Model-based multi-stage compression of human face images

Mustafa Sakalli, Hong Yan, Kin-Man Lam and Toshiaki Kondo
Department of Electrical Engineering. University of Sydney
Sydney, NSW 2006, Australia
sakalli@ee.usyd.edu.au

Abstract

This paper describes a multi-stage compression scheme of human face images. Snakes are employed in localisation of facial features and biorthogonal spline filters are used for the decomposition of segmented and normalised face images. Wavelet coefficients are vector quantized in different number of cascaded stages depending on their contribution to the subjective quality of the image.

1. Introduction

In this paper, the compression of human face images is considered in three stages: deformable templates, wavelet decomposition and interpolative classified vector quantization. Increasing size of applications and communication makes the compression essential for the bandwidth and storage, and for the possible advantages of the image processing with less computational burden and complexity in the transformed domain. Its potential applications spread from identification systems in the army, hospitals, universities, banks to medical image compression and videophones, etc..

The first stage is where each face is located and segmented into sub-components. This stage parameterises the salient features of eyes, nose and mouth by using deformable templates [1, 2]. The coarse and fine details of segmented images are evaluated in different resolutions by using subband coding scheme. Subbands and their residuals are vector quantized. Codebooks are generated for the subbands and for the interpolation errors of residuals in a multi-resolution sense for each layer of the decomposition. Residuals of the subbands are further processed to generate a gradually refined image in multi-stages for a given subjective quality. The next section presents three stages of the compression scheme.

2 Face Compression

The face compression scheme proposed here has three cascaded subsections.

2.1 Deformable Templates Stage

This stage is built to locate and to normalise the facial images to exploit their geometric redundancies for a limited number of vectors to be maintained for the multiresolution codebooks. A face is modelled as an elastic transparent rubber template in which some parts have limited elasticity. The local properties of nodes are assumed to have fixed shapes retaining their internal bodies when the rubber is aligned over the image. The eyes, mouth and nose are represented by the nodes and their spatial relation with springs as inter-components. The solution to minimise the total cost of the local fitness and the spatial features, yields the localisation of the salient features. The initial values of the parameters for templates are determined by preprocessing. In order to increase the convergence speed, snakes and corner detection [1, 2, 3], algorithms are also employed to locate head, mouth and eye boundaries. Following normalisation and segmentation of facial areas, the main focus of the coding scheme in very low bit rate coding is to minimise the quantization artifacts.

2.2. Subband Analysis of Normalized Images

This stage decorrelates the signal into subbands while mapping the coefficients with higher energy contents to the one side of the platform. Wavelet transforms convert the images jointly into subbands in space and in time.

All orthonormal filters allow the signal to be reconstructed perfectly. Orthogonal transforms preserve same wavelet basis functions in both stages, in analysis and in synthesis. They do not have the symmetry property and thus cause phase distortion. The

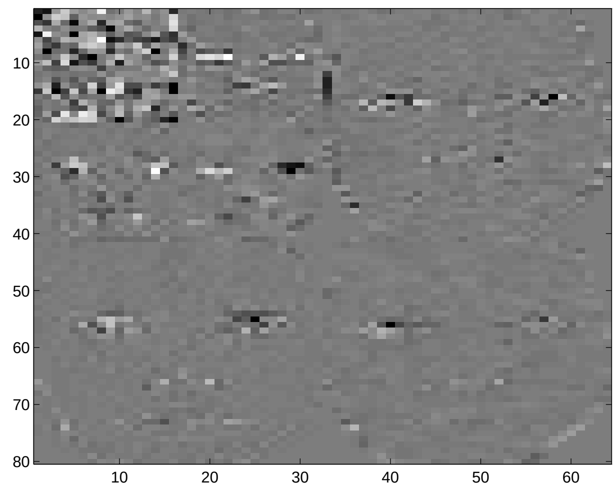

Figure 1. The 4 stage-depth wavelet (13 subbands) decomposition of all face images are decomposed before quantization. The coefficients of the high frequency bands are scaled for visual purposes.

symmetric orthogonal filters are not compactly supported, with the exception of Haar wavelets (HW). The HW introduces rather harsh edge effects on the reconstructed image [4, 6]. In practice, it is not necessary to use same bases functions in both stages, as long as wavelet filters accommodate mutual orthogonality, $< \psi_{j,k}\tilde{\psi}_{l,m} >= \delta_{j,k}\delta_{l,m}$, where $\tilde{\psi}_{l,m}$ is referred as dual wavelet of the mother wavelet function $\psi_{j,k}, j, k, l, m \in (Z)$. These filters are compactly supported FIR filters and can be cascaded in pyramidal structures without any need to phase compensation so that introduced aliasing cancel each-other at he synthesis stage. Image extensions are performed in whole-sample symmetry (WS) form [8]. The subband decomposition produces sub-images at different orientations. This property is exploited in ICVQ stage to generate a multiresolution classified codebook for subband images.

2.3 ICVQ of Subband Coefcients

The key issue in vector quantization is to obtain an optimum code book for a class of quasi-stationary image sources. The Generalized Lloyd Algorithm (GLA) [7] calculates the centroids of evenly spaced vectors and recursively minimises distortion between the input vectors and corresponding codewords. The initial codebook is generated as centroids of all N subsources, with, vectorial averaging $C_{0i} = \frac{1}{N}\sum_{i=1}^{N} S_i$. After this initial vector split by two, the operation of convergence recycles iteratively to an optimum codebook in two sub-steps: mapping the input source for a given decoder using $k-NN$ and then, checking if the decoder could be achievable for a given encoded data set, [7, 5].

Interpolation is applied for the all the HL and LH bands in vertical and horizontal directions respectively. An undersampled version of subbands is coded with DPCM. Then, the estimation errors and interpolation errors between DPCM coded samples are vector quantized. This is to decrease the existing statistical directional redundancies in the subbands. A set of classified residual codebooks are generated such as code books for the 2^{nd} lowest subbands of the normalised face and residual codebooks for the detail images of the eye and mouth.

3 Simulation

The images experimented are forward looking mugshot black and white face pictures, 8 bpp, 112x92 pixel obtained from Olivetti face data base. The background intensity is assumed to be uniform. The limits of the head boundaries are found with horizontal and vertical projections to initialise the snakes in close proximity of the facial borders. To decrease the influence of edge forces on snakes, blurring is performed by a 2-D Gaussian operator, then the intensity of blurring is gradually reduced. The bounded region of the face is segmented based on anthropometric information. The corners of the mouth and the eyes are detected with the algorithms introduced in [2]. All images are normalised to a constant number of pixel width between the center of each eye.

The normalised images are analysed until in a depth of four layers of the decomposition Fig.1. Since the quantization and truncation errors are amplified and spread wider over the surface, the length and the regularity of the synthesis filter are very important. The closer to the surface where high frequencies starts to be visible, the more local and less global artifacts should be expected. The subbands are rendered according to their energy content. If the absolute value of a wavelet coefficient is smaller than a pre-determined threshold, it is nullified. This is performed with a similar behaviour to the frequency sensitivity of the HVS modelled as the modulation transfer function [9]. It is observed that the mid-bands contribute more to the perceptual quality, that is why more bits are sacrificed for the third layer of the decomposition, especially for the eye and mouth images. Run-length coding is then performed. The reconstruction of the decoded wavelet coefficients is a time reversal of the analysis stage using synthesis filters.

4 Results and Concluding Marks

Four levels of reconstructed results are gradually presented for both the training Fig.2 and the test groups Fig.3. The first rows are a set of the original images. From the second to the last row the ranges of bpp including the side information for a range of PSNR are obtained as $0.0859 - 0.09$ bpp, $23.5 - 22.4$ dB; $0.138 - 0.141$ bpp, $24.3 - 25.6$ dB; $0.185 - 0.178$ bpp, $28.3 - 26.5$ dB; $0.22 - 0.24$ bpp, $31.4 - 29.6$ dB. The 7 to 10 subbands and their residuals are gradually introduced to the reconstruction. The side information includes the scaling and shifting values of the subbands in each stage before quantization for switched codebooks. The PSNR of test images are obtained $1.2 - 2.8$ dB less then those of training values in average. Runlength coding or Huffman coding didn't provide extra compression for each individual image. This is because, the information was already de-correlated not leaving enough redundancy to be exploited in entropy coding. When the indices of quantized vectors are concatenated as a bulk of 30 images, then an extra compression ratio of $1.2 - 1.3$ is provided.

Multi-level reconstruction is especially convenient for the progressive transmission of images for early recognition. The cost paid for the increasing quality is complexity and another 30-81 bytes in average for a desired signal quality. In conclusion, in this paper, we have developed a face compression scheme as an embedded structure of deformable templates, wavelet decomposition and interpolative classified residual vector quantization.

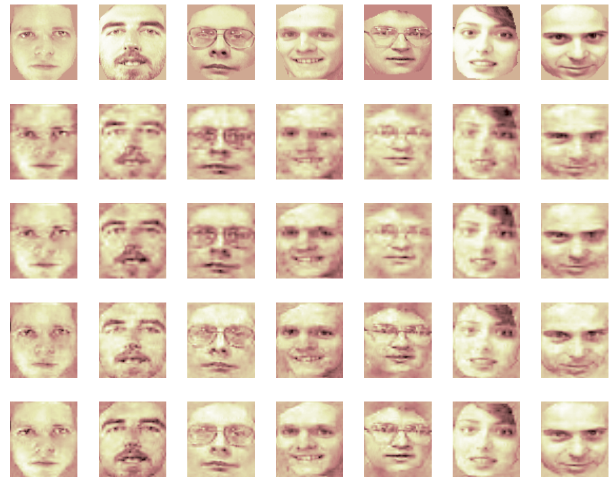

Figure 2. Stage by stage reconstruction of training groupfrom the 7 vector quantized subbands bands to 10 subbands, out of 13. Original images at the first row.

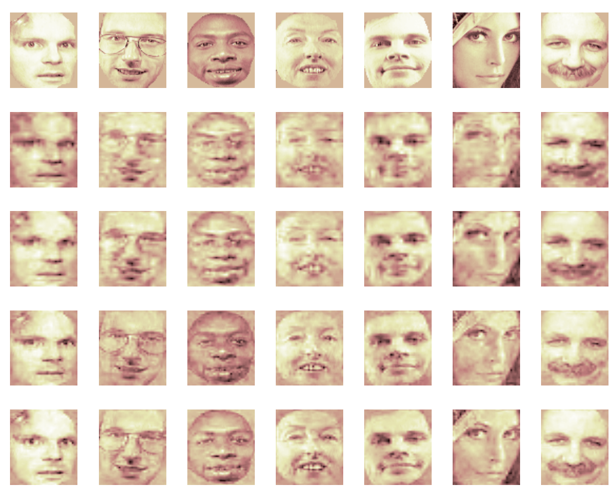

Figure 3. Reconstructed images of the test group, same as the Fig.2

References

[1] A.L. Yuille, P.W. Hallinan, D.S. Cohen, *Feature extraction from faces using deformable templates*, Int. J. of Comp. Vision. Vol.8, pp. 99-11. 1992.

[2] K.M. Lam, H. Yan, *Fast algorithm for locating head boundaries*, J. of Elec. Imag, 3(4), 352-359, 1994.

[3] M. Kass, A. Witkin, D. Terzopoulos, *Snakes: Active contour models*, Int. J. of Comp. Vision. Vol.1, pp. 321-331, 1987.

[4] C.K. Chui, *An introduction to wavelets*, Academic Press Inc. 1992.

[5] N.S. Jayant, P. Noll *Digital coding of waveforms*, Englewod Cliffs, NJ: Prentice Hall, 1984.

[6] A.N. Akansu, Haddad, *Multiresolution signal decomposition*, Academic Press, Inc., Boston, 1992.

[7] A. Gersho, R.M. Gray, *Vector quantization and signal compression*, Boston: Kluwer Acad. Publ., c1992.

[8] S.A. Martucci, R. M. Mersereau, *Digital filtering of images using the discrete sine or cosine transform*, Optical Eng. SPIE Vol. 35 pp. 119-127, 1996.

[9] N.B. Bill, *A visual model weighted cosine transform for image compression and quality assessment*, IEEE Trans. COM-33., pp. 551-557, 1985.