Enhanced Image Capture Through Fusion

Peter J. Burt and Raymond J. Kolczynski
David Sarnoff Research Center, Princeton NJ, 08543-5300

Abstract

Image fusion may be used to combine images from different sensors, such as IR and visible cameras, to obtain a single composite with extended information content. Fusion may also be used to combine multiple images from a given sensor to form a composite image in which information of interest is enhanced.

We present a general method for performing image fusion and show that this method is effective for diverse fusion applications. We suggest that fusion may provide a powerful tool for enhanced image capture with broad utility in image processing and computer vision.

1 Introduction

The development of new imaging sensors has brought with it a need for image processing techniques that can effectively fuse images from different sensors into a single composite for interpretation. To date fusion has been considered primarily as a means for presenting images to humans. For example, IR and visible images may be fused as an aid to pilots landing in poor weather, or CT and NMR images may be fused as an aid to medical diagnosis. It may be expected that fusion can become equally important in combining images, and hence in compressing source image data, for interpretation by computer vision systems.

An image fusion technique is successful to the extent that it creates a composite that retains all useful information from the source images, and does not introduce artifacts that could interfere with interpretation. Efforts to identify such techniques have only recently begun. The most direct approach to fusion is to sum and average the source images. Unfortunately this can produce unsatisfactory results. Features that appear in one source image but not in others are rendered in the composite at reduced contrast or superimposed on features from other images, as in a photographic double exposure.

Perhaps the most promising approaches to fusion that have been studied thus far are those that perform image combination in a pyramid transform domain. An image pyramid is first constructed for each source image, then a pyramid is formed for the composite image by selecting coefficients from the source image pyramids. Finally, the composite image is recovered through an inverse pyramid transform.

Several variations on pyramid-based fusion have been described. These differ in the type of pyramid transform used, and in the rules used to select transform coefficients that carry "salient" information for inclusion in the composite. The approach was first proposed as a model for binocular fusion in human stereo vision (Burt, 1984). This implementation used a Laplacian pyramid and a "choose max" selection rule that, at each sample position in the pyramid, copied that source pyramid coefficient with the maximum value to the composite pyramid. Toet (1989) proposed using a ratio of low pass pyramid and the same selection rule to fuse IR and visible image. Pavel, et. al. (1991) have used a related pyramid technique and a noise based selection rule to fuse millimeter wave sensor images with computer generated graphics. More recently fusion within a gradient pyramid was shown to provide improved stability and noise immunity (Burt, 1992).

These methods generally appear to provide good results. One limitation is in the fusion of patterns that have roughly equal salience but opposite contrast. This is a pathological case since image averaging results in pattern cancellation, and selection is unstable.

In this paper we present an extension to the pyramid approach to fusion. The modifications introduced are intended to further improve stability and noise immunity, and to address the pathological case of patterns with opposite contrast. We show diverse examples that suggest the proposed fusion approach is both general and robust.

We also suggest that fusion can be used as part of a general technique for capturing enhanced images for use in computer vision. A set of images are obtained with a single sensor while imaging conditions are sys-

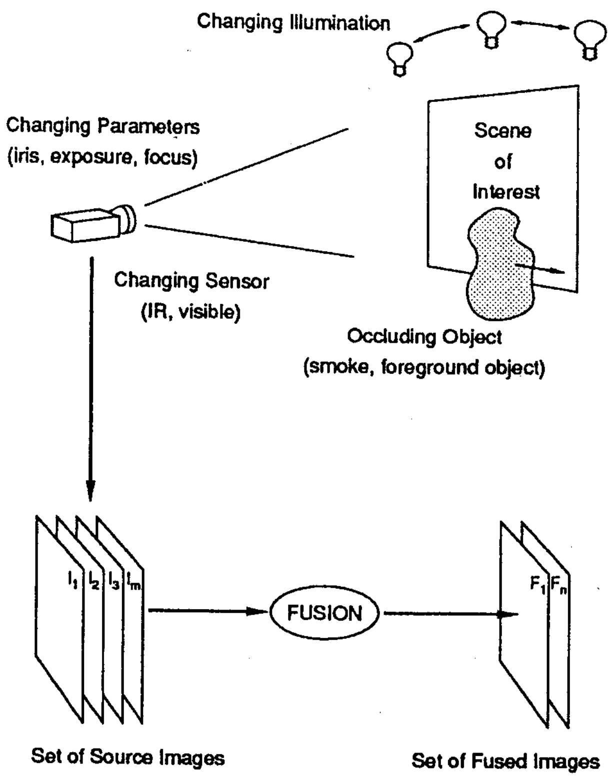


Figure 1: The fusion task.

tematically varied. These images are then fused to obtain a composite that can have greater information content than can be obtained directly from the sensor.

2 Problem Statement

A general view of image fusion is suggested in Figure 1. A set of 2 or more source images, $\mathbf{I}_1$, $\mathbf{I}_2$, ..., is obtained of a given scene viewed with different sensors or under different imaging conditions. Camera focus or exposure may be varied from image to image, or the positions of light sources may be changed. The source images may be obtained as objects move to reveal different portions of the background. In any case, each of the source images represents a partial view of the scene, and contains both "valid" and "invalid" data for the task at hand. The objective of image fusion is to combine the source images to form one composite image, $\mathbf{F}$, (or multiple composite images, $\mathbf{F}_1$, $\mathbf{F}_2$, ...) in such a way that valid data are retained from all source images, while invalid data are discarded. The combined image should also appear "natural" so that it can be readily interpreted by humans or machines using normal vision capabilities.

We assume that the source images are aligned prior to fusion. This can be achieved by obtaining all source images from the same position, and using the same sensor and optics. It can also be achieved by electronically transforming (warping) images taken from nearby positions so that they are in registration. (Automatic registration of images obtained from different sensors, or under different imaging conditions, is a challenging computer vision task in its own right, and is beyond the scope of the present paper.)

3 Pyramid-Based Fusion

The pyramid may be said to implement a "pattern selective" approach to image fusion (Burt, 1992). In effect, the composite images is constructed not a pixel at a time, but a feature at a time. Salient features are identified in each source image, then are copied, intact, to the composite image. Less salient features that may partially mask the more salient features are discarded. In this way features included in the composite are rendered at full contrast, and double exposure artifacts are avoided.

According to this view, the pyramid transform decomposes each source image into a set of component patterns, the basis functions of the transform. Pyramid basis functions are compact, self similar, and of many scales (i.e., wavelets). Pattern selection is then performed at each sample position of the pyramid for the composite image: the sample value at a given position is simply assigned the value of the corresponding sample in the source pyramids that is judged to have the highest salience. An inverse transform for the composite pyramid then combines the selected patterns in a smooth, natural-looking way.

A general framework for pyramid based image fusion is shown in Figure 2. This differs from earlier implementations in that it includes a local match measure, $\mathbf{M}_{AB}$, as well as area based saliency measures, $\mathbf{S}_A$ and $\mathbf{S}_B$.

To simplify notation we show the case in which there are just two source images, $\mathbf{A}$ and $\mathbf{B}$, and a single composite image, $\mathbf{C}$, although the methods described can be extended to larger numbers of source and composite images.

Image fusion is carried out in four steps: (i) construct a pyramid transform for each source image, (ii) compute match and saliency measures for the source images at each pyramid sample position, (iii) combine source pyramids to form a pyramid for the composite image, and (iv) recover the composite image through an inverse pyramid transform.

3.1 Pyramid transform

The fusion process begins with the construction of image pyramids $\mathbf{D}_A$ and $\mathbf{D}_B$ for the two source im-

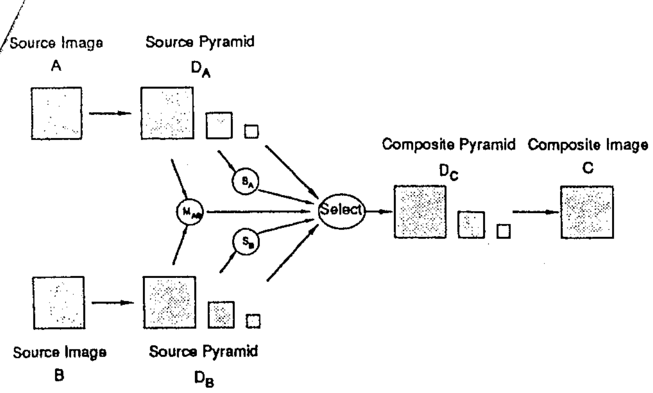


Figure 2: Fusion framework.

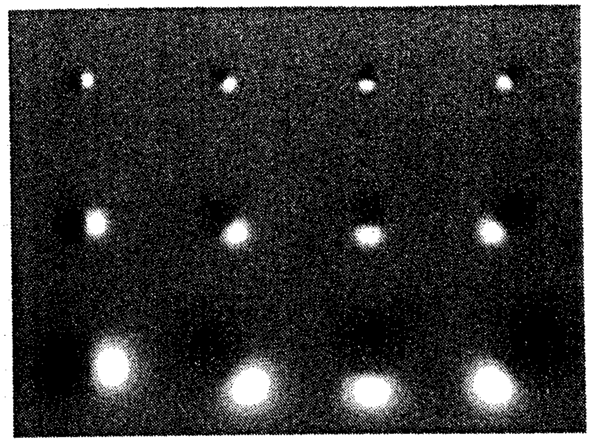


Figure 3: Gradient pyramid basis functions.

ages. Let $\mathbf{D}_I(mnk\ell)$ be the pyramid transform of image $\mathbf{I}(ij)$. The indices ij indicate sample position in the original image. The indices $mnk\ell$ indicate sample position, level, and orientation in the pyramid. To simplify notation let $\vec{m} = mnk\ell$ indicate a sample location in the pyramid.

Associated with each pyramid sample $\mathbf{D}(\vec{m})$ there is a basis function $\Psi(ij|\vec{m})$. Thus

$$I(ij) = \sum_{\vec{m}} \mathbf{D}_I(\vec{m})\Psi(ij|\vec{m}). \qquad (1)$$

In the present paper we implement fusion within a gradient pyramid. In this case basis functions are gradient-of-Gaussian patterns. Basis function for three successive pyramid levels and all four orientations are shown in Figure 3. Each basis function is a shifted, scaled and rotated copy of a single prototype pattern. The gradient pyramid transform is given in the Appendix.

3.2 Match and salience measures

In pattern-selective fusion the composite image is assembled from selected component patterns of the source images. In the pyramid implementation, the pyramid basis functions serve as the component patterns.

Here we define two distinct modes of combination: *selection* and *averaging*. At sample locations where the source images are distinctly different, the combination process selects the most salient component pattern from the source pyramids and copies it to the composite pyramid, while discarding less salient patterns. But at sample locations where the source images are similar, the process averages the source patterns. Again, selection avoid double exposure artifacts in the composite. Averaging reduces noise and provides stability where source images contain the same pattern information.

Pattern selective image fusion is guided by two measures: a match measure that determines the mode of combination at each sample position (selection or averaging), and salience measures that determine which source pattern is chosen in the selection mode.

Salience Measure: The salience of a particular component pattern is high if that pattern plays a role in representing important information in a scene. Salience is low if the pattern represents unimportant information, or, particularly, if it represents corrupted image data. Specific measures of salience may be based on criteria for the vision task at hand. In general a pattern may be expected to be important if it is relatively prominent in the image. Thus the amplitude of a pattern can be taken as a generic measure of its salience. Alternatively, the contrast of the component pattern with neighboring patterns can provide that measure. We define salience at sample $\vec{m}$ as a local energy, or variance, within neighborhood $\mathbf{p}$:

$$\mathbf{S}_I(m,n,k,\ell) = \sum_{\acute{m},\acute{n}} \mathbf{p}(\acute{m},\acute{n})\mathbf{D}_I(m+\acute{m},n+\acute{n},k,\ell)^2. \qquad (2)$$

In practice the neighborhood $\mathbf{p}$ is small, typically including only the sample itself (point case) or a 3 by 3 or 5 by 5 array of samples centered on the sample (area case).

Match Measure: The match measure is used to determine which of the two combination modes to use at each sample position, selection or averaging. The relative amplitudes of corresponding patterns in the two source pyramids can be used as a measure of their similarity, or match. Alternatively, the correlation between pyramids in the neighborhood of the source patterns can provide that measure. Here we define the match at sample $\vec{m}$ as a local normalized correlation within neighborhood $\mathbf{p}$:

$$\mathrm{M}_{AB}(\vec{m}) = \frac{2\sum_{\acute{m}\acute{n}} \mathbf{p}(\acute{m}\acute{n})\mathrm{D}_A(\vec{m}+\acute{m}\acute{n})\mathrm{D}_B(\vec{m}+\acute{m}\acute{n})}{\mathrm{S}_A(\vec{m})+\mathrm{S}_B(\vec{m})}. \tag{3}$$

Again, the neighborhood $\mathbf{p}$ may include only the given component pattern (point case) or it may include a local array of components (area case). $\mathbf{M}_{AB}$ has value 1 for identical patterns, value -1 for patterns that are identical except that they have opposite sign, and a value between -1 and 1 for all other patters. Unlike the usual definition of normalized correlation, it has a value less that 1 for patterns that are identical except for a scale factor.

3.3 Combine source pyramids

The pattern selective step of image fusion is repeated at each pyramid sample position. If the match measure between images is low at a given position, then the coefficient from the source pyramid with the highest salience is copied to the composite pyramid. If the match measure is high, then coefficients for both source pyramids are averaged to obtain the value inserted in the composite pyramid.

This combination rule can be stated as a weighted average in which weights depend on the match and saliency measures. At each position $\vec{m}$ we assign weights $\mathbf{w}_A$ and $\mathbf{w}_B$ to the source images, and define the combined result as

$$\mathrm{D}_C(\vec{m}) = \mathbf{w}_A(\vec{m})\mathrm{D}_A(\vec{m}) + \mathbf{w}_B(\vec{m})\mathrm{D}_B(\vec{m}). \tag{4}$$

The functional relationship between the weights and salience and match measures can take many forms within the present framework. It is only assumed that if the similarity is low, so the match measure is below a threshold α, then the weights are 1 and 0 (selection mode). On the other hand, if the similarity is high, with correlation near 1, then weights are roughly .5 and .5 (averaging mode). The weights increase and decrease monotonically between these extreme cases. In all cases the larger weight is assigned to the source pattern with larger salience value.

For the present we adopt a linear transition between the extreme cases noted above. At each position $\vec{m}$, if $\mathrm{M}_{AB} \leq \alpha$ then $\mathbf{w}_{min} = 0$ and $\mathbf{w}_{max} = 1$, else if $\mathrm{M}_{AB} \geq \alpha$ then:

$$\mathbf{w}_{min} = \frac{1}{2} - \frac{1}{2}\left(\frac{1-\mathrm{M}_{AB}}{1-\alpha}\right) \quad \text{and} \quad \mathbf{w}_{max} = 1 - \mathbf{w}_{min}. \tag{5}$$

The larger weight, $\mathbf{w}_{max}$, is assigned to the source image with the larger salience: If $\mathrm{S}_A > \mathrm{S}_B$ then $\mathbf{w}_A = \mathbf{w}_{max}$ and $\mathbf{w}_B = \mathbf{w}_{min}$, else $\mathbf{w}_A = \mathbf{w}_{min}$ and $\mathbf{w}_B = \mathbf{w}_{max}$.

3.4 Inverse Pyramid Transform

The final step in fusion is an inverse pyramid transform in which the combined image, $\mathbf{C}$, is recovered from its pyramid representation, $\mathbf{D}_C$. The inverse transform for the gradient pyramid is given in the Appendix.

4 Examples

Three examples will now be given to illustrate the fusion process defined above. While these relate to quite different applications, all use the same choice of fusion parameters. Match and saliency measures are computed within a 3 by 3 neighborhood ($\mathbf{p}$) and with the transition α between selection and fusion modes equal to .85.

4.1 Multi-sensor

Figure 4 shows the fusion of images obtained with different sensors. In this case a visible image of an airplane taxing on an airport runway, Fig. *a*, is fused with an IR image of the same scene, Fig. *b*. Note that each source shows certain aspects of the scene that are not visible in the other source.

Image fusion by simple pixel averaging and by pyramid based pattern selection are shown in Figures *c* and *d* respectively. Note that the pixel based method has a "muddy" appearance compared to the pattern based method. This is due primarily to the fact that averaging results in reduced contrast for all patterns that appear in only one source. Feature contrast is maintained in the pattern selective approach, and all significant features from both sources are retained in the composite.

Figures *e* and *f* show the relative weights given to the IR and visible source images in obtaining the fused result in Figure *d*. Gray indicates averaging while lighter darker indicate selection (weights near 1) and lighter regions rejection (weights near 0).

4.2 Multi-exposure

Figure 5 shows the fusion of images obtained with a single camera, but with different settings of the exposure. Here the source images, *a* and *b*, are of an outdoor scene that might be encountered in autonomous driving. The scene includes regions of bright sunlight and deep shadow that could not be imaged simultaneously due to limitations in the dynamic range of the camera. Figure *a* was obtained with exposure adjusted for the dark regions of the scene, and reveals a log in the shadow of trees. Figure *b* was obtained with exposure adjusted of the bright regions of the scene, and shows a grassy region in the foreground. The fused result,

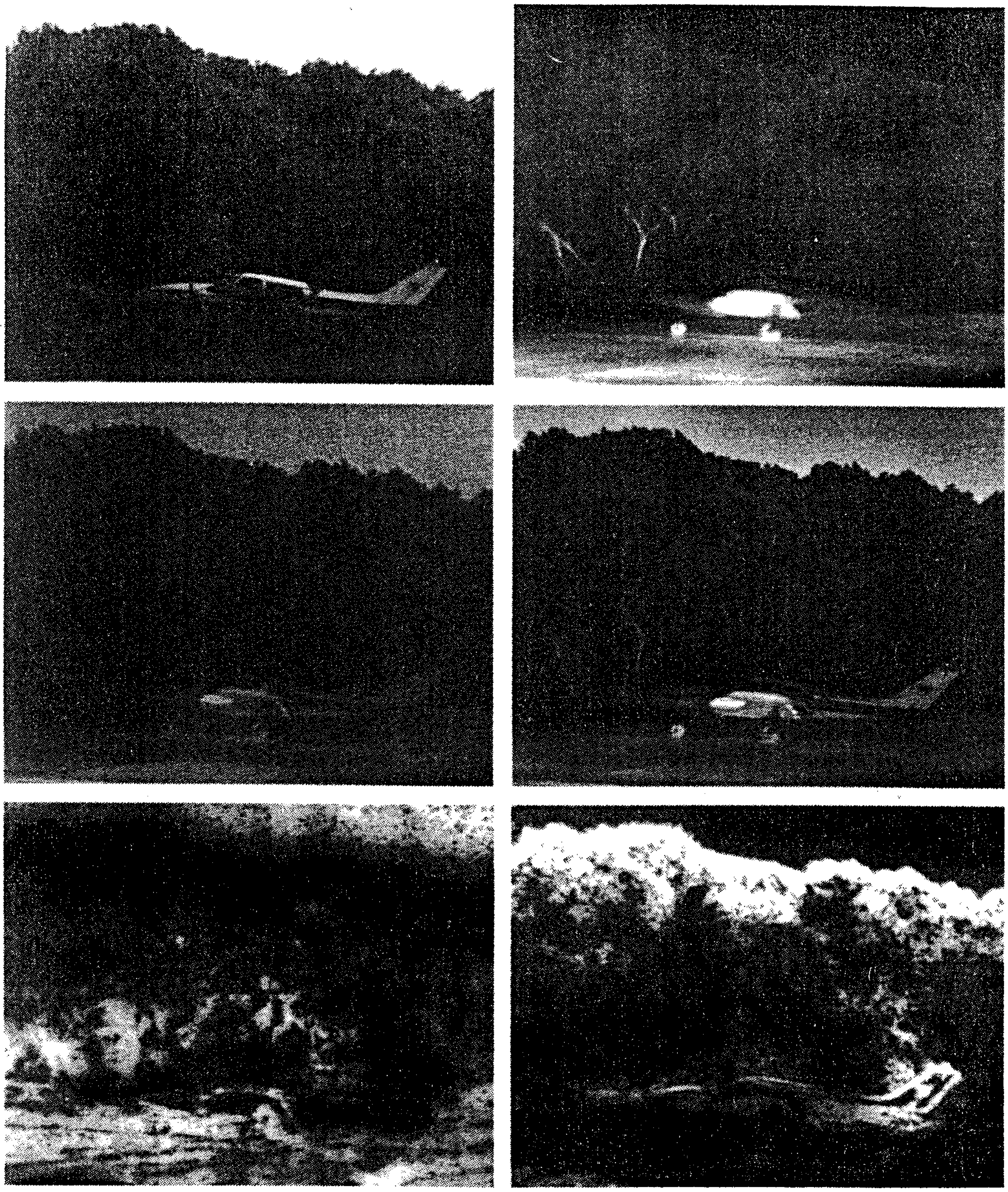

a b c d e f

Figure 4: Fusion of IR and visible images. (a) IR source image, (b) visible source image, (c) combined image obtained through pixel averaging, (d) combined image obtained through pyramid based pattern selective fusion, (e) a representation of weights given to the visible source image, and (f) weights given to the IR image in the pattern selective approach. (The source images in this example, Figs. *a* and *b*, were provided to us by the Westinghouse Electric Company.)

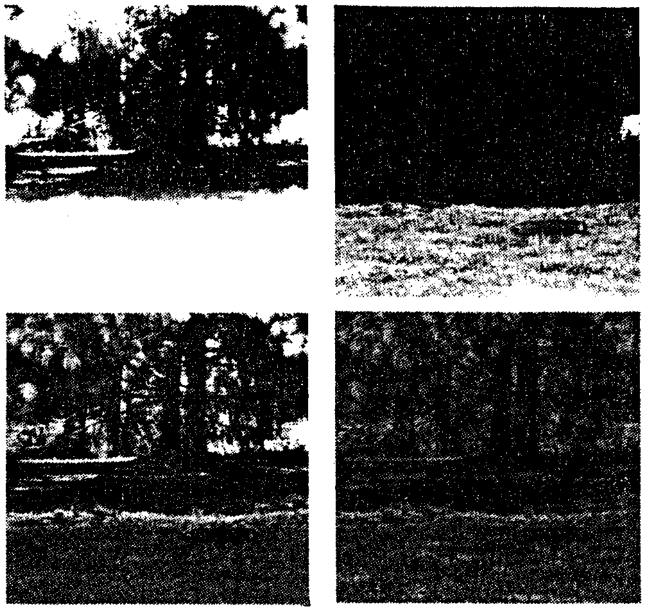


Figure 5: Fusion to extend sensor dynamic range. (*a* and *b*) Source images obtained with different camera exposure settings to observe patterns in shaded regions (*a*) and bright, sun lit regions (*b*). (*c*) The fused image includes detail from both regions. (*d*) Pyramid sample values are normalized and quantized to just 4 bits to demonstrate that a broad dynamic range scene can be represented by a narrow dynamic range signal without loss of critical detail.

Figure *c*, includes all of these regions within a single image. Figure *d* is the same as *c* except that pyramid sample values have been further normalized and quantized to just four bits. (The reconstructed image, however, extends over the full 8 bits of the original.) This technique can be used to compress a large dynamic range signal (in excess of 8 bits in this case) to a much smaller dynamic range signal while retaining texture and other pattern structure.

4.3 Multi-focus

An early application of image fusion was to extend the effective depth of field of a camera (Adelson, 1984). We show an example of this application with the fusion process defined here.

Figure 6 shows the fusion of images obtained with a single camera, but with different settings of the focus. The two source images, Figures *a* and *b*, are of a person sitting in front of a computer screen with the camera focused first at a relatively short distance, so the hands and key board appear sharp, and second at roughly twice the distance, so the person's head appears in focus. Figure *c* shows the result of fusing these two images. Note that both hands and face now appear in focus. Future workstations may include cameras to provide a means of visual input. This example suggests how such systems can maintain an extended depth of field. The technique may be more important, however, in industrial inspection applications where the need to image objects at very short distances frequently leads to difficulties in maintaining depth of field. Additional images might be obtained to extend the depth of field still further.

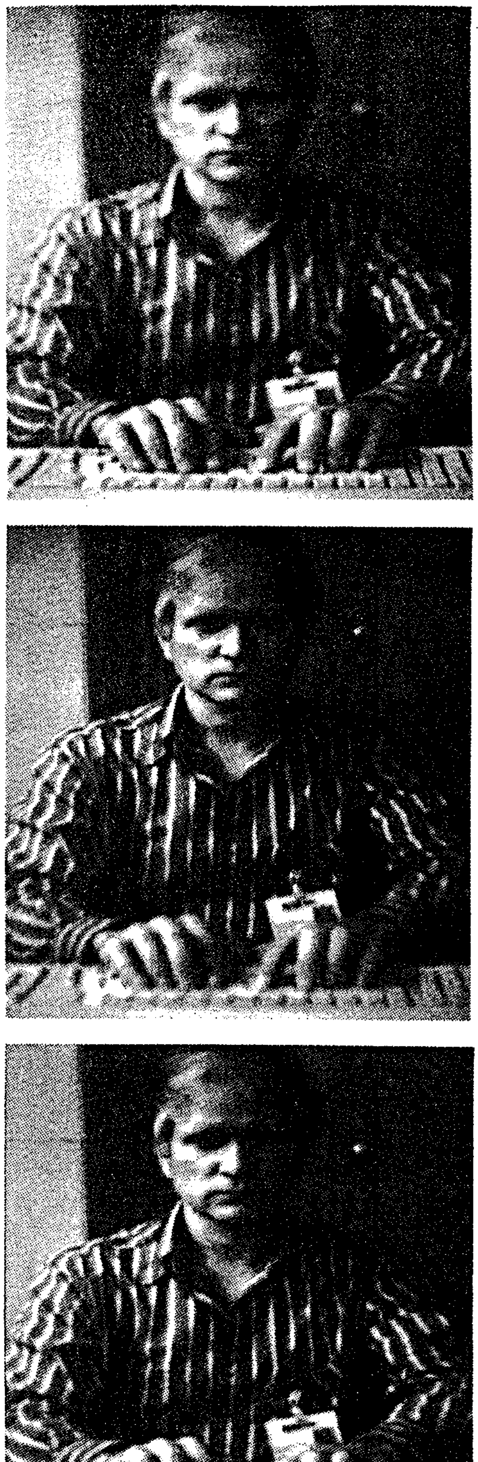
5 Performance

Ultimately the performance of a fusion process must be measured as the degree to which fusion enhances a viewers ability to perform particular tasks, such as aircraft landing or medical diagnosis. We have not yet attempted this type of task-specific evaluation.

Some performance characteristics of a fusion process can be measured independently of task. We have shown, for example, that the gradient pyramid provides more stable results than the Laplacian when used in video sequences and in the presence of motion or noise (Burt, 1992). This greater stability of the gradient is likely to be due in part to the fact that it represents the image in terms of four times as many basis functions as the Laplacian, so that any inappropriate frame-to-frame changes in the selection at any one sample position will have proportionally less affect on the fused result. In addition it is likely to be due to the fact that the gradient pyramid coefficients take their maximum values at the locations of edge patterns in the corresponding source images, while the Laplacian coefficient take their maximum value either side of an edge, and zero value at the edge itself. When sample amplitude is used as the measure of salience (as in past implementations) the gradient is likely to be most stable along edge features.

In the present implementation we define salience in terms of local variance, so that selection is likely to be less sensitive to shifts of patterns in the scene or to noise that is uncorrelated within the local regions used to define variance.

We have also found that the use of the match measures addresses another problem with prior implementations that used salience only. This problem arises when the source image patterns at a given location have roughly equal salience but opposite contrast (sign). Prior methods often averaged source patterns in such cases, resulting in loss of contrast or even cancellation in the composite image. Here reversed contrast is detected by the match measure, and component patterns are combined by selection rather than averaging. Contrast is retained.



a

b

c

Figure 6: Multi-focus example. (a and b) Source images obtained with a camera lens set to focus at different distances. (c) A fused image has an extended depth of field.

The fusion algorithm has two parameters: the size of the neighborhood $\mathbf{p}$ used in computing salience and match measures, and the parameter α that determines how similar two pattern must be for the combination to be one of averaging rather than selection. These parameters may be expected to influence contrast and stability of the fused results.

We have examined fusion performance for a range of parameter values with the following general results: First, it was found that the fusion algorithm is remarkably insensitive to changes in these parameters. Except in extreme cases results are virtually identical. Second, when the size of the neighborhood $\mathbf{p}$ is large, e.g., 5 by 5, there is some loss of contrast. This is also true when the fusion parameter α is small, e.g., .5 or less. Third, there is some loss in shift invariance and noise immunity when the size of support is reduced to just the sample itself, and α is large, e.g., 1. Thus there is a tradeoff between sharpness and shift invariance. While further study will be required to identify "optimal" parameters, it should be stressed again that these are subtle effects, and that the main result of the present study is that performance is good over a broad range of these parameters.

The general results of this study can be explained as follows. Increasing the size of the support used in computing the salience measure decreases the contribution of individual sample values, and hence decreases the likelihood that noise or shifts in the image pattern relative to the sample grid will result in a reversal in selection. Decreasing α has the effect of increasing the fraction of the component patterns that are combined through averaging rather than selection. Averaging also reduces sensitivity of the fusion process to small changes in sample values.

6 Composite Imaging

Image fusion has been viewed primarily as a means for extending the range of a sensor beyond its normal physical limitations, or for circumventing uncontrollable factors in the environment that interfere with imaging. Thus fusion can extend the depth of focus of a camera, or extend its dynamic range. And fusion can be used to remove foreground moving objects from a background scene of interest.

However it appears that there may be an equally important role for fusion in obtaining enhanced images even when the full signal range of interest can be recorded at once with a single sensor. For example, sensors are now becoming available that are sensitive both in the visible and IR portions of the spectrum. It might be assumed that use of such a sensor would eliminate the need for fusing IR and visible images as in our

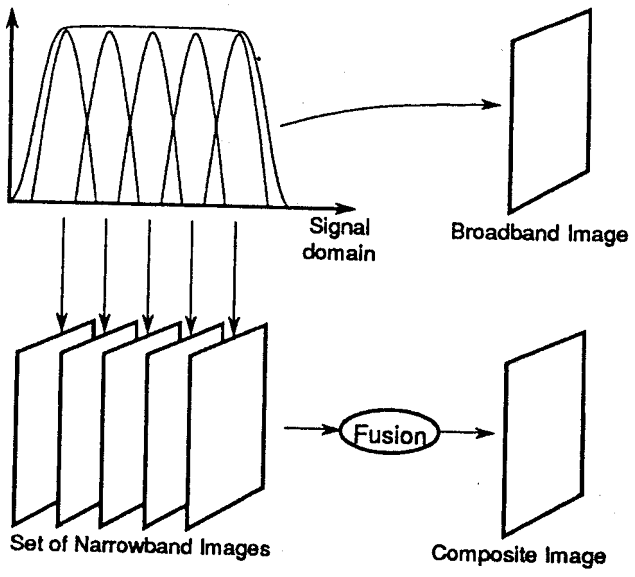


Figure 7: Extending image content through composite imaging.

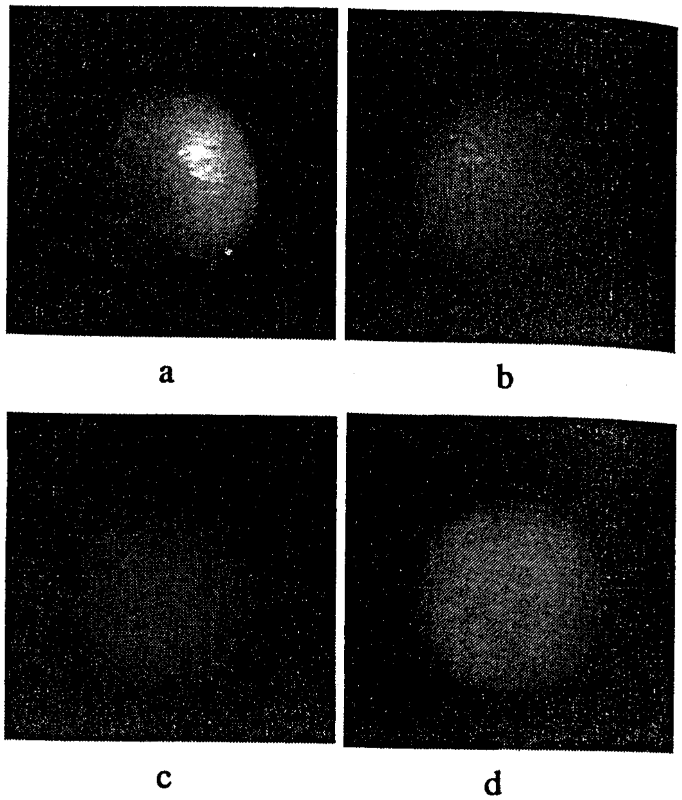


Figure 8: Multi-illumination example. (*a* and *b*) Two of four source images obtained with different light sources. (*c*) Pixel based fusion obtained by averaging all four source images. A single image obtained with all four lights on would produce a very similar result. (*d*) A fused image obtained with the pattern selective method and with highlights removed.

example, Figure 4. However the image obtained from such a "broadband" sensor may not contain as much useful information as an image obtained by fusing two separate "narrowband" IR and visible images. The broadband image would be similar to that obtained by averaging IR and visible images, Figure 4c, which appears less clear than the fused image, Figure 4d.

The potential advantage of obtaining a broadband image through the fusion of a set of narrowband images arises from the fact that features of interest in a scene can often be enhanced using controlled narrowband sensing. The imaging conditions used in obtaining each narrowband image can be controlled to enhance a particular set of features in the scene. Imaging conditions are adjusted from image to image to obtain a set of narrow band images that highlight all features of interest in turn, while avoiding background clutter when possible. Then fusion combines these results in such a way that each feature has the contrast that it had under most favorable narrowband imaging conditions. In broadband sensing, feature contrast is diminished as results obtained with most favorable signal conditions are effectively averaged with results obtained under less favorable imaging conditions. Just as narrowband images enhance signal components of interest, they also isolate corrupting signals (e.g., spectral highlights). The fusion process preserves the desired signal components as they appear under best image conditions, while isolating and discarding corrupting signal components.

In short, it may often be possible to obtain higher quality images for analysis by fusing a set of image obtained under deliberately restricted imaging conditions than by direct image capture with a broadband sensor. This ***composite imaging*** technique is summarized in Figure 7.

Composite imaging is illustrated in Figure 8. Here four (narrowband) images of an orange were obtained using four different light sources arranged in quadrants. Two of these source images are shown in Figs. *a* and *b*. Note that each enhances the surface texture in a different portion of the orange, and that each has a specular highlight. Figure *c* shows the result of averaging the four source images. Much the same (broadband) result would be obtained if a single image were obtained with all four lights on at once. Figure *d* shows the image obtained through pattern selective fusion of the four single light source images. Again we see that pixel averaging (or broadband image with all four lights on at once) results in loss of contrast in the surface texture. Surface texture is retained when pattern selective fusion is used. Note that samples that are corrupted by specular highlights have also been discarded in the fusion process, and that the composite image appears to be more uniformly illuminated with fusion than with all four lights are on at once.

7 Summary and Comments

We have presented a general approach to image fusion and have shown that it can be applied to diverse fusion tasks. Fusion is performed in a pyramid transform domain. The implementation described here extends prior algorithms in two important respects. First, a measure of pattern match has been introduced to control the mode used in image combination, selection or averaging. Second, both this match measure and the salience measures of past implementations are now defined as functions of the neighborhood of each pyramid sample rather than functions of only the sample itself. The neighborhood size can be small, 3 by 3 in the examples shown here.

These modifications address problems that are encountered with past implementations of pyramid-based fusion. In particular they provide greater shift invariance and immunity to video noise, and they provide at least a partial solution to the problem of combining components that have roughly equal salience but opposite contrast: mismatched patterns always are handled through selection, never averaging.

The fusion algorithm was found to perform well (based on visual inspection) for a range of tasks without requiring adjustment of the algorithm parameters. In fact, results were remarkably insensitive to changes in these parameters, suggesting that the procedure is both robust and generic.

We have also outlined a composite imaging technique that may provide a new and powerful tool for image capture. By fusing a set of images obtained under restricted, narrowband, imaging conditions it is often possible to construct an image that has enhanced information content (and reduce data) when compared to a single image obtained directly with a broadband sensor.

Appendix

A The Gradient Pyramid Transform

A gradient pyramid for image **I** can be obtained by applying a gradient operator to each level of its Gaussian pyramid representation. The image can be completely represented by a set of four such gradient pyramids, one each for derivatives in horizontal, vertical, and the two diagonal directions (Burt and Lee, 1988).

Here we first review definitions for Gaussian and Laplacian pyramid transforms, then we provide the additional steps needed for the gradient pyramid transforms.

A.1 Standard Gaussian and Laplacian Pyramids

Let $\mathbf{G}_k$ be the k^{th} level of the Gaussian pyramid for image **I**. Then $\mathbf{G}_0(i,j) \equiv \mathbf{I}(i,j)$ and for $k > 0$,

$$\mathbf{G}_k = [\mathbf{w} * \mathbf{G}_{k-1}]_{\downarrow 2}. \tag{A1}$$

Here **w** is the generating kernel, and the notation $[...]_{\downarrow n}$ indicates that the image array in brackets is subsampled by n.

Let $\tilde{\mathbf{L}}_k$ be the k^{th} level of the RE (reduce-expand) Laplacian pyramid (Burt, 1981). This is defined as the difference between successive levels of the Gaussian pyramid:

$$\tilde{\mathbf{L}}_k = \mathbf{G}_k - 4\mathbf{w} * [\mathbf{G}_{k+1}]_{\uparrow 2}. \tag{A2}$$

Here $[...]_{\uparrow n}$ indicates up sampling by n: $n-1$ rows and columns of zero value samples are inserted between the original rows and columns of the array in brackets. Convolution by **w** has the effect of interpolating the missing samples.

An image is recovered from its RE Laplacian pyramid by reversing these steps. Let $\hat{\mathbf{G}}$ be the Gaussian recovered from the Laplacian. Reconstruction requires all levels of the RE Laplacian, as well as the top level of the original Gaussian pyramid: $\hat{\mathbf{G}}_N = \mathbf{G}_N$, and for $k < N$,

$$\hat{\mathbf{G}}_k = \tilde{\mathbf{L}}_k + 4\mathbf{w} * [\hat{\mathbf{G}}_{k+1}]_{\uparrow 2}. \tag{A3}$$

Iterative application of this procedure yields $\hat{\mathbf{G}}_0$, the reconstructed version of the original image $\mathbf{G}_0$.

A.2 FSD Laplacian

Let $\mathbf{L}_k$ be the k^{th} level of the FSD (filter-subtract-decimate) Laplacian pyramid (Anderson, 1984). This is defined as the difference between $\mathbf{G}_k$ and the filtered copy of $\mathbf{G}_k$ prior to subsampling to form $\mathbf{G}_{k+1}$:

$$\begin{aligned} L_k &= \mathbf{G}_k - \mathbf{w} * \mathbf{G}_k \\ &= [1 - \mathbf{w}] * \mathbf{G}_k. \end{aligned} \tag{A4}$$

Note that in constructing the RE Laplacian a Gaussian level $\mathbf{G}_k$ is convolved with **w**, subsampled, then upsampled and convolved with **w** a second time before it is subtracted from itself to form $\tilde{\mathbf{L}}_k$. These results are changed only slightly if the resampling steps are skipped:

$$\begin{aligned} \tilde{\mathbf{L}}_k &\approx \mathbf{G}_k - \mathbf{w} * \mathbf{w} * \mathbf{G}_k \\ &= [1 - \mathbf{w} * \mathbf{w}] * \mathbf{G}_k \end{aligned}$$

$$= \quad [\mathbf{1}+\mathbf{w}] * [\mathbf{1}-\mathbf{w}] * \mathbf{G}_k. \qquad \text{(A5)}$$

Thus, to good approximation, levels of the FSD Laplacian can be converted to levels of the RE Laplacian through a simple filter convolution:

$$\tilde{\mathbf{L}}_k \approx [\mathbf{1}+\mathbf{w}] * \mathbf{L}_k. \qquad \text{(A6)}$$

More complex filters that can provide an arbitrarily exact conversion from $\mathbf{L}_k$ to $\tilde{\mathbf{L}}_k$ are given in Anderson (1984) and Burt and Lee (1988) but these are not required for most practical applications.

A.3 Gradient Pyramid

Assume the generating kernel $\mathbf{w}$ used in constructing the Gaussian pyramid is the 5 by 5 filter with binomial coefficients. Let $\dot{\mathbf{w}}$ be the 3 by 3 binomial filter:

$$\dot{\mathbf{w}} = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} \frac{1}{16} \qquad \text{(A7)}$$

and

(convolution

$$\mathbf{w} = \dot{\mathbf{w}} * \dot{\mathbf{w}} = \begin{bmatrix} 1 & 4 & 6 & 4 & 1 \\ 4 & 16 & 24 & 16 & 4 \\ 6 & 24 & 36 & 24 & 6 \\ 4 & 16 & 24 & 16 & 4 \\ 1 & 4 & 6 & 4 & 1 \end{bmatrix} \frac{1}{256}. \qquad \text{(A8)}$$

Let $\mathbf{D}_{k\ell}$ be the k^{th} level and ℓ^{th} orientation gradient pyramid image for $\mathbf{I}$. $\mathbf{D}_{k\ell}$ is obtained from $\mathbf{G}_k$ through convolution with gradient filter $\mathbf{d}_\ell$:

$$\mathbf{D}_{k\ell} = \mathbf{d}_\ell * [\mathbf{G}_k + \dot{\mathbf{w}} * \mathbf{G}_k]. \qquad \text{(A9)}$$

where

$$\begin{aligned} \mathbf{d}_1 &= \begin{bmatrix} 1 & -1 \end{bmatrix}, \\ \mathbf{d}_2 &= \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \frac{1}{\sqrt{2}}, \\ \mathbf{d}_3 &= \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \end{aligned}$$

and

$$\mathbf{d}_4 = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \frac{1}{\sqrt{2}}.$$

To reconstruct an image form its gradient pyramid, oriented Laplacian and FSD Laplacian pyramids are constructed as intermediate results. Note that the filter $(\mathbf{1}-\dot{\mathbf{w}})$ is identical to a sum of the oriented second derivative filters, scaled by $-\frac{1}{8}$:

$$\begin{aligned} \mathbf{1}-\dot{\mathbf{w}} &= \begin{bmatrix} -1 & -2 & -1 \\ -2 & 12 & -2 \\ -1 & -2 & -1 \end{bmatrix} \frac{1}{16} \\ &= -(\mathbf{d}_1 * \mathbf{d}_1 + \mathbf{d}_2 * \mathbf{d}_2 + \mathbf{d}_3 * \mathbf{d}_3 + \mathbf{d}_4 * \mathbf{d}_4)\frac{1}{8}. \end{aligned}$$

Note also that

$$\mathbf{1}-\mathbf{w} \;= \mathbf{1} - \dot{\mathbf{w}} * \dot{\mathbf{w}} \;= (\mathbf{1}-\dot{\mathbf{w}}) * (\mathbf{1}+\dot{\mathbf{w}}).$$

Each gradient pyramid level $\mathbf{D}_{k\ell}$ is converted to a corresponding second derivative pyramid (or oriented Laplacian) level $\vec{\mathbf{L}}_{k\ell}$:

$$\vec{\mathbf{L}}_{k\ell} = -\frac{1}{8}\mathbf{d}_\ell * \mathbf{D}_{k\ell}. \qquad \text{(A10)}$$

The oriented Laplacian pyramids are then summed to form an FSD Laplacian pyramid, $\mathbf{L}_k$:

$$\mathbf{L}_k = \sum_{\ell=1}^{4} \vec{\mathbf{L}}_{k\ell}. \qquad \text{(A11)}$$

Reconstruction is completed by converting the FSD Laplacian to the RE Laplacian, Eq. A6, and then converting the FSD to the Gaussian, Eq. A3.

References

[1] E.H. Adelson, Depth-of-focus imaging process method. *U.S. patent no. 4,661,986*, 1987.

[2] C.H. Anderson, An alternative to the Burt pyramid algorithm. *RCA correspondance*, 1984.

[3] P.J. Burt, Fast filter transforms for image processing. *Computer Graphics and Image Processing.* Vol. 16: 20-51, 1981.

[4] P.J. Burt, The Pyramid as a Structure for Efficient Computation, in *Multiresolution Image Processing and Analysis*, A. Rosenfeld, Ed., Springer-Verlag, New York, 1984.

[5] P.J. Burt, A gradient pyramid basis for pattern selective image fusion.. *Proc. of the Society for Information Display conference*, 1992 .

[6] P.J. Burt and W. A. Lee, A family of pyramid structures for multiresolution image processing. Sarnoff correspondance, 1988.

[7] M. Pavel, J. Larimer, and A. Ahumada, Sensor Fusion for Synthetic Vision, *Proc. AIAA Conf. on Computing in Aerospace 8*, Baltimore, October, 1991.

[8] A. Toet, Image fusion by a ration of low-pass pyramid. *Pattern Recognition Letters*, vol 9, pp 245-253, 1989.