

The Analysis of a Fractional Pixels Generated by a Combination of Linear FIR Filters in a Video Codec

Jinwuk Seok*, Kisong Yoon*, Bumho Kim*, Jeong-Woo Lee*

*Convergent Contents Research Department, Electronics and Telecommunications Research Institute (ETRI), 138 Gajeongno Yuseong-Gu Daejeon 305-700, Republic of KOREA

Jnwseok@etri.re.kr, ksyoon@etri.re.kr, mots@etri.re.kr,

Abstract— In this paper, we propose an analysis of fractional pixels which are interpolated from integer pixels using hierarchical linear FIR filters based on a differential structure of sub-pixels. In order to improve a video quality under a limited bitrate, most of video codecs employs fractional pixels (or samples) in a process of motion compensation or motion estimation. Even though the fractional pixels have been widely used in the field of video compression, it has not been so much researched to analyze a mathematical structure about the fractional pixels which are interpolated by linear FIR filters. Through the mathematical analysis, we provide a limitation of a resolution for fractional pixels for video codecs, such as h.264/AVC(Advanced video codec) or HEVC(High Efficiency Video Codec). Practically, we find a limitation of the resolution for fractional pixels is in unit of one octa-pixel (1/8-pixels) of the distance between integer pixel positions. Moreover, we find that it is not effective to employ fractional pixels beyond 1/16-pixels by the proposed analysis. The expected improvement of a compression performance using 1/8-pixels is about 35% at most, and the result of the expected improvement is very close to the previous results of research about the adoption of 1/8-pixel to video compression.

Keywords— Fractional pixel, Differential Structure, Motion Estimation, Video Codec

I. INTRODUCTION

In the viewpoint of video compression, a fractional-pixel resolution for a displacement vector provides an improvement of coding efficiency in the both sides of a bit-rate and distortion performance. Especially, a 1/4-pel displacement vector resolution evaluated by 2-phases of multiple finite impulse response (FIR) filters in the H.264/AVC, improves video quality better, in comparison to a previous version of other video standards. However, there still exists a significant problem about a high fractional-pixel resolution of a displacement vector. Is it possible that an employment of a higher fractional-pixel resolution always improves video quality? If a high fractional-pixel resolution does not guarantee a high video quality, what is the limitation of a fractional-pixel resolution? Preceding surveys provided by Thomas Wedi shows the solution of these two questions. When fractional-pixels are evaluated by multiple FIR filters, there exists the limitation of a fractional-pixel resolution with

a displacement vector resolution of 1/16-pel [1] [2] [3] [4] [5] [6] [7].

However, since the preceding surveys are based on an experimental investigation, it is not certain that the preceding limitation surveyed by Thomas Wedi is a general solution. Furthermore the preceding surveys do not show any factors for the limitation of a fractional-pixel. Therefore, based on a mathematical theory, we have to analyse these problems, and verify that the limitation of a fractional-pixel resolution is appropriate to the result of experimental investigation provided by Thomas Wedi. Since the preceding surveys provide sufficient experimental results about the limitation of the fractional-pixel resolution [8], we don't conduct computer simulation for the proposed analysis in this paper. We purely analyse which factor makes the limitation of the fractional-pixel resolution and show the level of the limitation through our analysis. This paper is organized as follows: Section 2 discusses the basic theory of the paper; In section 3, we verify a proof of the proposed theorem; Finally, section 4 provides the conclusions.

II. FUNDAMENTAL THEORY

In this section, we introduce a differential structure of fractional-pixels, which are interpolated from integer pixels [9], to analyse the structure of a fractional-pixel resolution. Suppose that linear n -tap FIR filters generate all the sub-pixels from corresponding integer pixels. In addition, each integer pixel is sampled with h such as $x_0 + h$ where $x_0 \in R^2$. Since we denote an integer pixel as $P(x_0)$, the integer pixels are denoted as $\{P(x) | s_0 \in Z, P(x_0 + s_0 h) \in R[0, 255]\}$, and a fractional-pixels are denoted as $P(x_0 + s_1 \frac{1}{2} h)$. For convenience, we assume that the dimension of a pixel parameter $x_0 \in R^2$ is 1-dimension such that the y -coordinate of a pixel is fixed at 0 such as $x_0 = (x_0, 0)$.

The following theorem is employed throughout this paper without proofs, to illustrate the upper bound of a fractional-pixels used in H.264/AVC or other video codecs [9].

Theorem 1. The norm of the derivative at a pixel $P(x_0)$ satisfies the following inequality:

$$\frac{P(x_0 + h^k) - P(x_0)}{h^k} = \frac{2^{k-1}}{h^0} \prod_{m=1}^{k-1} f_1^{k-m}(P(x_0 + h) - P(x_0)) + o(h^k) \quad (1)$$

where $o(h^k)$ is the high order term corresponding h^k such as $o(h^k) < \frac{\kappa(k-1)}{2} \lambda \kappa (\frac{1}{2})^k$, and the positive value κ is the supremum of the difference between a fractional-pixel and the integer pixel that is nearest the fractional-pixel such that $|P(x_0 + sh^k) - P(x_0)| < \kappa$. k is the number of interpolation process that generates a fractional-pixels. For instance, in the H.264/AVC, it requires 2 steps of up-sampling and low-pass filtering constructed with a multi-tab FIR filters. The number of interpolation processes is 2, i.e. $k = 2$, in the H.264/AVC.

Using the norm of the differential $dP(x_0)$ evaluated in the theorem 1, we can obtain the upper bound of the error $o(h^k)$ such as $\frac{\kappa(k-1)}{2} \lambda \kappa (\frac{1}{2})^k$. Theorem 1 shows that the fractional-pixel evaluated by the first order difference equation corresponding to a dominant coefficient of each FIR filter, has some error with the upper bound. $\frac{\kappa(k-1)}{2} \lambda \kappa (\frac{1}{2})^k$.

Considering the case of which the number of interpolation process is 1, and the coefficient of the FIR filter is $\{1/2, 1/2\}$. Due to the fact that the upper bound of the FIR filter in the above case is zero, the fractional-pixel generated by FIR filters can be correctly represented as the exact difference equation without any error, i.e. $o(h^k) = 0$. If the number of interpolation processes is not one, the difference equation of a fractional-pixel involves always an error term, i.e. $o(h^k) \neq 0$. Moreover, since this error $o(h^k)$ is not a monotone increasing function, the absolute value of the error term has an upper bound, which is derived by following theorem.

Theorem 2. The maximal integer value that maximize the supremum of the error $\frac{\kappa(k-1)}{2} \lambda \kappa (\frac{1}{2})^k$, is 3 or 4 for $Z^{++} = \{x \in Z \mid x \geq 1\}$

Proof To obtain the limit value, we differentiate the supremum of the error $\frac{\kappa(k-1)}{2} \lambda \kappa (\frac{1}{2})^k$ with respect to the number of interpolation process k , then the following equation is evaluated.

$$\begin{aligned} \frac{\partial o(k)}{\partial k} &= \lambda \kappa \frac{1}{2} \frac{\partial}{\partial k} \left(k(k-1) \left(\frac{1}{2} \right)^k \right) = \lambda \kappa \left(\frac{1}{2} \right)^{k+1} (2k-1 - \ln 2(k^2-k)) \\ &= \lambda \kappa \left(\frac{1}{2} \right)^{k+1} (2k-1 - \ln 2(k^2-k)) \\ &= \lambda \kappa \left(\frac{1}{2} \right)^{k+1} \frac{1}{\ln 2} \left(k^2 - (1 + \frac{2}{\ln 2})k + \frac{1}{\ln 2} \right) = 0 \end{aligned} \quad (2)$$

Since the solution of the second-order equation $k^2 - (1 + \frac{2}{\ln 2})k + \frac{1}{\ln 2} = 0$ is same to the solution of the partial differentiation $\frac{\partial o(k)}{\partial k}$, we should evaluate the solution of the second-order equation to obtain the maximum or the

minimum value of the error $o(k)$. The solution of the equation (2) is:

$$k = \frac{1}{2} \left(\left(1 + \frac{1}{\ln 2} \right) \pm \sqrt{\left(1 + \frac{1}{\ln 2} \right)^2 - \frac{8}{\ln 2}} \right). \quad (3)$$

Since the number of interpolation process should be an integer value, we have to find an integer value that is the nearest integer to the real solution (3). When the number of interpolation process is equal to 3 or 4, i.e. $k = 3$, or $k = 4$, the supremum of the error $o(h^k)$ is equal to $0.375 \cdot \lambda \kappa$ that is the maximum value of the $o(h^k)$ when the number of interpolation is an integer value. ■

Furthermore, when the number of interpolation process increases up to infinity, the supremum of $o(h^k)$ decreases down to zero, i.e.

$$\lim_{k \uparrow \infty} o(h^k) = \lim_{k \uparrow \infty} \frac{\kappa(k-1)}{2} \lambda \kappa \left(\frac{1}{2} \right)^k = 0. \quad (4)$$

III. VERIFICATION OF THE DERIVED UPPER BOUND FOR THE INTERPOLATION PROCESSES OF FIR FILTERS

In the previous chapter, we derive the upper bound for the number of interpolation process with hierarchical FIR filters generating a fractional-pixel with the differential form. However, even though the theorem 2 shows the upper bound of the supremum of the error corresponding to the differential expansion, the theorem 2 does not illustrate that the derived upper bound affects to the rate-distortion characteristic of a fractional-pixels generated by the hierarchical FIR-filters. If the difference between fractional-pixels and integer pixels is not bounded in spite of the bounded error by the supremum of the interpolation processes, the result of the previous chapter is meaningless.

In this chapter, through the analysis of differences between integer-pixels and fractional-pixels that are generated by hierarchical FIR-filters, we verify that the difference between fractional-pixels and integer pixels is also bounded by the upper bound for the number of the interpolation processes.

For the purpose of verifying the usefulness of the upper bound, we suppose that the object function that illustrates the difference between fractional-pixels and integer pixels, is a linear combination of the squares of pixel differences, such that $f(x, t, k) = L((P(x_0; t) - P(x_0 + h^k; t-1))^2)$, where $P(x; t)$ is a pixel data with a spatial coordinate x at time t , and L is a linear operator such that $\frac{\partial L(h(x))}{\partial x} = L \frac{\partial h(x)}{\partial x}$. In the viewpoint of video codecs, the motion estimation is a process which minimizes the object function that is a summation of the differences between pixels. Therefore, since a summation is a typical linear operation, it is feasible assumption that the object function is a linear combination of differences between pixels.

Definition 1. We define the index set $R(2^{-k})^2$ such that

$$R(2^{-k})^2 = \{x \in R^2 \mid x = x_0 + s \cdot h^k, s \in Z^2\}. \quad (5)$$

In addition, we define the time index t such that

$$\Lambda = \{t \in Z[0, n] \mid t \leq M\} \quad (6)$$

where M is the maximum positive number of frames in the memory of an H.264/AVC encoder.

Assumption 1. According to the result of the theorem 1, a fractional-pixel which is interpolated by hierarchical FIR-filters is evaluated by the following differential equation

$$P(x_0 + h^k; \cdot) = P(x_0; \cdot) + \bar{g}(k)(P(x_0 + h; \cdot) - P(x_0; \cdot)) \quad (7)$$

where the parameter $\bar{g}(k)$ is a brief of the term, $h^k \frac{2^{k-1}}{h^0} \prod_{m=1}^k f_1^{k-m}$

Theorem 3. The minimum difference of the object function $h(x, t, k) = \min_{x,t} f(x, t, k) - f(x, t, 1) \in C^2$ is a linear combination of the error $o(k)$ for the differential expansion of fractional-pixels.

Proof The value of object function $f(x)$ at $x_0 + sh^k$ can be evaluated as follows.

$$\begin{aligned} \bar{h}(x, t, k) &= f(x, t, k) - f(x, t, 1) \\ &= L\left((P(x_0; t) - P(x_0 + h^k; t-1))^2\right)_k \\ &\quad - L\left((P(x_0; t) - P(x_0 + h^k; t-1))^2\right)_1 \\ &= L\left((P(x_0; t) - P(x_0; t-1) \right. \\ &\quad \left. - \bar{g}(k)(P(x_0 + h; t-1) - P(x_0; t-1)) \right. \\ &\quad \left. + o(x_0, t-1, k))^2\right) \\ &\quad - (P(x_0; t) - P(x_0; t-1) - \bar{g}(k)P(x_0 + h; t-1) \\ &\quad - P(x_0; t-1))^2) \\ &= L(o(x_0, t-1, k) \cdot (P(x_0; t) - P(x_0; t-1) \\ &\quad - \bar{g}(k)(P(x_0 + h; t-1) - P(x_0; t-1)) + o(x_0, t-1, k))) \end{aligned} \quad (8)$$

For the evaluation of the equation (8), we set a function $\Psi(x_0, t, h)$ such that

$$\begin{aligned} \Psi(x_0, t, h) &= \\ P(x_0; t) - P(x_0; t-1) - \bar{g}(k)(P(x_0 + h; t-1) - P(x_0; t-1)) \end{aligned} \quad (9)$$

Then the equation (8) is evaluated as follows:

$$\begin{aligned} \bar{h}(x, t, k) &= L(o(x_0, t-1, k) \cdot (\Psi(x_0, t, h) + o(x_0, t-1, k))) \\ &= L\left((o(x_0, t-1, k) + \frac{1}{2}\Psi(x_0, t, h))^2 - \frac{1}{4}\Psi^2(x_0, t, h)\right) \end{aligned} \quad (10)$$

Since the object function is defined as $\min_{x,t} \bar{h}(x, t, k)$ for all coordinate $x \in R(2^{-k})^2$ and $t \in \Lambda$, the object function is evaluated by follows:

$$\begin{aligned} h(x, t, k) &= \min_{x,t} \bar{h}(x, t, k) \Big|_{o(x_0, t-1, k) = -\frac{1}{2}\Psi(x_0, t, h)} \\ &= -\frac{1}{4}L\Psi^2(x_0, t, h) \\ &= -L(o^2(x_0, t-1, k)) \end{aligned} \quad (11)$$

Consequently, the minimum difference of the object function is a linear combination of the error $o(k)$. ■

By the result of the theorem 3, since the minimum difference of the object function is an isomorphism with respect to the error $o(k)$, the minimum difference is also bounded if the error $o(k)$ is bounded. In the viewpoint of which the performance of video codec coincides with the distortion between an input picture and a reference picture, it means that we can evaluate the improvement of the video codec's performance, if we know an absolute value of the supremum of the expansion error $o(k)$. For instance, in the H.264/AVC, the number of interpolation process is 2. In addition, we assume that there is a video codec involving a fractional-pixels generated by interpolation processes of 3 steps. According to the theorem 1, the supremum value of the expansion error in the H.264/AVC is $0.25 \cdot \lambda \kappa$, and it is $0.375 \cdot \lambda \kappa$ for the video codec that the number of interpolation process is 3. Therefore, we can evaluate that the improvement of the performance in the video codec involving the fractional-pixels generated by interpolation of 3 steps is approximately 33.3%, in comparison to the H.264/AVC. This evaluation is almost equal to the results of experiment provided by the Thomas Wedi.

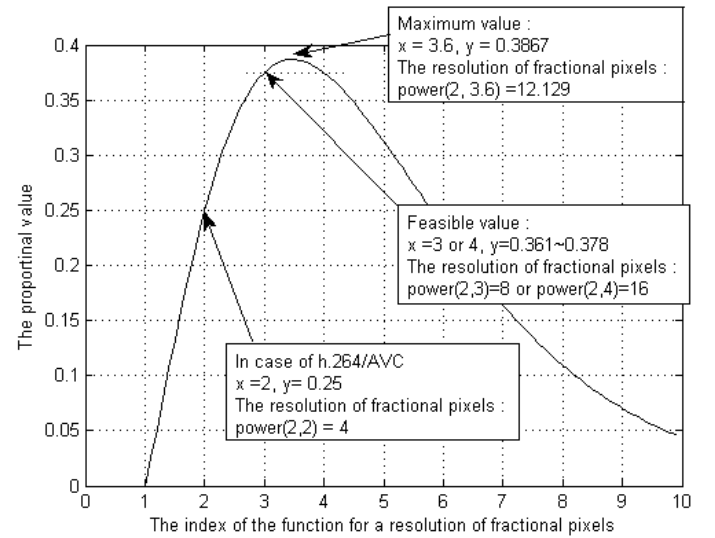


Figure 1. The plot of upper bound for error term $\frac{k(k-1)}{2} \lambda \kappa (\frac{1}{2})^k$, when $\lambda = 2, \kappa = 2$

IV. CONCLUSIONS

We propose an upper bound of fractional-pixel displacement vector resolution which is derived by hierarchical FIR-filters for H.264 video coding. According to the differential expansion of fractional-pixels which is interpolated by hierarchical FIR-filters at a fractional-pixel resolution, we derive the upper bound of the error terms which compensate the value of a fractional-pixel approximated by the first-order differential. Additionally, by the effect of the proposed upper bound, we show that the improvement of video quality is the maximum level at the number of interpolation processes of 3 or 4.

In the equation (10), to obtain the best video quality from a video encoder, the error of the differential expansion $o(x_0, t-1, k)$ should be equal to $-\frac{1}{2}\Psi(x_0, t, h)$. However, by many undetectable factors, it can be possible that the error $o(x_0, t-1, k)$ is not correctly equal to $-\frac{1}{2}\Psi(x_0, t, h)$. Especially, since the fractional-pixel of 1/16-pel is finer resolution than the resolution of the fractional-pixel of 1/8-pel, the possibility of minimizing the function $\bar{h}(x, t, k)$ is higher than that of the fractional-pixel of 1/8-pel. Thus, comparing to the fractional-pixel of 1/8-pel, the rate-distortion characteristics for the fractional-pixel of 1/16-pel shows same or better performance. This result fully coincides with T. Wedi's experimental results for the resolution of 1/8-pel and 1/16-pel.

Consequently, in this paper, we conclude that the improvement of video quality which depends on hierarchical linear FIR-filters is limited. To overcome this limitation of improvement in video quality, we should develop other theories of fractional-pixels, which can avoid the proposed upper bound based on the linear filter theory.

ACKNOWLEDGMENT

This work was supported by the IT R&D program of MCT/KOCCA[2-09-1205-001-10987-09-001 Development of DCI-compliant digital cinema distribution management and copyright protection technology].

REFERENCES

- [1] T. Wedi and H. G. Musmann, "Motion and Aliasing Compensated Prediction for Hybrid Video Coding", *IEEE Transactions on Circuits and Systems for Video Technology*, vol 13, pp 577-586, No. 7, July 2003.
- [2] Thomas Wedi, "Adaptive Interpolation Filters and High-Resolution Displacements for Video Coding", *IEEE Transactions on Circuits and Systems for Video Technology*, vol 16, pp 484-491, No. 4, April 2006.
- [3] S T. Wedi. "Adaptive 3D Interpolation Filter for Motion and Aliasing Compensated Prediction", in *Proc. Picture Coding Symposium (PCS 2003)*, Saint-Malo, France, April 2003.
- [4] T. Wedi, "Adaptive Interpolation Filter for Motion Compensated Prediction". *Proc. IEEE International Conference on Image Processing*, Rochester, New York USA, September 2002.
- [5] T. Wedi, "Adaptive Interpolation Filter for Motion and Aliasing Compensated Prediction", *IEEE Transactions on Circuits and Systems for Video Technology, Proc. Electronic Imaging 2002: Visual Communications and Image Processing*, San Jose, California USA, January 2002.
- [6] T. Wedi, "Adaptive Interpolation Filter for Motion Compensated Hybrid Video Coding", *Proc. Picture Coding Symposium*, Seoul, Korea, April 2001/
- [7] T. Wedi. "Hybrid Video Coding Based on High-Resolution Displacement Vectors", *Proc. Electronic Imaging 2001: Visual Communications and Image Processing*, San Jose, California USA, January 2001
- [8] J. V. T. of ISO/IEC MPEG ITU-T VCE. *Draft itu-t recommendation and final draft international standard of joint video specification*. ITU-T Rec. H.264 | ISO/IEC 14496-10 AVC, September 2001..
- [9] J. Seok, J. Lee, and C. Cho, "The differential structure of sub pixels interpolated from integer pixels using n-tab FIR filters for high definition H.264 video encoding", in *Proc. ACM Multimedia*, 2006, pp.667-670.