Multi-Model Local Illumination Compensation for Enhanced Compression Model

Jialin Li, *Student Member, IEEE*, Zhuoyuan Li, *Member, IEEE*, Yao Li, *Student Member, IEEE*, Dong Liu, *Senior Member, IEEE*, and Li Li, *Senior Member, IEEE*

***Abstract*—Inter prediction is a key technique for reducing temporal redundancy in video coding. However, irregular inter-frame illumination changes are hard to predict, yielding poorly predicted residuals that are difficult to compress. To address this, local illumination compensation (LIC) has been introduced into the next-generation exploratory coding model, which uses a linear model (LM) to approximate and compensate for inter-frame illumination variations, thereby reducing the prediction residual. Due to the simple design of LM and the least squares regression (LSR) solution in ECM, LIC suffers notable limitations, including the limited fitting capacity of the LM and the outlier sensitivity of LSR, which makes the compensation less robust to outliers. To address these limitations, we propose a multi-model local illumination compensation (MM-LIC) framework, which introduces the multi-model paradigm to optimize the model type (MMT) and model solution (MMS) jointly. In MMT, the candidate model types are extended beyond LM by incorporating a non-linear model (NLM) and a gradient linear model (GLM) to better handle diverse illumination variations. In MMS, ridge regression (RR) is further introduced to improve model solution accuracy and stability, especially when handling outliers. Meanwhile, in order to preserve decision accuracy without extra bit overhead, we design separable implicit decision mechanisms for the whole compensation process. The proposed method is implemented in ECM-8.0. Experimental results show that the proposed MM-LIC can achieve, on average, 0.45% and 0.23%, and achieve up to 0.68% and 0.40% BD-rate reduction under LDB and RA configurations, respectively, with only slight increases in time complexity. Furthermore, the proposed scheme also demonstrates its potential in the latest ECM version.**

***Index Terms*—Inter prediction, multi-model, local illumination compensation, model type, model solution, video coding.**

I. Introduction

In the modern video coding standards, the block-based hybrid coding framework has been widely adopted, such as H.265/HEVC [1] and H.266/VVC [2]. To reduce the temporal redundancy between video frames, inter prediction makes a major contribution to these video coding standards [3]–[6]. In general, inter prediction is employed to predict the current frame by leveraging previously coded frames, with its core implementation consisting of two key components: motion estimation (ME) and motion compensation (MC). In the predictive process, first, the current frame is usually divided into non-overlapping rectangular blocks as the basic units to match the prediction. For each block, ME is used to search for a motion vector (MV) that represents the displacement to its best-matching reference block in the reference frame. The MV is then used to perform MC and generate the prediction block.

While ME/MC is effective in modeling geometric motion, its prediction accuracy also relies on a photometric consistency assumption that the pixel intensity (or brightness) remains approximately unchanged across video frames after motion alignment. However, this assumption does not always hold in real-world videos. In practice, illumination may vary over time due to factors such as dynamic lighting, natural sunlight changes, or camera-side auto-exposure/auto-white-balance adjustments. Under these captured conditions, although ME/MC can find a correct motion correspondence, noticeable intensity discrepancies may still remain between the motion-compensated prediction and the original block. Such photometric mismatch typically increases the residual energy and degrades rate-distortion performance. Therefore, for sequences with inconsistent illumination, it is crucial to further adjust the motion-compensated prediction according to inter-frame illumination variations, which motivates the use of illumination compensation techniques. Along with the evolution of video coding standards, a variety of illumination compensation schemes have been proposed and can be broadly categorized into frame-level and block-level illumination compensation, depending on whether a single model is applied to the entire frame or adaptively estimated at a finer block-level granularity.

For frame-level illumination compensation schemes, a commonly used assumption is that illumination variations between adjacent frames can be modeled as a linear transform, as first proposed in [7]. Under this assumption, H.264/AVC [8] introduced the weighted prediction (WP) tool [9], where the reference slice is transformed by a scale and an offset to better match the target slice.

For block-level illumination compensation schemes, a block-level MC framework that jointly accounts for motion and illumination was early explored in [10], following a similar spirit to WP. Although WP can be signaled at a finer granularity, the bit overhead for transmitting the compensation parameters may impact its potential gains and thus limit its improvement. To reduce this overhead, localized WP (LWP) [11] and pixel-based illumination compensation (PBIC) [12] were proposed to model illumination variations using only one parameter of the linear model (either an offset or a scale), which can be implicitly derived at both the encoder and decoder. However, restricting the model to a single parameter substantially reduces its flexibility and often leads to

This work was supported in part by the National Key Research and Development Plan under Grant 2024YFF0505702. It was also supported by the CPU cluster built by MCC Lab of Information Science and Technology Institution, USTC. This article was recommended by Associate Editor Dr. Zhengxue Cheng. *(Corresponding author: Zhuoyuan Li.)*

The authors are with the MOE Key Laboratory of Brain-Inspired Intelligent Perception and Cognition, University of Science and Technology of China, Hefei 230027, China. (e-mail: {jialin.li, zhuoyuanli, mrliyao}@mail.ustc.edu.cn; {dongeliu, lil1}@ustc.edu.cn).

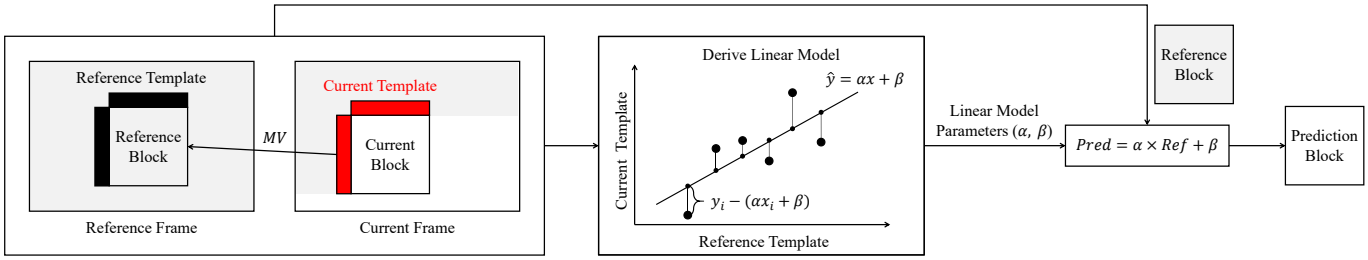


Fig. 1. Illustration of local illumination compensation (LIC). First, parameters of the linear model (LM) are derived based on the current block's template and the reference block's template by least squares regression (LSR). Then, the reference block indicated by the MV is adjusted by the derived LM to compensate for the inter-frame illumination variations. Finally, the prediction block obtained by LM adjustment is used as the final prediction of the current block.

inaccurate prediction under complex illumination changes. To overcome this limitation, inspired by illumination discrepancy compensation techniques developed for multiview video coding in 3D-HEVC [13]–[15], local illumination compensation (LIC) was proposed and explored in video coding standardization activities [16]. Unlike LWP and PBIC, LIC preserves a two-parameter linear model to capture inter-frame illumination variations, while still avoiding explicit signaling. Specifically, the model parameters are estimated via least squares regression (LSR) using the template extracted from the current block and its corresponding template in the reference frame, enabling identical parameter derivation at both the encoder and decoder.

The illumination compensation schemes reviewed above can alleviate inter-frame illumination inconsistency to some extent. However, their effectiveness is ultimately constrained by (i) the single linear assumption used to model illumination variations and (ii) the sensitivity of parameter estimation, issues that are particularly pronounced for LIC. Although the cooperation of a linear model (LM) and least squares regression (LSR) is efficient due to its simple applicability, LIC still exhibits intrinsic limitations in terms of the model type and the model solution. From the model type perspective, a single LM often lacks sufficient expressive power to capture sharp or spatially complex illumination variations, resulting in under-fitting and photometric mismatch after compensation. From the model solution perspective, when the correlation between the current block and its template is weak, the regression becomes more susceptible to outliers and ill-conditioning. This may produce unstable parameter estimates, resulting in prediction errors.

To address these limitations, motivated by our preliminary exploration of model formulation applied to the existing prediction pipeline, we propose a multi-model LIC framework, termed MM-LIC, which introduces the multi-model paradigm to optimize the existing LIC in Enhance Compression Model (ECM, the next-generation video coding exploration model beyond H.266/VVC) [17] in terms of model type and model solution, yielding multi-model type and multi-model solution, termed MMT and MMS. For the multi-model type, the candidate model types are extended beyond LM by incorporating non-linear model (NLM) [18]–[20] and gradient linear model (GLM) [21] to better handle diverse illumination variations. For the multi-model solution, in addition to the LSR, ridge regression (RR) [22], [23] is further introduced to improve the accuracy and stability of the model solution, which is insensitive to outliers. Meanwhile, in order to preserve decision accuracy without extra bit overhead, separate implicit decision mechanisms are designed for the selection of model type and model solution in multi-model LIC. For the model type, template cost-based model type selection is proposed to determine the optimal model type. For the model solution, content adaptive-based model solution selection is proposed to determine whether to use LSR or RR in the model solution. In summary, we make the following contributions that will be detailed in this paper:

- We propose a multi-model LIC (MM-LIC) framework to optimize the existing LIC in ECM.
- We propose a multi-model type (MMT) that the candidate model types are extended by incorporating NLM and GLM to better handle diverse illumination variations.
- We propose a multi-model solution (MMS) that further introduces RR to deal with outliers and improve the accuracy and stability of the model solution.
- We design separate implicit decision mechanisms for the selection of model type and model solution to preserve decision accuracy without extra bit overhead.

The rest of the paper is organized as follows. Section II reviews the related work of LIC. Section III clarifies the motivation of the MM-LIC framework. Section IV details the methodology of MM-LIC. Section V shows the experimental results and analysis. Section VI concludes the paper.

II. Related Work

In this section, we review the previous work that relates to our research from two aspects. First, we introduce the fundamental process of LIC. Second, we introduce some technologies related to LIC during its development in standards.

A. Local Illumination Compensation (LIC)

As illustrated in Fig. 1, LIC is an inter prediction technique to approximate the illumination variations between the current block and its reference block as a function of that between the current block's template and the reference block's template. The function can be denoted by a scale α and an offset β, which forms a LM to compensate for illumination variations. The LM can be expressed as:

$$Pred = \alpha \times Ref + \beta, \tag{1}$$

where Ref represents the reference block retrieved by MV in the reference frame. $Pred$ represents the final prediction of the current block. To avoid extra bit overhead, the parameters of LM are derived based on the current block's template and

the reference block's template by LSR. The least squares regression function can be expressed as:

$$\mathrm{LSR_{LM}}(\alpha,\beta)=\sum_{i=1}^{N}[Rec(i)-(\alpha\times Ref(i)+\beta)]^2, \quad (2)$$

by setting the derivative of Eq. (2) to zero, α and β can be expressed as:

$$\begin{aligned}\alpha&=\frac{N\times S_{xy}-S_x\times S_y}{N\times S_{xx}-S_x^2},\\ \beta&=\frac{S_{xx}\times S_y-S_x\times S_{xy}}{N\times S_{xx}-S_x^2},\end{aligned} \quad (3)$$

where S_x, S_y, S_{xx}, S_{xy} are expressed as:

$$\begin{cases}S_x &=\sum_{i=1}^{N}Ref(i)\\ S_y &=\sum_{i=1}^{N}Rec(i)\\ S_{xx}&=\sum_{i=1}^{N}Ref(i)^2\\ S_{xy}&=\sum_{i=1}^{N}Ref(i)\times Rec(i)\end{cases}, \quad (4)$$

where $Rec(i)$ and $Ref(i)$ are the i^{th} neighboring samples of the current block's template and reference block's template, and N is the total number of the paired neighboring samples.

B. Development of LIC in Standardization Activities

LIC is inspired by illumination-discrepancy compensation techniques developed for multiview video coding (e.g., 3D-HEVC [15]), where photometric inconsistencies across views (caused by camera/exposure differences) can notably impair inter-view prediction and thus motivate explicit illumination compensation [13]–[15]. Motivated by these observations, LIC was introduced and explored as a dedicated tool for compensating inter-frame illumination variations in the video coding standardization process via the ITU-T SG16/Q6 (VCEG) contribution VCEG-AZ06 [16]. Although LIC was not adopted into the final standards, it has been repeatedly investigated in exploration models, notably the Joint Exploration Model (JEM), an exploration platform beyond H.265/HEVC, and more recently the Enhanced Compression Model (ECM), an exploration platform beyond H.266/VVC, where it serves as a fundamental mechanism to refine motion-compensated prediction under illumination changes. Building upon this line of work, LIC was incorporated into the early ECM versions as a basic inter-frame prediction tool and applied to both Merge and AMVP modes to refine motion-compensated prediction under illumination changes [24]. Furthermore, a series of LIC-related enhancements have been proposed and progressively adopted into the advanced ECM versions to improve its effectiveness, including the non-local illumination compensation (NLIC) [25], LIC with BDMVR and BDOF [26], LIC with slope adjustment [27], bi-predictive LIC and OBMC with LIC [28], LIC flag derivation of merge candidates with template costs [29], etc. Collectively, these developments indicate that LIC has evolved into an effective prediction refinement mechanism in video coding standards.

In parallel with the development of LIC in standardization activities, a number of research works have also been devoted to further improving LIC. To improve LIC in scenes with sharp and complex illumination variations, Wang *et al.* proposed piecewise linear model-based LIC (PLMLIC) [30], which models the illumination relationship in a more flexible manner. Specifically, PLMLIC partitions neighboring samples into multiple groups using a classification strategy based on multiple reference lines and fits an individual linear model for each group to obtain piecewise parameters, thereby producing more accurate predictions.

In addition, LIC has also been extended to the intra coding, intra block copy with local illumination compensation (IBC-LIC) [31] was proposed to address illumination inconsistency between the current block and its reference block within the same frame. In IBC-LIC, the parameters of LM are derived in the same manner as in LIC. However, because screen content videos exhibit highly diverse textures and content, a single LM may be insufficient to compensate for the illumination variations between the current block and its intra reference. To address this limitation, an adaptive linear model for IBC-LIC (ALM-IBC-LIC) [32] was proposed, introducing a slope-adjustment parameter to improve adaptability. This parameter is determined at the encoder and signaled, enabling more accurate compensation under diverse illumination variations.

III. Motivation

While LIC has demonstrated its effectiveness in alleviating inter-frame illumination mismatch, its performance may degrade in challenging cases. In particular, a single compensation model may be insufficient to capture sharp and complex illumination variations, and parameter estimation may become unstable when the template correlation is weak. These observations indicate that further improving LIC requires enhancing both model expressiveness and estimation robustness.

To better quantify this improvement potential and to guide our subsequent design, we conduct a preliminary analysis under common test conditions (CTC) [33]. Specifically, we examine LIC-related samples extracted from the test sequences, including the coding blocks where LIC is activated, the corresponding templates used for parameter derivation, and the original blocks.

A. Preliminary Study on Model Type

Given that a single LM may be insufficient for complex illumination variations, we investigate the feasibility of enriching the model type in LIC by introducing multiple candidate models. Specifically, in addition to the conventional LM, we consider two-parameter gradient linear model (2P-GLM), three-parameter gradient linear model (3P-GLM), second-order non-linear model (2O-NLM), third-order non-linear model (3O-NLM) mentioned in [18]–[21]. The LM and the four additional candidates are formulated as:

$$\begin{cases}\textbf{LM}: & y=a_0+a_1x\\ \textbf{2P-GLM}: & y=a_0+a_1x_G\\ \textbf{3P-GLM}: & y=a_0+a_1x+a_2x_G\\ \textbf{2O-NLM}: & y=a_0+a_1x+a_2x^2\\ \textbf{3O-NLM}: & y=a_0+a_1x+a_2x^2+a_3x^3\end{cases}, \quad (5)$$

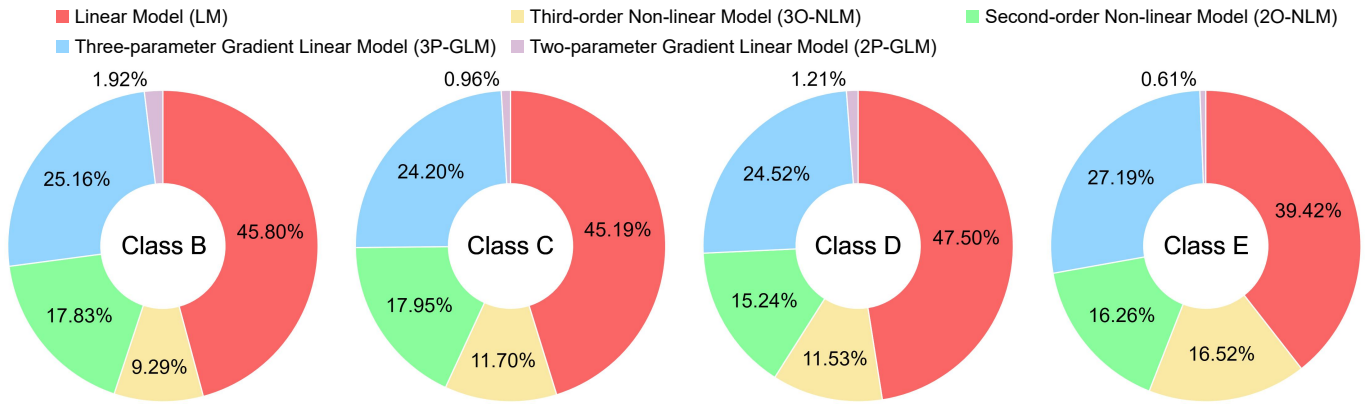


Fig. 2. The proportion of each model selected as the optimal model type for LIC. Candidate model types include linear model (LM), third-order non-linear model (3O-NLM), second-order non-linear model (2O-NLM), three-parameter gradient linear model (3P-GLM), and two-parameter gradient linear model (2P-GLM).

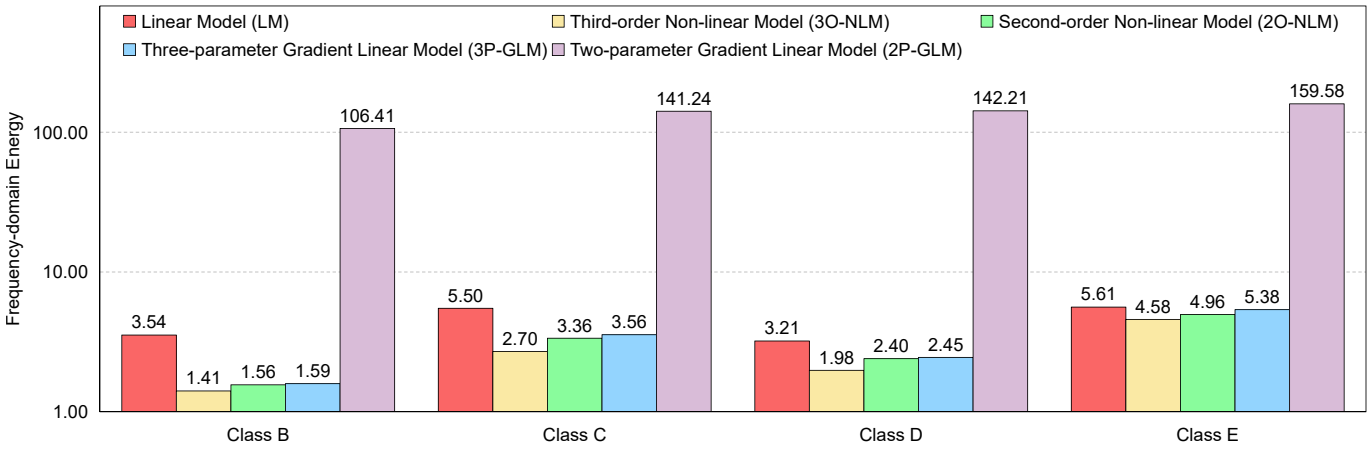


Fig. 3. The average frequency-domain energy of each model. Candidate model types include linear model (LM), third-order non-linear model (3O-NLM), second-order non-linear model (2O-NLM), three-parameter gradient linear model (3P-GLM), and two-parameter gradient linear model (2P-GLM).

where x and x_G denote the pixel value and the corresponding gradient of the reference block retrieved by the MV in the reference frame, and y denotes the final prediction of the current block.

For each sample pair in the dataset, we instantiate each candidate in Eq. (5) as the model type and solve its parameters using LSR, thereby obtaining the corresponding illumination-compensated prediction block for each model. We then compute the sum of absolute differences (SAD) between the illumination-compensated prediction block and the original block, and select the model with the minimum SAD as the optimal one for that sample. Based on these per-sample selections, we summarize the selection ratio of each model over the entire dataset. In addition, we derive the residual by subtracting the illumination-compensated prediction block from the original block, apply the discrete cosine transform (DCT) to the residual, and report the average frequency-domain energy for each model.

The proportion of each model selected as the optimal model for LIC is illustrated in Fig. 2. We can observe that LM is most frequently selected as the model type of LIC in video sequences across different resolutions, accounting for approximately 45%. The proportion of GLM is selected as the model type of LIC is approximately 25%, with 3P-GLM accounting for virtually the whole proportion and 2P-GLM accounting for a negligible proportion. The proportion of NLM is selected as the model type of LIC is approximately 30%, with 2O-NLM is roughly the same as 3O-NLM. The average frequency-domain energy of each model is illustrated in Fig. 3. We can observe that all models except the 2P-GLM have lower average frequency-domain energy compared with LM.

Based on the observation, for the GLM, because the proportion of 2P-GLM selected as the optimal model is negligible and 2P-GLM has extremely higher average frequency-domain energy compared with LM, 3P-GLM rather than 2P-GLM is selected as the potential model type for LIC. For the NLM, the M-order NLM is formulated as:

$$y = a_0 + a_1 x + a_2 x^2 + \cdots + a_M x^M, \tag{6}$$

the corresponding computational complexity can be expressed as $\mathcal{O}(NM^2+M^3)$, where N is the number of samples, and M is the order of the NLM. When $M = 2$, computational complexity $\mathcal{O}_{\text{2O-NLM}} = \mathcal{O}(4N + 8)$; When $M = 3$, computational complexity $\mathcal{O}_{\text{3O-NLM}} = \mathcal{O}(9N + 27)$. Therefore, although the proportion of 3O-NLM selected as the optimal model is only slightly lower than that of 2O-NLM and the average frequency-domain energy of 3O-NLM is slightly lower than that of 2O-NLM, the computational complexity of the model solution increases by nearly two times. Thus, 2O-NLM rather than 3O-NLM is selected as the potential model type for LIC.

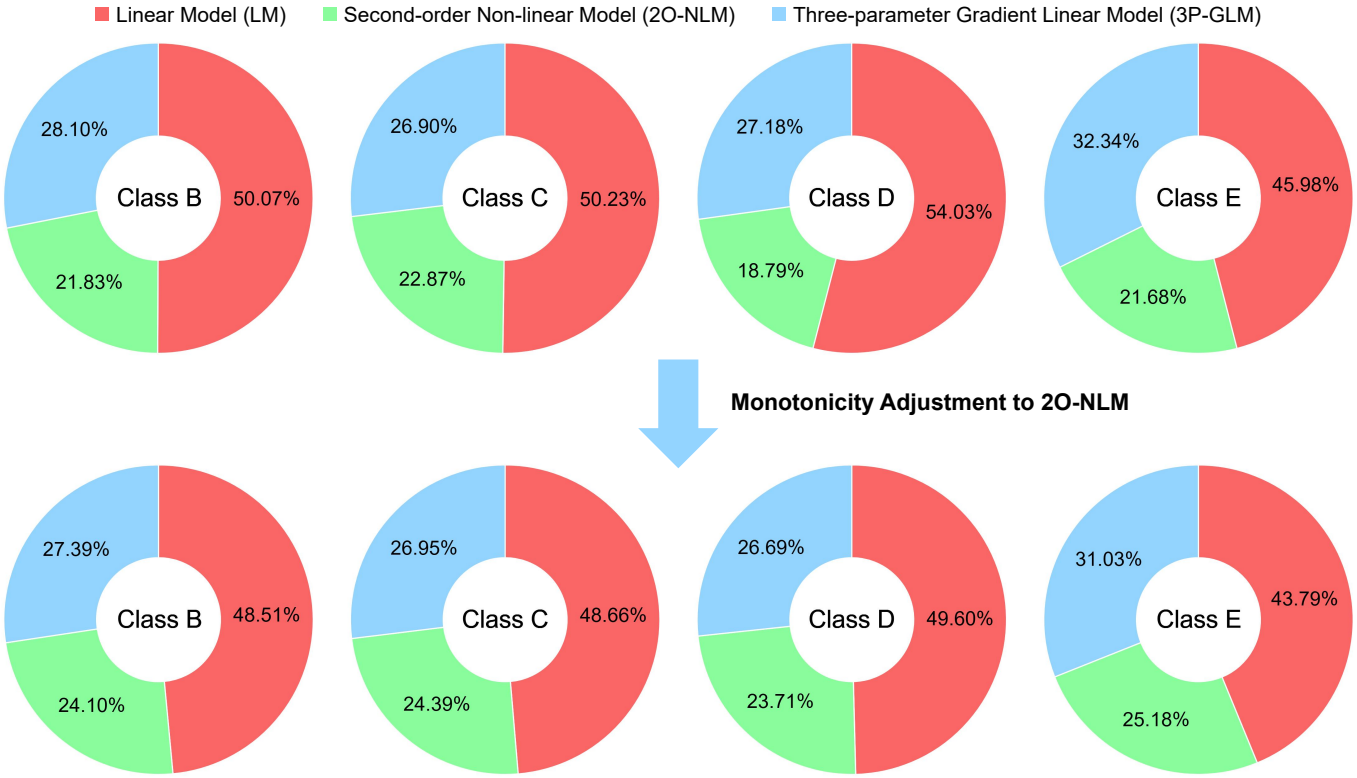


Fig. 4. The proportion of each model selected as the optimal model type for LIC before and after applying the monotonicity adjustment to the 2O-NLM. Candidate model types include linear model (LM), second-order non-linear model (2O-NLM), and three-parameter gradient linear model (3P-GLM). The pie charts above shows the result before applying monotonicity adjustment to 2O-NLM. The pie charts below shows the result after applying monotonicity adjustment to 2O-NLM.

Based on the above analysis, 3P-GLM and 2O-NLM are selected as the potential model types for LIC. Furthermore, for the 2O-NLM, we draw inspiration from prior experience in the low-light image restoration task, particularly the monotonicity-aware adjustment to quadratic illumination mapping adopted in [18]. Following this design intuition, the standard quadratic model in Eq. (5) is reformulated as $y = a_0 + x + a_1 x(1-x)$, where $a_1 \in [-1, 1]$. This reformulation enforces a monotonicity constraint during illumination compensation, making it more consistent with typical real-world illumination variations. Similar to before, the proportion of each model selected as the optimal model for LIC before and after applying the monotonicity adjustment to the 2O-NLM is statistically summarized again, as illustrated in Fig. 4. Before applying the monotonicity adjustment, we can observe that the proportions of LM, 2O-NLM, and 3P-GLM are 50%, 20% and 30%, respectively, which are essentially the same as those in Fig. 2. After applying the monotonicity adjustment, we can observe that the proportion of 3P-GLM remains largely unchanged, while the proportion of 2O-NLM increases and that of LM decreases. This indicates that, although the theoretical expressiveness of 2O-NLM decreases after monotonicity adjustment, the monotonicity-constrained 2O-NLM can better align with real-world scenarios and achieve more reliable illumination compensation. Following the above analysis, we keep 3P-GLM and monotonicity-constrained 2O-NLM as potential model types for LIC.

B. Preliminary Study on Model Solution

Beyond model expressiveness, the robustness of parameter estimation is another key factor for LIC. Since LSR can be sensitive to outliers when the template correlation is weak, we investigate the feasibility of introducing alternative solvers for more stable parameter estimation. Specifically, we construct a sub-dataset that emphasizes more challenging cases for parameter estimation by selecting sample pairs whose reference blocks exhibit relatively large intensity variance. Such blocks typically contain richer textures and larger local fluctuations, where template mismatches are more likely to occur and thereby make parameter fitting more susceptible to outliers in practice.

In addition to the conventional least squares regression (LSR) used in LIC, we consider lasso regression (Lasso) [34], [35], ridge regression (RR) [22], [23], and elastic net regression (EN) [36], [37] as alternative solvers due to regularization effect. The objective functions of LSR and the three additional model solvers are given by:

$$\begin{cases} J_{\text{LSR}}(\mathbf{w}) &= \sum_{i=1}^{N}(y_i - \mathbf{w}^T\mathbf{x}_i)^2 \\ J_{\text{Lasso}}(\mathbf{w}) &= \sum_{i=1}^{N}(y_i - \mathbf{w}^T\mathbf{x}_i)^2 + \lambda\|\mathbf{w}\|_1 \\ J_{\text{RR}}(\mathbf{w}) &= \sum_{i=1}^{N}(y_i - \mathbf{w}^T\mathbf{x}_i)^2 + \lambda\|\mathbf{w}\|_2^2 \\ J_{\text{EN}}(\mathbf{w}) &= \sum_{i=1}^{N}(y_i - \mathbf{w}^T\mathbf{x}_i)^2 + +\lambda_1\|\mathbf{w}\|_1 + \lambda_2\|\mathbf{w}\|_2^2 \end{cases}, \tag{7}$$

where $\mathbf{x}_i$ and y_i are the input vector and target value, $\mathbf{w}$ represents the parameter vector to be estimated, λ, λ_1 and λ_2 are regularization coefficients, and N is the number of paired neighboring samples.

Fig. 5. The average frequency-domain energy of residuals obtained with different solvers. Candidate model solutions include least squares regression, lasso regression, ridge regression, and elastic net regression.

Following the same evaluation protocol as in Sec. III-A, we fix the model type to LM and only vary the solvers to obtain the illumination-compensated prediction blocks. We then compute the residuals and report their average frequency-domain energy after DCT, which serves as an indicator of fitting quality and robustness. As shown in Fig. 5, we observe that using regularized solvers produces residuals with lower average frequency-domain energy than the plain LSR baseline, which demonstrates that introducing regularization (or, more generally, a more robust estimation strategy) helps stabilize parameter fitting under weak template correlation and mitigates the impact of outliers. However, among solvers with regularization terms, we observe that the corresponding average frequency-domain energy of lasso regression with an L_1 regularization term is only slightly lower than the plain LSR baseline, but ridge regression with an L_2 regularization term performs much better than lasso regression. Meanwhile, elastic net regression with the combination of L_1 and L_2 regularization terms also performs much better than lasso regression, but performs worse than ridge regression. This demonstrates that using only the L_2 regularization term yields superior performance in handling outliers. Following the above analysis, we keep ridge regression as a potential model solution for LIC.

IV. Multi-Model Local Illumination Compensation

In this section, we detail the proposed MM-LIC. First, the overall framework of MM-LIC is proposed. Second, the two core techniques of MM-LIC, MMT and MMS are introduced. Third, the separate implicit decision mechanisms designed for MMT and MMS are introduced. Finally, the integration of MM-LIC with existing inter modes in ECM is introduced.

A. Framework

As illustrated in Fig. 6, MM-LIC augments the existing LIC tool in ECM by introducing a multi-model paradigm from two complementary aspects: *model type* (MMT) and *model solution* (MMS), both determined implicitly without extra signaling. For a current block, its reference block and a corresponding pair of templates (current block's template and reference block's template) are used for parameter fitting and decision making. In MMT, a set of candidate compensation models (LM/NLM/GLM) is evaluated, and the optimal model type is selected via a template-cost criterion, i.e., the model that best matches the current block's template after compensation is preferred. In MMS, multiple solvers are considered for parameter estimation, and a content-adaptive solver decision is employed to select a more stable solver under weak correlation with outlier-contaminated templates. Finally, the selected model type and solver are used to estimate the compensation parameters from the template samples, and the resulting model is applied to the reference block to generate the illumination-compensated prediction block, which is then used as the final prediction for inter modes.

B. Multi-Model Type (MMT)

For the model type design of the existing LIC tool in ECM, only a linear model is used to approximate and compensate for the illumination variations between the current block and its reference block. While the linear assumption is simple and efficient, a single LM has limited expressive power and may be inadequate under diverse illumination changes. Therefore, in order to improve adaptability, we propose MMT to augment the candidate model set by incorporating a non-linear model and a gradient linear model as optional model types.

1) Non-linear Model (NLM): In general, a model's expressive power is closely related to its order: higher-order models typically provide greater flexibility and can represent more complex relationships. Accordingly, higher-order non-linear models are potential candidates for LIC in scenes with sharp and complex illumination changes. However, increasing the model order also raises the computational cost of parameter estimation. Balancing this complexity–accuracy trade-off, we adopt a quadratic non-linear model as the optional NLM in MMT, which can be expressed as:

$$Pred_n = \alpha_0 \times Ref_n^2 + \alpha_1 \times Ref_n + \beta. \tag{8}$$

To further ensure the monotonicity of the quadratic non-linear model when $Ref_n \in [0, 1]$, a corresponding adjustment as

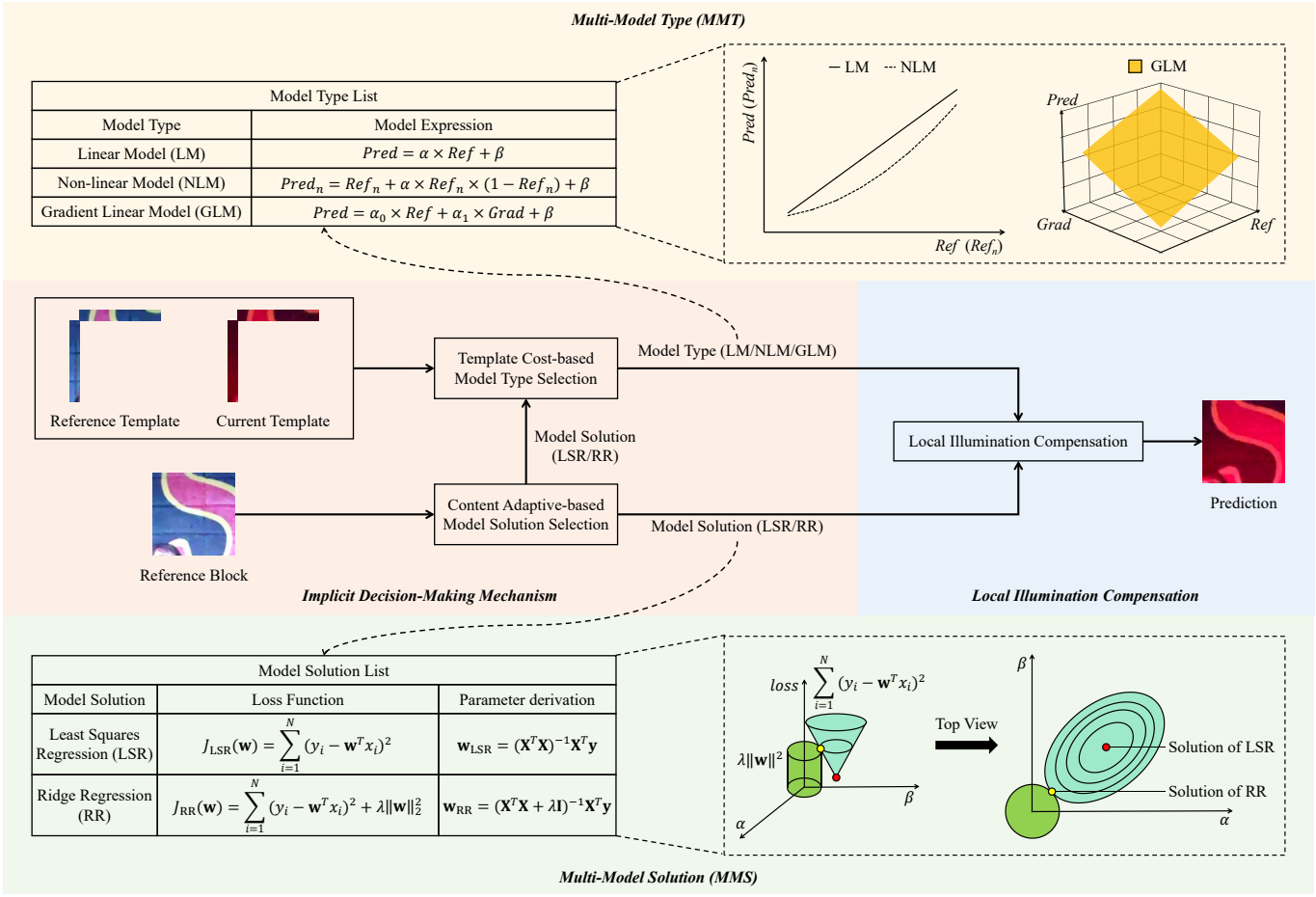


Fig. 6. Illustration of the proposed multi-model local illumination compensation (MM-LIC) framework. First, the reference block of the current block and a corresponding pair of templates (current block's template and reference block's template) are used to determine the model type and model solution for MM-LIC. Specifically, template cost-based model type selection is used to determine the optimal model type in the multi-model type (MMT). Content adaptive-based model solution selection is used to determine the optimal model solution in the multi-model solution (MMS). Second, the selected model type and model solution are used for local illumination compensation to generate the illumination-compensated prediction block. Finally, the illumination-compensated prediction block is used as the final prediction of the current block.

mentioned in [18] is required for the NLM in Eq. (8), which can be expressed as:

$$Pred_n = Ref_n + \alpha \times Ref_n \times (1 - Ref_n) + \beta, \tag{9}$$

where Ref_n represents the reference block retrieved by MV in the reference frame after normalization, and $Pred_n$ represents the final prediction of the current block after normalization. $\alpha \in [-1, 1]$ is a scale factor controlling the magnitude of the NLM, and β is the offset. The least squares regression function of NLM can be expressed as:

$$\mathrm{LSR}_{\mathrm{NLM}}(\alpha, \beta) = \sum_{i=1}^{N} [r(i) - (\alpha \times f(i) + \beta)]^2, \tag{10}$$

by setting the derivative of Eq. (10) to zero, α and β can be expressed as:

$$\begin{aligned} \alpha &= \frac{N \times S_{fr} - S_f \times S_r}{N \times S_{ff} - S_f^2}, \\ \beta &= \frac{S_{ff} \times S_r - S_f \times S_{fr}}{N \times S_{ff} - S_f^2}, \end{aligned} \tag{11}$$

where $f(i)$, $r(i)$, S_f, S_r, S_{ff}, S_{fr} are expressed as:

$$\begin{cases} f(i) &= Ref_n(i) \times (1 - Ref_n(i)) \\ r(i) &= Rec_n(i) - Ref_n(i) \\ S_f &= \sum_{i=1}^{N} f(i) \\ S_r &= \sum_{i=1}^{N} r(i) \\ S_{ff} &= \sum_{i=1}^{N} f(i)^2 \\ S_{fr} &= \sum_{i=1}^{N} f(i) \times r(i) \end{cases}, \tag{12}$$

where $Rec_n(i)$ and $Ref_n(i)$ are the i^{th} neighboring samples of the current block's template and reference block's template after normalization, and N is the total number of the paired neighboring samples.

2) Gradient Linear Model (GLM): In the LM, only the pixel values which represent low-frequency information are involved. This leads to the correlation between reference samples and current samples still not being fully exploited. Specifically, LM ignores the relative spatial variation information in reference samples, such as edge and gradient information. Therefore, in order to make full use of low-frequency and high-frequency information, gradient information is further introduced to form a three-parameter gradient linear model, which can be used as an optional model GLM in MMT. To

calculate the gradient of the reference block and its template, the first-order differential-based Prewitt operator [38] is introduced to calculate the gradient between the current pixel and adjacent pixels, which can be expressed as:

$$Grad(x,y) = \sqrt{(\mathbf{S} * \mathbf{P})^2 + (\mathbf{S} * \mathbf{P}^T)^2}, \tag{13}$$

where $\mathbf{S}$ is a 3×3 matrix consisting of the pixel at the (x, y) position and its surrounding pixels, and $\mathbf{P}$ is the Prewitt operator, which can be expressed as:

$$\mathbf{P} = \begin{vmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{vmatrix}. \tag{14}$$

Then, GLM can be expressed as:

$$Pred = \alpha_0 \times Ref + \alpha_1 \times Grad + \beta, \tag{15}$$

where Ref and $Grad$ represent the pixel value and gradient of the reference block retrieved by MV in the reference frame, $Pred$ represents the final prediction of the current block, α_0 and α_1 are the scale, and β is the offset. The least squares regression function of GLM can be expressed as:

$$\begin{aligned} &\mathrm{LSR}_{\mathrm{GLM}}(\alpha_0, \alpha_1, \beta) = \\ &\sum_{i=1}^{N} [Rec(i) - (\alpha_0 \times Ref(i) + \alpha_1 \times Grad(i) + \beta)]^2, \end{aligned} \tag{16}$$

by setting the derivative of Eq. (16) to zero, α_0, α_1 and β can be expressed as:

$$\begin{aligned} \alpha_0 &= \frac{\begin{vmatrix} N & S_z & S_y \\ S_x & S_{xz} & S_{xy} \\ S_y & S_{yz} & S_{yy} \end{vmatrix}}{\begin{vmatrix} N & S_x & S_y \\ S_x & S_{xx} & S_{xy} \\ S_y & S_{xy} & S_{yy} \end{vmatrix}}, \\ \alpha_1 &= \frac{\begin{vmatrix} N & S_x & S_z \\ S_x & S_{xx} & S_{xz} \\ S_y & S_{xy} & S_{yz} \end{vmatrix}}{\begin{vmatrix} N & S_x & S_y \\ S_x & S_{xx} & S_{xy} \\ S_y & S_{xy} & S_{yy} \end{vmatrix}}, \\ \beta &= \frac{\begin{vmatrix} S_z & S_x & S_y \\ S_{xz} & S_{xx} & S_{xy} \\ S_{yz} & S_{xy} & S_{yy} \end{vmatrix}}{\begin{vmatrix} N & S_x & S_y \\ S_x & S_{xx} & S_{xy} \\ S_y & S_{xy} & S_{yy} \end{vmatrix}}, \end{aligned} \tag{17}$$

where S_x, S_y, S_z, S_{xx}, S_{yy}, S_{xy}, S_{xz}, S_{yz} are expressed as:

$$\begin{cases} S_x &= \sum_{i=1}^{N} Ref(i) \\ S_y &= \sum_{i=1}^{N} Grad(i) \\ S_z &= \sum_{i=1}^{N} Rec(i) \\ S_{xx} &= \sum_{i=1}^{N} Ref(i)^2 \\ S_{yy} &= \sum_{i=1}^{N} Grad(i)^2 \\ S_{xy} &= \sum_{i=1}^{N} Ref(i) \times Grad(i) \\ S_{xz} &= \sum_{i=1}^{N} Ref(i) \times Rec(i) \\ S_{yz} &= \sum_{i=1}^{N} Grad(i) \times Rec(i) \end{cases}, \tag{18}$$

where $Rec(i)$ and $Ref(i)$ are the i^{th} neighboring samples of the current block's template and reference block's template, $Grad(i)$ is the i^{th} gradient of the reference block's template, and N is the total number of the paired neighboring samples.

C. Multi-Model Solution (MMS)

For the model solution design of the existing LIC in ECM, the model parameters are derived by the LSR, which minimizes the sum of squared error (SSE) between the predicted and target values. It can be generally expressed as:

$$J_{\mathrm{LSR}}(\mathbf{w}) = \sum_{i=1}^{N} (y_i - \mathbf{w}^T \mathbf{x}_i)^2, \tag{19}$$

where $\mathbf{x}_i$ and y_i are the input vector and target value, $\mathbf{w}$ represents the parameter vector to be estimated, and N is the total number of the paired neighboring samples. The model solution of LSR can be expressed as:

$$\mathbf{w}_{\mathrm{LSR}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}, \tag{20}$$

where $\mathbf{X}$ and $\mathbf{y}$ are the input matrix and target vector.

Although solving the model via LSR is simple and efficient, the resulting parameters can deviate substantially in certain scenarios. This issue becomes pronounced when the correlation between the current block and its template is weak: outliers can dominate the fitting process, yielding inaccurate parameters and thus poor compensation of inter-frame illumination variations. To improve robustness, we propose MMS, where RR is additionally incorporated as an optional solution for parameter estimation. Compared to the LSR, RR incorporates an L_2 regularization term to mitigate the influence of outliers. It can be generally expressed as:

$$J_{\mathrm{RR}}(\mathbf{w}) = \sum_{i=1}^{N} (y_i - \mathbf{w}^T \mathbf{x}_i)^2 + \lambda \|\mathbf{w}\|_2^2, \tag{21}$$

where $\lambda > 0$ is a regularization coefficient controlling the trade-off between fitting accuracy and parameter smoothness. When $\lambda = 0$, Eq. (21) and (22) degenerate to the Eq. (19) and (20). The model solution of RR can be expressed as:

$$\mathbf{w}_{\mathrm{RR}} = (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T \mathbf{y}, \tag{22}$$

where $\mathbf{I}$ is the identity matrix. When RR is used for the model solution, the solution procedures of each model in MMT are correspondingly adjusted. Meanwhile, to ensure that the solution better aligns with the actual process of illumination compensation, the L_2 regularization term of RR is also adjusted correspondingly.

For LM, the ridge regression function can be expressed as:

$$\begin{aligned} &\mathrm{RR}_{\mathrm{LM}}(\alpha, \beta) = \\ &\sum_{i=1}^{N} [Rec(i) - (\alpha \times Ref(i) + \beta)]^2 + \lambda \times (\alpha - 1)^2, \end{aligned} \tag{23}$$

by setting the derivative of Eq. (23) to zero, α and β can be expressed as:

$$\begin{aligned} \alpha &= \frac{N \times S_{xy} - S_x \times S_y + \lambda \times N}{N \times S_{xx} - S_x^2 + \lambda \times N}, \\ \beta &= \frac{S_{xx} \times S_y - S_x \times S_{xy} + \lambda \times (S_y - S_x)}{N \times S_{xx} - S_x^2 + \lambda \times N}. \end{aligned} \tag{24}$$

For NLM, the ridge regression function can be expressed as:

$$\mathrm{RR_{NLM}}(\alpha,\beta) = \sum_{i=1}^{N}[r(i)-(\alpha\times f(i)+\beta)]^2+\lambda\times\alpha^2, \tag{25}$$

by setting the derivative of Eq. (25) to zero, α and β can be expressed as:

$$\begin{aligned}\alpha &= \frac{N\times S_{fr}-S_f\times S_r}{N\times S_{ff}-S_f^2+\lambda\times N},\\ \beta &= \frac{S_{ff}\times S_r-S_f\times S_{fr}+\lambda\times S_r}{N\times S_{ff}-S_f^2+\lambda\times N}.\end{aligned} \tag{26}$$

For GLM, the ridge regression function can be expressed as:

$$\mathrm{RR_{GLM}}(\alpha_0,\alpha_1,\beta) = \sum_{i=1}^{N}[Rec(i)-(\alpha_0\times Ref(i)+\alpha_1\times Grad(i)+\beta)]^2+\lambda\times(\alpha_0-1)^2, \tag{27}$$

by setting the derivative of Eq. (27) to zero, α_0, α_1 and β can be expressed as:

$$\begin{aligned}\alpha_0 &= \frac{\begin{vmatrix} N & S_z & S_y\\ S_x & S_{xz}+\lambda & S_{xy}\\ S_y & S_{yz} & S_{yy}\end{vmatrix}}{\begin{vmatrix} N & S_x & S_y\\ S_x & S_{xx}+\lambda & S_{xy}\\ S_y & S_{xy} & S_{yy}\end{vmatrix}},\\ \alpha_1 &= \frac{\begin{vmatrix} N & S_x & S_z\\ S_x & S_{xx}+\lambda & S_{xz}+\lambda\\ S_y & S_{xy} & S_{yz}\end{vmatrix}}{\begin{vmatrix} N & S_x & S_y\\ S_x & S_{xx}+\lambda & S_{xy}\\ S_y & S_{xy} & S_{yy}\end{vmatrix}},\\ \beta &= \frac{\begin{vmatrix} S_z & S_x & S_y\\ S_{xz}+\lambda & S_{xx}+\lambda & S_{xy}\\ S_{yz} & S_{xy} & S_{yy}\end{vmatrix}}{\begin{vmatrix} N & S_x & S_y\\ S_x & S_{xx}+\lambda & S_{xy}\\ S_y & S_{xy} & S_{yy}\end{vmatrix}}.\end{aligned} \tag{28}$$

D. Implicit Decision Mechanisms in MM-LIC

For the model type and model solution design in MM-LIC, we provide multiple options in both MMT and MMS. Specifically, MMT supports three candidate model types, i.e., LM, NLM, and GLM, while MMS supports two parameter-estimation methods, i.e., LSR and RR. To maintain accurate decisions without introducing additional bit overhead, we design separate implicit decision mechanisms for MM-LIC.

1) Content Adaptive-based Model Solution Selection: For MMS, content adaptive-based model solution selection is proposed to determine whether to use LSR or RR in the model solution for MM-LIC. Specifically, the variance of the reference block $\mathrm{Var}(RefBlk)$ is used as a proxy to estimate the correlation between the current block and its template. The higher the variance of the reference block, the lower the correlation between the current block and its template. Outliers are more likely to be introduced in the model solution when the correlation is low. The variance is calculated as:

$$\mathrm{Var}(RefBlk) = \frac{\sum_{i=1}^{K}(RefBlk(i)-RefBlk_{\mathrm{avg}})^2}{K}, \tag{29}$$

where K is block size of the reference block, $RefBlk(i)$ is the i^{th} sample of the reference block, and $RefBlk_{\mathrm{avg}}$ is the average of the reference block. If $\mathrm{Var}(RefBlk)$ falls below a preset threshold, the default LSR remains unchanged in the model solution; otherwise, the LSR will be switched to the RR in the model solution to mitigate the impact of outliers.

2) Template Cost-based Model Type Selection: For MMT, as illustrated in Fig. 7, template cost-based model type selection is proposed to select the optimal model type for MM-LIC. First, the model solution decided by the content adaptive-based model solution selection is used to derive the parameters of each compensation model in MMT. Second, each model is applied to the reference block's template to get the template of the corresponding prediction. Third, a self-corrective decision strategy is applied to derive the template cost of each model in MMT. Specifically, in the model solution, L_2 loss is used as the metric to derive the model parameters. In the calculation of the template cost, to avoid overfitting and mitigate the impact of outliers, L_1 loss with greater robustness is selected. Therefore, instead of the MSE metric in LSR and RR, the SAD between the prediction template of each model and the current block's template is used as the distortion metric. The SAD is calculated as:

$$\mathrm{SAD} = \sum_{i=1}^{N}|Rec(i)-Pred(i)|, \tag{30}$$

where $Rec(i)$ and $Pred(i)$ are the i^{th} neighboring samples of the current block's template and prediction template, and N is the total number of the paired neighboring samples. Finally, the model corresponding to the minimum SAD is selected as the model type for MM-LIC.

E. Integration of MM-LIC with Existing Inter Modes in ECM

For the integration of MM-LIC, as illustrated in Algorithm 1, the existing inter modes (AMVP and Merge) [5] of ECM are modified accordingly from the perspective of the decoding process. In the decoding process, first, the motion data of the current coding block (CB) is obtained by parsing the bitstream. Second, for Merge or AMVP mode which is indicated by the $mergeFlag$, the decoded motion information mv is obtained via different approaches. For Merge mode, mv is derived with the $mergeIdx$. For AMVP mode, mv is derived with the $mvpIdx$ and motion vector difference mvd. Third, MC is performed based on the $(mv, refIdx)$ to get the reference block $RefBlk$ and its template $RefTmp$ in the reference frame. Fourth, separate implicit decision mechanisms are used to determine the model type and model solution of MM-LIC. To determine the model solution of MM-LIC, the variance of the reference block $\mathrm{Var}(RefBlk)$ is used to determine whether to use LSR or RR for the model solution. When the variance is larger than the threshold TH, the RR

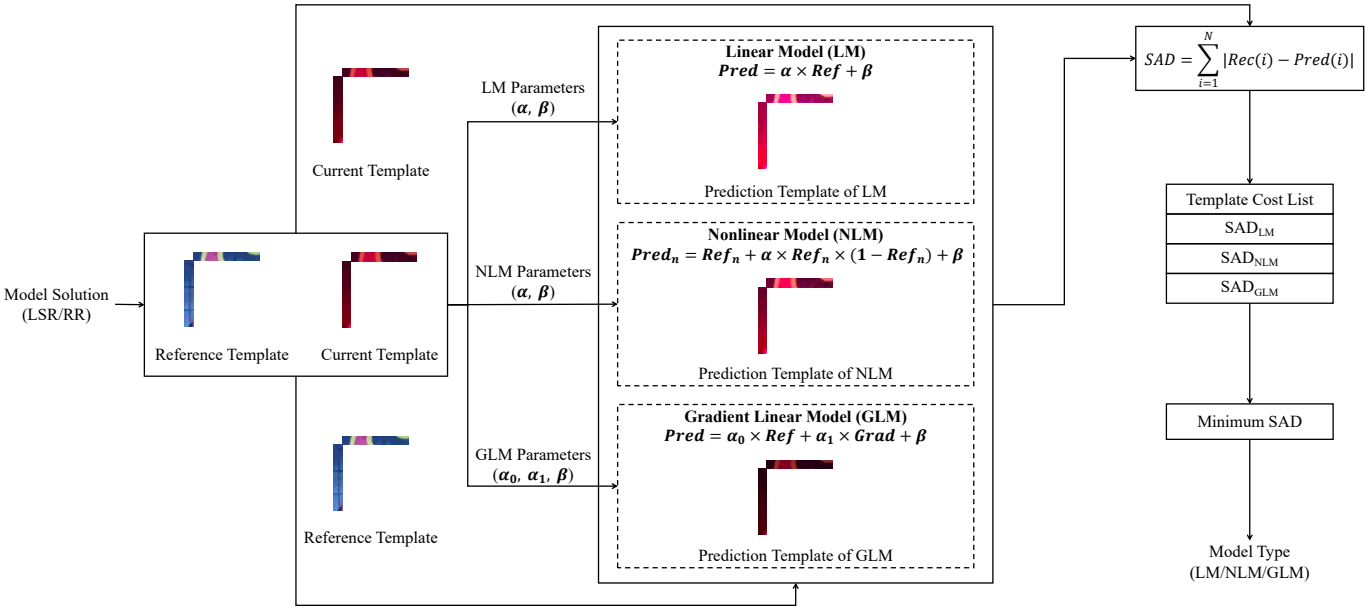


Fig. 7. Illustration of template cost-based model type selection. First, the model solution derived by the content adaptive-based model solution selection is used to estimate the compensation parameters of candidate compensation models (LM/NLM/GLM) from the template samples. Second, each resulting model is applied to the reference block's template to generate the corresponding illumination-compensated prediction template. Third, the SAD between the current block's template and the prediction template of each model is calculated to construct a template cost list. Finally, the model corresponding to the minimum SAD is selected as the model type for MM-LIC.

Algorithm 1 Integration of MM-LIC

Input: syntax elements of the CB
Output: prediction of the CB with MM-LIC: $Pred_{\text{MM-LIC}}$

if $mergeFlag == 1$ **then**
  $mv \leftarrow mergeIdx$
else
  $mv \leftarrow mvpIdx + mvd$
end if
$RefBlk, RefTmp \leftarrow$ MC $(mv, refIdx)$
Compute variance of reference block Var$(RefBlk)$
if Var$(RefBlk) > TH$ **then**
  $\text{MS}_{\text{MM-LIC}} \leftarrow$ RR
else
  $\text{MS}_{\text{MM-LIC}} \leftarrow$ LSR
end if
for each MT $\in \{\text{LM}, \text{NLM}, \text{GLM}\}$ **do**
  $PredTmp \leftarrow \text{MM-LIC}(RefTmp, \text{MT}, \text{MS}_{\text{MM-LIC}})$
  $\text{SAD}_{\text{MT}} \leftarrow \text{SAD}(PredTmp, CurTmp)$
end for
$\text{SAD}_{\text{min}} \leftarrow \min\{\text{SAD}_{\text{LM}}, \text{SAD}_{\text{NLM}}, \text{SAD}_{\text{GLM}}\}$
$\text{MT}_{\text{MM-LIC}} \leftarrow$ MT corresponding to SAD_{min}
$Pred_{\text{MM-LIC}} \leftarrow \text{MM-LIC}(RefBlk, \text{MT}_{\text{MM-LIC}}, \text{MS}_{\text{MM-LIC}})$
return $Pred_{\text{MM-LIC}}$

is used instead of the LSR, and vice versa. To determine the model type of MM-LIC, the candidate compensation models (LM/NLM/GLM) and the model solution previously obtained are applied to the reference block's template to get the prediction template $PredTmp$ of each model. Then, the SAD between the prediction template and the current block's template $CurTmp$ for each model is calculated. The model type corresponding to the minimum SAD is selected. Finally, the obtained model type and model solution are applied to the reference block to get the final prediction $Pred_{\text{MM-LIC}}$.

V. Experiment Results

A. Experimental Settings

To evaluate the proposed scheme, we implement MM-LIC in the Enhanced Compression Model (ECM) version 8.0 [39], which serves as a key exploration platform for next-generation video coding standardization. Experiments are conducted under two main profile configurations, Low-Delay B (LDB) and Random Access (RA), using the test sequences specified in [33]. Following the Common Test Conditions (CTC) [33], four quantization parameters (QPs) are evaluated: 22, 27, 32, and 37. For the measure of coding performance, the Bjontegaard Delta bit-rate (BD-rate) [40] is used as the objective metric.

B. Performance

In this subsection, we show the overall performance of the proposed MM-LIC. First, we show its coding performance. Second, we show the encoding/decoding complexity. Third, we test MM-LIC on the test sequence with lots of complex and sharp illumination variations. Finally, we test MM-LIC on latest ECM version.

1) Coding Performance: The R-D performance of the MM-LIC framework under LDB and RA configurations on the JVET common test sequences is illustrated in Table I. Y, U and V represent the R-D performance gain of the three channels of YUV. We can observe that the proposed MM-LIC can achieve, on average, 0.45% and 0.23% (marked by font bold), and achieve up to 0.68% and 0.40% (marked by underline) BD-rate reduction (Y component) under LDB and RA configurations, respectively. The experimental results

TABLE I
BD-RATE RESULTS OF PROPOSED MM-LIC COMPARED TO ECM-8.0 ANCHOR UNDER LDB AND RA CONFIGURATIONS ON CTC TEST SEQUENCES

Class	Configuration	Low-delay B (%)			Random Access (%)		
	Sequence	Y	U	V	Y	U	V
ClassA1 (3840x2160)	*Campfire*	—	—	—	-0.10%	0.19%	0.66%
	FoodMarket4	—	—	—	-0.23%	0.34%	0.33%
	Tango2	—	—	—	-0.31%	-1.78%	0.64%
	Average	—	—	—	**-0.21%**	**-0.41%**	**0.55%**
ClassA2 (3840x2160)	*CatRobot*	—	—	—	-0.30%	-0.46%	0.28%
	DaylightRoad2	—	—	—	-0.11%	-0.43%	0.04%
	ParkRunning3	—	—	—	-0.05%	-0.01%	-0.08%
	Average	—	—	—	**-0.16%**	**-0.30%**	**0.08%**
ClassB (1920x1080)	*BasketballDrive*	-0.08%	-0.57%	-0.54%	-0.26%	-0.50%	0.07%
	BQTerrace	-0.49%	-0.24%	0.30%	-0.29%	0.27%	0.23%
	Cactus	-0.43%	-0.32%	0.03%	-0.17%	-0.01%	0.10%
	MarketPlace	-0.34%	0.33%	0.28%	-0.03%	-0.42%	-0.48%
	RitualDance	-0.43%	-1.30%	0.24%	-0.20%	0.37%	-0.16%
	Average	**-0.35%**	**-0.42%**	**0.06%**	**-0.19%**	**-0.06%**	**-0.05%**
ClassC (840x480)	*BasketballDrill*	-0.68%	0.19%	-0.70%	-0.14%	0.18%	0.06%
	BQMall	-0.57%	-1.52%	-0.70%	-0.35%	-0.54%	-0.35%
	PartyScene	-0.25%	-0.63%	-0.23%	-0.32%	-0.12%	-0.29%
	RaceHorses	-0.62%	1.05%	1.30%	-0.23%	-0.05%	-0.24%
	Average	**-0.53%**	**-0.23%**	**-0.09%**	**-0.26%**	**-0.13%**	**-0.21%**
ClassD (416x240)	*BasketballPass*	-0.49%	-0.16%	-1.84%	-0.25%	-1.36%	-0.37%
	BlowingBubbles	-0.40%	-1.14%	-1.03%	-0.33%	-0.78%	-0.54%
	BQSquare	-0.49%	-1.13%	0.20%	-0.28%	-0.28%	-0.32%
	RaceHorses	-0.32%	1.79%	-1.96%	-0.40%	0.01%	-0.61%
	Average	**-0.43%**	**-0.16%**	**-1.16%**	**-0.32%**	**-0.60%**	**-0.46%**
ClassE (1280x720)	*FourPeople*	-0.50%	-0.45%	0.64%	—	—	—
	Johnny	-0.51%	-0.53%	-0.44%	—	—	—
	KristenAndSara	-0.59%	-1.65%	-0.95%	—	—	—
	Average	**-0.53%**	**-0.88%**	**-0.25%**	—	—	—
Overall		**-0.45%**	**-0.39%**	**-0.34%**	**-0.23%**	**-0.28%**	**-0.05%**

TABLE II
TIME COMPLEXITY OF PROPOSED MM-LIC COMPARED TO ECM-8.0 ANCHOR UNDER LDB AND RA CONFIGURATIONS ON CTC TEST SEQUENCES

Class	Low-delay B (%)		Random Access (%)	
	EncT	DecT	EncT	DecT
ClassA	—	—	106%	109%
ClassB	107%	112%	104%	106%
ClassC	103%	105%	103%	103%
ClassD	100%	99%	100%	101%
ClassE	104%	103%	—	—
Overall	**104%**	**105%**	**103%**	**105%**

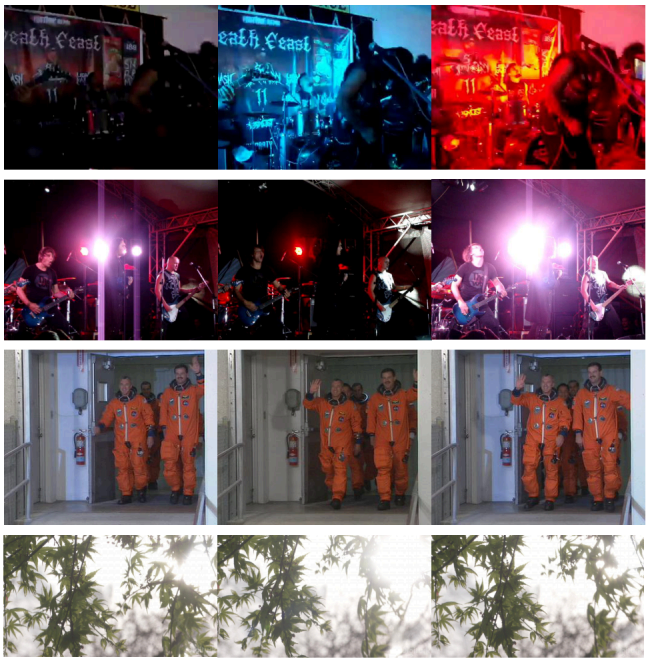


Fig. 8. Test sequences with specific scenarios, which contain lots of sharp and complex illumination variations. Each row represents three selected frames of each test sequence. From top to bottom: Frame 5-11-17 in LiveMusic-6a65 [41], Frame 11-16-21 in LiveMusic-2019 [41], Frame 2-11-20 in Crew [42], Frame 8-18-28 in SunBath [43].

show that the proposed MM-LIC framework performs well for sequences across different resolutions, which demonstrates the strong generalization capability of MM-LIC under both LDB and RA configurations. In addition, the experimental results show that the performance of MM-LIC is better under LDB configuration than that under RA configuration, which demonstrates a more significant role of MM-LIC under LDB configuration. The reason is that, although illumination variations are generally more pronounced in RA due to its hierarchical referencing with larger temporal distances, LIC is essentially an inter-prediction technique to refine the prediction retrieved by motion vectors and its effectiveness depends heavily on the quality of ME. From this perspective, LDB is generally more favorable than RA, as its reference frames are temporally closer to the current frame and therefore tend to be easier to estimate and better aligned with the current block than those under RA's hierarchical referencing structure.

2) Time Complexity: The encoding/decoding time complexity of the proposed MM-LIC is illustrated in Table II. EncT and DecT represent the encoding and decoding time ratio of the MM-LIC compared with the ECM-8.0 anchor. The codec time is tested on CPU, and the model is Intel(R) Xeon(R) CPU E5-2690 v4 @ 2.60GHz.

For the time complexity, the EncT/DecT is 104%/105% and 103%/105% under LDB and RA configurations, indicating only a slight runtime overhead compared to the anchor for both encoding and decoding. The increase mainly comes from the variance computation and multi-candidate evaluation in MMS and MMT at both encoder and decoder side. Specifically, MMS computes lightweight block statistics to adaptively select a more stable solver, and MMT evaluates multiple candidate model types on compact templates and selects the best one based on the template cost. Since the decisions are made implicitly to avoid additional signaling, the required on-the-fly decision and evaluation introduce extra computations.

3) Evaluation on Specific Scenarios: To further verify the effectiveness of MM-LIC, we test MM-LIC on some test sequences with specific scenarios, which contain lots of sharp and complex illumination variations. The test sequences mainly consist of live concert scenarios, camera flashes and natural sunshine scenarios from [41]–[43], a total of 12 test sequences with different resolutions across from 352x288 to 3840x2160. Some frames of the test sequences are shown in Fig. 8. Table III shows the performance of MM-LIC on these test sequences with specific scenarios. We can observe that the proposed MM-LIC can achieve, on average, 0.84% and 0.50%, and achieve up to 1.36% and 1.07% BD-rate reduction (Y component) under LDB and RA configurations. The MM-

TABLE III
BD-RATE RESULTS OF PROPOSED MM-LIC COMPARED TO ECM-8.0 ANCHOR UNDER LDB AND RA CONFIGURATIONS ON SPECIFIC SCENARIOS

Sequence	Resolution	Low-delay B (%)			Random Access (%)		
		Y	U	V	Y	U	V
Crew	*352x288*	-0.78%	1.48%	0.78%	-0.37%	-2.66%	0.04%
Foreman	*352x288*	-0.57%	-0.66%	-0.62%	-0.29%	0.95%	-1.67%
LiveMusic-265c	*480x360*	-1.19%	-1.16%	0.68%	-0.41%	0.97%	0.52%
LiveMusic-6a65	*480x360*	-0.80%	0.36%	1.95%	-1.07%	-0.06%	1.06%
LiveMusic-2b0f	*640x360*	-1.36%	-1.02%	-1.76%	-0.72%	-1.46%	-0.54%
LiveMusic-2019	*640x480*	-0.66%	-1.09%	-0.41%	-0.73%	-0.94%	-0.22%
LiveMusic-58fb	*640x480*	-0.81%	-0.08%	-0.73%	-0.37%	-1.23%	-0.82%
LiveMusic-58db	*1280x720*	-0.49%	-0.61%	-0.13%	-0.53%	-1.06%	0.20%
LiveMusic-6343	*1280x720*	-0.72%	-0.37%	0.43%	-0.38%	0.00%	0.93%
LiveMusic-14af	*1920x1080*	-0.92%	-0.71%	-0.84%	-0.25%	-0.18%	-0.29%
LiveMusic-28fe	*1920x1080*	-1.00%	-1.75%	-0.26%	-0.61%	-0.32%	-1.65%
SunBath	*3840x2160*	-0.75%	-0.32%	-0.99%	-0.31%	0.68%	0.46%
Overall		**-0.84%**	**-0.49%**	**-0.16%**	**-0.50%**	**-0.44%**	**-0.16%**

TABLE IV
BD-RATE RESULTS OF PROPOSED MM-LIC COMPARED TO ECM-19.0 ANCHOR UNDER LDB AND RA CONFIGURATIONS ON CTC TEST SEQUENCES

Class	Configuration	Low-delay B (%)			Random Access (%)		
	Sequence	Y	U	V	Y	U	V
ClassA1 (3840x2160)	*Campfire*	—	—	—	-0.12%	-0.09%	-0.33%
	FoodMarket4	—	—	—	-0.13%	0.29%	0.46%
	Tango2	—	—	—	-0.05%	0.15%	-0.13%
	Average	—	—	—	**-0.10%**	**0.12%**	**0.00%**
ClassA2 (3840x2160)	*CatRobot*	—	—	—	-0.22%	-0.35%	0.56%
	DaylightRoad2	—	—	—	-0.15%	-1.12%	-0.04%
	ParkRunning3	—	—	—	-0.10%	0.03%	-0.02%
	Average	—	—	—	**-0.16%**	**-0.48%**	**0.16%**
ClassB (1920x1080)	*BasketballDrive*	-0.26%	-0.81%	-0.13%	-0.19%	0.94%	-0.30%
	BQTerrace	-0.35%	0.60%	1.03%	-0.18%	-0.26%	-0.16%
	Cactus	-0.35%	0.02%	-0.38%	-0.16%	0.11%	-0.45%
	MarketPlace	-0.16%	-1.38%	-0.46%	-0.19%	0.81%	-0.43%
	RitualDance	-0.23%	-0.64%	0.14%	-0.17%	0.25%	-0.70%
	Average	**-0.27%**	**-0.44%**	**0.04%**	**-0.18%**	**0.37%**	**-0.41%**
ClassC (840x480)	*BasketballDrill*	-0.52%	-0.47%	-1.36%	-0.23%	-0.06%	-0.99%
	BQMall	-0.46%	-1.44%	1.22%	-0.30%	0.01%	0.12%
	PartyScene	-0.25%	-0.11%	-0.78%	-0.19%	-0.23%	-0.31%
	RaceHorses	-0.11%	-0.71%	-0.16%	-0.18%	0.48%	0.41%
	Average	**-0.33%**	**-0.68%**	**-0.27%**	**-0.23%**	**0.05%**	**-0.19%**
ClassD (416x240)	*BasketballPass*	-0.60%	-1.13%	-0.73%	-0.42%	-2.08%	0.60%
	BlowingBubbles	-0.31%	-1.52%	-0.33%	-0.18%	-1.26%	-0.53%
	BQSquare	-0.37%	-0.34%	-0.19%	-0.22%	-0.81%	-0.35%
	RaceHorses	-0.50%	0.66%	-1.29%	-0.21%	-0.48%	1.22%
	Average	**-0.45%**	**-0.58%**	**-0.63%**	**-0.26%**	**-1.16%**	**0.23%**
ClassE (1280x720)	*FourPeople*	-0.25%	0.27%	0.34%	—	—	—
	Johnny	-0.58%	-1.07%	0.13%	—	—	—
	KristenAndSara	-0.59%	1.11%	-0.56%	—	—	—
	Average	**-0.47%**	**0.10%**	**-0.03%**	—	—	—
Overall		**-0.37%**	**-0.43%**	**-0.22%**	**-0.19%**	**-0.19%**	**-0.07%**
EncT		**103%**			**104%**		
DecT		**104%**			**104%**		

LIC performs better on specific scenarios with more complex illumination variations, which demonstrates its stronger capability to handle challenging illumination variations and improve prediction accuracy in such cases.

4) Evaluation on Latest ECM Version: To further verify the effectiveness of the proposed scheme in the latest reference software, under the same experimental settings mentioned in Section V.A, MM-LIC is re-implemented in ECM-19.0 [44] and the corresponding performance is shown in Table IV. We can observe that the proposed MM-LIC can achieve, on average, 0.37% and 0.19% BD-rate reduction (Y component) under LDB and RA configurations, respectively, with only a slight increase in both encoding and decoding time complexity. Compared with the results of MM-LIC integrated into ECM-8.0, the BD-rate reduction achieved by MM-LIC in ECM-19.0 is degraded. This is mainly attributed to the integration of several advanced LIC-related tools in ECM-19.0, including NLIC [25], LIC with BDOF/BDMVR [26], LIC with slope adjustment [27], bi-predictive LIC [28], etc. These tools have already improved the prediction accuracy under illumination variations, thereby reducing the improvement potential of the proposed MM-LIC. In addition, these tools share a similar optimization objective with MM-LIC for prediction refinement under photometric inconsistencies, thereby resulting in a certain degree of performance overlap.

C. Ablation Study

1) Core Modules: To demonstrate the contributions of two core modules in MM-LIC, we conduct the ablation experiments on the proposed MMT and MMS.

First, we validate the effectiveness of MMT and the influence of the number of optional model types within MMT. In addition to the default MMT which includes LM, NLM, and GLM, we introduce two variants for comparison with MMT. For 2M-MMT (LM+NLM), only LM and NLM are optional model types within MMT. For 2M-MMT (LM+GLM), only LM and GLM are optional model types within MMT. As shown in Table V, we can observe that MMT, 2M-MMT (LM+NLM), 2M-MMT (LM+GLM) can achieve, on average, 0.32% and 0.17%, 0.19% and 0.10%, 0.18% and 0.11% BD-rate reduction (Y component) under LDB and RA configurations, respectively. Based on the experimental results, we draw three conclusions: (1) For the comparison of 2M-MMT and anchor, the experimental results show that incorporating NLM/GLM as an additional optional model type is superior to using LM alone. This demonstrates that incorporating NLM/GLM as an additional optional model type is beneficial to handle diverse illumination variations and improve the prediction accuracy. (2) For the comparison of MMT and anchor, the experimental results show that introducing both NLM and GLM as additional optional model types is superior to using LM alone. This demonstrates that incorporating both NLM and GLM as additional optional model types is beneficial to handle diverse illumination variations and improve the prediction accuracy. (3) For the comparison of MMT and 2M-MMT, the experimental results show that incorporating both NLM and GLM as additional optional model types is superior to incorporating only one of them. This demonstrates that the more optional model types within MMT, the more accurate the prediction becomes.

Second, we validate the effectiveness of MMS. As shown in Table V, we can observe that MMS can achieve, on average, 0.20% and 0.11% BD-rate reduction (Y component) under LDB and RA configurations, respectively. For the comparison of MMS and anchor, the experimental results show that incorporating RR as an additional optional model solution is superior to using LSR alone. This demonstrates that incorporating RR as an additional optional model solution is

TABLE V
ABLATION STUDY OF CORE MODULES ON ECM-8.0 UNDER LDB AND RA CONFIGURATIONS

Low-delay B (%)												
Class	MMS			MMT			2M-MMT (LM+NLM)			2M-MMT (LM+GLM)		
	Y	U	V	Y	U	V	Y	U	V	Y	U	V
ClassB	-0.20%	-0.22%	-0.25%	-0.23%	-0.06%	0.15%	-0.20%	-0.11%	-0.19%	-0.19%	0.22%	0.02%
ClassC	-0.24%	-0.40%	0.03%	-0.40%	-0.39%	0.04%	-0.24%	-0.19%	-0.09%	-0.22%	-0.09%	0.10%
ClassD	-0.13%	-0.05%	-0.32%	-0.31%	0.42%	-0.15%	-0.05%	-0.18%	-0.93%	-0.15%	-0.16%	-0.63%
ClassE	-0.22%	-0.17%	0.02%	-0.39%	-0.35%	-0.31%	-0.29%	0.09%	0.62%	-0.13%	-0.14%	0.49%
Overall	**-0.20%**	**-0.21%**	**-0.15%**	**-0.32%**	**-0.08%**	**-0.04%**	**-0.19%**	**-0.11%**	**-0.20%**	**-0.18%**	**-0.02%**	**-0.03%**

Random Access (%)												
Class	MMS			MMT			2M-MMT(LM+NLM)			2M-MMT(LM+GLM)		
	Y	U	V	Y	U	V	Y	U	V	Y	U	V
ClassA	-0.09%	-0.16%	0.18%	-0.19%	0.17%	0.36%	-0.14%	0.09%	0.27%	-0.10%	0.10%	0.32%
ClassB	-0.11%	0.12%	0.06%	-0.14%	-0.01%	0.16%	-0.05%	-0.13%	0.05%	-0.07%	0.03%	-0.18%
ClassC	-0.10%	-0.12%	-0.34%	-0.15%	-0.03%	-0.40%	-0.14%	-0.26%	-0.05%	-0.11%	-0.09%	0.02%
ClassD	-0.15%	-0.21%	-0.44%	-0.18%	-0.42%	-0.40%	-0.13%	-0.15%	-0.72%	-0.15%	-0.16%	-0.30%
Overall	**-0.11%**	**-0.09%**	**-0.09%**	**-0.17%**	**-0.04%**	**-0.01%**	**-0.11%**	**-0.09%**	**-0.06%**	**-0.11%**	**-0.01%**	**-0.01%**

TABLE VI
ABLATION STUDY OF NON-LINEAR MODEL ORDERS ON ECM-8.0 UNDER LDB AND RA CONFIGURATIONS

Low-delay B (%)												
Class	2M-MMT			3M-MMT			4M-MMT			5M-MMT		
	Y	U	V	Y	U	V	Y	U	V	Y	U	V
ClassB	-0.20%	-0.11%	-0.19%	-0.19%	0.36%	-0.08%	-0.21%	-0.11%	-0.12%	-0.20%	0.38%	-0.15%
ClassC	-0.24%	-0.19%	-0.09%	-0.33%	-0.50%	-0.07%	-0.33%	-0.08%	0.05%	-0.36%	-0.01%	0.08%
ClassD	-0.05%	-0.18%	-0.93%	-0.10%	-0.17%	-0.27%	-0.21%	-0.11%	-0.94%	-0.10%	-0.42%	0.08%
ClassE	-0.29%	0.09%	0.62%	-0.34%	-0.33%	0.49%	-0.28%	-0.46%	0.45%	-0.36%	-0.21%	0.55%
Overall	**-0.19%**	**-0.11%**	**-0.20%**	**-0.23%**	**-0.12%**	**-0.02%**	**-0.25%**	**-0.17%**	**-0.17%**	**-0.25%**	**-0.03%**	**0.09%**
EncT		**102%**			**105%**			**107%**			**109%**	
DecT		**103%**			**106%**			**109%**			**112%**	

Random Access (%)												
Class	2M-MMT			3M-MMT			4M-MMT			5M-MMT		
	Y	U	V	Y	U	V	Y	U	V	Y	U	V
ClassA	-0.14%	0.09%	0.27%	-0.09%	0.49%	0.60%	-0.09%	0.08%	0.76%	-0.05%	-0.04%	0.81%
ClassB	-0.05%	-0.13%	0.05%	-0.11%	0.00%	0.12%	-0.10%	0.02%	0.06%	-0.16%	-0.09%	0.08%
ClassC	-0.14%	-0.26%	-0.05%	-0.17%	-0.39%	-0.19%	-0.16%	-0.33%	-0.18%	-0.17%	-0.14%	-0.37%
ClassD	-0.13%	-0.15%	-0.72%	-0.19%	-0.41%	0.03%	-0.19%	-0.30%	-0.20%	-0.16%	0.00%	-0.13%
Overall	**-0.11%**	**-0.09%**	**-0.06%**	**-0.13%**	**-0.02%**	**0.19%**	**-0.13%**	**-0.10%**	**0.17%**	**-0.13%**	**-0.07%**	**0.17%**
EncT		**102%**			**105%**			**108%**			**110%**	
DecT		**104%**			**105%**			**109%**			**112%**	

beneficial to improve the stability of the model solution and the prediction accuracy.

2) Model Order: For the 2M-MMT (LM+NLM), in order to investigate whether introducing more higher-order models yields additional coding gains, we conduct the ablation experiments by progressively expanding the set of non-linear model candidates with MMT enabled alone. Specifically, in addition to the 2M-MMT which includes only the linear model and the second-order non-linear model, we introduce three additional variants: 3M-MMT (adding up to third-order), 4M-MMT (adding up to fourth-order), and 5M-MMT (adding up to fifth-order). The expression for the higher-order models is given in Eq. (6) and the corresponding performance is shown in Table VI. We can observe that as the number of higher-order models increases, the BD-rate reduction (Y component) achieved by 3M-MMT, 4M-MMT, and 5M-MMT reaches, on average, 0.23% and 0.13%, 0.25% and 0.13%, 0.25% and 0.13% under LDB and RA configurations, respectively. The experimental results show that introducing additional higher-order models yields further coding gains compared with the 2M-MMT, but the improvements gradually diminish as the number of higher-order models increases. In particular, the R-D performance remains largely unchanged from 4M-MMT to 5M-MMT. The reason is that, although higher-order models can better fit complex mapping relationships, the marginal benefit becomes negligible beyond a certain order. At the same time, including more higher-order models significantly increases encoding and decoding time complexity due to the additional parameter derivations and template cost computations in the implicit decision-making mechanism of MMT. Therefore, in order to achieve the best trade-off between coding efficiency and computational complexity, only the second-order non-linear model is adopted as the additional model type within the NLM component of MMT.

D. Robustness Analysis

1) Threshold in Model Solution Selection: To investigate the robustness of the content adaptive-based model solution

TABLE VII
ROBUSTNESS ANALYSIS TO THRESHOLD IN MODEL SOLUTION SELECTION ON ECM-8.0 UNDER LDB AND RA CONFIGURATIONS

Low-delay B (%)												
Class	**TH1**			**TH2**			**TH3**			**TH4**		
	Y	U	V	Y	U	V	Y	U	V	Y	U	V
ClassB	-0.19%	-0.33%	-0.25%	-0.21%	0.14%	0.02%	-0.21%	0.23%	-0.24%	-0.12%	-0.09%	0.23%
ClassC	-0.23%	-0.03%	0.64%	-0.32%	-0.23%	0.54%	-0.31%	-0.35%	-0.23%	-0.21%	-0.10%	0.23%
ClassD	-0.06%	0.70%	0.00%	-0.16%	0.00%	-0.96%	-0.18%	0.02%	-0.15%	-0.04%	0.22%	0.26%
ClassE	-0.26%	-0.37%	0.37%	-0.38%	-0.46%	-0.27%	-0.35%	-0.46%	0.27%	-0.08%	0.23%	0.37%
Overall	**-0.18%**	**-0.01%**	**0.15%**	**-0.26%**	**-0.10%**	**-0.15%**	**-0.25%**	**-0.10%**	**-0.12%**	**-0.12%**	**0.04%**	**0.26%**

Random Access (%)												
Class	**TH1**			**TH2**			**TH3**			**TH4**		
	Y	U	V	Y	U	V	Y	U	V	Y	U	V
ClassA	-0.12%	-0.07%	0.42%	-0.15%	-0.07%	0.20%	-0.16%	0.24%	0.45%	-0.10%	0.08%	0.38%
ClassB	-0.11%	0.12%	0.06%	-0.13%	0.20%	0.08%	-0.14%	0.09%	0.16%	-0.06%	-0.08%	0.16%
ClassC	-0.06%	-0.02%	-0.13%	-0.13%	-0.08%	-0.07%	-0.18%	-0.07%	-0.04%	-0.10%	-0.03%	-0.08%
ClassD	-0.12%	0.10%	0.48%	-0.13%	-0.26%	-0.29%	-0.15%	-0.13%	0.40%	-0.03%	-0.28%	-0.23%
Overall	**-0.10%**	**0.03%**	**0.22%**	**-0.14%**	**-0.04%**	**0.01%**	**-0.16%**	**0.06%**	**0.26%**	**-0.07%**	**-0.06%**	**0.10%**

TABLE VIII
ROBUSTNESS ANALYSIS TO TEMPLATE DEFINITION ON ECM-8.0 UNDER LDB AND RA CONFIGURATIONS

Low-delay B (%)												
Class	**Only Left Template**						**Only Above Template**					
	LIC			**MM-LIC**			**LIC**			**MM-LIC**		
	Y	U	V	Y	U	V	Y	U	V	Y	U	V
ClassB	0.08%	-0.09%	-0.43%	0.00%	0.21%	0.08%	0.17%	0.01%	0.07%	0.05%	0.14%	0.12%
ClassC	0.09%	-0.03%	-0.12%	-0.02%	0.37%	-0.02%	0.10%	0.46%	0.27%	-0.01%	0.69%	-0.21%
ClassD	0.22%	0.03%	0.03%	-0.06%	0.24%	0.28%	0.17%	0.61%	-0.47%	-0.05%	-0.13%	0.23%
ClassE	0.17%	0.39%	0.27%	0.04%	0.09%	-0.09%	0.13%	-0.06%	-0.22%	0.00%	-0.02%	-0.29%
Overall	**0.13%**	**0.04%**	**-0.11%**	**-0.01%**	**0.23%**	**0.07%**	**0.15%**	**0.26%**	**-0.07%**	**0.00%**	**0.18%**	**-0.01%**

Random Access (%)												
Class	**Only Left Template**						**Only Above Template**					
	LIC			**MM-LIC**			**LIC**			**MM-LIC**		
	Y	U	V	Y	U	V	Y	U	V	Y	U	V
ClassA	0.06%	0.13%	0.11%	0.02%	-0.04%	-0.14%	0.07%	-0.26%	0.12%	0.03%	-0.05%	-0.09%
ClassB	0.10%	0.06%	-0.14%	0.00%	-0.04%	0.07%	0.08%	-0.04%	-0.26%	-0.02%	-0.37%	-0.12%
ClassC	0.07%	-0.47%	0.21%	0.02%	-0.03%	0.18%	0.11%	-0.24%	-0.70%	0.00%	-0.14%	0.29%
ClassD	0.20%	0.01%	0.14%	0.07%	0.45%	0.22%	0.17%	-0.17%	-0.64%	0.06%	0.12%	-0.15%
Overall	**0.10%**	**-0.04%**	**0.07%**	**0.03%**	**0.06%**	**0.06%**	**0.10%**	**-0.18%**	**-0.31%**	**0.02%**	**-0.12%**	**-0.03%**

selection to different preset thresholds, we conduct experiments with MMS enabled alone and evaluate thresholds within various ranges: $TH1 \in [0, 50^2)$, $TH2 \in [50^2, 100^2)$, $TH3 \in [100^2, 150^2)$, and $TH4 \in [150^2, 200^2]$. The corresponding R–D performance of MMS under different thresholds is summarized in Table VII. We can observe that as the preset threshold increases, the BD-rate reduction (Y component) under $TH1$, $TH2$, $TH3$, and $TH4$ reaches, on average, 0.18% and 0.10%, 0.26% and 0.14%, 0.25% and 0.16%, 0.12% and 0.07% under LDB and RA configurations, respectively. More specifically, the R–D performance is relatively limited at an excessively small threshold ($TH1$), improves noticeably and remains stable within a moderate range ($TH2$ and $TH3$), and then degrades at an overly large threshold ($TH4$). This behavior is consistent with the underlying solver selection mechanism. Specifically, extreme thresholds tend to cause near-exclusive selection of a single solver, i.e., ridge regression at a small threshold and least squares regression at a large threshold. In contrast, thresholds within $[50^2, 150^2)$ allow a more balanced and content adaptive use of both solvers, thereby better exploiting their complementary strengths. The experimental results show that the preset threshold mainly serves as a switching cue between the two solvers, rather than a sensitive hyperparameter requiring careful tuning. As long as it lies within a reasonable range, the proposed content adaptive-based model solution selection remains stable and effective.

2) Template Definition: To investigate the robustness of the proposed MM-LIC to different template definitions, we conduct experiments with several template variants. Specifically, in addition to the default template definition that jointly uses both the above and left reconstructions, we further consider two alternative variants: using only the left reconstruction template and using only the above reconstruction template. Compared with the default definition, these variants provide less reference information and therefore serve as test cases. The corresponding R–D performance of MM-LIC under different template definitions is summarized in Table VIII. We can observe that, under these alternative template variants, the original LIC suffers noticeable performance degradation relative to the anchor (i.e., the original LIC with the default template definition), as expected due to the reduced reference information. In contrast, under the same template variants,

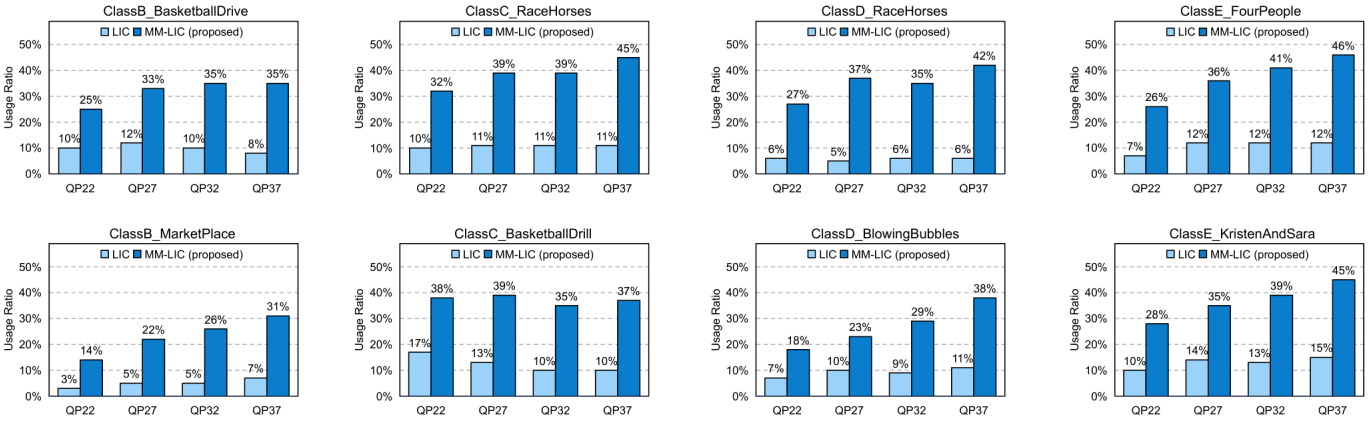


Fig. 9. Usage ratio of LIC and MM-LIC for sequences with different resolutions on ECM-8.0 (Configuration: LDB). The left column in each QP represents the usage ratio of LIC; the right column in each QP represents the usage ratio of the proposed MM-LIC.

(a) LIC (RitualDance, QP: 37, POC: 8) (b) MM-LIC (RitualDance, QP: 37, POC: 8)

Fig. 10. Mode selected results on ECM-8.0 (Configuration: LDB). (a) Mode selected results of the LIC. (b) Mode selected results of the proposed MM-LIC.

MM-LIC achieves negligible BD-rate loss or even slight gains relative to the anchor. More specifically, when only the left reconstruction is used as the template, the average BD-rate change (Y component) under the LDB and RA configurations is 0.13%/0.10% for LIC and -0.01%/0.03% for MM-LIC, respectively. When only the above reconstruction is used as the template, the corresponding BD-rate change (Y component) under the LDB and RA configurations is 0.15%/0.10% for LIC and 0.00%/0.02% for MM-LIC, respectively. The experimental results show that the effectiveness of the proposed MM-LIC does not depend on a specific template definition. Moreover, the consistently improved performance of MM-LIC under different template definitions demonstrates its robustness and general applicability.

E. Visual Analysis

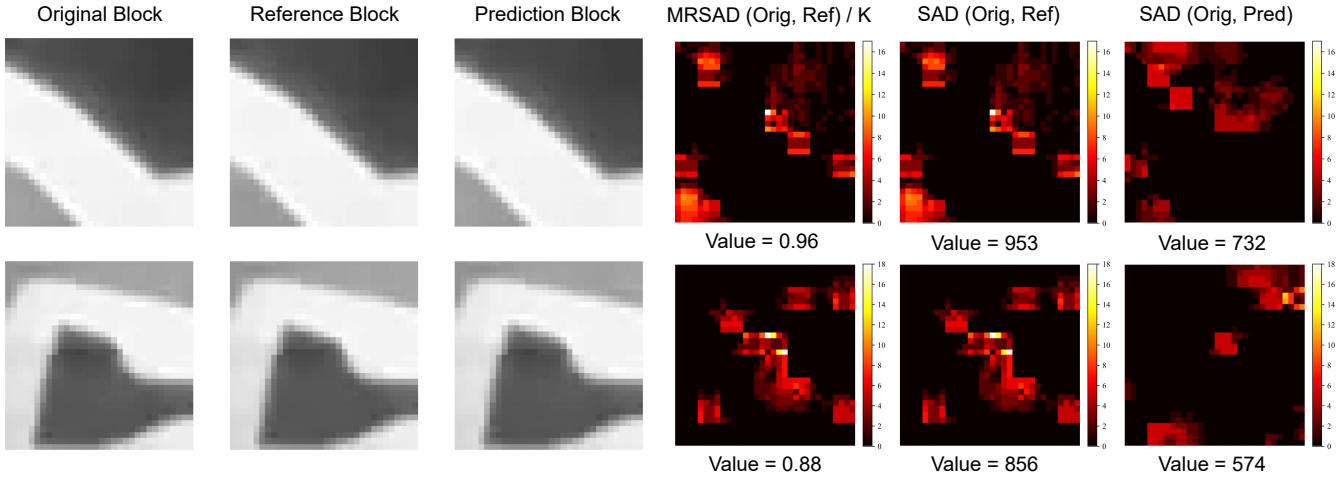
To observe the percentage of the original LIC and the proposed MM-LIC being selected, the usage ratio ($Ratio$) is used to validate the efficiency of mode selection, which is calculated by $Ratio = N/N_{total}$, where N indicates the number of coding units coded by the LIC or the proposed MM-LIC in B-frames of the test sequence, and N_{total} indicates the total number of coding units in B-frames of the test sequence. The $Ratio$ of LIC and MM-LIC in sequences with different resolutions under different QPs is shown in Fig. 9. We observe that the usage ratio of MM-LIC is significantly higher than that of LIC in sequences across different resolutions under different QPs. This demonstrates that, after introducing the proposed MM-LIC, a larger proportion of coding blocks can benefit from illumination compensation, leading to improved prediction quality.

In addition, Fig. 10 shows the mode selected results of the LIC and our proposed MM-LIC. Compared with LIC, MM-LIC possesses more flexible illumination variations description capability, which focuses on more diverse illumination variations with sharp and complex textures and content, such as the diverse illumination variations of the natural light in the test sequence “RitualDance”.

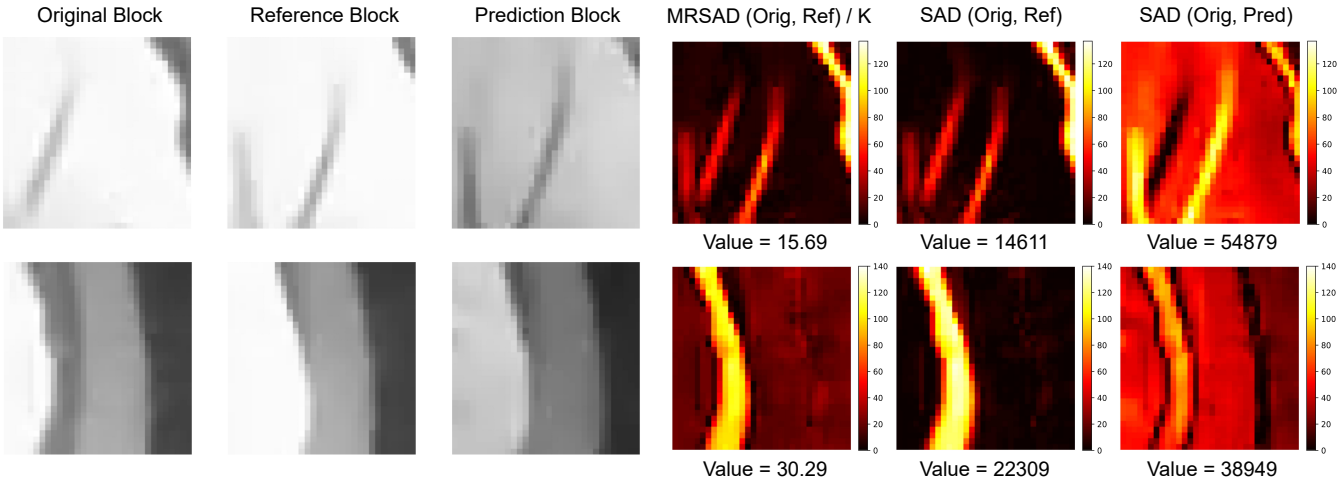
F. Limitation

Based on the above experiments, since MM-LIC refines the motion-compensated reference retrieved by motion vectors, its effectiveness is expected to depend on how accurately the reference block matches the target coding content, i.e., ME accuracy. When ME is accurate, MM-LIC produces more faithful refinement, whereas under inaccurate ME, the reference mismatch may amplify the prediction error and weaken the effect of MM-LIC, resulting in limited improvement in some cases.

To quantitatively analyze this limitation, we perform a statistical analysis on the blocks where MM-LIC is applied. First, for each block, the original block ($Orig$), the reference



(a) Visualization of MM-LIC prediction performance for some accurate motion estimation blocks in G_0.



(b) Visualization of MM-LIC prediction performance for some inaccurate motion estimation blocks in G_1.

Fig. 11. Limitation of MM-LIC from the perspective of motion estimation accuracy. (a) Prediction performance of MM-LIC in accurate motion estimation group G_0. (b) Prediction performance of MM-LIC in accurate motion estimation group G_1.

block before MM-LIC adjustment (Ref), and the prediction block after MM-LIC adjustment ($Pred$) are extracted to form data pairs. Second, the data pairs are categorized into two groups (G_0 and G_1) based on ME accuracy. Specifically, the mean-removed sum of absolute differences (MRSAD) between $Orig$ and Ref is used as the criterion to estimate ME accuracy, as it suppresses illumination differences and better reflects structural and motion alignment errors compared with the SAD. The MRSAD is calculated as:

$$\begin{aligned}&\text{MRSAD}\,(Orig, Ref)\\&= \sum_{i=1}^{K} \left|(Orig(i) - Orig_{\text{avg}}) - (Ref(i) - Ref_{\text{avg}})\right|,\end{aligned} \tag{31}$$

where K is the total number of samples in the block, $Orig(i)$ and $Ref(i)$ are the i^{th} samples of the original block and the reference block, respectively, and $Orig_{\text{avg}}$ and Ref_{avg} are the average values of the original block and the reference block, respectively. If $\text{MRSAD}\,(Orig, Ref)/K \leq 1$,the data pair is categorized into accurate-ME group (G_0); if $\text{MRSAD}\,(Orig, Ref)/K \geq 5$, the data pair is categorized into inaccurate-ME group (G_1). Third, within each group, the proportion of blocks where MM-LIC improves prediction accuracy is evaluated with SAD. When $\text{SAD}\,(Orig, Pred) < \text{SAD}\,(Orig, Ref)$, the corresponding data pair is considered to exhibit prediction improvement. The SAD is calculated as:

$$\begin{aligned}\text{SAD}\,(Orig, Ref) &= \sum_{i=1}^{K} |Orig(i) - Ref(i)|,\\ \text{SAD}\,(Orig, Pred) &= \sum_{i=1}^{K} |Orig(i) - Pred(i)|,\end{aligned} \tag{32}$$

where K is the total number of samples in the block, $Orig(i)$, $Ref(i)$, and $Pred(i)$ are the i^{th} samples of the original block, the reference block, and the prediction block, respectively. The results show that MM-LIC improves prediction accuracy for 86.53% of blocks in G_0, while the ratio drops to 65.67% in G_1. This indicates that MM-LIC is more effective when the reference is well aligned with the target block, but its benefit becomes limited under inaccurate ME. Fig. 11 shows typical examples in G_0 and G_1: when ME is accurate, MM-LIC produces more faithful refinement, whereas under inaccurate ME, the reference mismatch may amplify the prediction error and weaken the effect of MM-LIC, resulting in limited

improvement. This limitation suggests that future work can further explore tighter integration of MM-LIC with advanced motion vector refinement tools (e.g., decoder-side motion vector refinement (DMVR) and Bi-Directional Optical Flow (BDOF)) to achieve further coding gains.

VI. Conclusion

In this paper, a multi-model local illumination compensation framework is proposed to improve the prediction accuracy under diverse illumination variations in inter prediction. In MM-LIC, MMT and MMS are proposed to optimize LIC in terms of model type and model solution. Meanwhile, in order to preserve decision accuracy without extra bit overhead, separate implicit decision mechanisms are designed for MMT and MMS. We have conducted extensive experiments to verify the effectiveness of MM-LIC. Experimental results show that the proposed method achieves, on average, 0.45% and 0.23%, and achieves up to 0.68% and 0.40% BD-rate reduction compared to ECM-8.0 anchor under LDB and RA configurations, respectively. The encoding/decoding time complexity is only slightly increased under both LDB and RA configurations. Moreover, the proposed method performs better on specific scenarios with complex and sharp illumination variations and also demonstrates its potential in the latest ECM version.

In future work, first, we will explore the combination of the MM-LIC and the other inter prediction tools, such as the DMVR [45]–[47], affine mode [48], [49], geometric partitioning mode (GPM) [50], [51], and so on. Second, we will explore more flexible decision mechanisms in MM-LIC to further reduce the encoding and decoding time complexity.

References

[1] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649–1668, 2012.

[2] B. Bross, Y.-K. Wang, Y. Ye, S. Liu, J. Chen, G. J. Sullivan, and J.-R. Ohm, "Overview of the versatile video coding (VVC) standard and its applications," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 10, pp. 3736–3764, 2021.

[3] Y.-Q. Zhang and S. Zafar, "Predictive block-matching motion estimation for tv coding. ii. inter-frame prediction," *IEEE Transactions on Broadcasting*, vol. 37, no. 3, pp. 102–105, 2005.

[4] J.-Y. Lee, J.-K. Han, J.-G. Kim, and T. Q. Nguyen, "Efficient inter-view motion vector prediction in multi-view hevc," *IEEE Transactions on Broadcasting*, vol. 64, no. 3, pp. 666–680, 2017.

[5] W.-J. Chien, L. Zhang, M. Winken, X. Li, R.-L. Liao, H. Gao, C.-W. Hsu, H. Liu, and C.-C. Chen, "Motion vector coding and block merging in the versatile video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 10, pp. 3848–3861, 2021.

[6] H. Yang, H. Chen, J. Chen, S. Esenlik, S. Sethuraman, X. Xiu, E. Alshina, and J. Luo, "Subblock-based motion derivation and inter prediction refinement in the versatile video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 10, pp. 3862–3877, 2021.

[7] K. Kamikura, H. Watanabe, H. Jozawa, H. Kotera, and S. Ichinose, "Global brightness-variation compensation for video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, no. 8, pp. 988–1000, 1998.

[8] D. Marpe, T. Wiegand, and G. Sullivan, "The H.264/MPEG4 advanced video coding standard and its applications," *IEEE Communications Magazine*, vol. 44, no. 8, pp. 134–143, 2006.

[9] J. M. Boyce, "Weighted prediction in the H.264/MPEG AVC video coding standard," in *2004 IEEE International Symposium on Circuits and Systems (IEEE Cat. No. 04CH37512)*, vol. 3. IEEE, 2004, pp. III–789.

[10] N. M. Rodrigues, V. M. M. da Silva, and S. M. de Faria, "Hierarchical motion compensation with spatial and luminance transformations," in *Proceedings 2001 International Conference on Image Processing (Cat. No. 01CH37205)*, vol. 3. IEEE, 2001, pp. 518–521.

[11] P. Yin, A. M. Tourapis, and J. Boyce, "Localized weighted prediction for video coding," in *2005 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2005, pp. 4365–4368.

[12] C.-W. Seo and J.-K. Han, "Pixel based illumination compensation for inter prediction in HEVC," *Electronics letters*, vol. 47, no. 23, pp. 1278–1280, 2011.

[13] H. Liu, J. Jung, J. Sung, J. Jia, and S. Yea, "3D-CE2. h: Results of illumination compensation for inter-view prediction," *ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11 JCT3V-B0045*, 2012.

[14] K. Müller, H. Schwarz, D. Marpe, C. Bartnik, S. Bosse, H. Brust, T. Hinz, H. Lakshman, P. Merkle, F. H. Rhee *et al.*, "3D high-efficiency video coding for multi-view video and depth data," *IEEE Transactions on Image Processing*, vol. 22, no. 9, pp. 3366–3378, 2013.

[15] G. Tech, Y. Chen, K. Müller, J.-R. Ohm, A. Vetro, and Y.-K. Wang, "Overview of the multiview and 3D extensions of high efficiency video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 1, pp. 35–49, 2015.

[16] H. Liu, Y. Chen, J. Chen, L. Zhang, and M. Karczewicz, "Local illumination compensation," *ITU-T SG16/Q6 Doc. VCEG-AZ06*, 2015.

[17] Y. Ye, M. Karczewicz, Y. Huang *et al.*, "AHG12: on the status of the ECM software," *doucment JVET-W0049, ITU-T/ISO/IEC Joint Video Experts Team (JVET)*, 2021.

[18] C. Guo, C. Li, J. Guo, C. C. Loy, J. Hou, S. Kwong, and R. Cong, "Zero-reference deep curve estimation for low-light image enhancement," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 1780–1789.

[19] C. Li, C. Guo, and C. C. Loy, "Learning to enhance low-light image via zero-reference deep curve estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 8, pp. 4225–4238, 2021.

[20] C. Li, C. Guo, R. Feng, S. Zhou, and C. C. Loy, "Cudi: Curve distillation for efficient and controllable exposure adjustment," *arXiv preprint arXiv:2207.14273*, 2022.

[21] C.-W. Kuo, X. Li, X. Xiu, H.-J. Jhu, N. Yan, X. Wang, Y. Ye, J. Chen, and R.-L. Liao, "Gradient linear model for chroma intra prediction," in *2023 Data Compression Conference (DCC)*. IEEE, 2023, pp. 13–21.

[22] A. E. Hoerl and R. W. Kennard, "Ridge regression: applications to nonorthogonal problems," *Technometrics*, vol. 12, no. 1, pp. 69–82, 1970.

[23] D. W. Marquardt and R. D. Snee, "Ridge regression in practice," *The American Statistician*, vol. 29, no. 1, pp. 3–20, 1975.

[24] M. Coban, F. Léannec, K. Naser, J. Ström, and L. Zhang, "Algorithm description of Enhanced Compression Model 2 (ECM 2)," *JVET-Y2025-v2, Joint Video Experts Team (JVET) of ITU-T SG*, vol. 16, pp. 7–16, 2022.

[25] X. Xiu, C. Ma, N. Yan, H.-J. Jhu, C.-W. Kuo, W. Chen, and X. Wang, "Non-EE2: Enhancements on local illumination compensation," *JVET-AF0191*, October 2023.

[26] Y. Zhang, V. Seregin, H. Wang, Z. Zhang, C.-C. Chen, H. Huang, and M. Karczewicz, "Non-EE2: On LIC flag in merge mode," *JVET-AF0194*, October 2023.

[27] Y. Wang, K. Zhang, Y. He, H. Liu, and L. Zhang, "Non-EE2: Extension of local illumination compensation," *JVET-AF0200*, October 2023.

[28] X. Xiu, N. Yan, C. Ma, H.-J. Jhu, C.-W. Kuo, W. Chen, and X. Wang, "EE2-Test2.7: Improvements on local illumination compensation," *JVET-AD0213*, April 2023.

[29] N. Zhang, K. Zhang, H. Liu, Y. Wang, and L. Zhang, "Non-EE2: LIC flag derivation of merge candidates with template costs," *JVET-AE0109*, July 2023.

[30] C. Wang, X. Jin, H. Chen, H. Yang, K. Tong, and Y. Yang, "Piecewise linear model based local illumination compensation inter prediction for video coding," in *2022 Picture Coding Symposium (PCS)*. IEEE, 2022, pp. 385–389.

[31] Y. Wang, K. Zhang, and L. Zhang, "Local-aware intra block copy for video coding," in *2023 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2023, pp. 1–5.

[32] ——, "Adaptive linear model for intra block copy with local illumination compensation," in *2023 IEEE International Conference on Visual Communications and Image Processing (VCIP)*. IEEE, 2023, pp. 1–5.

[33] M. Karczewicz and Y. Ye, "Common test conditions and evaluation procedures for enhanced compression tool testing," *JVET-Y2017*, 2022.

[34] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 58, no. 1, pp. 267–288, 1996.

[35] C. Hans, "Bayesian lasso regression," *Biometrika*, vol. 96, no. 4, pp. 835–845, 2009.
[36] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 67, no. 2, pp. 301–320, 2005.
[37] ——, "Regression shrinkage and selection via the elastic net, with applications to microarrays," *JR Stat Soc Ser B*, vol. 67, pp. 301–20, 2003.
[38] J. M. Prewitt *et al.*, "Object enhancement and extraction," *Picture processing and Psychopictorics*, vol. 10, no. 1, pp. 15–19, 1970.
[39] https://vcgit.hhi.fraunhofer.de/ecm/ECM/-/tree/ECM-8.0.
[40] G. Bjontegaard, "Calculation of average psnr differences between rd-curves," *ITU-T SG16, Doc. VCEG-M33*, 2001.
[41] https://media.withyoutube.com.
[42] https://media.xiph.org/video/derf.
[43] https://ultravideo.fi.
[44] https://vcgit.hhi.fraunhofer.de/ecm/ECM/-/tree/ECM-19.0.
[45] H. Gao, X. Chen, S. Esenlik, J. Chen, and E. Steinbach, "Decoder-side motion vector refinement in VVC: Algorithm and hardware implementation considerations," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 8, pp. 3197–3211, 2020.
[46] H. Huang, Z. Zhang, V. Seregin, W.-J. Chien, C.-C. Chen, and M. Karczewicz, "Adaptive bilateral matching for decoder-side motion vector refinement in video coding," in *2022 Data Compression Conference (DCC)*. IEEE, 2022, pp. 01–07.
[47] Y. Jian, Y. Huang, Z. Lin, M. Lei, Y. Xue, L. Guo, and C. Zhou, "Asymmetric motion vector refinement for future video coding," in *2024 Data Compression Conference (DCC)*. IEEE, 2024, pp. 402–411.
[48] L. Li, H. Li, D. Liu, Z. Li, H. Yang, S. Lin, H. Chen, and F. Wu, "An efficient four-parameter affine motion model for video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 8, pp. 1934–1948, 2017.
[49] J. Chen, R.-L. Liao, Y. Ye, and X. Li, "Decoder-side affine model refinement for video coding beyond VVC," in *2023 Data Compression Conference (DCC)*. IEEE, 2023, pp. 248–257.
[50] Y. Kidani, H. Kato, K. Kawamura, and H. Watanabe, "Geometric partitioning mode with inter and intra prediction for beyond versatile video coding," *IEICE TRANSACTIONS on Information and Systems*, vol. 105, no. 10, pp. 1691–1703, 2022.
[51] Z. Li, Z. Yuan, L. Li, D. Liu, X. Tang, and F. Wu, "Object segmentation-assisted inter prediction for versatile video coding," *IEEE Transactions on Broadcasting*, 2024.

Jialin Li (Student Member, IEEE) received the B.S. degree in electronic information engineering from China University of Mining and Technology, Jiangsu, China, in 2023. He is currently pursuing the Ph.D. degree with the Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei, China. His current research interests include video coding and processing.

Zhuoyuan Li (Member, IEEE) received the B.S. degree in communication engineering from Southwest Jiaotong University, China, in 2020, and the Ph.D. degree in electronic engineering and information science from the University of Science and Technology of China (USTC), China, in 2025. He will join The Hong Kong Polytechnic University (PolyU) as a Postdoctoral Fellow in 2026.

His research interests include image and video coding. He has published over 10 papers in international journals and conferences. He won several technical challenges at ICIP 2024, MMSP 2024, and VCIP 2025. He received the PRCV 2025 Grand Challenge Outstanding Exploration Award and the NeurIPS 2025 Top Reviewer Award. He has submitted several technical proposals to ISO/IEC and AVS standardization activities.

Yao Li (Student Member, IEEE) received the B.S. degree in electronic engineering from Tianjin University, Tianjin, China, in 2022. He is currently pursuing the Ph.D. degree with the Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei, China. His current research interests include image/video processing and coding.

Dong Liu (Senior Member, IEEE) received the B.S. and Ph.D. degrees in electrical engineering from the University of Science and Technology of China (USTC), Hefei, China, in 2004 and 2009, respectively. He was a Member of Research Staff with Nokia Research Center, Beijing, China, from 2009 to 2012. He joined USTC as a faculty member in 2012 and became a Professor in 2020.

His research interests include image and video processing, coding, analysis, and data mining. He has authored or co-authored more than 200 papers in international journals and conferences. He has more than 30 granted patents. He has several technique proposals adopted by standardization groups. He received the 2009 IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY Best Paper Award, VCIP 2016 Best 10% Paper Award, and ISCAS 2022 Grand Challenge Top Creativity Paper Award. He and his students were winners of several technical challenges held in ICIP 2024, ISCAS 2023, ICCV 2019, ACM MM 2019, ACM MM 2018, ECCV 2018, CVPR 2018, and ICME 2016. He is a Senior Member of CCF and CSIG, an elected member of IVMSP-TC of IEEE SP Society, and an elected member of MSA-TC of IEEE CAS Society. He serves or had served as the Chair of IEEE 1857.11 Standard Working Subgroup (also known as Future Video Coding Study Group), an Associate Editor for IEEE TRANSACTIONS ON IMAGE PROCESSING, a Guest Editor for IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, an Organizing Committee member for ChinaMM 2024, VCIP 2022, ICME 2021, etc.

Li Li (Senior Member, IEEE) received the B.S. and Ph.D. degrees in electronic engineering from the University of Science and Technology of China (USTC), Hefei, Anhui, China, in 2011 and 2016, respectively. He was a visiting assistant professor at the University of Missouri-Kansas City from 2016 to 2020. He joined the department of electronic engineering and information science of USTC as a research fellow in 2020 and became a professor in 2022.

His research interests include image/video/point cloud coding and processing. He has authored or co-authored more than 80 papers in international journals and conferences. He has more than 20 granted patents. He has several technique proposals adopted by standardization groups. He received the Multimedia Rising Star 2023. He received the Best 10% Paper Award at the 2016 IEEE Visual Communications and Image Processing (VCIP) and the 2019 IEEE International Conference on Image Processing (ICIP). He serves as an associate editor for IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY and IEEE TRANSACTIONS ON MULTIMEDIA from 2024 to 2025.