Machine Vision and Applications (1990) 3:1–11

Machine Vision and Applications

Hierarchical Image Fusion

Alexander Toet
Institute for Perception TNO, Kampweg 5, Soesterberg NL-3769-DE, The Netherlands

Abstract: A hierarchical image fusion scheme is presented that preserves those details from the input images that are most relevant to visual perception. Results show that fused images present a more detailed representation of the scene and provide information that cannot be obtained by viewing the input images separately. Detection, recognition, and search tasks may therefore benefit from this fused image representation.

Key Words: sensor fusion, ratio of low-pass pyramid, mathematical morphology, contrast decomposition, multiresolution image representations

1 Introduction

A hierarchical scheme is presented for the fusion of signals from multiple imaging systems. The input of the algorithm can be an arbitrary number of simultaneously registered images. The only restriction of the method is that the input images must have some degree of spatial overlap. First, each input image is decomposed into a set of perceptually relevant pattern primitives. Pattern sets for the various source images are then combined to form a single set for the composite image. Finally, the composite image is reconstructed from its set of primitives. As a result, the output of the algorithm is a composite image that preserves those details from the input images that are most relevant to visual perception.

There are many imaging modes (e.g., direct view optics, television, forward-looking infrared, infrared search and track, microwave radar, millimeter wave radar, laser radar, synthetic aperture radar, laser rangefinder, acoustic transducer array, radio frequency interferometer, etc.) in current use. Systems that use a number of imaging systems severely increase the work load of a human operator. Moreover, a human observer cannot reliably integrate visual information by viewing multiple images separately and consecutively. The integration of information across multiple human operators is nearly impossible. Thus, a system that fuses images from multiple sensors into a single image is likely to be of great practical value.

The next section presents a hierarchical image fusion scheme. This scheme requires a complete description of the structure of the input images. A complete image representation can be obtained by studying the image structure over a range of scales (section 2.1). Such a multiresolution image description can be produced by repeated application of a size-limiting filter operator of a progressively increasing scale (section 2.2). A hierarchical image decomposition based on local luminance contrast can be obtained by computing the ratio between successive levels in a multiresolution description (section 2.3). Both linear (section 2.4) and morphological (section 2.5) filters can be used in the construction of a multiresolution image decomposition. Section 2.6 introduces an image fusion scheme based on a multiresolution contrast decomposition of the input images. Section 3 illustrates this fusion scheme by simulating the binocular perception of artificial input images (section 3.1) and by merging thermal and visual images (section 3.2). Finally, section 4 presents a discussion and some conclusions.

2 Multiresolution Image Fusion

The essential problem in merging images for visual display is "pattern conservation": merging must preserve detail in the resulting composite image while not introducting spurious pattern elements that could interfere with subsequent analysis. Simple methods to combine image details often create

Address reprint requests to: Alexander Toet, Institute for Perception TNO, Kampweg 5, Soesterberg NL-3769-DE, The Netherlands.

edge artifacts between regions taken from different images (e.g., cutting and pasting, sometimes followed by edge blurring) or may annihilate image details (e.g., gray-scale addition).

In this section we present a scheme to combine images from multiple sensors. The scheme employs a hierarchical representation of the input signals. The hierarchical representation is based on the fact that image structure depends on resolution.

2.1 The Resolution Dependency of Image Structure

When we zoom in on an image, we clearly see its substructure but we lose its outlines. Alternatively, when we blur an image to obtain a good overview of the depicted scene, details of that scene will be hard to recognize. It appears that relevant details of images exist only over a restricted range of scales.

When an image is blurred, it becomes less articulated because some of its extrema (light and dark blobs) disappear. As a result, any image can be represented by a juxtaposed and nested set of light and dark blobs, wherein each blob has a limited range of resolution in which it manifests itself (Koenderink 1984, Toet et al. 1984). Hence, a complete image description can be obtained by studying the image structure over a range of scales.

2.2 The Construction of a Multiresolution Image

In the sequel we will refer to filter operators that eliminate details smaller than a certain size as *size-limiting filters*. The difference of two size-limiting filters will be called a *size-selective filter*.

A family of images with progressively decreasing structural content (i.e., progressively less detail) can be produced by repeated application of a size-limiting filter operator of a progressively increasing scale.

A hierarchical image representation can be obtained by relating the descriptions generated by filters of increasing size-limits (O and Toet 1989). A trivial hierarchical relation is the one-to-one relation, whereby a correspondence is established between all points of successively filtered versions of the image.

Descriptions generated by large filter operators contain no small details. This consideration allows a progressive reduction of the sample frequency with a progressive increase of the filter size. Sampling induces a natural hierarchical relation.

A well-known hierarchical image representation is the quadtree or pyramid (e.g., Rosenfeld 1984). A pyramid is a sequence of images in which each image is a filtered and subsampled copy of its predecessor. Successive levels of a pyramid are generally reduced resolution versions of the input image (hence, the term "multiresolution" representation; e.g., Figure 5). However, the representation may also consist of descriptive information about certain image features (e.g., edges). In this case successive levels of the pyramid represent increasingly coarse approximations to these features (e.g., Figure 6).

A pyramidal image representation is constructed in the following way. The original image is the bottom or zero level P_0 of the pyramid. Each node of pyramid level l ($1 \leq l \leq N$ where N is the index of the top level of the pyramid) is obtained by sampling a filtered version of level $l - 1$. The process that generates each image in the sequence from its predecessor will be called a REDUCE operation since both the sample density and the resolution are decreased. Thus, for $1 \leq l \leq N$ we have

$$P_l = \text{REDUCE}(P_{l-1})$$

Pyramids convert local image features into global features. The global information can be used to impose constraints on local operations (Burt 1984, Crowley and Parker 1984). As a result, hierarchical operations can be more efficient than operations performed at a single level of resolution.

2.3 The Construction of a Multiresolution Image Decomposition

A hierarchical decomposition of an image into a set of light and dark blobs can be obtained by successive application of a range of size-selective filters. We noted before that size-selective filters are equivalent to a difference of size-limiting filters. As a result, a pyramidal image decomposition can be computed as the difference between successive levels in a pyramid of images, constructed with size-limiting filters. Because each level differs in sample density, it is necessary to interpolate new values between the given samples in an image at a higher pyramid level before this image can be subtracted from the image residing at a lower pyramid level.

Interpolation can be achieved simply by the EXPAND operation:

$$P_{l,0} = P_l$$

and

$$P_{l,k} = \text{EXPAND}(P_{l,k-1})$$

where $P_{l,k}$ represents the image obtained by k successive applications of the EXPAND operation to P_l.

Let P be a pryamid constructed with size-limiting

filters. An error pyramid E is a group of error images in which each image contains only details within a restricted range of sizes (i.e., a size selectively filtered set of images or a sieve (O and Toet 1989). As noted before, E can be computed as the difference between successive levels of P:

$$E(P)_l = P_l - \text{EXPAND}(P_{l+1}) \qquad \forall l \leq N$$

and

$$E(P)_N = P_N$$

When images are merged for visual display, details of the component images that are visually important must be preserved in the merged image. It is a well-known fact that the human visual system is sensitive to local luminance contrast (Koenderink and van Doorn 1978, 1982). If an image fusion scheme is to preserve visually important details it must exploit this fact.

This section presents an image decomposition scheme that is based on local luminance contrast (Toet 1989a). This scheme computes the ratio of the low-pass images at successive levels of the Gaussian pyramid. A multiresolution image decomposition based on local luminance contrast is obtained by computing a sequence of ratio images $R(P)_l$ defined by

$$R(P)_l = P_l/\text{EXPAND}(P_{l+1}) \qquad 0 \leq l \leq N - 1$$

and

$$R_N = P_N$$

Thus, every level R_l is a ratio of two successive levels in the size-limiting pyramid P.

Luminance contrast C is defined as

$$C = \frac{L - L_b}{L_b} = \frac{L}{L_b} - 1$$

where L denotes the luminance at a certain location in the image plane, L_b represents the luminance of the local background. Let $I_l(i, j) = 1\ \forall i, j, l$ represent the *unit pyramid*. When C_l is defined as

$$C_l = P_l/\text{EXPAND}(P_{l+1}) - I_l$$

we have

$$R_l = C_l + I_l$$

Therefore, we will refer to the sequence R_l as the *contrast pyramid*.

The contrast pyramid is a complete representation of the original image. P_0 can be recovered exactly by reversing the steps used in the construction of the pyramid:

$$P_N = R_N$$

and

$$P_l = R_l \cdot \text{EXPAND}(P_{l+1}) \qquad 0 \leq l \leq N - 1$$

This property of the contrast pyramid will be essential for the image merging scheme described subsequently.

2.4 Linear Filters

The application of linear filters in the construction of a size-limited pyramid results in a sequence of images in which each is a low-pass filtered and subsampled copy of its predecessor.

A popular low-pass filter is the convolution with a Gaussian kernel. The application of this type of filter in the pyramid construction results in the well-known Gaussian pyramid (Burt 1984). In this case the filtering and sampling operations can be combined into a single operation by computing the Gaussian weighted average.

Let array P_0 contain the original image. This array represents the bottom or zero level of the pyramid structure. Each node of pyramid level l ($1 \leq l \leq N$ where N is the index of the top level of the pyramid) is obtained as a Gaussian weighted average of the nodes at level $l - 1$ that are positioned within a window centered on that node. For a typical window of size 5×5 the REDUCE operation is defined by

$$P_l(i, j) = \sum_{m,n=-2}^{2} w(m, n) P_{l-1}(2i + m, 2j + n)$$

The weighting function $w(m, n)$ is separable:

$$w(m, n) = w'(m)w'(n)$$

with $w'(0) = a$, $w'(1) = w'(-1) = 0.5$, $w'(2) = w'(-2) = 0.5a$. A typical value of a is 0.4.

The corresponding EXPAND operation is given by

$$P_{l,k}(i, j) = 4 \cdot \sum_{m,n=-2}^{2} w(m, n) P_{l,k-1}\left(\frac{i + m}{2}, + \frac{j + n}{2}\right)$$

where only integer coordinates contribute to the sum.

The error pyramid corresponding to the Gaussian pyramid is known in the literature as the DOG (difference of Gaussians; e.g., Burton et al. 1986), DOLP (difference of low pass; e.g., Crowley and Parker 1984) or Laplacian pyramid (e.g., Burt and Adelson 1983).

2.5 Morphological Filters

Multiresolution structural image representation and decomposition schemes typically apply linear (low- or band-pass) filters with progressively increasing spatial extent to generate a sequence of images with progressively decreasing resolution (e.g., Burt and Adelson 1983; Burt 1984; Crowley and Parker 1984, Burton et al. 1986). Linear filter techniques alter the object intensities, and therefore the estimated location of their contours. As a result, decomposition schemes based on these techniques are of limited applicability to tasks involving precise measurement of object size and shape (O and Toet 1989). In contrast, morphological filters remove image details without adding a gray-scale bias (Serra 1982, 1988, Maragos 1987). They are therefore well suited for shape extraction.

The simplest morphological filters are the opening and closing transformations (Serra 1982). Opening and closing are dual operations, in that what one does to the image foreground the other does to the image background. By applying openings and closings iteratively the fore- and background of an image are treated in the same way. Sequential alternating application of opening and closing with the same structuring element removes details of the image (fore- and background) that are small relative to this structuring element. Using classical terminology, we can denote these alternating sequential filters as *morphological low-pass filters*. The filters are low pass because it is high-frequency fluctuation between a set and its complement, which is attenuated in the output image. A theoretical foundation for these operators was recently presented by Serra (1988).

In the sequel we will denote the opening with structuring element A and O_A and the closing with C_A. We will study the transformation generated by the successive application of the product of an opening and closing with a family of convex structuring elements. Let A be a convex structuring element. For any $l > 0$, $lA = A \oplus A \oplus \cdots \oplus A$ (l times) (where $\oplus$ denotes the Minkowski sum (Serra 1982) will also be convex. Let $F_{lA} = C_{lA}O_{lA}$ be the alternating iterative filter (Serra 1988) consisting of an opening followed by a closing. For $i < j$ we have $F_{jA}F_{iA} \geq F_{jA}$. Therefore, the smaller the structuring element is, the smaller are the details that are filtered or modified.

Before sampling images must be filtered (e.g., by opening or closing (Haralick et al. 1987a)). A sampled set can be reconstructed in two ways—by either a closing or a dilation. In both reconstructions the sample reconstructed image is equal to the sampled image. An image contains its reconstruction by closing and is contained in its reconstruction by dilation. Thus, morphological sampling cannot produce a reconstruction whose positional accuracy is better than the radius of the circumscribing disk of the structuring element used in the reconstruction process, in contrast to the sampling reconstruction process in signal processing from which only those frequencies below the Nyquist frequency can be reconstructed (Haralick et al. 1987a, Toet 1989b).

In this paper we will only use binary (i.e., bivalued or flat) structuring elements. The extension of our results to multivalued (function) structuring elements is straightforward. (For details see Serra 1982, Haralick et al. 1987b, Maragos 1987). Flat structuring elements are essential if a gray-scale transformation is to commute with the anamorphosis of a gray-scale image (e.g., contrast stretching). Furthermore, openings and closings by structuring elements with vertical sides such as cylinders and bricks maintain the vertical boundaries in the images they transform. Finally, flat structuring elements permit us to work level by level on gray-scale images that have been thresholded across several slices.

Next, we will describe the construction of a morphological size-limited pyramid. Let array P_0 contain the original image. This array will again represent the bottom level of the pyramid. Let $S_l \subset \mathbf{Z}^2$ (where $\mathbf{Z}$ denotes the set of integers) denote the sampling lattice corresponding with level l of the pyramid ($1 \leq l \leq N$). S_1 has grid size d_1 and S_l has grid size $d_l = 2^{l-1}d_1$. As a result, $S_i \supset S_j$ for $i \leq j$. Let $a \subset \mathbf{R}^2$ (where $\mathbf{R}$ denotes the set of real numbers) be a convex binary structuring element. The size-scaled discrete structuring element A_l applied at level l in the pyramid is equal to the restriction of the corresponding size-scaled continuous structuring element $2^{l-1}\ a$ to the sampling lattice S_l: $A_l = 2^{l-1}a|_{S_l}$. Because of the construction of the pyramid, we have $A_l = 2^{l-1}A_1|_{S_l} = 2A_{l-1}|_{S_l}$.

A morphological size-limited pyramid can be obtained by the successive application of an alternating iterative filter followed by sampling. The alternating iterative filter we adopted is given by

$$F_{A_l} = C_{A_l} O_{A_l} \qquad 1 \le l \le N$$

The iterative pyramid generation scheme is given by

$$P_l = \text{REDUCE}(P_{l-1}) = F_{A_l}(P_{l-1})|_{S_l} \qquad 1 \le l \le N$$

A square and flat (i.e., of uniform gray-value or "brick-like") structuring element was used with an initial size of 5×5. Its size was doubled for each layer in the stack.

The positional uncertainty introduced by morphological sampling may become fairly large in the discrete case (Toet 1989b). To eliminate positional errors in the EXPAND process, the sampling scheme was dropped altogether. In this case the EXPAND operation reduces to a simple projection:

$$P_{l,k}(i, j) = P_{l,k-1}(i, j)$$

This results in a stack of images with the same resolution but progressively diminishing structural content. The corresponding families of error and ratio images are easily obtained by respectively subtracting and dividing adjacent stack layers.

2.6 The Image Fusion Scheme

The image merging scheme can be cast into a three-step procedure (a schematical representation of this process is given by Figure 1). First, a contrast pyramid is constructed from each of the source images. The different source images are registered and have the same dimensions. The latter restriction is not very serious, as it can be shown that the method can also be applied to images with different definition regions as long as the intersection of these regions is not empty (Burt and Adelson 1985). Second, a contrast pyramid is constructed for the composite image by selecting values from corresponding nodes in the component error pyramids. The actual selection rule will depend on the application and may be based on individual node values or on masks or confidence estimates. For example, in case of the fusion of two input images A and B into a single output image C, and maximum absolute local lumi-

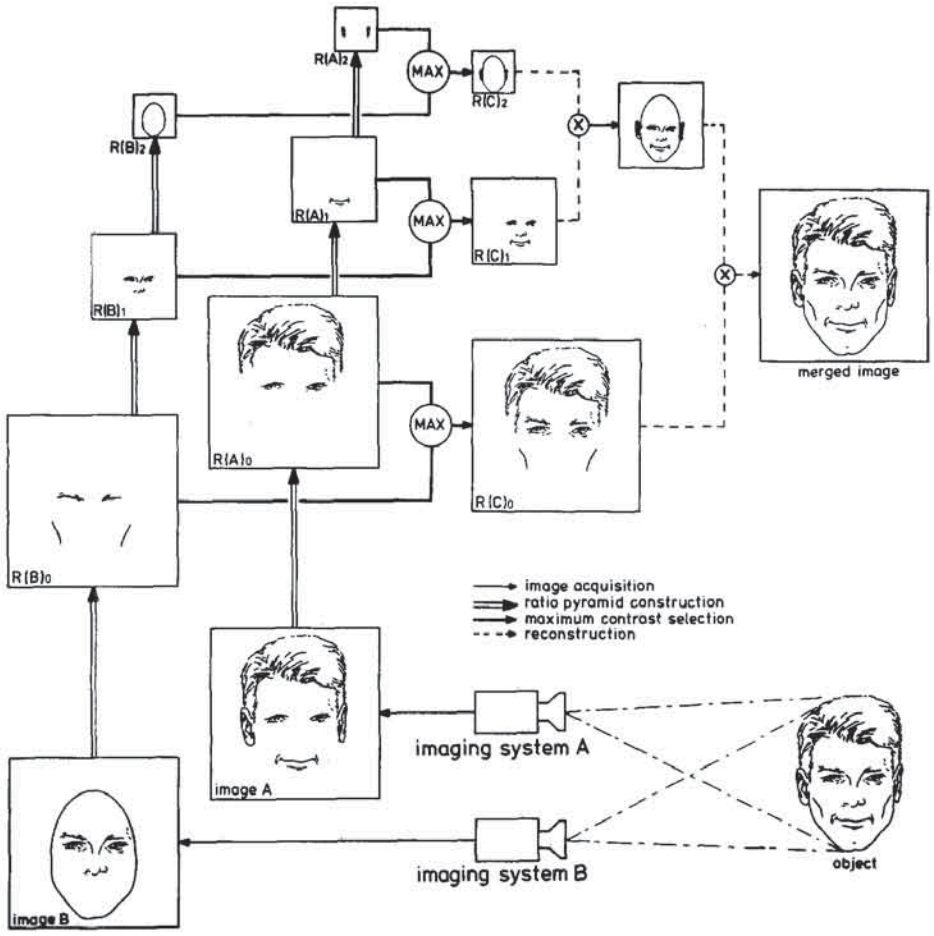


Figure 1. Schematical representation of the image fusion process.

nance contrast as a selection criterion, we have ($\forall i, j, l$)

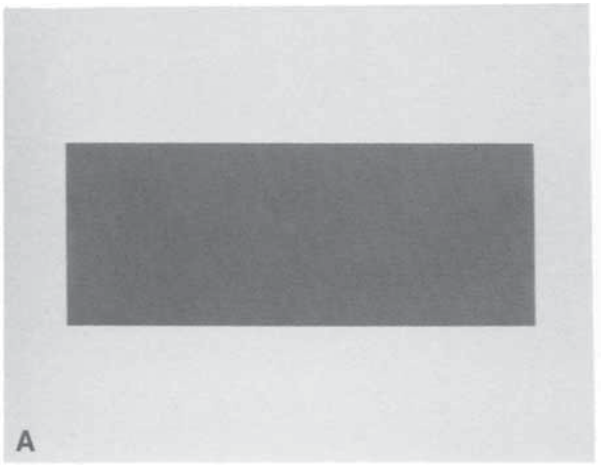
$$R(C)_l(i,j) = R(A)_l(i,j) \quad \text{if} \quad \|R(A)_l(i,j) - 1\| > \|R(B)_l(i,j) - 1\| = R(B)_l(i,j) \quad \text{otherwise}$$

where $R(A)$,$R(B)$ represent the ratio or contrast pyramids for the two source images and $R(C)$ represents the ratio pyramid for the fused output image. Finally, the composite image is recovered from its ratio pyramid representation through the corresponding reconstruction procedure.

3 Examples

3.1 Simulation of Binocular Perception

This section illustrates the image fusion scheme by applying it on artificial input images with a low amount of detail. The results show a certain analogy of the pyramid image merging scheme and the human visual system's ability to accumulate information from two monocular images into a perceptual whole.



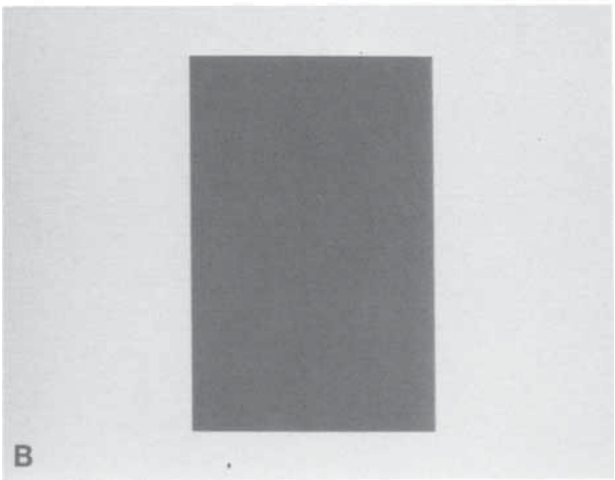


Figure 2. Two test images.

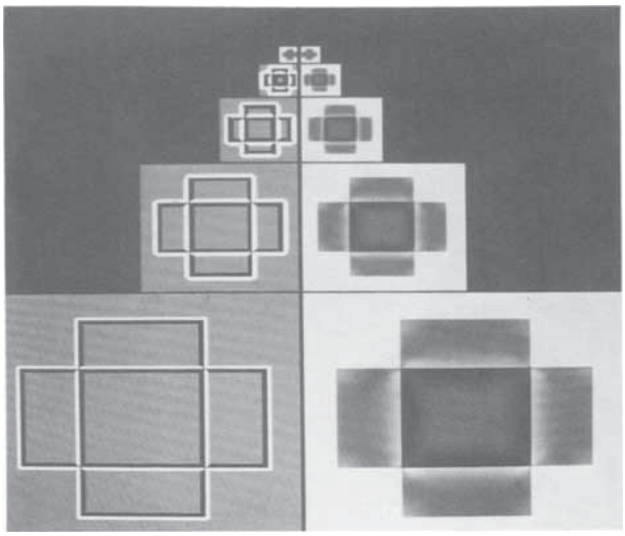


Figure 3. The composite Gaussian ratio pyramid (left column) of Figure 2 and the reconstructed composite Gaussian pyramid (right column). The lower level of the composite pyramid represents the final result of the image fusion scheme.

Supppose a horizontal gray bar (Figure 2a) is presented to the left eye of a human subject and a similar vertical bar (Figure 2b) is simultaneously presented to the right eye of the subject. The resulting binocular percept is very similar to the result of the merging scheme shown in the lower right corner of Figure 3. This result was obtained as follows. First the left and right bar images (denoted by L and R, respectively) are separately encoded as Gaussian ratio pyramids $R(L)$ and $R(R)$, respectively. A third so-called binocular pyramid $R(B)$ is constructed by selecting the node with maximum absolute contrast value from the corresponding nodes in the left and right monocular pyramids. That is, for all i, j, and l we have

$$R(B)_l(i,j) = R(L)_l(i,j) \quad \text{if} \quad \|R(L)_l(i,j) - 1\| > \|R(R)_l(i,j) - 1\| = R(R)_l(i,j) \quad \text{otherwise}$$

The binocular pyramid $R(B)$ is shown in the left column of Figure 3. The right column of this figure shows the reconstructed Gaussian pyramid for the binocular percept. The bottom layer of this pyramid represents the actual binocular combination. Notice that the edges of both images are preserved in the binocular image even though they are not specifi-

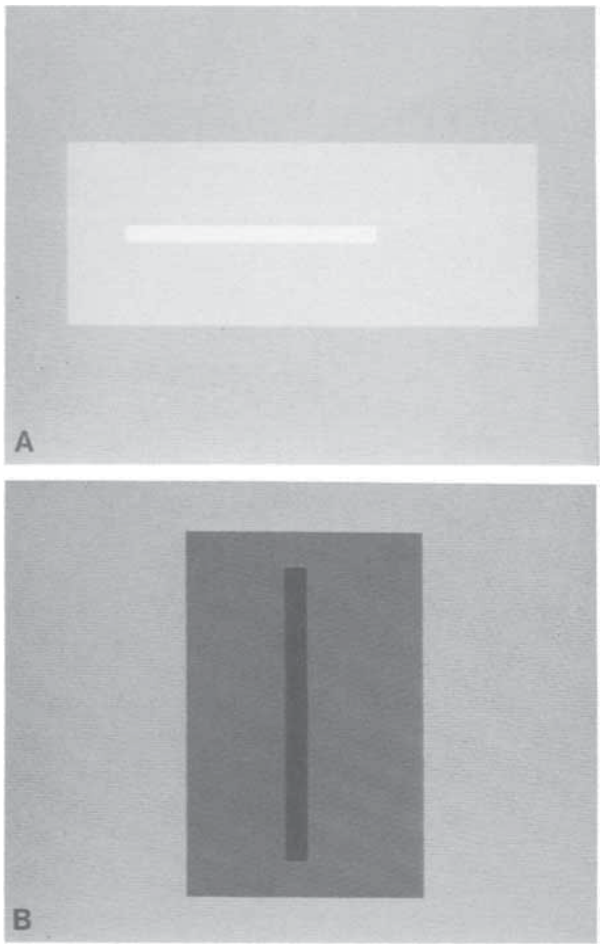


Figure 4. Two test images.

cally encoded. The shape and extent of the halo surrounding the central square represents the cumulative contribution of Laplacian filters of many spatial scales.

Figures 4a and 4b show two gray bars containing smaller bars in their interior. The absolute luminance contrast between the small and large bars in Figure 4 was chosen to be larger for the small bright bar in Figure 4a than for the small dark bar in Figure 4b. However, the absolute gray-value difference between the small and large bars was smaller for the bright bar in Figure 4a than for the dark bar in Figure 4b. The human visual system is only sensitive to local luminance contrast. Thus, if Figures 4a and 4b are presented simultaneously respectively to the left and right eye of a subject, the small bright bar of Figure 4a will dominate the perception of the small dark bar of Figure 4b in the resulting binocular percept. Figures 5 and 6 show respectively the Gaussian or size-limiting pyramid and the contrast or size-selective pyramid of both images from Figure 4. The composite contrast pyramid obtained by the

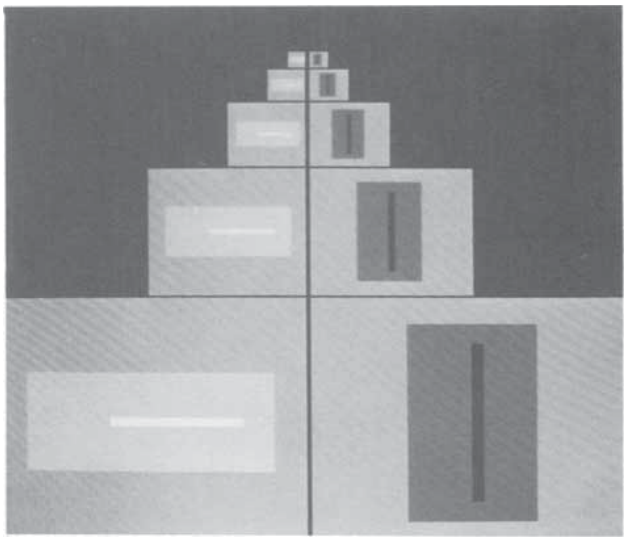


Figure 5. The Gaussian pyramids for the images of Figure 4.

maximum absolute contrast selection rule is shown on the left in Figure 7. The reconstructed composite Gaussian pyramid is shown on the right in Figure 7. Notice that the bright bar dominates the dark bar in this reconstruction in agreement with the representation by the human visual system (see Toet 1989a, c for an extensive comparison of the ratio and difference of Gaussian pyramids).

3.2 Fusion of Thermal and Visual Images

In this section we present some results of the application of the image fusion scheme on real images. First, we simultaneously recorded spatially regis-

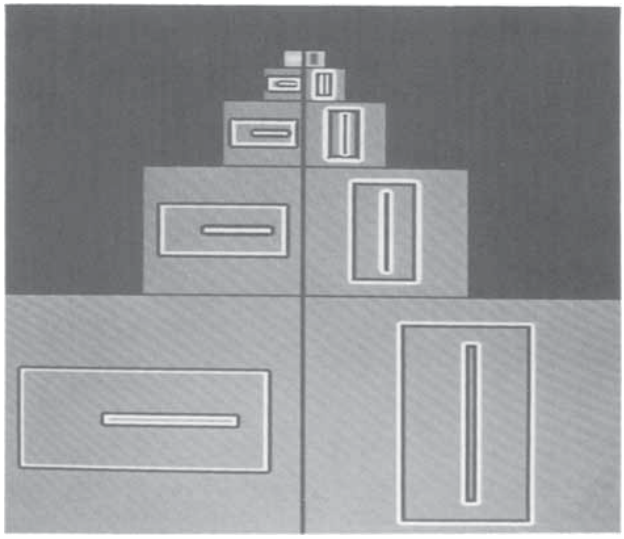


Figure 6. The Gaussian ratio pyramids for the images of Figure 4 constructed from Figure 5.

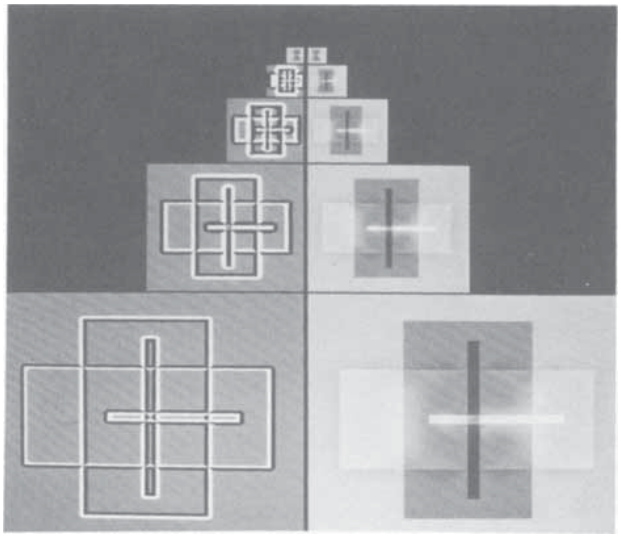


Figure 7. The composite Gaussian ratio pyramid (left column) and the reconstructed composite Gaussian pyramid (right column) for the images of Figure 4.

tered charge coupled device (CCD) and forward-looking infrared (FLIR) images on videotape. The images were thereafter digitized and brought in register. Finally, we digitally merged corresponding images using the hierarchical contrast merging scheme.

3.2.1 Aim of the experiment. 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 can sometimes easily be noticed in a thermal image. Incomplete representation of targets in thermal images may result from large temperature gradients within these objects. Also, targets in FLIR images may be hard to localize 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, for example, the remote control of vehicles, driving in hostile environments and surveillance.

3.2.2 Image acquisition. Figure 8 shows a schematic drawing of the experimental setup used to record the CCD and FLIR images. The CCD and FLIR 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 vertically erected 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) through adjusting the magnification, tilt, and direction off the CCD camera.

Prior to each recording session some shots of the reference coordinate 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 (SPIDER Working Group 1983). The bulbs were removed prior to recording objects in the scene. The signals from both cameras were recorded on synchronized videotape recorders.

3.2.3 Image merging. Figure 9 shows the CCD (Figure 9a) and FLIR (Figure 9b) images of a

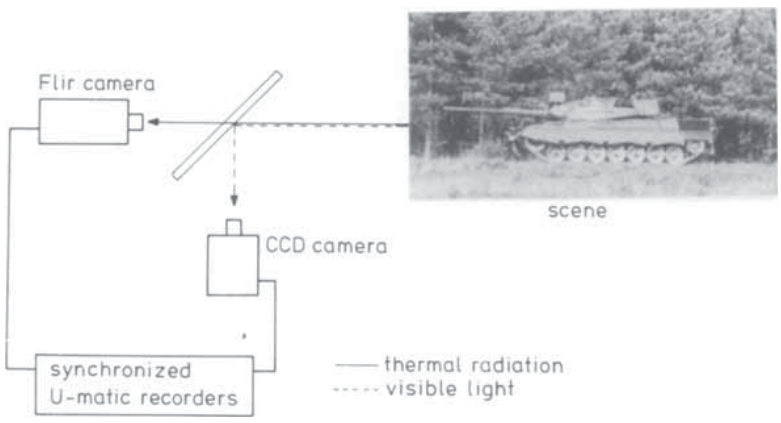


Figure 8. Schematical representation of the experimental setup: (—), thermal radiation; (---), visible light.

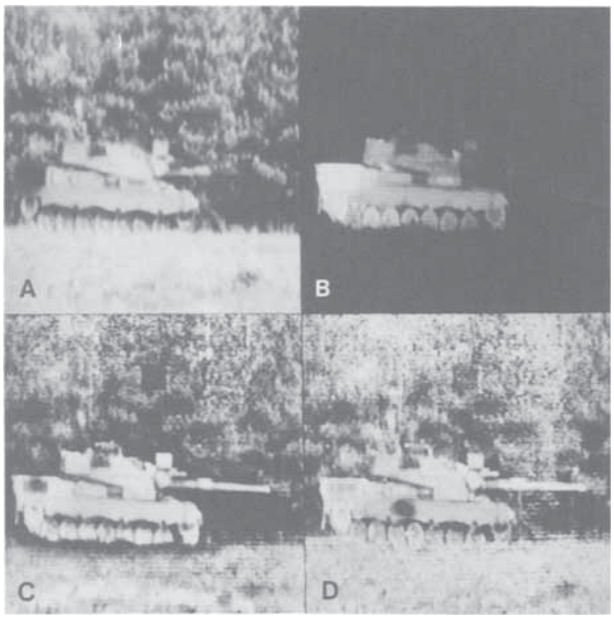


Figure 9. Original CCD (A) and FLIR (B) images with the results of the linear (C) and morphological (D) merging schemes.

tank. 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 these parts of the target have very low contrast in the visual image. However, in the thermal image these parts are clearly visible. The background, which is clearly visible in the CCD image, is nearly indistinguishable in the FLIR image.

Figure 10 shows the Gaussian ratio pyramids of the CCD (left column) and FLIR (right column) images from Figure 9. 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 Gaussian ratio pyramid of the FLIR image and only very weakly represented in the Gaussian ratio pyramid of the CCD image. However, the background that is accentuated in the Gaussian ratio representation of the CCD image is very noisy in the FLIR image.

In this example we used the maximum absolute contrast node selection rule. Figure 11 shows the composite Gaussian ratio pyramid (left column) and the reconstructed composite Gaussian ratio pyramid (right column). The bottom level of the composite Gaussian ratio pyramid represents the final result. This level is again shown in Figure 9c, together with the original input images (Figures 9a and b).

Figures 12 and 13 show respectively the ratio stacks of the morphological filtered CCD and FLIR images from Figure 9. Again, we used the maximum absolute contrast node selection rule in the merging scheme.

Figure 14 shows the composite morphological ratio stack. The final result of the morphological merging process is shown in Figure 9d, together with the original input images (Figures 9a and b) and the result of the Gaussian decomposition scheme.

Figure 9 convincingly demonstrates that the fused images contain those details from both input images that have maximum local contrast. Notice

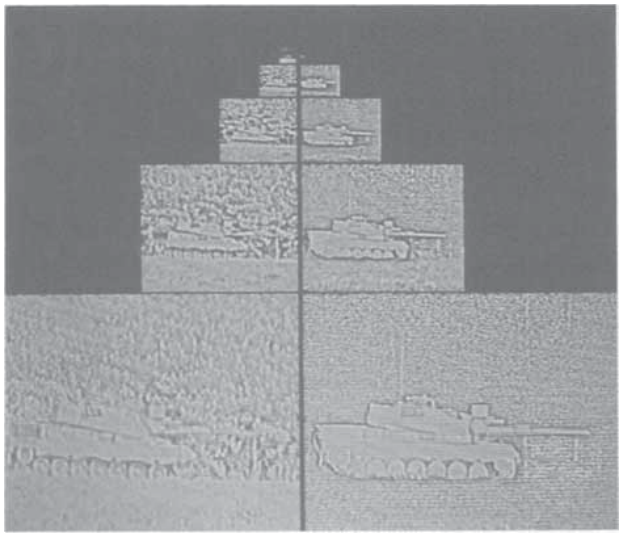


Figure 10. The Gaussian ratio pyramids of the CCD (left column) and FLIR (right column) images from Figure 9.

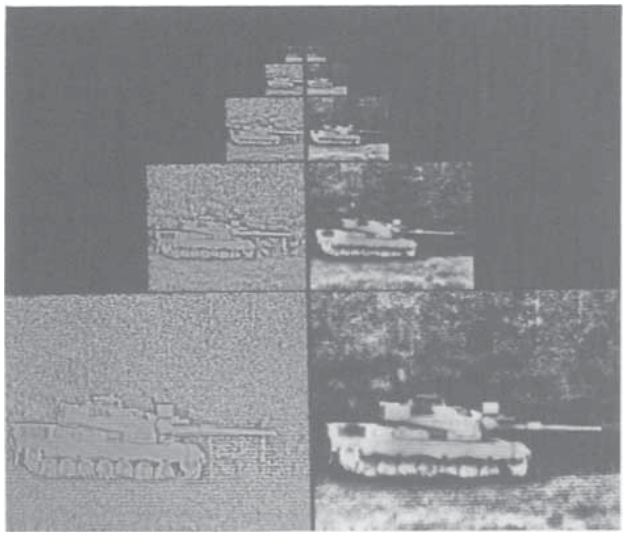


Figure 11. The composite Gaussian ratio pyramid (left column) and the reconstructed composite Gaussian ratio pyramid (right column) of the CCD and FLIR images from Figures 9a and b.

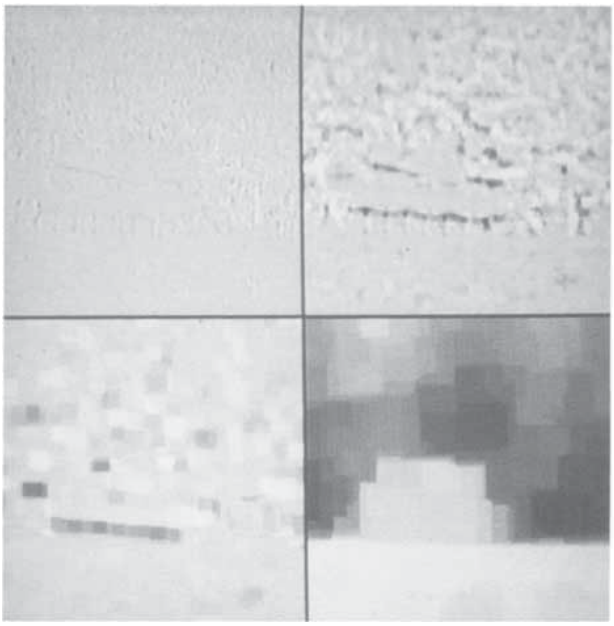


Figure 12. The ratio stack of the morphological filtered CCD image from Figure 9A.

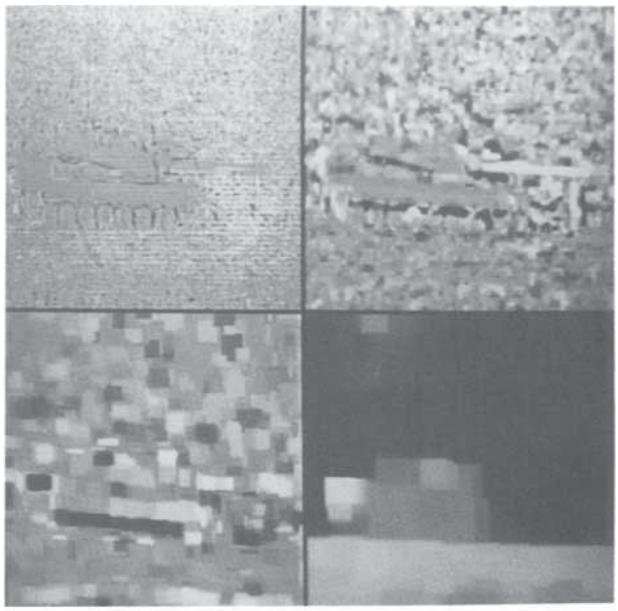


Figure 14. The composite morphological ratio stack of the CCD and FLIR images from Figures 9A and B.

that all (aforementioned) details that can only be obtained from a single image modality are clearly represented in the fused images.

Details in the composite image resulting from the morphological fusion scheme (Figure 9d) are more pronounced (better articulated) than their counterparts in the result from the linear fusion scheme (Figure 9c). This is a result of the fact that linear filters alter object intensities (blur image details), whereas morphological filters extract image details without adding a gray-scale bias.

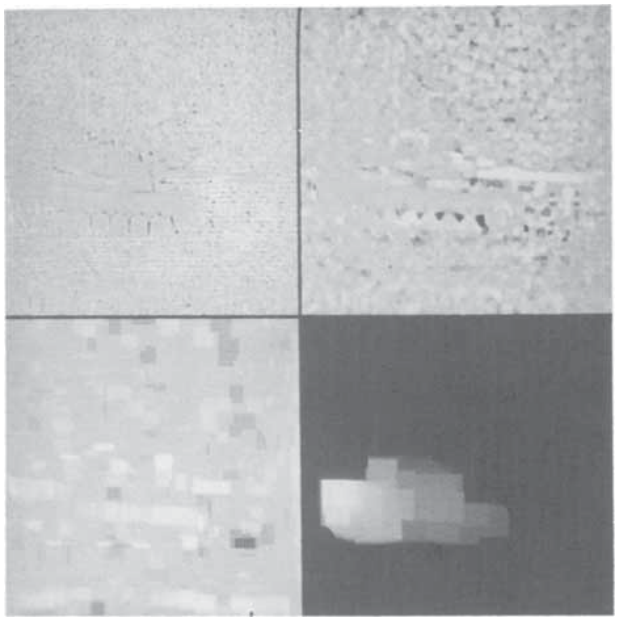


Figure 13. The ratio stack of the morphological filtered FLIR image from Figure 9B.

4 Discussion and Conclusions

This paper presents an image fusion method intended for human observation. The method is based on a hierarchical image decomposition that is constructed by repeated application of size-selective filters. The image fusion scheme preserves details of high local luminance contrast.

The superiority of the morphological multiresolution image decomposition over a conventional linear multiresolution image decomposition is demonstrated by the results of the hierarchical image fusion scheme. The images produced by the morphological fusion scheme appear more crispy than the images produced by the corresponding linear fusion scheme. This is a result from the fact that morphological filters preserve local luminance contrast, whereas linear filters blur local luminance contrast.

The morphological image decomposition scheme is well suited for real-time implementation when binary set structuring elements are used (Huang et al. 1989, Shih and Mitchell 1989). Because of the inherent congruences between the hierarchical morphological decomposition scheme and human visual perception, the method appears well suited to even-

tual integration into artificial intelligent computer vision systems.

Image fusion based on luminance ratios only fails when the mean local background luminance of all the input images is zero. A simple remedy is to add a constant background luminance to the composite image. This nonzero local background then serves as a pedestal for the contrast modulations.

The hierarchical image fusion scheme presented in this paper 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 will depend on the application (and can be anything from contrast enhancement to smoothing and thresholding). The method can be used to merge images from a variety of sensing modalities prior to display.

References

Burt PJ (1984) The pyramid as a structure for efficient computation. In: Rosenfeld A (ed) Multiresolution Image Processing and Analysis. Springer-Verlag, Berlin, pp 6–35

Burt PJ, Adelson EH (1983) The Laplacian pyramid as a compact image code. IEEE Transactions on Communications COM-31(4):532–540

Burt PJ, Adelson EH (1985) Merging images through pattern decomposition. In: Applications of Digital Image Processing VIII, Proceedings of SPIE 575, pp 173–181

Burton GJ, Haig ND, Moorhead IR (1986) A self-similar stack model for human and machine vision. Biological Cybernetics 53:397–403

Growley JL, Parker AC (1984) A representation for shape based on peaks and ridges in the difference of low-pass transform. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-6:156–170

Haralick RM, Lin C, Lee JSJ, Zhuang X (1987a) Multiresolution morphology. In: Proceedings of IEEE First International Conference on Computer Vision, IEEE Comp. Soc. Press, Washington, pp 516–520

Haralick RM, Sternberg SR, Zhuang X (1987b) Image analysis using mathematical morphology. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-9(4):532–550

Huang KS, Jenkins BK, Sawchuk AA (1989) Binary image algebra and optical cellular logic processor design. Computer Vision, Graphics and Image Processing 45:295–345

Koenderink JJ (1984) The structure of images. Biological Cybernetics 50:363–370

Koenderink JJ, Doorn AJ van (1978) Visual detection of spatial contrast; influence of location in the visual field, target extent and illuminance level. Biological Cybernetics 30:157–167

Koenderink JJ, Doorn AJ van (1982) Invariant features of contrast detection: An explanation in terms of self-similar detector arrays. Journal of Optical Society of America, 72:83–87

Maragos PA (1987) Tutorial on advances in morphological image processing and analysis. Optical engineering 26:623–632

O Ying-Lie, Toet A (1990) Mathematical morphology in hierarchical image representation. In: Proceedings of NATO ASI. The Formation, Handling and Evaluation of Medical Images. Springer-Verlag, New York. In press

Rosenfeld A (ed) (1984) Multiresolution Image Processing and Analysis. Springer-Verlag, New York

Serra J (1982) Image Analysis and Mathematical Morphology. Academic Press, New York

Serra J (ed) (1988) Alternating sequential filters. In: Image Analysis and Mathematical Morphology, Vol. 2: Theoretical Advances, pp 203–216. Academic Press, New York

Shih FY, Mitchell OR (1989) Threshold decomposition of gray-scale morphology into binary morphology. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-11(1):31–42

SPIDER Working Group (1983) SPIDER User's Manual. AIST MITI, Japan

Toet A (1989a) Image fusion by a ratio of low-pass pyramid. Pattern Recognition Letters 9:245–253

Toet A (1989b) A morphological pyramidal image decomposition. Pattern Recognition Letters 9:255–261

Toet A (1990) Morphological Multiresolution Image Representations. Report TNO-IZF Institute for Perception TNO

Toet A, Koenderink JJ, Zuidema P, Graaf CN de (1984) Image analysis-topological methods. In: DeConinck F (Ed) Information Processing in Medical Imaging. Martinus Nijhof, The Hague, pp 306–342