Integrating Rich Information for Video Recommendation with Multi-Task Rank Aggregation*

Xiaojian Zhao†, Guangda Li‡, Meng Wang‡,
Jin Yuan‡, Zheng-Jun Zha‡, Zhoujun Li†, Tat-Seng Chua‡
†State Key Laboratory of Software Development Environment
Beihang University
Beijing 100191, P.R. China
‡School of Computing
National University of Singapore
Computing 1, Singapore 117590
{zhaoxj01, guanda10, eric.mengwang, junzzustc}@gmail.com
{yuanjin, chuats}@comp.nus.edu.sg lizj@buaa.edu.cn

ABSTRACT

Video recommendation is an important approach for helping people to access interesting videos. In this paper, we propose a scheme to integrate rich information for video recommendation. We regard video recommendation as a ranking problem and generate multiple ranking lists by exploring different information sources. A multi-task rank aggregation approach is proposed to integrate the ranking lists for different users in a joint manner. Our scheme is flexible and can easily incorporate other methods by adding their generated ranking lists into our multi-task learning algorithm. We conduct experiments with 76 users and more than 10, 000 videos. The results demonstrate the feasibility and effectiveness of our approach.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval

General Terms

Algorithms, Experimentation, Performance

Keywords

Video Recommendation, Multi-Task Rank Aggregation, Rich Information

1. INTRODUCTION

With the rapid advances in storage devices, networks and compression techniques, video data from different domains are growing at an explosive rate. For example, the most popular video sharing site, Youtube[1], hosts over 150,000,000 videos [1]. As a result, users are often bewildered by the vast quantity of videos when they want to find some interesting video contents. Therefore, automatic video recommendation becomes highly desirable to tackle this information-overload problem by presenting lists of potentially interesting videos to the users.

*Area chair: Hari Sundaram

[1]http://www.youtube.com

MM'11, November 28–December 1, 2011, Scottsdale, Arizona, USA.

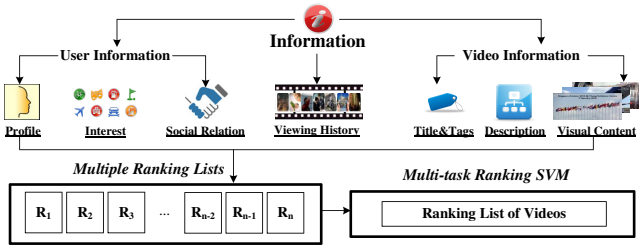

Figure 1: The schematic illustration of the proposed video recommendation scheme that explores multi-source information

Several efforts have been dedicated to video recommendation that explore different information sources. The recommendation services on commercial video websites, such as Youtube, Yahoo! Video[2] and Bing Videos[3], are usually built based on the textual information of videos or users' profiles (only for registered users) [12]. Some methods analyze the interest relationship of users by mining the viewing history [3] [15], while several others recommend videos based on the users' profile [11] or utilize users' viewing histories [19]. However, they all lack an unified scheme that is able to integrate multiple information sources for video recommendation.

In this paper, we propose a personalized video recommendation scheme by integrating rich information. Figure 1 demonstrates the schematic illustration of the approach. We regard video recommendation as a ranking task, i.e., to rank videos in a database with the expectation that interesting videos for a specific user can be prioritized in the list. We explore rich information of both the users and videos, including the education background, occupation, location, interests, past viewing history, social network of users; as well as the title, description, tags and visual content of videos. Multiple ranking lists are generated based on different information sources and we employ a multi-task rank aggregation approach to fuse the ranking lists to generate the final recommendation.

The main contributions of this paper are as follows:

(1) To the best of our knowledge, this is the first work that integrates rich information for video recommendation. We do not only provide a solution but also analyze the effectiveness of different information sources for video recommendation.

(2) We propose a multi-task rank aggregation algorithm that is able to accomplish user-adaptive rank aggregation. We show that it is superior to approaches that learn a global rank aggregation model as well as those that learn rank aggregation for each user individually.

[2]http://video.yahoo.com

[3]http://www.bing.com/videos/browse

2. RELATED WORK

Research on video recommendation mainly focuses on three typical approaches, namely, *collaborative filtering* (CF), *content-based filtering* (CBF), and *hybrid filtering* (HF) that combines the above approaches [2]. The CF approaches compare a user's ratings of videos with those of hundreds of others, find people who share similar preferences, and then recommend videos that are interesting for those people with similar preferences. For example, Setten et al. [14] proposed to use different social filtering methods to predict user interest based on other users' information, and designed a combination of prediction techniques. Baluja et al. [3] built a user-video graph which represents the co-view information among different users and its recommendation was performed by a graph propagation in which the label of each note was obtained from its neighbors. For CBF approaches, videos can be recommended based on previous user viewing information for the videos. For example, Mei et al. [13] presented an online video recommendation system, VideoReach, using multimodal relevance between videos and users' click-through data. They considered three modalities, textual, visual and aural, and combined the relevance scores from them by using attention fusion function. HF approaches combine the above two approaches in a single framework [2]. For example, Burke [4] employed mixture models which build the recommendation based on a linear combination, voting, or selection of the content-based prediction and collaborative prediction. Our proposed approach is a flexible approach and is able to integrate different methods. In fact, we can generate ranking lists with different methods or information sources and then employ the multi-task ranking approach to integrate all the ranking lists.

3. RANKING LIST GENERATION WITH RICH INFORMATION

In this section, we introduce the ranking lists generated with rich information. Here we organize them according to the involved information sources about the users including profile, viewing history and social network, as well as the videos. In addition, we also generate ranking lists based on collaborative filtering, the most typical recommendation approach.

3.1 Profile Based Ranking

We collect a set of profile for each user, including major, degree, occupation, interests and location. The user's profile information is represented with texts. Table 1 gives an example. Since we have the text information of each video, including title, description and tags, we can simply rank videos based on their textual similarity with user's profile information. Here we generate 3 ranking lists, which are based on user's interests, location and other information respectively.

3.2 History Based Ranking

The viewing history contains the list of videos that have been accessed by the user previously. It is able to reflect the interests of the user. We thus rank videos by estimating their similarities with the history. Here we consider three types of viewing history: recent (i.e., the most recent video viewed by the user), short-term history (the videos viewed on that day), and long-term history (all the past history). For each type of viewing history, we generate 2 ranking lists, one by measuring the visual similarity of videos and the other by textual similarity. Therefore, we generate 6 ranking lists based on the user's history information.

3.3 Social Network Based Ranking

Many recommendation algorithms, such as collaborative filtering, are built by mining the interest relationship of a large number of users [2] [10] [18]. But in fact, a user usually has close interests with his/her friends. For example, when a user wants to enjoy some videos, it is natural for him/her to turn to friends for suggestions. With the rapid advances of social network, we can easily collect information on a user's friends and explore their video viewing histories. Our approach is as follows. We denote $\mathcal{V}_i$ as the video set that has been viewed by the *i*-th friend. We integrate all the videos viewed by the users' friends into a video set $\mathcal{V}$, i.e., $\mathcal{V} = \mathcal{V}_1 \bigcup \mathcal{V}_2 \bigcup ..... \bigcup \mathcal{V}_r$, where r is the number of friends. We then generate ranking lists with two methods. For the first method, we set the ranking score of 1 to videos in $\mathcal{V}$ and the score of 0 to all the other videos. Since we can set $\mathcal{V}_i$ to either the recent one, short-term history, or long-term history of the *i*-th friend, we can obtain 3 ranking lists with this method. Note that these ranking lists are actually not stable as there are only two scores, 0 and 1, for the videos. But it is not a problem for generating the final recommendation list since our rank aggregation, which will be introduced in the next section, is a score-based fusion approach. In the second method, we still put the videos in $\mathcal{V}$ on top of the ranking list, but for the other videos, we rank them according to their similarities with $\mathcal{V}$. Since we can adopt visual and textual similarity respectively, so we can generate 3 ranking lists for each. Therefore, we generate 9 ranking lists based on users' social network information.

Table 1: The profile information of an exemplary user

Profile Item	Information
Major	Computer Science
Degree	Ph.D.
Occupation	Researcher
Interests	Reading;Swimming;Movie;Travel
Hometown	San Francisco, California
Working Place	Mountain View, California

3.4 Collaborative Filtering Based Ranking

Collaborative filtering is the most widely-adopted approach for recommendation. It is usually accomplished by mining a user-item matrix. There are two typical approaches, one is estimating the similarity of users [2] and the other is estimating the similarity of items. Here we employ these two methods to generate two ranking lists and integrate them into our approach.

4. MULTI-TASK RANK AGGREGATION

After generating multiple video ranking lists using different information sources, our next task is to aggregate these video lists into an optimized video list such that the top items can be recommended to the users. It can be formulated as a rank aggregation problem [9]. Rank aggregation methods usually can be categorized into two approaches, namely, the rank-based approach and score-based approach [16]. Here we adopt the score-based approach, i.e., we fuse ranking lists according to the ranking scores of each video instead of their ranking positions. There are two straightforward approaches, one is to learn a global rank aggregation model and the other is to learn a rank aggregation model for each user. However, for the first approach, the behavioral differences of different users are overlooked; while the second approach does not consider the dependency among users and this may make the learned model unreasonable if training data are limited to a user. Therefore, we propose a multi-task learning approach that is able to simultaneously learn models for multiple users such that the correlation of the users can be explored [7]. We build our approach based on ranking SVM [5], and the method is named multi-task ranking SVM.

We first introduce several notations. We denote $\mathbf{v}_i \succ \mathbf{v}_j$ and $(\mathbf{v}_i, \mathbf{v}_j) \in \mathcal{R}$, if a video $\mathbf{v}_i$ is ranked higher than a vdieo $\mathbf{v}_j$ in

an order $\mathcal{R}$. Otherwise, we denote $(\mathbf{v}_i, \mathbf{v}_j) \notin \mathcal{R}$. We assume for simplicity that $\mathcal{R}$ has strict ordering, which means that, for all pair $\mathbf{v}_i$ and $\mathbf{v}_j$ in $\mathcal{R}$, we have either $\mathbf{v}_i \succ \mathbf{v}_j$ or $\mathbf{v}_i \prec \mathbf{v}_j$. Let $\mathcal{R}^*$ be the optimal ranking of video in which the video is ordered perfectly according to the user's preference. A training set for multi-task ranking SVM is a set of partial orders $\mathcal{R}^* \subset \mathcal{R}$, which are the total number of pair wise orderings. The target of the multi-task ranking SVM is to learn a function f_k that satisfies $f_k(\mathbf{v}_i) > f_k(\mathbf{v}_j)$ for all pairs of $(\mathbf{v}_i, \mathbf{v}_j) \in \mathcal{R}$ given the k-th user. For simplicity, we assume $f_k(\mathbf{v}) = \mathbf{w}_k \cdot \mathbf{v}$. We assume that for every user, all $\mathbf{w}_k$ can be defined:

$$\mathbf{w}_k = \mathbf{w}_0 + \Delta\mathbf{w}_k \tag{1}$$

where $\mathbf{w}_0$ can be viewed as a common part of all users and $\Delta\mathbf{w}_k$ indicates the distinct difference of the k-th user. The vectors $\Delta\mathbf{w}_k$ are usually enforced to be *small*, i.e. we assume that the tasks are related in a way that the true models are all close to some model $\mathbf{w}_0$. We then estimate all $\Delta\mathbf{w}_k$ as well as $\Delta\mathbf{w}_0$ simultaneously. To this end, we solve the following optimization problem, which is analogous to SVM [7]:

$$\begin{aligned} &\min \sum_{k=1}^{K}\sum_{i=1}^{n_k} \xi_{ki} + \lambda_1 \sum_{k=1}^{K} ||\Delta\mathbf{w}_k||^2 + \lambda_2 ||\mathbf{w}_0||^2 \\ &s.t.\ (\mathbf{w}_0 + \Delta\mathbf{w}_k)(\mathbf{v}_{ki}^{(1)} - \mathbf{v}_{ki}^{(2)}) \geqslant 1 - \xi_{ki}\ \ \xi_{ki} \geqslant 0 \end{aligned} \tag{2}$$

In the above equation, n_k denotes the number of training sample pairs for the k-th user, $(\mathbf{v}_{ki}^{(1)}, \mathbf{v}_{ki}^{(2)})$ refers to the k-th training pair for the i-th user, λ_1 and λ_2 are the positive regularization parameters, and the ξ_{ki} are slack variables that measure the error of the final models $\mathbf{w}_k$.

It can be derived that the optimal solution to the optimization problem is:

$$\Delta\mathbf{w}_k^* = \tfrac{T}{2\lambda_1} \sum_{i=1}^{m} \alpha_{ik}(\mathbf{v}_{ik}^{(1)} - \mathbf{v}_{ik}^{(2)}) \tag{3}$$

$$\mathbf{w}_0^* = \tfrac{\lambda_1}{K\lambda_2} \sum_{k=1}^{K} \Delta\mathbf{w}_k^* \tag{4}$$

where α_{ik} is the nonnegative Lagrange multipliers. The dual formulation for the above problem is:

$$\begin{aligned} \max_{\alpha_{ik}} \{ &-\frac{1}{2} \sum_{i=1}^{m}\sum_{s=1}^{K}\sum_{j=1}^{m}\sum_{k=1}^{K} \alpha_{is}\alpha_{jk} K_{sk}(\mathbf{v}_{is}^{(1)} - \mathbf{v}_{is}^{(2)}, \mathbf{v}_{jk}^{(1)} - \mathbf{v}_{jk}^{(2)}) \\ &+ \sum_{i=1}^{m}\sum_{k=1}^{K} \alpha_{it} \} \end{aligned} \tag{5}$$

$$s.t.\ 0 \leq \alpha_{ik} \leq T/2\lambda_1$$

where

$$K_{sk}(\mathbf{y}, \mathbf{z}) := (\tfrac{\lambda_1}{K\lambda_2} + \delta_{sk})\mathbf{y} \cdot \mathbf{z},\ \ s, k = 1, \ldots, T \tag{6}$$

Accordingly, we can obtain the optimal $\mathbf{w}_0$ and $\Delta\mathbf{w}_k$. By performing rank aggregation with the weight vector $\mathbf{w}_0 + \Delta\mathbf{w}_k$, we can obtain the final ranking list for recommendation.

5. EXPERIMENTS

5.1 Experimental Settings

We conduct experiments with 76 participants from two countries. They are all active YouTube and Facebook users from different backgrounds. The video viewing behavior on YouTube of these users were tracked in a one-month period (from Dec. 2010 to Jan. 2011). It is shown that there are about 150 videos viewed per user on average. The profile and social relationship information among the 76 users are collected. It is shown that the number of friends of the users can vary from 3 to 18. The videos and their associated information, including title, description and tags are all collected. In this way, we have collected 11400 videos in all. For each user, we split the viewed videos into two parts, the first part is the videos viewed in the first two weeks and the second part is the videos viewed in the next two weeks. The second part is used for testing. That means, we regard videos in the second part as relevant samples for recommendation[4]. We assign the relevance scores of 1 and 0 to the videos in the second part and the other videos for a user, respectively.

Table 2: The comparison of exploring different information sources for video recommendation (For the history type, R, S and L indicate the recent video, short-term history, and long-term history respectively).

Ranking method		Textual Information			Visual Information		
User Information	interest	0.34			-		
	location	0.23			-		
	profile	0.22			-		
		R	**S**	**L**	**R**	**S**	**L**
History-Based		0.36	0.33	0.30	0.29	0.28	0.26
SN-Based		0.31	0.34	0.32	0.26	0.27	0.27
CF Based	item	0.22					
	user	0.23					
All Information		**0.41**					

In our ranking list generation method, we need to estimate the similarity of two sets of textual terms and the visual similarity of two videos. For calculating the similarity of two sets of textual terms (it needs to be used in our ranking list generation approach), we first compute the similarity of each pair of textual terms across the two sets based on Google distance [6] and then use the average result. To estimate the visual similarity of two videos, we first segment each video into shots and extract a representative key-frame from each shot. For each key-frame we extract 428-dimensional global visual features for each image [20], including 225-D block-wise color moments generated from a 5-by-5 fixed partition of the image [17], 128-D wavelet texture, and 75-D edge direction histogram [8] [21]. Then the visual similarity of two videos is calculated by averaging the similarities of all key-frame pairs across the two videos.

For the training of multi-task rank aggregation model, we further split the videos of the first two weeks for each user into two parts, one is the videos viewed in the first week and the rest videos form the second part. We generate ranking lists and learn the rank aggregation model based on the videos of the second part. The parameters λ_1 and λ_2 in Eq. 2 are jointly tuned with 5-fold cross-validation (note that here cross-validation is accomplished by dividing users into different folds). For performance evaluation metric, we adopt normalized discounted cumulative gain (NDCG) [18].

5.2 Experimental Results

We first compare our approach that integrates all information sources with methods that only use part of the information. We compare our approach with 17 methods that use only part of the information by combining different information sources. For the

[4]It is worth noting that this setting actually makes the performance underestimated, as the users may also be interested in the videos out of the second part. So a most rigorous approach for ground truth establishment is to let users label all videos with interestingness. But our approach is still reasonable for comparing different algorithms and it is a widely-adopted approach for avoiding the intensive labeling cost.

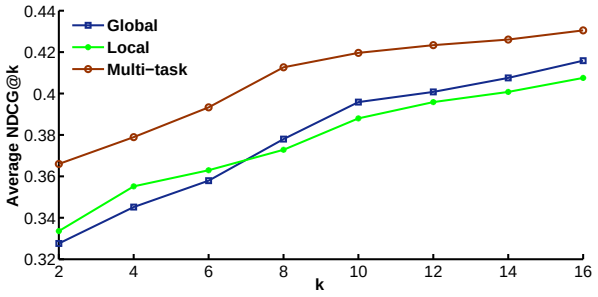

Figure 2: The performance variation of different ranking strategies in different depth

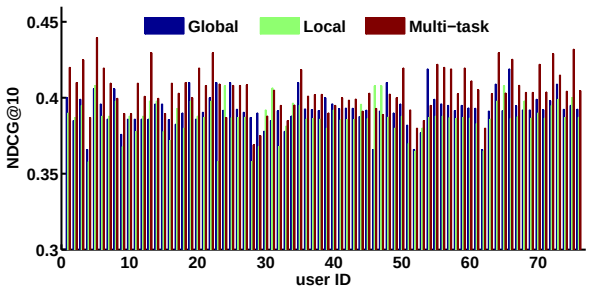

Figure 3: The performance comparison of the three rank aggregation methods for each user. NDCG@10 is used as the performance evaluation metric here.

methods that involve multiple ranking lists, we employ the multi-task rank aggregation approach to fuse them. Table 2 illustrates the comparison of average NDCG@10. From the results, we have the following observations. First, text-based recommendation consistently achieves better performance than visual content based recommendation. Second, among the three types of history information, recent based recommendation performs better than short-term history and long-term history. This is because many users' interests are temporally continuous and thus the continuously viewed videos are usually close. However, it is not the case for SN-based method because the longer viewing histories from friends' can better reveal the user's interest. Finally, we can see that integrating all information sources can achieve the best result.

We also compare our multi-task rank aggregation approach with the following two methods for rank aggregation:

(1) Learning a global rank aggregation model for all users, i.e., the weighting vectors of all users are identical.

(2) Learning a specific rank aggregation model for each user, i.e., $\mathbf{w}_0$ is removed in Eq. 2.

We denote our approach and these two methods as "Multi-Task", "Global" and "Local", respectively. Both of the "Global" and "Local" methods are based on Ranking SVM; the only difference is that, one learns a model for all users and the other learns a model for each indiviual user. In these two methods, the parameter of Ranking SVM is tuned to its optimal value. These three methods are used to fuse the 20 ranking lists introduced in Section 3. Figure 2 illustrates the comparison of average NDCG at different depths. We can see that our approach consistently outperforms the other two methods. Figure 3 illustrates the detailed NDCG@10 results for the 76 users. We can see that for most users our approach achieves the best results. But it is worth noting that, by adopting the multi-task rank aggregation approach, we also introduce a problem for new users. That is, how to learn rank aggregation model for the new user. We can adopt the following strategy. If a new user is associated with enough training data, we can learn a rank aggregation model for the user individually. Otherwise, we can employ several recently proposed model adaptation methods that are able to adapt the models of existing users to this new user.

6. CONCLUSION AND FURTHER WORK

In this paper, we proposed a video recommendation scheme that is able to integrate multiple information sources with a multi-task rank aggregation approach. Ranking lists are generated by exploring different information sources and we then fuse the ranking lists with a multi-task learning approach. We conducted experiments on more than 11,000 videos and the results demonstrated the feasibility and effectiveness of our approach. Our scheme is flexible and different ranking methods can be easily integrated as we only need to fuse several ranking lists in the aggregation step. In our future work, we will conduct experiments with more users and videos and we will also investigate the problem of new users with empirical justification.

7. ACHNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (61170189, 60973105), the Fund of the State Key Laboratory of Software Development Environment under Grant No. SKLSDE-2011ZX-03 and the Singapore National Research Foundation & Interactive Digital Media R&D Program Office, MDA under research grant (WBS:R-252-300-001-490).

8. REFERENCES

[1] Encyclopedia. *http://en.wikipedia.org/wiki/YouTube/*.

[2] G. Adomavicius and A. Tuzhilin. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6):734–749, 2005.

[3] S. Baluja, R. Seth, D. Sivakumar, Y. Jing, J. Yagnik, S. Kumar, D. Ravichandran, and M. Aly. Video suggestion and discovery for youtube: taking random walks through the view graph. In *WWW*, pages 895–904, 2008.

[4] R. Burke. Hybrid web recommender systems. *Lecture Notes in Computer Science*, pages 377–408, 2007.

[5] Y. Cao, J. Xu, T.-Y. Liu, H. Li, Y. Huang, and H.-W. Hon. Adapting ranking SVM to document retrieval. In *ACM SIGIR*, pages 186–193, 2006.

[6] R. L. Cilibrasi and P. M. Vitányi. The google similarity distance. *IEEE Transactions on Knowledge and Data Engineering*, 19(3):370–385, 2007.

[7] T. Evgeniou and M. Pontil. Regularized multi–task learning. In *ACM KDD*, pages 109–117, 2004.

[8] B. Geng, L. Yang, C. Xu, and X. Hua. Content-aware ranking for visual search. In *CVPR*, pages 3400–3407, 2010.

[9] B. Geng, L. Yang, C. Xu, and X.-S. Hua. Ranking model adaptation for domain specific search. *IEEE Transactions on Knowledge and Data Engineering*, 2010.

[10] X. Hu, L. Tang, and H. Liu. Enhancing accessibility of microblogging messages using semantic knowledge. In *ACM CIKM*, 2011.

[11] H. Luo, J. Fan, and D. A. Keim. Personalized news video recommendation. In *ACM MM*, pages 1001–1002, 2008.

[12] T. Mei and K. Aizawa. Video recommendation. In *Chapter of Internet Multimedia Search and Mining*. Bentham Science Publisher, 2011.

[13] T. Mei, B. Yang, X.-S. Hua, L. Yang, S.-Q. Yang, and S. Li. Videoreach: an online video recommendation system. In *ACM SIGIR*, pages 767–768, 2007.

[14] M. van Setten, M. Veenstra, A. Nijholt, and B. van Dijk. Prediction strategies in a TV recommender system-method and experiments. In *WWW*, pages 203–210, 2003.

[15] M. Wang and X.-S. Hua. Active learning in multimedia annotation and retrieval: A survey. *ACM Transactions on Intelligent Systems and Technology*, 2(2):10, 2011.

[16] M. Wang, X.-S. Hua, R. Hong, J. Tang, G. Qi, and Y. Song. Unified video annotation via multigraph learning. *IEEE Transactions on Circuits and Systems for Video Technology*, 19(5):733–746, 2009.

[17] M. Wang, X.-S. Hua, J. Tang, and R. Hong. Beyond distance measurement: constructing neighborhood similarity for video annotation. *IEEE Transactions on Multimedia*, 11(3):465–476, 2009.

[18] M. Wang, K. Yang, X.-S. Hua, and H. Zhang. Towards a relevant and diverse search of social images. *IEEE Transactions on Multimedia*, 12(8):829–842, 2010.

[19] B. Yang, T. Mei, X.-S. Hua, L. Yang, S.-Q. Yang, and M. Li. Online video recommendation based on multimodal fusion and relevance feedback. In *ACM CIVR*, pages 73–80, 2007.

[20] Z.-J. Zha, X.-S. Hua, T. Mei, J. Wang, and Z. Wang. Joint multi-label multi-instance learning for image classification. In *CVPR*, pages 1–8, 2008.

[21] Z.-J. Zha, L. Yang, T. Mei, M. Wang, and Z. Wang. Visual query suggestion. In *ACM MM*, pages 15–24, 2009.