International Journal of Applied Computer Technology and Information Systems: Volume 8, No.2, October 2018 - March 2019

SiamFishNet: The Deep Investigation of Siamese Fighting Fishes

Lawankorn Mookdarsanit[1*], Pakpoom Mookdarsanit[2]

[1*]Department of Business Computer, Faculty of Management Science
Chandrakasem Rajabhat University
Bangkok, Thailand
lawankorn.s@chandra.ac.th
[2]Department of Computer Science, Faculty of Science
Chandrakasem Rajabhat University
Bangkok, Thailand
pakpoom.m@chandra.ac.th

Abstract—The Siamese fighting fish (or betta) can be seen as ASEAN signature and heritage that should be preserved and collected in form of an intelligent model. This paper is proposed to create the R-CNN based model to investigate what the breed of an unknown betta image is (using only an image), called "SiamFishNet". The model is formulated from 87,560 betta images that cover 12 breed of bettas: (1.) Veil Tail, (2.) Crown Tail, (3.) Half-moon, (4.) Super-delta Tail, (5.) Rose Tail, (6.) Butterfly Tail, (7.) Dragon-scale, (8.) Bumble-bee, (9.) Paradise, (10.) Elephant-ear, (11.) Orchid, and (12.) Spade Tail, respectively. For verification, our model measured by average precision. The result shows that our model has the average precision as 84 %. ***(Abstract)***

Keywords-component; Fish Recognition; Region Proposals; Convolutional Neural Networks;

I. INTRODUCTION

The Siamese fighting fish (known as "betta") is a small colorful aggressive fish that can be often seen in ASEAN like Thailand, Laos, Malaysia, Cambodia, Indonesia and Vietnam. Since the first betta was originally found in any canals from Thailand (Venice of the East).

At early stage, native people were interested in the wild fighting game between 2 bettas, known as Siamese fighting fish(es). From the historical evidence, the King Nangklao of Thailand (Rama III) firstly launched licensing and collecting these bettas [1]. The king gave some bettas to Theodor Edvard [2] – a Danish zoologist and botanist during 1809–1860 who firstly published an article about Siamese fighting fish within his international encyclopedic archives. As of today, some fish fighting matches can be easily seen in many ASEAN countries that look like a general combat boxing.

To put it another way, Pierre Carbonnier – a French ichthyologist who initiated to see those colorful bettas as one of ornamental fishes inside the aquariums [3] and he also was awarded the Gold Medal for researching and breeding of bettas at European International Exhibition of sea and river industries at Paris in 1875. Later, some new colorful patterns of fishes from the breeding between colorful fishes and bettas were firstly done in Berlin during 1894-1896 by Paul Matte (all betta-samples in his experiment, collected and supported by Carbonnier) – a well-known German tropical fish breeder. In 1963, a scene of brutal fighting between bettas in an aquarium was renowned to the world from the film "James Bond: From Russia with Love" [4] with a dialogue "…Siamese fighting fish, a fascinating creatures, brave…" – Ernst Stavro Blofeld (a fictional character in James Bond, acted by Anthony Dawson). It is obviously seen that betta fighting and its beauty is absolutely one of ASEAN heritage.

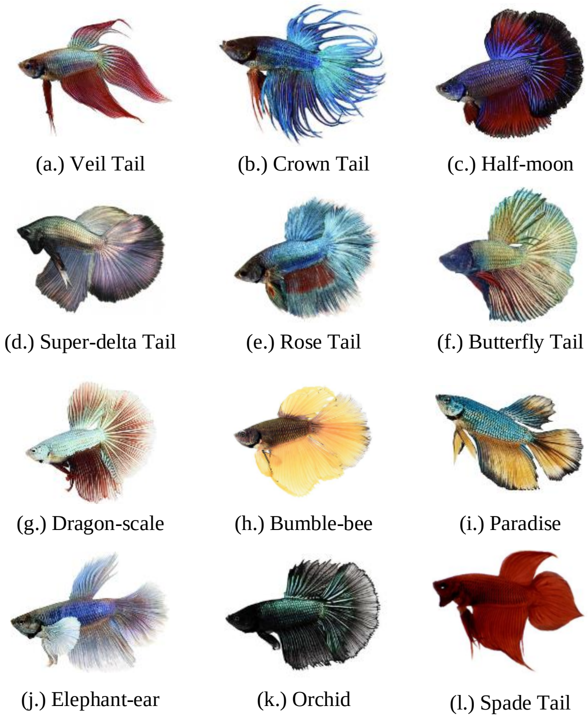

Figure 1. Taxonomic breed of bettas

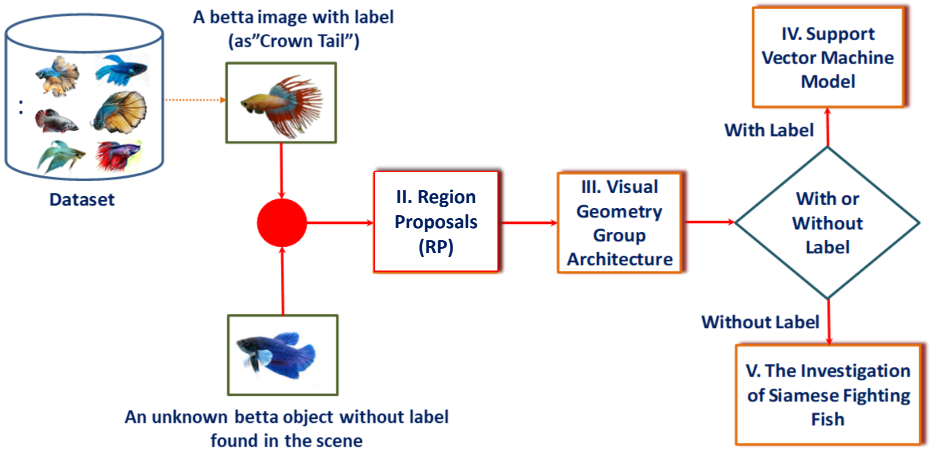

Figure 2. The organization of our R-CNN based "SiamFishNet" Framework

Since the taxonomic breed of bettas can be classified into 12 categories: (1.) Veil Tail, (2.) Crown Tail, (3.) Half-moon, (4.) Super-delta Tail, (5.) Rose Tail, (6.) Butterfly Tail, (7.) Dragon-scale, (8.) Bumble-bee, (9.) Paradise, (10.) Elephant-ear, (11.) Orchid, and (12.) Spade Tail. Each breed of betta has different physical appearance like color, shape, texture and size that most betta experts can investigate the breed of betta by their eyes. Different appearances make various pricing. To that end, it is feasible to use machine learning as an intelligent investigation of betta's appearance from an arbitrary image for online colorful betta website that can be processed on the high-performance cloud [5-6] infrastructure.

Due to computer vision [7-8] with machine learning [9-10] that is successful in many intelligent applications like Generic Scene Classification [11-13], Tourism Image Captioning [14-16], Heritage Recognition [17-19] Face Recognition [20-21], Semantic Folklore Retrieval [22-24], Food Recognition [25-27] and Remote Sensing [28-31], etc. In this paper, we propose a novel "Investigation of Siamese Fighting Fishes" that uses the framework of Region Convolutional Neural Networks (R-CNNs).

Since Convolution Neural Networks (CNNs) (known as AlexNet) outperformed the hand-crafted SIFT-based recognition [32] within a medium-sized codebook [33]. As well as traditional fish recognition [34-36] that CNN for this work is also better than traditional scheme in terms of accuracy and efficiency [37-38]. Notwithstanding, CNNs is mainly for image retrieval like generic fish classification. Moreover, the specific fish (specific bettas) recognition from the aquarium should be both localized and classified that is appropriately done by R-CNNs [39]. During the R-CNN training, the region proposals within a betta image are computed in term of fine-tuned CNN descriptor. And the descriptor with its label is trained to SVM classification. As well as testing, an unknown betta image is computed in term of CNN descriptor. And the SVM classifier investigates the label of image. Our model was formulated by the large-scale 87,560 betta images which only covered 12 breeds (labels or categories). All images were dedicated to an expert from the local community of betta breeders. We denominated our collection and model as "SiamFishNet" that has the average precision as 84%.

The organization of our R-CNN based intelligent model "SiamFishNet" is composed of "Region proposals" in section 2. Section 3 describes "Visual geometry group architecture". Section 4 and 5 as "Support vector machine model" and "The investigation of Siamese fighting fish" are explained about training and testing of SVM model that are shown in Figure 2.

II. Region Proposals

A selective search algorithm is proposed to segment those category-independent regions within a betta image called "Region Proposals (RP)" [40] before the Visual Geometry Group Architecture of CNN. Significantly, the RP selective search is totally based on "Color Similarity ($Color_{Sim}$)", "Texture Similarity ($Texture_{Sim}$)", "Size Similarity ($Size_{Sim}$)" and "Shape Consistency ($Shape_{Cons}$)" as (1).

$$Color_{Sim} \cap Texture_{Sim} \cap Size_{Sim} \cap Shape_{Cons} \qquad (1)$$

A. Color Similarity

Color similarity in each *i-th* region (r_i) is computed by one-dimensional color histogram for each channel using 25 bins ($C_i = \{c_i^1, c_i^2 c_i^3, \ldots, c_i^n\}$, where the dimension n =75). That is to say, the color similarity between the *i-th* region (r_i) and the *j-th* region (r_j) can be measured by (2).

$$Color_{Sim}(r_i, r_j) = \sum_{k=1}^{75} \min(c_i^{k1}, c_j^{k2}) \qquad (2)$$

B. Texture Similarity

This measurement uses the same concept as a well-known FAST description (a dimensional reduction version of SIFT-based approaches) that takes Gaussian derivatives among its 8 neighbor-pixels by $\sigma = 1$ in each channel using 10 bins ($T_i = \{ t_i^1, t_i^2 t_i^3, \ldots, t_i^n \}$, where $n = 240$). Texture similarity between the *i-th* region (r_i) and the *j-th* region (r_j) can be measured by (3).

$$Texture_{Sim}(r_i, r_j) = \sum_{k=1}^{240} \min(t_i^{k1}, t_j^{k2}) \qquad (3)$$

C. Size Similarity

Those small regions are merged together to ensure that object locations in any flexible pyramid scales from any parts of an image using the fraction of an image that the *i-th* region (r_i) and the *j-th* region (r_j) are joint together as described by (4).

$$Size_{Sim}(r_i, r_j) = \frac{size(r_i) + size(r_j)}{size(im)} \qquad (4)$$

where $size(im)$ is the size of an image (or a number of pixels)

D. Shape Consistency

Shape consistency measure how compatible the *i-th* region (r_i) and the *j-th* region (r_j) fit together. In case they are hardly touching together, they will be a form of strange region that must not be merged. The bounding box between r_i and r_j is defined as $BBox_{ij}$. The shape consistency can be measured by (5).

$$Shape_{Con}(r_i, r_j) = \frac{Size(BBox_{ij})}{Size(im)} - Size_{Sim}(r_i, r_j) \qquad (5)$$

III. Visual Geometry Group Architecture

The selected regions with their bounding boxes from previous RP are input to a fine-tuned CNN architecture named Visual Geometry Group (VGG) [41] to extract all CNN features as shown in Figure 3.

The VGG consists of low-level layers as Convolution-coupled with ELU layers and middle-level layers as Max-pooling layers (For R-CNN, the Support Vector Machines: SVMs is used as a model creation instead of Fully-Connected layers: FCs). The even number of layers in convolution and the odd layer number for max-pooling. The output nodes of previous convolution-coupled with ELU and max-pooling layers are grouped into 2D-plane feature mapping. Each layer plane is mostly inherited from the combination of previous layers.

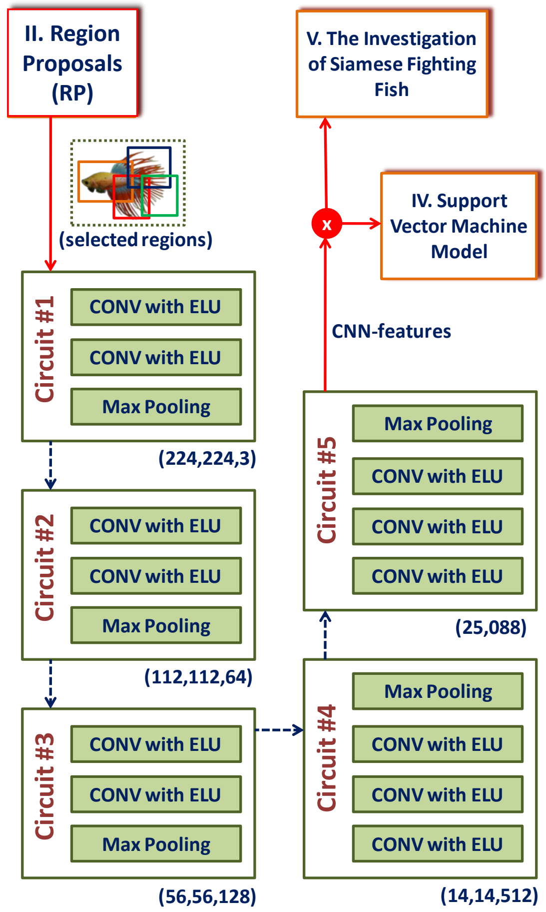

Figure 3. Visual Geometry Group (VGG) Architecture

A. Convolution-coupled with ELU Layers

The previous layers were multi-scalable mapped to formulate the output of kernels (also called "encoding"). Within the layer, each convolution is solved the vanish gradient problem by an ELU (Exponential Linear Unit) activation function [42]. Not only ELU totally speeds up the learning by adaptive parameter but also improves the accuracy. The ELU activation function of current *i-th* node (x_i^l) can be computed from the previous *i-th* node (x_i^{l-1}) using (6).

$$x_i^l = \begin{cases} x_i^{l-1} & , x \geq 0 \\ 0.1 * \exp(x_i^{l-1}), & x < 0 \end{cases} \qquad (6)$$

For the convolution computation can be visualized in Figure 4.

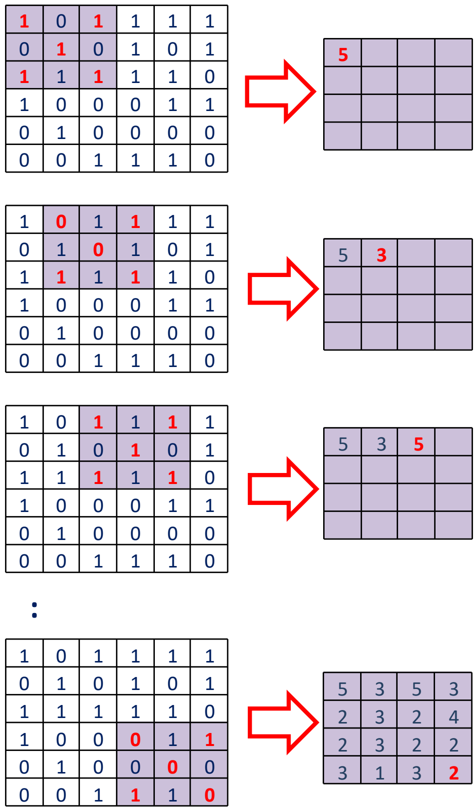

Figure 4. Visualization of Convolution Computation

B. *Max-pooling Layers*

A pooling layer operates the down sampled operation on the input maps. A popular max pooling was firstly done in AlexNet 2012. It was proved that the convolution layer after max pooling often yields the high accuracy of CNN-descriptors before they are trained into the machine model. To give an illustration, a max pooling can be shown in Figure 5.

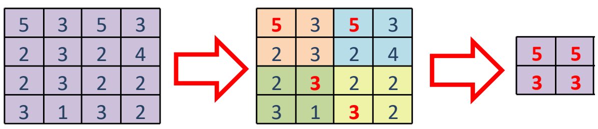

Figure 5. Visualization of Max Pooling

IV. SUPPORT VECTOR MACHINE MODEL

From the previous sections, a betta image is in form of CNN-descriptors with different sizes. Only the CNN-descriptors with a label input to this section to create the (Support Vector Machine) SVM [43] learning model that is composed of The combination of paired-feature from n-size, Label categorization and Linear Separation, respectively.

A. *The Combination of Paired-feature from n-size*

Each paired-feature is unordered and selected without replacement from the n-size of all features using (7).

$$\binom{n}{2} = \frac{n!}{(n-2)! \times 2!} \tag{7}$$

For instance, all paired-features from n-size can be shown in Figure 6.

$$\left\{ \begin{matrix} (F_1,F_2) & (F_1,F_3) & (F_1,F_4) & \dots & (F_1,F_n) \\ & (F_2,F_3) & (F_2,F_4) & \dots & (F_2,F_n) \\ & & (F_3,F_4) & \dots & (F_3,F_n) \\ & & & \dots & \vdots \\ & & & & (F_{n-1},F_n) \end{matrix} \right\} \equiv \forall \mathit{Paired\ Features}$$

Figure 6. The Combination of all Paired Features

B. *Label Categorization*

A large number of betta images in term of many CNN-descriptors with their labels are plotted in each paired-feature, called the categorization of labels. And the line between categorizations is drawn as Figure 7.

C. *Linear Separation*

The previous label categorization is mapped the non-linear line into linear line using the convolution of Gaussian Radial Basis Kernel (GRBK) function ($K(L_i, L_j)$) between neighbor labels $\{L_i, L_j\}$ as (8).

$$K(L_i, L_j) = \phi(L_i) \bullet \phi(L_j) = \exp\left(-\frac{|L_i - L_j|^2}{2\sigma^2}\right) \tag{8}$$

The visualization of linear separation mapping by GRBK Kernel function can be shown in Figure 8.

V. THE INVESTIATION OF SIAMESE FIGHTING FISHES

After the Support Vector Machine model has been created from CNN-features with their labels, the model is used to investigate an unknown betta image. Firstly, an unknown betta image input to the section 2 “Region Proposals (RP)” to select the region. After that, the selected regions are extracted by fine-tuned Visual geometry group (VGG) in section 3. Finally, this section investigate what breed of an unknown image is.

We use the average precision (AP) from 12 labels (or breed of bettas) as our model verification.

From Figure 9, Crown Tail and Elephant-ear have the highest precision as their distinctiveness. Veil Tail, Paradise and Spade Tail perform higher than 90%. In contrast, Half-moon, Super-delta Tail, Rose Tail and Butterfly tail sometimes are looked similar that made them lower than 80%. Moreover, the number of samples in Rose Tail, Super-delta Tail and Orchid are less than other labels that may cause the precision. And the AP of our model is 84% that is formulated from 87,560 betta images.

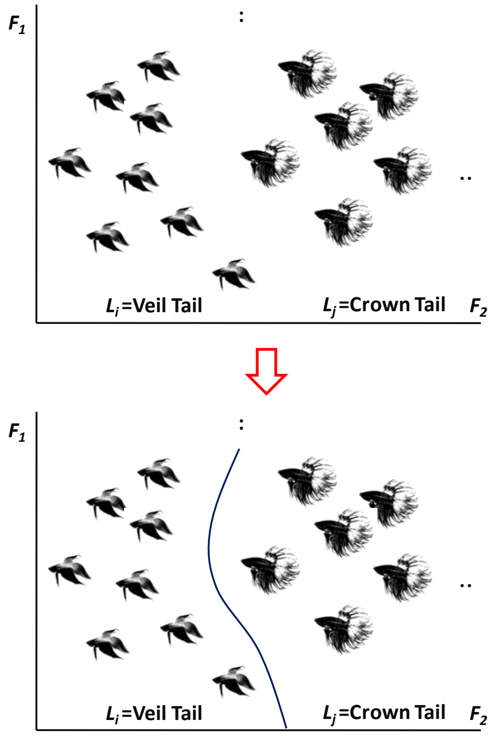

Figure 7. Label Categorization with Line

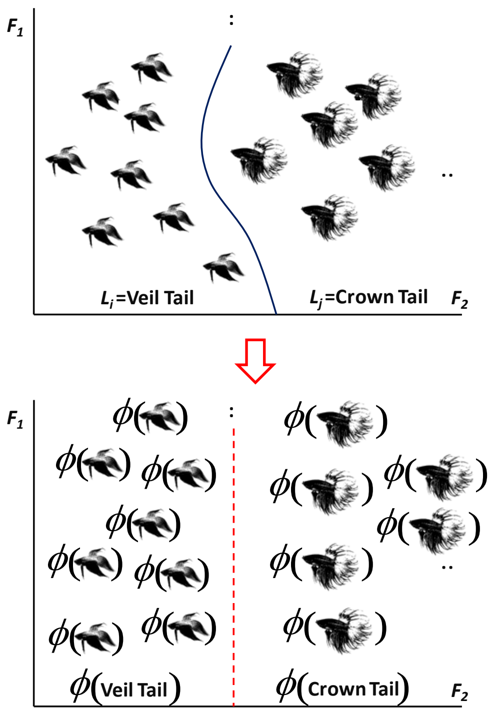

Figure 8. Mapping the Non-linear to Linear line using Kernel Function

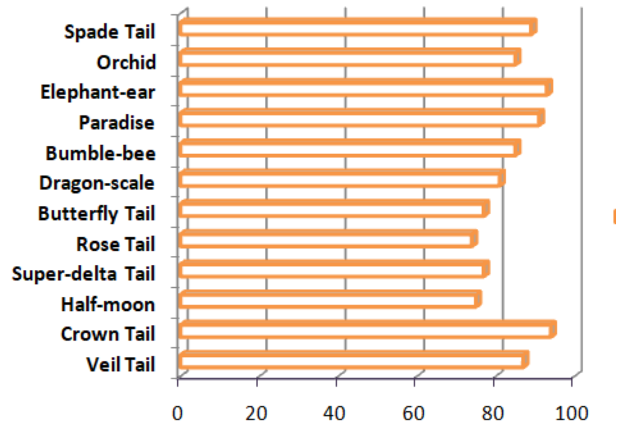

Figure 9. Model Verification by Average Precision

VI. CONCLUSION

This paper called "SiamFishNet" presents R-CNN model for betta (or Siamese fighting fish) categorization into 12 breeds (or labels) that can be divided into 2 parts: training and testing. For training, a betta image with its label firstly inputs to section 2 "Region proposals" to select some appropriated regions within an image. After that, the regions are model in term of CNN-descriptors by section 3 "Visual geometry group". Finally, the CNN-descriptors with its label is trained to create a computer model by section 4 "Support vector machine model". As well as testing, an unknown image inputs to section 2 and 3. And the model investigates what label of the betta image is. For verification, our model uses the average precision as 84%. The highest precision consists of Crown Tail and Elephant-ear. Since we use only 87,560 betta images to create a model. For future work, R-CNN has high precision but it is less time efficiency. Moreover, the higher precision needs more samples. We can use the generative adversarial networks (GANs) [44-45] to randomly generate more million betta images.

REFERENCES

[1] "The History of Betta Fighting Fish: Betta Origins," [Online]. Available: http://www.bettafishcenter.com/Betta-Origins.shtml. [Accessed: 7 August 2018].

[2] B. Beolens, M. Watkins and M. Grayson, The Eponym Dictionary of Reptiles. Baltimore: Johns Hopkins University Press, 2011, pp.47.

[3] "About Bettas and Betta History," [Online]. Available: https://www.chrisbettaworld.com/about-bettas/. [Accessed: 7 August 2018].

[4] "James Bond: From Russia with Love," [Online]. Available: http://www.007museum.com/from_russia_with_love.htm. [Accessed: 7 August 2018].

[5] J. Slabanja, B. Meden, P. Peer, A. Jaklič and F. Solina, "Segmentation and Reconstruction of 3D Models from a Point Cloud with Deep Neural Networks," 2018 International Conference on Information and Communication Technology Convergence (ICTC), Jeju, 2018, pp. 118-123.

[6] P. Mookdarsanit and S. Gertphol, "Light-weight operation of a failover system for Cloud computing," The 5th IEEE International Conference on Knowledge and Smart Technology, Thailand, 2013, pp. 42-46.

[7] C. Leng, H. Zhang, B. Li, G. Cai, Z. Pei and L. He, "Local Feature Descriptor for Image Matching: A Survey," in IEEE Access, 2018.

[8] P. Mookdarsanit, L. Soimart, M. Ketcham and N. Hnoohom, "Detecting Image Forgery Using XOR and Determinant of Pixels for Image Forensics," The 11th IEEE International Conference on Signal-Image Technology & Internet-Based Systems, Bangkok, 2015, pp. 613-616.

[9] L. Soimart and P. Pongcharoen, “Multi-row Machine Layout Design using Aritificial Bee Colony,” 2011 International Conference on Economics and Business Information, Bangkok, 2011.

[10] L. Soimart and P. Mookdarsanit, “An Admission Recommendation of High-school Students using Apriori Algorithm,” The 6th International Conference on Sciences and Social Sciences, 2016.

[11] X. Li, M. Larson and A. Hanjalic, "Geo-Distinctive Visual Element Matching for Location Estimation of Images," in IEEE Transactions on Multimedia, vol. 20, no. 5, pp. 1179-1194, May 2018.

[12] P. Mookdarsanit and M. Rattanasiriwongwut, “Location Estimation of a Photo: A geo-signature MapReduce Workflow,” in Engineering Journal, vol.21, no.3, pp.295-308, May 2017.

[13] P. Mookdarsanit and M. Ketcham, “Image Location Estimation of Well-known Places from Multi-source based Information,” The 11th International Symposium on Natural Language Processing, Ayutthaya, 2016.

[14] L. Soimart and P. Mookdarsanit, “Name with GPS Auto-tagging of Thai-tourist Attractions from An Image,” The 2017 Technology Innovation Management and Engineering Science International Conference, 2017.

[15] P. Mookdarsanit and L. Mookdarsanit, “Contextual Image Classification towards Metadata Annotation of Thai-tourist Attractions,” in ITMSoc Transactions on Information Technology Management, vol.3, no.1, pp.32-40, 2018.

[16] Z. Xu, J. Du, L. Ye and D. Fan, "Multi-feature indexing for image retrieval based on hypergraph," 2016 4th International Conference on Cloud Computing and Intelligence Systems (CCIS), Beijing, 2016, pp. 494-500.

[17] P. Mookdarsanit and M. Rattanasiriwongwut, “GPS Determination of Thai-temple Arts from a Single Photo,” The 11th International Conference on on Applied Computer Technology and Information Systems, Bangkok, Thailand, 2017, pp. 42-47.

[18] M. G. Abdelmonem, "Navigating virtual heritage applications for historic cities in the middle east," 2017 23rd International Conference on Virtual System & Multimedia (VSMM), Dublin, 2017, pp. 1-8.

[19] P. Mookdarsanit and M. Rattanasiriwongwut, “MONTEAN Framework: A Magnificent Outstanding Native-Thai and Ecclesiastical Art Network,” in International Journal of Applied Computer Technology and Information Systems, vol.6, no.2, pp.17-22, 2017.

[20] R. R. Atallah, A. Kamsin, M. A. Ismail, S. A. Abdelrahman and S. Zerdoumi, "Face Recognition and Age Estimation Implications of Changes in Facial Features: A Critical Review Study," in IEEE Access, vol. 6, pp. 28290-28304, 2018.

[21] L. Soimart and P. Mookdarsanit, “Gender Estimation of a Portrait: Asian Facial-significance Framework,” The 6th International Conference on Sciences and Social Sciences, Mahasarakham, 2016.

[22] P. Rakshit, S. Saha, A. Konar and A. K. Nagar, "Evolutionary Approach to Straight Line Approximation for Image Matching in Dance-Posture Recognition," 2018 IEEE Congress on Evolutionary Computation (CEC), Rio de Janeiro, Brazil, 2018, pp. 1-8.

[23] P. Mookdarsanit and L. Mookdarsanit, “An Automatic Image Tagging of Thai Dance’s Gestures,” Joint Conference on ACTIS & NCOBA, Ayutthaya, Thailand, January 2018, pp. 76-80.

[24] P. Mookdarsanit and L. Mookdarsanit, “A Content-based Image Retrieval of Muay-Thai Folklores by Salient Region Matching,” in International Journal of Applied Computer Technology and Information Systems, vol.7, no.2, pp.21-26, 2018.

[25] E. Aguilar, B. Remeseiro, M. Bolaños and P. Radeva, "Grab, Pay, and Eat: Semantic Food Detection for Smart Restaurants," in IEEE Transactions on Multimedia, vol. 20, no. 12, pp. 3266-3275, 2018.

[26] L. Soimart and P. Mookdarsanit, “Ingredients estimation and recommendation of Thai-foods,” in SNRU Journal of Science and Technology, vol.9, no.2, pp.509-520, Jul. 2017.

[27] P. Mookdarsanit and L. Mookdarsanit, “Name and Recipe Estimation of Thai-desserts beyond Image Tagging,” in Kasembundit Engineering Journal, vol.8, Special Issue, pp.193-203, May. 2018.

[28] L. Soimart and M. Ketcham, “An efficient algorithm for earth surface interpretation from satellite imagery,” in Engineering Journal, vol.20, no.5, pp.215-228, Nov. 2016.

[29] L. Soimart and M. Ketcham, “Hybrid of Pixel-based and Region-based Segmentation for Geology Exploration from Multi-spectral Remote Sensing,” The 11th International Symposium on Natural Language Processing, Ayutthaya, 2016.

[30] M. Wang, Z. Dong, Y. Cheng and D. Li, "Optimal Segmentation of High-Resolution Remote Sensing Image by Combining Superpixels With the Minimum Spanning Tree," in IEEE Transactions on Geoscience and Remote Sensing, vol. 56, no. 1, pp. 228-238, Jan. 2018.

[31] L. Soimart and M. Ketcham, “The Segmentation of Satellite Image Using Transport Mean-shift Algorithm,” 13th International Conference on IT Applications and Management (ITAM-13), 2015, pp. 124-128.

[32] L. Zheng, Y. Yang and Q. Tian, "SIFT Meets CNN: A Decade Survey of Instance Retrieval," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 5, pp. 1224-1244, 1 May 2018.

[33] A. A. Olaode, G. Naghdy and C. A. Todd, "Unsupervised Image Classification by Probabilistic Latent Semantic Analysis for the Annotation of Images," 2014 International Conference on Digital Image Computing: Techniques and Applications, Wollongong, NSW, Australia, 2014, pp. 1-8.

[34] B. İşçįMEN, Y. Kutlu, A. N. Reyhaniye and C. Turan, "Image analysis methods on fish recognition," 2014 22nd Signal Processing and Communications Applications Conference (SIU), Trabzon, 2014, pp. 1411-1414.

[35] Y. Nishida, T. Ura, T. Hamatsu, K. Nagahashi, S. Inaba and T. Nakatani, "Fish recognition method using vector quantization histogram for investigation of fishery resources," 2014 Oceans - St. John's, St. John's, NL, 2014, pp. 1-5.

[36] M. Chuang, J. Hwang and K. Williams, "A Feature Learning and Object Recognition Framework for Underwater Fish Images," in IEEE Transactions on Image Processing, vol. 25, no. 4, pp. 1862-1872, April 2016.

[37] K. Demertzis, L. Iliadis and V. Anezakis, "A deep spiking machine-hearing system for the case of invasive fish species," 2017 IEEE International Conference on Innovations in Intelligent Systems and Applications, Gdynia, 2017, pp. 23-28.

[38] K. Wang, K. D. Do and L. Cui, "An underwater electrosensor for identifying objects of similar volume and aspect ratio using convolutional neural network," 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems, Vancouver, BC, 2017, pp. 4963-4968.

[39] R. Girshick, J. Donahue, T. Darrell and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, 2014, pp. 580-587.

[40] J. R. R. Uijlings, K. E. A. van de Sande, T. Gevers and A. W. M. Smeulders, "Selective Search for Object Recognition," in International Journal of Computer Vision, vol. 104, no.2, pp. 154-171, 2013.

[41] K. Simonyan and A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition," The 3rd International Conference on Learning Representations, San Diego, CA, 2015.

[42] K. He, X. Zhang, S. Ren and J. Sun, "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification," 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, 2015, pp. 1026-1034.

[43] X. Qi, T. Wang and J. Liu, "Comparison of Support Vector Machine and Softmax Classifiers in Computer Vision," 2017 Second International Conference on Mechanical, Control and Computer Engineering, Harbin, 2017, pp. 151-155.

[44] G. Zhang, E. Tu and D. Cui, "Stable and improved generative adversarial nets (GANS): A constructive survey," 2017 IEEE International Conference on Image Processing, Beijing, 2017, pp. 1871-1875.

[45] X. Wu, K. Xu and P. Hall, "A survey of image synthesis and editing with generative adversarial networks," in Tsinghua Science and Technology, vol. 22, no. 6, pp. 660-674, December 2017.