

TRANG THÔNG TIN LUẬN ÁN

Tên đề tài luận án: Phân tích tín hiệu đa phân giải và áp dụng vào xử lý tiếng nói

Ngành: Khoa học máy tính

Mã số ngành: 9480101

Họ tên nghiên cứu sinh: Đỗ Đức Hào

Khóa đào tạo: 2019

Người hướng dẫn khoa học: TS. Trần Thái Sơn và TS. Châu Thành Đức

Cơ sở đào tạo: Trường Đại học Khoa học Tự nhiên, ĐHQG.HCM

1. TÓM TẮT NỘI DUNG LUẬN ÁN:

Luận án nghiên cứu về phương pháp phân tích tín hiệu đa phân giải, trong đó tập trung vào các phép biến đổi họ Chirplet và các ứng dụng vào xử lý tiếng nói.

Nội dung chính đầu tiên của luận án khai thác phép biến đổi Chirplet tuyến tính (Linear Chirplet Transform – LCT) và ứng dụng những phương pháp này vào một số các bài toán đơn giản trong lĩnh vực xử lý tiếng nói như nhận dạng giới tính, nhận dạng vùng miền.

Nội dung chính thứ hai tập trung vào nâng cao khả năng chống nhiễu cho đặc trưng thu được từ phép biến đổi Chirplet tuyến tính. Luận án đã tiến hành kết hợp phép biến đổi Chirplet tuyến tính với bộ tự mã hoá biến phân và bộ lọc Chebyshev để nâng cao chất lượng đặc trưng.

Nội dung chính thứ ba tập trung vào phân tích và làm rõ những giới hạn của LCT, qua đó mở rộng thành phép biến đổi Chirplet tuyến tính tổng quát (General Linear Chirplet Transform – GLCT). Sau đó, luận án kết hợp GLCT với phương pháp phân tích giá trị đơn rút gọn (Truncated Singular Value Decomposition - tSVD) để tạo thành phép biến đổi Chirplet tuyến tính tổng quát rút gọn (Compressed General Linear Chirplet Transform – CGLCT).

Nội dung chính thứ tư tập trung vào quá trình mở rộng từ phép biến đổi LCT thành phép biến đổi Chirplet đa thức (Polynomial Chirplet Transform – PCT) và phép biến đổi Chirplet nhiều hàm đa thức (Multi-Polynomial Chirplet Transform – MPCT).

2. NHỮNG KẾT QUẢ MỚI CỦA LUẬN ÁN:

Luận án đã đạt được một số kết quả mới, bao gồm:

Kết quả 1: Đề xuất thuật toán trích xuất đặc trưng tiếng nói dựa trên phép biến đổi LCT. Luận án đã sử dụng đặc trưng LCT với hệ số chirp âm vào các bài toán nhận dạng đơn giản như nhận dạng giới tính, nhận dạng vùng miền.

Kết quả 2: Nâng cao khả năng chống nhiễu cho phép biến đổi LCT. Cụ thể, luận án đã kết hợp phép biến đổi LCT với bộ tự mã hoá biến phân và bộ lọc Chebyshev để lọc nhiễu và tăng cường tiếng nói, qua đó nâng cao hiệu quả cho mô hình nhận dạng với dữ liệu đầu vào bị nhiễu.

Kết quả 3: Đề xuất hai thuật toán mở rộng khả năng của LCT bằng cách sử dụng nhiều hàm tuyến tính bao gồm GLCT và CGLCT. Thứ nhất, luận án kết hợp nhiều phép biến đổi LCT độc lập để tạo thành phép biến đổi GLCT. Sau đó, luận án đã đề xuất một phương pháp giảm chiều cho không gian đặc trưng thu được từ phép biến đổi GLCT bằng tSVD.

Kết quả 4: Đề xuất hai thuật toán nâng cao khả năng biểu diễn quy luật biến đổi của tín hiệu gồm PCT và MPCT. Từ đường tần số tức thời dạng tuyến tính, luận án đề xuất sử dụng đường tần số tức thời dạng đa thức để có thể biểu diễn được sự thay đổi phức tạp của tiếng nói.

3. CÁC ỨNG DỤNG/KHẢ NĂNG ỨNG DỤNG TRONG THỰC TIỄN HAY NHỮNG VẤN ĐỀ CÒN BỎ NGỜ CẦN TIẾP TỤC NGHIÊN CỨU

Mặc dù các đặc trưng đa phân giải họ Chirplet có chất lượng tốt, chúng vẫn gặp phải một số hạn chế cần được nghiên cứu và khắc phục như sau:

Đầu tiên, mặc dù giảm được độ phức tạp cho mô hình nhận dạng phía sau, quá trình xử lý của phép biến đổi Chirplet chưa tối ưu về tốc độ so với các phương pháp như ảnh phổ hay ảnh phổ mel. Do đó, hướng phát triển đầu tiên là thiết kế thuật toán tính toán nhanh, nhằm nâng cao tốc độ thực thi của phép biến đổi Chirplet, giúp ứng dụng trong thời gian thực và các hệ thống yêu cầu hiệu năng cao.

Thứ hai, với phép biến đổi CGLCT, dù hiệu quả tốt nhưng do đã qua phép biến đổi tSVD, thông tin không còn minh bạch như ban đầu, và khả năng giải thích được của đặc trưng suy giảm đáng kể. Vì vậy, hướng phát triển tiếp theo cho CGLCT là nâng cao khả năng giải thích của đặc trưng, có thể thông qua việc kết hợp với các kỹ thuật học máy giải thích được hoặc tối ưu hóa quá trình giảm chiều dữ liệu mà không làm mất mát thông tin quan trọng.

Cuối cùng, với phép biến đổi MPCT, thời gian thực thi và kích thước vector đặc trưng vẫn còn tương đối lớn. Do đó, hướng phát triển tiếp theo là giảm độ phức tạp cả không gian và thời gian cho quá trình xử lý. Điều này có thể đạt được bằng cách tối ưu hóa thuật toán, áp dụng các kỹ thuật giảm chiều hiệu quả hơn, hoặc sử dụng các phương pháp tính toán song song và tối ưu hóa phần cứng.

TẬP THỂ CÁN BỘ HƯỚNG DẪN

NGHIÊN CỨU SINH

TS. TRẦN THÁI SƠN

TS. CHÂU THÀNH ĐỨC

ĐỖ ĐỨC HÀO

**XÁC NHẬN CỦA CƠ SỞ ĐÀO TẠO
HIỆU TRƯỞNG**

THESIS INFORMATION

(1 – 1.5 A4 pages)

Thesis title: Multiresolution signal analysis and its applications in speech processing

Speciality: Computer science

Code: 9480101

Name of PhD Student: Hao Duc Do

Academic year: 2019

Supervisor: Dr. Son Thai Tran and Dr. Duc Thanh Chau

At: VNUHCM - University of Science

1. SUMMARY:

The dissertation investigates multi-resolution signal analysis methods, with a focus on the family of Chirplet transforms and their applications in speech processing.

The first main part of the dissertation explores the Linear Chirplet Transform (LCT) and applies this method to some basic tasks in speech processing, such as gender recognition and dialect recognition.

The second main part concentrates on enhancing the noise robustness of the features obtained from the Linear Chirplet Transform. The dissertation integrates the Linear Chirplet Transform with Variational AutoEncoders (VAE) and Chebyshev filters to improve feature quality.

The third main part focuses on analyzing and clarifying the limitations of the LCT, thereby extending it to the General Linear Chirplet Transform (GLCT). Subsequently, the dissertation combines the GLCT with Truncated Singular Value Decomposition (tSVD) to form the Compressed General Linear Chirplet Transform (CGLCT).

The fourth main part emphasizes extending the LCT to the Polynomial Chirplet Transform (PCT) and the Multi-Polynomial Chirplet Transform (MPCT).

2. NOVELTY OF THESIS:

The dissertation achieved four new results, including:

Result 1: A speech feature extraction algorithm based on the LCT transform was proposed. The dissertation utilized LCT-based features with negative chirp coefficients in simple recognition tasks such as gender and dialect recognition.

Result 2: Improved noise robustness of the LCT transform. Specifically, the dissertation integrated the LCT transform with a Variational AutoEncoder and a Chebyshev filter for noise suppression and speech enhancement, thereby improving the performance of recognition models under noisy input conditions.

Result 3: Proposed two algorithms to extend the capabilities of LCT through the use of multiple linear functions, resulting in the GLCT and CGLCT. First, multiple independent LCT transforms were combined to form the GLCT. Subsequently, the dissertation proposed a dimension reduction method for the GLCT-derived feature space using truncated Singular Value Decomposition (tSVD).

Result 4: Proposed two algorithms (PCT and MPCT) to enhance the representational power of signal variation patterns. Moving from a linear instantaneous frequency trajectory, the dissertation introduced polynomial instantaneous frequency trajectories to better capture the complex variations in speech.

3. APPLICATIONS/ APPLICABILITY/ PERSPECTIVE

Although the multi-resolution Chirplet family features exhibit high quality, they still face several limitations that warrant further investigation and improvement, as follows:

First, while these features reduce complexity for the subsequent recognition models, the Chirplet transform's processing speed is not yet optimized compared to methods like spectral images or mel-spectrograms. Thus, the first development direction involves designing fast computation algorithms to enhance the execution speed of the Chirplet transform, facilitating real-time applications and high-performance systems.

Second, with the CGLCT transform, although the effectiveness is good, the application of tSVD leads to reduced transparency of the original information and significantly diminishes the interpretability of the features. Therefore, the next development direction for CGLCT is to improve feature interpretability. This could be achieved by integrating explainable machine learning techniques or optimizing the dimensionality reduction process to avoid the loss of critical information.

Finally, for the MPCT transform, the execution time and feature vector size remain relatively large. Consequently, the next development direction is to reduce both spatial and temporal complexity in the processing. This can be achieved by optimizing the algorithm, applying more effective dimensionality reduction techniques, or employing parallel computing methods and hardware optimizations.

SUPERVISOR

PhD STUDENT

Dr. SON THAI TRAN

Dr. DUC THANH CHAU

HAO DUC DO

**CERTIFICATION
UNIVERSITY OF SCIENCE
PRESIDENT**