様　式　Ｃ－１９、Ｆ－１９、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成 ２７ 年　 ６ 月 １９ 日現在

機関番号：６２６１５
研究種目：基盤研究(C)
研究期間：2012～2014
課題番号：２４５００１３６

研究課題名（和文）Ｆａｃｅ　Ｉｎｄｅｘｉｎｇ　ｆｏｒ　Ｓｏｃｉａｌ　Ａｎａｌｙｓｉｓ

研究課題名（英文）Face Indexing for Social Analysis

研究代表者
ＬＥ　ＤＩＮＨ　ＤＵＹ（LE, DINH DUY）

国立情報学研究所・コンテンツ科学研究系・准教授

研究者番号：８０４５０１７０

交付決定額（研究期間全体）：（直接経費）　　4,100,000円

研究成果の概要（和文）：社会分析アプリケーションにおける大規模な映像アーカイブのための顔インデキシングの高度な技術を実現した。 顔検出、マッチング、および識別の性能は、大規模なマルチチャンネルビデオアーカイブの長期観察の利点を得ることによって改善され、その結果、数百万の大規模な顔のデータセットから顔検索および識別のためのスケーラブルな技術を開発した。

東京エリアの7チャンネルで放送される映像番組の10万時間以上のビデオアーカイブにこれらの技術が適用され、その後、政治家や有名人の頻繁な表示に基づいて、社会的な分析のためのアプリケーションが実証された。

研究成果の概要（英文）：We have developed advanced techniques for face indexing in large video archives with the target application being social analysis. As a result, we have improved performance of face detection, matching, and identification by taking the benefits from a long-term observation of large multi-channel video archives. We have developed a scalable technique for face retrieval and identification from large datasets of hundreds of millions of faces.
We have applied these techniques to the video archive of more than 100,000 hours of video programs broadcast on 7 channels in the Tokyo area. Then we have demonstrated an application for social analysis based on frequent appearance of politicians and celebrities.

研究分野：マルチメディア分析

キーワード：顔照合　顔インデキシング　映像マイニング　映像インデキシング　社会分析　クラスタリング

１．研究開始当初の背景

With the recent advances made in pattern recognition research, the human face is widely used in various applications since it provides a rich level of information for spotting the appearance of certain people of interest, such as government leaders in news video, or a hero in a movie, and is the basis for interpreting facts and events.

Leveraging NII's video archive of more than 100,000 hours of video programs broadcast on 7 channels in the Tokyo area (http://www.vpl.nii.ac.jp/tv-recs/index-ja.html), this research aims to develop advanced techniques for face indexing with the target application being social analysis.

Once such techniques applied to large-scale dataset such as the NII broadcast video archives, results can be useful for multiple purposes, such as face-name association, face retrieval, face-related survey, and social analysis, which are normally inaccessible due to the prohibitive costs of a manual analysis. One typical application is that we can evaluate correlation of a person with program types (e.g. politics, comedies, and dramas) and time period (e.g before and after an election), popularity of persons (recently popular comedians or actors/actresses) through total their appearance time.

２．研究の目的

Face indexing involves detecting, tracking, matching and naming faces appearing in multimedia databases. Most existing applications (e.g. the Google Image Search Engine) only involve in frontal face detection and face-name association is done by meta-tags such as the filename, text around images rather than visual content. Furthermore, most of these techniques are studied in controlled environments with static images that are not applicable to practical multimedia databases such as broadcast videos. Therefore, the challenges that will be attacked in this research are: *(i) how to take the benefits from a long-term observation of a large multi-channel video archive in order to significantly improve the performance of face detection and matching? (ii) how to effectively automate key labor-intensive tasks, such as making annotations for training a large number of person recognizers? and (iii) how to realize an actual working system for the purpose of social analysis that is scalable to huge data amount, i.e. hundreds of thousands of hours of video?*

３．研究の方法

(1) To improve the existing face detection techniques used for static images (e.g. Viola and Jones-like face detector[f]), the key idea is to make use of the rich amount of contextual information in video data, such as the textual, visual, and temporal information. In particular, the faces of each person in video sequences will simultaneously be detected, tracked, and grouped into face tracks. This method is able to not only handle frontal faces but also to handle profile faces and occluded faces.

(2) To improve the performance of face matching, the face tracks extracted by the above method are used instead of single faces. The advantage of the proposed approach is to take the benefits of a long-term observation of the co-occurrence between faces and names in large multi-channel video archives to improve the performance of name and face association. The key idea is we can obtain a cluster which is expected to correspond to a set of faces of one individual, but due to imperfect technologies, it may contain noises. However, if it has sufficiently many faces, and it's statistically sufficiently " correct" , we can approximately correctly observe correlations of a set of faces (corresponding to an individual) with metadata such as names in closed caption, program types, etc. In large video databases, these conditions are usually sastisfied.

(3) To minimize intensive labor tasks in making annotations for training a large number of person recognizers, a semi-supervised framework is proposed. The basic idea is to make use of the context and social information available on the Internet to obtain annotations that is then used to train person recognizers. Initially, these recognizers might not have high prediction accuracy due to noisy annotations, but by combining relevence feedback and result of name-face association in a boosting framework, the performance is significantly improved.

4．研究成果

(1) We have developed a face retrieval system using novel techniques for face track extraction and matching as shown in Figure 1.

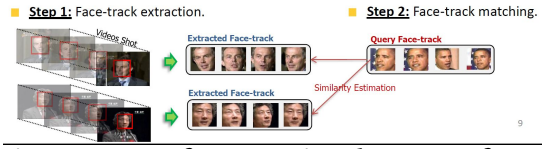


Figure 1 - A face retrieval system from videos using face tracks.

A simple yet efficient face track matching has been proposed as shown in Figure 2 and applied to the NII video archive.

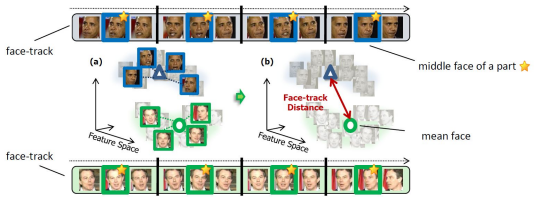
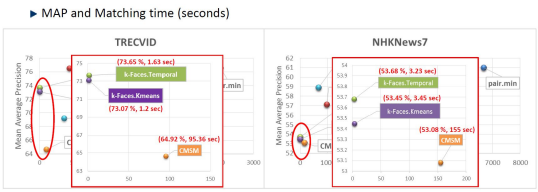


Figure 2 - Face track matching method.

In comparison with other state of the art methods, our method achives better performance as shown in Figure 3.

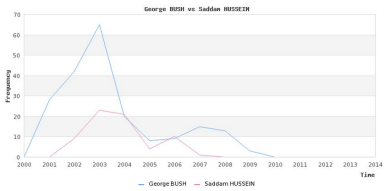


Approaches	TRECVID			NHKNews7		
	MAP(%)	Preprocessing Time	Matching Time	MAP(%)	Preprocessing Time	Matching Time
pair.min + *L*1	76.54	0.00	2544.73	60.99	0.00	6678.00
k – *pair.min* + *L*1 (*k*=20)	76.50	0.00	418.24	57.11	0.00	972.36
k-Faces.*Temporal* + *L*1 (*k*=20)	73.65	26.54	1.63	53.68	41.4	3.23
k-Faces.*KMeans* + *L*1 (*k*=20)	73.07	6380.50	1.20	53.45	14949.12	3.45
MSM	69.20	4454.10	347.39	58.92	4896.72	667.15
CMSM	64.62	4991.02	95.36	53.08	5841.80	155.40

Figure 3 - Comparison with other methods.

(2) We have developed an annotation framework and annotated face tracks extracted from video archives. The annotation data is made public available for research purpose (http://satoh-lab.ex.nii.ac.jp/users/ndthanh/NIIFacetrackDatasets/).

(3) We have developed a demo system as shown in Figure 4 for social scientists to analyze correlations between two persons based on their appearance on broadcast video programs of NHK News 7.Our system can search persons by using names such as "Naoto Kan" or search by images containing a face of the target person.



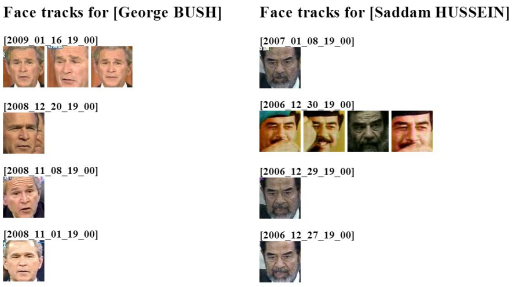


Figure 4 - A result of our face retrieval system.

5．主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 3 件）

(1) Bor-Chun Chen, Yan-Ying Chen, Yin-Hsi Kuo, Thanh Duc Ngo, Duy-Dinh Le, Shin'ichi Satoh, Winston H. Hsu, "Scalable Face Track Retrieval in Video Archives using Bag-of-Faces Sparse Representation", IEEE Trans. on Circuits and Systems for Video Technology, (2014) (Accepted) (査読有り).

(2) Klaus Schoeffmann, David Ahlström, Werner Bailer, Claudiu Cobârzan, Frank Hopfgartner, Kevin McGuinness, Cathal Gurrin, Christian Frisson, Duy-Dinh Le, Manfred Del Fabro, Hongliang Bai, "The Video Browser Showdown: A Live Evaluation of Interactive Video Search Tools", International Journal of Multimedia Information Retrieval, Vol.3, No.2, pp.113-123, (2014) (査読有り).

(3) Thanh Duc Ngo, Hung Thanh Vu, Duy-Dinh Le, Shin'ichi Satoh, "Face Retrieval in Large-Scale News Video Datasets", IEICE Transactions on Information and Systems, Vol.E96-D, No.8, pp. 1811-1825, (2013) (査読有り).

〔学会発表〕（計 7 件）

(1) Ngoc-Bao Nguyen, Vu-Hoang Nguyen, Thanh Duc Ngo, Duy-Dinh Le, Duc Anh Duong, "AttRel: An Approach To Person Re-Identification By Exploiting Attribute Relationships", International Conference on Multimedia Modeling (MMM), Jan. 5-7, 2015, Sydney, Australia.

(2) Thanh Duc Ngo, Vinh-Tiep Nguyen, Vu Hoang Nguyen, Duy-Dinh Le, Duc Anh Duong, Shin'ichi Satoh, “NII-UIT Browser: A multimodal Video Search System”, International Conference on Multimedia Modeling (MMM), Jan. 5-7, 2015, Sydney, Australia.
(3) Thanh Duc Ngo, Vu Hoang Nguyen, Vu Lam, Sang Phan, Duy-Dinh Le, Duc Anh Duong, Shin'ichi Satoh, “NII-UIT: A Tool for Known Item Search by Sequential Pattern Filtering”, International Conference on MultiMedia Modeling (MMM), Jan. 8-10, 2014, Dublin, Ireland.
(4) Vu-Hoang Nguyen, Thanh Duc Ngo, Kien Nguyen, Khang M. T. T Nguyen, Duy-Dinh Le, Duc Anh Duong, “Re-Ranking For Person Re-Identification”, International Conference of Soft Computing and Pattern Recognition (SoCPaR), Dec. 15-18, 2013, Hanoi, Vietnam.
(5) Vu Hoang Nguyen, Kien Nguyen, Duy-Dinh Le, Duc Anh Duong, Shin'ichi Satoh, “Person Re-Identification using Deformable Part Models”, International Conference on Neural Information Processing (ICONIP), Nov. 3-7, 2013, Daegu, Korea.
(6) Chi Nhan Duong, Pham Dinh Thang Cap, Thanh Duc Ngo, Duy-Dinh Le, Hoai Bac Le, Duc Anh Duong, Shin'ichi Satoh, “Robust Eye Localization in Video By Combining Eye Detector and Eye Tracker”, International Conference on Pattern Recognition (ICPR), Nov. 11-15, 2012. Tsukuba International Congress Center, Tsukuba, Ibaraki, Japan.
(7) Duy-Dinh Le, Shin'ichi Satoh, “Auto Face Re-Ranking By Mining the Web and Video Archives”, IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 16-21, 2012, Providence, Rhode Island, USA.

〔その他〕
ホームページ等
http://www.satoh-lab.nii.ac.jp/~ledduy/

6．研究組織

(1)研究代表者
ＬＥ　ＤＩＮＨ　ＤＵＹ（LE, DINH DUY）
国立情報学研究所・コンテンツ科学研究系
・准教授

研究者番号：80450170

(2)研究分担者
なし

(3)連携研究者
なし