行政院國家科學委員會專題研究計畫 成果報告

架構在階層式雙向預測影片上的多重描述編碼法之設計
研究成果報告(精簡版)

計 畫 類 別 ：個別型
計 畫 編 號 ：NSC 99-2221-E-009-139-
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 ：蔡文錦

計畫參與人員：碩士班研究生-兼任助理人員：黃致遠
碩士班研究生-兼任助理人員：楊巧安
碩士班研究生-兼任助理人員：蕭成憲
碩士班研究生-兼任助理人員：林宗翰

公 開 資 訊 ：本計畫可公開查詢

中 華 民 國 100 年 11 月 22 日

中文摘要： 多重描述編碼法(Multiple Descriptor Coding，簡稱 MDC)就是一種將影片切割成多個低頻寬子影片的視訊編碼法，由於各個子影音檔在傳輸時可走不同路徑，封包遺失可以藉由其他子影片檔的內容來加以修復，大幅減低封包遺失對影片品質的影響，因此非常適合用於網路上的影音串流的應用中。

多重描述編碼法在實際的應用中未被廣為使用的原因，主要在於編碼效能差。如果將影片檔的資料任意切割成多份分別壓縮，則個別子影片的壓縮效率必然很差，因為資料之間的相互關係(co-relation)被切割分散在不同子影片中，無法充分利用，造成整體編碼的資料量(bitrate)遠高於 H.264 標準編碼(single description coding) 產生的資料量。有鑑於此，我們希望提出一種多重描述編碼法，來有效減少多餘的資料量(redundant bit-rate)。Hierarchical B-frame prediction structure (簡稱 hierarchical B)的架構是一種有高效能的運動預測方式，其藉著階層式的雙向預測來達到比傳統 B frames 有更好的編碼效能，因此，本計畫的多重描述編碼法擬植基於 hierarchical B 的架構上來提升編碼效能，並加以改善其容錯能力。

此研究計畫中，我們完成以下兩個目標: 1. 研究架構於 hierarchicalB 上的多重描述編碼法； 2. 研究架構於 hierarchical B 上的錯誤隱藏方法。目標一主要是採用混和的方式來結合多種切割法於 hierarchical B 的架構上，來互補各切割法的優缺點，相較於目前多數的多重描述編碼方法架構在 non-hierarchical B 上且只使用一種特性的切割，我們的方法確實有較佳的編碼效能與容錯能力。目標二主要是著重在 hierarchical B 架構上錯誤隱藏方法(error concealment)之設計。由於 hierarchical B 是階層式的架構，不同 frame 在不同階層上有不同的重要性及參考方式，我們考慮這些特性來設計適合的錯誤隱藏方法，而非直接使用傳統的方式。在兼顧視訊資料內容的特性及網路狀況，我們的錯誤隱藏方法的確能達到相當不錯的容錯效果。

中文關鍵詞： 階層式雙向預測視訊編碼, 多重描述編碼, 不對等保護, 空間切割, 時間切割

英文摘要： Multiple description coding (MDC) is a technique of striping a video sequence into two or more descriptions in such a way that each description is independently decodable. There have been a number of proposals for MDC coding, each providing their own tradeoff between coding efficiency and error

resilience. However, the solutions proposed so far still suffer from poor coding efficiency. Therefore, in this project, we provide a MDC solution to meet the requirements of video streaming applications. Hierarchical B-frame prediction structure (called hierarchical B) is a technique in which B frames are arranged in a hierarchical way with two-way prediction so that coding efficiency can be increased. In this project, we would like to build our MDC method on it.
There are two goals that we have achieved in this project: 1. Design a coding efficient MDC methods based on hierarchical B architecture, 2. Design error concealment methods based on hierarchical B architectures. In the first goal, we focus on the design of MDC methods which improve coding efficiency by using hierarchical B and hybrid segmentation methods, where the hybrid segmentation means to combine several segmentation techniques (e.g., spatial, temporal, or frequency segmentations) on a single video stream. In the second goal, we focus on designing error concealment methods which combines different techniques and take hierarchical B structure, network conditions, and video content characteristics into considerations to improve the error resilience of the proposed MDC methods.

英文關鍵詞： Multiple description coding (MDC), hierarchical B pictures, spatial splitting, temporal splitting, duplication.

目錄：

報告內容

1. 前言

Multiple description video coding (MDC) is one of the approaches for reducing the detrimental effects caused by transmission over error-prone networks. In this project, a MDC model based on hierarchical B pictures is proposed to optimize the tradeoff between coding efficiency and error resilience. The model produces two descriptors by applying different MDC techniques such as duplication, spatial splitting and temporal splitting on the different frames of video sequences, taking into account unequal importance of frames at different hierarchical levels. Duplication (high redundancy) is for key frames; spatial splitting (medium redundancy) for reference B frames; and temporal splitting (low redundancy) for non-reference B frames. As a consequence, better error resilience can be achieved at high coding efficiency.

2. 研究目的

During data transmission, packets may be dropped or damaged, due to channel errors, congestion, and buffer limitation. Moreover, the data may arrive too late to be used in real-time applications. In the case of transmission of compressed video, this loss may result in a completely damaged stream at the decoder side. For real-time applications, since retransmission is often not acceptable, error resilience (ER) and error concealment (EC) techniques are required for displaying a pleasant video signal despite the errors and for reducing distortion introduced by error propagation.

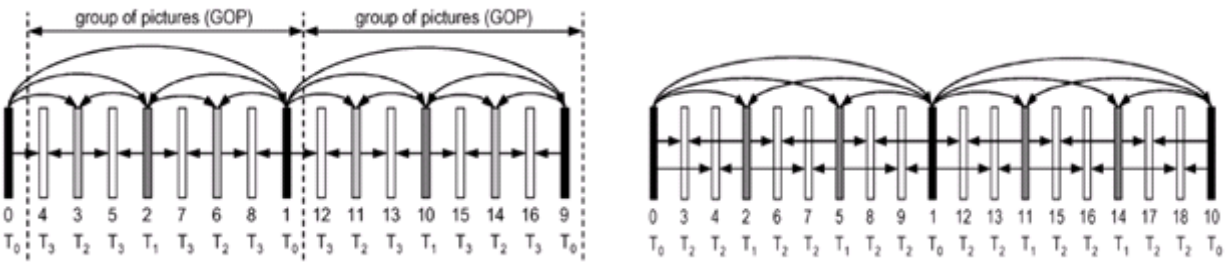
Several ER methods have been developed, such as *forward error correction* (FEC), *intra/inter coding mode selection*, *layered coding*, and *multiple description coding* (MDC). This project is concerned with MDC. Multiple description coding is a technique that encodes a single video stream into two or more equally important sub-streams, called *descriptions*, each of which can be decoded independently. Different from the traditional single description coding (SDC) where the entire video stream is sent in one channel, in MDC, these multiple descriptions are sent to the destination through different channels, resulting in much less probability of losing the entire video stream (all the descriptions), where the packet losses of all the channels are assumed to be independently and identically distributed. Due to effectiveness in providing error resilience, a variety of researches on different MDC approaches had been proposed. These approaches can be intuitively classified through the stage where it split the signal, such as, frequency domain [1], spatial domain [2], and temporal domain [3]. In our previous works [4], a hybrid MDC method has been proposed, which applies MDC first in spatial domain to split motion compensated residual data, and then in frequency domain to split quantized coefficients. The results in [4] show that, by properly utilizing more than one splitting technique, the hybrid MDC can improve error-resilient performance.

Although a variety of MDC approaches have been proposed, most of them were built upon

conventional H.264/AVC coding structure and did not utilize hierarchical B-picture prediction. In a hierarchical B-picture prediction framework, the B frames at the coarser temporal levels can to be used as reference for the B frames at the finer temporal levels and therefore, the coding efficiency can be further improved. Compared with classical H.264/AVC, the improvement can be more than 1dB as described in [5]. In [6], an MDC based on hierarchical B pictures was proposed, where two descriptions are generated by duplicating the original sequence and then coded by hierarchical B structure with staggered key frames in the two descriptions. By using different QPs at different levels, their approach enables each frame to have two different quality fidelities in different descriptions. When two descriptors are received, their approach simply selects the frame with high-fidelity, or uses a linear combination of the high-fidelity and low-fidelity frames to generate a better reconstruction. When only one descriptor is received, the lost frame is recovered by copying from the corresponding frame in the other descriptor. It can be seen that although their MDC approach employs hierarchical B-pictures to improve coding efficiency, it still suffers from high bit-rate redundancy by duplicating the original sequence to two descriptions. This paper presents a MDC based on hierarchical B pictures with unequal error protection considered. Our approach employed duplication, spatial splitting, and temporal splitting for the frames at different hierarchical levels to provide unequal redundancy to frames with different fidelity requirements.

3 文獻探討

A typical hierarchical prediction framework with 4 dyadic hierarchy stages is illustrated in Fig. 1(a), where the key frames (I or P frames) are coded in regular intervals. A key frame and all frames that are temporally located between the key frame and the previous key frame form a group of picture (GOP). The remaining B frames are hierarchically predicted using two reference frames from the nearest neighboring frames of the previous temporal level. In Fig.1, B^i denotes the B frames at level *i*. It should be noted that the usage of hierarchical coding structure is not restricted to be the dyadic case. Fig.1(b) shows the example of a non-dyadic hierarchical structure with 3 levels. For the optimized encoding, it is better to set smaller QPs for the frames that are referenced by other frames. In the Joint Scalable Video Model 11 (JSVM11) [7], QPs of the B frames at level-1 equal to the QPs of the I/P frames plus 4, and the QPs at level-*i* increase by 1 from level-($i-1$), with $i \geqq 2$.



(a) Dyadic (b) Non-dyadic

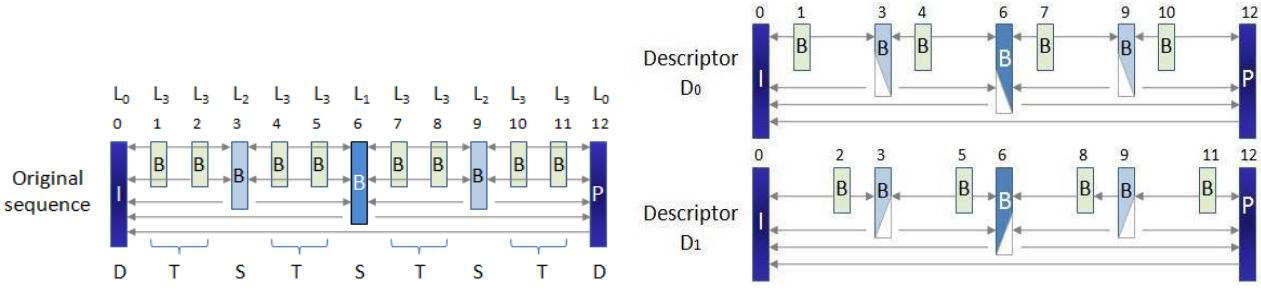
Fig. 1 Hierarchical B-picture prediction structure

4 研究方法

4.1 Hierarchical B-picture Coding

In hierarchical B-picture prediction framework, the frames at lower hierarchical levels can be used as reference for the frames at higher hierarchical levels. Due to this dependency, the decoding quality of a frame strongly depends on the quality of the frames at its previous hierarchical level of the same GOP. The lower level at which a frame is lost, the more frames that will be corrupted. As an example in Fig.1(a), the loss of an I or P picture will directly affect 7 other frames, while the loss of a level-1, level-2, and level-3 B-pictures will directly affect 4, 2, and 0 other frames, respectively. Based on this observation, the proposed MDC aims at providing unequal redundancy for the hierarchical B pictures, taking into account the unequal importance of the frames at different hierarchical levels.

The proposed MDC model is illustrated in Fig.2, where a non-dyadic hierarchical B-picture structure with 4 levels is used. We refer to the I/P frames at the lowest hierarchical level as *key frames*; the B frames at intermediate levels as *reference B frames* (*RB frames*) because they are used as reference; and the B frames at the highest level as *non-reference B frames* (*NRB frames*) because they are not used as reference. As the Fig.2(a) shows, we apply *duplication* (denoted by D) on key frames for providing the highest error resilience; *spatial-splitting* (S) on RB frames for modest error resilience; and *temporal-splitting* (T) on NRB frames for the lowest error resilience. The resulting two descriptions are illustrated in Fig.2(b), where the rectangles with a missing corner represent incomplete frames (due to spatial splitting). It can be seen that, due to different MDC methods applied, the frames at different hierarchical levels have unequal redundancy to provide robustness again errors. Assuming that description D0 is lost; the lost key-frames (0 and 12) can be easily reconstructed at decoder by using the same frames in description D1; the partially lost level-1 and level-2 frames (3, 6 and 9) can be estimated by using the information of their counterparts in description D1; while the lost level-3 frames (1, 4, 7 and 10) which are not in D1, can only be estimated by using other frames. The estimation methods will be discussed later in next section.



(a) Original sequence (b) The resulting two descriptors

Fig. 2 Proposed MDC based on hierarchical B-picture prediction.

4.2 Estimation of Loss Description

Taking advantages of different MDC methods applied on the frames at different hierarchical

levels, different estimation methods are designed for different frames. Table II summarizes the cases for different estimation methods to be applied, where S denotes the spatial method, T the temporal method, and D the duplication method. The columns describe the two loss cases; while the rows describe three types of frames.

Table I. Summary of the cases for different estimation methodsbv

Estimation methods		Descriptor status	
		One-descriptor loss	Two-descriptor loss
Frame type	Key frame	D	T
	RB frame	S	T
	NRB frame	T	T

4.2.1 One Descriptor Loss

In case of one-descriptor loss, since the lost key-frames can be reconstructed by simply using the duplicated version in the other descriptor, it is marked as **D** in Table I. As for RB frames, since they are split in the spatial domain, one-descriptor loss only causes partial-frame loss. In this case, spatial method (marked as **S** in Table I) is applied to estimate the lost part. After the received descriptor has been entropy decoded, de-quantized, and inversely transformed, the Spatial Merger will apply polyphase inverse permutation on the resulting data and then the residual pixels will be distributed like a checkerboard inside the macroblock as shown in Fig. 3, where each lost residual pixel has four available neighboring pixels. Our spatial estimation uses *bilinear interpolation* to reconstruct the lost residual pixels, as shown in Equation (1) where $f'_{j,i}$ is the reconstructed value of the residual pixel in column i and row j. Since neighboring pixels have high spatial correlation, spatial estimation should be efficient.

$$f'_{j,i} = (f_{j+1,\,i} + f_{j-1,\,i} + f_{j,\,i+1} + f_{j,\,i-1}) / 4 \qquad (1)$$

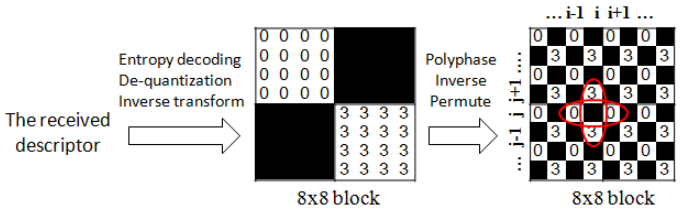


Fig. 3 Spatial concealment by bilinear interpolation.

As for NRB frames, one-description loss will result in whole frame loss because they are split in the temporal domain. In this case, a temporal estimation method (marked as **T** in Table I) is applied to reconstruct the lost frame. Since the temporal method is also adopted for all types of frames in case of two-description loss, we describe it in the next subsection.

4.2.2 Two Descriptor Loss

In case of two-description loss, it will result in whole-frame loss regardless of frame types. For

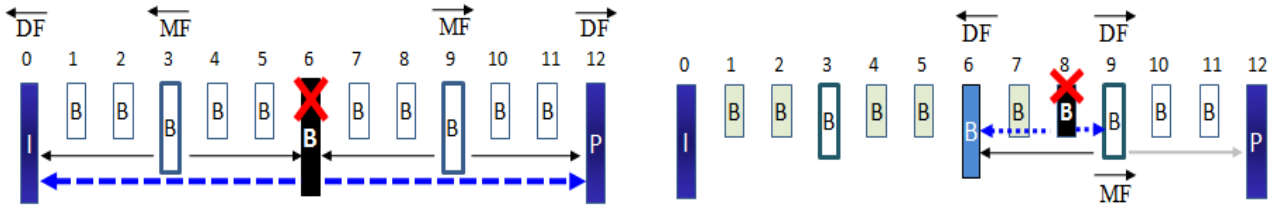
whole-frame loss, each block in the lost frame is recovered based on temporal correlation since all the neighboring blocks are also lost. We refer to the pictures whose pixels are used to predict the missing pixels as the *data prediction frame* (DF) and the pictures whose block motions are used to predict the motion of the missing blocks as the *motion prediction frame* (MF). In our method, DF can be different from MF. Besides, the proposed methods adopt bi-directional motion-compensated signal to recover missing pixels. Thus, we need to select two DFs: a *backward DF* and a *forward DF* (denoted by $\overleftarrow{\mathrm{DF}}$ and $\overrightarrow{\mathrm{DF}}$, respectively); and two MFs: a *backward MF* and a *forward MF* (denoted by $\overleftarrow{\mathrm{MF}}$ and $\overrightarrow{\mathrm{MF}}$, respectively) for a lost picture. Since the data correlation among pictures involved tends to considerably weaken as the temporal distances among these pictures become longer, for a lost picture, it is better to choose the nearest pictures in display order to serve as its DFs. However, to serve as DFs requires that these pictures are decoded earlier than the lost picture. Based on the hierarchical B-picture structure, for a lost picture, we select its reference frames in backward and forward directions as its $\overleftarrow{\mathrm{DF}}$ and $\overrightarrow{\mathrm{DF}}$, respectively.

As for MFs, they are selected differently from DFs. In case of frame loss, even though the frames later than the lost frame (in decoding order) cannot be decoded before the lost frame is recovered, the motion information of these frames is obtainable. Therefore, the MFs need not to be located earlier than the lost picture in decoder order. Instead of using temporal direct mode (TDM) technique which adopts reference pictures as MFs, we choose pictures at higher levels because these pictures are temporally nearer to the lost picture in display order. As an example in Fig.4(a), if the frame 6 is lost, we will select its reference frames (0 and 12) as its DFs, but frames 3 and 9 as its MFs. In Fig. 4(a), if frame 3 is lost, we will select its reference frames 0 and 6 as DFs, but frames 2 and 4 as MFs. This selection policy is applied to all frames except NRB frames which are at the highest level within the hierarchical structure. For NRB frames, the MFs are selected from their reference frames at the previous level of the lost picture. Fig.4(b) illustrates the case of NRB frame loss, where frame 8 is the lost frame. In this case, frames 6 and 9 will serve as the DFs, and frame 9 (which is at previous level of frame 8) will serve as the MF. Similarly, if frame 10 is lost, its DFs will be frames 9 and 12, and its MF will be frame 9. Specifically, for the lost picture F_t^l at time instant t with hierarchical level l, we select its $\overleftarrow{\mathrm{MF}}$ and $\overrightarrow{\mathrm{MF}}$ as

$$\overleftarrow{\mathrm{MF}} = \begin{cases} F_{t_{nb}}^{l+1} & for\ l_{base} \le l < l_{top} \\ F_{t_{ref}}^{l-1} & if\ F_{t_{ref}}^{l-1}\ exists\ for\ l = l_{top} \end{cases} \qquad (2)$$

$$\overrightarrow{\mathrm{MF}} = \begin{cases} F_{t_{nf}}^{l+1} & for\ l_{base} \le l < l_{top} \\ F_{t_{ref}}^{l-1} & if\ F_{t_{ref}}^{l-1}\ exists\ for\ l = l_{top} \end{cases} \qquad (3)$$

where l_{base} denotes the base level (key-frame level) and l_{top} the top level (NRB-frame level). $F_{t_{\mathrm{nb}}}^{l+1}$ and $F_{t_{\mathrm{nf}}}^{l+1}$ denote F_t^l's nearest backward and forward frames at level *l+1*, respectively. $F_{t_{ref}}^{l-1}$ denotes the F_t^l's reference frame at level *l-1*.



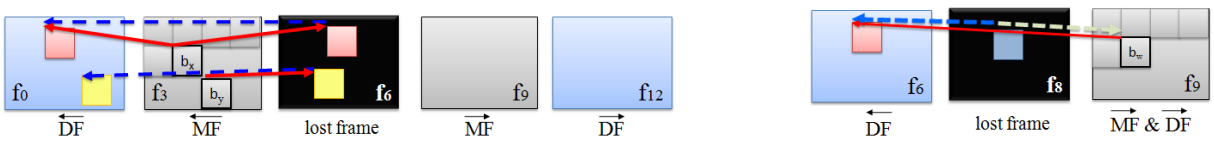
(a) DF and MF selection for RB frames (b) DF and MF selection for NRB frames

Fig.4 DF and MF selection for temporal estimation method.

After determining DFs and MFs, the motion vectors in MFs will be used to estimate the missing motion vectors (pointing to DFs from the lost frame). When the lost frame is a RB frame, since its MFs are located in between DFs and the lost frame (see Fig. 4(a)), the motion vectors are composed if the block in MFs has two motion vectors, or extrapolated if the block has only one motion vector. The motion vector derivation corresponding to Fig.4(a) is illustrated in Fig.5(a), where the two motion vectors of b_x in f_3 are composed and the motion vector of b_y is extrapolated, so that the derived motion vectors will point to f_0 from f_6. The motion vectors pointing to f_{12} from f_6 can also be derived in a similar manner using motion vectors of f_9. On the other hand, when the lost frame is a NRB frame, since one MF is used for two DFs located on different sides of the lost picture (see Fig.4(b)), the motion vectors in the MF are interpolated as illustrated in Fig.5(b) where the motion vector of block b_w is interpolated to obtain two motion vectors respectively pointing to f_6 and f_9 from f_8. Let $\overleftarrow{mv}$ and $\overrightarrow{mv}$ denote the derived motion vectors pointing to $\overleftarrow{\mathrm{DF}}$ and $\overrightarrow{\mathrm{DF}}$ from the lost frame, respectively. For a lost frame, after all the motion vectors in its MFs have been composed, extrapolated, or interpolated, the missing pixels on the lost frame can be classified into four types: the pixels associated with one or more $\overleftarrow{mv}$, the pixels with one or more $\overrightarrow{mv}$, the pixels with both $\overleftarrow{mv}$ and $\overrightarrow{mv}$, and the pixels without $\overleftarrow{mv}$ and $\overrightarrow{mv}$. For a pixel P in the lost picture, we recover it by the predicted signal $\tilde{P}$ obtained as follows

$$\tilde{P}(x) = \begin{cases} \sum_i \overleftarrow{\mathrm{DF}}(x + \overleftarrow{mv}_i) & ;\ if\ P \text{ has } \overleftarrow{mv} \text{ only} \\ \sum_i \overrightarrow{\mathrm{DF}}(x + \overrightarrow{mv}_i) & ;\ if\ P \text{ has } \overrightarrow{mv} \text{ only} \\ w_0 \sum_i \overleftarrow{\mathrm{DF}}(x + \overleftarrow{mv}_i) + w_1 \sum_i \overrightarrow{\mathrm{DF}}(x + \overrightarrow{mv}_i) & ;\ if\ P \text{ has } \overleftarrow{mv} \text{ and } \overrightarrow{mv} \\ w_0 \overleftarrow{\mathrm{DF}}(x)) + w_1 \overrightarrow{\mathrm{DF}}(x)) & ;\ otherwise \end{cases} \quad (4)$$

Here, x is spatial coordinate of P. w_0 and w_1 are the weighting values, which are set in inverse proportion to the temporal distances of $\overleftarrow{\mathrm{DF}}$ and $\overrightarrow{\mathrm{DF}}$, respectively, from the lost picture.



(a) Motion composition and extrapolation for a RB frame (b) Motion interpolation for a NRB frame

Fig. 5 Temporal estimation using bi-directional predicted signal.

5 結果與討論

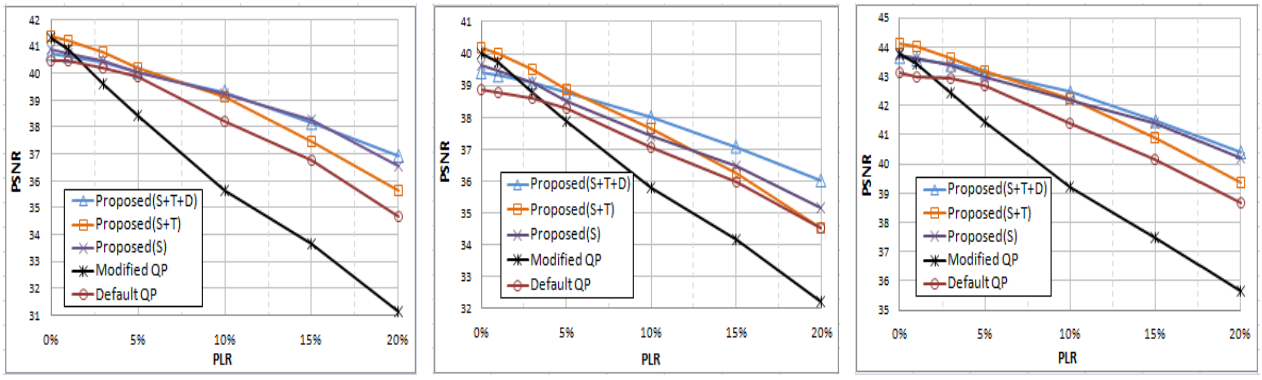
In this section, the proposed MDC model is examined. To see the effects of different MDC techniques adopted in our model, experiments were conducted for three variations of proposed MDC model: *proposed (S)*, *proposed (S+T)*, and *proposed (S+T+D)*. The proposed (S) stands for the method which adopts spatial splitting only. It applies spatial splitting in the residual domain for all frames, regardless of hierarchical levels. The proposed (S+T) stands for the method which adopts two kinds of splitting: temporal splitting for top-level frames (*i.e.*, NRB frames) and spatial splitting for others. The proposed (S+T+D) stands for the full version of proposed method, which adopts temporal splitting for top-level NRB frames, spatial splitting for RB frames, and duplication for base-level key frames. We compare our three methods with Zhu *et al*'s method [6] which generates two descriptors by duplicating the original sequence and then coded by hierarchical B structure with staggered key frames in the two descriptions. This approach is characterized by that each frame at level 0, 1, or 2 of description 1 will be at level 3 of description 2 and vice versa, resulting in two fidelities of each frame in two descriptions. Two variations, *default_QP* and *modified_QP*, in their literature are adopted in our comparison. The default_QP follows the QP assignment rules specified in JSVM11[7] as described in Section II, while the modified_QP modifies the QPs of top-level frames to 51 in order to reduce bitrate redundancy. The results in [6] show that rate-distortion performance of center decoder can be improved remarkably by modified_QP in comparison to default_QP. In this section, their packet-loss performances are also examined. Table II lists the error concealment methods used by these MDC methods, where D**'** means the error concealment method in [6], where in case of one-descriptor loss, the lost frame is recovered by the duplicated version in the other description. D**'** is distinguished from D because the duplicated frame is at the same level in our approach, but at a different level in Zhu *et al*'s approach. Since Zhu *et al*. did not provide solutions for two-description loss, our temporal estimation method is adopted for fair comparison. The five MDC methods are implemented based on H.264 reference software, JM 16.0[8].

All the methods encode video sequences using hierarchical B-picture structure of four levels to generate two descriptors. The three proposed methods adopt a non-dyadic structure which allows temporal splitting on NRB frames as depicted in Fig.2; while Zhu *et al*.'s two methods adopt a dyadic structure which ensures that each frame has two different fidelities in the two descriptions.

5.1 Packet Loss Performance

The five MDC methods were examined in a packet-loss scenario where various packet-loss rates, ranging from 0% to 20%, are adopted. We use one packet for each frame of each descriptor. Fig.6 shows the results for four CIF sequences, *Foreman*, Coastguard, *Mobile*, and *News*. The R_{2D} in

Fig.6 denotes the total bit-rate of two descriptions. For each sequence, two kinds of R_{2D} are used. In each case of Fig.6, the five methods encode the sequence using the same R_{2D} for fair comparison. The results are the averages of 100 independent runs. It can be seen that, in the case of PLR=0%,



(a) Foreman (CIF, R_{2D} = 1500kbps) (b) Coastguard (CIF, R_{2D} = 2800kbps.) (c) News (CIF, R_{2D} = 700kbps)

Fig. 6 Performance comparison in packet-loss environments.

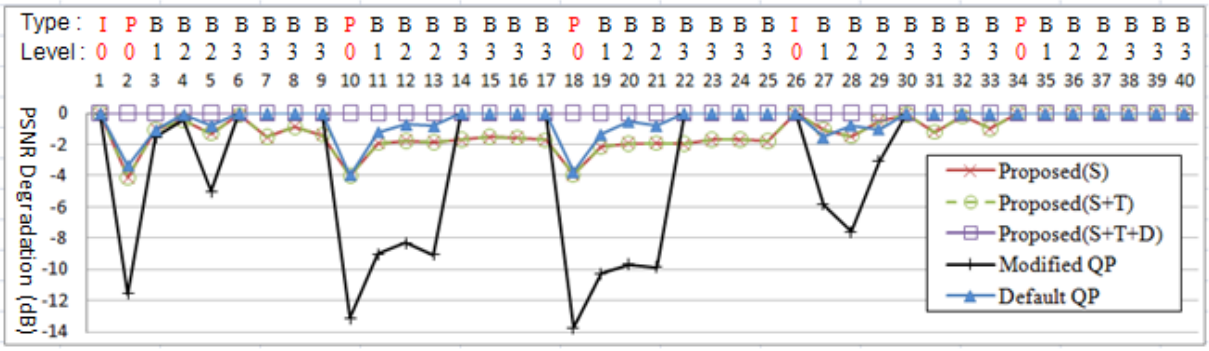
modified_QP and proposed (S+T) have the best performance and default_QP has the worst performance among all methods. This is due to that the default_QP duplicates the entire sequence to two descriptions and therefore, suffers from considerable bit-rate redundancy. By providing poorer picture quality at the lowest level, the modified_QP can effectively reduce the bitrate and thus achieve a better performance at PLR=0. As PLR increases, however, the modified_QP curves drop much more quickly than others for all sequences, showing that the poorer quality at the lowest level will strongly affect error-concealment effectiveness and thus, degrade the performance. Compared with modified_QP, default_QP performed much better as PLR increases. However, the duplication method used in default_QP still cannot avoid quality degradation in recovering lost frames because the same frames in two descriptions are at different levels with different fidelities. The degraded error-concealment performance and the high bitrate-redundancy result in the worse performance of default_QP, compared with the three proposed methods.

Among these methods, proposed (S+T+D) has the overall best performance. Although proposed (S) performed slightly better than proposed (S+T+D) for *foreman*, it performed much worse than proposed (S+T+D) for *mobile* and *coastguard*. This is due to that spatial estimation cannot recover lost data well for these sequences when there is packet loss. With temporal splitting on NRB frames, proposed (S+T) reduces bit-rate redundancy and hence, improves the R-D performance at low PLR, but still cannot solve the problem for high PLR. By duplicating key frames, the proposed (S+T+D) can alleviate this problem effectively. When packet loss rate is low (PLR<5%), the proposed (S+T+D) performed equally to, or slightly worse than, proposed (S+T) and proposed (S). This stems from the fact that the scheme of key-frame duplication adopted in the proposed (S+T+D) cannot take much effect in error concealment when PLR is low. As the PLR increases, however, proposed (S+T+D) outperformed others noticeably. This is due to that the key frames in proposed

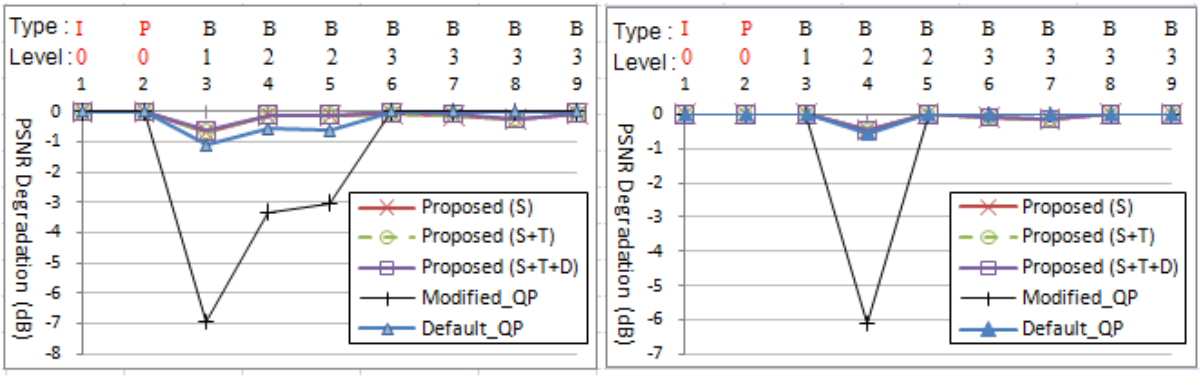
(S+T+D) can be recovered without quality loss once they are lost. Since key- frames have the maximum number of frames depending on it, the duplication of key-frame can suppress error propagation effectively and improve performance substantially. We will discuss the error propagation issue further in later section. To summarize, the overall results demonstrate that, by adopting spatial splitting, temporal splitting and duplication for the frames at different levels, the proposed (S+T+D) optimizes the trade-off between bit-rate redundancy and error-resilient capability and therefore, achieves the best performance among the five MDC methods.

5.2 Error Propagation Effects

This section presents the frame-by-frame comparison of error propagation effects using different MDC methods. The effects of error propagation were examined for a single frame loss occurring at different hierarchical levels of *Mobile* sequence at QP=28. We use one packet for each frame in each descriptor and the error propagation results of a single packet loss at different hierarchical levels are shown in Fig.7, where we renumber the selected frames according to decoding order. Figs.7(a-c) show the results of the frame loss occurring at levels 0, 1 and 2, respectively. In Fig.7, y-axis denotes PSNR degradation and x-axis the frame number (in decoding order). From Fig.7(a) it is observed that almost all the methods suffer from severe error propagation for the P-frame loss, except the proposed (S+T+D). This is due to that the proposed (S+T+D) duplicates key-frames to



(a) Frame loss at level 0 (the 2nd frame is lost)



(b) Frame loss at level 1 (the 3rd frame is lost) **(c)** Frame loss at level 2 (the 4th frame is lost)

Fig.7 Frame-by-frame comparison.

two descriptors and thus, when only one of them is loss, the other can be used to reconstruct frame without quality degradation and error propagation. In both proposed (S) and proposed (S+T) methods, key-frames are spatially split to two descriptors and hence, the P-frame loss in one descriptor will cause partial-frame loss which is recovered by using spatial estimation, suffering from quality degradation and error propagation. As for default_QP, although it duplicates the entire sequence to two descriptors, the same frames in the two descriptors are at different levels and thus, the lost key frame can only be recovered by the corresponding low quality frame in the other descriptor. This also results in quality degradation and error propagation. It is worth to mention that even though the quality degradation of default_QP in Fig.7(a) is smoother than those of proposed (S) and proposed (S+T), it is at the cost of bit-rate redundancy. That is why default_QP has worse performance than proposed (S) and (S+T) as shown in Fig.6. Compared with default_QP, modified_QP suffers from much severe quality degradation because the top-level frame used to recover the lost key frame has been set to QP=51 to reduce the bit-rate. Compared with Fig.7(a), the results in Figs.7(b) and (c) show that when the fame loss occurs at level 1 or 2, the error propagation effects are substantially reduced for all the methods and the performance gaps between different methods are also decreased. To summarize, the results in Fig.7 show that quality degradation and error propagation in the hierarchical prediction structure are affected by key frames most, and level-1 and level-2 frames the second. By taking into account the importance of frames at different levels, proposed (S+T+D) optimizes the trade-off between coding efficiency and error resilience and achieves the overall best performance.

參考文獻

[1] O. Campana, R. Contiero, “An H.264/AVC Video Coder Based on Multiple Description Scalar Quantizer,” IEEE Asilomar Conference on Signals, Systems and Computers (ACSSC), 2006.

[2] R. Bemardini, M. Durigon, R. Rinaldo, L. Celetto, and A. Vitali, “Polyphase Spatial Subsampling Multiple Description Coding of Video Streams with H.264,” Proceedings of IEEE Intel. Conf. on Image Processing (ICIP), Oct. 2004.

[3] J. G. Apostolopoulos, “Error-Resilient Video Compression Through the Use of Multiple States,” Proceedings of IEEE Intel. Conf. on Image Processing (ICIP), Vol. 3, 2000.

[4] C. W. Hsiao and W. J. Tsai, “Hybrid Multiple Description Coding Based on H.264,” IEEE Trans. on Circuits and Syst. for Video Technol., Vol. 20, No.1, Jan. 2010.

[5] H. Schwarz, D. Marpe , T. Wiegand, “Analysis of hierarchical B pictures and MTCF,” IEEE International Conference on Multimedia & Expo, ICME ’06, pp. 1929-1932

[6] C. Zhu and M. Liu, “Multiple description video coding based on hierarchical B pictures,” IEEE Trans. Circuits and Systems for Video Technology, vol.19, No.4, April 2009.

[7] J. Reichel, H. Schwarz, and M. Wien, Joint Scalable Video Model 11 (JSVM 11), Joint Video Team, Doc. JVT-X202, Jul. 2007.
[8] H.264/AVC Ref. Software – JM, http://iphome.hhi.de/ suehring/tml/.
[9] W. J. Tsai and Hao-Yu You, "Multiple description video coding based on hierarchical B pictures using unequal redundancy," To appear in IEEE Trans. on Circuits and Systems for Video Technology.

計畫成果自評

- 本計畫的目標是設計一架構於 Hierarchical B 上的多重描述演算法，並結合不對等保護的觀念使整體 rate-distortion 效果能達最好。我們已依計畫目標完成方法的設計及相關的實驗，而研究成果也已被國際期刊 *IEEE Trans. on Circuits and Systems for Video Technology* 所接受[9]。
- 完成 Hierarchical B 架構上多重描述切割法(MDC)的技術之研究
  Hierarchical B 架構並不同於一般的編碼架構，其 Hierarchical 的特色會影響或限制各種多重描述切割法的整體表現。本計畫植基於此架構上，依據不同階層的特色與不同切割法的特性，設計出最佳的混合式 MDC，主要是結合空間域、時間域和複製法的切割法。其中，「複製法」因能提供最完善的 error resilient，我們將此方法套用在最需要保護的 key frame level，來減少傳輸錯誤時所帶來嚴重的 error propagation。「空間域切割」法適用於一般的 reference frame，所以我們將此套用於中間層的 RB-frame，藉其優越的錯誤補償機制，發揮空間域切割法的最佳效益。「時間域切割」法是套用在最高層的 frame，因為在傳輸錯誤發生時，位於最高層 frames 可從其他層的 frames 取得眾多 motion vectors 來做為錯誤補償的估算，更提升了補償之後的視覺品質。實驗結果顯示，我們所提出的混和式 MDC 將不同的切割法對應到 hierarchical B 架構下不同的層級中，都能發揮他們的最佳效益，因此，和其他的方法比起來都有較好的效果。
- 完成不同 MDC 切割法之 redundant bit-rate 關係的實驗與分析
  Redundant bit-rate 是用來比較各種 MDC 切割法優劣的衡量標準之一，在實驗結果中，我們發現用於一般編碼方式的 MDC 切割法，套用在 Hierarchical B 架構上後，redundant 的比例並不能如同先前未套用在 Hierarchical B 架構上所能達到的比例，甚至有很大的落差。因此，若只使用單一種切割法(如:時域切割法、空間域切割法、複製法)，在 Hierarchical B 架構上是非常不適用，而混合式的多重描述切割法則能改善此問題，降低 redundant 的比例，達到一般 MDC 所要求的標準。
- 完成 hierarchical B 架構上，空間域錯誤隱藏方法之研究
  在經過不同方法的實驗之後，我們選用 Bilinear Interpolation 的方法來補償空間域(Residual)錯誤的部分，Bilinear Interpolation 方法是取鄰近的資訊做雙向線性預測，因空間域上鄰近的 Pixel 有些許的相關性，以棋盤狀的方式來做雙向線性預測的錯誤隱藏方式，充分利用了鄰近相關的特性，而預測出較準確的 pixel value。從實驗結果中發現，此方法不需要過度複雜的運算，運算時間少又能提供不錯的品質。
- 完成 hierarchical B 架構上，時間域錯誤隱藏方法之研究

我們根據 hierarchical B 的架構，研究出不同於傳統的時間域錯誤隱藏方法，此錯誤隱藏方法可分成兩部分:1.提供 RB-frame 的重建，2. 提供 NRB-frame 的重建，依此兩類 frame 的特性，所參照的 motion vector 種類有所不同；RB-frame 在錯誤隱藏中所需要的 motion vector 是使用還未 decoded frame 的 motion vector 資訊，而 NRB-frame 則是使用最鄰近 frame 的 motion vector 當作參照。

使用了上述所提到的 motion vector，我們在補償錯誤的方法中，以 pixel 為單位做雙向 motion vector interpolation and extrapolation。實驗結果顯示,在發生 whole frame loss 的情況下，我們的方法能提供最佳的畫面品質。

- 完成各種錯誤隱藏方法在不同視訊內容特性與傳輸通道的實驗與分析

  在實驗結果中，我們所提出的空間域錯誤隱藏方法對於不同類型的視訊內容，並沒有很明顯的差別，residual 的錯誤補償的優劣主要是取決於視訊壓縮的 QP 大小，而不會因為動態、靜態或複雜的視訊內容有明顯的變動。

  時間域的錯誤隱藏方法則對於靜態的視訊內容能有較好的效益，相反的動態的視訊內容因有太多大幅度的 motion vector，或是 scene change 的情況過多，都會直接影響到時間域錯誤隱藏的效益，所以動態的視訊內容較不適用時間域的錯誤隱藏方法。

  為了模擬不同的傳輸通道的情況，我們設計了 0%~20%的 random loss rate 來實驗，在網路傳輸條件較為可靠時，整體效益的表現以時間域的錯誤隱藏方法為佳，但若是像無線傳輸這類型會發生較高 loss rate，則以空間域的效益較高。

- 完成 hierarchical B 架構上，發生傳輸錯誤的實驗與分析

  在 hierarchical B 的架構中，若是發生了傳輸錯誤，所造成的 error propagation 遠比一般傳統架構中來的嚴重許多，在實驗結果中發現，不同層的錯誤發生都會有著不同大小的影響程度，進而導致每層的擁有不同的重要性，皆需不同的保護機制，也就是 UEP(unequal error protection)，而這也是設計最佳切割演算法中考量的重要一環。

國科會補助計畫衍生研發成果推廣資料表

日期:2011/08/31

國科會補助計畫	計畫名稱: 架構在階層式雙向預測影片上的多重描述編碼法之設計		
	計畫主持人: 蔡文錦		
	計畫編號: 99-2221-E-009-139- 學門領域: 影像處理		
研發成果名稱	(中文) 架構於Hierarchical B上的多重描述視訊編碼法		
	(英文) Multiple description coding method based on Hierarchical B-picture structure for video streaming applications		
成果歸屬機構	國立交通大學	發明人(創作人)	蔡文錦,游灝瑜
技術說明	(中文) 本技術的內容主要包含一個架構在階層式視訊編碼(hierarchic B coding) 上的混合式多重描述編碼法 (hybrid MDC),其主要是結合複製法、空間域、和時間域的切割法來產生兩個描述檔(descriptor)。其中,「複製法」是用在最需要保護的key frame level,把完整的key frame (I or P frame) 複製一份讓兩個描述檔各自擁有一份以提供最完善的error resilience。「空間域切割」法是用於hierarchic B 中間層的可作為參考用的 B-frames (稱 RB-frame),主要是把 frame 上的點重排後分成兩部份,分別給兩個描述檔,當有一個描述檔遺失時,可藉由空間域錯誤補償機制,提供中度的error resilience。「時間域切割」法是套用在hierarchic B 最高層的不作為參考用的 B-frames (稱 NRB-frame),主要是交錯地將不同NRB-frame 分給不同描述檔, 因此,當有一個描述檔遺失時,必須用不同張的 NRB-frame 藉由時間域錯誤補償機制來加以重建資料。		
	(英文) Based on Hierarchical B-picture prediction structure, our multiple description coding (MDC) generates two descriptors by applying duplication for key frames (i.e., I or P-frame) which require the highest error resilience; apply spatial-splitting for Reference-B frames (RB-frame) which require modest error resilience; and apply temporal-splitting (T) for non-reference B- frames (NRB-frame) which require the lowest error resilience. With the proposed MDC, different estimation methods are designed for different frames. In case of one-descriptor loss in key frame level, the lost key frames can be found in the other descriptor; in case of one-descriptor loss in RB-frame level, the loss RB-frame can be estimated by using spatial estimation method; in case of one-descriptor loss in NRB-frame, the loss NRB-frame can be estimated by using temporal estimation method.		
產業別	資訊服務業		
技術/產品應用範圍	視訊串流應用如網路電視		
技術移轉可行性及預期效益	可轉移至視訊串流應用如網路電視的產業, 提供其know-how.		

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

計畫主持人：蔡文錦	計畫編號：99-2221-E-009-139-
計畫名稱：架構在階層式雙向預測影片上的多重描述編碼法之設計	

成果項目			量化			單位	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）
			實際已達成數（被接受或已發表）	預期總達成數(含實際已達成數)	本計畫實際貢獻百分比		
國內	論文著作	期刊論文	0	0	100%	篇	
		研究報告/技術報告	0	0	100%		
		研討會論文	0	0	100%		
		專書	0	0	100%		
	專利	申請中件數	0	0	100%	件	
		已獲得件數	0	0	100%		
	技術移轉	件數	0	0	100%	件	
		權利金	0	0	100%	千元	
	參與計畫人力（本國籍）	碩士生	4	4	100%	人次	
		博士生	0	0	100%		
		博士後研究員	0	0	100%		
		專任助理	0	0	100%		
國外	論文著作	期刊論文	1	1	100%	篇	已為期刊：IEEE Transactions on Circuits and Systems for Video Technology 所接受（尚未出版）
		研究報告/技術報告	0	0	100%		
		研討會論文	0	0	100%		
		專書	0	0	100%	章/本	
	專利	申請中件數	0	0	100%	件	
		已獲得件數	0	0	100%		
	技術移轉	件數	0	0	100%	件	
		權利金	0	0	100%	千元	
	參與計畫人力（外國籍）	碩士生	0	0	100%	人次	
		博士生	0	0	100%		
		博士後研究員	0	0	100%		
		專任助理	0	0	100%		

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。)	本計畫的研究成果已為頂級國際期刊 IEEE Transactions on Circuits and Systems for Video Technology 所接受, 因此具國際影響力, 而所設計的方法可實際用於視訊串流的應用中, 改善錯誤容錯能力, 因此具有產業的應用性.

	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)

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)

本研究成果可應用在實際的視訊串流應用中,因此有其應用價值,而所提的方法使用的技術極為創新,已為國際期刊 IEEE Transactions on Circuits and Systems for Video Technology 所接受(尚未刊登), 因此極具學術價值.