using the original MBE pitch analyser, but these errors are effectively removed by using the improved pitch analyser (unbroken). We also found that the improved pitch detection algorithm performed very well, even without pitch tracking, therefore the algorithmic coding delay was shortened to 32ms. A full-duplex MBELP coder employing the improved pitch detection algorithm was successfully implemented in real time on a single C31 DSP, and pitch estimation requires only 16% of the processor resource.

Conclusion: We propose a new error measure for spectrum matching in MBE pitch analysis of speech. By applying a corrective measure to the original error measure for pitch estimation, gross pitch errors can be effectively removed. The corrective measure is based on a sum-of-product formula which facilitates fast searching of the optimum pitch period by using partial-sum comparisons.

© IEE 1996
5 March 1996
Electronics Letters Online No: 19960593

C.-F. Chan and E.W.M. Yu (*Department of Electronic Engineering, City University of Hong Kong, 83, Tat Chee Avenue, Kowloon, Hong Kong*)

References

1 GRIFFIN, D.W., and LIM, J.S.: 'Multi-band excitation vocoder', *IEEE Trans. Acoust., Speech Signal Process.*, 1988, **ASSP-36**, (8), pp. 1223–1235

2 YU, E.W.M., and CHAN, C.F.: 'Efficient multiband excited linear predictive coding of speech at 1.6kbps'. Eurospeech-95, 1995, pp. 685–688

Motion estimation with object based regularisation

S. Panis and J.P. Cosmas

Indexing terms: Motion estimation, Dynamic programming

A dynamic programming based matching method for motion estimation, that optimises a Bayesian maximum likelihood function in a 3-D optimisation space, is presented. The Bayesian function consists of a matching cost and an object based 2-D regularisation cost. The method gives results more accurate than block-based matching since the motion boundaries are close to the actual object boundaries.

Introduction: A dynamic programming-based matching method for motion estimation is presented which optimises a Bayesian maximum likelihood function that consists of two parts: a matching cost and a 2-D regularisation cost that enforces the motion vectors to be a monotonic function only within objects, thereby allowing an edge preserving transition between objects and hence object-based regularisation. A 3-D optimisation space is used to accommodate the two degrees of motion freedom.

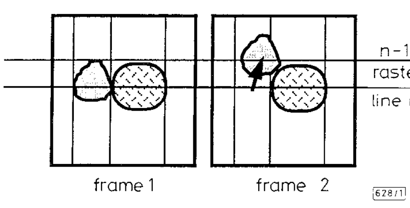

Fig. 1 *Quick and random motion of objects*

Three-dimensional optimisation: If matching of a raster line from one image to the next is attempted, and one of the objects moves as shown in Fig. 1, then dynamic programming cannot correctly match raster line n in frames 1 and 2. This is apparent in Fig. 2*a* where the raster lines n from the two consecutive frames is plotted against each other. The dynamic programming will give no-matching results in this case. To prevent this, the raster line of frame 2 is composed of a collection of segments from other raster lines as in Fig. 2*b*. In Fig. 2*b*, a segment from raster line n–1 (Fig. 1) was used to compose the matching raster line of frame 2, and the vertical difference per section was recorded. Composition was achieved by having a 3-D optimisation space for dynamic programming. One axis of the optimisation space is the raster line and the other two axes are the possible horizontal and vertical motion vector axes.

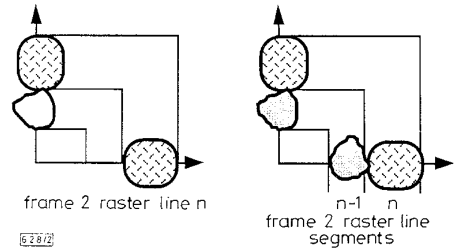

Fig. 2 *Non matchable and matchable patterns*
a Non matchable, *b* Matchable

Cost function: The total cost function for dynamic programming is a combination of a normalised matching cost and a regularisation cost. The normalised matching cost (C_{NMC}) is determined by the quality of matching. The regularisation cost consists of two parts. One part enforces the vectors across a surface to be a monotonic function (monotonicity constrain [1]) and the other part controls the strength of the monotonicity function so that its effect is dependent on the luminance variations and hence acts strongly inside objects but lightly at the edges. Cox *et al.* [2] claim that enforcement of regularisation may harm the results at the edges where they are most accurate owing to the rich texture. Geiger *et al.* [1] state that regularisation is necessary in order to obtain accurate results even where matching is poor, e.g. within objects with poor texture. The cost function presented here takes account of both points of view. By reducing the effect of the monotonicity cost at the edges, the dynamic programming can use the matching cost C_{NMC} as the main decision cost. Since matching near edges is usually better than within a surface, the C_{NMC} has very high confidence at the edges and this leads the program to a correct decision without regularisation influence. A typical problem occurs at two consecutive pixels that belong to two different objects with large motion vector difference. If no consideration for edges was there, the monotonicity constraint alone would impose vector uniformity, shooting the total cost high and resulting in very large punishment. The area would have been interpreted as unmatched.

The regularisation cost function is defined as

$$f(x, s) = \frac{(\mu\sqrt{|x|} + \varepsilon|x|)}{e^{\beta s^2}} \quad x \geq 0, s \geq 0 \qquad (1)$$

where x is the amount of motion vector jump between consecutive pixels of the raster line and s is the normalised luminance gradient extracted using a Sobel edge detector. The plot of the cost function is shown in Fig. 3. The weights μ, ε, β are empirically determined constants. For $s \leq 1$ i.e. when there is practically no edge, the function is punishing only motion vector jumps and hence acting purely as a monotonicity constraint [1]. For $s > 1$, i.e. when there is a strong edge, the effectiveness of the motion vector jump function is reduced to allow a jump without strong punishment. According to Fig. 3, as s increases the smoothness criterion is pulled down to allow jumps between consecutive pixels. The total cost $T_n(j, i)$ is calculated with the recursive function

$$\begin{aligned} T_n(j, i) &= C_{NMC_n}(j, i) + f(x_h, s) + f(x_v, s) \\ &\quad + T_{n-1}(j_{best}, i_{best}) \qquad \text{for } n > 1 \\ T_n(j, i) &= C_{NMC_n}(j, i) \qquad \text{for } n = 1 \end{aligned} \qquad (2)$$

where x_h is the horizontal motion jump between the current candidate horizontal motion vector and the horizontal motion vector of the previous column, x_v is the respective vertical motion jump, n is the current column number, j is the candidate horizontal motion vector and i is the candidate vertical motion vector. (j_{best},i_{best}) is the best predecessor vector.

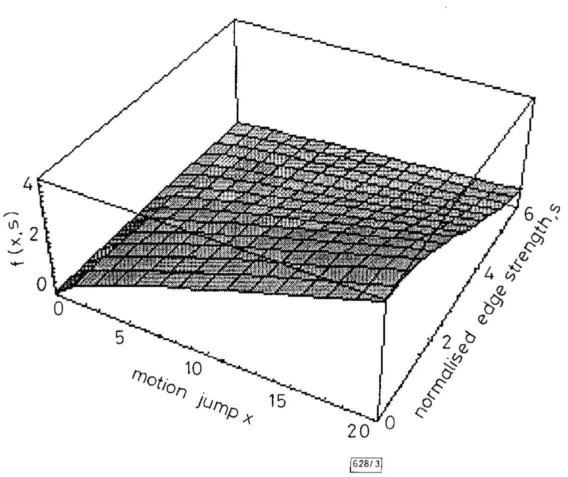

Fig. 3 *Cost function for motion estimation*

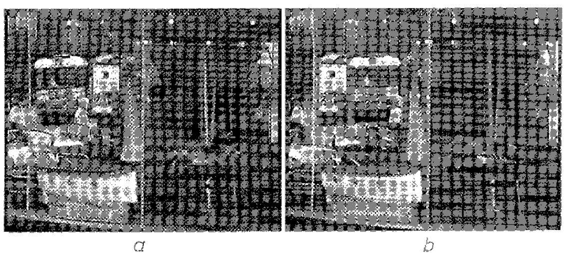

Fig. 4 *'Manege' sequence*

a First frame, *b* Second (edited) frame

Results: In Fig. 4*a* and Fig. 4*b*, frames 1 and 2, respectively, of the sequence 'Manege' [3] are shown where, after editing in the second frame, the bus and the bus stop were swapped. The motion estimation program was first run without edge consideration ($\beta = 0.0$) i.e. with the monotonicity constraint applied uniformly on the entire scanline. Motion results were obtained for the bus, whereas the bus stop was classified as an area with no match owing to punishment of the large motion vector difference at the boundary between the bus and the bus stop. Hence a cheaper alternative path was to ignore the pixels that belonged to the bus stop, and jump until pixels with motion vectors comparable to those of the bus were met.

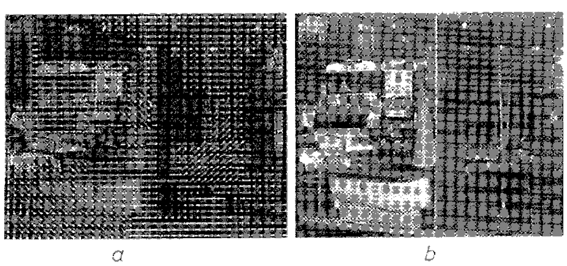

Fig. 5 *Motion vector field with edge consideration and reconstructed image*

The program with edge consideration $\beta = 0.1$ was run on the two frames, and the horizontal motion vector field is shown in Fig. 5*a*. For the bus stop, there is a calculated motion value (arrows pointing left) because the regularisation cost at the boundary between the bus and the bus stop was pulled down by the edge factor, therefore the cost was reduced such that it could allow transition from a bus pixel to a bus stop pixel without a jump, and without shooting the total cost up so that it was very high. Fig. 5*b* shows a reconstruction of frame 1 from frame 2 using the motion vector field.

An advantage of this method over the block-based matching is that the motion edges are closer to the actual luminance edges, which is a valuable feature for segmentation and motion compensation in object based coding. Owing to the object base accuracy of the motion vectors, block artefacts and the mosquito effect are not present. The proposed algorithm was used in the object based coder of [4] and experiments showed that errors owing to motion compensation were considerably lower than when block based matching was used, with the performance of the coder being better than that of MPEG-2.

Acknowledgments: This work was done within the framework of RACE (Research on Advanced Communications in Europe) under contract R2045 'DISTIMA' (DIgital STereoscopic IMaging and Applications).

© IEE 1996
7 March 1996
Electronics Letters Online No: 19960598

S. Panis (*Siemens AG, ZFE T SN 2, 81730 Munich, Germany*)

J.P. Cosmas (*Queen Mary and Westfield College, Electronic Engineering Department, Mile End Road, London E1 4NS, United Kingdom*)

References

1 GEIGER, D., LANDENDORF, B., and YUILLE, A.: 'Occlusions and binocular stereo', *Int. J. Comput. Vis.*, 1995, **14**, pp. 211–226
2 COX, I.J., HINGORANI, S., MAGGS, B.M., RAO, S.B.: 'Stereo without regularization'. NEC Research Institute Report, October 21, 1992
3 CCIR 601 format sequences shot within the EU project RACE DISTIMA
4 PANIS, S., ZIEGLER, M., and COSMAS, J.P.: 'Object-oriented monoscopic/stereoscopic image coder'. Int. Conf. Signal Process., Limassol, Cyprus, June 1995

Trellis decoding of combined diversity-coding scheme (MLSD) for fading channels

K. Mohamed-Pour and P.G. Farrell

Indexing terms: Trellis coded modulation, Fading, Diversity reception

The performance of a receiver using a combination of soft-decision decoding and diversity reception is investigated for nonselective multipath Rayleigh fading channels. A new scheme for soft-diversity, soft-decision detection, maximum likelihood selection and decoding (MLSD), is introduced, in which decisions on the diversity channels and decoding are carried out simultaneously by using a trellis and the Viterbi algorithm.

Introduction: Diversity reception is one way to improve the reliability of communication without increasing either the transmitted power or the bandwidth. Coding is another advantageous way which can be used for power-limited and/or bandlimited channels [1]. In this Letter a combination of diversity detection and decoding, especially in the soft-decision decoding case, is introduced which improves the performance of the transmission of data over a multipath fading channel, compared with separate diversity detection and decoding.

Diversity with soft decision decoding: Supposing an L-diversity scheme, the system is modelled as L independent multipath Rayleigh channels corrupted by additive white Gaussian noise $n(t)$. The L receivers are assumed to employ matched filters to get the maximum signal/noise ratio at sampling times. Values $\{r_{ij}\}$, $i = 1,2,..., L$; $j = 1,2,..., n$, are the unquantised samples of the outputs. These samples are then quantised to Q levels and then accumulated in an $L \times n$ buffer matrix as $\{y_{ij}\}$ values which can be used in a diversity combination scheme, involving a code of length n.

'Soft-diversity decoding' refers to any decoding process on the soft values of the outputs of the channels in a diversity system which may involve different combination techniques. Two such techniques are: