

氏 名 Le Dinh Duy

学位（専攻分野） 博士（情報学）

学位記番号 総研大甲第 1001 号

学位授与の日付 平成 18 年 9 月 29 日

学位授与の要件 複合科学研究科 情報学専攻
学位規則第 6 条第 1 項該当

学位論文題目 HUMAN FACE PROCESSING TECHNIQUES WITH
APPLICATION TO LARGE SCALE VIDEO INDEXING

論文審査委員 主 査 教授 佐藤 真一
教授 上野 晴樹
教授 杉本 晃宏
助教授 片山 紀生
助教授 北本 朝展
教授 井宮 淳（千葉大学）

論文内容の要旨

Human faces play an important role in efficiently indexing and accessing video contents, especially in large scale broadcasting news video databases. It is due to faces are associated to people who are related to key events and key activities happening from all over the world. There are many applications using face information as the key ingredient, for example, video mining, video indexing and retrieval, person identification and so on. However, face appearance in real environments exhibits many variations such as pose changes, facial expressions, aging, illumination changes, low resolution and occlusion, making it difficult for current state of the art face processing techniques to obtain reasonable retrieval results.

This thesis studies human face processing techniques whose target is to efficiently apply to a general framework for large scale video mining and indexing. In this framework, faces firstly are extracted, filtered and normalized from video sequences by using a fast and robust face detector. Next, similar faces are grouped into clusters. Then, these face clusters are labeled by the person names extracted from the video transcripts.

To extract faces from video, we propose a multi-stage approach that uses cascades of classifiers to yield a coarse-to-fine strategy to reduce significantly detection time while maintaining a high detection rate. This approach is distinguished from previous work by two features. First, we use a cascade of AdaBoost classifiers that is trained to be invariant to translation up to 25% of the original window size to detect quickly face candidate regions. Second, we use SVM classifiers which reuse the features selected by AdaBoost in the previous stage for robust classification and simple training. Reusing these features brings to two advantages: (i) These features do not need to be re-evaluated because they have already been evaluated. (ii) By using SVM classifiers with powerful generalization, using too many features in the cascade is avoided, with the important results of saving training time and avoiding over-fitting.

Furthermore, to help to reduce the training time, we propose two feature selection methods that quickly select a small and optimal subset of features by using mutual information and feature variance. In the feature selection method using mutual information, we propose using a more efficient discretization method that uses minimum description length principle (MDLP) to estimate probability densities of continuous random variables. This approach can be considered as a generalization of previous ones that mainly use a single threshold for discretization. In the other feature selection method, features are selected based on their distances to principle components computed by PCA (principle component analysis) from the data distribution. Using this approach, the final classifier is able to run faster than that using the traditional

PCA-based feature extraction method since it avoids computation cost of the subspace projection. These proposed feature selection methods are integrated seamlessly and efficiently into the multi-stage based framework for face detection described above.

The organization of the extracted faces is usually done automatically by using a clustering method. In many video indexing applications, k-means clustering is very common. However, it suffers from a number of serious drawbacks. For example, it can not be applied to general similarity measures; the number of clusters must be provided in advance; it generates many bad clusters when the input data is noisy; and it is not scalable to handle large datasets. Instead, we propose using the relevant set correlation (RSC) clustering model from which the GreedyRSC clustering heuristic derived. This clustering model can help to avoid all the problems of k-means clustering. Furthermore, it is very efficient in finding high quality clusters in such noisy datasets as face datasets extracted from video. These high quality clusters along with person names extracted from video transcripts are useful to identify important people appearing frequently in video databases that can be done by an association method based on the statistical machine translation.

The proposed techniques are integrated in developing a video indexing and retrieval system that can help users to access and navigate contents in news video databases easily and quickly. The system can show representative names and faces appearing in videos ranked by their occurrence frequency, and access to related news stories by using these faces or names. Furthermore, it can show possible associations between names and faces. Our approach is generic and has the potential to handle very large scale video datasets effectively and efficiently.

論文の審査結果の要旨

本博士論文は、画像・映像中の顔を対象とし、顔の検出・照合・同定、並びにこれらの技術を大規模映像アクセスに利用するための、映像マイニング・インデクシング技術について広範囲に検討し、まとめたものである。

顔情報は、映像検索などの映像インデクシングのために重要な情報源となりうる。特に大規模ニュース映像では、全世界で起こる重要なイベントや行動において、人物情報およびその顔情報はきわめて重要である。顔情報を利用した映像応用としては、映像マイニング、映像インデクシング、映像検索、人物同定などがあげられる。しかし、映像中に現れる顔は、姿勢変動、表情、照明条件、低解像度、隠れなどにより、さまざまな変動を受け、現状の顔画像処理技術の適用を困難にしている。本論文では、現状の顔画像処理にまつわるさまざまな技術を網羅的に調査すると同時に、映像中の顔画像処理により適した技術について検討し、有効な映像応用システムの実現性について探求した。

本論文における主な研究成果は以下の通りである。

(1) 映像中の顔画像処理に適した特徴抽出並びに特徴選択

顔検出や顔照合に適した画像特徴量を探求するために、Wavelet、Local Binary Patterns(LBP)、エッジ方向ヒストグラムなど、種々の方法について網羅的に調査し、適用する画像の品質、検出や照合などの利用局面に応じた利害得失などについて明らかにした。また、顔画像処理に適した、条件付相互情報量に基づく特徴選択手法、並びに主成分分析に基づく特徴選択手法を提案した。

(2) 多段構成による高速顔検出

機械学習アルゴリズムの AdaBoost ならびに SVM を構成要素とし、多段構成の構造を持つ高速顔検出手法を提案した。提案手法は、現在知られているもっとも高速で検出精度も高い顔検出手法と比べて、遜色ない、あるいは勝っている動作速度並びに検出精度を実現すると同時に、学習時間を大幅に削減することに成功した。

(3) 顔画像による大規模映像インデクシング・映像検索

顔検出、顔照合のための特徴量、並びにクラスタリング技術を組み合わせ、数百時間規模の米国並びに日本のニュース映像から抽出した顔情報を対象とした映像インデクシング・映像検索システムを実現した。指定した期間で話題となった人物の顔画像を一覧する機能、ある人物がどのようなニュースで扱われているかを調査する機能、顔と名前の対応付けを行う機能などにより、効率よく大規模ニュース映像を探索できることを示した。

平成18年7月26日に、博士論文公聴会にて出願者による博士論文の内容の発表、並びに審査委員および聴講者による質疑応答が行われた。その結果を受けて、非公開にて、審査委員による審査が行われた。論文内容の発表では、研究の背景と狙い、本論文が提案する映像インデクシングのための顔画像処理技術の特徴並びに関連技術との比較などが報告された。質疑応答に対し、出願者は的確かつ明快に回答した。その結果、出願者は情報学並びに関連する分野に関し、学位を授与するに十分な学識を有するものと判断し、合格とした。