BANDWIDTH REDUCTION SCHEMES FOR MPEG-2 TO H.264 TRANSCODER DESIGN

[†]*Xianghui Wei, Wenqi You, Guifen Tian, Yan Zhuang, Takeshi Ikenaga, Satoshi Goto*

Graduate School of Information, Production and Systems, Waseda University
2-7 Hibikino Wakamatsu-ku Kitakyushu, 808-0135, Fukuoka, Japan
phone: + (81) 93-692-5298, fax: + (81) 93-692-5298, [†]email: wxh@ruri.waseda.jp
http://www.waseda.jp/ips/

ABSTRACT

In MPEG-2 to H.264 transcoding, MPEG-2 motion vector (MV) reuse is an effective technique to simplify motion estimation (ME) processing. The irregularity of MPEG-2 MV also brings difficulty in applying data reuse method for hardware design, which plays a critical role for bandwidth reduction. In this paper, two search window reuse methods are introduced for HDTV application. The Level C method utilizes horizontal search window reuse for most macroblock (MB), while the Level C+ method is designed for ultra-low bandwidth applications by reusing both horizontal and vertical search window. Experiment results show both methods can achieve a low bandwidth level with negligible video quality loss.

1. INTRODUCTION

Video transcoding performs operations to transform one compressed video stream to another [1]. These operations include bit rate, frame rate, spatial resolution, coding syntax and content transform, etc. In [2], transcoding has been classified into two categories: homogeneous and heterogeneous transcoding. Two types of transcoder structure are often used: close-loop and open-loop. This paper focuses on heterogeneous transcoding between MPEG-2 and H.264 video coding standard with a close-loop structure.

Many efforts have been contributed to develop an efficient MPEG-2 to H.264 transcoding algorithm. The MPEG-2 motion vector reuse is a commonly employed method to reduce computation of ME. In [3], the decoded MPEG-2 motion vector is used as one of motion vector predictors (MVP) in enhanced predictive zonal search (EPZS) algorithm. In [4], MEPG-2 motion vector is also reused as MVP. Then MVP selection, motion vector refinement and a top-down splitting strategy for sub-block motion vector re-estimation are applied. In MVP selection, motion costs of MPEG-2 motion vector and motion vectors from neighboring blocks are compared. The motion vector with the smallest cost is chosen as search center.

A critical issue in hardware design is bandwidth reduction because external bandwidth is a limited resource in hardware. Several methods have been proposed for full search block matching algorithms (FSBMA). In [5], four search region data reuse methods from Level A to Level D have been discussed. In [6], a Level C+ algorithm and its associated n-stitched zig-zag scan method have been introduced. This method utilizes horizontal and partially vertical overlapping area of search window by stitching several MB vertically.

All these existing search window reuse method is based on regular search position in reference frame. As for MPEG-2 to H.264 transcoding, it is difficult to directly apply these methods, because introducing MPEG-2 motion vector as search center will lead to irregular overlapping between successive search windows. In this paper, two search window reuse methods, Level C and Level C+, are proposed for MPEG-2 to H.264 transcoder design. In section 2, the proposed search window reuse schemes are introduced. Experiment results will be shown in section 3. In section 4, we will draw some conclusions.

2. SEARCH WINDOW REUSE SCHEMES FOR TRANSDOER DESIGN

Motion estimation is the most computation and data intensive part in video coding system. In transcoding, MV from MPEG-2 reused as search center in ME module of H.264 encoder is a simple but effective technique to simplify ME procedure, because MPEG-2 MV can precisely indicate the matching position in reference frame. Thus many computations can be saved. In [3], the decoded MPEG-2 MV is reused as one of MVP to accelerate motion estimation procedure in H.264 encoder. In [4], MPEG-2 MV is also reused as MVP and then a top-down splitting procedure is applied to decide sub-block MV.

In transcoder design, the kernel problem for applying search window reuse method is to address the irregularity of MPEG-2 MV. The proposed Level C method utilizes the similarity of neighboring MV to regularize the search center position for most MBs. The proposed Level C+ method applies a pre-processing procedure, called motion vector clipping, to handle this problem. These two methods will be discussed in detail in following sections.

In following discussion, search range is defined as $[-p_h, p_h)$ in horizontal direction and $[-p_v, p_v)$ in vertical direction; MB size is defined as $N \times N$. Accordingly, search window size is $(sr_h + N - 1)(sr_v + N - 1)$, where $sr_h = 2p_h$ and $sr_v = 2p_v$. The raster-scan is assumed to be the processing order for H.264/AVC ME.

2.1 Level C Search Window Reuse Scheme for Transcoder Design

2.1.1 Overall Algorithm

The Level C scheme is based on the fact that neighboring MV often have similar value. If the MV difference between successive MVs is less than a threshold, they are assumed to the same and are set as search center for successive macro-blocks. After the regularization processing, search window reuse is available for most macro-block. The Level C algorithm is as following

for each MB_i in sequence
{
  if $|\Delta\overrightarrow{mv}| = |\overrightarrow{mv}_i - \overrightarrow{mv}_{i-1}| < t$ then
  {
    $x_i = x_{i-1} + 16$
    $y_i = y_{i-1}$
    $SW_{i-1} = \mathbf{SearchWindow}(\overrightarrow{mv}_{i-1});$
    $SW_{reuse} = SW_i \bigcap SW_{i-1};$
    $SW_{loaded} = SW_i - SW_{reuse};$
    $SW_i = SW_{loaded} \bigcup SW_{reuse};$
  }
  else
    $SW_i = SW_{loaded} = \mathbf{SearchWindow}(\overrightarrow{mv}_i);$
  $\mathbf{MotionEstimation}(SW_i);$
}

The raster scan is assumed to be the processing order for H.264 ME. $\overrightarrow{mv}_i$ and $\overrightarrow{mv}_{i-1}$ are current and previous MPEG-2 MV; (x_i, y_i)

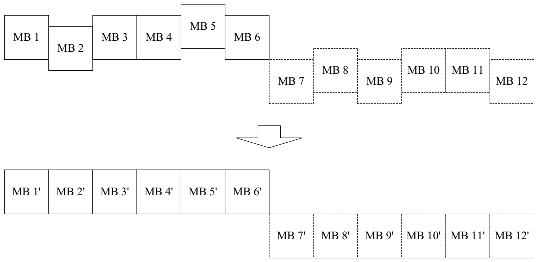

Figure 1: An example of MV regularization.

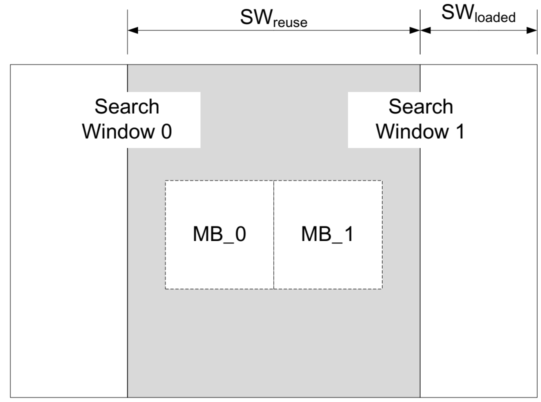

Figure 2: Level C data reuse scheme.

and (x_{i-1}, x_{i-1}) are their coordinates; t is a predefined threshold which is set to 6; **SearchWindow** is the function to determine search window (SW_i) based on $\overrightarrow{mv}_i$ as shown in Equation 1; **MotionEstimation** represents the function of motion estimation performed within SW_i.

$$\begin{aligned} SW_i &= \mathbf{SearchWindow}(\overrightarrow{mv}_i) \\ &= \{(x,y) | x \in [x_i - p_h, x_i + p_h + N - 1), \\ &\qquad y \in [y_i - p_v, y_i + p_v + N - 1)\} \end{aligned} \tag{1}$$

Figure 1 shows an example of MB sequences before and after MV regularization. The regularized MBs can utilize Level C search window reuse scheme, which is shown in Figure 2.

2.1.2 *Threshold Determination*

The reuse-rate p_{reuse} in the proposed method is defined as

$$p_{reuse} = \frac{n_{reuse}}{N_{total}}$$

where n_{reuse} is the number of MB which can utilize the Level C search window reuse under the proposed algorithm framework. N_{total} is the total number of MB in the whole sequence.

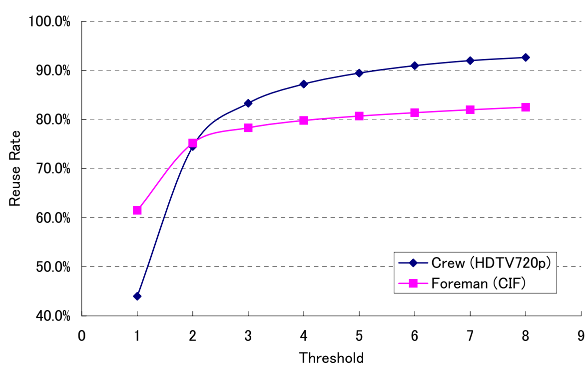

Figure 3: Threshold value versus reuse-rate.

Obviously p_{reuse} determines bandwidth level in the proposed algorithm, because higher p_{reuse} means more MB can utilize search window reuse. So we must test which threshold value can achieve a usable reuse-rate. Experiments are conducted in Foreman (CIF) and Crew (HDTV720p) sequence. Fig.3 shows the experiment result. It is observed that about 80% (Foreman) and 90% (Crew) MB can be regularized if the threshold is set to 6. Further increasing the threshold will lead to slowly increase in reuse-rate; this is mainly due to the sparse distribution of MB whose MV difference is larger than the threshold.

2.1.3 *Reuse Rate*

The next critical issue is that how many MB can utilize search window reuse after MV regularization, which determines the external bandwidth level. The reuse rate of four CIF sequences and five HDTV720p sequences are tested. The results are shown in Table 1. It is observed that most of MB both in CIF and HDTV720p sequences can utilize search window reuse under the proposed method. In general, the HDTV720p sequence achieves higher reuse-rate than CIF sequence, because HDTV720p sequence has higher resolution and thus a lower difference between pixels of neighboring MB. So the MV similarity in HDTV720p sequence is higher than CIF.

Table 1: Search window reuse rate comparison.

Sequence		Reuse rate
HDTV720p	Crew	90.7%
	KnightShield	91.3%
	City	92.8%
	Harbour	93.2%
	Parkrun	97.5%
CIF	Canoa	74.0%
	Container	95.5%
	Foreman	81.9%
	Stefan	85.5%

2.1.4 *Performance Evaluation*

Two factors must be taken into consideration to evaluate performance of data reuse scheme: on-chip memory size for reference frame and redundancy access factor [6]. The on-chip memory size represent required buffer of candidate block for data reuse. The redundancy access factor R_α evaluates external bandwidth and is defined as the number of reference pixel be loaded for each MB pixel. $\overline{R}_\alpha$ of the proposed Level C method is calculated as the expectation of R_α of all MB, which is shown as following

$$\begin{aligned} \overline{R}_\alpha &= \mathbf{E}(R_\alpha) \\ &= p_{reuse} \cdot R_{\alpha(Level\,C)} + (1 - p_{reuse}) \cdot R_{\alpha(No\,Reuse)} \end{aligned} \tag{2}$$

where p_{reuse} is the reuse rate mentioned above. $R_{\alpha(Level\,C)}$ and $R_{\alpha(No\,Reuse)}$ are calculated as following

$$R_{\alpha(Level\,C)} = \frac{N(sr_v+N-1)}{N^2}$$
$$R_{\alpha(No\,Reuse)} = \frac{(sr_h+N-1)(sr_v+N-1)}{N^2} \quad (3)$$

The on-chip memory size for this method is equal to 2 times of the size of search window, which is $2(sr_h+N-1)(sr_v+N-1)$. This is because current reference buffer must be flushed when the coordinate difference is larger than the threshold; and the next MB must load reference pixel from another memory.

2.2 Level C+ Search Window Reuse Scheme for Transcoder Design

Although the method proposed in last section can utilize horizontal search window reuse for most MB, further improvement is possible by utilizing both horizontal and vertical search window reuse. The fundamental idea of the proposed method is to load search window for a group of MB stitched together. MB of the same group can reuse search window horizontally and vertically. By this way, the average reference pixel for each MB can be reduced. This method is called Level C+ scheme in this paper. It is primarily designed for ultra-low bandwidth applications, while on-chip memory is not a main constraining factor.

2.2.1 *Motion Estimation Unit (MEU)*

The MB group in the proposed method is defined as Motion Estimation Unit (MEU), which is the unit of reference pixel loading and motion estimation. It is composed of v MB in vertical direction and h MB in horizontal direction as shown in Equation 4. Figure 4 shows an example of MEU.

$$MEU = \{MB_{ij} | 0 < i \le v, 0 < j \le h\}$$
$$\overrightarrow{mv}_{MEU} = \frac{1}{v\cdot h}\sum_{i=1}^{v}\sum_{j=1}^{h}\overrightarrow{mv}_{ij} \quad (4)$$

To load search window for MEU, search center of MEU ($\overrightarrow{mv}_{MEU}$) must be first determined, which is calculated as average $\overrightarrow{mv}_i$ of each MB in MEU. In this step, we make the assumption that each MB in MEU has the same MV as $\overrightarrow{mv}_{MEU}$; thus after the search center of MEU is determined, the boundary of search window of MEU (SW_{MEU}) is also determined.

2.2.2 *Overall Algorithm*

The proposed Level C+ method is described in detail as following

SW_{MEU} = $\mathbf{SearchWindowMEU}(\overrightarrow{mv}_{MEU})$;
for each $MB_{ij} \in MEU$
{
  $SW_{ij} = \mathbf{SearchWindow}(\overrightarrow{mv}_{ij})$;
  if $SW_{ij} \subset SW_{MEU}$ then
    $\mathbf{MotionEstimation}(SW_{ij})$;
  else
  {
    $\overrightarrow{mv}'_{ij} = \mathbf{MVClipping}(\overrightarrow{mv}_{ij})$;
    $SW'_{ij} = \mathbf{SearchWindow}(\overrightarrow{mv}'_{ij})$;
    $\mathbf{MotionEstimation}(SW'_{ij})$;
  }
}

where **SearchWindow** is the function to determine search window (SW_{ij}) based on $\overrightarrow{mv}_{ij}$ as shown in Equation 1; **SearchWindowMEU** is the function to determine search window for MEU (Equation 5); **MVClipping** represents the processing of motion vector clipping. It performs on $\overrightarrow{mv}_{ij}$ which goes beyond the boundary of SW_{MEU} to make sure that each MB in MEU can load reference pixels from the same on-chip memory; thus both horizontal and vertical search window reuse is available. Figure 5 shows an example.

$$\begin{aligned} SW_{MEU} &= \mathbf{SearchWindowMEU}(\overrightarrow{mv}_{MEU}) \\ &= \{(x,y)| \\ &\quad x \in [x_{MEU}-p_h, x_{MEU}+p_h+N\cdot h-1), \\ &\quad y \in [y_{MEU}-p_v, y_{MEU}+p_v+N\cdot v-1)\} \end{aligned} \quad (5)$$

2.2.3 *Performance Evaluation and Optimization of MEU Size*

The redundancy access factor R_α of the proposed Level C+ method is calculated as following

$$R_\alpha = \frac{(sr_h+h\cdot N-1)(sr_v+v\cdot N-1)}{v\cdot h\cdot N^2} \quad (6)$$

The size of SW_{MEU} is $(sr_h+h\cdot N-1)(sr_v+v\cdot N-1)$, which is also equal to the size of on-chip memory ($M_{on-chip}$).

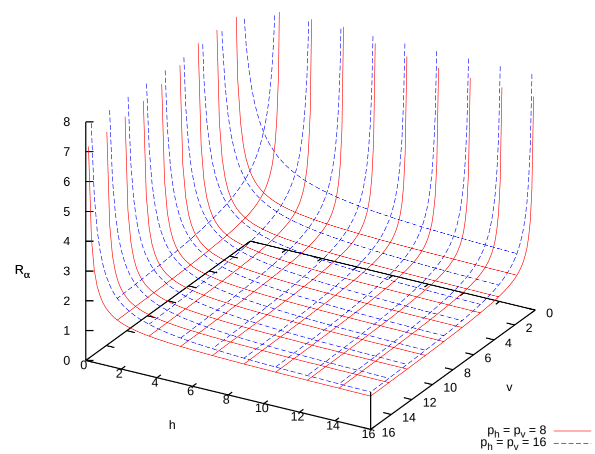

Figure 6: Bandwidth comparison under different v, h, p_h and p_v.

In Figure 6, the relationship between v, h and R_α is presented. It is observed that R_α will reach a very low level (approaching 1) if v and h are large enough. But too large v and h will lead to unacceptable size of on-chip memory. A proper size of MEU must reach a balance between external bandwidth and on-chip memory. This problem can be addressed by solving the following optimization problem.

$$\begin{aligned} \min_{v,h} \quad & z = f(v,h) = R_\alpha + \frac{1}{K}\cdot M_{on-chip} \\ \text{s.t.} \quad & 0 < v \le N_v, \\ & 0 < h \le N_h, \\ & N_v \mod v = 0, \\ & N_h \mod h = 0, \\ & v > h. \end{aligned} \quad (7)$$

The z is the cost function of variable v and h, which is composed of redundancy access factor R_α (Equation 6) and $M_{on-chip}$ as mentioned in last section. K is the normalizer of on-chip memory, which is set to the on-chip memory size of standard H.264 IME module (4 MB) as shown in Table 3. N_v and N_h is the number of MB in vertical and horizontal direction. $p_v = p_h = 16$ is set for HDTV720p sequence and $p_v = p_h = 8$ is set for CIF sequence.

$N_v \mod v = 0$ and $N_h \mod h = 0$ must be satisfied for a convenient hardware implementation. $v > h$ is set because the horizontal variation of motion vector is larger than in vertical direction. The optimization result for HDTC720p is $v = 9, h = 4$ and for CIF is $v = 6, h = 2$.

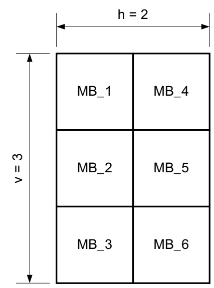

Figure 4: An example of MEU ($v = 3$, $h = 2$).

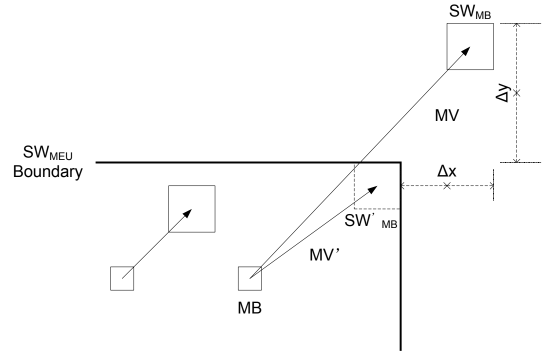

Figure 5: An example of motion vector clipping.

3. EXPERIMENT RESULT

Experiments are conducted with CIF and HDTV720p sequence. The MPEG-2 reference software [7] and the H.264 reference software JM 11 [8] are used to construct the transcoding platform and to simulate the proposed search window reuse algorithm. In MPEG-2 encoder, search range is set to $[-32, 32)$ and $[-64, 64)$ for CIF and HDTV720p sequence. All sequences are coded with IPPP pattern in MPEG-2 encoder. The MV of Intra MB in P-picture is set to the same value as previous MV; and the MV of P_Skip MB in P-picture is set to zero.

3.1 Video Quality Comparison for Level C and Level C+ Scheme

Video quality must be compared between ME algorithms with and without search window reuse scheme, because both the proposed Level C and Level C+ schemes will lead to video quality change. It is introduced by modification of MV. This section presents experiment results on video quality comparison. The test sequences are four CIF and four HDTV720p sequences. Three transcoding algorithms are compared using the method proposed by [9]. The first method is served as baseline.

- A transcoding algorithm reusing MPEG-2 MV as search center. No search window reuse scheme is implemented. (Baseline)
- A transcoding algorithm reusing MPEG-2 MV as search center. The proposed Level C search window reuse scheme is applied.
- A transcoding algorithm reusing MPEG-2 MV as search center. The proposed Level C+ search window reuse scheme is applied.

Table 2 shows the video quality comparison. It is observed that there exists very small difference of video quality. The HDTV720p sequences have smaller difference because of its higher similarity of MV among neighboring MB. Video quality improvement is observed in Level C+ scheme for sequence with fast motion and complicate background, for example in Football and Stefan (CIF). This is because the motion vector clipping processing produces a smoother MV field by punishing MV going beyond of SW_{MEU}. [10] indicates that a smooth motion vector field can represent object motion more accurate.

3.2 Bandwidth Comparison

The bandwidth and on-chip memory of five IME modules are compared based on equation of [6], section 2.1 and 2.2. The test sequence is HDTV720p. The search range is set to $[-64, 64)$ for H.264 and $[-16, 16)$ for transcoder.

- **H.264 (Level C)** A regular H.264 integer motion estimation (IME) module with Level C scheme.
- **H.264 (Levle C+)** A regular H.264 IME module with Level C+, 2 and 4 MB stitched vertically.
- **Transcoder (No Reuse)** A transcoder IME module without any data reuse scheme.
- **Trandscoer (Level C)** A transcoder IME module with Level C. Reuse-rate is 90%.
- **Trandscoer (Level C+)** A transcoder IME module with Level C+ ($v = 9, h = 4$).

Table 3: Bandwidth and on-chip memory comparison.

		R_α	On-chip Memory (Kbyte)
H.264 (Levle C)		8.94	20.0
H.264 (Level C+)	2 MB	4.97	24.7
	4 MB	2.98	35.6
Transcoder (No Reuse)		8.63	2.2
Transcoder (Level C)		3.51	4.3
Transcoder (Level C+)		1.80	16.2

It is observed from Table 3 that the redundancy access factor R_α of the proposed Level C scheme for transcoding is almost equal to standard H.264 IME with Level C scheme. But the on-chip is much smaller (only 17.4% (2 MB) and 12.1% (4 MB) of standard H.264 IME). The redundancy access factor R_α of the proposed Level C+ scheme is smaller than 2, which means that the proposed method needs no more than two reference pixels for motion estimation of each MB pixel. The on-chip memory of the proposed method is 65.6% (2 MB) and 45.5% (4 MB) of standard H.264 with Level C+, which is acceptable.

4. CONCLUSION

In this paper, Level C and Level C+ search window reuse scheme for MPEG-2 to H.264 transcoding are proposed. The Level C scheme can achieve the same level of bandwidth as standard H.264 IME with Level C scheme with a much smaller on-chip memory. The Level C+ method can achieve a very low external bandwidth ($R_\alpha < 2$) with an acceptable on-chip memory by applying both horizontal and vertical search window reuse. Compared with the standard Level C+ method which loads search window for a group of MB connected vertically, the proposed one loads search window for a group of neighboring MB both in horizontal and vertical direction. Combined with reduced search range introduced by MPEG-2 MV, ultra-low bandwidth can be achieved. Experiment results show that both methods have negligible video quality degradation.

The memory index problem introduced by reusing MPEG-2 MV is important because less memory access will reduce power dissipation. For the proposed Level C method, on-chip memory access is the same as the standard Level C method, while the memory update can be controlled by a modified ping-pang strategy. That is when one memory is accessed, the other one is updated. For the proposed Level C+ method, on-chip memory access times can be reduced by further regularizing the position of search window in the same row. Considering the similarity between successive search windows, this regularization will not greatly degrade video quality.

Table 2: Video quality comparison.

Sequence		Level C		Level C+	
		ΔPSNR	ΔBitrate	ΔPSNR	ΔBitrate
CIF	Container	-0.0004 dB	+0.0007%	-0.0000 dB	+0.0000%
	Football	-0.0086 dB	+0.1314%	+0.0349 dB	-0.6404%
	Foreman	-0.0002 dB	+0.0055%	-0.0020 dB	+0.0312%
	Stefan	-0.0079 dB	+0.1698%	+0.0294 dB	-0.5952%
HDTV720p	City	-0.0002 dB	+0.0073%	+0.0031 dB	-0.0951%
	Harbour	-0.0011 dB	+0.0433%	-0.0004 dB	+0.0101%
	Knightshields	+0.0019 dB	-0.0359%	+0.0003 dB	+0.0051%
	Parkrun	-0.0009 dB	+0.0518%	+0.0003 dB	-0.0051%

In the proposed Level C method, the threshold is fixed. This could be further improved by defining a adaptive threshold based on video context. In addition, the threshold can be chosen as different value for vertical and horizontal direction. We will testify this idea in our further work.

Our future work will involve a hardware implementation of the proposed search window reuse schemes which is included in a MPEG-2 to H.264 transcoder chip design for HDTV application.

REFERENCES

[1] J. Xin, C.-W. Lin, and M.-T. Sun, "Digital video transcoding," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 84–97, Jan. 2005.

[2] I. Ahmad, X. Wei, Y. Sun, and Y.-Q. Zhang, "Video transcoding: an overview of various techniques and research issues," *IEEE Trans. on Multimedia*, vol. 7, no. 5, pp. 793–804, Oct. 2005.

[3] X. Lu, A. Tourapis, P. Yin, and J. Boyce, "Fast mode decision and motion estimation for h.264 with a focus on mpeg-2/h.264 transcoding," *Proc. of ISCAS2005*, vol. 2, pp. 1246–1249, 23-26 May 2005.

[4] Z. Zhou, S. Sun, S. Lei, and M.-T. Sun, "Motion information and coding mode reuse for mpeg-2 to h.264 transcoding," *Proc. of ISCAS2005*, vol. 2, pp. 1230–1233, 23-26 May 2005.

[5] J.-C. Tuan, T.-S. Chang, and C.-W. Jen, "On the data reuse and memory bandwidth analysis for full-search block-matching vlsi architecture," *IEEE Trans. on Circ. and Syst. for Video Tech.*, vol. 12, no. 1, pp. 61–72, Jan. 2002.

[6] C.-Y. Chen, C.-T. Huang, Y.-H. Chen, and L.-G. Chen, "Level c+ data reuse scheme for motion estimation with corresponding coding orders," *IEEE Trans. on Circ. and Syst. for Video Tech.*, vol. 16, no. 4, pp. 553–558, Apr. 2006.

[7] "Mpeg-2 reference software, test model 5 (tm5) [online]. available: http://www.mpeg.org/mpeg/mssg/," 2008.

[8] "H.264/avc reference software jm11.0 [online]. available: http://www.iphome.hhi.de/suehring/tml/download/," 2008.

[9] G. Bjontegaard, "Calculation of average psnr differences between rd-curves," *13th VCEG-M33 Meeting*, pp. 1–4, Apr. 2001.

[10] H.-M. Wong, O. Au, C.-W. Ho, and S.-K. Yip, "Enhanced predictive motion vector field adaptive search technique (e-pmvfast)-based on future mv prediction," *Proc. of ICME2005*, pp. 6–8, July 2005.