Sādhanā, Vol. 25, Part 1, February 2000, pp. 37–43. © Printed in India

A study of correlation technique on pyramid processed images

M SANKAR KISHORE and K VEERABHADRA RAO

Research Centre Imarat, Vignyana Kancha, Hyderabad 500 069, India

MS received 4 August 1999

Abstract. The pyramid algorithm is potentially a powerful tool for advanced television image processing and for pattern recognition. An attempt is made to design and develop both hardware and software for a system which performs decomposition and reconstruction of digitized images by implementing the Burt pyramid algorithm.

In this work, an attempt is also made to study correlation performance on reconstructed images. That is, the reference image is taken from the original image and correlation is performed on expanded images of the same size. Similarly, correlation performance study is carried out on different pyramid-processed levels. In this paper results are presented in terms of RMS error between original and expanded images. Only still images are considered, and the hardware is designed around an i486 processor and software is developed in PL/M 86.

Keywords. Image processing; correlation; pyramid processing.

1. Introduction

Currently, research is aimed at understanding how the human visual system interprets and processes image information, that is, how the eye perceive edges, noise, colour, motion etc. in static and dynamic images. It is felt that if this process were better understood, we could develop image analysis, enhancement and coding techniques that would make it possible to extract and preserve image information that the eye is more responsive to at the expense of less-relevant information. This knowledge will guide us to a more systematic and optimized approach for solving sophisticated image-processing tasks.

Each spatial-frequency band in the human visual system is responsive to approximately one octave of spatial frequency and the centre frequency of each band differs from its neighbour by roughly a factor of two. Studies (Peter & Edward 1983) suggest that there are approximately five to seven bands or channels spanning the spatial frequency range of the human visual system. Furthermore, there is evidence that the spatial frequency information in a particular band is processed independently of the information in any other band. Also, the signal information in a particular band is computed using a small subregion of the image. Thus, spatial frequency information at a given point in an image is determined by using very localized image areas rather than the entire image field. To start with we have considered still-pictures only. A straightforward approach that uses standard, digital, multi-band, filtering techniques would require complex high-speed circuitry. Typically, the lower spatial frequency bands would be unnecessarily maintained at the original image sample density. This redundant information is a waste of both memory storage and processing time. In the following paragraph, we explain image reduction through a Gaussian pyramid.

1.1 *Gaussian pyramid*

The Burt pyramid (James and Roger 1986) is a computationally efficient algorithm for generating two-dimensional bandpass representations of images. The pyramid decomposition process is fully invertible so that the original image can be exactly recovered from its bandpass components. Invertibility makes this algorithm ideal for image-enhancing operations where individual bands can be manipulated prior to reconstruction.

The Burt pyramid is an algorithm used to separate an image, into a set of contiguous spatial frequency bands. The original size image which may be 512×512 pixels or $2^9 \times 2^9$, in size is filtered by applying a low pass filter which passes only half the frequency bands. Since the maximum frequency in the new image is half that of the original, the image may be resampled by eliminating half the samples in both directions and producing a four-fold reduction in data. The new image, which can be represented by 256×256 pixels, can itself be filtered. This process is known as a reduction operation and may be repeated in a pipeline fashion. The first step is to low-pass filter the original image g_0 to obtain image g_1. We say that g_1 is a reduced version of g_0 in both resolutions. In a similar way we form g_2 as a reduced version of g_1, and so on. Filtering is performed by a procedure equivalent to convolution with one of a family of symmetric weighting functions. The sequence of images $g_0, g_1, g_2, \ldots, g_n$ is called a Gaussian pyramid.

Suppose the image g_0 is represented initially by the array of C columns and R rows of pixels. Each pixel represents the light intensity at the corresponding image point with an integer I between 0 and $K - 1$, K being the number of gray levels. This image becomes the bottom or zero level of the Gaussian pyramid. Pyramid level 1 contains image g_1, which is a reduced or low-pass filtered version of g_0. Each value within level 1 is computed as a weighted average of values in level 0 within a 5×5 window. Each value in level 2, representing g_2, is then obtained from values in level 1 by applying the same pattern of weights. The size of the weighting function is not critical. We have selected the 5×5 pattern because it provides filtering at low computational cost. The Gaussian images are generated according to the REDUCE function, defined as

$$g_k\,(i, j) = \mathrm{REDUCE}\, g_{k-1}, 0 < K < Q \text{ for } Q \text{ level}$$

which is short hand notation for

$$g_k(i, j) = \sum_{m=-2}^{2} \sum_{n=-2}^{2} W\,(m, n)\, g_{k-1}(2i + m, 2j + n)$$

For all samples i,j where $0 < I < C_k$ and $0 < j < R_k$. Here C_k, R_k are the columns and row dimensions of image k. $W(m, n)$ is a two dimensional filtering or neighborhood window function and is constrained by

$$W(m,n) = W(m) \times W(n), \quad \text{filter separability},$$
$$W(i) = W(-i), \quad \text{filter symmetry},$$
$$\sum_{i=-2}^{2} W(i) = 1, \quad \text{normalized},$$
$$W(0) + 2\sum_{i \text{ even}} W(i) = 2\sum_{i \text{ odd}} W(i), \quad \text{equal contribution}.$$

Let $W(0) = a$; substituting this in the above set of equations, we get

$$W(1) = b = 0.25 = W(-1) \quad \text{and}$$
$$W(2) = c = 0.25 - a/2 = W(-2).$$

As explained by Peter & Edward (1983), the shape of the equivalent weighting functions converges rapidly to a characteristic form with successively higher levels of the pyramid so that only its scale changes. However, this shape does depend on the choice of $W(0) = an$ in the generating kernel. Characteristic shapes for four choices of a are explained by Peter & Edward (1983). The equivalent weighting functions are Gaussian-like, when $a = 0.4$. When $a = 0.5$ the shape is triangular; when $a = 0.3$ it is flatter and broader than the Gaussian. With $a{=}0.6$ the central positive mode is sharply peaked and is flanked by small lobes. The equation can be written as a matrix product

$$g_k\,(i,j) = W^T G W, \tag{1}$$

where W is a 1×5 and G is a 5×5 matrix, and

$$W = [W(-2)W(-1)W(0)W(+1)W(+2)]^T.$$

The EXPAND process is repeated for the entire set of Gaussian images according to

$$g_{k,n}(i,j) = 4\sum_{m=-2}^{2}\sum_{n=-2}^{2} W(m,n)g_{k,n-1}[(I-m)/2,(j-n)/2] \tag{2}$$

for nodes i,j where $0 \ll i \ll C_{k-n}$ and $0 \ll j \ll R_{k-n}$, and only for integer values of $(i-m)/2$ and $(j-n)/2$.

The correlation technique is selected and explained below to study the degradation effect on the reduction/expansion process.

1.2 *Correlation algorithm* (Anuta 1970; Pratt 1974; Boland et al 1977)

Extensive study of the literature qualitatively indicates that the correlation algorithm (direct method) is best for scene-matching, and low signal-to-noise ratio images and hence this method is selected for the present analysis. This template-matching technique is based on the classical correlation measure. The normalized correlation between reference R and search window Sij is computed for all (i, j).

$$C(i,j) = \sum\sum R(l,m)Sij(l,m) \Big/ \left[\sum\sum R^2(l,m)Sij^2(l,m)\right]^{1/2} \tag{3}$$

where $\sum\sum$ indicates $\sum_{l=1}^{M}\sum_{m=1}^{M}$.

The peak of $C(i, j)$ gives the point of registration $C(i, j) = 1$ for an exact match. Normalization of correlation function is essential because the unnormalized correlation function has the disadvantage that it can give a false peak even when the reference lies in the search image without any distortion when there are some areas in the search image with very high gray levels as compared to the area containing the reference image. To overcome this problem, each window is normalized by the RMS value of that window.

2. System design and analysis

The hardware has been designed and developed using an i486 DX 25 MHz processor. Internal cache memory is used to increase the processing speed. The system has the required memory for storing data/results and ROM space for storing firmware. It has two serial links with other subsystems. The software has been developed in PL/M 86 language (Intel user's guide 1985).

The application program is developed to generate reduced levels of images by applying (1) and expanded levels of images by applying (2). A visible CCD camera is connected to the system. Still-pictures are captured and stored in the high-speed memory. The CPU generates both the reduced and expanded versions of images for a particular value of the equivalent weighting function and transmits them to the PC through a serial link for analysis.

For the purpose of the present analysis, a 128×128 size image is considered. Reduced images are generated at level 1 (65×65), level 2 (33×33), level 3 (17×17), level 4 (9×9) and level 5 (5×5) from the 128×128 image by using the Gaussian pyramid algorithm and expanded back to 128×128 for a particular weighting function. Similarly, reduced/expanded images have been generated with other weighting functions as when a=0.2, 0.3, 0.4, 0.5 and 0.6.

$Lr(l, m)$ is the pixel value of the original image and $Lre(l, m)$ is the corresponding pixel value of the expanded image obtained from reduced images. The RMS error is given by

$$\text{RMS error} = \left[(1/M^2)\sum_{l=1}^{M}\sum_{m=1}^{M}[Lr(l,m) - Lre(l,m)]^2\right]^{1/2}. \tag{4}$$

In the present case M=128. This study has been carried out on a number of images for each level and each weighting function, and the RMS errors obtained are tabulated in tables 1 to 4. To study the correlation algorithm performance on pyramid processed images, a reference image of size (16×16) is taken from the original 128×128 image and an attempt is made to match it on the expanded 128×128 size of the image by applying direct correlation algorithm, (3). These results are also tabulated in the tables 1 to 4. In tables 1 to 4, @ – represents correct correlation match has taken place, # – represents correlation mismatch. The first value @ 11.79 under column I of table 1 indicates that the match has occurred and the RMS error obtained with (4) is 11.79.

As the information content in the reduced image of size is 5×5 negligible, the 5×5 image case is not considered. The direct method is based on the minimization of mean square error, which is a better criterion in the presence of noise than the absolute error criterion. Between the direct method and the modified direct method, the latter performs better generally because it responds to variations of pixel intensities rather than the actual

Table 1. RMS error at level – 1.
128 × 128 (original) ⇒ 65 × 65 (reduced) ⇒ 128 × 128 (expanded); I – $a = 0.2$, $b = 0.25$, $c = 1.5$; II – $a = 0.3$, $b = 0.25$, $c = 0.1$; III – $a = 0.4$, $b = 0.25$, $c = 0.05$; IV – $a = 0.5$, $b = 0.25$, $c = 0.1$; V – $a = 0.6$, $b = 0.25$, $c = -0.05$

Picture Image no.	I	II	III	IV	V
1	@ 11.79	@ 11.83	@ 11.93	@ 12.30	@ 12.96
2	@ 11.59	@ 11.367	@ 11.40	@ 11.86	@ 12.77
3	@ 9.21	@ 8.94	@ 8.86	@ 8.96	@ 9.355
4	@ 10.37	@ 10.22	@ 10.21	@ 10.358	@ 10.75
5	@ 12.01	@ 11.78	@ 11.83	@ 12.03	@ 12.27
6	@ 7.18	@ 6.72	@ 6.535	@ 6.50	@ 6.99
7	@ 7.72	@ 7.44	@ 7.34	@ 7.42	@ 7.68
8	@ 17.41	@ 17.47	@ 17.60	@ 17.93	@ 19.27
9	@ 15.77	@ 15.97	@ 16.25	@ 17.07	@ 17.64
10	@ 11.77	@ 11.44	@ 11.53	@ 11.76	@ 12.60

@ – correct correlation match

values of the pixel intensities and this is a better criterion for problems of pattern recognition. However, the results seem to be data dependent sometimes. In level 1, all images (expanded) have good contrast and the reference image is registered at the exact place. In the case of defocused images, error increases linearly with weighting functions. In the case of focused images, errors follow a parabolic shape. The dip may correspond to II or I. The minimum error is mostly at II or the error difference between II and III and II and IV is very little. Hence optimum results can be achieved for III. In level 2, in all cases, the correct match has taken place, but in the case of defocused images, registration has taken place from III, IV and V. In these cases RMS error increases linearly. In level 3, in general, in all images no match has occurred and IV and V are better because of the edge preservation. Level 4 is not suitable for the match type of application.

Table 2. RMS error at level – 2.
128 × 128 (original) ⇒ 65 × 65 ⇒ 33 × 33 ⇒ 65 × 65 ⇒ 128 × 128 (expanded). I – $a = 0.2$, $b = 0.25$, $c = 1.5$; II – $a = 0.3$, $b = 0.25$, $c = 0.1$; III – $a = 0.4$, $b = 0.25$, $c = 0.05$; IV – $a = 0.5$, $b = 0.25$, $c = 0.1$; V – $a = 0.6$, $b = 0.25$, $c = -0.05$

Picture Image no.	I	II	III	IV	V
1	@ 18.13	@ 19.36	@ 19.80	@ 20.49	@ 21.90
2	@ 18.61	@ 18.77	@ 19.18	@ 19.76	@ 20.99
3	@ 20.40	@ 20.80	@ 21.24	@ 22.00	@ 23.50
4	@ 19.75	@ 20.27	@ 20.92	@ 22.07	@ 23.80
5	@ 19.27	@ 19.80	@ 20.49	@ 21.30	@ 22.62
6	@ 20.48	@ 20.96	@ 21.70	@ 22.60	@ 24.46
7	@ 24.25	@ 24.86	@ 25.63	@ 26.63	@ 28.27
8	@ 18.70	@ 19.12	@ 19.74	@ 20.65	@ 22.12
9	@ 17.89	@ 17.65	@ 18.13	@ 18.77	@ 19.90
10	@ 21.70	@ 22.03	@ 22.50	@ 23.26	@ 24.40

@ – as in table 1

Table 3. RMS error at level – 3.
128 × 128 (original) ⇒ 65 × 65 ⇒ 33 × 33 ⇒ 17 × 17 ⇒ 33 × 33 (expanded) ⇒ 65 × 65 (expanded) ⇒ 128 × 128 (expanded). I – $a = 0.2$, $b = 0.25$, $c = 1.5$; II – $a = 0.3$, $b = 0.25$, $c = 0.1$; III – $a = 0.4$, $b = 0.25$, $c = 0.05$; IV – $a = 0.5$, $b = 0.25$, $c = 0.1$; V – $a = 0.6$, $b = 0.25$, $c = -0.05$.

Picture Image no.	I	II	III	IV	V
1	# 20.71	# 21.28	# 21.92	@ 23.13	@ 25.44
2	@ 28.22	@ 28.79	@ 29.55	@ 30.84	@ 33.41
3	# 18.91	# 19.12	# 19.73	@ 20.17	@ 21.83
4	# 25.63	@ 26.23	@ 27.06	@ 28.30	@ 30.56
5	# 26.85	# 27.15	# 27.63	@ 20.48	@ 30.33
6	@ 25.32	@ 25.81	# 26.48	@ 27.88	# 30.09
7	@ 21.57	@ 22.01	@ 22.78	@ 23.80	@ 25.70
8	# 16.18	# 16.19	@ 16.08	@ 16.56	@ 17.73
9	# 17.39	# 17.50	@ 17.83	@ 18.33	@ 20.30
10	# 19.61	@ 15.31	# 11.98	@ 13.20	@ 13.58

@ – as in table 1; # – correlation mismatch

3. Conclusions

The pyramid processing system has been designed using an i486 processor. The techniques/methods discussed in the above sections have been implemented both for hardware and software in the new design. All the control logic has been fused in EPLDs and firmware in EEPROMs to maintain the security of the design. Reduced and expanded images are generated from a still picture. In the system, provision is made to select different weighting functions and levels for generating the expanded images. The system works satisfactorily. From the results, it is known that for scene-matching type of applications levels 1 and 2 are sufficient. With levels 3 and 4 the probability of the matching reduces. If the RMS error is greater than 10 (i.e. gray value difference between the original and the

Table 4. RMS error at level – 4.
128 × 128 (original) ⇒ 65 × 65 ⇒ 33 × 33 ⇒ 17 × 17 ⇒ 9 × 9 ⇒ 17 × 17 (expanded) ⇒ 33 × 33 (expanded) ⇒ 65 × 65 (expanded) ⇒ 128 × 128 (expanded). I – $a = 0.2$, $b = 0.25$, $c = 1.5$; II – $a = 0.3$, $b = 0.25$, $c = 0.1$; III – $a = 0.4$, $b = 0.25$, $c = 0.05$; IV – $a = 0.5$, $b = 0.25$, $c = 0.1$; V – $a = 0.6$, $b = 0.25$, $c = -0.05$

Picture Image no.	I	II	III	IV	V
1	# 19.31	# 20.06	@ 20.39	@ 21.12	@ 25.58
2	# 32.07	@ 33.31	@ 34.02	@ 34.97	# 37.34
3	@ 26.48	@ 27.10	# 28.28	@ 26.39	# 27.26
4	@ 44.31	@ 45.73	@ 49.27	# 50.61	# 50.61
5	# 34.25	# 31.84	@ 27.32	@ 36.96	@ 37.24
6	@ 26.23	# 26.80	# 26.97	# 27.47	@ 29.79
7	@ 36.57	# 37.07	# 37.66	# 38.52	# 41.87
8	@ 29.89	# 30.18	# 30.78	@ 31.56	# 3418
9	@ 29.06	# 29.13	# 29.32	# 30.03	# 32.87
10	@ 17.34	# 17.54	# 16.56	# 16.55	@ 34.09

@, # – as in table 3

expanded of same size) then matching may not be possible. Further, matching type of application weighting functions $a = 0.5$ and 0.6 are better than $a = 0.2$, 0.3 and 0.4 because of the edged nature of the function.

We wish to express our sincere gratitude to K V S S Prasad Rao for encouragement. We are grateful to our colleagues at RCI and to N R Iyer, Defence Research & Development Laboratory for his cooperation with this work.

References

Anuta P E 1970 Spatial registration of multispectral and multitemporal digital imagery using fast fourier transform techniques. *IEEE Trans. Geosci. Electron.* GE-8: 353–368

Boland J S, Pinson L J, Kane G R, Honnell M A, Peters E G 1977 Automatic target handoff using correlation techniques. Technical report, Auburn University, Alabama, pp 57–63

James H A, Roger F B 1986 A two-dimensional real time video pyramid processor. *RCA Rev.* 47: 3–31

Peter J B, Edward H A 1983 The Laplacian pyramid as a compact image code. *IEEE Trans. Comput.* 31: 512–540

Pratt W K 1974 Correlation techniques of image registration. *IEEE Trans. Aerosp. Electron. Syst*. AES-10: 353–350

Intel user's guide 1985 Intel PL/M 86. Intel Literature Department, USA