行政院國家科學委員會專題研究計畫 成果報告

新興高效能視訊編碼技術
研究成果報告(精簡版)

計 畫 類 別 ：個別型
計 畫 編 號 ：NSC 99-2221-E-009-017-
執 行 期 間 ：99年08月01日至100年07月31日
執 行 單 位 ：國立交通大學資訊工程學系（所）

計畫主持人：彭文孝

計畫參與人員：碩士班研究生-兼任助理人員：吳崇豪
碩士班研究生-兼任助理人員：黃嘉彥
碩士班研究生-兼任助理人員：陳孟傑
碩士班研究生-兼任助理人員：李宗霖
碩士班研究生-兼任助理人員：曾于真
碩士班研究生-兼任助理人員：陳彥宇
碩士班研究生-兼任助理人員：吳牧軒
博士班研究生-兼任助理人員：陳俊吉
博士班研究生-兼任助理人員：詹家欣
博士班研究生-兼任助理人員：陳漪紋

報 告 附 件 ：出席國際會議研究心得報告及發表論文

處 理 方 式 ：本計畫可公開查詢

中 華 民 國 100年10月31日

附件一

行政院國家科學委員會補助專題研究計畫 ■成果報告 □期中進度報告

新興高效能視訊編碼技術

計畫類別：■個別型計畫　□整合型計畫
計畫編號： 99-2221-E-009-017
執行期間： 2010 年 08 月 01 日至 2011 年 07 月 31 日

執行機構及系所：國立交通大學多媒體工程研究所

計畫主持人：彭文孝
共同主持人：
計畫參與人員：詹家欣、李宗霖、陳俊吉、陳漪文、吳牧軒
吳崇豪、黃嘉彥、陳孟傑、曾于真、陳彥宇

成果報告類型(依經費核定清單規定繳交)：■精簡報告　□完整報告

本計畫除繳交成果報告外，另須繳交以下出國心得報告：
■赴國外出差或研習心得報告
□赴大陸地區出差或研習心得報告
□出席國際學術會議心得報告
□國際合作研究計畫國外研究報告

處理方式：除列管計畫及下列情形者外，得立即公開查詢
■涉及專利或其他智慧財產權，□一年■二年後可公開查詢

中　華　民　國　100　年　10　月　24　日

行政院國家科學委員會補助專題研究計畫成果報告

Bi-prediction Combining Template and Block Motion Compensations

計畫編號：99-2221-E-009-017
執行期限：2010年08月01日至 2011年07月31日
主持人：彭文孝 教授 國立交通大學資訊工程系所
計畫參與人員：詹家欣、李宗霖、陳俊吉、陳漪文、吳牧軒、吳崇豪、黃嘉彥、陳孟傑、曾于真、陳彥宇 國立交通大學資訊工程系所

一、中文摘要

關鍵詞： 雙向預測、交疊區塊動作補償、樣板核對
此篇報告提出一個如同單向預測僅需一個動作向量傳輸負擔的雙向預測機制。此方法透過參數化交疊區塊動作補償的方法，不僅結合樣板核對的動作向量，且重新設計了區塊動作搜索的機制，而其中所使用的最佳視窗函數是基於模型的框架所設計的。此外，我們結合了可適性動作合併機制達到了無樣板核對的實作。根據 JCT-VC 所提出之一般測試條件，三種分別代表不同效能及複雜度取捨的演算法實作於 TMuC-0.9_HM 參考軟體。相較於對照組的壓縮結果，我們的技術平均能夠節省 2.2%的資料量，最小與最大的節省分別為 0.2%與 4.7%。

二、英文摘要

Keywords：Bi-prediction, OBMC, Template Matching
This report introduces a bi-prediction scheme with only a motion overhead as for unidirectional prediction. It combines motion vectors found by template and block matchings with the overlapped block motion compensation (OBMC). An optimal window function is designed based on a model-based framework. Additionally, the concept of adaptive motion merging is incorporated to enable a template-matching-free implementation. Three algorithms featuring different performance and complexity trade-offs are implemented using the TMuC-0.9_HM software and tested with the common test conditions. Relative to the anchor, the best of them achieves an average BD-rate saving of 2.2%, with a minimum of 0.2% and a maximum of 4.7%.

三、研究成果

1. INTRODUCTION

Using the data accessible to the decoder for motion inference has recently emerged as a promising technique for the next generation video coding standard. This notion has been applied to design a wide variety of methods. Typical examples are template matching prediction (TMP) [5] and adaptive motion merging [11]. The former estimates the motion vector (MV) for a target block on the decoder side by minimizing the matching error over the reconstructed pixels in its immediate inverse-L-shaped neighborhood (usually termed the *template*), while the latter adaptively reuses motion parameters for its neighboring blocks. It is interesting to note that the overlapped block motion compensation (OBMC) [8], first proposed one and half decades ago, also follows the same rationale. It views the received motion data as a source of information about the motion field and forms a better prediction of a pixel's intensity based on its own and nearby block MVs.

Motivated by these observations, we are led to develop a biprediction scheme, which incurs only the overhead for unidirectional prediction. The idea is to combine MVs resulting from the template and block matchings with OBMC. Of particular interest in this combination is that the template MV is inferred on the decoder side. Thus it has only to signal one block MV while attaining biprediction performance. The same notion is further generalized to allow a template-matching-free implementation, which replaces the template MV with a decoded MV specified by a mechanism similar to motion merging.

The choice of a proper window function is critical in these applications. We approach this problem through the parametric OBMC framework in [3]. The resulting window function is shown to resemble a particular type of geometry motion partitioning [6] with its MVs located on the diagonal running from the upper left to the lower right. Particularly, asymmetric-like partitioning [7] arises when the template region is of rectangular shape and is located to the left or on the top of a target block.

Based on these notions, three algorithms featuring different performance and complexity trade-offs are implemented using the TMuC software [4] and tested with the common test conditions [2]. When compared with the TMuC_0.9-hm anchor in common test conditions, the best of them is observed to have 0.2-4.7% BD-rate savings [1], with an average of 2.2%. The average decoding time increases by 10% while the encoding time doubles. Remarkably, this particular implementation can produce an effect similar to having one of the asymmetric/geometry motion partitions coded in merge mode, the concept of which is known as partial motion merging.

The rest of this report is organized as follows: Section 2 briefly reviews the POBMC framework in [3]. Sections 3 and 4 introduce the proposed technique and its low-complexity implementation, respectively. Section 5 provides experimental results for various test conditions. Finally, this report is concluded with a summary of our works.

2. PARAMETRIC OBMC

In [8], OBMC is introduced to provide a better prediction of a pixel's intensity value $I(s)$ based on motion-compensated signals $\{\bar{I}(\mathrm{s}+\mathrm{v}(\mathrm{s}_i))\}_{i=1}^{L}$ derived from its own and nearby block MVs $\{\mathrm{v}(\mathrm{s}_i)\}_{i=1}^{L}$. Essentially, these MVs are considered to be plausible hypotheses for its true motion; their weights $\mathrm{w}=(w_1, w_2, ..., w_L)$ are chosen to minimize the mean squared prediction error subject to the unit-gain constraint:

$$\mathrm{w}^* = \arg\min_{\mathrm{w}} \xi(\mathrm{w}) \ s.t. \sum_{i=1}^{L} \omega_i = 1, \tag{1}$$

where

$$\xi(\mathrm{w}) = \mathrm{E}\left\{ \left(I(s) - \sum_{i=1}^{L} \omega_i \bar{I}(\mathrm{s}+\mathrm{v}(\mathrm{s}_i)) \right)^2 \right\}. \tag{2}$$

In applying OBMC to variable block-size motion partition, our previous work [3] proposes a parametric window solution. Using the motion model in [12], the $\xi(\mathrm{w})$ at pixel s is shown, under some mild conditions, to have the following form:

$$\xi(\mathrm{w}) = \sum_{i=1}^{L} \varepsilon \omega_i^2 r^2(\mathrm{s};\mathrm{s}_i), \tag{3}$$

where ε indicates the randomness of the motion and intensity fields; $\{r(\mathrm{s};\mathrm{s}_i)\}_{i=1}^{L}$ are the L_2 distances between s and its nearby block centers $\{\mathrm{s}_i\}_{i=1}^{L}$ where $\{\mathrm{v}(\mathrm{s}_i)\}_{i=1}^{L}$ are sampled. Upon setting the gradient of with respect to w to 0, the optimal weights become

$$\omega_i^* = \frac{r^{-2}(\mathrm{s};\mathrm{s}_i)}{\sum_{i=1}^{L} r^{-2}(\mathrm{s};\mathrm{s}_i)}, 1 \le i \le L. \tag{4}$$

The significance of this result is that it requires only the geometry relations of pixel s and its nearby block centers $\{\mathrm{s}_i\}_{i=1}^{L}$ to obtain $\{\omega_i^*\}_{i=1}^{L}$. This remarkable property allows MVs associated with any motion partitions to be incorporated for OBMC.

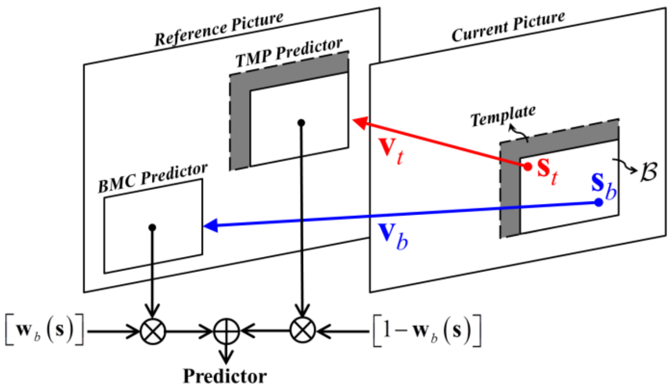


Fig. 1. Joint application of TMP and BMC.

3. COMBINING TEMPLATE AND BLOCK MOTION COMPENSATIONS

In this section, we present the proposed bi-prediction scheme and the design of its window function.

3.1. Concept of Operation

Fig. 1 depicts its concept of operation. Like the conventional biprediction, it predicts a target block based on two predictors. These predictors however are weighted in a pixel-adaptive manner using POBMC, with one of them derived from a MV v_t found by template matching [5] and the other from the usual motion compensation. Since v_t can be inferred on the decoder side, this scheme has only to signal motion parameters for or one block MV (denoted as v_b). Additionally, we restrict v_b to be unidirectional in order to minimize the overhead.

3.2. Optimized Block Matching Criterion

In this application, v_t cannot be specified discretionarily. It is thus important to find a v_b that, when applied jointly with B, minimizes the prediction error over the target block B. Such v_b can be described in symbols as follows:

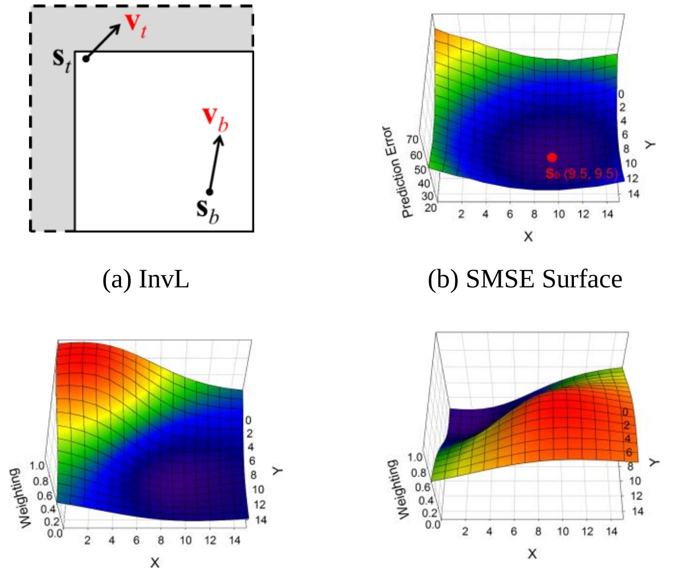
$$\mathrm{v}_b^* = \arg\min_{\mathrm{v}_b} \sum_{\mathrm{s}\in \mathrm{B}} \left(I(\mathrm{s}) - \omega_t(\mathrm{s})\tilde{I}(\mathrm{s}+\mathrm{v}_t) - \omega_b(\mathrm{s})\tilde{I}(\mathrm{s}+\mathrm{v}_b) \right)^2 , \tag{5}$$

where $\omega_b(\mathrm{s})$ denotes its window function, $\omega_t(\mathrm{s}) = 1 - \omega_b(\mathrm{s})$, and $\tilde{I}$ is the reference picture. Clearly, $\omega_b(\mathrm{s})$ has a decisive effect on the value of v_b and thus on the prediction performance of this scheme.

Instead of using an iterative procedure to determine $\omega_b(\mathrm{s})$ (and thus $\omega_t(\mathrm{s})$), as is the case with the OBMC [8], we approach this problem by resorting to the parametric framework in Section 2. To proceed, v_t is approximated as the pixel true motion $\mathrm{v}(\mathrm{s}_t)$ at the template centroid s_t , a justification of which can be found in [10]. However, we avoid making the same approximation for v_b because its search criterion is no longer to minimize the sum of squared prediction errors[1] (cf. (5)). It is replaced instead by the true motion $\mathrm{v}(\mathrm{s}_b)$ of some unknown pixel s_b in B. We now cast the problem of determining $\omega_b(\mathrm{s})$ as the search for an s_b in B that minimizes the sum of mean squared prediction errors (SMSE) over B:

$$\sum_{\mathrm{s}\in \mathrm{B}} \left(I(\mathrm{s}) - \omega_t(\mathrm{s})\tilde{I}(\mathrm{s}+\mathrm{v}_t) - \omega_b(\mathrm{s})\tilde{I}(\mathrm{s}+\mathrm{v}_b) \right)^2 \tag{6}$$

[1]A block MV approximates the pixel true motion at the block center only if its search criterion is to minimize the sum of squared prediction errors [12][9].



(c) 1 - $\omega_b^*(s)$ (d) $\omega_b^*(s)$

Fig. 2. SMSE surface as a function of S_b's location, and the optimal window functions associated with v_t and v_b, respectively.

Each term in the summation of (6) is simply the mean squared prediction error, produced by OBMC based on $v(s_t)$ and $v(s_b)$, at some pixel s and can be modeled with (3). For a given s_b, (6) is minimized when each operand reaches its minimum, that is, according to (4), when $\omega_b(s) = \omega_b^*(s) = r^2(s;s_t) / (r^2(s;s_t) + r^2(s;s_b))$. Noting this, we have:

$$s_b^* = \arg\min_{s_b \in B} \sum_{s \in B} \varepsilon\left((\omega_t^*(s))^2 r^2(s;s_t) + (\omega_b^*(s))^2 r^2(s;s_b)\right). \tag{7}$$

Due to its non-linear nature, s_b^* has to be found by numerical simulations, i.e., we have to compute SMSE for every admissible location of s_b. Fortunately, the computation is tedious but not difficult, and can be made offline. Once it has been solved, the $\omega_b^*(s)$ is immediate by (4). Applying this window function to (5), we get an optimized block matching criterion, with which a v_b^* approximating the pixel true motion at s_b^* can be found.

To get a sense of where s_b^* should be located, Fig. 2 (b) plots the SMSE as a function of its location. As can be observed, the SMSE becomes smaller when s_b sits in the bottom right quarter; a further precise calculation shows that its optimal location occurs at point (9.5, 9.5) for a 16×16 target block. This is of no surprise because v_t, located at the template centroid (1.9, 1.9), has a higher correlation with the motion field in the upper left quarter and thus contributes more to minimizing the errors there. Intuitively, s_b should be so placed that the errors in the remaining part of B can be minimized.

3.3. Window Functions

Fig. 2 (c) and (d) plot the window functions, $1-\omega_b^*(\mathrm{s})$ and $\omega_b^*(\mathrm{s})$, for v_t and v_b, respectively. As can be seen, their waveforms suggest a special type of geometry motion partition [6] with two MVs located on the diagonal running from the upper left to the lower right. Following the same line of derivation, we can obtain window functions for other template designs. Some results are given in Fig. 3. In particular, asymmetric-like motion partitions [6][7] result when the template region locates directly to the left or above a target block (See Fig. 3 (a) and (b)). Two conceptual differences however are to be noted. First, unlike explicit geometry or asymmetric partitions, these implicit "soft" partitions incur less motion overhead (only one MV is to be signaled). Second, there is a strong interdependency between the transmitted and inferred MVs due to OBMC (cf. (5)).

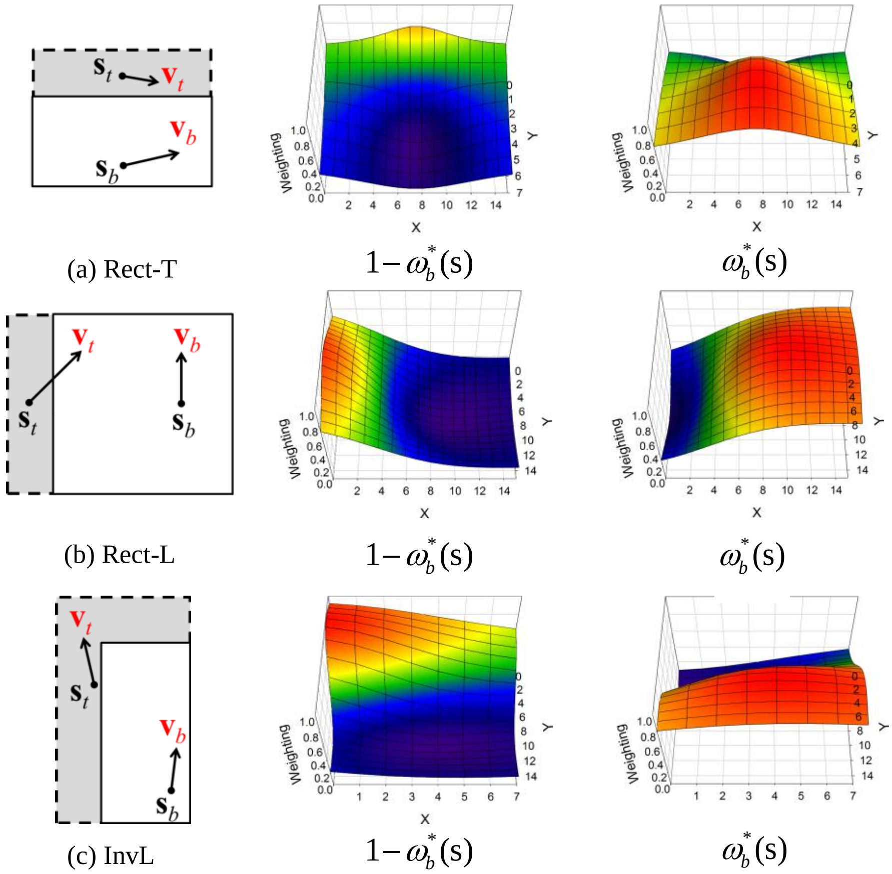


Fig. 3. Window functions for typical template designs.

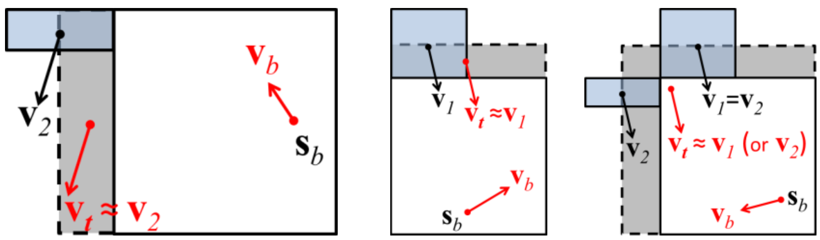


Fig. 4. Adaptive motion merging and the approximation of v_t.

4. EXTENSIONWITH MOTION MERGING

Recognizing that performing template matching not only increases the decoding complexity but also complicates the pipeline design of the decoder, we additionally propose a low-complexity and TMPfree implementation. This is accomplished by replacing the template MV v_t by one of those decoded MVs from neighboring partitions (cf. Fig. 4). In this way, the need to perform TMP is waived at the cost of extra bits. To minimize the overhead, we adopt the same signaling mechanism as for motion merging [11]. When enabled, it sends one flag to indicate the choice of the first available neighboring partition, to the left or above the current one, of which its motion parameters are reused as v_t. Depending on the inference direction, a separate window function is applied for OBMC. For example, in Fig. 4, if v_t is deduced from v_1 (respectively, v_2), the window created for the Rect-T (respectively, Rect-L) template is selected. In other cases where v1 and v_2 are identical, the one for the InvL template is used instead. Essentially, we treat v_1 and v_2 as if they were the pixel true motion associated with the corresponding template centroids. Because these assumptions may not always hold true, we let the encoder to switch adaptively between this proposed mode and the usual inter mode.

5. EXPERIMENTS

Based on the proposed scheme (referred hereafter to as TB-mode), we develop three algorithms featuring different performance and complexity trade-offs. Extensive experiments are carried out using the TMuC-0.9_HM software [4] and the common test conditions [2] to compare their BD-rate savings relative to anchor encodings with TB-mode disabled. Rough estimates of their complexity characteristics are made by showing the encoding and decoding time ratios relative to anchor settings. All three algorithms use parametric window functions. Weighting coefficients are rounded offline into 16-bit integers and stored in tables for reducing computational complexity. The following summarizes these algorithms:

- **Algo. #1** applies TB-mode to 2N×2N prediction units (PUs)[2] only. The v_t is found by performing shape-adaptive template matching in a search range of MVp±8 pixels. For each 2N×2N PU, one flag is sent to switch adaptively between TBmode and the usual inter mode. When the former is chosen, it codes two extra bits (at most) to specify the template shape (InvL, Rect-T or Rect-L). A separate set of window functions are designed for each distinct 2N×2N PU size.
- **Algo. #2** simplifies Algo. #1 by signaling v_t with motion merging mechanism (See Section 4).
- **Algo. #3** extends Algo. #2 to all PU sizes (2N×2N, 2N×N, N×2N, N×N). However, the flag indicating TB-mode is sent at coding unit (CU) level—i.e., all PUs in a CUmust be coded either in TB-mode or in the usual inter mode. When TB-mode is in use, each of its PU needs one more flag to signal how v is derived and thus which window function to apply.

5.1. Compression Performance and Complexity

Table 1 presents the average BD-rate savings of these algorithms in different test classes and configurations. As can be seen, Algo. #1 has an average BD-rate (Y) saving of 1.3% over all test cases, with a minimum of 0.2% and a maximum of 3.4%. Due to the extra computations needed for template matching, it doubles the decoding time while increasing the encoding time by about 70%.

Algo. #2 gives better coding performance at a much faster decoding speed than Algo. #1. Averagely, it performs 0.1% better in terms of BD-rate saving (Avg:1.4%; Min: 0.0%; Max: 3.3%) with an average decoding time only 4% slightly longer than anchors'. In tests with High Efficiency configurations, its decoder can run even a bit faster. As expected, without having to perform template matching, its decoding complexity drops significantly. But, somehow surprisingly, the v_t signaled by motion merging seems to outperform that inferred by template matching. The reason may be twofold. First, the motion merging signaling mechanism incurs less overhead: while it needs only one extra bit to signal v_t, Algo. #1 requires, on average, more than one bit to indicate the template shape. Second, template matching may result in poor v_t due to coding noise. Recall that its search criterion is to minimize the error over the reconstructed pixels. Nevertheless, we believe Algo. #1 has plenty of room for further improvement.

Algo. #3 delivers the best compression performance, with an average BD-rate saving of 2.2% and a maximum saving up to 4.7%. This is attributed to the PU-adaptive window selection. But also because of this feature, the encoder has to perform, for each PU, one extra motion search to evaluate TB-mode, which accounts for the doubled encoding time. By contrast, the increase in decoding time relative to Algo. #2 is minor.

Table 1. BD-rate savings and processing time ratios

Random Access	**High Efficiency**			**Low Complexity**		
Algo.	#1	#2	#3	#1	#2	#3
Class A	-1.3	-1.3	-1.9	-1.1	-1.5	-2.4
Class B	-1.0	-1.1	-1.6	-0.8	-1.1	-1.6
Class C	-1.5	-1.5	-2.3	-1.2	-1.5	-2.4
Class D	-1.4	-1.6	-2.5	-0.9	-1.3	-2.2
All	-1.3	-1.4	-2.1	-1.0	-1.3	-2.1
Enc. Time [%]	171	126	208	202	142	263
Dec. Time [%]	179	95	99	230	114	123
Low Delay	**High Efficiency**			**Low Complexity**		
Algo.	#1	#2	#3	#1	#2	#3
Class B	-1.5	-1.5	-2.1	-1.0	-0.8	-1.4
Class C	-1.9	-1.8	-2.7	-1.0	-1.0	-1.8
Class D	-1.5	-1.8	-2.8	-0.9	-1.0	-1.6
Class E	-2.9	-3.1	-4.3	-1.7	-1.7	-2.6
All	-1.9	-1.9	-2.8	-1.1	-1.1	-1.8
Enc. Time [%]	145	116	172	163	127	203
Dec. Time [%]	182	97	100	223	111	121

Table 2. Comparison of table sizes.

Algo.	**PU Shape**	**Template Shape**	**Table Size**
#1	2N×2N	InvL	5.6 Kbytes
#2	2N×2N	InvL, Rect-T/-L	11.0 Kbytes
#3	All PU shapes	InvL, Rect-T/-L	19.2 Kbytes

[2]CU, the basic compression unit as the MB in AVC, can have various sizes but is restricted to be a square shape. PUs are the various MB partitions having a square or rectangular shape with several sizes.

5.2. Mode Distribution

Fig. 5 charts the mode distributions of different schemes, including the anchor. The percentage of a prediction mode is segmented into several bars according to the CU size; the length of each bar represents its average spatial coverage (in units of pixels) in CUs of some specific size. From the figure, two observations can be made. First, when Skip and Merge modes are excluded, TB-mode tends to dominate over other inter modes, especially in Algo. #2 and #3. This confirms its effectiveness in providing similar or better prediction efficiency with less overhead. Second, the percentage of Merge mode drops moderately when TB-mode is enabled, which implies that a considerable number of PUs can benefit from motion refinement. Besides, we also observe that TB-mode increases a bit the use of larger CUs.

6. CONCLUSION

Summarizing, in this report, we propose a bi-prediction scheme, which combines MVs found by template and block matchings with an optimized OBMC window function. Since the template MV is inferred on the decoder side, it has only to signal one block MV. This notion is further generalized by incorporating adaptive motion merging to allow a template-matching-free implementation. Three algorithms featuring different performance and complexity tradeoffs are presented; they all show moderate coding gains. The best of them produces an effect similar to performing partial motion merging for geometry/asymmetric motion partitions.

四、結果與討論、計畫成果自評

本計畫有以下幾類成果。第一類為提出一個如同單向預測僅需一個動作向量傳輸負擔的雙向預測機制。透過參數化交疊區塊動作補償結合樣板核對的動作向量，且重新設計了區塊動作搜索的機制，而其中所使用的最佳視窗函數是基於模型的框架所設計。本實驗室亦有發表提案參與 JCT-VC 的 CfP 競賽，在 27 個參賽隊伍中，取得了第 12 名的成績，其中競賽單位包括 Samsung、Qualcomm 等知名大廠。第二類為在計畫執行期間參與了 JCT-VC 標準會議，總計有 9 件標準會議的貢獻文件。第三類為計畫執行過程中所獲得之研究成果論文，已發表於(含接受)國際學術會議及國際學術期刊，共有 2 篇期刊論文與 3 篇研討會論文。第四類成果為一篇美國專利案正在申請中。此外，參與計畫之研究人員亦獲得了許多視訊壓縮相關之最新技術之設計經驗，達到人才培育之目的。

綜合評估：本計畫發表了許多具有學術價值的成果，其中在國際標準會議上也有貢獻，並藉此機會培育出許多視訊壓縮領域之人才，不論是對學術相關研究或是與國內業界合作進行相關產業之新產品開發，都有正面的價值。

五、參考文獻

[1] G. Bjontegaard, “Improvements of the BD-PSNR Model,” *VCEGAI11*,Jul. 2008.

[2] F. Bossen, “Common Test Conditions and Software Reference Configurations,” *JCTVC-C500*, Oct. 2010.

[3] Y. W. Chen and et al., “A Parametric Window Design for Overlapped Block Motion Compensation with Variable Block-size Motion Estimates,” *IEEE MMSP*, 2009.

[4] JCT-VC, “TMuC-0.9_HM,” https://hevc.hhi.fraunhofer.de/svn/svn_TMuCSoftware/branches/0.9-hm/.

[5] S. Kamp and et al., “Decoder Side Motion Vector Derivation for Inter Frame Video Coding,” *IEEE*

ICIP, 2008.

[6] M. Karczewicz and et al., “Video Coding Technology Proposal by Qualcomm Inc.” *JCTVC-A121*, Apr. 2010.

[7] K. McCann and et al., “Samsung’s Response to the Call for Proposals on Video Compression Technology,” *JCTVC-A124*, Apr.2010.

[8] M. T. Orchard and G. J. Sullivan, “Overlapped Block Motioin Compensation: An Estimation-Theoretic Approach,” *IEEE Trans. on Image Processing*, vol. 3, no. 5, pp. 693–699, May 1994.

[9] B. Tao and M. Orchard, “A Parametric Solution for Optimal Overlapped Block Motion Compensation,” *IEEE Trans. on Image Processing*, vol. 10, no. 3, pp. 341–350, Mar. 2001.

[10] T.-W. Wang and et al., “Analysis of Template Matching Prediction and its Application to Parametric Overlapped Block Motion Compensation,” *IEEE ISCAS*, 2010.

[11] M. Winken and et al., “Description of Video Coding Technology Proposal by Fraunhofer HHI,” *JCTVC-A116*, Apr. 2010.

[12] W. Zheng and et al., “Analysis of Space-dependent Characteristics of Motion-compensated Frame Differences based on a Statistical Motion Distribution Model,” *IEEE Trans. on Image Processing*, vol. 11, no. 4, pp. 377–386, Mar. 2002.

行政院國家科學委員會補助國內專家學者出席國際學術會議報告

100 年 10 月 19 日

報告人姓名	陳俊吉	服務機構及職稱	國立交通大學資訊工程學系 博士生
會議 時間 地點	99 年 12 月 7 日 至 99 年 12 月 10 日 日本 名古屋	本會核定補助文號	
會議名稱	(中文) 2011 影像編碼研討會 (英文) 2011 Picture Coding Symposium		
發表論文題目	(中文) 位元平面之壓縮取樣搭配貝氏估計解碼法於失真壓縮之應用 (英文) Bit-Plane Compressive Sensing with Bayesian Decoding for Lossy Compression		

報告內容應包括下列各項:

一、參加會議經過

2010年影像編碼研討會(Picture Coding Symposium)研討會於12月7日至12月10日在日本名古屋舉行,本屆PCS' 2010最後僅接受來至世界各地論文發表共157篇,依論文特色與領域共分成22個技術主題,在25個平行場次進行。同時本次研討會在議程的設計上,除了一般性的論文發表外,大會亦特別安排了兩場Tutorial Sessions與三場Keynote Speech,分別介紹三個極具潛力及前瞻性的研究領域,參與這三位大師的演說,不僅瞻仰大師的丰采,亦提升了筆者於學術領域的廣度,著實獲益良多。

此外,於四天的議程安排中,大會並聘請兩位專家學者,針對影像處理相關領域的特定主題安排兩場訓練課程(Tutorials)。主題分別為:(1) Evolutive Video Coding - From Generic Algorithm towards Content-Specific Algorithm、(2) Quality Assessment for Image Compression Purpose。PCS' 2011 接受之論文內容涵蓋數位影像處理、新一代視訊壓縮技術 (High Efficiency Video Coding)、電腦視覺與圖學、與 3D 電視與自由視角電視等相關領域,不但針對相關技術的介紹,亦加深理論於實務設計上的考量。其餘三天為論文口頭報告(Lecture Session)與壁報發表(Poster Session),大會總共包含 22 個技術主題,共 25 場次分 5 個口頭報告場次及 20 個壁報發表場次同時進行,與會者可依興趣與專長自由參與感興趣的場次聽講及發問。其中 10 個主題分別為:(1) 3DTV、(2) Free-viewpoint Television、(3) Beyond H.264/MPEG-4 AVC、(4) Coding of still and moving pictures、(5) Model-based and synthetic coding、(6) Distributed source coding、(7) Image and video processing、(8) Multimodal coding and processing、(9) Very high-resolution imaging, coding and processing、(10) Multi-view video processing and coding、(11) Representation, analysis and coding of 3D scenes、(12) Virtual/augmented/mixed reality、

(13) Subjective and objective quality assessment metrics and methods、(14) Joint source and channel coding、(15) Error robustness, resilience and concealment、(16) Transcoding and transmoding、(17) Coding for mobile, IP and sensor networks、(18) Coding and processing for database applications、(19) Protection and integrity of visual data、(20) Implementation architectures and VLSI、(21) New applications and techniques for visual data processing、(22) Standards for visual data coding。

筆者於與會期間，多次遇見來自中港臺新加坡等地的華人學者，可感覺到華人於國際學術舞台參與度與交流之不斷提升，也對來自台灣多位學者於此會議之貢獻感到驕傲。與會期間亦認識來自歐美等地的學者，不但領受學術之交流，亦感覺到主辦單位籌備之用心，讓每位參與者皆有賓至如歸之感覺。

二、與會心得

影像編碼研討會為一歷史悠久(已超過 40 年)，且其規模亦不斷成長的國際著名研討會，最近幾年輪流於世界各大洲等地舉辦，是實至名歸且具影響與號召力的的國際研討會，對於從事視訊壓縮、影像處理、3D 視訊處理等相關研究的學者而言，每年的參與應可獲得相當程度的幫助，同時亦於第一時間接觸到最新且最熱門研究主題(例如：高效能視訊壓縮、3D 視訊壓縮)。整體而言，此會議所包含的研究領域相當廣泛、論文內容豐富與質精，是一非常值得參加的會議。筆者此次參加此會議，不僅皆獲得不少與本人研究相關的知識，更獲得同領域學者於學術研究上的建議與新知。很幸運可於與會同時認識了多位國內外相關領域之學者，對於知識的擴展與友誼的增進方面均覺受益良多，實已達到了學術交流之目的，在此特別感謝國科會所給予之補助。

三、考察參觀活動(無是項活動者省略)

無。

四、建議

基於參與此次 PCS’2010 會議的經驗，筆者之建議如下：第一天的訓練課程通常包含新且最熱門研究主題，經常是與會學者非常有興趣參加的，因主講者針對演講之主題所整理的資料，通常不但具有一定程度的啟發性與前瞻性，並且有助於個人研究，亦可能提供為未來之教學教材內容，若能將參與訓練課程之費用也列入補助項目，將是一非常正面且實質的鼓勵。

五、攜回資料名稱及內容

本次會議所有的論文皆收錄在一片光碟內，每一位與會學者皆可擁有一片。筆者所攜回之資料為：(1) PCS‘2010 CD 論文集一片、(2) PCS‘2010 紙本論文集一本、(3) WPCIP‘2010 紙本論文集一本、(4) 近期相關領域之論文徵稿資料。

六、其他

無。

國科會補助計畫衍生研發成果推廣資料表

日期:2011/03/04

<table>
<tr><td rowspan="3">國科會補助計畫</td><td colspan="3">計畫名稱: 新興高效能視訊編碼技術</td></tr>
<tr><td colspan="3">計畫主持人: 彭文孝</td></tr>
<tr><td colspan="3">計畫編號: 99-2221-E-009-017-　　　　學門領域: 訊號處理</td></tr>
<tr><td rowspan="2">研發成果名稱</td><td colspan="3">(中文) 畫面估測系統及其估測方法</td></tr>
<tr><td colspan="3">(英文) Frame Prediction System and Prediction Method Thereof</td></tr>
<tr><td>成果歸屬機構</td><td>國立交通大學</td><td>發明人(創作人)</td><td>彭文孝, 陳滂紋</td></tr>
<tr><td rowspan="2">技術說明</td><td colspan="3">(中文) 本發明係揭露一種畫面估測系統及其估測方法，其包含初始模組初始化一具有複數個像素之第一畫面區塊。提供模組係提供一第二畫面區塊之第一重心點及第一運動向量。位置查找模組根據第一重心點查找一位置點，並根據各該像素對第一重心點及位置點之關係，各別對應產生第一權重值和第二權重值。向量查找模組根據第一重心點、第一運動向量、位置點、第一權重值及第二權重值，查找該複數個像素於第一畫面區塊中具有最低像素強度誤差值之第二運動向量。處理模組根據該些運動向量及該些權重值，依序運算各該像素之一預測強度值。</td></tr>
<tr><td colspan="3">(英文) The present invention discloses a frame prediction system and a prediction method thereof. An initializing module initializes a first image block having a plurality of pixels. A providing module provides a first centroid and a first motion vector of a second image block. The location lookup module finds a location according to the first centroid, and generates a first weight and a second weight respectively according to a relationship between each of the pixels, the first centroid and the location. A vector lookup module finds a second motion vector, which gives a minimum pixel intensity error for the plurality of pixels in the first image block according to the first centroid, the first motion vector, the location, the first weight and the second weight. A processing module sequentially calculates a plurality of predictive intensity values according to the motion vectors and the weights.</td></tr>
<tr><td>產業別</td><td colspan="3">研究發展服務業</td></tr>
<tr><td>技術/產品應用範圍</td><td colspan="3">視訊影像編解碼器</td></tr>
<tr><td>技術移轉可行性及預期效益</td><td colspan="3">此一發明所推廣之技術已由ISO/IEC MPEG與ITU-T VCEG之新一代高效能視訊編碼國際標準所採用</td></tr>
</table>

註：本項研發成果若尚未申請專利，請勿揭露可申請專利之主要內容。

99 年度專題研究計畫研究成果彙整表

<table>
<tr><td colspan="3">計畫主持人：彭文孝</td><td colspan="5">計畫編號：99-2221-E-009-017-</td></tr>
<tr><td colspan="8">計畫名稱：新興高效能視訊編碼技術</td></tr>
<tr><td colspan="3" rowspan="2">成果項目</td><td colspan="3">量化</td><td rowspan="2">單位</td><td rowspan="2">備註（質化說明：如數個計畫共同成果、成果列為該期刊之封面故事...等）</td></tr>
<tr><td>實際已達成數（被接受或已發表）</td><td>預期總達成數(含實際已達成數)</td><td>本計畫實際貢獻百分比</td></tr>
<tr><td rowspan="12">國內</td><td rowspan="4">論文著作</td><td>期刊論文</td><td>0</td><td>0</td><td>100%</td><td rowspan="3">篇</td><td></td></tr>
<tr><td>研究報告/技術報告</td><td>0</td><td>0</td><td>100%</td><td></td></tr>
<tr><td>研討會論文</td><td>0</td><td>0</td><td>100%</td><td></td></tr>
<tr><td>專書</td><td>0</td><td>0</td><td>100%</td><td></td><td></td></tr>
<tr><td rowspan="2">專利</td><td>申請中件數</td><td>0</td><td>0</td><td>100%</td><td rowspan="2">件</td><td></td></tr>
<tr><td>已獲得件數</td><td>0</td><td>0</td><td>100%</td><td></td></tr>
<tr><td rowspan="2">技術移轉</td><td>件數</td><td>0</td><td>0</td><td>100%</td><td>件</td><td></td></tr>
<tr><td>權利金</td><td>0</td><td>0</td><td>100%</td><td>千元</td><td></td></tr>
<tr><td rowspan="4">參與計畫人力（本國籍）</td><td>碩士生</td><td>0</td><td>0</td><td>100%</td><td rowspan="4">人次</td><td></td></tr>
<tr><td>博士生</td><td>0</td><td>0</td><td>100%</td><td></td></tr>
<tr><td>博士後研究員</td><td>0</td><td>0</td><td>100%</td><td></td></tr>
<tr><td>專任助理</td><td>0</td><td>0</td><td>100%</td><td></td></tr>
<tr><td rowspan="12">國外</td><td rowspan="4">論文著作</td><td>期刊論文</td><td>2</td><td>2</td><td>100%</td><td rowspan="3">篇</td><td></td></tr>
<tr><td>研究報告/技術報告</td><td>9</td><td>9</td><td>100%</td><td></td></tr>
<tr><td>研討會論文</td><td>3</td><td>3</td><td>100%</td><td></td></tr>
<tr><td>專書</td><td>0</td><td>0</td><td>100%</td><td>章/本</td><td></td></tr>
<tr><td rowspan="2">專利</td><td>申請中件數</td><td>1</td><td>1</td><td>100%</td><td rowspan="2">件</td><td></td></tr>
<tr><td>已獲得件數</td><td>0</td><td>0</td><td>100%</td><td></td></tr>
<tr><td rowspan="2">技術移轉</td><td>件數</td><td>0</td><td>0</td><td>100%</td><td>件</td><td></td></tr>
<tr><td>權利金</td><td>0</td><td>0</td><td>100%</td><td>千元</td><td></td></tr>
<tr><td rowspan="4">參與計畫人力（外國籍）</td><td>碩士生</td><td>7</td><td>7</td><td>100%</td><td rowspan="4">人次</td><td></td></tr>
<tr><td>博士生</td><td>3</td><td>3</td><td>100%</td><td></td></tr>
<tr><td>博士後研究員</td><td>0</td><td>0</td><td>100%</td><td></td></tr>
<tr><td>專任助理</td><td>0</td><td>0</td><td>100%</td><td></td></tr>
</table>

其他成果 (無法以量化表達之成果如辦理學術活動、獲得獎項、重要國際合作、研究成果國際影響力及其他協助產業技術發展之具體效益事項等，請以文字敘述填列。)	1. 提出一個如同單向預測僅需一個動作向量傳輸負擔的雙向預測機制。透過參數化交疊區塊動作補償結合樣板核對的動作向量，且重新設計了區塊動作搜索的機制，而其中所使用的最佳視窗函數是基於模型的框架所設計。 2. 本實驗室亦有發表提案參與 JCT-VC 的 CfP 競賽，在 27 個參賽隊伍中，取得了第 12 名的成績，其中競賽單位包括 Samsung、Qualcomm 等知名大廠。

	成果項目	量化	名稱或內容性質簡述
科教處計畫加填項目	測驗工具(含質性與量性)	0	
	課程/模組	0	
	電腦及網路系統或工具	0	
	教材	0	
	舉辦之活動/競賽	0	
	研討會/工作坊	0	
	電子報、網站	0	
	計畫成果推廣之參與（閱聽）人數	0	

國科會補助專題研究計畫成果報告自評表

請就研究內容與原計畫相符程度、達成預期目標情況、研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）、是否適合在學術期刊發表或申請專利、主要發現或其他有關價值等，作一綜合評估。

1. 請就研究內容與原計畫相符程度、達成預期目標情況作一綜合評估

■達成目標

□未達成目標（請說明，以 100 字為限）

□實驗失敗

□因故實驗中斷

□其他原因

說明：

2. 研究成果在學術期刊發表或申請專利等情形：

論文：■已發表 □未發表之文稿 □撰寫中 □無

專利：□已獲得 ■申請中 □無

技轉：□已技轉 □洽談中 ■無

其他：（以 100 字為限）

1. JCT-VC 標準會議，總計有 9 件標準會議的貢獻文件
2. 2 篇國際期刊論文與 3 篇國際研討會論文
3. 一篇美國專利案正在申請中

3. 請依學術成就、技術創新、社會影響等方面，評估研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）（以 500 字為限）

本計畫有以下幾類成果。第一類為提出一個如同單向預測僅需一個動作向量傳輸負擔的雙向預測機制。透過參數化交疊區塊動作補償結合樣板核對的動作向量，且重新設計了區塊動作搜索的機制，而其中所使用的最佳視窗函數是基於模型的框架所設計。本實驗室亦有發表提案參與 JCT-VC 的 CfP 競賽，在 27 個參賽隊伍中，取得了第 12 名的成績，其中競賽單位包括 Samsung、Qualcomm 等知名大廠。第二類為在計畫執行期間參與了 JCT-VC 標準會議，總計有 9 件標準會議的貢獻文件。第三類為計畫執行過程中所獲得之研究成果論文，已發表於(含接受)國際學術會議及國際學術期刊，共有 2 篇期刊論文與 3 篇研討會論文。第四類成果為一篇美國專利案正在申請中。此外，參與計畫之研究人員亦獲得了許多視訊壓縮相關之最新技術之設計經驗，達到人才培育之目的。

綜合評估：本計畫發表了許多具有學術價值的成果，其中在國際標準會議上也有貢獻，並藉此機會培育出許多視訊壓縮領域之人才，不論是對學術相關研究或是與國內業界合作進行相關產業之新產品開發，都有正面的價值。