, Member, IEEE *, Student Member, IEEE* *, Student Member, IEEE*
, Member, IEEE

***Abstract*—In this paper, we propose a region-of-interest (ROI) based HEVC coding approach for conversational videos, with a novel hierarchical perception model of face (HP model), to improve the perceived visual quality of state-of-the-art HEVC standard. In contrast to the previous ROI-based video coding approaches, this novel HP model allows the unequal importance of facial features (e.g., the eyes and mouth) within the facial region, by generating a pixel-wise weight map. Bene tting from such a perception model, the adaptive coding tree unit (CTU) partition structure is developed to alleviate the encoding complexity of HEVC, without any degradation of the visual quality in facial regions, especially in the regions of facial features. Subsequently, for the rate control in HEVC a weight-based uni ed rate-quantization (URQ) scheme, instead of the conventional pixel-based URQ scheme, is proposed to adaptively adjust the value of quantization parameter (QP). Such an adaptive adjustment of QPs is capable of allocating more bits to the face/facial features with respect to our HP model, and as a result, the visual quality of face, in particular facial features, can be enhanced for conversational HEVC coding. Finally, the experimental results show that the perceived visual quality of our approach is greatly improved, with even less encoding time, for conversational video coding on the HEVC platform.**

***Index Terms*—HEVC, perceptual video compression, teleconferencing, rate distortion.**

N fl

fi

fi

fi

■ fi

fi

fi

fi

Akiyo

Yan ×

×

fl

Akiyo *Yan*

fi

fi

fi

fi

fi

fi

fi

fi

A. The Necessity of Hierarchical Perception Model of Face

fi

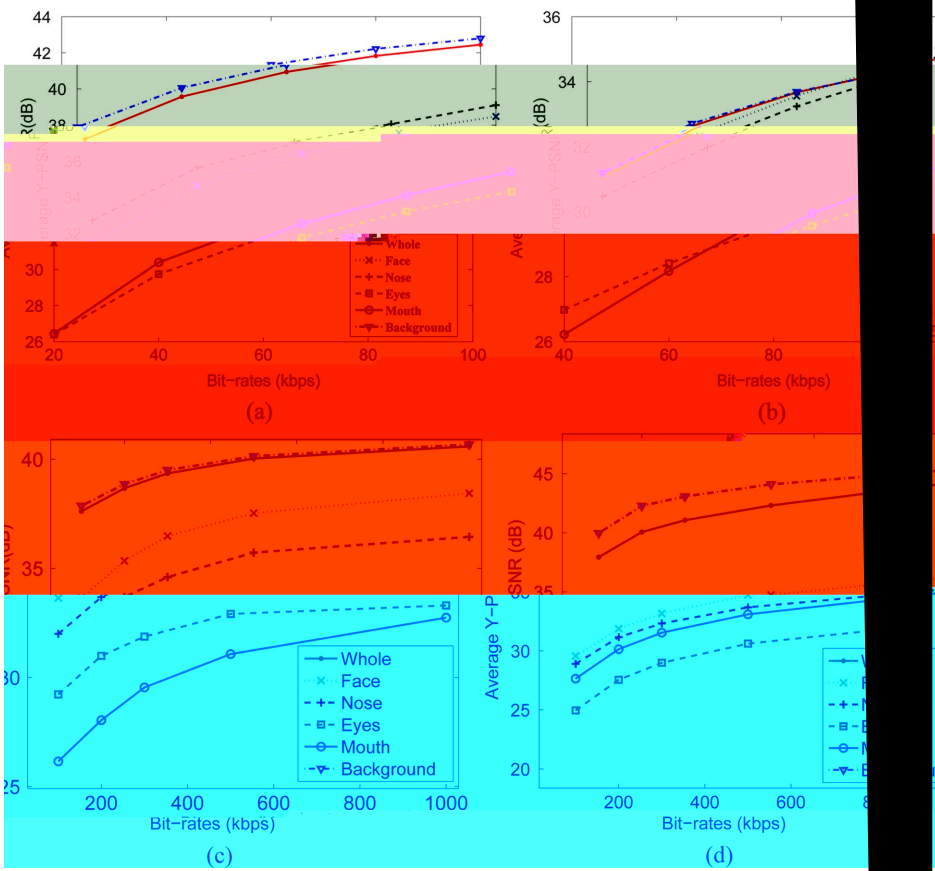
Akiyo *Foreman* *Simo*

fi *Yan*

fi

fi

Foreman



Akiyo *Foreman* *Yan* *Simo*

$\{\bar{\mathbf{p}}_t\}_{t=1}^{T}$

$\mathbf{p}_t = s\mathbf{R}(\bar{\mathbf{p}}_t +$

B. The Proposed Hierarchical Perception Model of Face

s

$\mathbf{q}$

prior

fi

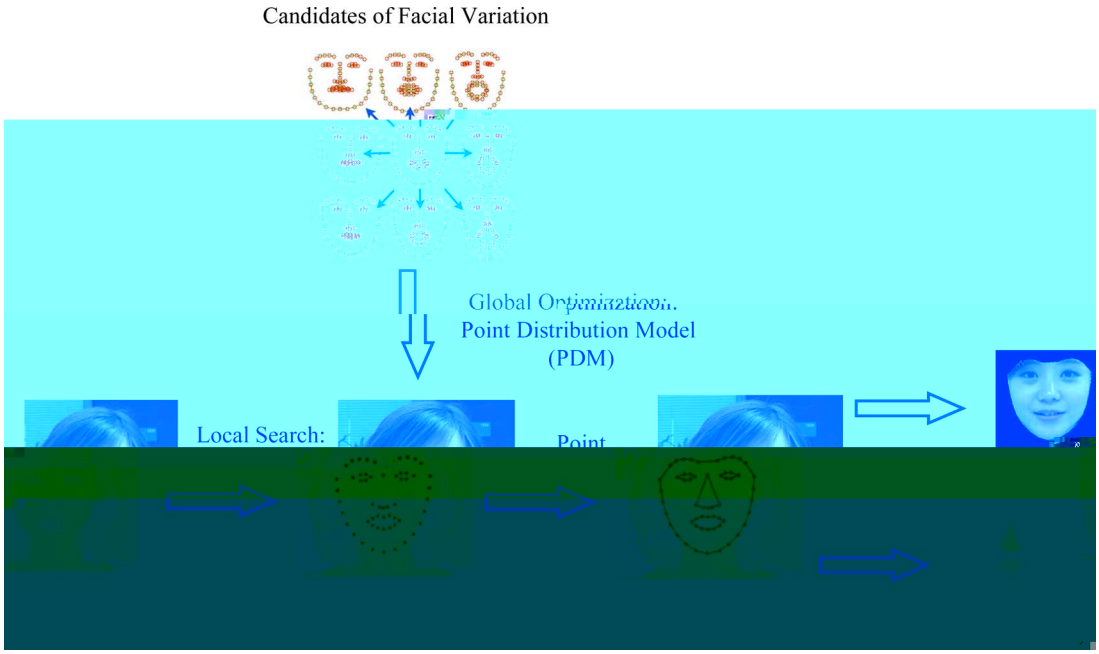
fi

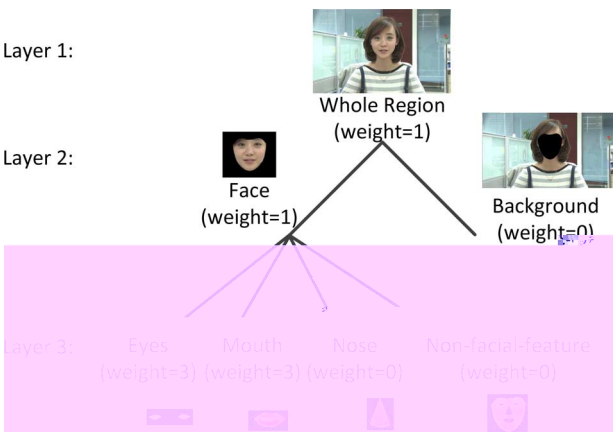
fi

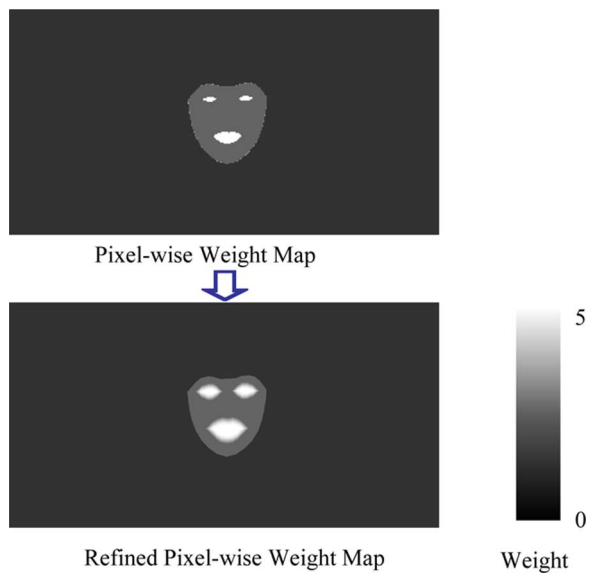
$\min_{\mathbf{p}_t} \sum_{=1}^{T}$ p

fi

fl







fi

fi

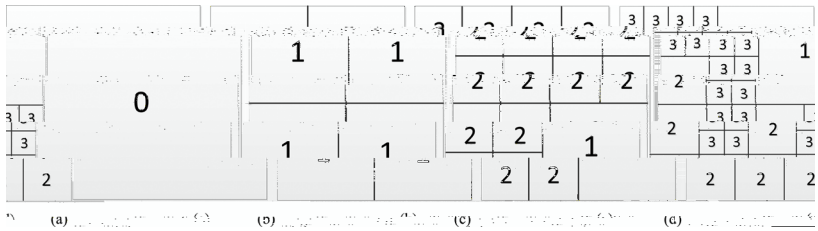
$(= 3)$

fi

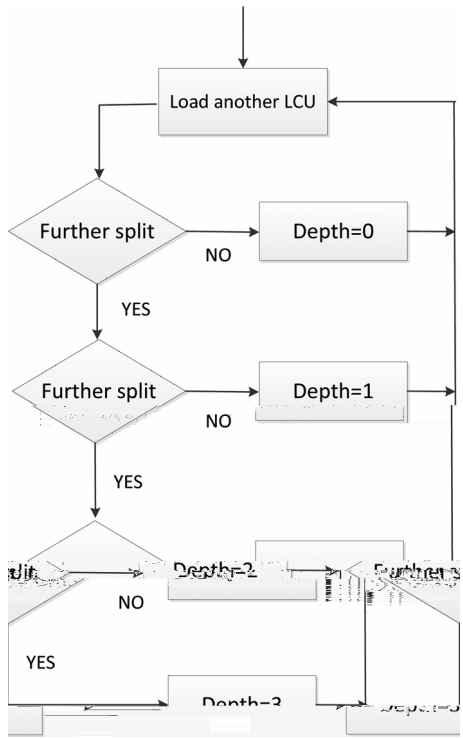
fi

fi

fi



×



$\{w_n\}_{n=1}^{N}$

N

Δd_n n

v_i i

σ_i

v_i i

$$\Delta w_n = v_i e^{-\frac{1}{2}(\frac{\Delta d_n}{\sigma_i})^2},$$

$v_i > 0$

× ×

A. The CTU Partition Structure in HEVC

×

$i = 1$ $v_1 = 3; i = 2$
$v_2 = 3; i = 3$ $v_3 = 0$

fl

fi

fi

fi

B. The Proposed ROI-Based Adaptive CTU Partition Structure

fi λ_j

j

$$\lambda_j = \frac{1}{M} \sum_{n \in \mathbf{n}_j} w_n,$$

$\mathbf{n}_j$ j

M

$\{\lambda_j\}_{j=1}^{J}$ J

$\{z_j\}_{j=1}^{J}$

$$z_j = \begin{cases} 1 & \text{if } \lambda_j \leq \theta_1 \\ 2 & \text{if } \theta_1 < \lambda_j \leq \theta_2 \\ 3 & \text{if } \lambda_j > \theta_2 \end{cases}$$

a b fi

M
$\mathrm{QP_j}$

$\mathrm{QP_j}$
$$= \frac{a \cdot \mathrm{MAD_{pred,j}} + \sqrt{a^2 \cdot \mathrm{MAD_{pred,j}}^2 + 4b \cdot \mathrm{MAD_{pred,j}} \cdot \frac{T_j}{M}}}{\frac{2T_j}{M}}.$$

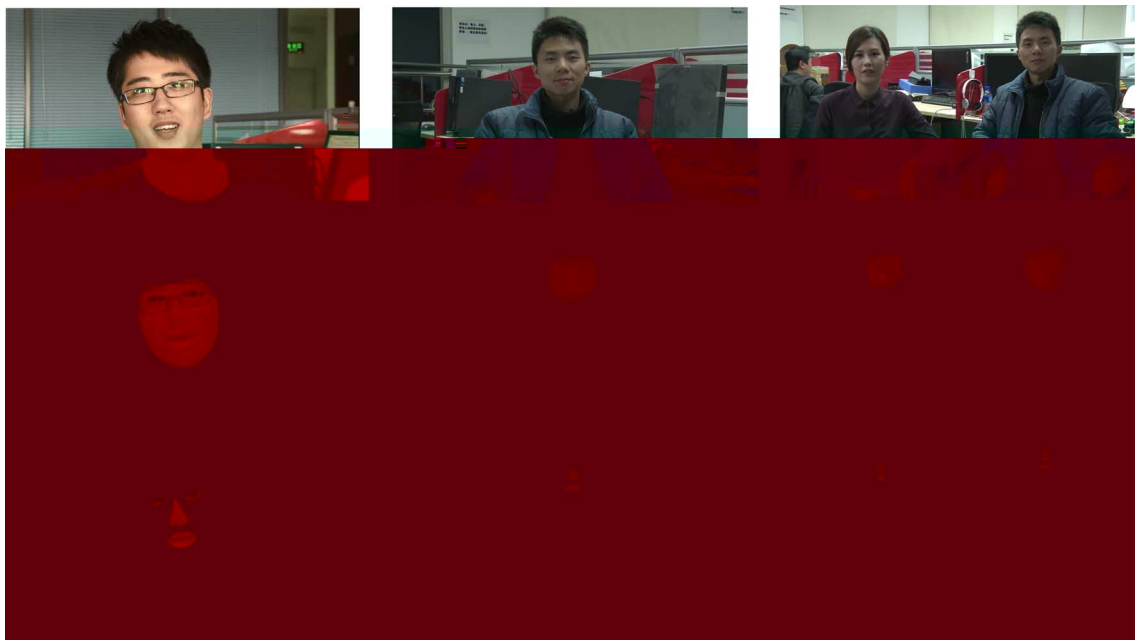
$\mathrm{QP_j}$
T_j T_j

$$T_j = \beta \cdot \hat{T}_j + (1 - \beta) \cdot \tilde{T}_j,$$

$\tilde{T}_j$
$\hat{T}$

B_j

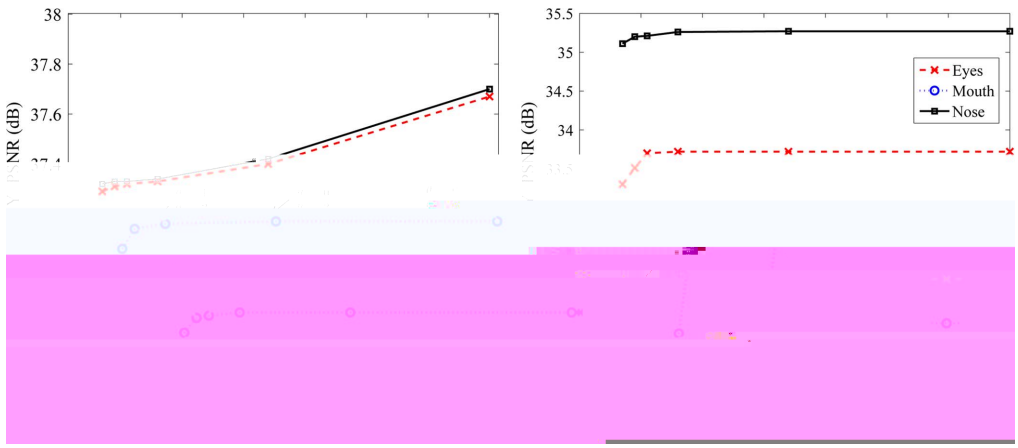
$$\hat{T}'_j = \sum_{n \in \mathbf{n}}$$



Simo Lee Couple ×

fi fi

Simo Lee Couple



Yan

Video Sequences	*Akiyo*					*Foreman*				
Bit-rates (kbps)	20	40	60	80	100	40	60	80	100	120
Encoding time reduction (%)	19.4	22.1	23.8	20.0	20.6	22.2	22.1	21.5	21.8	22.9
Average Y-PSNR improvement in face (dB)	1.15	1.32	1.43	1.59	1.61	1.13	1.43	1.51	1.38	1.16

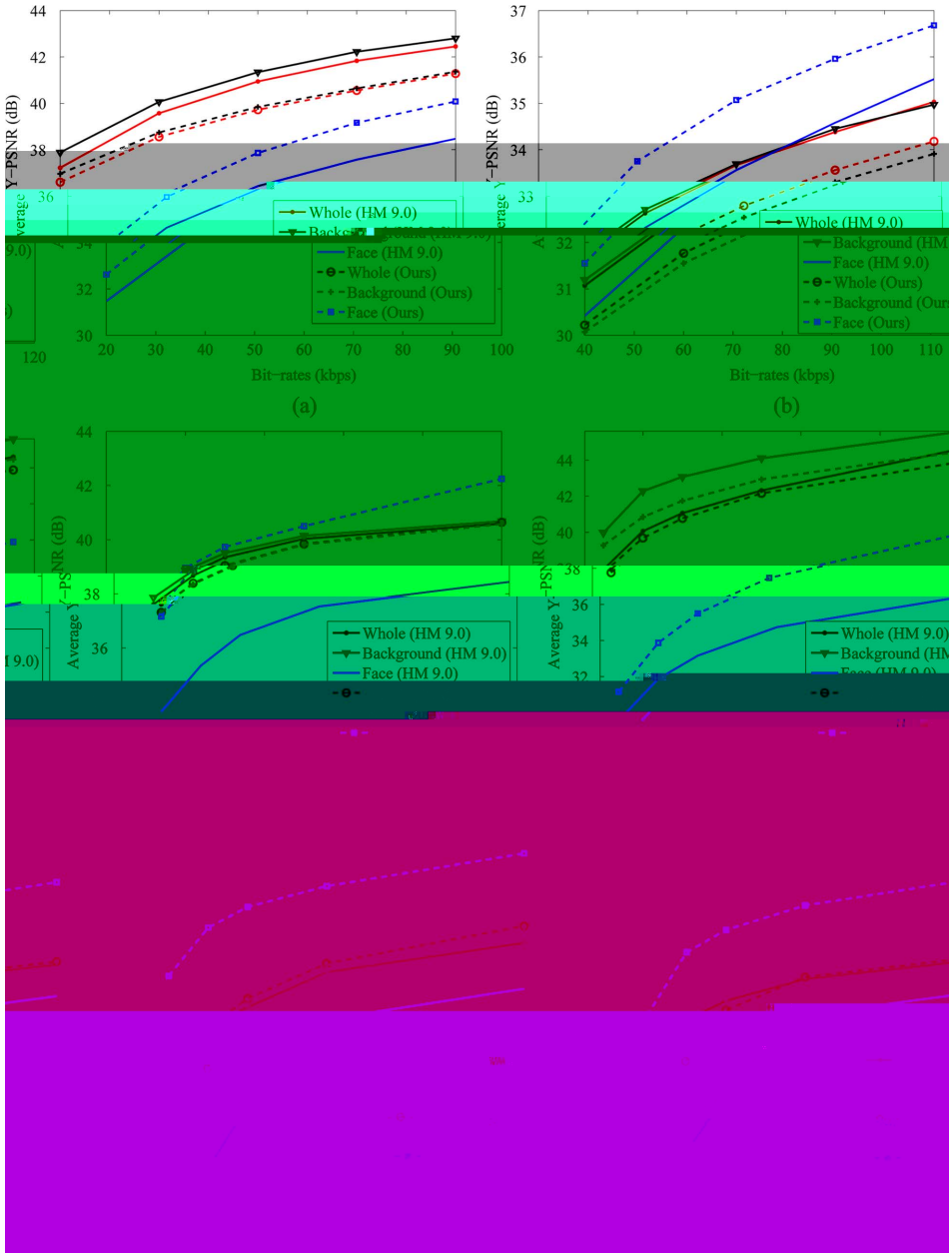
Video Sequences	*Yan*					*Simo*				
Bit-rates (kbps)	100	200	300	500	1000	100	200	300	500	1000
Encoding time reduction (%)	54.0	54.5	53.4	51.8	53.0	58.7	57.1	56.0	54.8	53.8

fi

θ_1 θ_2

θ_1

θ



Akiyo *Foreman* *Yan* *Simo* *Lee* *Couple*

×

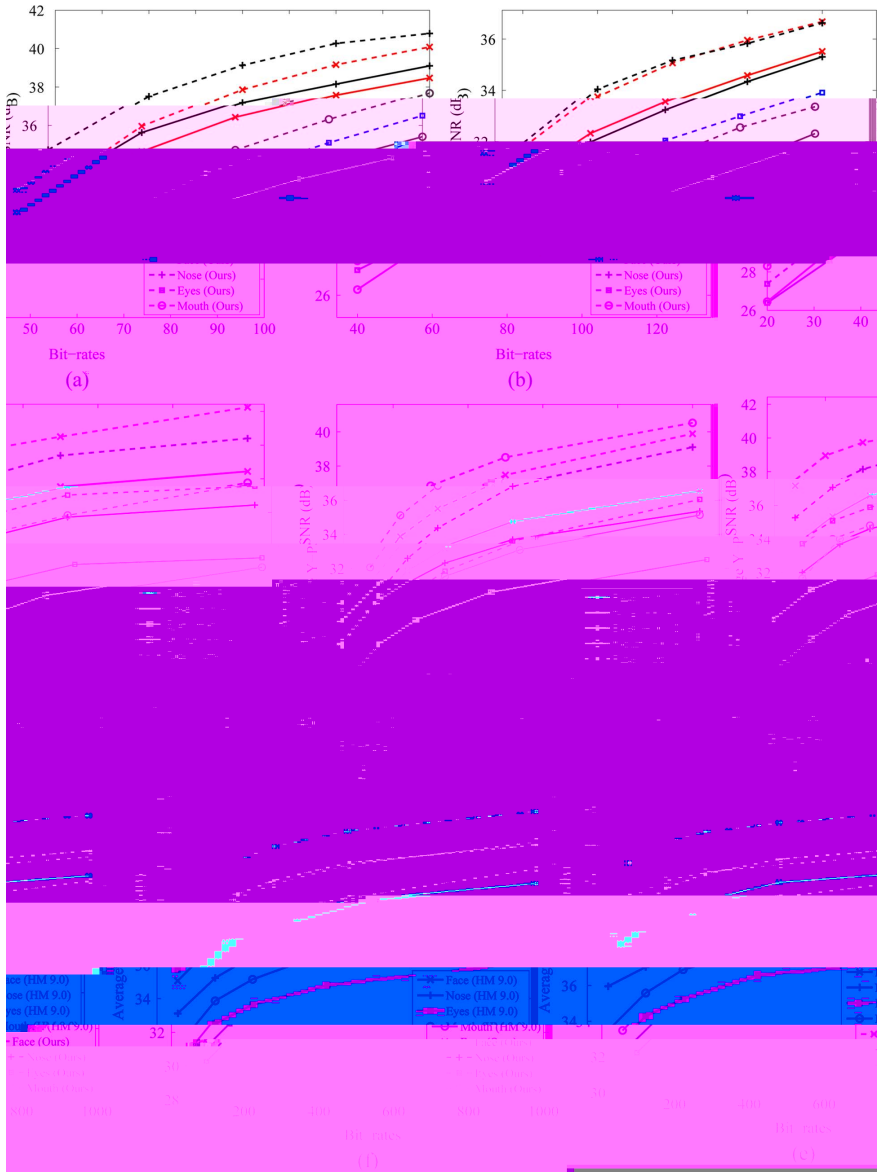
fi

Akiyo *Foreman*

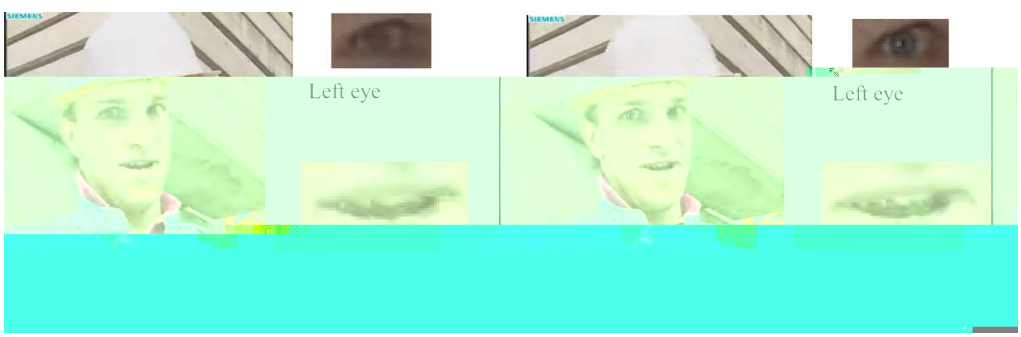
Yan *Simo* *Lee*

Couple

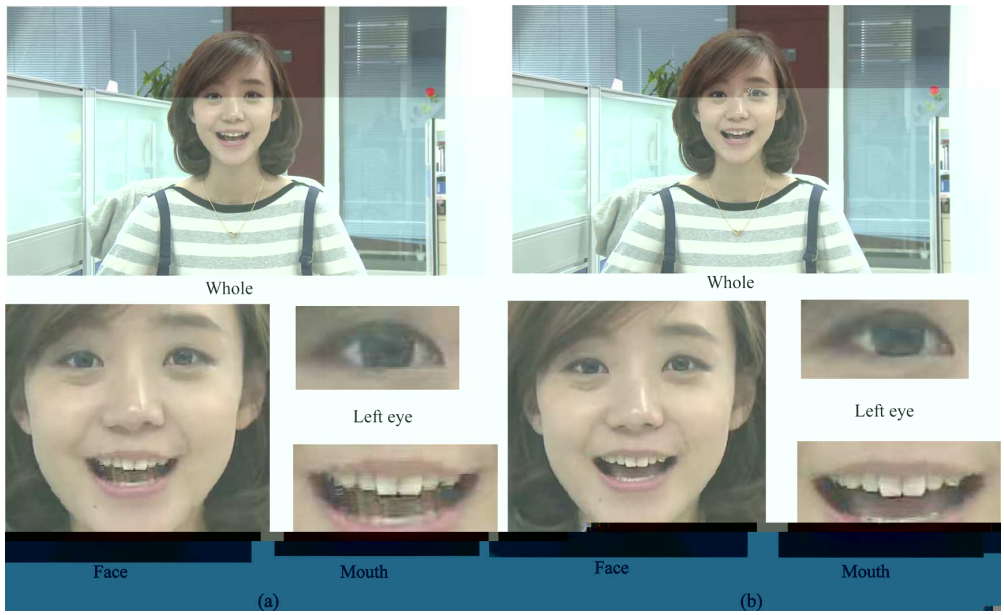
″



Akiyo *Foreman* *Yan* *Simo* *Lee* *Couple*



Foreman



Yan

Sequences	Resolution	Bit-rates (kbps)	HM 9.0 DMOS	Our DMOS	DMOS Difference
Akiyo	352×288	20	59.06	50.16	-8.90
		40	34.66	23.45	-11.21
Foreman	352×288	60	73.44	62.93	-10.51
		80	57.62	43.71	-13.91
Yan	1920×1080	100	71.88	46.15	-25.73
		300	46.22	31.46	-14.76
Simo	1920×1080	100	71.63	57.73	-13.90
		300	54.35	38.46	-16.10
Lee	1920×1080	100	67.23	40.62	-26.61
		300	45.41	29.15	-16.26
Couple	1920×1080	100	73.78	46.41	-27.37
		300	47.39	28.16	-19.23

fi

fl

fi

fi

IEEE Trans. Circuits Syst. Video Technol.

IEEE Trans. Circuits Syst. Video Technol.

, Foundations of Vision

Per-

cept. Motor Skills

IEEE J. Sel. Topics Signal Process.

IEEE Consumer Electron. Mag.

IEEE Trans. Circuits Syst. Video Technol.

IEEE Trans. Circuits

IEEE Trans. Cir- *Syst. Video Technol.*

cuits Syst. Video Technol.

IEEE

Trans. Circuits Syst. Video Technol.

IEEE Trans. Pattern Anal. Mach. Intell.

Mai Xu

Proc. ICCV

Proc. SPIE

Comptuer Graphics

Proc. ICIP

IEEE Trans. Image Process.

Xin Deng

Proc. ICPR

Image Vis. Comput.

IEEE Trans.

Circuits Syst. Video Technol.

IEEE Trans. Pattern Anal. Mach. In-

tell.

fi

Vis. Res.

IEEE Trans. Circuits Syst. Video Technol.

Shengxi Li

Signal Process.: Image Commun.

Proc. VCIP

, Handbook of Face Recogntition

IEEE Trans. Circuits Syst. Video Technol.

Proc. ICME

fl

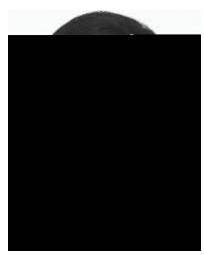
ACM Trans. Mul-

timedia Comput.

IEEE Trans. Multimedia

fi

J. Vis. Commun. Image Represent.



Zulin Wang

fi

fi *IEEE J. Sel.*

Topics Signal Process.

IEEE Trans. Pattern Anal. Mach. Intell.