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," IEEE Trans. Commun. COM-31, 532-540 (1983).
3. P. J. Burt, "The pyramid as a structure for efficient computation," in *Multi-resolution Image Processing and Analysis,* A. Rosenfeld, ed., Springer-Verlag, Berlin (1984).
4. J. J. Koenderink and A. J. van Doom, "Visual detection of spatial contrast; influence of location in the visual field, target extent and illuminance level," Biol. Cybem. 30, 157-167 (1978).
5. G. J. Burton, N. D. Haig, and 1. R. Moorhead, "A self-similar stack model for human and machine vision," Biol. Cybem. 53, 397-403 (1986).
6. E. Adelson, C. H. Anderson, J. R. Bergen, P. J. Burt, and J. M. Ogden, "Pyramid methods in image processing," RCA Engineer 29, 33-41 (1984).
7. A. Toet, "Image fusion by a ratio of low pass pyramid," Pattern Rec. Lett. 9(4), 245-253 (1989).

Alexander Toet received his Ph.D. degree in physics from the University of Utrecht, The Netherlands, in 1987. He has published over 20 papers on spatial vision, image processing, and neural networks. His current research interests are spatial localization in the human visual system and the application of mathematical morphology in hierarchical image representations.

Lodewijk J. van Ruyven received his Ph.D. degree in physics from the Technical University of Eindhoven, The Netherlands, in 1964. He has done research on optoelectronic semiconductor devices for IR imaging systems, light conversion, detection, and luminescence. He is presently engaged in the automatic detection of counterfeit banknotes for the Dutch Central Bank.



J. Mathee Valeton studied physics at the University of Utrecht, The Netherlands. He obtained his Ph.D. degree in visual electrophysiology in 1983. Since then he has worked on spatial vision and the evaluation of electro-optical imaging systems performance in battlefield conditions.