

基于特征值置换的子空间语音增强算法

孙成立^{a,b}, 穆俊生^b

(南昌航空大学 a. 图像处理与模式识别重点实验室; b. 信息工程学院, 南昌 330063)

摘 要: 为提高重构语音质量, 提出一种子空间语音增强算法。在内嵌式预白化子空间方法的基础上, 用特征值置换增强语音效果。将纯净语音和噪声的协方差矩阵进行广义特征值分解后, 大特征值分量包含纯净语音信息, 而小特征值分量包含噪声信息, 待特征值排序后, 用相邻的大特征值分量置换小特征值分量, 可有效提高语音帧的相关性, 获得更好的增强语音信号。相比传统的子空间方法, 该算法适用于强噪声环境, 能显著提高信噪比。

关键词: 语音质量; 语音增强; 子空间方法; 特征值分解; 语音帧

中文引用格式: 孙成立, 穆俊生. 基于特征值置换的子空间语音增强算法[J]. 计算机工程, 2016, 42(2): 272-277, 282.

英文引用格式: Sun Chengli, Mu Junsheng. Subspace Speech Enhancement Algorithm Based on Eigenvalue Substitution[J]. Computer Engineering, 2016, 42(2): 272-277, 282.

Subspace Speech Enhancement Algorithm Based on Eigenvalue Substitution

SUN Chengli^{a,b}, MU Junsheng^b

(a. Key Laboratory of Image Processing and Pattern Recognition;

b. School of Information Engineering, Nanchang Hangkong University, Nanchang 330063, China)

【Abstract】 A subspace speech enhancement approach is proposed for a better reconstructed speech quality. On the basis of Generalized Eigenvalue Decomposition (GEVD) subspace approach for speech enhancement, speech quality is improved by eigenvalue substitution. The study shows that after the generalized eigenvalue decomposition for the covariance matrices of the clean speech signal and the noise, the larger eigenvalue component contains the main information of speech signal, while the smaller eigenvalue component contains noise information. After the eigenvalue ordering, the smaller eigenvalue and its corresponding eigenvectors are replaced with the larger ones so that there is a better enhanced speech signal and the correlation of the speech frame can be improved greatly. Compared with other subspace methods, the proposed method can restrain strong noise more effectively, and get a higher signal noise ratio.

【Key words】 speech quality; speech enhancement; subspace approach; eigenvalue decomposition; speech frame

DOI: 10.3969/j.issn.1000-3428.2016.02.048

1 概述

语音信号在传输过程中容易受到背景噪声的干扰, 语音增强是解决噪声污染的有效方法, 它的目标是尽可能从带噪语音信号中恢复出纯净的语音信号, 改善语音听觉质量以减轻听觉上的疲劳。经典的语音增强算法包括谱减法、维纳滤波法和子空间算法等。谱减法^[1-3]是最简单常见的语音增强算法, 但是它容易产生音乐噪声。维纳滤波法^[4-6]是一种基于最小均方准则的方法, 通过设置滤波器系数尽可能消除

噪声。该方法可以有效地抑制音乐噪声, 但是低信噪比下语音失真度较大。子空间^[7-10]方法是近年兴起的一种语音增强算法, 它利用投影矩阵把含噪语音投影到 2 个相互正交的子空间上, 分别为信号子空间(对应纯净语音信号)和噪声子空间(对应噪声信号), 通过将噪声子空间置零, 从语音子空间中获得增强的语音信号。但由于噪声是分布在整个空间的, 语音子空间也不可避免地残留一些噪声成分, 因此, 传统子空间方法不能完全消除噪声, 尤其在低信噪比下增强语音中往往残留有较多的噪声成分。

基金项目: 国家自然科学基金资助项目(61362031); 江西省自然科学基金资助项目(20142BAB207002); 江西省教育厅科技基金资助项目(GJJ14520)。

作者简介: 孙成立(1975-), 男, 副教授, 主研方向为语音信号处理、语音识别、图像处理与识别; 穆俊生, 硕士研究生。

收稿日期: 2014-12-31 **修回日期:** 2015-03-02 **E-mail:** 794054754@qq.com

本文在内嵌式预白化子空间方法的基础上, 提出用特征值置换改善增强语音效果, 对纯净语音和噪声的协方差矩阵进行广义特征值分解后, 通过用相邻的大特征值分量置换小特征值分量, 提高获取语音帧的相关性, 进而改善降噪效果。

2 背景介绍

子空间算法^[11]根据最小二乘估计准则, 对含噪语音进行奇异值分解 (Singular Value Decomposition, SVD), 忽略较小的奇异值, 从中恢复出语音信号, 起初这种方法只适合白噪声的情况。文献[12]将基于奇异值分解算法推广到有色噪声环境下, 利用奇异值分解算法, 将对噪声信号的白化作为子空间算法的一部分, 并采用最小方差准则估计语音信号子空间。文献[13]提出采用 Karhunen-Loeve 变换 (Karhunen-Loeve Transform, KLT) 进行子空间分解的方法, 通过对有色噪声进行预白化处理, 使得子空间算法可以处理有色噪声。文献[14]假设语音信号的特征向量可以近似地使有色噪声的协方差矩阵对变化, 提出了一种自适应的 KLT 方法, 但这种方法为次优估计方法。文献[7]利用广义特征值分解 (Generalized Eigenvalue Decomposition, GEVD) 给出了一种最优的估计, 即内嵌式预白化子空间方法。文献[9]提出了一种新的子空间优化准则重构误差方差 (Variance of Reconstruction Error, VRE), 用于平衡噪声抑制和语音失真。

纯净的语音信号通常具有短时平稳性, 相邻帧之间的数据具有连续性, 表现为由语音帧向量组成的语音矩阵具有较强的相关性。由于时域上背景噪声的随机性, 表现为噪声和纯净语音的叠加会降低语音信号的相关性。从图 1 可以看出, 和纯净语音相比, 加入噪声后使得含噪语音的相关性严重下降。在现有的子空间算法中, 只考虑噪声的消除, 并没有考虑如何提高降噪语音的相关性, 因而不能达到良好的语音增强效果。横坐标表示计算相关度用到的前后语音帧。

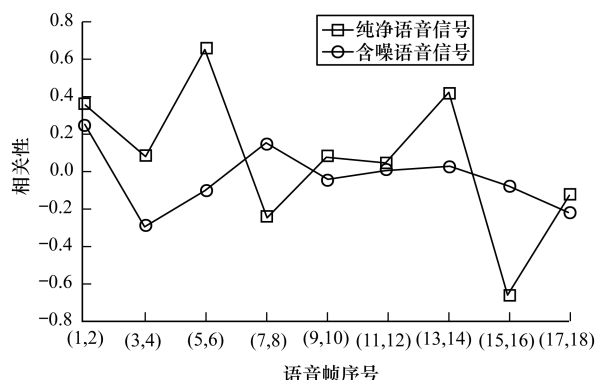


图 1 纯净语音信号和含噪语音信号帧相关性对比

3 子空间语音增强方法

子空间方法的基本原理是将带噪语音信号的向量空间分解为相互正交的 2 个子空间, 其中一个只包含噪声, 另一个同时包含语音和噪声。通过将噪声子空间置零, 得到纯净语音信号的近似估计结果。

假设纯净语音信号 s 与噪声 n 互不相关且为零均值, 那么含噪语音信号 y 可以表示为:

$$y(t) = s(t) + n(t) \quad (1)$$

纯净语音信号 s 可由含噪语音信号 y 和估计线性评估器 H 得到:

$$\hat{s} = Hy \quad (2)$$

估计误差为:

$$r = \hat{s} - s = (H - I)s + Hn = r_s + r_n \quad (3)$$

其中, r_s 代表信号失真; r_n 表示残余噪声。

定义信号失真能量为 $\bar{r}_s^2$, 残余噪声能量为 $\bar{r}_n^2$, 即:

$$\bar{r}_s^2 = E[r_s^T r_s] = \text{tr}(E[r_s^T r_s]) = \sigma^2 \text{tr}(H H^T) \quad (4)$$

$$\begin{aligned} \bar{r}_n^2 &= E[r_n^T r_n] = \text{tr}(E[r_n^T r_n]) \\ &= \text{tr}((H - I)R_s(H - I)^T) \end{aligned} \quad (5)$$

其中, σ^2 为一常数; R_s 为纯净语音信号的协方差矩阵; I 为单位矩阵。

把残余噪声能量控制在某一阈值之下同时保证信号失真最小, 即:

$$\min \bar{r}_s^2 \text{ subject to: } \bar{r}_n^2 \leq K\sigma^2 \quad (6)$$

在以上条件约束下, 最优线性评估器 H 可以通过求下式的极值得:

$$\begin{aligned} \ell(H, \mu) &= \bar{r}_s^2 + \mu(\bar{r}_n^2 - K\sigma^2) \\ &= \text{tr}((H - I)R_s(H - I)^T) \\ &\quad + \mu(\sigma^2 \text{tr}(H H^T) - K\sigma^2) \end{aligned} \quad (7)$$

其中, $\mu \geq 0$ 为拉格朗日系数。

通过对上式关于 H 求偏导, 可以得到最优线性评估器:

$$H_{\text{opt}} = R_s(R_s + \mu R_n)^{-1} \quad (8)$$

其中, $R_n = \sigma^2 I$ 为噪声的协方差矩阵。

对于高斯白噪声来说, 最优线性评估器可以表示为:

$$H_{\text{opt}} = U\Delta_s(\Delta_s + \mu\Delta_n)^{-1}U^T \quad (9)$$

其中, U 为特征向量矩阵; 而 Δ_s 为由含噪语音信号协方差矩阵特征值分解得到的特征值矩阵; Δ_n 为由噪声协方差矩阵特征值分解得到的特征值矩阵。上式是白噪声^[15]环境下的最优线性评估器, 但却不适用于更具一般性的有色噪声。文献[6]利用广义特征值分解 (GEVD) 对有色噪声的情况给出了一种最优估计。

$$\begin{aligned} H_{\text{opt}} &= \mathbf{R}_n \mathbf{V} \mathbf{A}_s (\mathbf{A}_s + \mu \mathbf{I})^{-1} \mathbf{V}^T \\ &= \mathbf{V}^{-T} \mathbf{A}_s (\mathbf{A}_s + \mu \mathbf{I})^{-1} \mathbf{V}^T \end{aligned} \quad (10)$$

4 特征值分量置换原理

4.1 特征值分量置换算法

图 2 为特征值置换方法的流程图。首先对含噪语音进行加窗和分帧,考虑到语音信号前后的连续性,设置了帧移,其长度是帧长的一半。经过上述处理后的含噪语音可以视作一个矩阵,矩阵的每一列代表不同的语音帧。接下来为了求取 $\mathbf{R}_n^{-1} \mathbf{R}_s$ 协方差矩阵,需要估计噪声和含噪语音的协方差矩阵。对于估计的协方差矩阵 $\mathbf{R}_n^{-1} \mathbf{R}_s$,经过广义特征值分解、特征值分量置换,以及广义逆矩阵变换,可以求得线性评估器 $\mathbf{H}_{\text{opt}}$ 。再经过一个反叠加过程,就可以得到合成的增强语音。

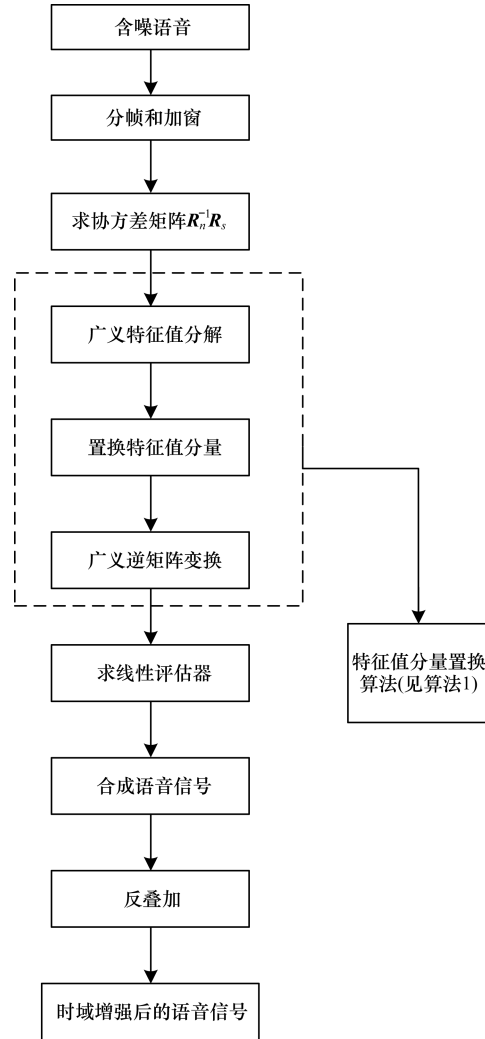


图 2 基于特征值置换的子空间语音增强方法流程

传统的子空间方法中,通过下式同时对角化噪声和纯净语音的协方差矩阵求取广义特征值分量,有:

$$\mathbf{V}^T \mathbf{R}_s \mathbf{V} = \mathbf{A}_s, \mathbf{V}^T \mathbf{R}_n \mathbf{V} = \mathbf{I} \quad (11)$$

其中, $\mathbf{V}$ 和 $\mathbf{A}_s$ 分别为噪声协方差矩阵 $\mathbf{R}_n$ 和纯净语音协方差矩阵 $\mathbf{R}_s$ 的广义特征值矩阵和广义特征向量。进而借助式(10)估计线性评估器 $\mathbf{H}_{\text{opt}}$ 。然而,利用这种方法没有考虑到叠加噪声的影响。实际上,噪声会使含噪语音和纯净语音的特征向量投影方向严重偏离,从而造成增强语音相关性降低。不同于传统方法,基于特征值置换的子空间方法首先对协方差矩阵 $\mathbf{R}_n^{-1} \mathbf{R}_s$ 进行广义特征值分解,然后按特征值大小排序。为提高增强语音的相关性,对特征值分量进行置换。由于通常大特征值分量主要包含语音信号,而小特征值分量主要包含噪声^[7],用相邻大特征值分量置换小特征值分量,能够有效抑制噪声,提高语音听觉质量。

经特征值分量置换后,广义特征值分量由 $\mathbf{V}$ 和 $\mathbf{A}_s$ 变为 $\mathbf{V}_s$ 和 $\mathbf{A}$ 。由于特征值分量置换后的特征向量矩阵 $\mathbf{V}_s$ 可能不再可逆,引进广义逆变换^[16]。广义逆变换指出,对于常规不可逆矩阵,通过广义逆运算,也可以求得其逆矩阵。最后求得线性评估器为:

$$\mathbf{H}_{\text{opt}} = \mathbf{V}_s^{-T} \mathbf{A} (\mathbf{A} + \mu \mathbf{I})^{-1} \mathbf{V}_s^T \quad (12)$$

其中, $\mu \geq 0$ 为拉格朗日系数^[7]。

合成增强后的语音信号 x 为:

$$x = \mathbf{H}_{\text{opt}} y = \mathbf{V}_s^{-T} \mathbf{A} (\mathbf{A} + \mu \mathbf{I})^{-1} \mathbf{V}_s^T y \quad (13)$$

算法 特征值分量置换算法

(1) 广义特征值分解:

$$[\mathbf{V}, \mathbf{D}] = \text{evd}(\mathbf{R}_n^{-1} \mathbf{R}_s);$$

(2) 自动确定阈值 S :

$$P = \sum 10 \lg \sum \frac{\lambda_i}{N};$$

while($-5 < P < -0.5$) $S = 0.001$;

while($-0.5 < P < 7$) $S = 0.0003$;

(3) 特征值分量按从大到小排序:

while($t > 0$)

$$\mathbf{V}[:, t] = \mathbf{V}[:, \text{index}(m)];$$

$$\mathbf{D}[t] = \mathbf{D}[m];$$

$$\mathbf{V}[:, t] = \mathbf{V}[:, \text{index}(m)];$$

end

(4) 置换特征值分量:

$$i = \frac{t}{2};$$

while($i > 0 \&\& (\mathbf{D}[2i] - \mathbf{D}[2i-1] < S)$)

$$\mathbf{V}[:, 2i-1] = \mathbf{V}[:, 2i];$$

$$\mathbf{D}[:, 2i-1] = \mathbf{D}[:, 2i];$$

end

其中, $\sum \lambda_i$ 表示前 N 个大于 0 的特征值之和; N 表示含噪语音信号的帧长; t 表示特征值分解后的特征值个数; m 为特征值分解后特征值的序号。

特征值置换(ES)方法用相邻大特征值分量置换小特征值分量,和传统的内嵌式预白化子空间(GKLT)方法相比,可以求得更精确的线性评估器,进而合成增强效果更好的语音信号。

4.2 特征值置换阈值 S 的估计

将特征值分量按从小到大排序后,相邻的特征值的相近程度用 $S_i = |\lambda_i - \lambda_{i-1}|$ 表示, λ_i 表示含噪语音协方差矩阵经特征值分解并按从小到大排序后第 i 个特征值。以 $S = 0.05$ 为例,它表示如果2个特征值距离如果在0.05之内,用大的特征值分量置换小的特征值分量。如果在0.05之外,则不做任何处理。实验测得,当 S 在0.000 1~0.07之间变化时,可获得较好的语音增强效果。 S 值过小,在低信噪比情况下仍会有大量残余噪声, S 值过大,在高信噪比环境下会造成严重的语音失真。因此, S 的选择非常重要,实际中 S 的选择与信噪比有关,这是因为,在低信噪比情况下,噪声信息比较少,取较大的 S 值可以用相对更大的特征值分量替代小的特征值分量,可以更好地抑制噪声,在高信噪比环境下,噪声信息比较少,取较大的 S 值会引起语音失真。

在传统的子空间方法中,信噪比用 $SNR \approx$ 前 N 个特征值之和/帧长来估计^[7],通过此方法可以估计出一个未知含噪语音的信噪比。借助这种关系,根据信噪比来估计 S 值:

$$\begin{cases} S = 0.001 & -5 < SNR < -0.5 \\ S = 0.0003 & -0.5 < SNR < 7 \end{cases} \quad (14)$$

$$SNR = \frac{\sum_{i=1}^N \lambda_i}{N} \quad (15)$$

其中, $\sum \lambda_i$ 表示前 N 个大于0的特征值之和; N 表示含噪语音信号的帧长; $-5 \sim -0.5$ 表示在低信噪比环境下, $-0.5 \sim 7$ 表示在高信噪比环境中。

5 仿真实验与结果分析

为了验证 ES 方法的语音增强效果,选择 NOISEX-92^[14] 噪声库中的4种不同噪声信号,分别为白噪声、粉红噪声、F16 战斗机舱噪声和 hfchannel 高频信道噪声,以 NOISEUS 库^[17] 的 sp01~sp30 文件作为基线纯净语音,每种噪声以不同的信噪比叠加到纯净语音,形成测试含噪语音。含噪语音的采样频率均为8 kHz,采样长度为21 000。矩形窗的长度为300,帧移为50%。

图3比较了 GKLT 方法和 ES 方法时域增强后的语谱图。含噪语音 sp02, $SNR = -5$ dB,噪声类型为 white 噪声。从图3(c)和图3(d)可以看出, GKLT 方法仍存在大量残余噪声。它和别的子空间方法一样,只是去掉了对应噪声的子空间部分,而对

应语音信号的子空间部分的噪声没有做任何处理,这将影响语音信号听觉质量及信噪比得分。

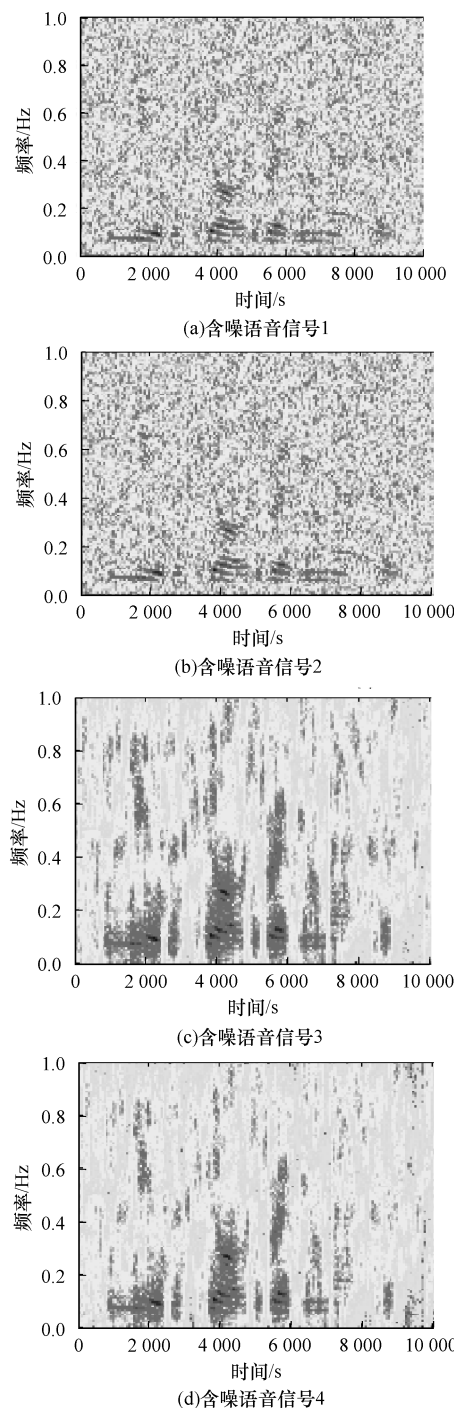


图3 改进前后语谱对比

图4为经 GKLT 和 ES 方法处理后的增强语音的帧相关度对比图,含噪语音 sp02, $SNR = -5$ dB,噪声类型为 white 噪声。横坐标为相关度计算用到的前后帧序号。在绝大多数情况下,ES 比 KLT 的帧相关度有显著提高。GKLT 方法的实质是对含噪语音信号进行 V^T 变换, V^T 即为将含噪信号经投影得到的信号子空间部分。通过 ES 方法置换相近的

特征值分量后,得到的特征向量矩阵 V_s 存在相同的列向量,经过转置变换得到存在相同行向量的矩阵 V_s^T 。语音信号是分帧处理的,那么含噪语音信号经过 V_s^T 变换得到的增强后的语音信号 $V_s^T y$ 存在相同数值的列向量。在出现的相同数值中,其中一个原有的,而另一个是用大特征值分量取代小特征值分量得到的,所以语音帧相关系数会变大(假设语音帧列向量零均值,相关系数即为对应 2 帧列向量的点积),即语音信号相关性变大。

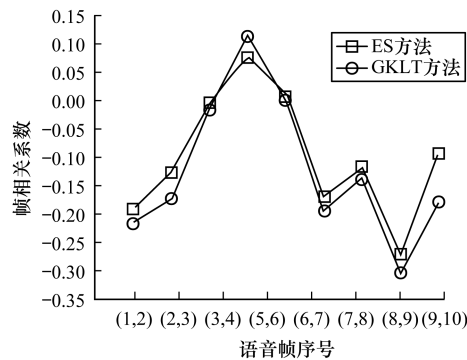


图 4 改进前后语音帧相关系数对比

下面给出作为参考的其他经典的语音增强方法: GKLT^[17], mss-map^[18], SSPARA98^[19], MMSE2004^[20]。

图 5 和图 6 为各种语音增强方法的增强效果在白噪声环境下的对比,噪声类型为 white 噪声。PESQ (Perceptual Evaluation of Speech Quality)^[21] 反映增强语音的听觉质量,segSNR (segmental SNR)^[22] 反映噪声的去除程度。从图中可以明显看出,ES 方法的 PESQ 和 segSNR 得分在各种信噪比下都优于 GKLT 方法和其他参考语音增强方法。说明在不同信噪比下,ES 方法都提高了语音听觉质量并且可以大程度去除语音信号中的残余噪声。图 7 为 PESQ 和 segSNR 得分对比,噪声类型为 white 噪声。从图中可以看出,与参考方法相比,ES 方法能够显著提高 segSNR 得分,具有很好的降噪性能。同时该方法还能适当提高 PESQ 性能,保持较好的语音听觉质量。表 1 为各种语音增强方法在粉红噪声环境下强增效果的对比。从表 1 可以看出,相比其他方法,ES 方法的 PESQ 及 segSNR 得分都有所提高,该结果表明

所提方法在有色噪声环境下具有较好的语音增强效果。

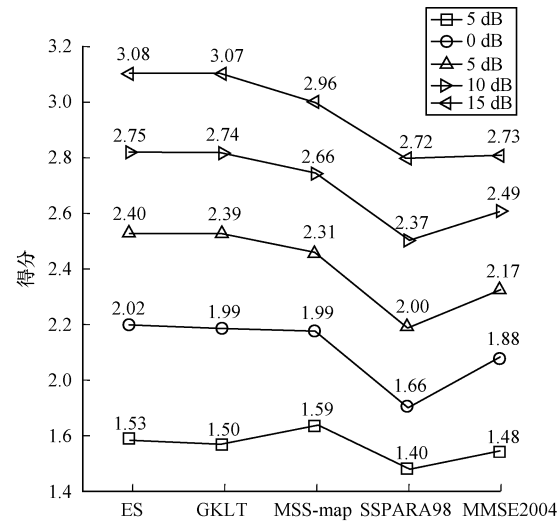


图 5 各种语音增强方法的 PESQ 得分对比

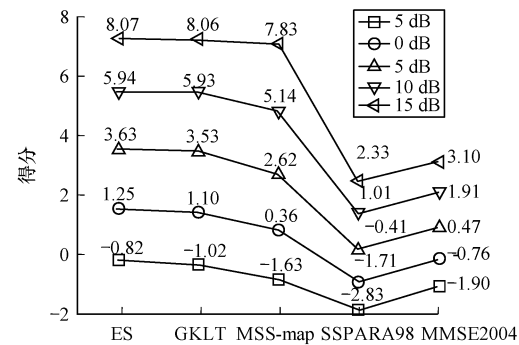


图 6 各种语音增强方法的 segSNR 得分对比

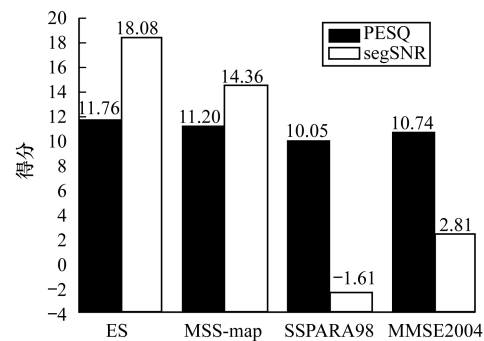


图 7 各种语音增强算法在不同信噪比下的总得分对比

表 1 粉红噪声环境下各种语音增强算法效果对比

方法	-5 dB		0 dB		5 dB		10 dB		15 dB	
	PESQ	segSNR	PESQ	segSNR	PESQ	segSNR	PESQ	segSNR	PESQ	segSNR
ES 方法	1.686 9	-1.224 3	2.078 6	0.781 4	2.415 6	3.098 7	2.766 5	5.379 1	3.126 0	7.462 6
GKLT 方法	1.672 3	-1.276 4	2.068 7	0.775 3	2.412 1	3.097 2	2.754 8	5.378 1	3.125 8	7.462 6
Mss-map 方法	1.751 7	-1.832 6	2.078 1	0.071 9	2.405 2	2.437 6	2.745 0	5.085 1	3.058 3	8.028 7
SSPARA98 方法	1.550 3	-2.737 6	1.836 7	-1.524 3	2.134 4	-0.222 6	2.502 7	1.160 2	2.771 5	2.290 6
MMSE2004 方法	1.718 9	-1.792 6	1.992 1	-0.695 8	2.303 0	0.634 1	2.569 4	1.965 2	2.789 4	3.132 5

表 2 为高频信道噪声环境下各种语音增强方法在不同信噪比下增强效果的对比。可以看出:在不同信噪比下,ES 方法的 segSNR 得分均优于其他的参考方法,表明 ES 方法能够有效去除噪声。并且 ES 方法的 PESQ 得分也优于其他的比较方法,说明 ES 方法

还可以有效提高语音听觉效果。表 3 为 F16 战斗机驾驶舱噪声环境下语音增强效果的对比。从中可以明显看出,ES 方法在各信噪比下的综合性能都优于 GKLT 方法以及其他参考方法,低信噪比环境下尤为明显,表明该方法适合工作于复杂的背景噪声环境。

表 2 高频信道噪声环境下各种语音增强算法效果对比

方法	- 5 dB		0 dB		5 dB		10 dB		15 dB	
	PESQ	segSNR	PESQ	segSNR	PESQ	segSNR	PESQ	segSNR	PESQ	segSNR
ES 方法	1.485 7	- 1.248 7	1.954 2	0.793 6	2.346 0	3.272 7	2.678 2	5.699 8	3.010 2	7.890 0
GKLT 方法	1.462 9	- 1.405 1	1.943 3	0.704 3	2.333 3	3.221 8	2.675 8	5.687 1	3.007 8	7.866 3
Mss-map 方法	1.540 9	- 1.913 9	1.929 8	0.056 9	2.284 6	2.305 8	2.636 3	4.797 2	2.931 6	7.536 5
SSPARA98 方法	1.388 8	- 2.826 4	1.689 3	- 1.697 0	2.011 8	- 0.385 9	2.366 0	1.033 5	2.700 6	2.262 1
MMSE2004 方法	1.543 1	- 1.960 5	1.859 4	- 0.879 2	2.148 4	0.365 9	2.486 7	1.849 5	2.731 1	2.994 4

表 3 F16 噪声环境下各种语音增强算法效果对比

方法	- 5 dB		0 dB		5 dB		10 dB		15 dB	
	PESQ	segSNR	PESQ	segSNR	PESQ	segSNR	PESQ	segSNR	PESQ	segSNR
ES 方法	1.655 1	- 1.220 7	2.069 8	0.733 5	2.432 4	2.990 2	2.760 6	5.281 3	3.127 9	7.424 3
GKLT 方法	1.632 4	- 1.306 1	2.052 8	0.678 8	2.422 3	2.976 7	2.755 0	5.268 5	3.126 9	7.423 1
Mss-map 方法	1.738 7	- 1.799 8	2.083 8	0.166 2	2.434 9	2.536 7	2.786 6	5.146 2	3.105 3	8.070 7
SSPARA98 方法	1.649 6	- 2.594 0	1.748 6	- 1.414 8	2.206 8	- 0.063 6	2.560 2	1.194 3	2.827 5	2.362 3
MMSE2004 方法	1.742 6	- 1.874 9	1.821 0	- 0.741 0	2.348 5	0.653 6	2.623 1	1.904 4	2.826 9	3.073 4

6 结束语

本文提出一种基于特征值置换的子空间方法。该方法通过特征值置换改善增强语音的相关度,进而提高语音增强效果。实验结果表明,该方法适用于稳态噪声和色噪声,在低信噪比环境下也具有较好的语音增强性能。下一步将继续改进该方法,提高特征值分量置换阈值的估值准确性,在 ES 方法框架中融合更多的语音及噪声的先验知识。

参考文献

[1] Zhang Yi,Zhao Yunxin. Real and Imaginary Modulation Spectral Subtraction for Speech Enhancement [J]. Speech Communication,2013,55(4) :509-522.

[2] Boll S F. Suppression of Acoustic Noise in Speech Using Spectral Subtraction [J]. IEEE Transactions on Acoustics,Speech and Signal Processing,1979,27(2) :113-120.

[3] 陈欢,邱晓晖.改进谱减法语音增强算法的研究[J].计算机技术与发展,2014,(4) :69-71.

[4] Wiener N. Extrapolation,Interpolation,and Smoothing of Stationary Time Series[M]. New York,USA:Wiley,1949.

[5] El-Fattah M A A,Dessouky M I. Speech Enhancement with an Adaptive Wiener Filters [J]. International Journal of Speech Technology,2014,17(1) :53-64.

[6] 郭利华,马建芬.具有高可懂度的改进的维纳滤波的语音增强算法[J].计算机应用与软件,2014,31(11) :155-157.

[7] Loizou Y H. A Generalized Subspace Approach for Enhancing Speech Corrupted by Colored Noise [J], IEEE Transactions on Speech and Audio Processing,

2003,11(4) :334-341.

[8] 崔晓.拉格朗日乘子受控子空间语音增强[J].计算机工程与应用,2014,50(9) :182-185.

[9] Saadoune A, Amrouche A, Selouani S A. Perceptual Subspace Speech Enhancement Using Variance of the Reconstruction Error [J]. Digital Signal Processing, 2014,24(1) :187-196.

[10] 贾海蓉,张雪英,白静.联合听觉掩蔽效应的子空间语音增强算法[J].计算机工程,2011,37(8) :259-261.

[11] Dendrinis M,Bakamidis S,Garayannis G. Speech Enhancement from Noise: A Regenerative Approach[J]. Speech Communication,1991,10(2) :45-57.

[12] Ephraim Y, Malah D. Speech Enhancement Using a Minimum Mean Square Error Short-time Spectral Amplitude Estimator[J]. IEEE Transactions on Acoustics, Speech,and Signal Processing,1984,32(6) :1109-1121.

[13] Jensen S H,Hansenn P C,Hansen S D,et al. Reduction of Broad-band Noise in Speech by Truncated QSVD[J]. IEEE Transactions on Speech Audio Processing, 1995,3(6) :439-448.

[14] Rezayee A,Gsxot S. An Adaptive KLT Approach for Speech Enhancement[J]. IEEE Transactions on Speech and Audio Processing,2001,9(2) :87-95.

[15] Hermus K,Wambacq P, Hamme H V. A Review of Signal Subspace Speech Enhancement and Its Application to Noise Robust Speech Recognition [J]. EURASIP Journal on Advances in Signal Processing, 2007,(1) :1-15.

[16] Chang Jing,Li Zhibin. The Generalized Inverse Eigenvalue Problem for Generalized Periodic Jacobi Matrices [C]// Proceedings of International Conference on Computer Design and Applications. Washington D. C. , USA: IEEE Press,2010:354-356.

(下转第 282 页)

与此同时,本文算法抽取的第 553 帧、588 帧和 634 帧,反映出男女主角从相见到接吻以及后方男子经过的过程。而对比算法仅抽取第 584 帧和 634 帧作为关键帧,并未表现出后方男子经过的完整过程。除此之外,对比算法所得结果在第 279 帧后出现冗余。如表 6 所示,2 种算法的 FID 值和 VSE 值基本相同,但本文算法的结果获得了更好的用户主观评价。事实上,本文算法对物体运动有较好的鉴别能力,能准确表达镜头内物体的运动信息。

5 结束语

本文提出了一种基于镜头划分的视频关键帧抽取方法。实验结果表明,通过选取最优 Tsallis 熵指数, Jensen-Tsallis 距离在辨识视频图像颜色分布的距离上更具优势。基于 Jensen-Tsallis 距离的关键帧抽取算法能够有效、精确地检测出视频内含有运动物体的子镜头,从而得到更具代表性的关键帧序列。在今后的工作中,考虑到小波变换所得图像特征更接近于人们视觉方面对图像的感受,将尝试利用小波变换来处理视频图像,进行关键帧抽取,同时也将考虑提取视频中的运动要素来选取关键帧。

参考文献

- [1] Truong B T, Venkatesh S. Video Abstraction: A Systematic Review and Classification [J]. ACM Transactions on Multimedia Computing, Communications, and Applications, 2007, 3(1).
- [2] Money A G, Agius H. Video Summarization: A Conceptual Framework and Survey of the State of the Art [J]. Journal of Visual Communication and Image Representation, 2008, 19(2): 121-143.
- [3] 丁洪丽, 陈怀新. 基于镜头内容变化率的关键帧提取算法 [J]. 计算机工程, 2009, 35(13): 225-227.
- [4] Mentzelopoulos M, Psarrou A. Key-frame Extraction Algorithm Using Entropy Difference [C]//Proceedings of the 6th ACM SIGMM International Workshop on Multimedia Information Retrieval. New York, USA: ACM Press, 2004: 39-45.
- [5] Cernekova Z, Pitas I, Nikou C. Information Theory-based Shot Cut/Fade Detection and Video Summarization [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2006, 16(1): 82-91.
- [6] Omidyeganeh M, Ghaemmaghami S, Shirmohammadi S. Video Keyframe Analysis Using a Segment-based Statistical Metric in a Visually Sensitive Parametric Space [J]. IEEE Transactions on Image Processing, 2011, 20(10): 2730-2737.
- [7] Janvier B, Bruno E, Pun T, et al. Information-theoretic Temporal Segmentation of Video and Applications: Multiscale Keyframes Selection and Shot Boundaries Detection [J]. Multimedia Tools and Applications, 2006, 30(3): 273-288.
- [8] He Yun, Hamza A B, Krim H. A Generalized Divergence Measure for Robust Image Registration [J]. IEEE Transactions on Signal Processing, 2003, 51(5): 1211-1220.
- [9] Xu Qing, Wang Pengcheng, Long Bin, et al. Selection and 3D Visualization of Video Key Frames [C]//Proceedings of IEEE International Conference on Systems Man and Cybernetics. Washington D. C., USA: IEEE Press, 2010: 52-59.
- [10] Lienhart R, Peiffer S, Effelsberg W. Video Abstracting [J]. Communications of the ACM, 1997, 40(12): 54-62.
- [11] Harvda J, Charvát F. Quantification Method of Classification Processes. Concept of Structural α -entropy [J]. Kybernetika, 1967, 3: 30-35.
- [12] Cover T M, Thomas J A. Elements of Information Theory [M]. [S. l.]: John Wiley & Sons, 2012.
- [13] Chang Hyun-sung, Sull S, Lee S U. Efficient Video Indexing Scheme for Content-based Retrieval [J]. IEEE Transactions on Circuits and Systems for Video Technology, 1999, 9(8): 1269-1279.
- [14] Liu Tiecheng, Kender J R. Computational Approaches to Temporal Sampling of Video Sequences [J]. ACM Transactions on Multimedia Computing Communications and Applications, 2007, 3(2): 601-604.
- [15] Liu Tianming, Zhang Hongjiang, Qi Feihu. A Novel Video Key-frame-extraction Algorithm Based on Perceived Motion Energy Model [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2003, 13(10): 1006-1013.
- [17] Hu Yi, Loizou P C. Subjective Evaluation and Comparison of Speech Enhancement [J]. Speech Communication, 2007, 49(7): 588-601.
- [18] Yang Lu, Loizou P C. Estimators of the Magnitude-Squared Spectrum and Methods for Incorporating SNR Uncertainty [J]. IEEE Transactions on Audio, Speech and Language Processing, 2011, 19(5): 1123-1137.
- [19] Sim B L, Tong Y C, Chang J S, et al. A Parametric Formulation of the Generalized Spectral Subtraction Method [J]. IEEE Transactions on Speech And Audio Processing, 1988, 6(4): 328-337.
- [20] Cohen I. Speech Enhancement Using a Noncausal A Priori SNR Estimator [J]. IEEE Signal Processing Letters, 2004, 11(9): 725-728.
- [21] Rix A W, Beerends J G, Hollier M P, et al. Perceptual Evaluation of Speech Quality (PESQ): An Objective Method for End-to-End Speech Quality Assessment of Narrow-band Telephone Networks and Speech Codecs [C]//Proceedings of International Conference on Acoustics, Speech and Signal Processing. New York, USA: ACM Press, 2001: 749-752.
- [22] Loizou P C. Speech Enhancement: Theory and Practice [M]. Boca Raton, USA: CRC Press, 2007.

编辑 顾逸斐

编辑 顾逸斐

(上接第 277 页)