# Online Learning Based Face Distortion Recovery for Conversational Video Coding

Xi Wang[*+], Li Su[+], Qingming Huang[*+], Guorong Li[+] and Honggang Qi[+]

[*] *Key Lab of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China*

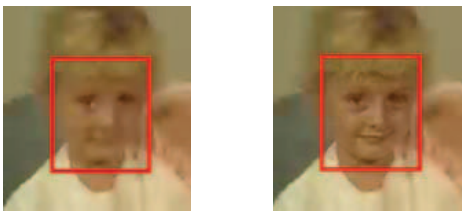
[+] *University of Chinese Academy of Sciences, Beijing, China*

*{xwang, lsu, qmhuang, grli, hgqi}@jdl.ac.cn*

In a video conversation, the participants usually remain the same. As the conversation continues, similar facial expressions of the same person would occur intermittently. However, the correlation of similar face features has not been fully used since the conventional methods only focus on independent frames. We set up a face feature database and updated it online to include new facial expressions during the whole conversation. At the receiver side, the database is used to recover the face distortion and thus improve the visual quality. Additionally, the proposed method brings small burden to update the database and is generic to various CODEC.

The proposed algorithm is composed of two parts: the face database training at the sender side and the face distortion recovery at the receiver side. At the sender side, the original video and the reconstructed video can be accessed both. To obtain an optimal database, face alignment is performed for each frame. The face regions in the original and distorted video frames are decomposed into overlapping n × n blocks and their positions are recorded. Each training unit is composed of three items, original block($o$), reconstructed block($r$) and the center coordinate($c$). It can be written as $U = (o, r, c)$. The training units are collected from the first $m$ frames to initialize the database. Firstly, we classify the training units into $K$ groups according to the reconstructed blocks, by using $K$-means. Then, if the variance of units coordinates in one group is too large, we further separate it into new groups. After initialization, we update the database online. For the $i^{\text{th}}$ group, the group center is $\bar{U}_i = (\bar{o}_i, \bar{r}_i, \bar{c}_i)$ and the number of training units is $T_i$ .When a new training unit $U_c = (r_c, o_c, c_c)$ arrives, the most correlated group is searched according to the reconstructed block similarity and the position constraint$\{min\|r_c - \bar{r}_i\|^2, \|c_c - \bar{c}_i\|^2 < D\}$. Let $E = max(\|r_c - \bar{r}_i\|^2, \|o_c - \bar{o}_i\|^2)$, if $E > \varepsilon$ we create a new group and add $U_c$ to this group. Else if E ≤ ε and $T_i$ is smaller than a giving constant integer $R$, we add $U_c$ to the $i^{\text{th}}$ group and update the group center $\bar{U}_i$ and set $T_i = T_i + 1$. While $T_i$ arrives at $R$, we put the group center $\bar{U}_i = (\bar{o}_i, \bar{r}_i, \bar{c}_i)$ into the database and send it to the receiver side. This group will not be updated anymore. The online training processing works till the end of the conversation. At the receiver side, the face regions in the decoded frames are divided into non-overlapping n × n blocks($r_d$) and their center coordinates($c_d$) is recorded. For the decoded block($r_d, c_d$), we search the most similar unit $U_j = (o_j, r_j, c_j)$ in the database by the condition $\{min\|r_d - r_j\|^2, \|c_d - c_j\|^2 < D\}$. If $\|r_d - r_j\|^2 \le \varepsilon$, block $r_d$ will be recovered by the corresponding original block $o_j$ ($r_d = o_j$). Else, $r_d$ keep unchanged. Finally, the recovered frames are displayed.

We compare the proposed method with H.264/AVC reference software JM17.1 on test sequences in CIF format. Our method obtains averagely 6.05db gain in face regions and the additional burden for the database is averagely 1.96 units per frame.



(a) JM (b) Proposed

Figure 1. Subjective quality comparison

This work was supported in part by National Basic Research Program of China (973 Program): 2009CB320906, in part by National Natural Science Foundation of China: 61025011, 61001177 and 61001108.

1068-0314/13 $26.00 © 2013 IEEE
DOI 10.1109/DCC.2013.105