

Multiple User Behavior Learning for Enhancing Interactive Image Retrieval

Yiu-ming Cheung, Fellow, IEEE, and Sheung Wai Chan

Department of Computer Science, Hong Kong Baptist University
Hong Kong SAR, China

ymc@comp.hkbu.edu.hk, cs_swchan@comp.hkbu.edu.hk

Abstract. The existing image retrieval approaches focus on the behavior of a single user only in each query without considering the correlation of the behaviors of multiple users in performing similar queries. In fact, users would have similar behaviors while they have similar expectations during queries. Accordingly, this paper therefore proposes the interactive image retrieval framework with the Similar Behavior Learning model. The framework consists of two stages. In the first stage, the framework retrieves images with the content-based feature vector as preliminary query result for user selection. In the second stage, the SBL model determines the similarity of the user behavior and annotates label code to the selected images instantly. The images are indexed by label code can be retrieved more efficiently. Meanwhile, the selected images in preliminary result are used as additional information for retrieving better results at the end of the current query. Experiments show the promising results.

Keywords: Multiple User Behavior Learning, Content-based Image Retrieval

1 Introduction

The information of user behavior is important for image retrieval (IR), which can help the retrieval result much closer to human cognition. In the literature, the methods of IR learning can be summarized in two lines: supervised learning and reinforcement learning. Both learning methods require to collect user behavior after each query iteration. In supervised learning, the logs of user behavior in each iteration can be considered as labeled data in order to get the optimized parameters of the algorithm. The supervised learning framework needs to be trained by using labeled data initially before the system operation. In reinforcement learning, the framework keeps learning through user behavior in each iteration. The learning outcome from the model can contribute to retrieval performance improvement in the subsequent iterations. Both of them target to extract knowledge from user behavior for fine-tuning the system. However, an image retrieval system where user behavior information is involved still suffers from the following issues: (1) System fine-tuning in each iteration is based on user behavior from a single user in each query without considering the correlation of the behaviors of multiple users in performing similar queries. In fact, users would have similar behaviors while they have similar expectations during queries in using the IR system. (2) The image retrieval system is used by multiple users, the type and ordering of image queries are

unpredictable. How to identify similar queries from different user behavior and assign correct label code to the selected images instantly in each iteration become a challenge.

To this end, an interactive image retrieval framework with Similar Behavior Learning (SBL) model is proposed. The query process of the proposed framework consists of two stages: In the first stage, the framework generates preliminary query based on the image sample provided by user and the learning result from SBL model. In the second stage, final query is retrieved based on the image selection in preliminary query. User is required to select images from preliminary query result for collecting the information of user behavior. The information is used to retrieve the final query result. Simultaneously, the SBL model determines the similarity of the current query with previous queries according to the user behavior in preliminary query stage. The learning outcome of SBL mode is represented as label code and recorded in database. As aforementioned, users would have similar behaviors when they have similar expectations. If user behavior is detected similar to a previous user behavior record, the selection of the images in the current iteration is considered as the same concept with the selected image in the previous record. The same label code is assigned to the selected images. With continuous learning from different users by SBL model, the images in dataset are annotated with proper label code for future usage.

The contributions of the proposed framework include: (1) User behavior information by multiple users are considered. The proposed SBL model can annotate the images which are select in the query by determining the similarity of user behavior in each iteration instantly. (2) The information of user behavior is presented as a label code and annotate the selected images in every iteration. Based on the annotated images which are learned by SBL model cumulatively in each iteration, the performance of retrieval results keeps improving. Also, the annotated images can help to improve the retrieval process more efficient.

2 Related Work

The proposed IR system collects the information from the aspect of the objective and subjective way for retrieving the expected images. The content-based [12,9] information is considered as objective information which can be extracted from any images. Content-based information usually is visual information that can be extracted from images itself, like their shape, color or texture [23,10]. Consistent content can be extracted from the same image if the same feature extraction methods are used. However, there is a semantic gap issue [2,16] that the content-based information may not match with human cognition. For example, a red apple and a red car may be considered as similar images because they have similar color information. To overcome the limitation, researches [21,19] used hybrid-based methods for combining context-based [5,11,3] and content-based information to retrieve their expected images.

Besides the content-based information, user behavior can be considered as objective information that is much closer to human cognition. To fully utilize the information from user behavior, some researches have already applied the user behavior with supervised learning method. In the research [20], the correlation between interaction signal and user examination analysis is conducted in Web image search. They proposed a grid-

based user browsing model (GUBM) which is inspired by analyzing commercial search logs. The model can capture user behaviors like cursor hovering and alleviate position bias during the interaction. The model can estimate images of topical relevance level and image quality. In the research [1], user behavior data has been utilized, like click-through features, browsing features, query-text features and the human judges with a six-point scale ranging, which is saved as logs file from search engine. The logs file need to prepare for analyzing in order to optimize the ranking function.

Reinforcement learning (RL) [15,6,14] is one of research areas in machine learning. Comparing with supervised learning, RL does not require prepared label data for initial training, the performance of the framework with RL keeps changing sequentially through continuously learning. The typical example is Q-learning [18], the system generates Q-Table that records all the states of each iteration, give an award for the positive statues and penalty for the negative status. After a certain number of iterations, the system can learn a perfect Q-Table what is the best way to achieve the expected result. In the research [13], they have applied a weighted trace transform to solve the problems of IR. Through the techniques of reinforcement learning, the parameters in the algorithm keep fine-tuning in order to improve the performance. In the research [4], they have provided interactive interface allows users to select relevant images to direct the exploration of the image space during the query. Users are able to explore the images by scoring them which are currently on display. In the research [22], they mainly focus on integrating relevance feedback techniques by using reinforcement learning. The retrieval system does a relevance learning based on the previous query. The learning scheme keeps going to the next iteration as long-term learning.

3 The Proposed Approach

In this paper, we propose a two-steps interactive image retrieval framework with SBL model. In the proposed SBL model, the information of user behavior not only benefits the current iteration but also the following iterations as long-term learning. Meanwhile, the proposed framework design can encourage the users to provide valuable feedback which is the user behavior for bringing positive impact in the learning. The proposed SBL model is applied to interactive IR as an example for illustrating. The framework can be divided into two parts: two-step interactive image retrieval framework and SBL model.

3.1 Two-Step Interactive Image Retrieval System

The query process of interactive image retrieval consists of two steps, preliminary query and final query. Let q be the expected image sample that is provided by user. Images are described by the 109-dimension feature vector which is used in the paper [7] as subjective information. The 109-dimension feature vector $f(q)$ is extracted from the provided image sample and $f^n(q)$ be the elements in this feature vector is presented. x_j represents the images in dataset, $f(x_j)$ be 109-dimension feature vector of dataset image and its elements of feature vector $f^n(x_j)$. The distance of images based on Euclidean distance by equation 1.

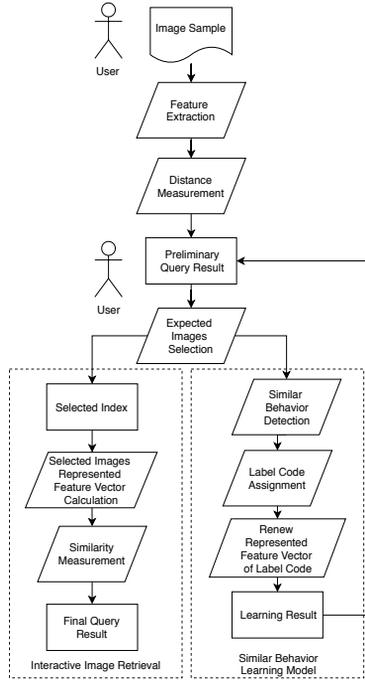


Fig. 1: The flowchart of the interactive image retrieval framework with SBL model.

$$D(q, x_j) = \sum_{n=1}^N (f^n(q) - f^n(x_j))^2 \quad (1)$$

In the proposed framework, user is able to select similar expected images from the preliminary query result. Not all similar expected images are required to be selected. The number of image selection in the preliminary query result only affects the learning rate, the speed of image annotation in the dataset. The selection in the preliminary query result is defined as user behavior in this paper. The basic procedure of the proposed framework is shown in figure 1.

At the left of the figure 1, user provides an image sample for the query. The feature vector of the image sample is extracted. The framework calculated the distance between the image sample and the images in the dataset. The similar images are shown in the result of preliminary query. User can select their expected images from the result of preliminary query. Then the selected images are considered as new image samples input and retrieve again. The more image samples are provided to the framework, the more accurate result is shown in final query result. Meanwhile, the SBL model can utilize the information of user behavior for long-term learning in order to improve the following iterations. The proposed two-step interactive image retrieval framework encourages user to provide more information in order to improve the retrieval accuracy of the result in

the current iteration. Furthermore, it also provides fine-tuned preliminary query result in the next iteration based on the user behavior from previous iterations after learning from SBL model.

3.2 Similar Behavior Learning (SBL) Model

In this framework, the SBL model is integrated into the query process, as shown in figure 1. The learning is started to process after the user behavior is collected in the preliminary query of each iteration as an input. Let s_p be the selected images P in preliminary query result. When the number of iterations = 1, all images in preliminary query result are based on the similarity between q and X by equation 1. All selected images from preliminary query result S are assigned label code $l_p = 1$ and calculate represented feature vector save as $mf(l_p)$ by equation 2 for initialization.

$$mf^n(l_p) = \frac{1}{p} \sum_{p=1}^P f^n(s_p) \quad (2)$$

When the number of iterations > 1 , the images in the preliminary query result come from $imageset_A$ and $imageset_B$. $imageset_A$ contains the images which is similar to q by measuring distance with equation 1. $imageset_B$ contains the images which have proper label code l . Proper l means that the represented feature vector $mf(l)$ with corresponding label code have highest similarity from q , it shows as equation 3.

$$l = \arg \max_{l_p} D(q, mf(l_p)) \quad (3)$$

According to the user behavior in preliminary query result, there are two cases. In figure 2, the images marked as "A" come from $imageset_A$ and "B" come from $imageset_B$. The images with the red rectangle are the selection in the preliminary query result. Let G_A and G_B be the number of images selected in $imageset_A$ and $imageset_B$ in P respectively. After the selection in the preliminary query result, the SBL model assigns exiting or new label code to the selected images according to the similarity of user behavior, Case 1 or Case 2. The label code annotation based on user behavior in SBL model is shown as algorithm ??.

The selected images in the preliminary query result are annotated label code according to different cases as follows:

Case 1 - ($G_A > 0$): Only the images in $imageset_A$ are selected, see the example of case 1 in figure 2. We can consider that the images with proper l ($imageset_B$) which have the highest similarity with query images are not the images that are expected by user. The images assigned label code l can be determined as non-expected image. As a result, the SBL model determined that the current user behavior is not similar to any past cases. The selected images in preliminary query result in the current stage should be assigned a new label code and saved in the dataset.

Case 2 - ($G_A > 0$ and $G_B > 0$): The images in both $imageset_A$ and $imageset_B$ are selected, see the example of case 2 in figure 2. We can consider that the images in $imageset_A$ are similar to the images in $imageset_B$ while SBL model determined that current user behavior is similar with a user behavior case from previous iterations. The selected images should be assigned same l with $imageset_B$. The selection of l is based on the similarity between the query image q and represented feature vector of each assigned label code $mf(l_p)$.

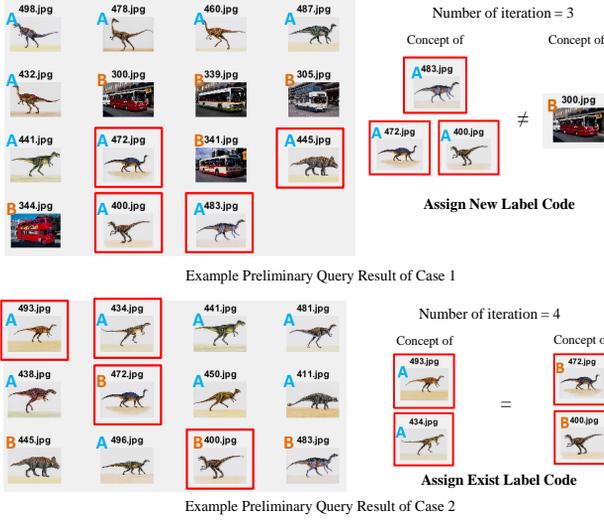


Fig. 2: In case 1, only the images retrieved with content-based information (marked as A) are selected. SBL model determined that the current user behavior is not similar to previous cases. A new label code is assigned to those selected images. In case 2, images retrieved with content-based information (marked as A) or label code learned by SBL model (marked as B) are both selected. SBL model determined the current user behavior is similar to previous queries. Corresponding label code is assigned to those selected images

The images of l_p have been changed, so the represented feature vector $mf(l_p)$ of the corresponded label code should be updated. The updated process is based on the confidence of the images in the l_p . The confidence of image s_p with label code l_p can be represented as $c(s_p)$. The update of $mf(l_p)$ is shown in equation 4.

$$mf^n(l_p) = \frac{1}{\sum c(s_p)} \sum_{p=1}^P c(s_p) f^n(s_p) \quad (4)$$

The confidence c is the frequency of images selected in the history of user behavior. The number of times is selected with the same label code can be considered as the correctness of the label code assignment. In other words, the images with higher c

illustrate that their feature vectors can more represent their corresponding l_p because they are assigned the same label code repeatedly. Hence, more information of images with higher c should be provided to contribution in represented feature vector calculation of corresponding l_p . In this paper, the number of $imageset_A$ is defined as double of $imageset_B$ that it shows in preliminary query result. The ordering of images in $imageset_A$ and $imageset_B$ in the preliminary query result is randomly assigned. It can avoid determining Case 1 or Case 2 incorrectly in case of user selects too few images or just select the first few images. Also, if the images in $imageset_B$ are not be selected, it can provide more evidence to explain that the label code of $imageset_B$ are not what user expected in the condition of the number of $imageset_A$ is already double of $imageset_B$. In this iteration, the selected images in $imageset_A$ should be assigned a new label code.

In the proposed framework, the changing of learning from the user behavior in each iteration is relatively small. It can avoid overwriting the effects of earlier learning. From equation 4, the $mf(l_p)$ is determined by $c(s_p)$ and $f(s_p)$, so individual error events do not have a fatal impact on the whole learning process.

No. of Iteration of Queries	Query Image ID (Concept)	Selected Images in Preliminary Query Result	Similar Behavior Learning Log		
			Annotated Label Code	Image ID of Label Code	Confidence of Images of Label Code
1	300 (Bus)	300, 309, 341, 305	1	300, 305, 399, 341	1, 1, 1, 1
2	344 (Bus)	305, 339, 344, 300	1	300, 305, 339, 341, 344	2, 2, 2, 1, 1
3	400 (Dinosaur)	400, 472, 483, 445	1	300, 305, 339, 341, 344	2, 2, 2, 1, 1
			2	400, 445, 472, 483	1, 1, 1, 1
4	450 (Dinosaur)	472, 493, 434, 400	1	300, 305, 339, 341, 344	2, 2, 2, 1, 1
			2	400, 434, 445, 472, 483, 493	2, 1, 1, 2, 1, 1
5	327 (Bus)	352, 341, 318, 300	1	300, 305, 318, 339, 341, 344, 352	3, 2, 1, 2, 2, 1, 1
			2	400, 434, 445, 472, 483, 493	2, 1, 1, 2, 1, 1
6	530 (Elephant)	588, 586, 564, 529	1	300, 305, 318, 339, 341, 344, 352	3, 2, 1, 2, 2, 1, 1
			2	400, 434, 445, 472, 483, 493	2, 1, 1, 2, 1, 1
			3	529, 564, 586, 588	1, 1, 1, 1
7	207 (Rome)	269, 211, 231	1	300, 305, 318, 339, 341, 344, 352	3, 2, 1, 2, 2, 1, 1
			2	400, 434, 445, 472, 483, 493	2, 1, 1, 2, 1, 1
			3	529, 564, 586, 588	1, 1, 1, 1
			4	211, 231, 269	1, 1, 1
8	600 (Flower)	605, 609, 644, 623	1	300, 305, 318, 339, 341, 344, 352	3, 2, 1, 2, 2, 1, 1
			2	400, 434, 445, 472, 483, 493	2, 1, 1, 2, 1, 1
			3	529, 564, 586, 588	1, 1, 1, 1
			4	211, 231, 269	1, 1, 1
			5	605, 609, 623, 644	1, 1, 1, 1
9	605 (Flower)	615, 600, 609, 623	1	300, 305, 318, 339, 341, 344, 352	3, 2, 1, 2, 2, 1, 1
			2	400, 434, 445, 472, 483, 493	2, 1, 1, 2, 1, 1
			3	529, 564, 586, 588	1, 1, 1, 1
			4	211, 231, 269	1, 1, 1
			5	600, 605, 609, 615, 623, 644	1, 1, 2, 1, 2, 1
10	440 (Dinosaur)	436, 483, 400, 499	1	300, 305, 318, 339, 341, 344, 352	3, 2, 1, 2, 2, 1, 1
			2	400, 434, 436, 445, 472, 483, 493, 499	3, 1, 1, 1, 2, 2, 1, 1
			3	529, 564, 586, 588	1, 1, 1, 1
			4	211, 231, 269	1, 1, 1
			5	600, 605, 609, 615, 623, 644	1, 1, 2, 1, 2, 1

Table 1: The log of SBL model with different concept images query by different users

4 Experiments

1000 images from Corel dataset [8,17] with 10 concepts [People, Beach, Rome, Bus, Dinosaur, elephant, flower, house, mountain, food] are utilized for testing our framework. The image samples of dataset are shown in figure 3.

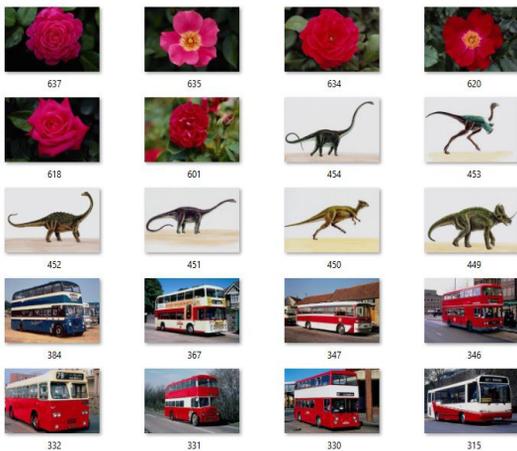


Fig. 3: Partial image samples in Coref dataset

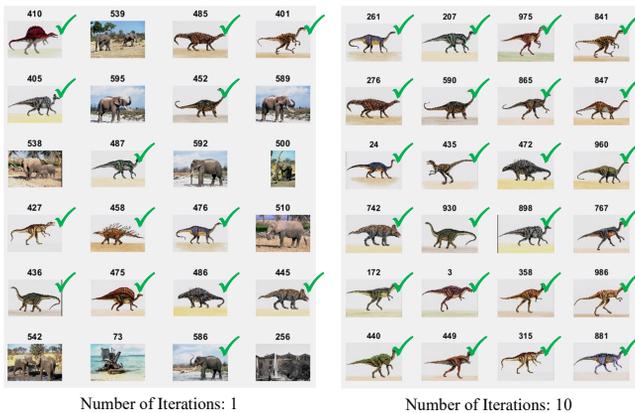


Fig. 4: Correctness comparison of 1st iteration and 10th iteration after learning with SBL model

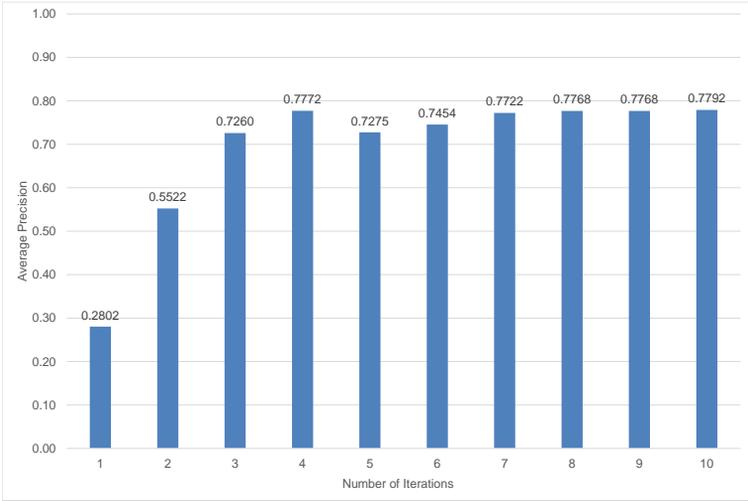


Fig. 5: Average precision of queries with same concept images in 10-times iteration.

In this paper, two experiments are conducted to prove the accuracy improvement by using the proposed framework with the SBL model. In the first experiment, the performance of expected image retrieval after learning via the SBL model is focused. 10 images with the same concept provided by users as an input for the experiment. Users select their expected image samples from preliminary query result. Based on the user behavior in the selection of preliminary query result in each iteration, the selected images are assigned label code that can help to improve the performance of expected image retrieval. In the second experiment, we simulate different users to do queries with 10 image samples in 5 concepts. The queries in each iteration whether it is the same concept or different user or not are randomly assigned. So image samples with same concept are not provided continually for simulating a real situation.

5 Evaluation

The average precision are used as evaluation metrics for 2 experiments comparing the improvement of retrieval performance. The equation of average precision is shown in equation 5.

$$\begin{aligned}
 P(g) &= \frac{\text{Relevant images in Top } g}{g} \\
 R(g) &= \frac{\text{Relevant images in Top } g}{\text{Relevant images in dataset}} \\
 AP &= \sum P(g) \Delta R(g)
 \end{aligned} \tag{5}$$

In experiment 1, the retrieval results show great improvement after 10-times iteration from Figure 4. In this experiment, we can see 58% of the images are correctly

retrieved in the top 24 final query result in the 1st iteration. The images which are correctly retrieved in the top 24 increase to 100% in the 10th iteration. From figure 5, we can see that the average precision of the final query result is increasing through the proposed SBL model in 10-time iteration. We can see that the increase in average precision becomes stable after the 4th iteration.

In experiment 2, table 1 shows the number of iterations, images with the concept provided in each query, the selection of image ID of preliminary query result in each iteration, and Similar Behavior Learning log. In column "Similar Behavior Learning Log", it records changing of the image ID with assigned label code l_p and the corresponding confidence value $c(s_p)$ in each iteration. The image ID represents as bold means the data updated in each iteration. From the table, we can see that the selected images with different concepts are annotated different label codes while the image with the same concept is annotated the same label code correctly.

6 Conclusion and Future Work

In this paper, we explored using user behavior from multiple users as additional subjective information for learning for improving the performance of image retrieval. We propose a two-step interactive image retrieval framework with SBL model that it has combined the information from the subjective and objective way. Preliminary query step is added to collect valuable information: user behaviors. Based on the user behavior collected in preliminary query stage, the proposed SBL model can identify the similarity of user behavior from previous cases instantly. It annotated new or existed label codes for the selected images as index for next iteration. The database is updated every iteration after the learning process of SBL model. It can benefit the following iterations with faster speed in generate preliminary query result for user selection. With the annotated images which are learned by SBL model, further improvements are consistently observed in the experiments. The framework is general and easy to be adapted, it can be combined with various feature extraction methods or distance measurement methods according to the purpose of the application.

Our future works as follows. There is always a challenge to understand user expectations with limited information. Users can use text, samples, sketch or hybrid methods to express their expectations of retrieval results. The way to collect expectations can be various. In the future, we will explore various descriptors to express images, such as shape, color and texture. We can assign label code to each descriptor for image annotation in detail. With the annotated label at the descriptor level, we can construct a more flexible and accurate query for fulfilling users' expectations.

References

1. Agichtein, E., Brill, E., Dumais, S.: Improving web search ranking by incorporating user behavior information. In: Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval. pp. 19–26. ACM (2006)
2. Franzoni, V., Leung, C.H., Li, Y., Mengoni, P., Milani, A.: Set similarity measures for images based on collective knowledge. In: International Conference on Computational Science and Its Applications. pp. 408–417. Springer (2015)
3. Gudivada, V.N., Raghavan, V.V.: Content based image retrieval systems. *Computer* **28**(9), 18–22 (1995)
4. Hore, S., Tyrvaainen, L., Pyykko, J., Glowacka, D.: A reinforcement learning approach to query-less image retrieval. In: International Workshop on Symbiotic Interaction. pp. 121–126. Springer (2015)
5. Huang, Y., Wu, Q., Song, C., Wang, L.: Learning semantic concepts and order for image and sentence matching. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 6163–6171 (2018)
6. Kaelbling, L.P., Littman, M.L., Moore, A.W.: Reinforcement learning: A survey. *Journal of artificial intelligence research* **4**, 237–285 (1996)
7. Khodaskar, A., Ladke, S.: Content based image retrieval with semantic features using object ontology. *International Journal of Engineering Research & Technology* (2012)
8. Li, J., Wang, J.Z.: Automatic linguistic indexing of pictures by a statistical modeling approach. *IEEE Transactions on pattern analysis and machine intelligence* **25**(9), 1075–1088 (2003)
9. Sidhu, S., Saxena, J.: Content based image retrieval a review. *International Journal Of Research In Computer Applications And Robotics* **3**(5), 84–88 (2015)
10. Singha, M., Hemachandran, K.: Content based image retrieval using color and texture. *Signal & Image Processing* **3**(1), 39 (2012)
11. Smeulders, A.W., Worring, M., Santini, S., Gupta, A., Jain, R.: Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis & Machine Intelligence* (12), 1349–1380 (2000)
12. Solli, M., Bergstrom, S.: Image retrieval and processing systems and methods (Jan 15 2019), uS Patent App. 10/180,950
13. Srisuk, S., Fooprateepsiri, R., Petrou, M., Waraklang, S., Sunat, K.: A general framework for image retrieval using reinforcement learning. *The Image and Vision Computing* pp. 26–28 (2003)
14. Sutton, R.S., Barto, A.G.: Reinforcement learning: An introduction. MIT press (2018)
15. Sutton, R.S., Barto, A.G., et al.: Introduction to reinforcement learning, vol. 135. MIT press Cambridge (1998)
16. Wang, C., Zhang, L., Zhang, H.J.: Learning to reduce the semantic gap in web image retrieval and annotation. In: Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval. pp. 355–362. ACM (2008)
17. Wang, J.Z., Li, J., Wiederhold, G.: Simplicity: Semantics-sensitive integrated matching for picture libraries. *IEEE Transactions on Pattern Analysis & Machine Intelligence* (9), 947–963 (2001)
18. Watkins, C.J., Dayan, P.: Q-learning. *Machine learning* **8**(3-4), 279–292 (1992)
19. Westerveld, T.: Image retrieval: Content versus context. In: Content-Based Multimedia Information Access-Volume 1. pp. 276–284. Le Centre de hautes études internationales d’informatique documentaire (2000)
20. Xie, X., Mao, J., de Rijke, M., Zhang, R., Zhang, M., Ma, S.: Constructing an interaction behavior model for web image search. In: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval. pp. 425–434. ACM (2018)

- 495 21. Xu, G., Li, J., Xu, C., Wang, Q.: Content-based image retrieval improved by incorporating 495
496 semantic annotation via query expansion. *International Journal of Signal Processing, Image* 496
497 *Processing and Pattern Recognition* **8**(12), 59–66 (2015) 497
498 22. Yin, P.Y., Bhanu, B., Chang, K.C., Dong, A.: Integrating relevance feedback techniques for 498
499 image retrieval using reinforcement learning. *IEEE transactions on pattern analysis and* 499
500 *machine intelligence* **27**(10), 1536–1551 (2005) 500
501 23. Yue, J., Li, Z., Liu, L., Fu, Z.: Content-based image retrieval using color and texture fused 501
502 features. *Mathematical and Computer Modelling* **54**(3-4), 1121–1127 (2011) 502
503 503
504 504
505 505
506 506
507 507
508 508
509 509
510 510
511 511
512 512
513 513
514 514
515 515
516 516
517 517
518 518
519 519
520 520
521 521
522 522
523 523
524 524
525 525
526 526
527 527
528 528
529 529
530 530
531 531
532 532
533 533
534 534
535 535
536 536
537 537
538 538
539 539