Stereo-Based Stochastic Noise Compensation Based on Trajectory GMMs

Author : Heiga Zen, Yoshihiko Nankaku, Keiichi Tokuda

Professor:陳嘉平

Reporter :吳柏鋒

This paper proposes a novel stereo-based stochastic noise compensation technique based on trajectory GMMs. Although the GMM-based noise compensation techniques such as SPLICE work effective, their performance sometimes degrades due to the inappropriate dynamic characteristics caused by the frame-by-frame mapping. While the use of dynamic feature constraints on the mapping stage can alleviate this problem, it also introduces an inconsistency between training and mapping. The recently proposed trajectory GMM-based feature mapping technique can solve this inconsistency while keeping the benefits of the use of dynamic features, and offers an entire sequence-level transformation rather than the frame-by-frame mapping. Results from a noise compensation experiment on the AURORA-2 task show that the proposed trajectory GMM-based noise compensation technique outperforms the conventional ones.

This paper proposes a novel stereo-based stochastic noise compensation technique based on trajectory GMMs. Although the GMM-based noise compensation techniques such as SPLICE work effective, their performance sometimes degrades due to the inappropriate dynamic characteristics caused by the frame-by-frame mapping. While the use of dynamic feature constraints on the mapping stage can alleviate this problem, it also introduces an inconsistency between training and mapping. The recently proposed trajectory GMM-based feature mapping technique can solve this inconsistency while keeping the benefits of the use of dynamic features, and offers an entire sequence-level transformation rather than the frame-by-frame mapping. Results from a noise compensation experiment on the AURORA-2 task show that the proposed trajectory GMM-based noise compensation technique outperforms the conventional ones.

This paper proposes a novel stereo-based stochastic noise compensation technique based on trajectory GMMs. Although the GMM-based noise compensation techniques such as SPLICE work effective, their performance sometimes degrades due to the inappropriate dynamic characteristics caused by the frame-by-frame mapping. While the use of dynamic feature constraints on the mapping stage can alleviate this problem, it also introduces an inconsistency between training and mapping. The recently proposed trajectory GMM-based feature mapping technique can solve this inconsistency while keeping the benefits of the use of dynamic features, and offers an entire sequence-level transformation rather than the frame-by-frame mapping. Results from a noise compensation experiment on the AURORA-2 task show that the proposed trajectory GMM-based noise compensation technique outperforms the conventional ones.

This paper proposes a novel stereo-based stochastic noise compensation technique based on trajectory GMMs. Although the GMM-based noise compensation techniques such as SPLICE work effective, their performance sometimes degrades due to the inappropriate dynamic characteristics caused by the frame-by-frame mapping. While the use of dynamic feature constraints on the mapping stage can alleviate this problem, it also introduces an inconsistency between training and mapping. The recently proposed trajectory GMM-based feature mapping technique can solve this inconsistency while keeping the benefits of the use of dynamic features, and offers an entire sequence-level transformation rather than the frame-by-frame mapping. Results from a noise compensation experiment on the AURORA-2 task show that the proposed trajectory GMM-based noise compensation technique outperforms the conventional ones.

This paper proposes a novel stereo-based stochastic noise compensation technique based on trajectory GMMs. Although the GMM-based noise compensation techniques such as SPLICE work effective, their performance sometimes degrades due to the inappropriate dynamic characteristics caused by the frame-by-frame mapping. While the use of dynamic feature constraints on the mapping stage can alleviate this problem, it also introduces an inconsistency between training and mapping. The recently proposed trajectory GMM-based feature mapping technique can solve this inconsistency while keeping the benefits of the use of dynamic features, and offers an entire sequence-level transformation rather than the frame-by-frame mapping. Results from a noise compensation experiment on the AURORA-2 task show that the proposed trajectory GMM-based noise compensation technique outperforms the conventional ones.

簡介

- 以GMMs為基礎的立體聲連續隨機特徵對映技術已應用到噪音補償上

- 此對映技術分為三個階段:

(1)由一GMMs集合計算噪音和乾淨特徵間的聯合機率密度函式

(2)從給定的一噪音，可從聯合機率密度函式估算出乾淨特徵的條件機率密度函式

(3)根據條件機率密度函式何時造成最小的MMSE，來決定乾淨特徵的對映

簡介

- 此對映技術在音框與音框間的動態特性對映上，效果較不佳

- 因此又使用以GMMs為基礎且對動態特徵作限制的特徵對映技術，來改善此問題。但在訓練與對映上仍然有可能會產生不一致

簡介

- 最近提出的軌跡模型，此模型又稱為軌跡HMMs，可以在不需要任何額外參數情況下，得到在音框方面假設獨立狀態條件下的輸出機率和分段的HMM統計常數

- 以這概念為基礎，提出以GMMs軌跡為基礎的連續隨機特徵對映技術

MFCC向量序列

$$\boldsymbol{x} = \left[\boldsymbol{x}_1^\top, \ldots, \boldsymbol{x}_T^\top\right]^\top, \quad \boldsymbol{y} = \left[\boldsymbol{y}_1^\top, \ldots, \boldsymbol{y}_T^\top\right]^\top, \qquad (1)$$

其中 $\boldsymbol{x}_t$ 和 $\boldsymbol{y}_t$ 分別表示在第 t 個音框的<u>噪音</u>和<u>乾淨</u>MFCC靜態特徵向量且其維度為M，T 為音框總數

聯合機率密度函式(JPDF)

- 由軌跡GMMs獲得的 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 的聯合機率密度函式定義：

$$p(\boldsymbol{z} \mid \lambda) = \sum_{\forall \boldsymbol{q}} P(\boldsymbol{q} \mid \lambda)\, p(\boldsymbol{z} \mid \boldsymbol{q}, \lambda), \tag{2}$$

其中

$$P(\boldsymbol{q} \mid \lambda) = \prod_{t=1}^{T} c_{q_t}$$ （ 即混合事前機率的乘積 ）

$$p(\boldsymbol{z} \mid \boldsymbol{q}, \lambda) = \mathcal{N}\left(\boldsymbol{z} ; \bar{\boldsymbol{z}}_{\boldsymbol{q}}, \boldsymbol{P}_{\boldsymbol{q}}\right) \tag{3}$$

$$\boldsymbol{R}_{\boldsymbol{q}} \bar{\boldsymbol{z}}_{\boldsymbol{q}} = \boldsymbol{r}_{\boldsymbol{q}} \tag{4}$$

其中 $\boldsymbol{R}_{\boldsymbol{q}} = \boldsymbol{W}^{\top} \boldsymbol{\Omega}_{\boldsymbol{q}} \boldsymbol{W} = \boldsymbol{P}_{\boldsymbol{q}}^{-1}$

$$\boldsymbol{r}_{\boldsymbol{q}} = \boldsymbol{W}^{\top} \boldsymbol{\Omega}_{\boldsymbol{q}} \boldsymbol{\mu}_{\boldsymbol{q}}$$

$$\boldsymbol{\mu}_{\boldsymbol{q}} = \left[\boldsymbol{\mu}_{q_1}^{\top}, \ldots, \boldsymbol{\mu}_{q_T}^{\top}\right]^{\top} \quad \boldsymbol{\mu}_i = \left[\boldsymbol{\mu}_i^{(\boldsymbol{x})\top}, \boldsymbol{\mu}_i^{(\boldsymbol{y})\top}\right]^{\top}$$

$$\boldsymbol{\Omega}_{\boldsymbol{q}} = \operatorname{diag}\left[\boldsymbol{\Omega}_{q_1}, \ldots, \boldsymbol{\Omega}_{q_T}\right], \quad \boldsymbol{\Omega}_i = \begin{bmatrix} \boldsymbol{\Omega}_i^{(\boldsymbol{xx})} & \boldsymbol{\Omega}_i^{(\boldsymbol{xy})} \\ \boldsymbol{\Omega}_i^{(\boldsymbol{yx})} & \boldsymbol{\Omega}_i^{(\boldsymbol{yy})} \end{bmatrix}$$

對映(Mapping)

- 條件機率(重寫式子(3))

$$p(z \mid q, \lambda) = \mathcal{N}\left(\begin{bmatrix} x \\ y \end{bmatrix} ; \begin{bmatrix} \bar{x}_q \\ \bar{y}_q \end{bmatrix}, \begin{bmatrix} P_q^{(xx)} & P_q^{(xy)} \\ P_q^{(yx)} & P_q^{(yy)} \end{bmatrix} \right) \quad (5)$$

其中

$$P_q^{(xx)} = C_q^{(xx)^{-1}}, \quad P_q^{(yy)} = C_q^{(yy)^{-1}},$$

$$P_q^{(xy)} = -R_q^{(xx)^{-1}} R_q^{(xy)} C_q^{(yy)^{-1}} = P_q^{(yx)^{\top}}$$

$$\bar{x}_q = P_q^{(xx)} \left(r_q^{(x)} - R_q^{(xy)} R_q^{(yy)^{-1}} r_q^{(y)} \right)$$

$$\bar{y}_q = P_q^{(yy)} \left(r_q^{(y)} - R_q^{(yx)} R_q^{(xx)^{-1}} r_q^{(x)} \right)$$

$$C_q^{(xx)} = R_q^{(xx)} - R_q^{(xy)} {R_q^{(yy)}}^{-1} R_q^{(yx)}$$

$$C_q^{(yy)} = R_q^{(yy)} - R_q^{(yx)} {R_q^{(xx)}}^{-1} R_q^{(xy)}$$

$$R_q^{(xx)} = {W^{(x)}}^\top \Omega_q^{(xx)} W^{(x)},$$

$$R_q^{(yy)} = {W^{(y)}}^\top \Omega_q^{(yy)} W^{(y)},$$

$$R_q^{(xy)} = {W^{(x)}}^\top \Omega_q^{(xy)} W^{(y)} = {R_q^{(yx)}}^\top$$

$$\Omega_q^{(xx)} = \mathrm{diag}\left[\Omega_{q1}^{(xx)}, \ldots, \Omega_{qT}^{(xx)}\right],$$

$$\Omega_q^{(yy)} = \mathrm{diag}\left[\Omega_{q1}^{(yy)}, \ldots, \Omega_{qT}^{(yy)}\right],$$

$$\Omega_q^{(xy)} = \mathrm{diag}\left[\Omega_{q1}^{(xy)}, \ldots, \Omega_{qT}^{(xy)}\right] = \Omega_q^{(yx)}$$

$$
\begin{aligned}
\boldsymbol{r}_{\boldsymbol{q}}^{(x)} &= W^{(x)^{\top}} \left(\boldsymbol{\Omega}_{\boldsymbol{q}}^{(xx)} \boldsymbol{\mu}_{\boldsymbol{q}}^{(x)} + \boldsymbol{\Omega}_{\boldsymbol{q}}^{(xy)} \boldsymbol{\mu}_{\boldsymbol{q}}^{(y)} \right), \\
\boldsymbol{r}_{\boldsymbol{q}}^{(y)} &= W^{(y)^{\top}} \left(\boldsymbol{\Omega}_{\boldsymbol{q}}^{(yy)} \boldsymbol{\mu}_{\boldsymbol{q}}^{(y)} + \boldsymbol{\Omega}_{\boldsymbol{q}}^{(yx)} \boldsymbol{\mu}_{\boldsymbol{q}}^{(x)} \right), \\
\boldsymbol{\mu}_{\boldsymbol{q}}^{(x)} &= \left[\boldsymbol{\mu}_{q_1}^{(x)^{\top}}, \ldots, \boldsymbol{\mu}_{q_T}^{(x)^{\top}} \right]^{\top}, \\
\boldsymbol{\mu}_{\boldsymbol{q}}^{(y)} &= \left[\boldsymbol{\mu}_{q_1}^{(y)^{\top}}, \ldots, \boldsymbol{\mu}_{q_T}^{(y)^{\top}} \right]^{\top}.
\end{aligned}
$$

對映(Mapping)

- 在給定 x 和 λ 下，y 的條件機率:

$$p(y \mid x, \lambda) = \sum_{\forall q} \gamma_q \cdot p(y \mid x, q, \lambda)$$

其中 γ_q 是 q 的事後機率

$$p(y \mid x, q, \lambda) = \mathcal{N}\left(y ;\ \tilde{y}_q, \tilde{P}_q^{(yy)}\right)$$

$$\tilde{y}_q = \bar{y}_q + P_q^{(yx)} C_q^{(xx)} \left(x - \bar{x}_q\right)$$

$$\tilde{P}_q^{(yy)} = P_q^{(yy)} - P_q^{(yx)} C_q^{(xx)} P_q^{(xy)}$$

以最小均方差(MMSE)為基礎的對映

- 可以由乾淨的靜態MFCC向量序列 $\hat{y}$ 來估算MMSE:

$$\begin{aligned}\hat{y} &= E[y \mid x] = \int p(y \mid x, \lambda)\, y\, dy \\ &= \int \sum_{\forall q} \gamma_q \cdot p(y \mid x, q, \lambda)\, y\, dy \\ &= \sum_{\forall q} \gamma_q \cdot \tilde{y}_q,\end{aligned}$$

其中 $E[\cdot]$ 是平均值的期望值

實驗

- 使用**AURORA-2**語料庫

- AURORA-2是由一連續英文數字句所組成，並人工加入8種不同的SNR(-5dB, 0dB, 5 dB, 10dB, 15dB, 20dB, clean)到測試集A與B中

實驗

- 實驗中針對四種對映技術做比較:

  (1)**GMM-Static:**

  基本的GMM對映(MMSE對映+SPLICE)，應用在13維的靜態MFCC向量上

  (2)**GMM-Complete:**

  類似SPLICE，在13維的靜態MFCC向量加入動態特徵之後再進行以GMM為基礎的對映

實驗

(3)**GMM-Dynamic:**

13維的靜態MFCC向量在靜態和動態特徵之間具有限制下的GMM對映

(4)**Trajectory GMM:**

以軌跡GMM為基礎的對映對13維靜態MFCC向量作轉換

實驗結果(1)(2)

Aurora 2 Clean Training + GMM-Static	A					B					C			Overall	Percentage
	Subway	Babble	Car	Exhibition	Average	Restaurant	Street	Airport	Station	Average	Subway M	Street M	Average	Average	Improvement
Clean	98.86	99.00	98.96	99.23	99.01	98.86	99.00	98.96	99.23	99.01	99.14	98.97	99.06	99.02	-0.56%
20 dB	97.33	98.10	97.70	97.53	97.67	96.99	96.46	96.93	96.91	96.82	96.75	95.01	95.88	96.97	37.63%
15 dB	94.60	95.98	96.27	95.62	95.62	95.00	92.59	93.14	94.35	93.77	93.06	89.81	91.44	94.04	51.53%
10 dB	90.27	90.81	90.75	91.51	90.84	87.38	83.07	85.09	85.75	85.32	81.27	77.45	79.36	86.34	56.58%
5 dB	80.50	72.61	77.27	80.13	77.63	69.63	59.31	63.14	64.83	64.23	54.31	53.66	53.99	67.54	45.45%
0 dB	59.16	35.70	46.82	61.06	50.69	37.70	29.14	28.96	29.99	31.45	25.64	26.03	25.84	38.02	25.02%
-5 dB	26.90	4.05	16.13	32.12	19.80	7.55	10.01	3.13	1.88	5.64	11.54	12.85	12.20	12.62	4.52%
Average	84.37	78.64	81.76	85.17	82.49	77.34	72.11	73.45	74.37	74.32	70.21	68.39	69.30	76.58	
	48.77%	57.38%	53.71%	57.15%	54.69%	52.20%	27.53%	43.21%	42.23%	41.96%	11.95%	6.73%	9.33%		41.36%

Aurora 2 Clean Training + GMM-Complete	A					B					C			Overall	Percentage
	Subway	Babble	Car	Exhibition	Average	Restaurant	Street	Airport	Station	Average	Subway M	Street M	Average	Average	Improvement
Clean	98.96	99.00	98.96	99.23	99.04	98.96	99.00	98.96	99.23	99.04	99.14	98.97	99.06	99.04	1.31%
20 dB	97.88	98.31	98.06	97.84	98.02	97.48	97.64	97.82	97.87	97.70	97.42	97.22	97.32	97.75	53.55%
15 dB	96.75	96.83	96.96	96.88	96.86	96.28	96.46	96.36	95.53	96.16	94.96	94.89	94.93	96.19	69.46%
10 dB	93.37	94.23	94.93	93.64	94.04	91.71	91.57	91.05	90.71	91.26	87.66	86.43	87.05	91.53	72.89%
5 dB	88.12	84.89	88.04	86.27	86.83	80.47	78.96	79.54	78.25	79.31	70.53	68.95	69.74	80.40	66.78%
0 dB	69.76	53.63	63.88	70.32	64.40	50.51	51.09	50.52	46.13	49.56	39.58	41.57	40.58	53.70	43.89%
-5 dB	32.98	18.65	23.83	32.34	26.95	14.92	18.71	12.44	11.82	14.47	17.75	19.50	18.63	20.29	12.85%
Average	89.18	85.58	88.37	88.99	88.03	83.29	83.14	83.06	81.70	82.80	78.03	77.81	77.92	83.92	
	64.52%	71.22%	70.49%	68.18%	69.03%	64.75%	56.20%	63.76%	58.75%	61.13%	35.07%	34.53%	34.80%		59.72%

實驗結果(3)(4)

Aurora 2 Clean Training + GMM-Dynamic

	A					B					C			Overall	Percentage
	Subway	Babble	Car	Exhibition	Average	Restaurant	Street	Airport	Station	Average	Subway M	Street M	Average	Average	Improvement
Clean	98.93	99.00	98.96	99.23	99.03	98.93	99.00	98.96	99.23	99.03	99.14	98.97	99.06	99.04	0.75%
20 dB	98.16	98.19	98.27	98.21	98.21	97.45	97.49	98.03	97.22	97.55	97.54	97.37	97.46	97.79	55.32%
15 dB	96.65	97.19	97.55	96.95	97.09	95.67	95.77	95.76	94.57	95.44	95.39	94.44	94.92	95.99	68.34%
10 dB	94.11	94.59	95.59	93.98	94.57	90.08	90.15	90.01	89.76	90.00	89.32	86.91	88.12	91.45	73.22%
5 dB	88.73	84.85	88.70	86.58	87.22	78.26	77.09	77.54	75.29	77.05	75.65	70.56	73.11	80.33	67.11%
0 dB	71.48	52.24	65.05	68.13	64.23	46.27	46.70	46.47	41.10	45.14	43.75	42.20	42.98	52.34	42.43%
5 dB	35.46	17.26	24.16	32.46	27.34	12.59	18.14	14.94	11.20	14.22	18.27	19.20	18.74	20.37	12.96%
Average	89.83	85.41	89.03	88.77	88.26	81.55	81.44	81.56	79.59	81.03	80.33	78.30	79.31	83.58	
	66.65%	70.89%	72.16%	67.55%	69.63%	61.07%	51.77%	60.56%	54.00%	57.14%	41.87%	35.95%	38.91%		58.88%

Aurora 2 Clean Training + Trajectory GMM

	A					B					C			Overall	Percentage
	Subway	Babble	Car	Exhibition	Average	Restaurant	Street	Airport	Station	Average	Subway M	Street M	Average	Average	Improvement
Clean	98.93	99.00	98.96	99.23	99.03	98.93	99.00	98.96	99.23	99.03	99.14	98.97	99.06	99.04	0.75%
20 dB	97.97	98.16	98.18	98.40	98.18	97.39	97.58	97.61	97.16	97.44	97.67	97.43	97.55	97.76	54.73%
15 dB	96.81	97.16	97.46	96.79	97.06	95.18	95.98	95.05	94.63	95.21	96.01	94.74	95.38	95.98	68.73%
10 dB	94.66	94.77	95.44	94.72	94.90	88.73	90.45	89.53	90.44	89.79	90.30	87.30	88.80	91.63	74.16%
5 dB	90.60	84.49	90.67	87.84	88.40	77.40	77.42	76.86	77.82	77.38	78.39	71.95	75.17	81.34	68.99%
0 dB	76.05	54.23	70.03	73.25	68.39	47.34	48.40	48.67	47.21	47.91	49.34	45.56	47.45	56.01	46.93%
-5 dB	41.57	15.72	30.15	39.46	31.73	12.19	17.17	13.48	11.94	13.70	20.02	20.56	20.29	22.23	15.05%
Average	91.22	85.76	90.36	90.20	89.38	81.21	81.97	81.54	81.45	81.54	82.34	79.40	80.87	84.54	
	71.21%	71.59%	75.52%	71.68%	72.54%	60.36%	53.14%	60.52%	58.20%	58.29%	47.81%	39.20%	43.50%		61.30%