

音源復元のための映像中の微小振動方向の解析*

安見祐亮, 滝口哲也, 有木康雄 (神戸大)

1 はじめに

本稿では、音響信号の影響により発生した（音源から離れた）物体の微小振動をハイスピードカメラで捉えることにより、映像から音響信号を復元する手法を検討する．従来、遠距離から物体の振動を取るための手法においてレーザー光などが用いられていた．近年、文献 [1] では、映像中の目に見えない物体の微小振動をハイスピードカメラで捉える手法が提案されている．この手法では、物体を撮影するための光以外に必要なものが無いので、センサなどの追加部品を必要とせず、また、カメラの物体への向きに関する大きな制約もない．監視や安全保障などの分野にて、応用が期待されるほか、物体の振動に対する特性の計測にも利用が可能である．本稿では、抽出する振動方向による復元音の変化を調べる．

2 音の復元について

2.1 従来研究

画像の輝度プロファイルを画像中の座標 x によって $f(x)$ と表現する．このとき、画像をフーリエ級数分解によって複素正弦波の集合で表すと次のようになる [2]．

$$f(x) = \sum_{\omega=-\infty}^{\infty} A_{\omega} e^{i\omega x} \quad (1)$$

このとき、フレーム t における画像中の点の時間移動が置換関数 $\delta(t)$ で表されるとすると、画像中の点の移動は位相における変化にのみ表れ移動後の画像は次のように表される [2]．

$$f(x + \delta(t)) = \sum_{\omega=-\infty}^{\infty} A_{\omega} e^{i\omega(x + \delta(t))} \quad (2)$$

よって物体の振動を取り出す際には、参照フレームとフレームとの位相差を取り出せばよい．

文献 [1] においては、Complex Steerable Pyramid [3] を特徴量として用いることにより、局所的な位相差を取り出している．これにより、画像はスケール、オリエンテーションごとにウェーブレット分解される．この特徴量を用いて画像の移動量を求めることによりその原因となった音を復元している．また、この手法においてはノイズ処理として、パワースフィラによるナイキスト周波数の $1/20$ 以下の周波数の

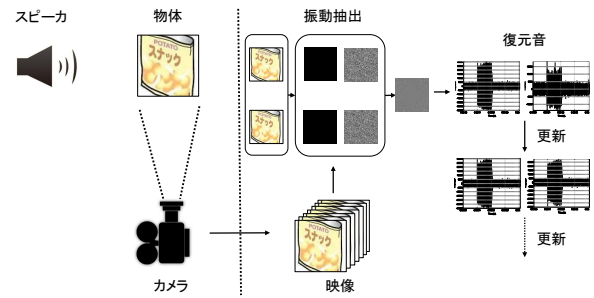


Fig. 1 Process of sound recovery[1]

カットと、スペクトルサブトラクション [4] や音声強調 [5] による処理が行われている．

2.2 提案手法

本稿では、物体の振動方向を求めるために画像の周波数領域の各周波数 (ω_x, ω_y) ごとの位相変化を音の復元に利用する手法を提案する．これは周波数領域の各周波数成分で振動を抽出することにより、それぞれが特定の方向の変化を取り出せるようにするためである．まず、映像の各フレームをフーリエ変換し周波数シフトを行ったものを $IM(\omega_x, \omega_y, t)$ とする． t はフレーム番号である．そして各周波数成分ごとにオイラーの公式から位相 $\phi(\omega_x, \omega_y, t)$ と振幅 $A(\omega_x, \omega_y, t)$ に分解する．

$$IM(\omega_x, \omega_y, t) = A(\omega_x, \omega_y, t) e^{i\phi(\omega_x, \omega_y, t)} \quad (3)$$

各信号 $s(\omega_x, \omega_y, t)$ は直前のフレームとの位相差 $\phi_v(\omega_x, \omega_y, t)$ と振幅の時間平均 $\bar{A}(\omega_x, \omega_y)$ の自乗との積から各フレーム間の移動量として求められる．それぞれ以下の式で表される．

$$\phi_v(\omega_x, \omega_y, t) = \phi(\omega_x, \omega_y, t) - \phi(\omega_x, \omega_y, t-1) \quad (4)$$

$$s(\omega_x, \omega_y, t) = \bar{A}(\omega_x, \omega_y)^2 \phi_v(\omega_x, \omega_y, t) \quad (5)$$

また、各信号はその時点までの移動量を足し合わせることで最初のフレームからの移動量に変換される． $s'(\omega_x, \omega_y, t)$ は変換後の各信号である．

$$s'(\omega_x, \omega_y, t) = \sum_t s(\omega_x, \omega_y, t) \quad (6)$$

*Analyzing a subtle vibration direction in a video for sound recovery. by Yūsuke Yasumi, Tetsuya Takiguchi, Yasuo Ariki (Kobe University)

また特定の方向の画像全体の移動量を求めるために、次のようなフィルタ $g(\omega_x, \omega_y)$ で各信号を重み付けして足し合わせる。 θ は π から $-\pi$ で振動の角度を表す。

$$g(\omega_x, \omega_y) = \cos(\theta + \arctan(\omega_y/\omega_x)) \quad (7)$$

こうして求められたものを復元音 $S(\theta, t)$ とする。また、振動方向を求めるために最初に復元音を $\theta = 0, -\pi/2$ の二つの方向について求める。次に得られた二つの復元音と各周波数成分における信号の相関を求め、最も相関の大きな角度の振動方向の復元音を求める。これを二つの角度の差が 0 か π になるまで繰り返すことにより、最も音を復元できる角度を求める。相関は復元音において最もエネルギーの大きな長さ l の区間 m での値を以下の式から求めた。

$$\sum_{t=t_{m_1}}^{t_{m_l}} S(\theta, t)s(\omega_x, \omega_y, t) \quad (8)$$

3 実験

3.1 実験条件

使用した映像は、菓子の包装を拡大した映像で、カメラに対し垂直方向からスピーカを用いて音を当てて撮影した。映像は 256×256 ピクセルで、フレームレートが 8000 [Hz] である。また、ノイズ除去についてはバタワースフィルタによる 55 [Hz] 以下の信号のカットオフとスペクトルサブトラクション [4] を行った。音には 2000 [Hz] の正弦波を用いた。比較のための従来手法には文献 [1] の手法を基に参照フレームを直前のフレームに固定し、最後に (6) 式と同様の更新により音を復元する方法を用いた。このとき特徴量のスケールが 4 、オリエンテーションが 2 である。提案手法においては、収束するまで更新するのではなく一度の更新でどれだけ変化するのかを求めた。

3.2 実験結果

従来手法、提案手法による復元音の Segmental SNR (SSNR)[6] を Table 1 に示す。実験結果より、更新によってノイズが軽減されているのがわかる。また、更新の前は二つの復元音は逆位相であったが更新後は同位相になった。このことから、抽出する振動方向によって復元音は同位相か逆位相のいずれかになると考えられる。

4 おわりに

今回は、映像の振動方向を求めその方向の振動を復元音として抽出する方法を提案し、抽出する振動方向による復元音の違いを確認する実験を行った。今後は、この手法を基に映像のみから音源定位が行えるのかについて考える。

Table 1 SSNR of recovered sound

	SSNR [dB]
Conventional method	5.4827
Proposed method ($\theta = 0$, before update)	9.0295
Proposed method ($\theta = -\pi/2$, before update)	2.3992
Proposed method ($\theta = 0$, after update)	9.0419
Proposed method ($\theta = -\pi/2$, after update)	7.3259

参考文献

- [1] Abe Davis *et al.*, “The Visual Microphone: Passive Recovery of sound from Video,” ACM Transactions on Graphics, 33(4), pp. 79:1-79:10, 2014.
- [2] Neal Wadhwa *et al.*, “Phase-Based Video Motion Processing,” ACM Transactions on Graphics, 32(4), 2013.
- [3] Javier Portilla *et al.*, “A Parametric Texture Model Based on Joint Statistics of Complex Wavelet Coefficients,” International Journal of Computer Vision 40(1), pp. 49-71, 2000.
- [4] Steven F. Boll, “Suppression of Acoustic Noise in Speech Using Spectral Subtraction,” IEEE Transactions on Acoustics, Speech and Signal Processing, vol. 27, no. 2, pp. 113-120, 1979.
- [5] Philipos C. Loizou, “Speech enhancement based on perceptually motivated bayesian estimators of the magnitude spectrum,” Speech and Audio Processing, IEEE Transactions on Speech and Audio Processing, vol. 13, no. 5, pp. 857-869, 2005.
- [6] John H. L. Hansen *et al.*, “An Effective Quality Evaluation Protocol for Speech Enhancement Algorithms,” In Proc. of the International Conf. on Speech and Language Processing, pp. 2819-2822, 1998.