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 １６ 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：２４５００１３５

研究課題名（和文）ラベル伝播による画像データセットにおける顔への自動ラベル付け手法

研究課題名（英文）Unsupervised Techniques for the Propagation of Label Information in Image Datasets

研究代表者
Ｍｉｃｈａｅｌ　Ｅ．Ｈｏｕｌｅ（HOULE, Michael E.）

国立情報学研究所・大学共同利用機関等の部局等・客員教授

研究者番号：９０３９９２７０

交付決定額（研究期間全体）：（直接経費）　　3,900,000円

研究成果の概要（和文）：顔画像の自動認識という研究テーマは、多くの努力にもかかわらず、非常に困難である。クエリとして顔画像を処理する時に、現在の一番効果的な顔認識アルゴリズムは、固定数の得られた顔画像の間でクエリと似ている画像を同定すると言う手法である。一般的な画像データベースには、検索された顔と似ているインスタンスが複数含まれる事が多い。このプロジェクトの目標は、類似画像の集合の中でアノテーションを共有することによって、タグ情報を持つデータベース画像の数を増やすことである。

研究成果の概要（英文）：Despite much effort by researchers, the problem of automatic recognition of face images remains very difficult. When processing face image queries, the most effective of the current face recognition algorithms determine as the query result a set of images that most closely resemble several of a fixed set of labeled examples. Typically, image databases contain multiple images that may be similar to a given query. The goal of this project is to increase the number of database images with label information, by sharing face image annotations with unlabeled images that are similar to them.

研究分野：データマイニング

キーワード：ラベル伝播　分類アルゴリズム　類似検索　画像データ

１．研究開始当初の背景

Computer users around the world share images and videos on a daily basis, of such events as graduations, weddings, birthdays, and travel. City centers, sensitive sites, business and residential buildings are being monitored daily through cameras for security reasons. In news, multimedia data represent the main information channel. In all these cases, the search by users for particular images is mainly done manually. The general problem with images and videos is that their digital representations do not convey any meaningful information about their content. An interpretation needs to be added to the raw data either manually or automatically. Manual annotation of images and video is tedious, and automatic inference of the semantics of images and videos is not an easy task.

In several major applications of images and videos, the semantics involve the recognition of human faces. Face annotation can be cast as a standard classification problem: Given a set of face images labeled with the person's identity (the gallery set) and an unlabeled set of face images from the same group of people (the probe set), we seek to identify each person in the probe images. So far this problem has been tackled from the pattern classification and machine learning point of view, where it is known as *face recognition*. Classifiers are known to work better when trained with large training sets. Unfortunately, large training sets are generally very expensive to compute, due to the high degree of human intervention required.

２．研究の目的

Consider a large set of images of N individuals (e.g., from family photo albums or news images databases), from which M frontal faces are detected. After manual tagging of a small number of faces (with the names of the individual), the tag information is to be propagated to untagged faces. In order to decide which tag to apply to an individual face, candidates must be identified and ranked. Each detected face is associated with a face descriptor (high dimensional vector), based on which pair-wise face similarity scores can be computed using state-of-the-art face matching techniques. However, current state of the art methods for face matching are not well-suited for ranking purposes – the scores are generally reliable only for relatively close matches, and are less discriminatory for partial matches (such as when the face is partially obscured, or viewed in a different pose or from a different angle).

The annotation of image objects, such as the faces of individuals or famous landmarks, can be cast as a standard classification problem: Given a set of image objects with descriptive labels attached (the gallery set), and an unlabeled set of similar objects (the probe set), we seek to identify as many objects as possible from the probe set. The aim of this project is to reduce the high cost of human intervention by relying on the redundancy of faces in the probe set to propagate the identities provided by the gallery set.

３．研究の方法

Consider a large set of images (e.g., from family photo albums or news images databases), from which M objects of some targeted type are detected. Examples of such objects include images of faces of individuals, and images of household objects, both taken from a variety of angles and lighting conditions. After manual tagging of a small number of objects (with names or other object attributes), the tag information is to be propagated to untagged objects. In order to decide which tag to apply to an individual object, candidates must be identified and ranked. Each detected object is associated with a descriptor (high dimensional vector), based on which pair-wise similarity scores can be computed using state-of-the-art matching techniques. However, current state of the art methods for image matching are not well-suited for ranking purposes – the scores are generally reliable only for relatively close matches, and are less discriminatory for partial matches (such as when the object is partially obscured, or viewed in a different pose or from a different angle, or under different lighting conditions) .

Recently, researchers had proposed the use of matching graphs as a means for expressing relationships among images. We planned to adopt this approach for our tag propagation problem. In our model, each object is represented by a vertex of a graph. If the similarity between a tagged

object $O_t$ and an untagged object $O_u$ is greater than a threshold $T_{tu}$, an edge is drawn from $O_t$ to $O_u$. If the similarity between two untagged objects $O_u$ and $O'_u$ is greater than a threshold $T_{uu}$, edges are drawn from both $O_t$ to $O'_t$ and $O'_t$ to $O_t$. Finally, self-edges are added to all tagged faces. The result is a directed graph within which influences can be propagated in a manner similar to that of the PageRank web page ranking model. Unlike PageRank in which the score of a node is averaged by the number of its forward links and propagated to the linked node, in our model, neighbors of an untagged object $O_u$ have equal rights to vote and their scores should be aggregated and propagated differently. The voting and propagation mechanisms have a profound influence on performance, and thus were expected to require extensive discussion and collaboration among the project members for their design. Our main objective was to develop different integration models and compare them experimentally on a number of large, real-world datasets.

Our original plan for the later years of the project was to extend our investigation so as to take visual feature information into account, using the bag-of-visual-words model. In a manner similar to before, we would build a graph with two distinguished edge sets – one based on face similarity and the other on visual word similarities. The main research question becomes how to remodel voting and propagation strategies to take into account two forms of similarity simultaneously. There are very many possible candidate models to consider, and extensive experimentation will be required.

４．研究成果

(1) In the first year of the project, we proposed an influence propagation strategy, *SW-KProp*, that requires no human intervention beyond the initial labeling of a subset of the images. *SW-KProp* distributes semantic information within a similarity graph defined on all images in the database: each image iteratively transmits its current label information to its neighbors, and then readjusts its own label according to the combined influences of its neighbors. *SW-KProp* influence propagation can be efficiently performed by means of matrix computations, provided that pairwise similarities of images are available. We also propose a variant of *SW-KProp* which enhances the quality of the similarity graph by selecting a reduced feature set for each prelabeled image and rebuilding its neighborhood. The performances of the *SW-KProp* method and its variant were evaluated against several competing methods on classification tasks for three image datasets: a handwritten digit dataset, a face dataset and a web image dataset. For the digit images, *SW-KProp* and its variant performed consistently better than the other methods tested. For the face and web images, *SW-KProp* outperformed its competitors for the case when the number of prelabeled images was relatively small. The performance was seen to improve significantly when the feature selection strategy was applied. This work was published in the top-ranked international journal ACM TOMCCAP (now known as ACM TOMM). A preliminary version was presented at the top-ranked international conference ACM Multimedia in November 2011, after the initial application for this grant, but before the formal start in April 2012).

After the first year of the project, the outcomes obtained to that point (on similarity graph construction with local feature selection) led us to believe that the performance of our label propagation method depends crucially on the quality of the similarity graph used. The methods developed for label propagation turned out to be generally applicable, but with widely-varying performance across data sets and data types. For the latter half of the project, we therefore turned our attention to the problem of how to enhance the similarity graph construction so as to achieve better quality results for the most difficult classes of objects (such as images).

(2) In data mining applications such as subspace clustering or feature selection, including the feature selection methods used in our most successful label propagation strategy from the first year of the project, changes to the underlying feature set can require the computationally-expensive reconstruction of search indices to support fundamental data mining tasks. These methods require the efficient on-demand computation of similarity values using subsets of the original feature set. For such situations, multi-step search approaches have been

proposed that can accommodate changes in the underlying similarity measure without the need to rebuild the index. In a paper published at the premier IEEE ICDM international conference in data mining, we presented a heuristic multi-step search algorithm that utilizes a measure of intrinsic dimension, the generalized expansion dimension (GED), as the basis of its search termination condition. Compared to the current state-of-the-art method, our experimental results showed that our heuristic approach is able to obtain significant improvements in both the number of candidates and the running time, while losing very little in the accuracy of the query results.

This solution was extended to the problem of Flexible Aggregate k-Nearest Neighbor (FANN) search, in which queries are specified by multiple objects rather than just a single object. This capability would allow similarity graphs to be constructed in which the influences of multiple labeled instances could be combined. Two papers arising from this work on FANN are currently in submission, one to the international journal IEEE TKDE (as a requested "minor revision") and the other to the SISAP 2015 international conference. Within this year, we plan to evaluate the performance of FANN in the construction of similarity graphs for label propagation.

Another approach to more flexible construction of similarity graphs is to support efficient subspace query processing, in which only some of the features are allowed to contribute to the similarity score (with these features being determined by the user when specifying the query). Although this query variant had been considered by other researchers over the past decade, previous solutions have been impractically expensive. Our solution, based on lower-bounding distances and multi-step search, outperformed competing methods by a very wide margin. This work was presented at the SISAP 2014 international conference, where it received the Best Paper award. A more complete version of the paper has been submitted to an internal journal.

(3) Another line of research was to allow dynamic modification of the similarity graph in order to enhance the quality of image labeling produced by the propagation process. For each instance of a tagged image, the relevance of each of its features is assessed according to the discriminability of the similarity measure over all other tagged images, with respect to the single feature. Then, the tagged image is associated with a reduced set of its own most discriminative features. These features are then used to customize the similarity values between tagged and untagged faces in the graph construction.

In an international journal paper (IJMIR) and a top-ranked international conference paper (ICM), we proposed the *NNF-Descent* heuristic for the efficient construction of similarity graphs based on nearest-neighbor and feature descent, in which selective sparsification of feature vectors is interleaved with neighborhood refinement operations in an effort to improve the semantic quality of the result. A variant of the Laplacian Score is proposed for the identification of noisy features local to individual images, whose values are then set to 0 (the global mean value after standardization). We showed through extensive experiments on several datasets that *NNF-Descent* is able to increase the proportion of semantically-related images over unrelated images within the neighbor sets, and that the proposed method generalizes well for other types of data which are represented by high-dimensional feature vectors. The similarity graphs produced, when evaluated on the label propagation tasks, reduced the overall classification error by a huge margin (typically, by a proportion of 30%).

(4) Although these outcomes were extremely effective, we have found that the best practical results have been obtained by combining strategies (feature reduction, voting strategies, etc.). During the term of the project, some of the later ideas (FANN, subspace similarity search) have not been able to be tested in the label propagation task. We plan to continue this very fruitful direction of research in 2015.

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　2 件）

1. M. E. Houle, X. Ma, V. Oria and J. Sun, "Improving the quality of

K-NN graphs for image databases through vector sparsification: application to image databases", International Journal on Multimedia Information Retrieval, refereed, 3(4), September 2014, pp. 259-274, DOI 10.1007/s13735-014-0067-7 .

2. M. E. Houle, V. Oria, S. Satoh and J. Sun, "Annotation propagation in image databases using similarity graphs", ACM Transactions on Multimedia Computing, Communications and Applications (ACM TOMCCAP), refereed, 10(1), Article 7, December 2013, 21 pages, DOI 10.1145/2487736 .

〔学会発表〕（計　3 件）

1. M. E. Houle, X. Ma, V. Oria and J. Sun, "Efficient algorithms for similarity search in axis-aligned subspaces", Proc. 7th International Conference on Similarity Search and Applications (SISAP), refereed, Springer LNCS 8821, 29-31 October 2014, Los Cabos (Mexico), 12 pages, DOI 10.1007/978-3-319-11988-5 .
2. M. E. Houle, X. Ma, V. Oria and J. Sun, "Improving the quality of K-NN graphs for image databases through vector sparsification", Proc. 5th ACM International Conference on Multimedia Retrieval (ICMR), refereed, 1-4 April 2014, Glasgow (UK), pages 89-96, DOI 10.1145/2578726.2578730 .
3. M. E. Houle, X. Ma, M. Nett and V. Oria, "Dimensional testing for multi-step similarity search", Proc. 12th IEEE International Conference on Data Mining (ICDM), refereed, 10-13 Dec. 2012, Brussels (Belgium), pp. 299-308, DOI 10.1109/ICDM.2012.91 .

〔図書〕（計　0 件）

〔産業財産権〕
○出願状況（計　0 件）

○取得状況（計　0 件）

６．研究組織
(1)研究代表者
フール　マイケル　E.（HOULE, Michael E.）
国立情報学研究所・大学共同利用機関等の部局等・客員教授
研究者番号：90399270

(2)研究分担者
佐藤　真一（SATOH, Shin'ichi）
国立情報学研究所・コンテンツ科学研究系・教授
研究者番号：90249938