Pattern Recognition Letters 2 (1983) 43–46
North-Holland

October 1983

*Hierarchical image representation for compression, filtering and normalization**

Joseph NAOR and Shmuel PELEG
Department of Computer Science, The Hebrew University of Jerusalem, 91904 Jerusalem, Israel

Received 7 February 1982
Revised 28 March 1983

Abstract: Applications of pyramid data structure to image compression and image filtering are described. The compression scheme fits visual perception models, and filtering is computationally faster than filtering in the frequency domain. Image normalization is also possible.

Key words: Pyramid, hierarchical structures, compression, filtering, normalization.

1. Introduction

Image compression and filtering in pyramid data structures are discussed. The pyramid definitions come from Burt (1980) and are described in this section. Applications to compression, filtering and normalization are developed in the following sections.

A new image compression scheme was described by Adelson and Burt (1981). In this method, a low-pass filtered image is subtracted from the original image, giving an uncorrelated difference image. Given the original image g_0, it is low-pass filtered to yield g_1. The difference is given by $l_0 = g_0 - g_1$. The difference l_0 is a decorrelated matrix and can be encoded in fewer bits, and g_1 can be encoded in a coarser sampling. Thus, coding l_0 and g_1 rather than g_0 can result in image compression. The process can be iterated, and g_1 can be represented by it's low-pass filtered version g_2 and the difference $l_1 = g_1 - g_2$. Repeating this process yields a sequence of difference pictures, $l_0, l_1, \ldots, l_N$, each one smaller than it's predecessor.

Every g_i above is obtained from g_{i-1} using a convolution with a Gaussian-like weighting function w of size $2m+1$ by $2m+1$, giving this structure the name *Gaussian pyramid.* Let $W(x)$ be a function defined only at integer values of x, then in the one-dimensional case, for $0 < l \leq N$:

$$g_l(x) = \sum_{i=-m}^{m} w(i) g_{l-1}(x - ir^{l-1}).$$

The distance between elements is multiplied by r from one level to the next, while the number of elements in each level is reduced by a factor of r and by r^2 in the two-dimensional case.

When using $m=2$ the weighting function w has the form

$$w(0) = a,$$

$$w(-1) = w(1) = 0.5,$$

$$w(-2) = w(2) = \tfrac{1}{2}a.$$

A typical value of a is 0.4. In the two-dimensional case w is separable: $w(i,j) = w(i)w(j)$.

To obtain the sequence $l_0, l_1, \ldots, l_{N-1}$ of difference pictures, the expansion of an $N \times N$ picture into an $rN \times rN$ is defined as follows:

$$\text{expand}(g_l(x,y)) = r^2 \sum_{i=-m}^{m} \sum_{j=-m}^{m} w(i,j) g_{l+1}\left(\frac{x-i}{r}, \frac{y-j}{r}\right)$$

* This research has been supported by the U.S. – Israel Binational Science Foundation Grant No. 2442/81.

0167-8655/83/$3.00 © 1983, Elsevier Science Publishers B.V. (North-Holland)

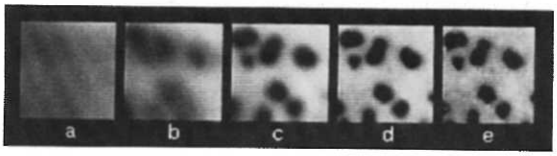

Fig. 1. Reconstruction from the Laplacian pyramid. (a) Top level only, (b) Two top levels, (c) Three top levels, (d) Four top levels, (e) All five levels.

where the sum includes only those terms for which $(x-i)/r$ and $(y-j)/r$ are both integers.

The sequence l_i is then defined by

$$l_i = g_i - \text{expand}(g_{i+1}) \quad \text{for } 0 \leq i < N,$$

$$l_N = g_N.$$

In this sequence too, the number of elements is reduced by a factor r^2 from l_i to l_{i-1}. Every level is a difference of two levels in the Gaussian pyramid, making it equivalent to a convolution with a Laplacian-like filter, and thus the structure is called the *Laplacian pyramid*.

The original image g_0 can be recovered by reversing the generation steps of the Laplacian pyramid:

$$g_N = l_N,$$

$$g_i = l_i + \text{expand}(g_{i+1}) \quad \text{for } 0 \leq i < N.$$

It is interesting to note that Laplacian operators were suggested as important steps in human visual perception (Marr (1982)). An example of recovering a picture from the Laplacian pyramid using this scheme is shown in Figure 1.

2. Compression

Compression can be achieved by requantization of the values in the Laplacian pyramid since they are mostly decorrelated. The quality of the reconstructed image will depend on the technique used for quantization and on the number of quantization steps. Deciding on the number of quantization steps may depend on visual perceptual knowledge: less steps will be chosen in the pyramid levels corresponding to least sensed frequencies. The perceptual question will not be addressed here, but given the number of quantization steps in each level, requantization is found. We propose the use of the well-known optimal quantization for this purpose, using the following definitions:

$p(x)$ – the gray level probability density function.

k – the number of quantization steps.

z_i – decision limits.

q_i – new quantization values.

The quantization changes the gray level x, such that $z_i \leq x \leq z_{i+1}$, into q_{i+1}.

For the quantization to be optimal, in the least squares norm, we minimize

$$\delta_q^2 = \sum_{i=1}^{k} \sum_{x=z_{i-1}}^{z_i} (x-q_i)^2 p(x)\, dz.$$

This expression is minimized for

(1) $z_i = \frac{1}{2}(q_i + q_{i+1}), \quad i = 1, \ldots, k,$

(2) $q_i = \sum_{x=z_{i-1}}^{z_i} x p(x) \Big/ \sum_{x=z_{i-1}}^{z_i} p(x), \quad i = 1, \ldots, k.$

No analytical solution is known for these equations, and we applied the following iterative algorithm to find a solution.

The z_i are initialized arbitrarily so that there is at least one element between each z_i and z_{i+1}. The new quantization levels q_i are computed using equation (2) and then new decision levels z_i are computed using (1). The process iterates until the

Fig. 2. Compression using the Laplacian pyramid. (a) Four quantization steps at all levels (2.66 bits/pixel), (b) Two steps at the lowest level, four steps at all other levels (1.66 bits/pixel), (c) Two steps at the lowest level, eight at all other levels (2 bits/pixel), (d) Two steps at all levels (1.33 bits/pixel).

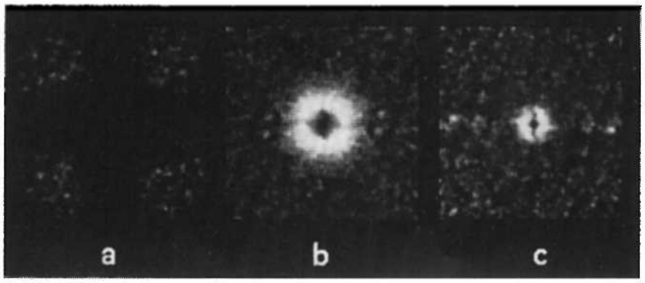

Fig. 3. Effect of the Gaussian pyramid levels in the Fourier spectrum. It was obtained by dividing the Fourier spectrum of the original picture by the Fourier spectrum of the filtered picture. (a) Elimination of lowest level preserves low frequencies, (b) Elimination of the second level attenuates an interior frequency band, (c) Elimination of the third level attenuates smaller interior frequencies.

new decision levels z_i are equal to the previous ones. This is similar to the ISODATA (Duda and Hart (1973)) algorithm. To prevent the problem of a section $[z_i, z_{i+1}]$ becoming empty, the new boundaries are not computed in parallel but sequentially. Results of this compression are displayed in Figure 2. Examples are shown for different numbers of quantization steps. Minimizing the quantization steps of the lowest level is the most significant, since this level has 75% of all pyramid elements. Choosing four quantization steps for the entire pyramid results in 2 bits per element in each level or 2.66 bits/pixel in the entire pyramid. Reducing the number of quantization steps to 2 in the lowest level only, leaving other levels with 4 quantization steps, results in a compression of 1.66 bits/pixel.

3. Filtering

Every level of the Gaussian pyramid is a low-pass version of the level underneath, filtered using a Gaussian-like filter. When we take the difference of two Gaussian operators, we get a band limited operator which is Laplacian-like. In the Laplacian pyramid, the bottom level is a high-pass filter while the middle levels are band-pass filters. The top level is a low-pass filter. When a picture is reconstructed, the Laplacian pyramid's levels are added up. We define a filtering process by using a weighted sum in the reconstruction:

$$f_N = a_N l_N,$$

$$f_i = a_i l_i + \text{expand}(f_{i+1}) \quad \text{for } 0 \le i < N,$$

$$F = f_0$$

where F is the result of the filtering. The a_i are the filtering coefficients and N is the number of pyramid levels. This filtering corresponds to frequency filtering in the Fourier domain. The frequency bands covered by the pyramid levels are displayed in Figure 3. It can be seen that by eliminating the lowest level all low frequencies are preserved, while high frequencies are reduced. Elimination of internal pyramid levels attenuates frequencies in a circular region which grows smaller for higher levels. The coefficients will be given in a vector $A = (a_N, a_{N-1}, \dots a_2, a_1)$. When $A = (1, 1, \dots, 1)$ the original picture will be reconstructed. Examples of filtered reconstruc-

Fig. 4. Filtering in the Laplacian pyramid. (a) Original image, (b) High-pass filter, $A = (1,0,0,1,2,2)$, (c) Band-pass filter, $A = (1,0,2,2,0.5,0.5)$, (d) Low-pass filter, $A = (1,2,0.7,0.5,0.3,0.1)$.

tions using different coefficient vectors corresponding to frequency bands may be seen in Figure 4. The advantage of our filtering technique is that its computational complexity is $O(n)$ in comparison with the FFT algorithm whose computational complexity is $O(n \log n)$.

4. Picture normalization

A Laplacian pyramid defines for every picture some frequency components, where the lower pyramid levels correspond to high frequencies and the higher pyramid levels correspond to lower frequencies. Each level l_i can have a measure m_i associated with it. A normalization filtering will be a filtering for which this measure will have prespecified values. One such filtering can be to equalize m_i for all levels, being some 'frequency equalization'.

As an example, define e_i to be the energy per element at each pyramid level:

$$e_i = \frac{1}{MN} \sum_{m=1}^{N} \sum_{n=1}^{N} l_i(m, n)^2$$

where M and N are the number of rows and columns respectively. To equalize the energy at each level, the following filtering vector is used:

$$A = (a_N, a_{N-1}, \ldots, a_1)$$

where

$$a_i = \sqrt{\frac{e}{e_i}}$$

and e is the average energy in the pyramid. An example of a reconstruction such that the energy in all levels is equal is shown in Figure 5. Fig. 5a was produced by multiple blurring of a picture using 2 by 2 averaging, and Fig. 5b shows the effect of 'energy equalization' on the blurred picture, and much of the blur is eliminated.

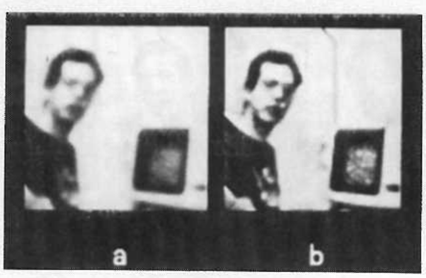

Fig. 5. Energy equalization. (a) Blurred picture obtained by three applications of 2×2 averaging, (b) Reconstruction of (a) such that the energy for all levels is equalized.

5. Concluding remarks

Efficient compression, filtering and picture normalization were exhibited in the Gaussian and Laplacian pyramids. The compression results are better than original results in Adelson and Burt (1981), and approach the quality of other compression methods. The filtering process is computationally faster than filtering in the frequency domain, and should be preferred when there is no need for directional filtering.

References

[1] Adelson, E.H. and P.J. Burt (1981). Image data compression with the Laplacian pyramid. In: *IEEE Conference on Pattern Recognition and Image Processing*, Dallas, pp. 218–223.

[2] Burt, P.J. (1980). Fast hierarchical correlations with Gaussian-like kernels, *5-th Int. Conf. on Pattern Recognition*, Miami, pp. 828–830.

[3] Duda, R.O. and P.E. Hart (1973). *Pattern Classification and Scene Analysis*. Wiley, New York.

[4] Marr, D. (1982). *Vision*. Freeman, San Francisco.