氏　　名	LI, Feng			
学位の種類	博士(情報科学)			
学位記番号	博情第 422 号			
学位授与年月日	令和元年 9 月 24 日			
論文題目	Separation of Singing Voice from Music using Robust Principle Component Analysis and its Extensions			
論文審査委員	主査	赤木　正人	北陸先端科学技術大学院大学	教授
		党　　建武	同	教授
		鵜木　祐史	同	教授
		吉高　敦夫	同	准教授
		LI, Junfeng	中国科学院音響研究所	教授

論文の内容の要旨

The development in multimedia technologies has promoted dramatically the rapid growth of music data in recent years. There are various different applications for people's demands in music such as information retrieval, identification and handing. However, singing voice and background music are related to each other in the mixed music, the mutual interference has brought huge obstacles to music information processing. The problem of how to extract the audio information from music has become an important research topic. As the part of music information retrieval, the technologies of singing voice separation are facing unprecedented challenge.

The objective of this research is to deal with the problem of singing voice separation from monaural recordings. It is even more difficult than multichannel since the spatial information cannot be applied in the separation procedure. Singing voice separation is a technique for separating or extracting singing voice from a musical mixture, which has found many applications in the wide areas such as singer identification, singing evaluation and query by humming. This is a relatively easy separation task of the human auditory system, but it becomes more difficult when we attempt to simulate this problem in a computational method. To achieve the task of singing voice separation, this study mainly focuses on robust principal component analysis (RPCA) and its extensions.

RPCA has been recently proposed of popularization and effectiveness way of separation approach that separates singing voice and accompaniment from a mixture music. It decomposes a given amplitude spectrogram (matrix) of a mixture signal into the sum of a low-rank matrix (accompaniment) and a sparse matrix (singing voice). Since musical instruments reproduce nearly the same sounds every time, a given note is played in a given song, the magnitude spectrogram of these sounds can be considered as a low-rank structure. Singing voice, in contrast, varies significantly, but has a sparse distribution in the spectrogram domain to its harmonic structure.

Although RPCA is an effective approach to separate singing voice from the mixed audio signal, it fails when there are significant differences in dynamic range among the different background instruments. Some instruments, such as drums, correspond to singular values with tremendous dynamic range; because it uses nuclear norm to estimate the rank of the low-rank matrix, RPCA algorithm over-estimates the rank of a matrix that includes drum sounds. The accuracy of such separation results thus decreases, as drums may be placed in the sparse subspace instead of being low-rank. Thus, it motivates us to describe exactly the separated low-rank matrix.

To overcome the disadvantage of RPCA for singing voice separation, two extensions of RPCA algorithm are proposed in this dissertation. One is called weighted robust principal component analysis (WRPCA). It uses different weighted values to describe the low-rank matrix for singing voice separation. Additionally, incorporating the proposed WRPCA with gammatone auditory filterbank for singing voice separation. The significance of WRPCA can describe different low-rank matrix under the conditions of human's auditory perceptual properties. Because the cochleagram is derived from non-uniform time-frequency transform whereas time-frequency units in low-frequency regions have higher resolutions than in the high-frequency regions, which closely resembles the functions of the human ear. Therefore, it is promising to separate singing voice via sparse and low-rank decomposition on cochleagram instead of the spectrogram. %However, WRPCA suffers from high computational cost due to computing the singular value decomposition at each iteration. Hence, the running time of WRPCA is slower than RPCA.

Another extension of RPCA with rank-1 constraint called constraint RPCA (CRPCA). It utilizes the rank-1 constraint minimization of singular values in RPCA instead of minimizing the nuclear norm for separating singing voice from the mixture music. Thus, it not only provides a robust solution to large dynamic range differences among instruments but also reduces the computation complexity. Then, incorporating the proposed CRPCA with gammatone auditory filterbank on cochleagram for singing voice separation. In addition, constructing coalescent masking and vocal activity detection on CRPCA method to constrain the temporal segments that allowed to constrain singing voice from the mixed music datum. Finally, combining F0 and non-negative rank-1 constraint RPCA, which incorporates F0 and non-negative rank-1 constraint minimization of singular values in RPCA instead of minimizing the nuclear norm.

In conclusion, this dissertation proposes two extensions of the effective optimization algorithms concentrating on RPCA for singing voice separation. One is using different weighted value for describing the separated low-rank matrix. The other is exploring rank-1 constraint minimization of singular value in RPCA. In terms of source-to-artifact ratio, the previous is better than the later. However, CRPCA obtains better separation quality than WRPCA in singing voice separation. The outcomes of this research contribute to further improving the technologies related to music information retrieval. Additionally, the potential contribution of this research is to deal with the

problems of noise reduction and speech enhancement by using the separated low-rank and sparse model. Since the background noise is assumed as the part of low-rank component and the human speech is regarded as the part of sparse component.

Keywords: Singing voice separation, robust principal component analysis, weighted, rank-1 constraint, F0.

論文審査の結果の要旨

本論文は，モノラル録音された伴奏つきの歌唱音楽から歌声のみの分離を行うための手法に関する研究報告である。

マルチメディア技術の発展は，近年音楽データ処理の急速な成長を促進した。情報の検索，識別，配布など，音楽処理に対する人々の要求は膨らむ一方である。しかし一方で，過去にモノラル録音された音楽では歌声と伴奏とは互いに関連しており，これらの相互干渉が音楽情報処理に障害をもたらしている。歌声の分離は，歌唱音声の識別，歌唱の評価，ハミングによる検索など幅広い分野で多くの用途があることが知られている。このため，音楽から音声をどのように抽出するかという問題は，重要なトピックとなっている。マルチチャネル録音であれば，ICA などの方法も考えられるが，モノラル録音では空間情報を分離手順に適用することができないので，マルチチャネルよりもさらに困難な問題である。

本論文では，この問題を克服するために，ロバスト主成分分析（RPCA）とその拡張に焦点を当てている。RPCA は，歌声と背景楽器音がミックスされたオーディオ信号から歌声を分離するための効果的なアプローチであり，与えられた混合信号の振幅スペクトログラム（行列）を低ランク行列（背景楽器音）とスパース行列（歌声）の和に分解する。しかし，背景楽器音のダイナミックレンジに大きな違いがある場合（たとえばドラムなどの音が背景音にある場合）は，この音がスパース部分空間に配置される可能性があり，分離結果の精度を低下させる。

本論文では，このような問題点を持つ RPCA アルゴリズムに対して，二つの拡張が提案されている。一つ目の拡張は，分離された低ランク行列を記述するために異なる加重値を用いる方法（WRPCA）である。二つ目の拡張は，RPCA における特異値のランク 1 制約の最小化を適用する方法（CRPCA）である。評価実験の結果，WRPCA，CRPCA 双方とも，基準となる RPCA に比較して高い分離精度を示した。ソース対アーティファクト比の点では，WRPCA が CRPCA より優れている。一方，CRPCA は歌声分離において WRPCA よりも優れた分離品質が得られることも確認された。本研究の成果は，音楽情報検索に関する技術のさらなる向上に貢献できること，さらに雑音低減および音声強調の問題にも適用できることである。

以上のように，本論文は，従来知られていた RPCA の手法に対して有効な拡張法を提案

し，実験で有効性を実証したしたものであり、学術的に貢献するところが大きい。よって博士（情報科学）の学位論文として十分価値あるものと認めた。