2015 International Symposium on Nonlinear Theory and its Applications
NOLTA2015, Kowloon, Hong Kong, China, December 1-4, 2015

Similarity-based Image Retrieval by Self-Organizing Map with Refractoriness — New Distance between Image Features —

Hideto NAKAJIMA and Yuko OSANA

School of Computer Science, Tokyo University of Technology
1404-1 Katakura, Hachioji, Tokyo, 192-0982, Japan Email: osana@stf.teu.ac.jp

Abstract—In most of the conventional image retrieval systems, the Euclidean distance is used in order to compare image features. However, in the Euclidean distance, 1 (which means the feature is included) and 0 (which means the feature is not included) are treated equally. In this paper, we propose a similarity-based image retrieval by self-organizing map with refractoriness using new distance between images features. We carried out a series of computer experiments and confirmed the effectiveness of the proposed system.

1. Introduction

Recently, some similarity-based image retrieval systems which make use of the flexible information processing ability of artificial neural networks have been proposed[1]-[7]. Most of these systems use the Euclidean distance in order to compare image features. However, in the Euclidean distance, 1 (which means the feature is included) and 0 (which means the feature is not included) are treated equally.

In this paper, we propose a similarity-based image retrieval by self-organizing map with refractoriness using new distance between images features.

2. Image Features

Color, color and size of artifacts, shape (distance from circumcircle)[8], SIFT (Scale-Invariant Feature Transform)[9], HOG (Histograms of Oriented Gradients)[10], spectrum, LBP (Local Binary Pattern)[11] and keywords are used as image features.

2.1. Color

Each image is divided into some regions by the K-means algorithm[12]. Then, normalized average x, y and z coordinates in the HSV color space on each region are trained in the self-organizing map, and they are used as image feature. Here, two self-organizing maps for natural objects and artifacts are used. In the proposed system, the features on color are calculated per nine areas.

The feature vector on the color of the natural objects at the area s in the image p, $\boldsymbol{x}^{N(p,s)}$ is given by

$$x_i^{N(p,s)} = g\left(\sum_{l\in C_s^N} x_i^{N(p,s,l)}\right) \quad (1)$$

$$g(u) = \begin{cases} 1 & (u > 0) \\ 0 & (u = 0) \end{cases} \quad (2)$$

where C_s^N is the set of the regions which belong to the area s for the natural objects, $x_i^{N(p,s,l)}$ is the output of the neuron i in the self-organizing map which learns the color of the natural objects when the color information at the region l which belongs to the area s in the image p is given.

The feature vector on the color of the artifacts at the area s in the image p, $\boldsymbol{x}^{A(p,s)}$ is given by

$$x_i^{A(p,s)} = g\left(\sum_{l\in C_s^A} x_i^{A(p,s,l)}\right) \quad (3)$$

where C_s^A is the set of the regions which belong to the area s for the artifacts, $x_i^{A(p,s,l)}$ is the output of the neuron i in the self-organizing map which learns the color of the artifacts when the color information at the region l which belongs to the area s in the image p is given.

2.2. Color and Size of Artifacts

The feature vector on color of artifacts $\boldsymbol{x}^{A2(p)}$ is calculated by

$$x_i^{A2(p)} = \sum_{l\in C^A} x_i^{A(p,l)} r^{A(p,l)} \quad (4)$$

where C^A is the set of artifact areas, $x_i^{A(p,l)}$ is the output of the neuron i of the self-organizing map which learns the color of the artifacts when the color information at the region l which belongs to the area s in the image p is given, $r^{A(p,l)}$ is the rate in the whole artifacts areas of the area l of the image p.

The feature on size of artifacts $x^{A3(p)}$ is calculated by

$$x^{A3(p)} = \sum_{l\in C^A} S^{(p,l)} \Big/ \sum_{l} S^{(p,l)} \quad (5)$$

where $S^{(p,l)}$ is the number of pixels of the region l of the image p.

2.3. Shape (Distance from Circumcircle)

As the method to describe the shape, we use the distance between the point on the circumcircle and the edge of the object [8]. In this method, first, the center of the object is found by the moments, and the circumcircle centering on the point is drawn. Then the distance from the point on the circumference to the edge of the object toward the center is calculated.

2.4. SIFT (Scale-Invariant Feature Transform)

SIFT is an algorithm to detect and describe local features in images. The feature vector on SIFT is generated based on the idea of the Bag-of-Features[13].

2.5. HOG (Histograms of Oriented Gradients)

HOG is feature descriptors which is used for object detection. The feature vector on HOG is generated based on the idea of the Bag-of-Features as similar as the feature vector on SIFT.

2.6. Spectrum

An image spectrum is used as one of image features. Here, the calculated spectrum is divided into $N^k \times N^l$ areas, and binarized average spectrum in each area is used.

2.7. LBP (Local Binary Pattern)

LBP is the texture features which uses patterns that shows the magnitude relation of local brightness in an image. In this system, normalized histogram of LBP is used as one of image features.

2.8. Keywords

Keywords such as sky, cloud also can be used as the query.

3. Similarity-based Image Retrieval using Self-Organizing Map with Refractoriness

3.1. Structure

The proposed image retrieval system is based on the self-organizing map with refractoriness[2] and it has two layers; (1) Input Layer and (2) Map Layer. The neurons in the Input Layer receives the feature vector of key image(s) and the neurons in the Map Layer whose connection weights are similar to the input feature vector fires. In the proposed system, each neuron in the Map Layer corresponds to one of the stored images.

3.2. Learning Process

In the learning process of the proposed system, image features of the images to be stored are trained in the self-organizing map with refractoriness.

Step 1 : Extraction of Artifacts

In **Step 1**, the original image is divided into some regions by the K-means algorithm, and whether or not artifacts are included is judged for each divided area.

Step 2 : Generation of Feature Vectors

In **Step 2**, the image features are extracted from the images to be stored, and the feature vectors are generated.

Step 3 : Learning of Self-Organizing Map with Refractoriness

In **Step 3**, the feature vectors generated in **Step 2** are trained in the self-organizing map with refractoriness.

3.3. Image Retrieval Process

In the proposed system, the following five search requests are considered as similar as the conventional system[7], and the feature vector is generated for the key image based on the search request which is selected by a user.

(1) Retrieval of images which are similar to key image

(2) Retrieval of images which are similar to a part of key image

(3) Retrieval of images which have similar feature of same positions

(4) Retrieval of images which have similar features to common features in plural key images

(5) Retrieval of images which include similar artifacts in different position

(1) Generation of Feature Vector

The feature vector is generated for the key image(s) based on the search request which is selected by a user.

If Search Request 3 is selected, the feature vector on color information of natural objects is given by

$$x_i^{N(p,s)} = \begin{cases} -1 & \text{(if all regions in area } s \text{ are outside the selected part by users)} \\ g\left(\sum_{l \in C_s^N} x_i^{N(p,s,l)}\right) & \text{(otherwise)} \end{cases} \quad (6)$$

where C_s^N is the set of the regions which belong to the area s for the natural objects, $x_i^{N(p,s,l)}$ is the output of the neuron i in the self-organizing map which learns the color of the natural objects when the color information at the region l which belongs to the area s in the image p is given. In the proposed system, if all regions in area s are outside the selected part by users are set to -1. In the same way, the feature vector for the color information of artifacts is generated. And, spectrum and LBP are not used.

If Search Request 4 is selected, the feature vector is generated from the plural key images. In the proposed system, first, the feature vector for each key image $\boldsymbol{x}^{(p)}(p = 1, \cdots, N^{key})$ is generated. Here, N^{key} is the number of key images. Then, the feature vector for the plural key images $\boldsymbol{x}$ is generated from $\boldsymbol{x}^{(p)}$ as follows:

$$x_i = \begin{cases} x_i^{(p)} & \left(\sum_{p=1}^{N^{key}-1} \sum_{q=p+1}^{N^{key}} |x_i^{(p)} - x_i^{(q)}| = 0\right) \\ -1 & \text{(otherwise)} \end{cases} \quad (7)$$

As shown in Eq.(7), the proposed system uses only the common features in all key images.

(2) Image Retrieval

The image retrieval process of the proposed system has four steps.

Step 1 : Input of Feature Vector

Image features of the key image(s) are given to the Input Layer.

Step 2 : Calculation of Internal States of Neurons in Map Layer

When the image feature of the key image(s) is given to the Input Layer, the internal state of the neuron i of the module y in the Map Layer at the time t, $u_i^y(t)$ is calculated by

$$u_i^y(t) = \frac{\sum_{f \in C_F} S^f(\boldsymbol{w}_i^y, \boldsymbol{x})}{F'} - \alpha \sum_{d=0}^{t-1} k_r^d x_i^{MAP}((t-1)-d) \quad (8)$$

where F' is the number of image features which are used in the retrieval process, C_F is the set of features which are used in the retrieval process, α is the scaling factor, k_r is the damping factor, and $x_i^{MAP(y)}(t)$ is the output of the neuron i of the module y in the Map Layer at the time t. $S^f(\boldsymbol{w}_i^y, \boldsymbol{x})$ is the similarity the weight vector of the neuron i of the module y in the Map Layer $\boldsymbol{w}_i^y$ and the input $\boldsymbol{x}$, and is defined by

$$S^f(\boldsymbol{w}_i^y, \boldsymbol{x}) = \begin{cases} \dfrac{2N_f^{C_1}(\boldsymbol{w}_i^y, \boldsymbol{x}) + k_w(N_f^{C_2}(\boldsymbol{w}_i^y, \boldsymbol{x}) + N_f^{C_3}(\boldsymbol{w}_i^y, \boldsymbol{x}))}{N_f^w(\boldsymbol{w}_i^y) + N_f^k(\boldsymbol{x})} & (f = 11) \\ \dfrac{2N_f^{C_1}(\boldsymbol{w}_i^y, \boldsymbol{x})}{N_f^w(\boldsymbol{w}_i^y) + N_f^k(\boldsymbol{x})} & (\text{otherwise}) \end{cases} \quad (9)$$

where f (= 1, ..., F) is an image feature (1 : color (natural objects), 2 : color (artifacts), 3 : color ratio of artifacts, 4 : size of artifacts, 5 : shape (distance from circumcircle), 6 : SIFT, 7 : HOG, 8 : spectrum (natural objects), 9 : spectrum (artifacts) 10 : LBP, 11 : keywords).

In Eq.(9), $N_f^w(\boldsymbol{w}_i^y)$ is the number of elements whose value are 1 corresponding to the image feature f in the weight vector $\boldsymbol{w}_i^y$, and it is given by

$$N_f^w(\boldsymbol{w}_i^y) = \sum_{j: x_j \in C_f} w_{ij}^y. \quad (10)$$

$N_f^k(\boldsymbol{x})$ is the number of elements whose value are 1 corresponding to the image feature f in the feature vector $\boldsymbol{x}$, and it is given by

$$N_f^k(\boldsymbol{x}) = \sum_{j: x_j \in C_f \text{and} x_j \neq -1} x_j. \quad (11)$$

$N_f^{C_1}(\boldsymbol{w}_i^y, \boldsymbol{x})$ is the number of elements whose value are 1 corresponding to the image feature f in the weight vector $\boldsymbol{w}_i^y$ and the feature vector $\boldsymbol{x}$, and it is given by

$$N_f^{C_1}(\boldsymbol{w}_i^y, \boldsymbol{x}) = \sum_{j: x_j \in C_f \text{and} x_j \neq -1} w_{ij}^y x_j. \quad (12)$$

Step 3 : Calculation of Outputs of Neurons in Map Layer

Table 1: Precision, Recall and F-measure (without Keywords).

		Precision	Recall	F-measure
Search Request 1	Conventional	0.981	0.467	0.633
	Proposed	0.942	0.645	0.766
Search Request 2	Conventional	0.819	0.554	0.658
	Proposed	0.693	0.800	0.743
Search Request 3	Conventional	0.795	0.627	0.701
	Proposed	0.847	0.721	0.779
Search Request 4	Conventional	0.764	0.477	0.587
	Proposed	0.799	0.820	0.809
Search Request 5	Conventional	0.953	0.461	0.622
	Proposed	0.966	0.783	0.865

Table 2: Precision, Recall and F-measure (with Keywords).

		Precision	Recall	F-measure
Search Request 1	Conventional	0.992	0.824	0.900
	Proposed	0.960	0.855	0.904
Search Request 2	Conventional	0.908	0.840	0.873
	Proposed	0.818	0.856	0.837
Search Request 3	Conventional	0.884	0.775	0.826
	Proposed	0.877	0.885	0.881
Search Request 4	Conventional	0.994	0.955	0.974
	Proposed	0.980	0.981	0.980
Search Request 5	Conventional	0.968	0.835	0.897
	Proposed	0.981	0.826	0.897

The output of the neuron i of the module y in the Map Layer at the time t, $x_i^{MAP(y)}(t)$ is calculated by

$$x_i^{MAP(y)}(t) = \begin{cases} 1 & (i = c^{(y)} \text{ and } u_i^y(t) > \theta_{s1} \text{ and } S_{min}^{y(i)} > \theta_{s2}) \\ 0 & (\text{otherwise}) \end{cases} \quad (13)$$

where θ_{s1} is the threshold for the internal state of the neuron, and θ_{s2} is the threshold for the similarity of each feature. $S_{min}^{y(i)}$ is given by

$$S_{min}^{y(i)} = \min_f(\boldsymbol{w}_i^y, \boldsymbol{x}). \quad (14)$$

In the proposed system, each stored image corresponds to a neuron in the Map Layer. So, the images corresponding to the fired neurons in the Map Layer are output.

Step 4 : Repeat

Steps 2 and **3** are repeated.

4. Computer Experiment Results

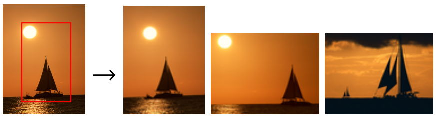
Tables 1 and 2 show the precision, the recall and F-measure of the proposed system and the conventional system[7] which stores 550 images. Figure 1 shows the retrieval results of the proposed system and the conventional system[7] which stores 550 images. From these results, we confirmed that similar images can be searched with the proposed system more correctly than the conventional system.

5. Conclusions

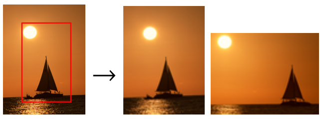
In this paper, we have proposed the similarity-based image retrieval by self-organizing map with refractoriness us-

(a) Search Request 1 (Proposed System)

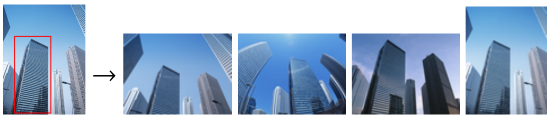
(b) Search Request 1 (Conventional System[7])



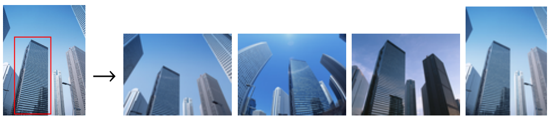
(c) Search Request 2 (Proposed System)



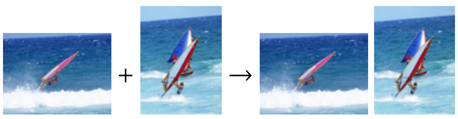
(d) Search Request 2 (Conventional System[7])



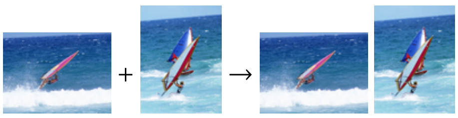
(e) Search Request 3 (Proposed System)



(f) Search Request 3 (Conventional System[7])

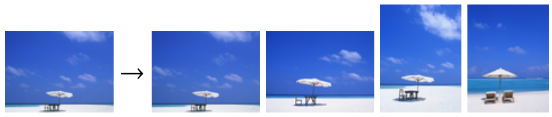


(g) Search Request 4 (Proposed System)



(h) Search Request 4 (Conventional System[7])

(i) Search Request 5 (Proposed System)



(j) Search Request 5 (Conventional System[7])

Figure 1: Retrieval Results.

ing new distance between images features. We carried out a series of computer experiments and confirmed that images can be searched with the proposed system more correctly than the conventional system.

References

[1] J. T. Laaksonen, J. M. Koskela, S. P. Laakso and E. Oja: "PicSOM—content-based image retrieval with self-organizing maps," Pattern Recognition Letters, Vol.21, No.13, 14, pp.1199–1207, 2000.

[2] H. Mogami, M. Otake, N. Kouno and Y. Osana : "Self-organizing map with refractoriness and its application to image retrieval," Proceedings of IEEE International Joint Conference on Neural Networks, Vancouver, 2006.

[3] K. Nagashima, M. Nakada and Y. Osana: "Similarity-based image retrieval by self-organizing map with refractoriness," Proceedings of IEEE International Joint Conference on Neural Networks, Orlando, 2007.

[4] Y. Kuramochi, R. Fukazawa and Y. Osana: "Similarity-based image retrieval considering artifacts from a part of images by self-organizing map with refractoriness," Proceedings of IEEE International Conference on System, Man and Cybernetics, Seoul, 2012.

[5] I. Kitano and Y. Osana : "Similarity-based image retrieval considering position fluctuation by self-organizing map with refractoriness," Proceedings of IASTED Artificial Intelligence and Applications, Innsbruck, 2013.

[6] T. Yamane, R. Ohno and Y. Osana : "Similarity-based image retrieval considering size and color of artifacts by self-organizing map with refractoriness," Proceedings of IEEE International Conference on System, Man and Cybernetics, Manchester, 2013.

[7] I. Miura and Y. Osana : "Similarity-based image retrieval considering shape and texture by self-organizing map with refractoriness," Proceedings of International Symposium on Nonlinear Theory and its Applications, Luzern, 2014.

[8] K. Kushima, H. Akama, S. Kon'ya, H. Kimoto and M. Yamamuro, "ExSight: An Object Based High Performance Image Retrieval System," Information Processing Society of Japan, Vol.40, No.2, pp. 732–741, 1999.

[9] D. G. Lowe : "Object recognition from local scale-invariant features," Proceeding of IEEE International Conference on Computer Vision, pp.1150–1157, 1999.

[10] N. Dalal, and B. Triggs : "Histograms of oriented gradient for human deteciton," Proceeding of IEEE Conference on Computer Vision and Pattern Recognition, pp.886–893, 2005.

[11] T. Ojala, M. Pietikäinen and D. Harwood : "A comparative study of texture measures with classification based on feature distributions," Pattern Recognition, Vol.29, pp.51–59, 1996.

[12] J. B. MacQueen : "Some methods for classification and analysis of multivariate observations," Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability, Vol.1, No.297, University of California Press, 1967.

[13] G. Csurka, C.R.Dance, L.Fan and C.Bray : "Visual categorization with bags of Keypoints," Proceedings of European Conference on Computer Vision, pp.1–22,2004.