KDDI Research

Lossless Video Coding Based on Probability Model Optimization Utilizing Example Search and Adaptive Prediction

Kyohei Unno[†‡], Koji Nemoto[‡], Yusuke Kameda[‡],
Ichiro Matsuda[‡], Susumu Itoh[‡], Sei Naito[†]

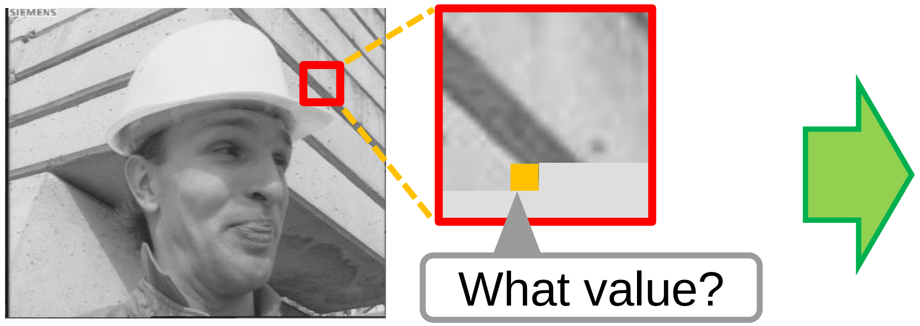
† KDDI Research, Inc.

‡ Tokyo University of Science

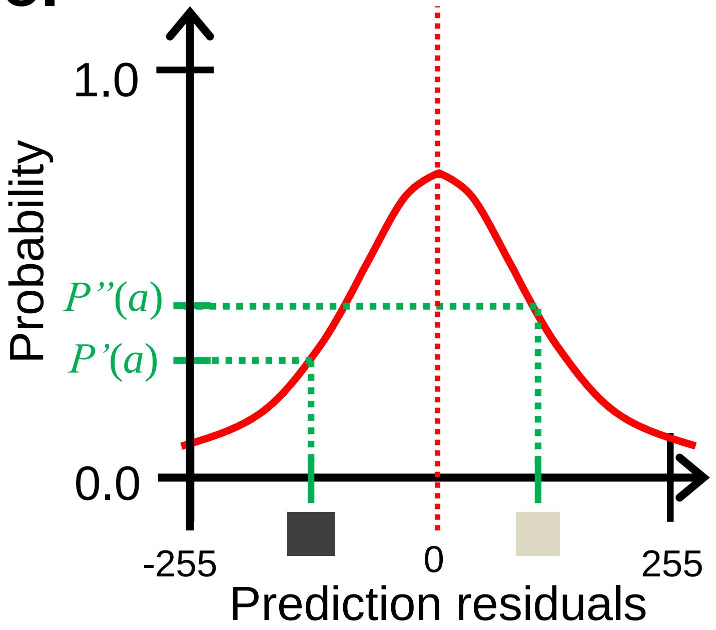
Presenter : Kyohei Unno

Background

- Conventional video coding methods typically consist of two stages.
  1. De-correlation (e.g. Prediction to get residuals)
  2. Entropy coding (e.g. Arithmetic coding)
- In general, probabilities of de-correlated signals are modeled as single-peaked symmetric functions centered at zero.



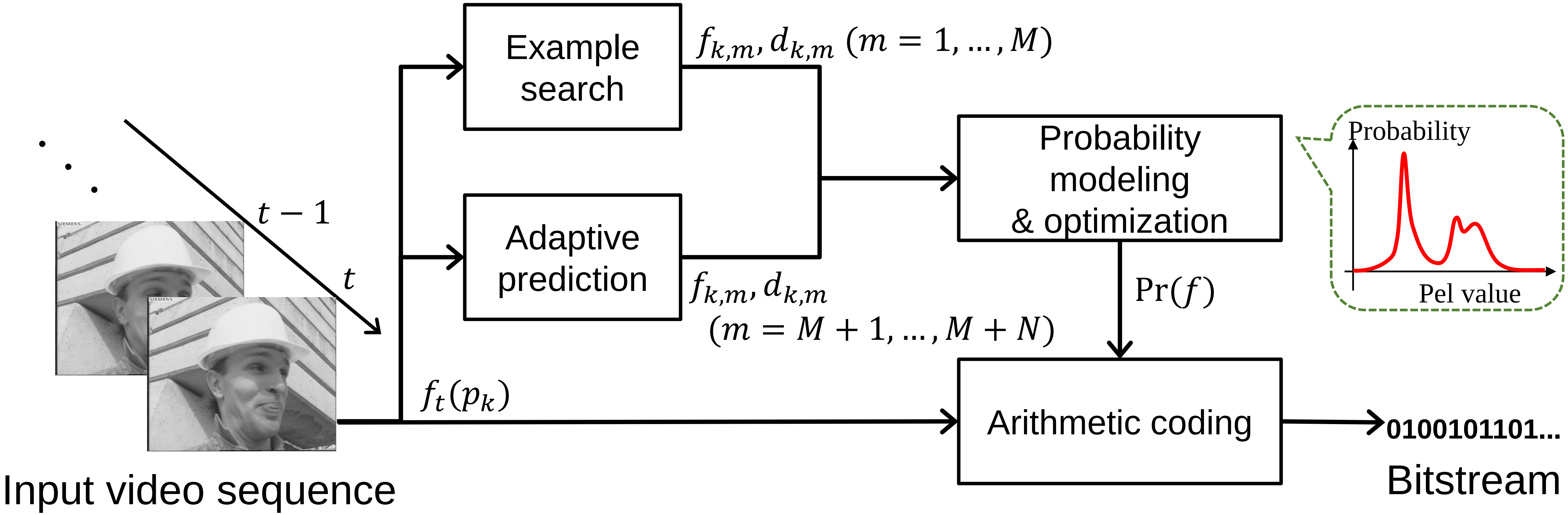
Actual value may be ■ or ■



Such a model can not work well at texture boundaries ☹.

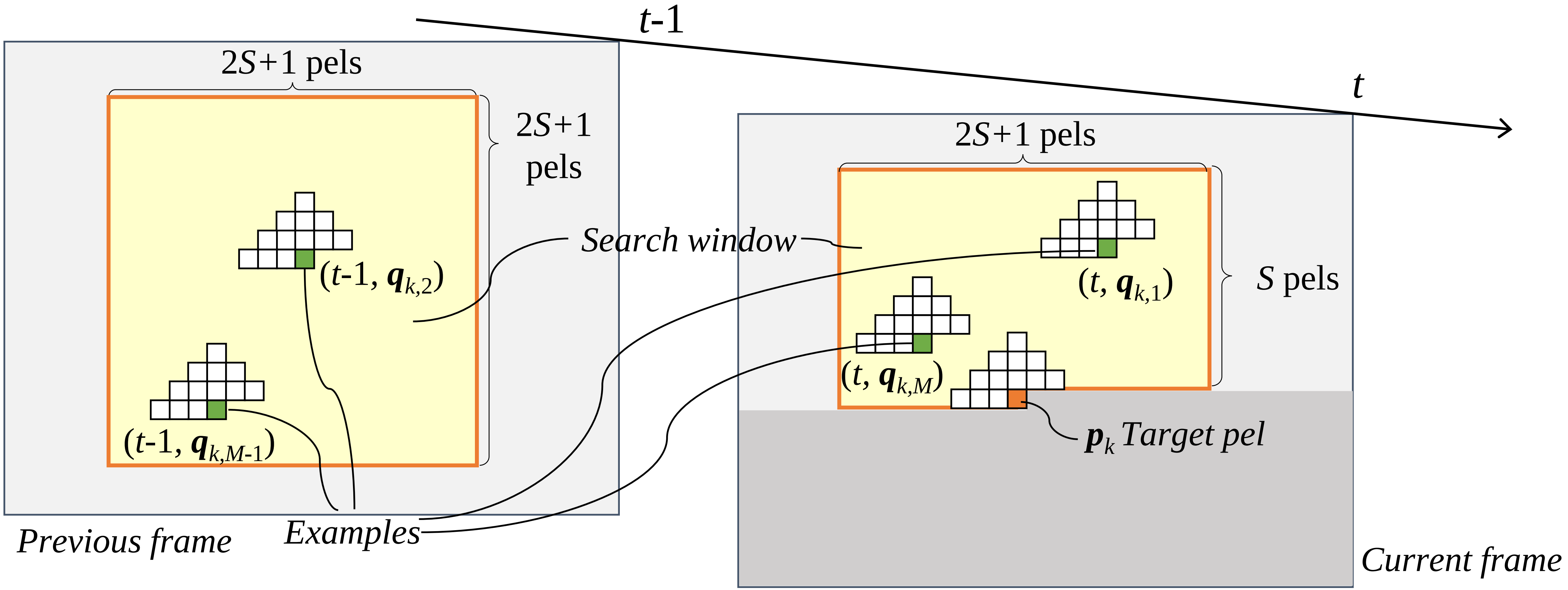
Overview of the proposed method

- Proposed method estimates probability of pel values directly with multi-peaked Gaussian mixture model (GMM) pel-by-pel.
- Center position $f_{k,m}$ and reliability $d_{k,m}$ (related to variance) of each Gaussian are estimated by example search and adaptive prediction.



Example search

- M pels $\{\boldsymbol{q}_{k,1}, \dots, \boldsymbol{q}_{k,M}\}$ that have smaller template matching costs are collected as examples from the search window.
  - The search window is set to both the current frame and the previous frame.



Example search

- Center position $f_{k,m}$ and reliability metric $d_{k,m}$ for each gaussian are given by the examples and the corresponding cost, respectively
  - Center position $f_{k,m}$: Example pel value with local mean compensation.

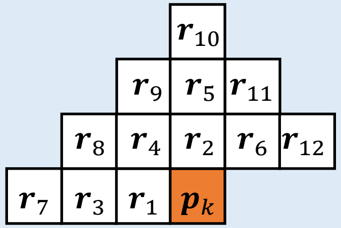
$$f_{k,m} = f_{t-\tau}(\boldsymbol{q}_{\boldsymbol{k},\boldsymbol{m}}) - \mu_{t-\tau}(\boldsymbol{q}_{\boldsymbol{k},\boldsymbol{m}}) + \mu_t(\boldsymbol{p}_{\boldsymbol{k}})$$

Pel value of $\boldsymbol{q}_{\boldsymbol{k},\boldsymbol{m}}$

Local means of neighboring pels at $\boldsymbol{p}_{\boldsymbol{k}}$ and $\boldsymbol{q}_{\boldsymbol{k},\boldsymbol{m}}$

  - Reliability metric $d_{k,m}$: Template matching cost of the k-th example.

$$d_{k,m} = \left[\sum_{i=1}^{12} w_i \cdot \left(f_{t-\tau}(\boldsymbol{q}_{\boldsymbol{k},\boldsymbol{m}} + \boldsymbol{r}_i) - \mu_{t-\tau}(\boldsymbol{q}_{\boldsymbol{k},\boldsymbol{m}}) - f_t(\boldsymbol{p}_k + \boldsymbol{r}_i) + \mu_t(\boldsymbol{p}_k)\right)^2\right]^{\frac{1}{2}} + \lambda_d \cdot \left\|\boldsymbol{q}_{\boldsymbol{k},\boldsymbol{m}} - \boldsymbol{p}_k\right\|_1$$



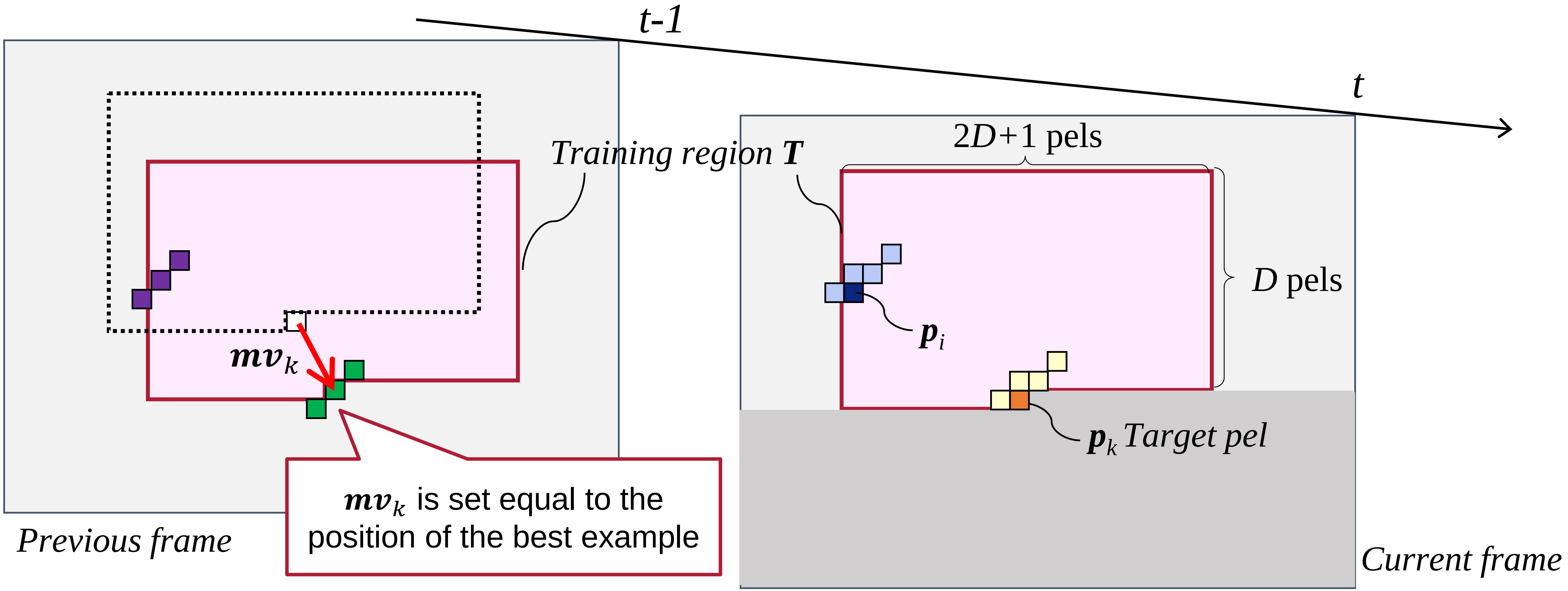
Weighted SSD with local mean compensation

Regularization term

- The number of examples M is optimized for every region of 64 x 64 pels.

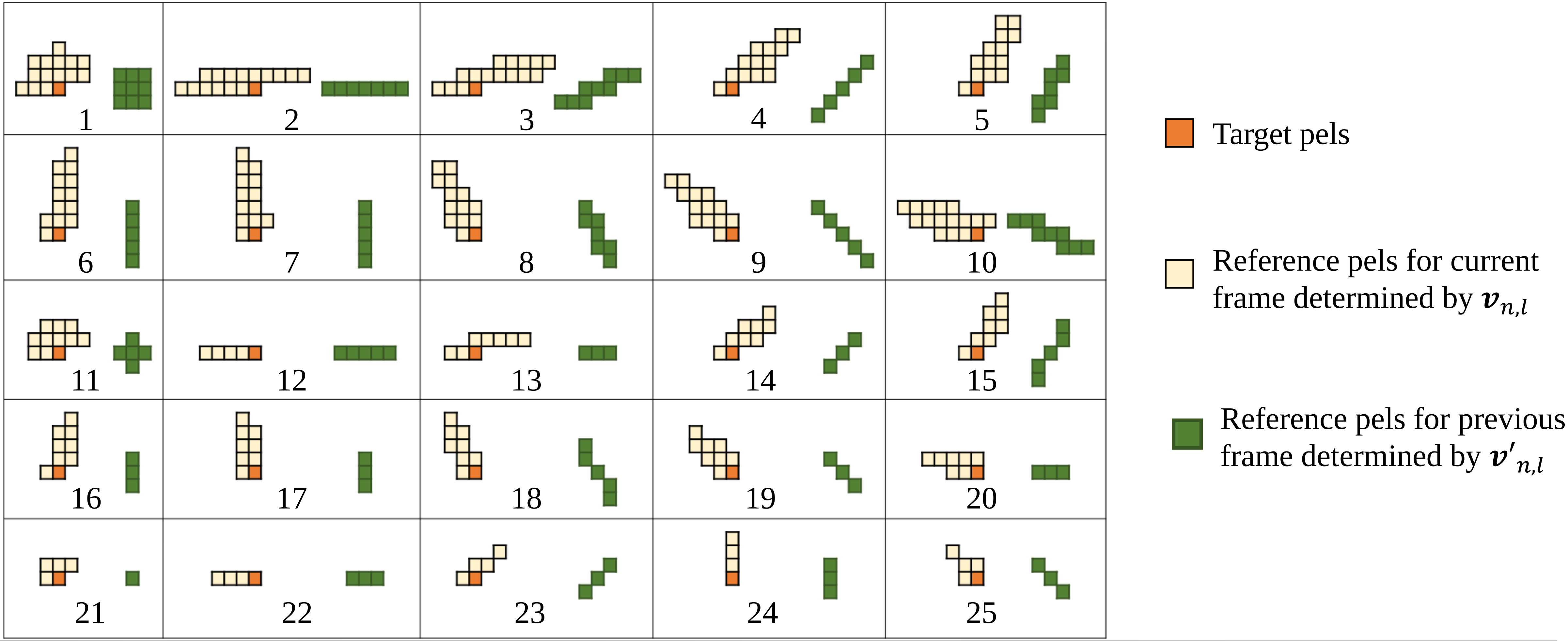
Adaptive prediction

- N kinds of affine predictors (i.e. linear + constant terms) are trained pel-by-pel by the weighted least-square method.
  - Predictors refer pels in the current frame and the previous frame simultaneously.



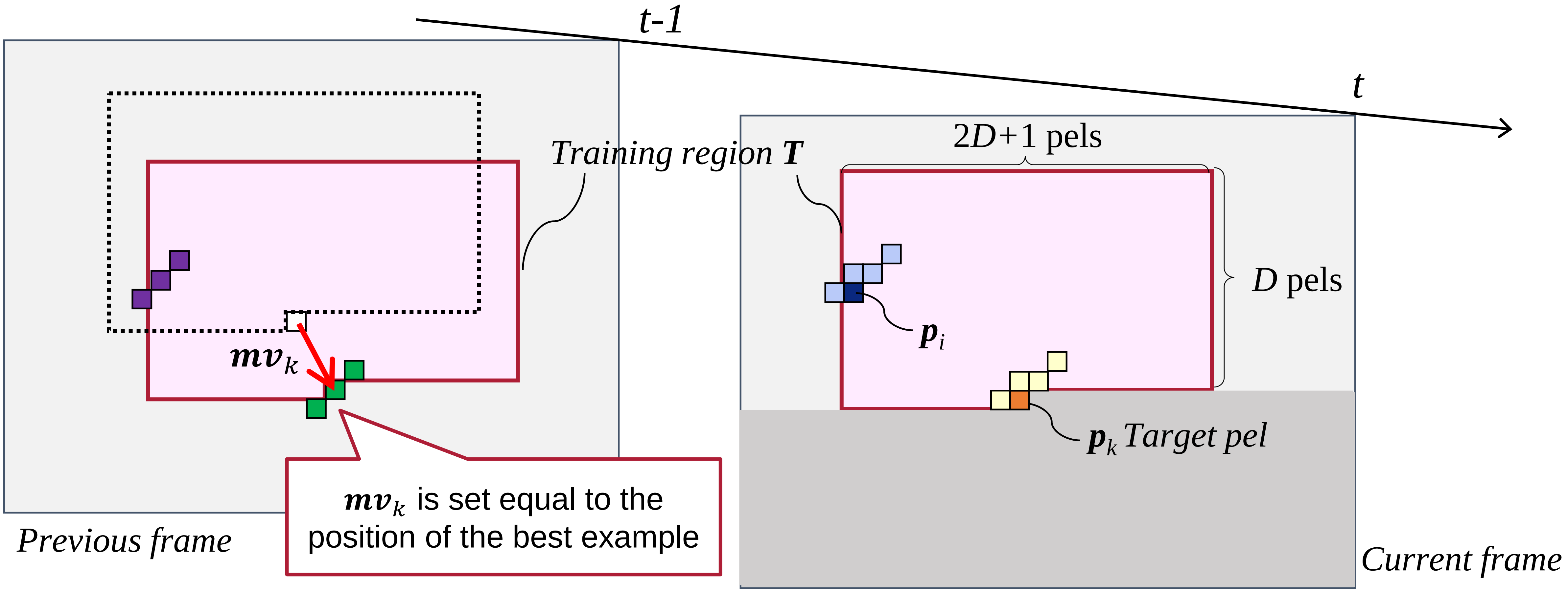
Adaptive prediction

- We use $N = 25$ kinds of predictors for each pel.
  - Each predictor has unique arrangement of reference pels.



Adaptive prediction

- N kinds of affine predictors (i.e. linear + constant terms) are trained pel-by-pel by the weighted least-square method.
  - Predictors refer pels in the current frame and the previous frame simultaneously.



Adaptive prediction

- Center position $f_{k,m}$ and reliability metric $d_{k,m}$ for each gaussian are given as the predicted value and the training error, respectively
  - Center position $f_{k,m}$: Predicted value.

Number of reference pels

$$f_{k,M+n} = b_{n,0} + \sum_{l=1}^{L_n} b_{n,l} \cdot f_t(\boldsymbol{p}_k + \boldsymbol{v}_{n,l}) + \sum_{l=1}^{L'_n} b_{n,L_n+l} \cdot f_{t-1}\left(\boldsymbol{p}_k + \boldsymbol{mv_k} + \boldsymbol{v'}_{n,l}\right)$$

Arrangement of reference pels

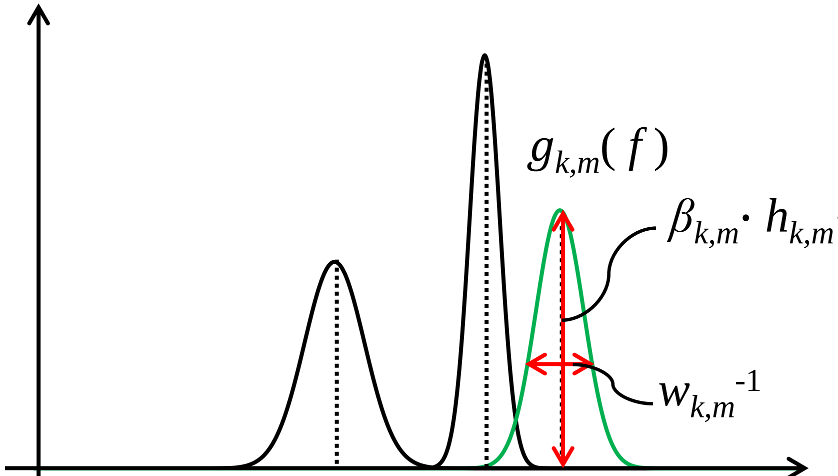
  - Reliability metric $d_{k,m}$: Training error.

$$d_{k,M+n} = \left(\sum_{\boldsymbol{p}_i \in \boldsymbol{T}} \left(\left(f_t(p_i) - \widehat{f}_t(p_i, n) \right)^2 / \sigma_{n,i}^2 \right) \Big/ \sum_{\boldsymbol{p}_i \in \boldsymbol{T}} \left(1/\sigma_{n,i}^2 \right) \right)^{\frac{1}{2}}$$

Squared training prediction error by predictor n

Squared error between neighboring pels of $\boldsymbol{p_k}$ and $\boldsymbol{p_i}$

Probability Modeling



$$P(f \mid \boldsymbol{F}_k, \boldsymbol{D}_k, \boldsymbol{u}_k) = \sum_{m}^{M+N} g_{k,m}(f) + \epsilon$$

$$g_{k,m} = \beta_{k,m} \cdot h_{k,m} \cdot w_{k,m} \cdot \exp(-w_{k,m}^2 \cdot (f - f_{k,m})),$$

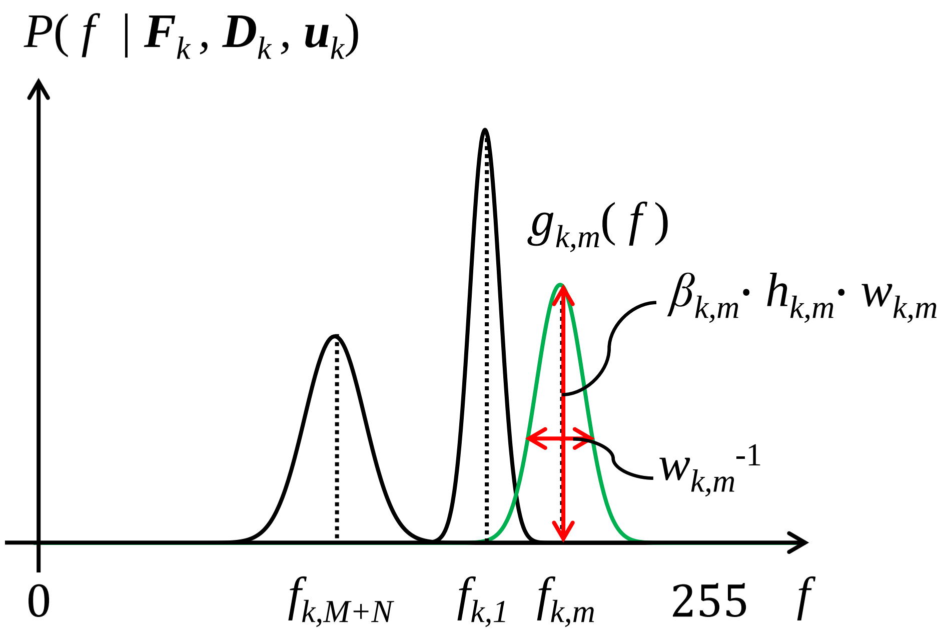
$$h_{k,m} = \exp(-a_1 \cdot d_{k,m}),$$

$$w_{k,m} = a_0 \cdot \exp(-a_2 \cdot d_{k,m}) \cdot \exp(-a_3 \cdot u_{k,m})$$

$$Bitrate(\boldsymbol{p}_k) = -\log_2 \frac{P(f(\boldsymbol{p}_k) \mid \boldsymbol{F}_k, \boldsymbol{D}_k, \boldsymbol{u}_k)}{\sum_{f=0}^{255} P(f \mid \boldsymbol{F}_k, \boldsymbol{D}_k, \boldsymbol{u}_k)}$$

- **GMM is defined by center positions $\boldsymbol{F}_k = \{f_{k,m}\}$, reliability metrics $\boldsymbol{D}_k = \{d_{k,m}\}$, and local feature $\boldsymbol{u}_k = \{u_{k,m}\}$ $(m = 1, \dots, M + N)$.**
  - $\{f_{k,m}, d_{k,m}, u_{k,m} \mid m = 1, \dots, M\}$ **come from example search.**
  - $\{f_{k,m}, d_{k,m}, u_{k,m} \mid m = M+1, \dots, M+N\}$ **come from adaptive prediction.**
  - $\boldsymbol{u}_k$ **is local average of bitrates in the causal area and used for context modeling.**
- **Shape of the probability model is controllable by model parameters $a_0, \dots, a_3$.**

Probability Optimization



$$P(f \mid \boldsymbol{F}_k, \boldsymbol{D}_k, \boldsymbol{u}_k) = \sum_{m}^{M+N} g_{k,m}(f) + \epsilon$$

$$g_{k,m} = \beta_{k,m} \cdot h_{k,m} \cdot w_{k,m} \cdot \exp(-w_{k,m}^2 \cdot (f - f_{k,m})),$$

$$h_{k,m} = \exp(-a_1 \cdot d_{k,m}),$$

$$w_{k,m} = a_0 \cdot \exp(-a_2 \cdot d_{k,m}) \cdot \exp(-a_3 \cdot u_{k,m})$$

$$Bitrate(\boldsymbol{p}_k) = -\log_2 \frac{P(f(\boldsymbol{p}_k) \mid \boldsymbol{F}_k, \boldsymbol{D}_k, \boldsymbol{u}_k)}{\sum_{f=0}^{255} P(f \mid \boldsymbol{F}_k, \boldsymbol{D}_k, \boldsymbol{u}_k)}$$

- Model parameters $a_0, \ldots, a_3$ and the number of examples M are optimized to minimize bitrate in every region of 64 x 64 pels.
  - $a_0, \ldots, a_3$ are optimized by non-linear optimization technique (the quasi-Newton method).
  - Optimal setting of $M (= 0{\sim}31)$ is chosen in the optimization process.
  - The number of predictors $N (= 25)$ is fixed. Optimization of N is a future task.
- $a_0, \ldots, a_3$ and M are sent to a decoder by fixed-length coding.

Experimental results

KDDI Research

- Test sequences
  - JVET CTC Class B, C, D, and E sequences.
    - Align with Low-Delay P condition.
  - Converted from 10 bits to 8 bits by 2 bits right shift.
  - Only luma signal is tested.
  - The first 15 and 30 frames are used for Class B and the other classes, respectively.
    - Due to the long processing time of the proposed method.

- Comparison with lossless mode of HEVC RExt and VVC
  - HEVC RExt : HM 16.20
  - VVC : VTM 8.2
  - The number of reference pictures is restricted to 1 for both HM and VTM.

JVET Test sequences

Class B
1920x1080


MarketPlace


RitualDance


Cactus


BasketBallDrive


BQTerrace

Class C
832x480


BasketBallDrill


BQMall


PartyScene


RaceHorses

Class D
416x240


BasketBallPass


BQSquare


BlowingBubbles


RaceHorses

Class E
1280x720


FourPeople


Johnny


KristenAndSara

Experimental results

- **Test sequences**
  - **JVET CTC Class B, C, D, and E sequences.**
    - Align with Low-Delay P condition.
  - **Converted from 10 bits to 8 bits by 2 bits right shift.**
  - **Only luma signal is tested.**
  - **The first 15 and 30 frames are used for Class B and the other classes, respectively.**
    - Due to the long processing time of the proposed method.

- **Comparison with lossless mode of HEVC RExt and VVC**
  - **HEVC RExt : HM 16.20**
  - **VVC : VTM 8.2**
  - **The number of reference pictures is restricted to 1 for both HM and VTM.**

Experimental results

KDDI Research

■ Results

- Proposed method achieves about 9% and 5% better coding rates than HEVC RExt and VVC, respectively.
- Over 13% bitrate reductions from HEVC RExt are observed for RaceHorses (Class C) and all of Class E sequences.

■ Weakness of the proposed method is its processing time.

- Our implementation has not yet been tuned for the processing time.

Comparison of coding rates (bits/pel).

Test sequences		Proposed	HEVC RExt	VVC
MarketPlace	Class B (1920x1080)	3.229	3.675	3.532
RitualDance		1.612	2.036	1.940
Cactus		3.814	4.120	3.946
BasketballDrive		3.473	3.785	3.602
BQTerrace		3.647	3.879	3.749
BasketballDrill	Class C (832x480)	2.866	3.157	3.039
BQMall		3.404	3.660	3.505
PartyScene		3.828	3.973	3.794
RaceHorses		3.329	3.902	3.793
BasketballPass	Class D (416x240)	2.068	2.182	2.114
BQSquare		3.459	3.703	3.426
BlowingBubbles		3.844	3.889	3.767
RaceHorses		3.329	3.658	3.571
FourPeople	Class E (1280x720)	2.508	2.902	2.768
Johnny		2.500	2.882	2.771
KristenAndSara		2.376	2.741	2.618
Average		3.080	3.384	3.246
Bitrate reduction (vs HEVC RExt)		-8.97%	0.00%	-4.07%

Conclusion

KDDI Research

- **Proposed method**
  - **Proposed method estimates probability of pel values directly with multi-peaked Gaussian mixture model (GMM) pel-by-pel.**
  - **Center position $f_{k,m}$ and reliability $d_{k,m}$ (related to variance) of each Gaussian are estimated by example search and adaptive prediction.**
  - **Model parameters of GMM $a_0, \dots, a_3$ and the number of examples M are optimized to minimize bitrate in every region of 64 x 64 pels.**
  - **The proposed method achieves better bitrate in comparison with VVC and HEVC lossless mode.**
- **Future works**
  - **Reducing the processing time.**
    - Adaptive prediction and optimization waste the most part of encoding time.
  - **Adaptive setting of the number of predictor N.**
    - It might be effective for both improving coding performance and reducing processing time.

Acknowledgement

This work was supported by Ministry of Internal Affairs and Communications (MIC) of Japan (Grant no. JPJ000595).