Dynamic Reconstruct for Network Photograph Exploration

T.RAJESH[#1], A.RAVI[#2]
Asst. Professor in MCA[#1], Asst. Professor in IT[#2]
Malineni Lakshmaiah Engineering College
S.Konda, Prakasam Dist., A.P., India.

ABSTRACT—Photograph sear reconstruct methods usually fail to capture the user's intention when the query termism ambiguous. Therefore, reconstruct with user interactions, or active reconstruct, is highly demanded to effect very improve the search performance. The essential problem in active reconstruct is how to target the user's intention. To complete this goal, this paper presents a structural information based sample selection strategy to reduce the user's labeling efforts. Furthermore, to localize the user's intention in the visual feature e space, a novel local-global discriminative dimension reduction algorithmic proposed. In this algorithm, a sub manifold is leer need by transferring the local geometry and the discriminate vet information from the labeled photographs to the whole (global) photograph database. Experiments on both synthetic datasets and a real Network photograph sear chdatasetde menstruate he effectiveness of the proposed active reconstruct scheme, including both the structural information based active sample selection strategy and the local-global discriminative dimension reduction algorithm.

Index Terms—Active reconstructs, local-global discriminative (LGD) dimension reduction, structural information (SInfo) based active sample selection, network photograph sear reconstruct.

I. INTRODUCTION

CURRENTLY, most of the popular commercial Network photograph exploration engines, e.g., Microsoft's Live Photograph Exploration and Google Photograph Exploration, are built for "query by keywords"scenario. That is, a user provides keyword, e.g., "panda", then the exploration engine returns corresponding photographs by processing the associated xtual information, e.g., file name, surrounding text, URL, etc. Although text-based exploration techniques have shown their effectiveness in the document exploration, they are problematic when applied to the photograph exploration. There are two main problems. One is the mismatching between photographs and their associated textual information, resulting into irrelevant photographs appearing in the exploration results. For example, an photograph which is irrelevant to "panda" will be mistaken as a relevant photograph if there is a word "panda" existing in its surrounding text. The other problem is that the textual information is insufficient to represent the semantic content of the photographs. The same query words may refer to photographs that are semantically different, e.g., we cannot differentiate an animal panda photograph from an photograph for a person whose names Panda, just with the text word "panda".

Because the textual information is insufficient for semantic photograph retrieval, a natural recourse is the visual information. Recently a dozen of photograph/video rerunning methods [6], have been proposed to exploit the usage of the visual information for refining the text-based exploration result. Most of these reconstruct methods utilize the visual information in an unsupervised and passive manner. The only exception is the Intent Exploration [6], which reorders the text-based exploration result by using query by example (QBE), with the query photograph specified by the user from the initial text-based exploration result.

Although Intent Exploration[6] can be deemed as a simplified version of dynamic reconstruct, i.e., the user's intention is defined by only one query photograph, it cannot work well when the user's intention is too complex to be represented bygone photograph. As shown in Fig. 3, the query relevant photographs for "Animal" vary largely both in visual appearance and features, thus we cannot represent "Animal" only with one photograph. Instead, our proposed active reconstruct method can learn the user's intention more extensively and completely.

A. Dynamic User's Labeling Information Collection

To collect the labeling information from user sufficiently, a ne w structural information (SInfo) based strategy is proposed to dynamically select the most informative query photographs.

B. Visual Characteristic Localization

To localize the visual characteristics of the user's intention, we proposea novel local-global discriminative (LGD) dimension reduction algorithm. Basically, we assume that the quarreled vent photographs, which represent the user's intention, are lying one low-dimensional submanifoldof the original ambient (visual feature) space.

II. A CTIVE RECONSTRUCTFOR NETWORKPHOTOGRAPHEXPLORATION

Fig. 1 shows the proposed general framework for dynamic reconstruct in Network photograph exploration. Take the query term "panda" as an example. When "panda" is submitted to the Network photograph exploration engine, an initial text-based exploration result is returned to the user , as shown in Fig. 1(a) (only the top nine photographs are given for illustration). This result is unsatisfactory because both person and animal photographs are retrieved as top results. This is caused by the ambiguity of the query term. Without the user interactions, it is impossible to eliminate this ambiguity. In particular , which kind of photographs, animal panda or person whose name is Panda, are user's intention? Therefore, traditional reconstruct methods, which improve the initial exploration results by only utilizing the visual property of photographs, cannot achieve good performances.

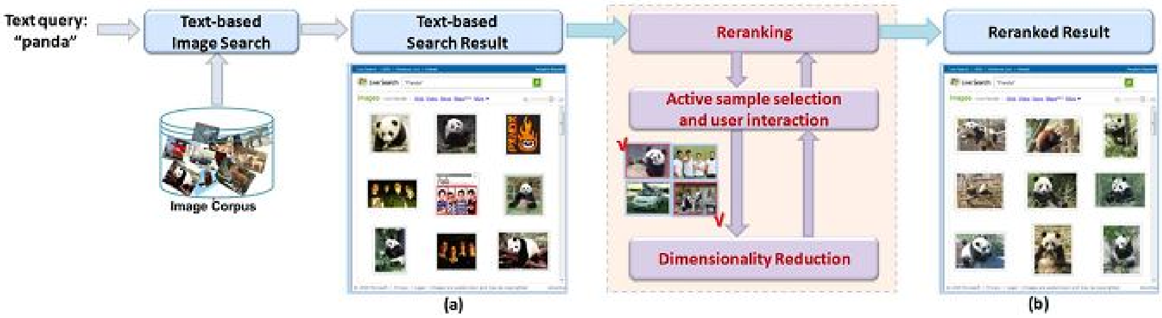


Fig. 1. Framework for active reranking illustrated with the query "panda". When the query is submitted, the text-based image search engine returns a coarse result (a). Then the active reranking process is adopted to obtain a more satisfactory result (b), by learning the user's intention.

In summary , there are two key steps in learning the user's intention, i.e., the dynamic sample selection strategy and the dimension reduction algorithm. This paper implements these two steps via new S Info sample selection strategy and a novel LGD dimensionreduction algorithm, as will be discussed in Sections III and IV, respectively.

III. SI NFODYNAMIC SAMPLE SELECTION

An S Info dynamic sample selection strategy is presented to learn the user's intention efficiently which selects photographs by considering not only the ambiguity but also the representativeness in the whole photograph database. Ambiguity and representativeness are two important aspects in dynamic sample selection. Labeling a sample which is more ambiguous will bring more information. On the other side, the information provided by individual sample can be shared by its neighbors. Therefore, the more representative samples are preferred for labeling. In SInfo, the ambiguity of an photograph is measured by the entropy of the relevance probability distribution while the representation vanes is measured by the density.

A. Ambiguity

The ambiguity denotes the uncertainty whether an photograph is relevant or not. It can be estimated via various sophisticated learning methods, e.g., support vector machine (SVM) [35], transductive SVM (TSVM) [18] and the harmonic Gaussian filed method[42], by conducting a binary classification task.

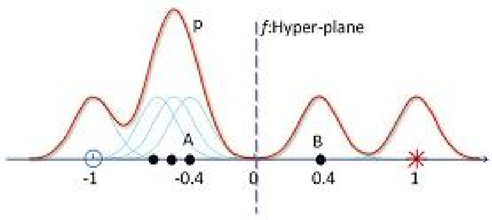


Fig. 2. Because "A" and "B" have the same distance to the hyper-plane (dashed line), they have an identical ambiguity. However, the more representative sample "A" is more preferable to "B".

B. Representativeness

Besides the ambiguity, representativeness, an important property but not well studied before, is also taken into account. Apart from the unreliable estimation led by insufficient labeled photographs, the ambiguity measures the importance of the photograph itself only. Once the Network photograph exploration system gets the labeling information of an photograph, it is very important to consider how many other photographs can share the labeling information with the labeled one. For example, given two unlabeled samples with the identical ambiguity, labeling the more representative one, i.e., many samples are distributed around it, will bring more information and achieve a better reconstruct performance.

C. Dynamic Sample Selection

Since the most informative photographs should meet both ambiguity and representativeness simultaneously, the structural information of photograph , can be measured by the product of the two terms, i.e.,

$$\mathrm{SI}\,(I_i) = p(I_i)H(I_i).$$

Then the most informative photograph is selected from the unlabeled photograph set according to

$$I^* = \arg\max_{I_i \in \mathcal{U}} \mathrm{SI}\,(I_i).$$

IV. LGDDIMENSION REDUCTION

In reconstruct, the photographs returned for a certain query term are represented by low-level visual features, i.e., with the -dimensional visual feature for photograph . The performance of reconstruct is usually poor because of the gap between the low-level visual features and high-level semantics.

We have three types of photographs: labelledrelevant, la-belled irrelevant, and unlabeled. There fore, networkuild 3 types of patches, which are:

1) local patches for labeled relevant images to represent the local geometry of them and the discriminative information to separate relevant photographs from irrelevant ones,

2) local patches for labeled irrelevant images to represent the discriminative information to separate irrelevant photographs from relevant ones, and

3) global patches for both labelledand unlabeled images for transferring both the local geometry and the discriminative information from all labelled photographs to the ulna-belled ones.

For convenience, we use superscript "-" to denote the la-belled relevant photographs and "+" to denote the labelledirrele-vant ones. If there is no superscript, it refers to an arbitrary photograph which may be labelledrelevant, labelledirrelevantor un-labelled.

A. Local Patches for Labelled Relevant Photographs

BDA, a popular dimension reduction algorithm for photograph retrieval, assumes that all query relevant samples are alike while each irrelevant sample is irrelevant in its own way [41]. Thus, the relevant samples are required to be close to each other in the projected subspace. However, this assumption is usually unreliable in Network photograph exploration.

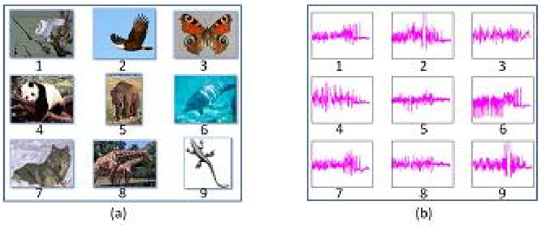


Fig. 3. For query "animal", the query relevant images vary largely in both appearance (a) and visual features (b). In (b), the utilized 428-D visual features include 225-D color moment, 128-D wavelet texture and 75-D edge distribution histogram.

B. Local Patches for Labelled Irrelevant Images

Discriminative information is also partially encoded in allirrelevant photographs, so we construct local patches for labelled irrelevant photographs by separating each irrelevant photograph from all relevant photographs. Because each irrelevant photograph is irrelevant in its own way , it could be unreasonable to keep the local geometry of the irrelevant photographs. In this paper, we model the local patch for the lo w-dimensional representation of labelled irrelevant photograph as

$$\min - \sum_{j=1}^{k} \|\mathbf{y}_i^- - \mathbf{y}_{i_j}\|^2 = \min \operatorname{tr}\left(\mathbf{Y}_i^- \mathbf{L}_i^- (\mathbf{Y}_i^-)^T\right).$$

C. Global Patches for All Photographs

In dynamic reconstruct, users would like to label only a small number of photographs, so it is la viand unreasonable to abandon a large number of unlabelled photographs. With only the labelled photographs, the learned subspace will bias to that spanned by these labelled photographs and cannot generalize well to the large amount of unlabeled data.

We have tried many different dimension reduction algorithms and the results are illustrated in Fig. 4(b)–(k). For each dimension reduction algorithm, we have computed the projection plane(the upper part of subfigure)and the projected 2-D data (the lower part of subfigure). With these conventional algorithms, the relevant and irrelevant samples are overlapped in the projected subspace and the submanifoldof the relevant samples is not well preserved, as illustrated in the figure. This is caused by the problems existing in these algorithms as aforementioned.

To avoid these problems, the proposed LGD learns the sub manifold by transferring both the local geometry and the discriminative information from labelled samples to all unlabeled samples. Global patches are built for each sample (including both labelled and unlabeled) to complete the cross domain knowledge transferring process. According to the alignment scheme in [40], the global patch for the lo w-dimensional representation of the photograph is modeled in a similar way to local patches.

$$\max \operatorname{tr}((\mathbf{y}_i - \mathbf{y}^m)(\mathbf{y}_i - \mathbf{y}^m)^T)$$

D. Patch Coordinate Alignment

Each patches its own coordinate system. With the calculated local and global patches, we can align them together in to a consistent coordinate. For each photograph

$$(\mathbf{S}_i)_{pq} = \begin{cases} 1, & \text{if } p = F_i(q) \\ 0, & \text{else} \end{cases}$$

V. BAYESIAN RECONSTRUCT

To verify the effectiveness of the proposed dynamic reconstruct method, we apply the SInfo dynamic sample selection strategy and the LGD dimension reduction algorithm to reconstruct. Both SInfo and LGD are general and can be directly applied to various reconstruct algorithms, e.g., Visual Rank[17]. In this paper, we take the Bayesian reconstruct as the basic reconstruct algorithm for illustration.

Usually, several interaction rounds are performed to achieve a satisfactory performance. There fore, in next interaction round, SInfo and LGD are performed with the new obtained in the last round. The overall procedure of our dynamic reconstruct is summarized as follows:

VI. EXPERIMENTSON SYNTHETIC DATASETS

In this section, we used three synthetic datasets to illustrate the effectiveness of the SInfo sample selection strategy, as shown in Fig. 5 (top). In each dataset, the relevant samples are marked with red stars ("*") while the irrelevant ones are marked with blue circles ("o"). The initial ranking score list was set randomly since we had note xtual information to simulate the text-based exploration process. At the be ginning stage, one relevant and one irrelevant sample were randomly selected as the labelled set and the rest were taken as the unlabeled. The initial reran ked results ["RerankInitial" curve in Fig. 5 (bottom)] were obtained by reconstruct without user interactions. Parameters in each method were determined empirically in this paper to achieve its best performance.

In each interaction round, only one sample was selected for labeling. For each dataset, we have given the ranked results after 4 interaction rounds with different dynamic sample selection.

We compared SInfo with other three sample selection strategies, i.e., "Error Reduction" [43], "Most Uncertain" [4] and "Random". In "Most Uncertain", the most ambiguity samples are selected for interaction according to (3). While in "Random", the query samples are selected randomly. The comparison results, as shown in Fig. 5 (bottom), demonstrate that the proposed strategy outperforms the rival methods consistently on all three datasets. This is because "Error Reduction" and "Most Uncertain "suffer from the small sample size problem. SInfo is more robust because it takes both ambiguity and representativeness into consideration, and thus alleviates the influence of the small sample size problem.

VII. EXPERIMENTSON NETWORKPHOTOGRAPHEXPLORATION DATASET

We also conducted experiments on a real Network photograph exploration dataset. In this dataset, there are 105 queries selected seriously from a commercial photograph exploration engine query log as well as popular tags of Flickr. These queries cover a large range of topics, including named person, named object, general object and scene. For each query, a maximum of 1 000 photographs returned by commercial photograph exploration engines, i.e., Google, Live and Yahoo, were collected as the initial text-based exploration results. This dataset contains 94341 photographs in total. For each query, three participants were asked to judge whether the returned photographs are query relevant or irrelevant. An photograph is labelled as query relevant if at least two of the three participants judged it as relevant, and vice versa .

A. Dynamic Reconstruct With SInfo

In this section, we will in vest iGATE the effectiveness of S info sample selection strategy and compare it with other three methods: "Error Reduction" , "Most Uncertain" [4], and "Random." To be noted, here both the reconstruct and the dynamic sample selection were conducted in the original feature space. The effectiveness of the LGD dimension reduction algorithm will be discussed in Section VII-B, in comparing with other representative ones.

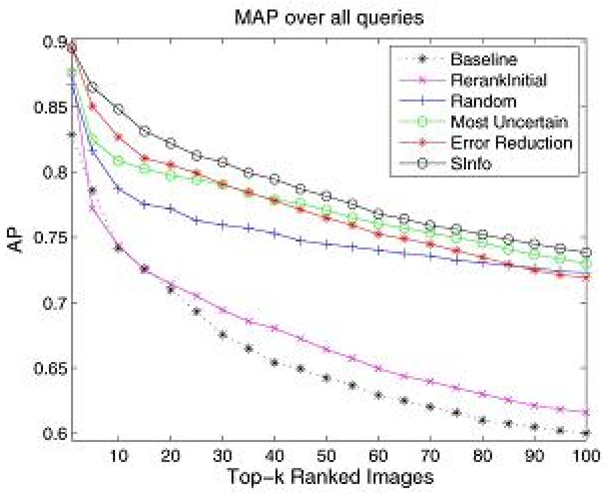


Fig. 6. MAP over all queries with different sample selection strategies.

B. Dynamic Reconstruct With LGD

To test the effectiveness of LGD discussed in Section IV, we conducted the dynamic reconstruct in the projected subspace by using different dimension reduction algorithms. The SInfo sample selection strategy was adopted in this experiment.

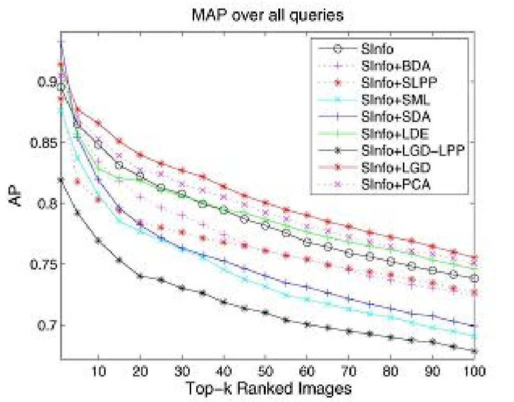


Fig. 7. MAP over all queries with different dimension reduction algorithms.

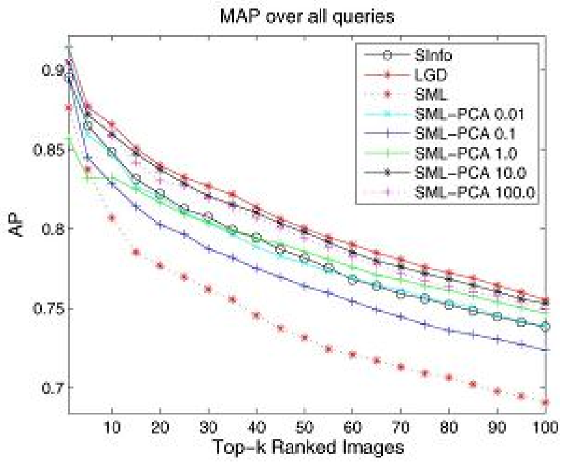


Fig. 8. Performance of SML-PCA.

C. LGD With Random Sample Selection

In Section VII-B, we have shown that, when samples are selected via SInfo, the performance of reconstruct conducted in the original feature space, i.e., the "SInfo" curve in Fig. 7, is consistently improved when LGD is utilized. As illustrated in Fig. 7, "SInfo+LGD" performed better than "SInfo". To verify the sensitivity of LGD to sample selection strategy, we further conducted experiments for LGD when samples were randomly selected.

VIII. PARAMETER SENSITIVITY

In this section, we analyze the sensitivity of important parameters in SInfo and LGD for dynamic reconstruct. The analyses are performed based on the experiments conducted on the Network mage exploration dataset. The experiments are conducted with S info dynamic sample selection and LGD dimension reduction, if not explicitly stated otherwise.

A. Evaluation on Ambiguity Trade-Of f Parameter

The in (3) plays an important role in balancing the ambiguity estimation, which is one of the two critical aspects in SInfo.

B. Evaluation on Local Patch Trade-Of f Parameter We also in vest gated the influence of the trade-off parameter in (7) for LGD when building the local patch for labelled relevant photographs.

C. Evaluation on Local-Global Patch Trade-Of f Parameter

Both the local and global patches reflect data information from different aspects. To in vest iGATE the contributions of these two parts, we have tested the performance of LGD with different trade-offs . When , only local patches are utilized.

D. Evaluation on Number of Inter action Rounds for Dynamic Sample Selection

Morel belled photographs will bring more information and thus a better performance can be achieved. However, users usually lose their patience after a few interaction rounds. Therefore, it is important to find out a good trade-off between the reconstruct performance and the number of the interaction rounds. In this experiment, we investig acted the performance of reconstruct with interaction rounds varying from 1 to 20.In each round, 5 photographs are selected via S Info for labeling. LGD is adopted to learn the effective subspace for reconstruct.

E. Influence of Labelled Image Size on Model Parameters

In Sections VIII-B and C, we have discussed the influence of parameters and in LGD to the reconstruct performance when20photographs (4 interaction rounds with 5 photographs labelled per round) are labelled. In this section, we turn to in vest iGATE the influence of the number of labelled photographs on these model parameters. Fig. 16showsthe performance curves of with different number of labelled photographs while illustrates that of C.

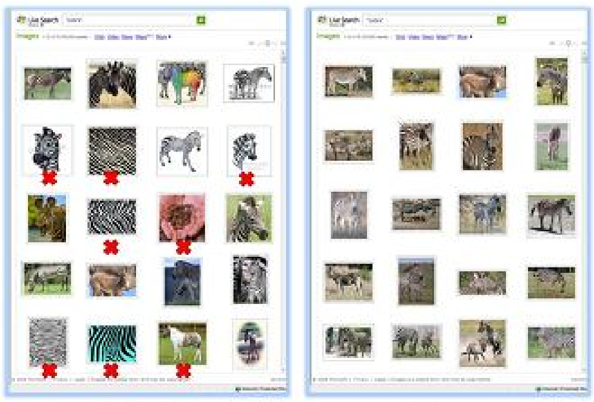
The in (13) is utilized to control the influence of the global patches. Fig. 17(a)shows that a larger is preferred when fewer mages are labelled. With few labelled photographs, little information is contained in them and thus the global patches play the main role. Fig. 17(d)shows that when the number of labelled photographs is augmented, the discriminative information and the local geometry become robust and thus a smaller provides better performance.

F. Evaluation on Dimension of the Projected Subspace

LGDaims to learn a submanifold from the ambient visual feature space to express the user's intention. To find outa proper dimension of the projected feature, the following experiment has been done to in vestigate the influence of the dimension.

Text-based search result Reranked result

Fig. 19. Query "George W. Bush".



Text-based search result Reranked result

Fig. 20. Query "zebra".

IX. CONCLUSION

This paper has presented a novel dynamic reconstruct framework for Network photograph exploration by using user interactions. To target the user's intention effectively and efficiently, we have proposed an dynamic sample selection strategy and a dimension reduction algorithm, to reduce labeling efforts and to learn the visual characteristics of the intention respectively. To select the most informative query photographs, the structural information based dynamic sample selection strategy takes both the ambiguity and the representativeness into consideration. To learn the visual characteristics, a new local-global discriminative dimension education algorithm transfers the local information in the domain of the labelled photographs domain to the whole photograph database. The experiments on both synthetic datasets and a real Network photograph exploration dataset have demonstrated the effectiveness of the proposed dynamic reconstruct scheme, including both the sample selection strategy and the dimension reduction algorithm.

R EFERENCES

[1] D. Cai, X. He, J. Han, and H.-J. Zhang, "Orthogonal laplacianf aces for face recognition, " IEEE Trans. Image Process., pp. 3608–3614, 2006.

[2] D. Cai, X. He, and J. Han, UsingGraphModelfor Face Analysis, Tech. Rep.,2005, Comput. Sci. Dept.Univ. Illinois, Urbana-Champaign.

[3] D. Cai, X. He, and J. Han, "Semi-superviseddiscriminantanalysis," in Proc.IEEE Int. Conf.ComputerVision , 2007, pp. 1–8.

[4] E. Y. Chang, S. Tong, K. Goh, and C.-W. Chang, "Support vector machine concept-dependent dynamic learning for photograph retrieval, " IEEE Trans. Multimedia, 2005.

[5] H.-T.Chen,H.-W. Chang, and T. L. Liu, "Local discriminant embed- ding and its variants, " in IEEE Int. Conf.ComputerVision and Pattern Recognition , 2005, pp. 846–853.

[6] J. Cui, F. Wen, and X. Tang, "Real time google and live photograph exploration re-ranking, " presented at the ACM Int. Conf.Multimedia, 2008.

[7] R. A. Fisher, "The use of multiplemeasurements in taxonomic problems," Ann. Eugen., pp. 179–188, 1936.

[8] Y. Fuand T. Huang, "Photograph classification using correlation tensor analysis," IEEE Trans. Ima ge Process., pp. 226–234, 2008.

[9] Y. Fu, S. Yan, and T. Huang, "Correlation metric for generalized feature extraction, " IEEE Trans. Pattern Anal. Mach. Intell. , pp. 2229–2235, 2008.

[10] X. He,D. Cai, and J. Han, "Learninga maximummargin subspace for photograph retrieval, " IEEE Trans. Knowl. DataEng. , pp. 189–201, 2008.

[11] X. Heand P. Niyogi,"Locality preserving projections, " Adv. Neural Inf. Process. Syst. , 2003.