Content-Based Representation For Moth Recognition And Retrieval: A Review

[1]Nur Fatin Mohd Hamran, [2]Mas Rina Mustaffa, [3]Shyamala Doraisamy

[1,2,3]Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor Malaysia
*Corresponding author: nurhamcd@gmail.com

Abstract:

There are numerous of moths on the earth and their presence is useful to our life especially as life indicators. Most of the entomologists have problems to recognize moth's species because each of them has their own color, texture and shape. The varieties of color, texture and shape of moths has increased the researcher's attention to improve the method in recognizing the moth's species. This study investigates the effectiveness of Bag of Visual Words (BOVW) representation for recognizing the moth's species. BOVW is a simple and effective representation. This representation broadly used especially in computer vision and object recognition. Local descriptors, clustering approaches, and word-image histograms in regards to BOVW for image classification and retrieval are studied. There is a contradiction between BOVW models about spatial information. The extension of BOVW to consider the spatial information is believed able to contribute to a more effective representation for moth recognition and retrieval.

Keywords: Bag of Visual Words, moth, recognition, spatial information, retrieval

1. Introduction

Due to the increasing number of images from time to time makes the image retrieval process become more challenging. Therefore, it is important to systematically organize the collection of images in the database (Ghosh, 2013; Osman & Mustaffa, 2016). The retrieval of images in the database can be through query; text-based and content-based format. Text-based image retrieval has been introduced in 1970's. The development of the text-based image retrieval is basically the idea of storing keywords or descriptions of image content, which are either manually created and entered by human annotators or automated by computer algorithms (Sheikh, Lye, Mansor, & Fauzi, 2011; Osman & Mustaffa, 2016). Although text-based image retrieval offers the best idea but it has limitation in terms of different perceptions between human for the very same image. This will lead to a mismatch with the image annotations, hence affecting the accuracy (Datta et al., 2008; Osman & Mustaffa, 2016). Therefore, to overcome the limitation, the content based image retrieval (CBIR) has been developed in 1990's. CBIR is the process of image retrieval based on low level features of the given query image. This method represents image based on its content instead of using keywords, tags, or descriptions which is very difficult to explain an image. There are three levels of features in CBIR which are low level, middle level and high level features. CBIR able to extracts low level features such as color, texture and shape efficiently. By using the Bag of Visual Words (BOVW), CBIR can relates low level features to high level features. BOVW representation is one of the middle level features. This representation can extract compact fixed-length descriptor from the images. Therefore, this model is useful especially for moth recognition because it allows a better deal with image size changes due to the camera distances

A. Bag of Visual Words (BOVW)

BOVW is a histogram of local features. For tackling the problem of video retrieval, Sivic and Zisserman (2003) has proposed the bag of visual words. Csurka et al. (2004) also presented this model and this model is widely used in object recognition and image classification. BOVW is inspired by bag of words (BOW). The BOVW model is used because it is simple and very effective. The computer vision task performances have been widely acknowledged over the years, for example in medical image classification, object recognition, image retrieval, animal-activity recognition and animal recognition. Several ways have been introduced in the previous works to implement the BOVW model, but generally there are three steps involved in BOVW; (i) the set of image patches of training dataset are represented by local feature descriptor or feature vector, for example SIFT, (ii) the set of local feature descriptor are clustered; the set of cluster centers are called visual vocabulary or codebook while each of the cluster centers is called visual word or code word, and (iii) replaced the visual regions of each images with the nearest visual word (assignment step), the attribute vector is built to represent the images in histogram. This histogram representation can be used for image classification, retrieval or recognition tasks.

B. Moth

Moths are the group of insects that belong to the order Lepidoptera, meaning 'scaly-winged'. Both butterflies and moths share the same form of insects group. There are more species of moths with approximately 160,000 species discovered on earth compared to 17,500 species of butterflies. Moth gives many

benefits and is important in our life especially in education. For the educational aspect, moths have been used to observe the impact of habitat loss and also climate change. Besides, insects' migration can be known by looking at the intricate wing patterns of the moths. Sometimes, the present of moths give us the clues to change our environment by monitoring the moth's numbers and ranges, for example the air pollution, pesticides and also climate change. Although most of the commonly seen moths have dull colors and lack of pattern on the wings, but there are a lot of moth's species with unique and varieties of colors, pattern and also shape. Therefore, due to the growth in the computer vision and pattern recognition, it gives a better deal for identifying and recognizing the insect's species.

2. Literature Review

Few studies have been done based on BOVW model. Firstly, Jiang, Yang, Ngo and Hauptmann (2010) conduct the comprehensive study on the Bag of Visual Words representation. Since two key points may be allocated into the same cluster even if they are different, the discriminative power of size of vocabulary is reduced. Although BOVW have limitation regarding the size of vocabulary, but in term of semantic and form, BOVW model is analogous to the bag of words. Besides, another work has been done by Zhang et al. (2016). To measure the discriminative power of a code word in the dictionary, the authors has proposed the BOVW-based medical image retrieval method. However, similar to the study that has been done by Jiang et al. (2010), the discriminative power of visual words and the representation to identify the similar images is decreased due to the noise and redundancy of visual dictionary used for building the BOVW model. Tu et al. (2016) come out with the idea to propose the automatic detection of damaged rooftops areas based on BOVW model by using super pixel method for image segmentation. This is because BOVW is an effective representation for areas detection especially from the high resolution aerial image.

Several works have been done based on the local descriptor and BOVW technique (Pires et al., 2016; Wen, Guyer & Li, 2009). Typically, local descriptor is distinctive, robust to occlusion and do not require segmentation.

Most of BOVW models focus on building, scene and face, not the animal or insects recognition. Because of that, Mansourian et al. (2015) proposed the animal recognition based on BOVW representation because BOVW model is invariant to scale and orientation and it will gives the excellent performance in animal recognition. In a similar spirit, Wen et al. (2013) proposed the BOVW approach to recognize cow's basic sow activities and presented the modified spatial temporal local binary pattern (MLBP-TOP) in the study.

For the insect recognition, Feng, Bhanu & Herathy, (2016) and Wen, Guyer & Li, (2009) proposed different methods to identify insects. Feng et al. (2016) introduced an image descriptor based on Semantically Related Visual (SRV) attributes to identify the pattern on the moth's wing while Wen et al. (2009) proposed a local feature based method to identify insect in orchard and automatically classify the harmful insect.

However, the previous works show that BOVW model does not consider the spatial information because histogram of visual words representation neglects the spatial information of the visual words (Franco et al., 2017; Lopez-Monroy et al., 2016; Anwar, Zambanini & Kampel, 2015; & Farhangi, Soryani, Mohsen, & Fathy, 2013). Spatial information is useful in several computer vision tasks. Khan et al.(2015) and Wang et al. (2016) have proven that by adding the spatial information can improve the visual words in the BOVW representation. However, spatial information is not very effective when using the large dictionaries compared to the small dictionaries (Khan et al., 2015). Wang et al. (2016) introduced the novel method in image classification by combined the salient region with spatial geometry structure for

visual words enhancement. In a similar spirit, Ali et al. (2016), come out with the ideas to construct the histogram of triangular regions by adding the spatial information to the bag of features representation. The proposed method gives the better result in image retrieval by reducing the semantic gap.

3. Discussion

Based on the previous works that have been done, BOVW model is the best representation for moth recognition due to its characteristic which is invariant to scale and orientation. BOVW is not affected by the object orientation and position at all. However, spatial information is one of the main issues in BOVW representation. In object recognition, spatial information can improve and give the outstanding performance of recognition. Therefore, the extension of BOVW is believed will be important to provide better representation for moth species which will lead to a more effective recognition and retrieval.

4. Conclusion

Computer vision and pattern recognition are becoming more popular among the researchers over the years. Although recognition technique attracts a lot of attention but there are still not many works have been done for moth recognition. Several works have been done to the recognition of insects for example butterflies, grasshoppers, etc. but there are still a lot of rooms for improvement. One of the improvements that can be done is by using the Bag of Visual Words (BOVW) representation which is the popular model for image recognition. BOVW is one of the successful representations in classifying images according to the objects they contain. However, this model has no enough coverage on accounting for spatial information. Further efforts should be carried out to improve BOVW so that spatial information could be well presented, hence leading to a more effective representation for moth recognition and retrieval.

Acknowledgements

References

[1] Ali, N., Bajwa, K. B., Sablatnig, R. & Mehmood, Z. (2016). Image Retrieval by Addition of Spatial Information Based on Histogram Of Triangular Regions. Computer and Electrical Engineering, 54, 539-550. http://dx.doi.org/10.1016/j.compeleceng.2016.04.002

[2] Anwar, H., Zambanini, S. & Kampel, M. (2015). Efficient Scale- and Rotation-Invariant Encoding of Visual Words for Image Classification. IEE Signal Processing Letters, 22(10). doi:10.1109/LSP.2015.2432851

[3] Csurka, G., Dance, C. R., Fan, L., Willamowski, J. & Bray, C. (2004). Visual Categorization with Bags of Keypoints. In Workshop on Statistical Learning in Computer Vision, ECCV. (pp. 1–22).

[4] Farhangi, M. M., Soryani, M., & Fathy, M. (2013). Informative Visual Words Construction to Improve Bag of Words Image Representation. IET Image Processing, 8(5), 310-318. doi:10.1049/iet-ipr.2013.0449

[5] Franco, A., Maltoni, D. & Papi, S. (2017). Grocery Product Detection and Recognition. Expert System with Application, 81, 163-176. http://dx.doi.org/10.1016/j.eswa.2017.02.050

[6] Feng, L., Bhanu, B. & Heraty, J. (2016). A Software System for Automated Identification and Retrieval of Moth Images Based on Wing Attributes. Pattern Recognition. 51, 225-241. http://dx.doi.org/10.1016/j.patcog.2015.09.012

[7]   Ghosh, M., Mukherjee, J. & Parekh, R.   (2013). Fish Shape Recognition Using Multiple Shape Descriptors. International Journal of Computer Applications, 73(16).

[8]   Jiang, Y. -G., Yang, J., Ngo, C. -W. & Hauptmann. A. G. (2010). Representation of Keypoint- based Semantic Concept Detection: A Comprehensive Study. IEEE Transactions on Multimedia, 12(1). doi:10.1109/TMM.2009.2036235

[9]   Khan, R., Barat, C., Muselet, D. & Ducottet, C. (2015). Spatial Histogram of Soft Pairwise Similar Patcher to Improve the Bag- of- Visual- Words Model. Computer Vision and Image Understanding, 132, 102-112. http://dx.doi.org/10.1016/j.cviu.2014.09.005

[10] López-Monroy, A. P., Montes-y-Gómez, M., Escalante, H. J., Cruz-Roa, A., & González, F. A. (2016). Improving The Bovw Via Discriminative Visual N-grams and MKL Strategies. Neurocomputing, 175, 768-781. http://dx.doi.org/10.1016/j.neucom.2015.10.053

[11] Mansourian, L., Abdullah, M. T., Abdullah, L. N., Azman, A., & Mustaffa, M. R. (2016). BoVW Model for Animal Recognition: An Evaluation on Sift Feature Strategies. In Badioze Zaman, H., Robinson, P., Smeaton, A. F., Shih, T. K., Velastin, S., Jaafar, A., & Mohamad Ali, N. (Eds.). Advances in Visual Informatics (pp. 227-236). Heidelberg: Springer International Publishing.

[12] Osman, N. S. & Mustaffa, M. R. (2015). A Review on Content-Based Image Retrieval Representation and Description for Fish. Proceedings of the 2015 4th International Conference on Advanced Computer Science Applications and Technologies (ACSAT 2015), 8-10 December 2015, Kuala Lumpur, Malaysia (pp. 39-44). IEEE.

[13] Pires, R. D. L. et al. (2016). Local Descriptor for Soybean Disease Recognition. Computer and Electronic in Agriculture, 125, 48-55. http://dx.doi.org/10.1016/j.compag.2016.04.032

[14] Sheikh, A. R., Lye, M. H., Mansor, S., & Fauzi, M. F. A. (2011). A Content Based Image Retrieval System for Marine Life Images. 2011 IEEE 15th International Symposium on Consumer Electronics (ISCE), 29–33. http://doi.org/10.1109/ISCE.2011.5973777

[15] Sivic, J. & Zisserman, A. (2003). Video Google: A Text Retrieval Approach to Object Matching in Videos. Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV'03). IEEE.

[16] Tu, J., Sui, H., Feng, W., Sun, K., & Hua, L. (2016). Detection of Damaged Rooftop Areas from High-Resolution Aerial Images Based on Visual Bag-of Words Model. IEEE Geoscience and Remote Sensing Letters, 13(12). doi:10.1109/LGRS.2016.2614298

[17] Wang, R., Ding, K., Yang, J. & Xue, L. (2016). A Novel Method for Image Classification Based on Bag of Visual Words. J. Vis. Commun. Image R., 40, 24-33. http://dx.doi.org/10.1016/j.jvcir.2016.05.022

[18] Wen, C., Zhou, X., Zhou, C., Chen, X. & Guo, Q. (2013). Bag of Visual Words for Cows' Basic Activity Recognition. 2013 International Conference on Information Science and Cloud Computing.

[19] Wang, J., Ji, L., Liang, A. & Yuan, D. (2012). The Identification of Butterfly Families Using Content Based Image Retrieval. Biosystem Engineering III. 24-32. doi:10.1016/j.biosystemseng.2011.10.003

[20] Wen, C., Guyer, D. E. & Li, W. (2009). Local Feature-Based Identification and Classification for Orchard Insects. Biosystem Engineering, 104, 299-307. doi:10.1016/j.biosystemseng.2009.07.002

[21] Zhang, F., et al. (2016). Dictionary Pruning with Visual Word Significance for Medical Image Retrieval. Neurocomputing, 177, 75-88. http://dx.doi.org/10.1016/j.neucom.2015.11.008

[22] Zhu, C., & Peng, Y. (2017). Discriminative Latent Semantic Feature Learning for Pedestrian Detection. Neurocomputing, 238, 126-138. http://dx.doi.org/10.1016/j.neucom.2017.01.043