한국산학기술학회논문지
Vol. 10, No. 2, pp. 274-279, 2009

Hybrid Algorithm for Interpolation Based on Macro-block Gray Value Gradient under H.264

Shi Wang[1], Hongxin Chen[1], Hyeon-Joong Yoo[2] and Hyongsuk Kim[1*]

H.264하에서 마크로 블록 그레이 값의 미분을 사용한 인터폴레이션

왕실[1], 진홍신[1], 유현중[2], 김형석[1*]

Abstract H.264 suggests applying a 2-D 6-tap wiener filter to realize the interpolation for half-pixel positions, followed by a bilinear interpolation to get the data of quarter-pixels precision. This method is comparatively simpler; however, it only considers the affection of 4-connection neighborhood ignoring the influence that comes from the changing rate between respective neighborhoods. As a result, it has the characteristics of a Low-pass filter under the risk of losing high-frequency weights. The Cubic interpolation uses the gray-values within the larger regions of points to be sampled for interpolation. Nevertheless, the cubic interpolation is more complicated and computational. We give a deep analysis on the features while applying both bilinear and cubic interpolation in H.264 presenting a proper selection of interpolation algorithm with respect to specific distribution of gray-value in a certain grand block. The experiments point out that load for motion searching and interpolation are reduced when promoting the precision of interpolation simultaneously.

Key Words : H.264, Half-pixel, Wiener filter, Cubic interpolation

요 약 H.264에서는 2-D 6-tap wiener filter가 1/2 화소의 위치로부터 1/4 화소 위치를 보간해 내는데 사용되고 있다. 이 방법은 비교적 간단하지만 이웃 화소들의 변화율을 무시하고, 4 방향의 이웃 화소들의 영향만 고려하게 되므로, 결과적으로 저역 필터의 특성을 갖게 된다. 그런데, 큐빅 보간 법을 사용하면 보다 넓은 영역의 화소 값을 고려하여 보간하는 장점은 있지만 계산이 복잡하다는 단점이 있다. 이 연구에서는 H.264에서 bilinear와 cubic 보간 법을 사용할 경우의 특성들을 해석하여, 임의의 큰 블록에 보다 적합한 보간 방법을 자동 선택할 수 있도록 하였다. 실험에서 물체영상의 움직임 탐색과 보간에 요구되는 계산 량을 보간의 정밀화를 통하여 대폭 감소시킬 수 있음을 보였다.

1. Introduction

H.264 standard is a new generation video compress standard composed of ITU-T and ISO/IEC, which supports multi-frame selection and up to 7 motion estimation modes, consequently [1][2]. Andmotion estimation takes a larger computation in H.264 coders. Various searching algorithms which are applied to reduce the motion estimation could be divided into: (1) Full Searching (FS) reducing load of searching by assistance strategy. For instance, choosing a dynamic searching centre or jumping out the matching searching in advanced to reduce the calculation for block matching, and so on. (2) Partial Searching utilizing reduction of searching forthe spatial position candidates or special principle for matching. But the faults of them are the calculation is too

[1]Division of Electronics and Information Eng., Graduate School, Chonbuk National University,
[2]Dept. of Information and Telecommunication Eng., Sangmyung University,
*Corresponding Author: Hyongsuk Kim(hskim@chonbuk.ac.kr)
Received January 12, 2009 Revised February 12, 2009 Accepted February 18, 2009

complex, the calculation quantity is too large. So the improvement of motion search algorithm based on blocks looks rather important. The improvement is mainly modified from the following three faces: the choice of initial search points, the matching rule and the search strategy. Numerous scholars produce lots of new Fast Integer Motion Search, mainly including three steps method, altogetherclutches the direction search method, Two-dimensional algorithm drop method, four steps method, Diamond method and so on [2]. The improved algorithm produced in this paper mainly improves efficiency of algorithm from search strategy, and then reduces search time of the motion estimation.

2. Bilinear Interpolation and Applications in H.264

The value derived from bilinear interpolation depends on average sum of 4 adjoined grid points. In another word, it is based on distance between the interpolated pointand neighbor points to decide the corresponding weights and compute interpolation value ultimately. For linear function interpolation, function $f(x)$ defined within the rectangle N should be taken into consideration, the formula is:

$$f(x)=\sum_{n\subset N} h(x,n)f(n) \tag{1}$$

x is a point among the region of N, $h(x,n)$ is the linear interpolated weight function. A typical interpolated weight function is Lagrange polynomial:

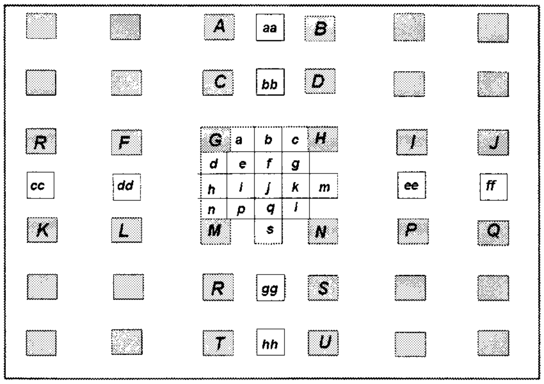
$$h(x,n)=\prod_{i\neq n}\frac{x-i}{n-i} \tag{2}$$

The linear interpolated weight function of Equation (2) changes into:

$$h(x,n)=\begin{cases}1-|x-n| & n-1<x<n+1\\ 0 & n+1\le x \text{ or } x\le n-1\end{cases} \tag{3}$$

H.264 applies a Wiener filter to complete interpolation for positions of half-pixels, and further to get quarter pixels precision by bilinear interpolation.

As shown in Fig.1, capitals A, B, C, D are entire pixels. b, h, m, s and aa, bb, gg, hh are half-pixels gained by the Wiener filter. The rest quarter-pixels of a,c,d,n,f,j,k,q are gained through interpolation between entire pixels and neighbor half-pixels.



[Fig.1] Pixel positions

The formulas are as following:

$$\begin{aligned} a&=(G+b+1)>>1 & i&=(h+j+1)>>1\\ c&=(H+b+1)>>1 & c&=(b+j+1)>>1\\ d&=(G+h+1)>>1 & k&=(j+m+1)>>1\\ h&=(M+h+1)>>1 & q&=(j+s+1)>>1 \end{aligned} \tag{4}$$

The others:

$$\begin{aligned} e&=(b+h+1)>>1 & p&=(h+s+1)>>1\\ g&=(b+m+1)>>1 & r&=(m+s+1)>>1 \end{aligned} \tag{5}$$

From the equation above, to obtain a quarter-pixel interpolation, 2 addition and 1shift compute are needed. Even though the whole procedure is simple, the frequency spectrum of bilinear interpolation is similar to a low-pass filter, which leads to an incorrect interpolation for high-frequency signals. However, it is right high-frequency signals that reflect edge information in images mostly, hence the edge information of the image is weakened under such situation and image turns to be relatively blurred.

3. Cubic Interpolation and Applications in H.264

Cubic interpolation is also widely applied in numerical analysis besides bilinear interpolation. According to interpolation theory [6], 2-Dimention image cubic interpolation of even-interval samples can be presented as following:

$$S(x,y)=\sum_{l}\sum_{m}C(x_l,y_m)h(|x-x_l|)h(|y-y_l|) \quad (6)$$

Where $s(x,y)$ is defined as interpolated point, $c(x_l,y_m)$, sample value of pixel (x_l,y_m), and h is the function for interpolation weights. The selection of h is the core of Cubic interpolation.

Single-variable interpolation weight Selection in H.264 mainly lies on Coons - Gordon surface, which is a bi-variable interpolation could be constructed by an approximately repeated single-variable interpolation.

Single-Variable Cubic interpolation core value [3]:

$$h(x)=\begin{cases}(A+2)\cdot x^3-(A+3)\cdot x^3+1 & 0\le x<1\\ A\cdot x^3-5A\cdot x^3+8A\cdot x-4A & 1\le x<2\\ 0 & \end{cases} \quad (7)$$

Practically, A is chosen to be 1/2. In Cubic interpolation utilized in image processing, it is necessary to expand a one-dimension Cubic interpolation to a 2-dimension Cubic interpolation, and then take use of equation(5), picking up the 4-by-4 pixel close to the pixel to be detected for interpolation.

$$\begin{aligned}a&=F\\ b&=0.5625(F+G)-0.0625(E+H)\\ c&=0.5625(F+J)-0.0625(B+N)\\ d&=0.0625\times0.0625(A+D+M+P)+0.5625\times0.5625(F+G+L+K)-0.5625\times0.0625(F+G+L+K)(B+C+E+I+H+L+N+O)\end{aligned} \quad (8)$$

Generally, Cubic interpolation possesses wider frequency-band than that of the bilinear method [4][5]. Besides, it preserves the high-frequency parts generating higher precision but more complex in calculation.

4. Selection of Proper Interpolation

In mathematics, bilinear interpolation is an extension of linear interpolation for interpolating functions of two variables on a regular grid. The key idea is to perform linear interpolation first in one direction, and then again in the other direction. Bicubic interpolation is an extension of cubic interpolation for interpolating data points on a two dimensional regular grid. The interpolated surface is smoother than corresponding surfaces obtained by bilinear interpolation or nearest-neighbor interpolation. Compared with the two algorithms discussed above, bilinear interpolation is very effective to Low-frequency signals (the regions where gray-values distribute regularly) besides not being so computational, cubic convolution is acting on high-frequency parts (the regions where gray-value distribute irregularly), especially on protecting the edge regions of images, but it is rather computational. To balance theinterpolation precision and complexity, a good trend is proposed as to pick up different interpolation according to gray-value characteristics in different parts of blocks.

Grounded on statistical theory, square difference can be used as a threshold to judge the regularity of gray-value distribution in a block. However, it requires large amount of computation on all the pixels in the blocks according to:

$$S\wedge=\sum X*X/N \quad (9)$$

In the other hand, the edge parts of images which is the most effective to human's vision concentrates in a certain part of a block, where cubic interpolation or some other high-precision algorithms are adopted; while in other parts where gray-values distribute regularly, bilinear interpolation is good enough. So this simple selection principle combines the strong points of both bilinear and cubic interpolation.

In each block, distance between two neighbor pixels is certainty. To equal-interval sampling data, an interpolation can be represented as:

$$f(x)=\sum c_k h(x-x_k) \quad (10)$$

Where is an interpolation kernel, is weight coefficient, is the number of data used for convolution. As a matter of fact, is always symmetric in practical, that is to present $h(x)=h(-x)$, is the sampling value. From (9), it is clear that when takes different functions, the interpolation result will be different, so the algorithm alters. Actually, we can infer that:

$$f(x)=f(x_k)=f(x_k+1) \quad (11)$$

Regardless of which kind of function is being applied. Evidently, if the gray-values of reference points are equal, then the gray-value of the interpolated points will be the same with that of reference ones, generating no changes judging by human eyes. So if we want to produce obvious changes which could be perceived by naked eyes, it is necessary to make a step-change between the interpolated points and reference point in adjacency. The stronger the changes produce stronger vision perception.

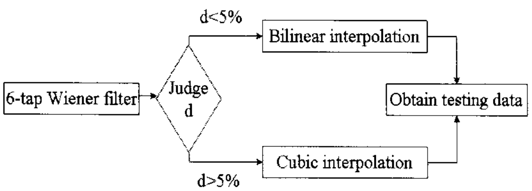
Suppose that an interpolated point's gray-value is $f(x)$ the changing gradient is

$$d = (f(x) - \sum C_k A(x_k)) / \sum C_k A(x_k) \tag{12}$$

C_k is a weight coefficient, and $A(x_k)$ is a gray-value adjoining to point with gray-value equal to $f(x)$. From the equation (11), we can find that weight coefficient and the selection for the numbers of adjacent points are crucial to the precision of gradient d.

Under H.264, during the procedure of searching for motion, the first interpolated value $f(x)$ can be obtained through a 6-tap wiener filter. Taking computation complexity into consideration, we only take two integer pixels that are most close to $f(x)$. Take Fig.1 as an example, for point h we take the integer point G and M while for point b, G and H in horizontal direction will be picked up which is most closed in vertical direction, and C_k is replaced by 1/2 here. When d< 5 %, the gray-value of $F(x)$ does not change greatly, to which human's eyes does not react sensitively enough. Approximately, assume the half-pixel found by motion search changes little compared with the adjacent gray-values. In practical, there is no much difference in result, whatever any algorithm adopted in second interpolation.

The structure described details as following:



[Fig. 2] Proposed algorithm

Consequently, after the first interpolation of a block finishing finding the best matching pixels, if the gray-values of the pixels are consist to that of the neighborhood region, choosing bilinear algorithm for next interpolation; or else cubic method would be better. In this way, the results costs less time for computation without loss of interpolated quality in the meantime.

5. Experiment & Analysis

5.1 Experiment Environment

To validate the effecting form the algorithm proposed in this paper, we applied the Coding program supplied by TML-8, utilizing the interpolation algorithm proposed only for motion compensation parts. In our experiment, the standard test sequence "Miss American"was chosen as the source image, mostly on that it features for slow motion and comparatively simple. Each sequence adopts the format of QCIF 50fpr (176×144 Pixels, Y :U :V = 4:2:0), testing on a Pentium4 PC. During coding procedure, all the motion estimation applied matching-fit, based on SAD principle, Using Fast-Diamond searching for motion searching. Simultaneously, Fixed-Step strategy was used for removing the difference derived from the control by various code rates. The performance comparison between Bi-linear algorithm and the proposed algorithm was measured based on (Peak Signal-to-Noise) PSNR and compression ratio.

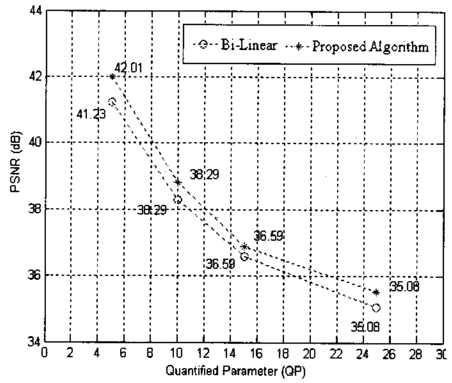
5.2 Comparison on PSNR

PSNR is the average Peak value of SNR for multi-frame images, which is a significant metric for the entire coder.

It is expressed as : (dB)

$$\text{PNSR} = \frac{1}{L}\sum 10\log\frac{255^2 MN}{\sum_{m=1}^{M}\sum_{n=1}^{N}[fk(m,n) - fk'(m,n)]^2} \tag{13}$$

Where the size of the image is $M \times N$, L is the number of frame, of original image at the point of (m,n), $f_k(m,n)$ is the pixel value of the k_{th} frame in image. The difference between Bi-linear and proposed algorithm is shown in Fig.3.



[Fig. 3] Comparison of PSNR between Bi-linear and Proposed Algorithm

From the data shown in the Fig.3, with the increase of quantified parameters, the number of zero-block increases accordingly, hence, PSNR will decrease in a certain extent.

While the quantified parameter of "Miss America image" sequence is 5, the PSNR obtained from a simple bi-linear interpolation will be raised to around 0.8 dB. As the quantified parameters increase, PSNR from both algorithm declines largely. However, the PSNR got from the algorithm proposed in this paper is obviously 0.2~0.6 dB higher than regular algorithm

5.3 Compress Ratio

[Table 1] Length variety of code sequence from 2 typical algorithms under different Qp. (Unit: byte)

	Bi-Linear	Proposed Algorithm	variety (%)
5	47680	46513	−2. 4
10	21025	20779	−1. 2
15	12607	12510	−0. 8
25	8477	8196	−3. 3

Before the both algorithms are selected, a Wiener filter must be used at first, so we skip this step here. After the judgment of the branch, the fit algorithm should be chosen. Fig.4 is the actual results of enlarged pictures. Each picture takes 25 samples, size of 100 ×100 pixels. Different algorithms are used to enlarge these pictures to 800 ×600 pixels. Experiment results just focus on the face part that contains many details.

(a) (b) (c) (d)

[Fig. 4] Experiment results of (a) Cubic interpolation, (b) the Interpolation of proposed method, and (c) Bilinear Interpolation Compared with Original picture (d)

Though one more offset judgment than single algorithm seems to increase complexity of the program, actually it equals to a single calculation, unparallel to computation cost by Cubic interpolation. From Fig. 4, it performed almost the same effecting with Cubic but much better than linear method. Obvious computation speed difference shows: the speed is raised up to 20%-30% when competed with utilizing Cubic interpolation alone. What is more, when judging on the ground of threshold, it quit automatically from searching once d is less than a certain extent, which also saves plenty of time for interpolation. The decision of exact value for threshold needs to be discussed.

6. Conclusion

In this paper, we discussed a hybrid algorithm combining the advantages of Cubic and Bi-linear interpolation, against the limitation by using a single interpolation. Precision of gradient is adopted to judge the changes between the interpolated points and reference point in adjacency. Then the appropriate interpolation method will be chosen based on it. The results of our experiment proved it effective to improve image quality objectively and enhance the coding compress ratio, while brings no more coding consumption.

References

[1] Stockhammer T, Hannuksela M M,Wi – egand T. H.264 /AVC in Wireless Environments[J]. IEEE Trans, Circuits Syst, Video Technol, 2003, 13 (6): 657 -673.

[2] I. Lille-Langoy, et al. Telenor. H.26L Proposal and

Software ITU-T SG16/Q15doc. Q15-H-08 [S]. Berlin, 1999-08.

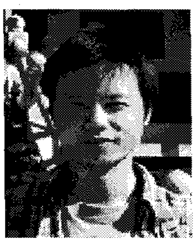
[3] Schreiber William F, et al. Transformation between continuous and discrete representations of image: A Perceptual Approach. [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1985, PAMI-7: 178-186.

[4] Wan Shuai, Wang Xindai. A new Generation Video Compress Standard H. 264. [J] China Cable TV, 2003, 9 (18): 32 - 34.

[5] Zhen Xiang, Ye Zhi yuan, Zhou Bing feng. A Summary for Core Tech in Protocol of JVT. Software Transaction, 2004, 15 (1): 58 - 68.

[6] Chih-Hung Kuo, Meiyin Shen, C.-C. Jay Kuo. Fast motion search with efficient inter-prediction mode decision for H.264. [J] Visual communication & Image Representation. April. 2005

Shi Wang [Regular Member]



- 2000 ~ 2004 : Wuhan University of Technology, China, B.S.
- 2004 ~ 2007 : Huazhong University of Science and Technology, China, M.S.
- 2007 ~ Present : Div. of Electronics and Information Eng., Chonbuk National University, PhD.

<Research Interests>
Multimedia and Communications System, Segment-Based 3-D Matching, Robot Vision.

Hongxin Chen [Regular Member]

- 2000 ~ 2004: Wuhan University of Technology, China, B.S.
- 2004 ~ 2007: Wuhan University of Technology, China, M.S.
- 2007 ~ Present: Div. of Electronics and Information Eng., Chonbuk National University, PhD.

<Research Interests>
Multimedia and Communications System, Robot Vision, Remote Control.

Hyeon-Joong Yoo [Regular Member]

- 1982 ~ 1989:Researcher Agency for Defense Development
- 1996: Senior Researcher, Industrial Electronics Dept., Samsun Co.
- 1996 ~ Present: Dept. of Information and Telecommunication Eng. Sangmyung University.

<Research Interests>
Pattern Recognition, Image Processing, Neural Networks

Hyongsuk Kim [Regular Member]

- Jan. 1982 ~ Aug. 1992: Senior Researcher, Agency for Defense Development
- May. 1992: University of Missouri, Columbia, PhD.
- Sep. 1993 ~ present: Professor, Div. of Electronics and Information Eng., Chonbuk National University

<Research Interests>
Robot Vision, Sensors of Robots, Analog Parallel Processing System