IMAGE RETRIEVAL BASED ON LRGA ALGORITHM AND RELEVANCE FEEDBACK FOR INSECT IDENTIFICATION

Susumu Genma[†]*, Takahiro Ogawa*[‡] *and Miki Haseyama*[‡]

[†]School of Engineering, Hokkaido University
N-13, W-8, Kita-ku, Sapporo, Hokkaido, 060-8628, Japan
[‡]Graduate School of Information Science and Technology, Hokkaido University
N-14, W-9, Kita-ku, Sapporo, Hokkaido, 060-0814, Japan
E-mail: {genma, ogawa}@lmd.ist.hokudai.ac.jp, miki@ist.hokudai.ac.jp

ABSTRACT

This paper presents an image retrieval method based on local regression and global alignment (LRGA) algorithm and relevance feedback for insect identification. Based on LRGA algorithm, the proposed method enables estimation of ranking scores for image retrieval in such a way that the neighborhood structure of the database can be optimally preserved. This is the biggest contribution of this paper. Then our method measures relevance between the query image and all the images in the database and realizes retrieval of images based on the measured relevance. Furthermore, if positively labeled images obtained by a user are available, they are used as the query relevance information for the relevance feedback to improve the retrieval results. Experimental results show the effectiveness of our method.

Index Terms— Image retrieval, LRGA, local structure, relevance feedback, insect identification.

1. INTRODUCTION

There are many kinds of insects on the earth. Identifying the type of these insects will lead to developments in various fields such as biology and agriculture. For instance, insect identification contributes to enriching the collection of specimens [1, 2]. In addition, in the field of agriculture, there is a problem of damage to crops by pests [2, 3]. In order to solve this problem, pest control is important. However, when farmers cannot know whether the insects are pests or not, it is difficult to take appropriate measures. Thus, it is useful to support the identification of insects for selecting appropriate control methods.

From the above background, various methods to support identification based on image processing technology have been proposed [4–8]. Many of these methods are based on classification or retrieval. In classification-based methods [5, 7, 8], a large amount of labeled data is required for accurate classification. However, since it is difficult to prepare enough number of labeled data for each species, these methods are not appropriate for insect identification.

On the other hand, in the retrieval-based methods [9–11], given a query image by a user, images in the database are ranked according to their relevance to the query image, and the top-ranked ones are provided. By showing images of similar insects, i.e., several candidates, the user can identify his/her interested insect from the retrieval results. Most of retrieval-based methods are distance-based ranking methods, which usually perform retrieval according to the Euclidean distance between the query image and images in the database, either in the original feature space or in a lower dimensional space of visual features [12, 13]. In addition, relevance feedback (RF) has been proven helpful for retrieval [14–20]. RF is to ask the user to provide feedback regarding the relevance of the current retrieval results and perform learning from the feedback to achieve an improved performance in the next round. We have also proposed a retrieval method that can performs image retrieval based on RF [21]. In this method, the distance between images based on visual features is integrated with the feedback to achieve highly accurate identification support.

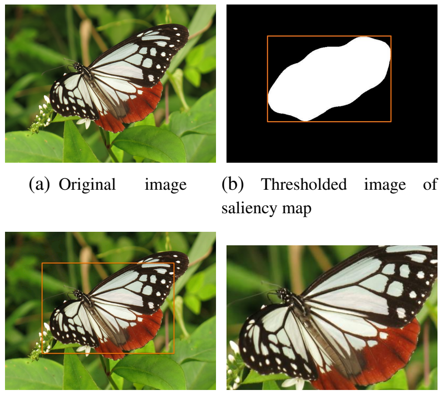
It should be noted that distance-based ranking methods and their variants including RF only focus on the pairwise similarity between the query image and images in the database, but the similarities between images in the database are not considered, which usually degrades their effectiveness [15]. Therefore, it is essential to consider the database distribution[1]. In contrast, a transductive method for ranking was proposed in [22]. Transductive methods rank the images with respect to the intrinsic database distribution. Transductive ranking algorithms are intrinsically different from the distance-based ranking methods because they exploit the database distribution for ranking rather than only considering the pairwise distances between the query image and images in the database. In recent years, transductive ranking algorithms have been used in various applications such as image retrieval [15], shape retrieval [23] and cross-media retrieval [20]. These results have demonstrated that it is beneficial to utilize the database distribution for ranking.

In this paper, we present a new insect image retrieval method considering database distribution and introducing RF. First, we retrieve images using a new transductive ranking algorithm based on local regression and global alignment (LRGA) [24]. For each data point[2], LRGA employs a local linear regression model to predict the ranking scores of its neighboring points. In order to assign an optimal ranking score to each data point, LRGA considers a unified objective function to globally align local linear regression models from all the data points. Second, we also introduce RF, which reflects the user's evaluation on the retrieval results. RF considers relevant information provided by the current user and can improve the accuracy of the current retrieval. Therefore, by the combination use of

This work was partly supported by JSPS KAKENHI Grant Numbers JP24120002, JP17H01744.

[1]In this paper, the positional relationships of each image in the database in the feature space is called a "database distribution".

[2]In this paper, the point at which the visual feature of each image is located in the feature space is called a "data point".



(a) Original image (b) Thresholded image of saliency map

(c) Image surrounding insects in a rectangular region (d) Extracted image of insect region

Fig. 1. Example of insect region extraction based on visual saliency.

the ranking algorithm and RF described above, our method realizes successful insect image retrieval.

2. OUR IMAGE RETRIEVAL METHOD

The image retrieval method based on LRGA ranking algorithm and RF is presented in this section. First, the proposed method focuses on characteristics of insect images and calculates their visual features based on visual saliency. Next, the obtained visual features are then used to perform image retrieval with the LRGA ranking algorithm, where respective images in the database are ranked by database distribution. Then we reperform image retrieval with the LRGA ranking algorithm using RF.

This section is organized as follows. In **2.1**, we explain the calculation of visual features from insect images based on visual saliency. In **2.2**, we describe the image retrieval based on LRGA ranking algorithm and RF, which is the main contribution of this paper.

2.1. Visual Feature Calculation based on Saliency

In this subsection, we describe how to calculate visual features from insect images. First, in insect images, there are backgrounds unnecessary for representing visual characteristics of the insects. In order to reduce the effect of the backgrounds, we extract the insect regions from the insect images by a simple but effective scheme. Since the target insect is generally located at the center of the image, a saliency map weighted at the center of the original image (Fig.1 (a)) is calculated based on our previously reported method [25]. After thresholding (Fig. 1 (b)) is applied to the calculated saliency map, the minimum rectangle region (Fig.1 (c)) surrounding the obtained region is extracted as the insect region (Fig.1 (d)). This makes it possible to suppress the effect of background regions unnecessary for the feature calculation of insects. Thus, the visual feature calculation can be performed from the extracted insect region. Note that the detaile of visual features are shown in section 3.

2.2. Image Retrieval based on LRGA Ranking Algorithm and RF

In this subsection, we explain the retrieval method based on LRGA and RF. LRGA can update the rank of similar image to upper rank based on evaluation presented by user. Therefore, we can effectively combine LRGA and RF. First, a set of calculated visual feature vectors is defined as $\mathcal{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_N\}$ (N being the total number of images in the database). The ranking score of each image is defined as $\boldsymbol{f} = [f_1, f_2, \dots, f_N]^T \in \mathbb{R}^N$. Specifically, we consider the following optimization problem:

$$\min_{\boldsymbol{f}} \sum_{i=1}^{N} (f_i - y_i)^2, \tag{1}$$

where y_i is defined as follows:

$$y_i = \begin{cases} 1 & \text{if } \boldsymbol{x}_i \text{ is the feature vector of the query example} \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Equation (1) is an optimization problem in order to make the query image consistently top ranked. Since the query image contains the user's retrieval intention, the ranking score is always considered to be high. We introduce a matrix $\boldsymbol{V} = \mathrm{diag}(V_{ii})$ to make the ranking score for the query image high. Using $\boldsymbol{V}$, Eq. (1) is rewritten as follows:

$$\min_{\boldsymbol{f}} \sum_{i=1}^{N} V_{ii}(f_i - y_i)^2 = \min_{\boldsymbol{f}} (\boldsymbol{f} - \boldsymbol{y})^T \boldsymbol{V} (\boldsymbol{f} - \boldsymbol{y}), \tag{3}$$

where $\boldsymbol{y} = [y_1, y_2, \dots, y_N]^T \in \mathbb{R}^N$. Here V_{ii} is a very large value[3] when $\boldsymbol{x}_i$ is the feature vector of the query image, and $V_{ii} = 1$ otherwise in order to make ranking score for the query image high.

Next, we consider another optimization problem. In many real applications, the local manifold structure is more important than the global structure [26]. Meanwhile, it has been reported that the local learning algorithms often outperform global learning algorithms [27, 28]. To make use of the database distribution, we employ the local structure of $\boldsymbol{x}_i$. First, let us denote $\mathcal{L}_k(\boldsymbol{x}_i) = \{\boldsymbol{x}_i, \boldsymbol{x}_{i_1}, \boldsymbol{x}_{i_2}, \dots, \boldsymbol{x}_{i_k}\}$ as the set of k-nearest neighbors of $\boldsymbol{x}_i$ and $\boldsymbol{x}_i$ itself. Then we consider the following linear regression model as the prediction of the ranking score for $\mathcal{L}_k(\boldsymbol{x}_i)$:

$$h_i(\boldsymbol{x}_j) = \boldsymbol{w}_i^T \boldsymbol{x}_j + b_i, \ \boldsymbol{x}_j \in \mathcal{L}_k(\boldsymbol{x}_i), \tag{4}$$

where $\boldsymbol{w}_i \in \mathbb{R}^d$ (d being the dimension of visual features) represents the projection vector and $b_i \in \mathbb{R}$ represents the bias term. Using this linear model, we consider the following minimization problem:

$$\min_{\boldsymbol{f}_{(i)}, b_i, \boldsymbol{w}_i} \sum_{\boldsymbol{x}_j \in \mathcal{L}_k(\boldsymbol{x}_i)} (\boldsymbol{w}_i^T \boldsymbol{x}_j + b_i - f_j)^2 + \lambda \boldsymbol{w}_i^T \boldsymbol{w}_i, \tag{5}$$

where $\boldsymbol{f}_{(i)} = [f_i, f_{i_1}, f_{i_2}, \dots, f_{i_k}]^T$ is a vector including the ranking scores of all data points in $\mathcal{L}_k(\boldsymbol{x}_i)$, and the second term is a regularization term for avoiding overfitting. Equation (5) can be rewritten as follows:

$$\min_{\boldsymbol{f}_{(i)}, b_i, \boldsymbol{w}_i} ||\boldsymbol{X}_i^T \boldsymbol{w}_i + b_i \mathbf{1}_{k+1} - \boldsymbol{f}_{(i)}||^2 + \lambda \boldsymbol{w}_i^T \boldsymbol{w}_i, \tag{6}$$

where $\boldsymbol{X}_i = [\boldsymbol{x}_i, \boldsymbol{x}_{i_1}, \boldsymbol{x}_{i_2}, \dots, \boldsymbol{x}_{i_k}]$ is a matrix including all the data points in $\mathcal{L}_k(\boldsymbol{x}_i)$, $\mathbf{1}_{k+1} \in \mathbb{R}^{k+1}$ is a column vector whose elements are all ones. Then, by setting the derivatives of Eq. (6) to be zero with respect to $\boldsymbol{w}_i$ and b_i, we can obtain the solution of $\boldsymbol{w}_i$

[3] It is theoretically ∞. In this work, we set it to 1.0×10^4.

and b_i as follows:

$$\boldsymbol{w}_i = (\boldsymbol{X}_i\boldsymbol{U}\boldsymbol{X}_i{}^T + \lambda\boldsymbol{I}_{k+1})^{-1}\boldsymbol{X}_i\boldsymbol{U}\boldsymbol{f}_{(i)}, \tag{7}$$

$$b_i = \frac{1}{k+1}(\mathbf{1}_{k+1}^T\boldsymbol{f}_{(i)} - \mathbf{1}_{k+1}^T\boldsymbol{X}_i^T\boldsymbol{w}_i), \tag{8}$$

where $\boldsymbol{U} = \boldsymbol{I}_{k+1} - \frac{1}{k+1}\mathbf{1}_{k+1}\mathbf{1}_{k+1}^T$ is a centering matrix, and $\boldsymbol{I}_{k+1} \in \mathbb{R}^{(k+1)\times(k+1)}$ is the identity matrix. By substituting the obtained $\boldsymbol{w}_i$ and b_i into Eq. (6), the optimization problem is rewritten as follows:

$$\begin{aligned}\min_{\boldsymbol{f}_{(i)}} \Big[&||\boldsymbol{U}\boldsymbol{X}_i^T(\boldsymbol{X}_i\boldsymbol{U}\boldsymbol{X}_i{}^T + \lambda\boldsymbol{I}_{k+1})^{-1}\boldsymbol{X}_i\boldsymbol{U}\boldsymbol{f}_{(i)} - \boldsymbol{U}\boldsymbol{f}_{(i)}||^2 \\ &+ \lambda\boldsymbol{f}_{(i)}^T\boldsymbol{U}\boldsymbol{X}_i^T(\boldsymbol{X}_i\boldsymbol{U}\boldsymbol{X}_i{}^T + \lambda\boldsymbol{I}_{k+1})^{-2}\boldsymbol{X}_i\boldsymbol{U}\boldsymbol{f}_{(i)}\Big].\end{aligned} \tag{9}$$

Furthermore, based on [24], Eq. (9) can be rewritten as follows:

$$\min_{\boldsymbol{f}_{(i)}} \boldsymbol{f}_{(i)}^T\boldsymbol{M}_i\boldsymbol{f}_{(i)}, \tag{10}$$

where $\boldsymbol{M}_i \in \mathbb{R}^{(k+1)\times(k+1)}$ is defined as follows:

$$\boldsymbol{M}_i = \boldsymbol{U} - \boldsymbol{U}\boldsymbol{X}_i^T(\boldsymbol{X}_i\boldsymbol{U}\boldsymbol{X}_i{}^T + \lambda\boldsymbol{I}_{k+1})^{-1}\boldsymbol{X}_i\boldsymbol{U}. \tag{11}$$

Next, by using a matrix $\boldsymbol{O}_i \in \mathbb{R}^{N\times(k+1)}$ defined below, $\boldsymbol{f}_{(i)}^T$ can be expressed as $\boldsymbol{f}_{(i)}^T = \boldsymbol{f}^T\boldsymbol{O}_i$:

$$(\boldsymbol{O}_i)_{pq} = \begin{cases} 1 & p = (\boldsymbol{v}_i)_q \\ 0 & \text{otherwise,} \end{cases} \tag{12}$$

where $\boldsymbol{v}_i = [i, i_1, i_2, \cdots, i_k]$ is a vector containing the indices of images in $\mathcal{L}_k(\boldsymbol{x}_i)$, and $(\boldsymbol{v}_i)_q$ is the q-th element of $\boldsymbol{v}_i$. As a result, Eq. (10) can be rewritten as follows:

$$\min_{\boldsymbol{f}_{(i)}} \boldsymbol{f}^T\boldsymbol{O}_i\boldsymbol{M}_i\boldsymbol{O}_i^T\boldsymbol{f}. \tag{13}$$

By solving the minimization problem considered in this set $\mathcal{L}_k(\boldsymbol{x}_i)$ for all images, it is possible to assign an optimal ranking score to each data point. The minimization problem can be written as follows:

$$\min_{\{\boldsymbol{f}_{(i)}\}_{i=1}^N} \sum_{i=1}^{N} \boldsymbol{f}^T\boldsymbol{O}_i\boldsymbol{M}_i\boldsymbol{O}_i^T\boldsymbol{f} = \min_{\boldsymbol{f}} \boldsymbol{f}^T\boldsymbol{M}\boldsymbol{f}, \tag{14}$$

where $\boldsymbol{M}$ is defined as follows:

$$\boldsymbol{M} = [\boldsymbol{O}_1, \boldsymbol{O}_2, \cdots, \boldsymbol{O}_N] \begin{bmatrix} \boldsymbol{M}_1 & & \\ & \ddots & \\ & & \boldsymbol{M}_N \end{bmatrix} [\boldsymbol{O}_1, \boldsymbol{O}_2, \cdots, \boldsymbol{O}_N]^T. \tag{15}$$

By combining Eq. (1) and Eq. (14), we can obtain the following problem:

$$\min_{\boldsymbol{f}} (\boldsymbol{f} - \boldsymbol{y})^T\boldsymbol{V}(\boldsymbol{f} - \boldsymbol{y}) + \boldsymbol{f}^T\boldsymbol{M}\boldsymbol{f}. \tag{16}$$

The optimal solution to Eq. (16) can be obtained by solving the following linear equation for $\hat{\boldsymbol{f}}$:

$$(\boldsymbol{M} + \boldsymbol{V})\hat{\boldsymbol{f}} = \boldsymbol{V}\boldsymbol{y}. \tag{17}$$

The database is sorted in descending order according to the obtained ranking score $\hat{\boldsymbol{f}} = [\hat{f}_1, \hat{f}_2, \cdots, \hat{f}_N]^T$ and the top-m ranked images are returned as the retrieval result to the user.

Next, after the user evaluates the results, our method performs RF. For each feature vector $\boldsymbol{x}_i$, we update y_i as follows.

$$y_i^* = \begin{cases} 1 & \text{if } \boldsymbol{x}_i \text{ is the feature vector of the positive image} \\ & \text{marked by the user} \\ 0 & \text{otherwise.} \end{cases} \tag{18}$$

where the positive image is an image similar to the query judged by the user. Then we solve Eq. (16) and (17) by using the updated $\boldsymbol{y}^* = [y_1^*, y_2^*, \cdots, y_N^*]^T$ again to rerank the retrieved images for a new round of retrieval.

$$\min_{\boldsymbol{f}^*} (\boldsymbol{f}^* - \boldsymbol{y}^*)^T\boldsymbol{V}^*(\boldsymbol{f}^* - \boldsymbol{y}^*) + \boldsymbol{f}^{*T}\boldsymbol{M}\boldsymbol{f}^*, \tag{19}$$

$$(\boldsymbol{M} + \boldsymbol{V}^*)\hat{\boldsymbol{f}}^* = \boldsymbol{V}^*\boldsymbol{y}^*. \tag{20}$$

Finally we obtain reranking scores $\hat{\boldsymbol{f}}^* = [\hat{f}_1^*, \hat{f}_2^*, \cdots, \hat{f}_N^*]^T$. We sort database in descending order according to the obtained reranking score $\hat{\boldsymbol{f}}^*$ and the top-m ranked images are returned as the retrieval result to the user.

3. EXPERIMENTS

In this section, we conduct experiments to confirm the effectiveness of our proposed method. In **3.1**, we shown the database and visual features used in the experiments. In **3.2**, we shown results of the experiments.

3.1. Database and Visual Features

In this section, we conduct experiments to confirm the effectiveness of our proposed method. In this experiment, a total of 10482 insect images were used as a database. Insect images in the database are divided into two subsets. The first subset is a sum of 100 labeled insect images composed of 20 kinds of insects, and the second subset is a sum of 10382 unlabeled insect images. We prepared insect images of the first subset from national museum of nature and science[4], and the second subset was obtained by collecting up to 500 insect images from Flickr[5] for each order such as 'lepidoptera', 'coleoptera', belonging to the 'insecta' [29].

For the query image, ten kinds of species were randomly selected from the first subset. The number of correct images among the retrieval target images is four. We calculated 4096 dimensional features from the fc7 layer of AlexNet [30] based on [31] for the visual features, and the number of round of retrieval for RF is one.

3.2. Experimental Results

For the evaluation, we utilized mean average precision (MAP) defined as follows:

where

$$\text{MAP@}n = \frac{1}{M}\sum_{i=1}^{M} \text{AP}(i)\text{@}n, \tag{21}$$

$$\text{AP}(i)\text{@}n = \sum_{t=1}^{n} \frac{\alpha_t \text{Precision@}t}{N_p}, \tag{22}$$

[4]http://www.kahaku.go.jp/
[5]https://www.flickr.com/

Fig. 2. Query example.

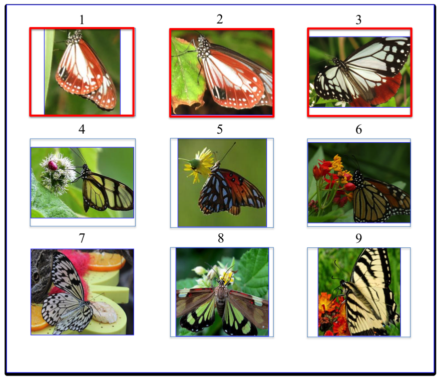


Fig. 3. Retrieval results obtained by the proposed method. (Correct images are indicated by thicker borders around the retrieved images.)

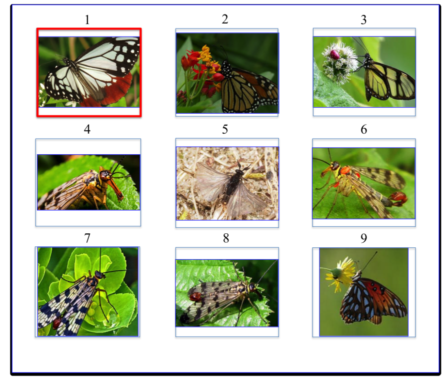


Fig. 4. Retrieval results obtained by the comparative method 1. (Correct images are indicated by thicker borders around the retrieved images.)

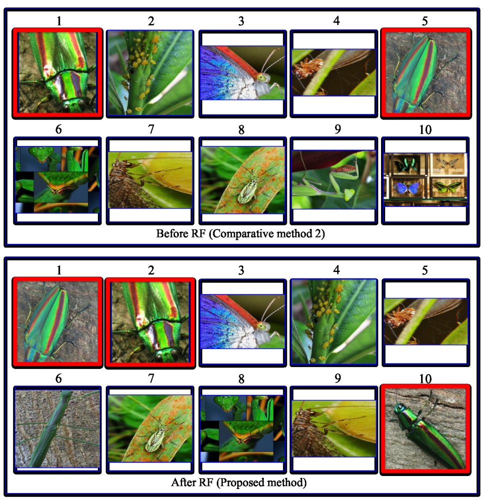


Fig. 5. Top 10 retrieved images before and after RF.

$$\text{Precision@}t = \frac{N_p}{t}, \tag{23}$$

and t is the number of retrieval results, N_p is the sum of correct images included in t, and $\alpha_t = 1$ when t-th image is a correct image and $\alpha_t = 0$ otherwise, then M is the sum of the queries, n is the sum of retrieval results.

In order to verify the effectiveness of the proposed method, we compared its performance with those of the following three methods.

- Comparative method 1 : A method which obtains the retrieval results by only using the distance-based ranking method and RF [21].
- Comparative method 2 : A method which obtains the retrieval results by only using the LRGA ranking algorithm without RF.
- Comparative method 3 : A method which obtains the retrieval results by only using the distance-based ranking method without RF.

Figure 2 shows a query example. The experimental results are shown in Figs. 3 and 4. There results show top 9 retrieved images by the proposed method and comparative method 1 from the query image shown in Fig. 2 as a query example. Correct images are indicated by thicker borders around the retrieved images. The results show that the proposed method provides more correct images than comparative method 1. Next, Fig. 5 shows top 10 retrieved images before and after RF using another query example image. Figure 5 shows that the proposed method provides more correct images than comparative method 2. Furthermore, Fig. 6 shows that MAP@n($n = 1, 2, \cdots, 100$) of the proposed method is higher than those of other comparative methods. Therefore, from these results, we can confirm that the novelties of the proposed method can accurately provide the insect retrieval results.

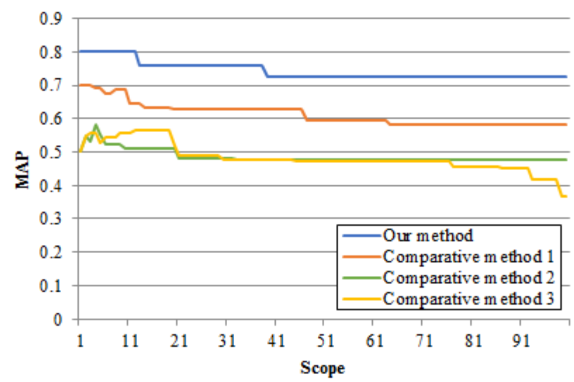


Fig. 6. MAP@n obtained by our method and three comparative methods.

4. CONCLUSIONS

In this paper, we have presented a method for insect image retrieval using LRGA ranking algorithm and RF. In the feature space on the database including the query, the proposed method estimates ranking scores considering neighborhood structure of the database. Next, by using RF, we reperform the LRGA ranking algorithm to rerank the retrieval results for a new round of retrieval. The experimental results show that our method realizes successful retrieval results.

5. REFERENCES

[1] J. L. Edwards, M. A. Lane, and E. S. Nielsen, "lnteroperability of Biodiversity Databases : Biodiversity lnformation on Every Desktop," *Science*, vol. 289, no. September, pp. 2312–2314, 2000.

[2] P. J. D. Weeks, M. A. O'Neill, K. J. Gaston *et al.*, "Automating insect identification: exploring the limitations of a prototype system," *Journal of Applied Entomology*, vol. 123, no. 1, pp. 1–8, 1999.

[3] J. D. Lattin and L. Knutson, "Taxonomic information and services on arthropods of importance to human welfare in Central and South America," *Food and Agriculture Organization (FAO) of the United Nations, Plant Protection Bulletin*, vol. 30, no. 01, pp. 9–12, 1982.

[4] N. Larios, H. Deng, W. Zhang *et al.*, "Automated insect identification through concatenated histograms of local appearance features: Feature vector generation and region detection for deformable objects," *Machine Vision and Applications*, vol. 19, no. 2, pp. 105–123, 2008.

[5] T. Arbuckle, S. Schröder, V. Steinhage *et al.*, "Biodiversity informatics in action: identification and monitoring of bee species using ABIS," *EnviroInfo*, vol. 1, pp. 425–430, 2001.

[6] A. T. Watson, M. A. O'Neill, and I. J. Kitching, "Automated identification of live moths (Macrolepidoptera) using Digital Automated Identification SYstem (DAISY)," *Systematics and Biodiversity*, vol. 1, no. 3, pp. 287–300, 2003.

[7] M. Do, J. Harp, and K. Norris, "A test of a pattern recognition system for identification of spiders," *Bulletin of Entomological Research*, vol. 89, no. 03, pp. 217–224, 1999.

[8] C. Wen and D. Guyer, "Image-based orchard insect automated identification and classification method," *Computers and Electronics in Agriculture*, vol. 89, pp. 110–115, 2012.

[9] a.W.M. Smeulders, M. Worring, S. Santini *et al.*, "Content-based image retrieval at the end of the early years," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1349–1380, 2000.

[10] a. Pentland, R. W. Picard, and S. Sclaroff, "Photobook: Content-based manipulation of image databases," *M.I.T. Media Laboratory Percaptual Computing Technical Report*, vol. 255, no. 3, pp. 233–254, 1993.

[11] T. León, P. Zuccarello, G. Ayala *et al.*, "Applying logistic regression to relevance feedback in image retrieval systems," *Pattern Recognition*, vol. 40, no. 10, pp. 2621–2632, 2007.

[12] Jing Huang, S. Kumar, M. Mitra *et al.*, "Image indexing using color correlograms," *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 191, no. 3-4, pp. 762–768, 1994.

[13] G. Chechik, U. Sharma V. anf Shalit, and S. Bengio, "Large Scale Online Learning of Image Similarity Through Ranking," *J. Mach. Learn. Res.*, vol. 11, pp. 1109–1135, 2010.

[14] D. Cai, X. He, and J. Han, "Semi-supervised discriminant analysis," *Proceedings of the IEEE International Conference on Computer Vision*, 2007.

[15] J. He, M. Li, H.-J. Zhang *et al.*, "Manifold-ranking based image retrieval," *ACM Multimedia*, pp. 9–16, 2004.

[16] X. He, W.-Y. Ma, and H.-J. Zhang, "Learning an image manifold for retrieval," *Proceedings of the 12th annual ACM international conference on Multimedia - MULTIMEDIA '04*, p. 17, 2004.

[17] J. Yu and Q. Tian, "Learning image manifolds by semantic subspace projection," *Proceedings of the 14th annual ACM international conference on Multimedia - MULTIMEDIA '06*, p. 297, 2006.

[18] M. S. Lew, N. Sebe, C. Djeraba *et al.*, "Content-Based Multimedia Information Retrieval: State of the Art and Challenges," *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 2, no. 1, pp. 1–19, 2006.

[19] Y. Rui and T. Huang, "Optimizing learning in image retrieval," *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, vol. 1, pp. 236–243, 2000.

[20] Y. T. Zhuang, Y. Yang, and F. Wu, "Mining semantic correlation of heterogeneous multimedia data for cross-media retrieval," *IEEE Transactions on Multimedia*, vol. 10, no. 2, pp. 221–229, 2008.

[21] S. Genma, T. Ogawa, and M. Haseyama, "Image Retrieval for Identification of Insects Based on Saliency Map and Distance Metric Learning," *2016 IEEE 5th Global Conference on Consumer Electronics (GCCE)*, pp. 1–2, 2016.

[22] D. Zhou, J. Weston, A. Gretton *et al.*, "Ranking on Data Manifolds," *Advances in Neural Information Processing Systems*, vol. 16, no. c, pp. 177–186, 2003.

[23] X. Bai, X. Yang, L. J. Latecki *et al.*, "Learning context-sensitive shape similarity by graph transduction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 5, pp. 861–874, 2010.

[24] Y. Yang, D. Xu, F. Nie *et al.*, "Ranking with local regression and global alignment for cross media retrieval," *Proceedings of the seventeen ACM international conference on Multimedia - MM '09*, no. September, p. 175, 2009.

[25] S. Takahashi, T. Ogawa, and M. Haseyama, "Feature Extraction for Road Structure Assessment based on Visual Saliency of Distress Image," *International Workshop on Advanced Image Technology*, pp. P. 2—-B. 6, 2016.

[26] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding." *Science (New York, N.Y.)*, vol. 290, no. 5500, pp. 2323–2326, 2000.

[27] L. Bottou and V. Vapnik, "Local Learning Algorithms," *Neural Computation*, vol. 4, no. 6, pp. 888–900, 1992.

[28] M. Wu and B. Schölkopf, "Transductive classification via local learning regularization," *Aistats*, pp. 628–635, 2007.

[29] M. Ohara, "Class Insecta," *Union of Japanese Societies for Systematic Biology, Japanese Biota Species Number Survey*, 2003.

[30] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Advances In Neural Information Processing Systems*, pp. 1–9, 2012.

[31] J. Donahue, Y. Jia, O. Vinyals *et al.*, "DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition," *Icml*, vol. 32, pp. 647–655, 2014.