A Novel Approach to Image Search System by using Markov Stationary Features

Thi Thi Zin, Pyke Tin, Takashi Toriu
Graduate School of Engineering
Osaka City University
Osaka, Japan
thithi@info.eng.osaka-cu.ac.jp

Hiromitsu Hama
Research Center for Industry Innovation
Osaka City University
Osaka, Japan

***Abstract*— The number of searchable images available is now in the hundreds of millions, and the amount is continuously growing. Two prominent sources contributing to this trend are personal albums (e.g., Flickr, Piczo, Picasa web album and photo.net) and general-purpose image collections (e.g., Google Images and Yahoo! Images). Users of most large image collections often face the problem of how to retrieve relevant images from the collection. One of the main obstacles is that the system does not necessarily understand the user's intentions behind a search. In this paper, we propose a concepts oriented Markov stationary feature to represent images and also present a classifier based scheme for web image search systems. First, a set of query images is explored to learn a Markov stationary feature about the concept by using a correlated random walk model, in which the spatial co-occurrence of the bag-of-features representation and the concept information are integrated. The resulting feature vectors are relatively low in dimensions compared to those in other systems and are then applied to search web images. The experimental results demonstrate that our system is capable of retrieving web images that belong to the same category.**

Keywords- image search system; Markov stationary features; co-occurrence matrix; similarity; random walk model

I. INTRODUCTION

Information on the internet is shifting from text-based to multimedia-based with a large amount of visual and audio data. The development of tools such as digital cameras and scanners, which convert analog data into digital data, has accelerated the increase in multimedia information on the internet, and widened internet bandwidth has dramatically improved access. These changes have demonstrated the need for current internet search systems to improve their search engines to include multimedia data such as images, music, and videos. Among these, images are most numerous, requiring a more efficient searching technique. To better support image indexing and retrieval at semantic level, the research on web image searching or collecting are really necessary. To enable web image search effectively, a suitable image representation to interpret the semantics of images is one of important and challenging problems.

On the other hand many applications such as the digital libraries, image search engines, logo, and trademarks search systems require effective and efficient image retrieval techniques to access the images based on their contents [1]. Such technique computes relevance of the query and database images based on the visual similarity of the low-level features (e.g., color, texture, shape, edge, etc.) derived entirely from the images [1, 2]. In image retrieval research, researchers are moving from keyword based, to content based then towards semantic based image retrieval and the main problem encountered in the content based image retrieval research is the semantic gap between the low-level feature representing and high-level semantics in the images.

A key function in the image search system is feature extraction. A feature is a characteristic that can capture a certain visual property of an image either globally for the whole image, or locally for objects or regions. Some key issues are the following. First, how to present image contents for the extracted features. Second, how to determine the similarity between images based on their extracted features. One technique for these issues is using a vector model. This model represents an image as a vector of features and the difference between two images is measured via the distance between their feature vectors.

In this paper, we propose a novel image representation which we name as the concept oriented Markov stationary features to retrieve relevant images from the large scale image databases such as World Wide Web. By exploring the basic idea of the correlated random walk model, the proposed Markov stationary feature not only extends the bag-of-features representation with spatial structure information as in [3-5] but also involves the concept information of images into the final representation. During the search process, the feature vector of the query image is computed and matched against those features in the database. This paper is organized as follows. The related works are described in section II. The overview of the proposed image search system is discussed in section III. The experimental results are presented in section IV followed by conclusions and future work in section V.

II. RELATED WORKS

In the content-based system, queries such as query by example image, query by sketch and drawing, and query by selected color and texture patterns are supported. The visual features include color, texture, and shape [3]. Color is represented by color set, texture based on co-occurrence and probability matrix, and spatial relationship between image regions [4–5]. The photobook system is composed of a set of interactive tools for browsing and searching images. It supports query by example. The images are organized in three sub-books from which shape, texture, and face appearance features are extracted respectively [6]. The differences between our system and the previous systems are in feature extraction and query strategy. For feature extraction, we propose a novel Markov Stationary feature vector. The spatial information of an image is automatically obtained using a unique unsupervised segmentation algorithm [7] in combination with the Markov chain technique. Our query strategy includes a color filter and a spatial filter, which greatly reduces the search range and therefore speeds up the query.

Even though, considerable amount of work had already been done for texture images, lot of issues exists in extracting the orientation, directionality and regularity features for the image retrieval. Central to these issues are image similarity measures also called ‘distance functions’, or more generally in information theory, ‘distortion measures’, that quantify how well one image matches another. During the past decade, a number of similarly measures have been proposed to compare multi-dimensional sequences. Mean distance [8] is the most generic among them and it is firstly extends the traditional Euclidean distance for time series to support multi-dimensional sequence matching. Other classical examples, such as histogram representations, e.g. color histogram, histogram of local binary patterns [9, 10], and Bag-of-Words based on SIFT features [11] have been widely used for visual recognition, content based image retrieval, and video content analysis. The inability of conventional histogram features to convey spatial and temporal contextual information, however, greatly limits their discriminating power. Layout histograms and multi-resolution histograms [12] are the pioneering attempts to incorporate spatial contextual information for improving the discriminating capability of the histogram features. Instead of the indirect use of spatial contextual information, coherence vector [13] and autocorrelation [14] were proposed to encode local spatial contextual information directly into histograms. In [15, 16], we introduced the spatial co-occurrence matrix based Embedded Markov Chain (EMC) model to encode the intra-histogram-dominant color bin and inter-histogram-dominant color bin relationships into histograms where the initial and steady state distributions of the Markov chain model are combined to form the new Markov Feature Vector (MFV).

III. The Proposed Image Search System

The architecture of our system is shown in Fig. 1. There are two main components in the system. The first component is the visual content extraction and indexing. Each image in the image database is analyzed and the color and spatial information are generated using the color label histogram computation algorithm [17–19] and the unsupervised segmentation algorithm [7] respectively. The obtained features are transformed in a Markov stationary feature database and organized in an efficient way for query retrieval. The second component is the query engine. It consists of a query user interface and a query processing subcomponent. Query by example image is supported in the system. When a user issues a query through the query user interface, the query processing subcomponent computes the Markov stationary similarity measure between the query image and each image in the search range. Two filters, the color filter and the spatial filter, are used to reduce the search range. The top k images similar to the query image are displayed in the query user interface.

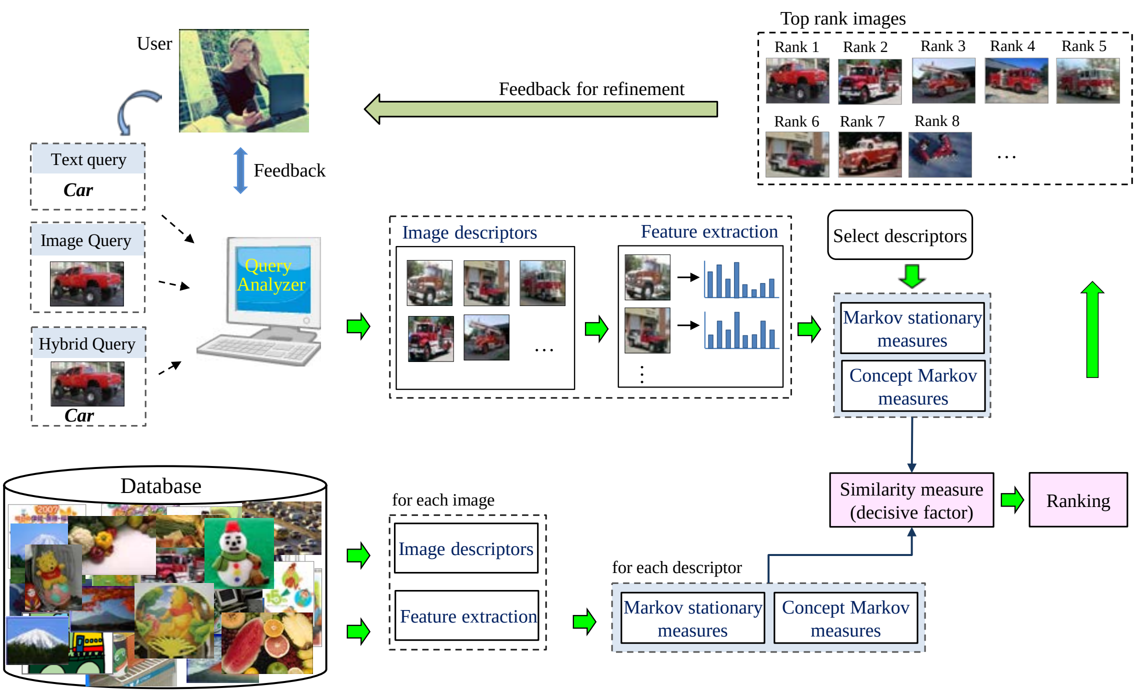

Figure 1. Overview of proposed image search system.

A. General Concepts of Markov Stationary Feature

1) Notation, Definitions and Properties: We consider a finite lattice *S* with a neighborhood structure $N = \{N(x, y) \mid (x, y) \in S\}$, where every *N*, is a subset of the lattice *S*. The number of elements of a subset $A \subset S$ will be denoted by $n(A)$. The following definition is basic to the new framework.

a) Definition 2.1.

Let A_1, A_2 be two subsets of *S*. Then the co-function set of A_1 with respect to A_2 for the neighborhood structure is a subset of *S* defined by

$$CF(A_1, A_2, N) = \bigcup_{(x,y)\in A_1} [N(x, y) \cap A_2]. \quad (1)$$

b) Definition 2.2.

The co-function set measure $cm(A_1, A_2)$ is defined by

$$cm(A_1, A_2) = \sum_{(x,y)\in A_1} [N(x, y) \cap A_2] \quad (2)$$

which is the number of elements in $CF(A_1, A_2, N)$.

c) Definition 2.3.

The co-function matrix is formed by using the co-function set measures $cm(L_i, L_j)$ for all possible levels L_i in the given information and will be denoted by $CF = [f_{ij}]$. The co-function set can be defined in terms of mathematical morphological dilation [10] as:

$$CF(A_1, A_2, N) = [N(x, y) \oplus A_1] \cap A_2, \quad (3)$$

where $f_{ij} = \mathrm{cm}(L_i, L_j)$ and *N* is the structuring element. Thus we have that the co-function set of A_1 with respect to A_2 on a lattice with neighborhood structure *N* is equivalent to the dilation $N \oplus A_1$ followed by intersection with A_2.

2) Relationship of Co-function Matrix to Co-occurrence Matrix: The co-occurrence matrix of an image is an estimation of the second-order joint probability density of the intensity changes between pair of pixels, separated by a given distance at a certain orientation.

Let *I* be a gray scale image coded on *L* grey levels.

Let $p = (x, y)$ be the position of a pixel in *I* and

$d = (dx, dy)$ be a translation vector along θ direction.

Typically, the distance is one pixel and the orientation is quantized into four different orientations $(\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ)$. Therefore, four co-occurrence matrices are generated and they are summed to obtain a rotation invariant matrix. Each of co-occurrence matrix, denoted by $C[i, j, d, \ldots]$ is a $L \times L$ matrix whose (i, j) element is the number of pairs of pixels separated by the translation vector d that have the pair grey level (L_i, L_j).

Thus we have,

$$C[i, j, d, \ldots] = card\{(p, p + d) \in I \times I \mid I[p] = L_i \wedge I[p + d] = L_j\} \quad (4)$$

It can be seen that the co-occurrence matrix in (4) is just a particular case of our general co-function set defined in (2) and (3).

3) Relationship of Co-function Matrix to Gradient Orientations and MRF: We first introduce a Co-occurrence Matrix for Histograms of Oriented Gradients (CMHOG) which is a building block of pairs of gradient orientations. Since the pair of gradient orientations has more information than a single one. CMHOG can express shapes in more detail than HOG, which uses single gradient orientation. It can be seen that the CMHOG is just a particular type of our newly developed co-function matrix with the disk shape SE. In addition, it is worth to point out that the conventional energy function for MRF model [20] can be obtained by using the cliques as SEs. It is important to note that the co-function set is set-theoretic, or more precisely, a graph-theoretic concept and does not necessarily involve the actual coloring of the lattice. Although the previous example uses binary (black and white) colors to represent the subsets A_1 and A_2, these subsets can be defined by properties other than the gray levels. For instance, the set A_1 can be the gradients of pixels in the image and A_2 the complementary set of A_1.The co-function set of A_1 with respect to A_2 for a four-nearest neighbor *N* gives the HOGs of predefined cells or blocks.

4) Illustrations: We now give an illustrative example for the concepts, definitions and relations to conventional approaches which we have developed so far. This is an example of forming a co-function set using dilation. In Fig. 2(a), a 7x7 lattice, *S*, is shown with the lattice point values 0, 1, 2, 3, 4 and 5.

Let $A_1 = \{a_1 \mid a_1 \in S, a_1 = 0\}$,

$A_2 = \{a_2 \mid a_2 \in S, a_2 = 2\}$,

N be the SE (Fig. 2(b).

The dilation, $A_1 \oplus N$ is given in Fig. 2(c). Intersecting this set with A_2 gives the co-function set $CF(A_1, A_2, N)$ in Fig. 2(d).

B. Functional Markov Stationary Measures

The Markov Feature Vector (MFV) [15, 21] was recently proposed to characterize spatial co-occurrence of histogram patterns based on a sequence of embedded Markov Chain models, which has been shown to be generally superior over the coherence vector and autocorrelation histogram by incorporating both intra-bin and inter-bin co-occurrence information for visual representation. The calculation of MFV was done by using symmetric nature of co-occurrence matrix. In this section we will develop more general and robust similarity measure to be named as Functional Markov Stationary Measures (FMSM) based on the co-function set concepts described in Section III (A).

Let $CF = [f_{ij}]$ be the co-function matrix of size $K \times K$ for given image I. We then normalize the matrix to produce a stochastic matrix

$$SM = [s_{ij}] \tag{5}$$

where $s_{ij} = f_{ij} \Big/ \sum_j f_{ij}$.

It can be seen that the stationary distribution of the transition matrix be k-dimensional row vector, denoted as $\pi = (\pi_1, \pi_2, \ldots, \pi_k)$ satisfies

$$\begin{aligned} &\pi \times SM = \pi \qquad \text{or} \\ &\pi[SM - I] = 0, \end{aligned} \tag{6}$$

where I is a unit matrix whose diagonal elements are 1's and others are 0's. Then the functional stationary distribution will be derived by using basic theory of matrix and its determinant and will be defined as the Functional Markov Stationary Measures (FMSM). By direct substitution method, we can prove Theorem 3.1.

1) *Theorem 3.1.* Let c_{ij} be the absolute value of cofactor of determinant of matrix $SM - I$ in (6).

The distribution π, defined as

$$\pi_i = c_{ij} \Big/ \sum_{j=0}^{K} c_{ij} \,,$$

is a solution to the transition matrix *SM* defined in (6), namely, $\pi \times SM = \pi$.

3	1	1	1	3	2	2
0	0	2	2	2	2	5
0	0	3	3	2	1	1
1	1	2	2	3	4	4
0	0	0	1	1	0	4
1	2	3	3	2	1	2
0	0	1	2	1	5	1

(a)

(x, y-1)	
(x, y)	(x+1, y)

(b)

3	**1**	1	1	3	2	2
0	**0**	**2**	2	2	2	5
0	**0**	**3**	3	2	1	1
1	**1**	**2**	2	3	**4**	4
0	**0**	**0**	**1**	1	0	**4**
1	**2**	3	3	2	1	2
0	**0**	**1**	2	1	5	1

(c)

3	1	1	1	3	2	2
0	0	**2**	2	2	2	5
0	0	3	3	2	1	1
1	1	**2**	2	3	4	4
0	0	0	1	1	0	4
1	**2**	3	3	2	1	2
0	0	1	2	1	5	1

(d)

Figure 2. Example of forming co-function set by morphological dilation: (a) gray-level lattice, (b)Structuring Element (SE), (c) the dilation set of all 0's with SE, *N*, (d) the co-function set of all 0's with respect to the 2's.

This gives that π with $\pi_i = c_{ij} \Big/ \sum c_{ij}$ is the stationary distribution of co-function process.

In practice, the co-functional Markov Stationary Measure (FMSM), is defined as the combination of the initial distribution $\pi(0)$ and the stationary distribution π. Here the initial distribution cannot be ignored since a Markov chain is determined by both its transition matrix and by its initial distribution. By using the dominant color concepts the FMSM is also can be computed as shown in Fig. 3.

2) *Similarity Measures:* During synthesis, it is important to have an accurate measure to determine how close the output result matches the input. In our work, the similarity between two images is measured by the sum of the distances between their corresponding FMSM characteristics, where the distance of two matrices is the chi square distance of the two matrix vectors defined in basic statistics theory.

C. Concept-oriented Functional Markov Stationary Measures

We now propose the Concept-oriented Functional Markov Stationary Measures (CFMSM) to integrate the concept information with the spatial structure information of images. Using the same symbols as above, the proposed features obeys

$$\pi(N+1) = \alpha \times SM \times \pi(N) + (1-\alpha) I\, v \tag{7}$$

where v is the concept information vector, SM is the transition matrix and α is the parameter to leverage the influence of concept information and spatial structure information. After several iterations according to (7), the stationary distribution of u should satisfy.

$$\pi = \alpha \times SM \times \pi + (1-\alpha) v, \tag{8}$$

which has a convergent solution as follows:

$$\pi = (1-\alpha)(I - \alpha P)^{-1} v, \tag{9}$$

And we define the solution in (9) as the proposed CFMSM. In this way, the new representation of Markov stationary is not only affected by the spatial structure information but also by the concept information of images. Yet, the Markov stationary still keeps simplicity, compactness, and robustness.

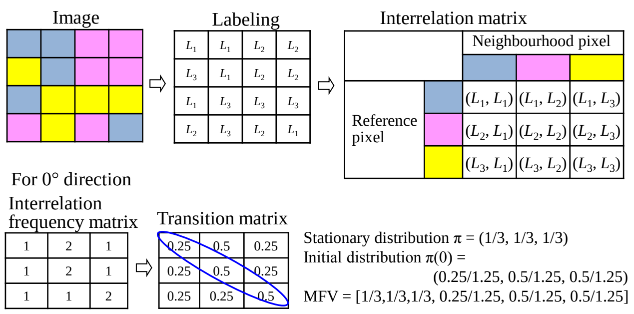

Figure 3. Computation of Markov stationary measures.

Equations (7) and (8) have the same formation to the correlated random walk. Markov stationary algorithm in [15] has been proven very effective in many applications. The addition of concept information also makes the Markov stationary more resistant to noise because FMSM is only relevant to the transition matrix calculated within one image, while the CFMSM incorporates the concept information as a prior knowledge besides the transition matrix within one image. The concept information is calculated using many images belonging to the same concept. The novelty of the proposed CFMSM lies in two aspects.

Firstly, the CFMSM extends the bag-of-features based image representation, and also owns the discriminative power as the local features which are resistant to occlusions and within-class shape variations. Secondly, it combines the concept information with the spatial structure information of images and goes one step further beyond FMSM.

IV. Experimental Results

We evaluated the performance of the proposed system by applying our method on a data set which contains 1000 color images and 50 images are chosen under structured sampling techniques. These randomly chosen images are used to serve as queries. After the classification step of the pixels, we keep the coordinates of pixels of the scanned areas. Afterwards, the dominant gradient orientations are extracted. For simplicity, four dominant orientations cases are performed. We then deduce the co-function matrices according to disk shape SE to compute the corresponding initial and stationary distributions. Finally these distributions are employed as the CFMSM features enabling to retrieve the most similar images. Various poses and different sizes are tested for similarity measures. Some extreme cases such as deformed and highly deteriorated images are used for target images. Some outputs are shown in Fig. 4.

Figure 4. Example of retrieved images (the first top 10 images).

Two statistical metrics were used to evaluate system performance, based on two measures frequently used in information search systems namely recall and precision. Recall indicates the proportion of relevant images retrieved from the database when answering a query. Precision is the proportion of the retrieved images that are relevant to the query. A high value of precision therefore denotes that the top-ranked images are relevant. The user starts retrieving by selecting the region of interest. The system returns a set of most similar images containing the selected queries even they are contained in a partial segments. To further show the performance of our approach, we have compared our approach with a global descriptors method. The evaluation is done for 50 image queries. In each experiment, one query image was randomly selected from the database and matched against the rest of the database. In our approach, the average precision rate of 80 % for a recall of 50% is achieved. The empirical results also confirm that the proposed system is invariant to rotation, translation and scaling.

V. Discussions and Conclusion

This paper has proposed an efficient and effective image retrieval framework based on automatically generated visual features analogous to its counterpart, text retrieval domain. Based on a new set-theoretic neighborhood based framework and shown its application to information similarity modeling, the basic tools: the “co-function set” which describes the relative presence of a set in the neighborhood of another set and the measure of this set, the “co-function measure” are designed and an embedded Markov chain is constructed on top of the co-occurrence matrix where a query expansion is performed by exploiting its topology preserving nature.

A sparse vector representation is also obtained by considering both local and global weight based on the frequency of occurrence of visual features in the individual image as well as in the collection. To reduce the feature dimensionality, Markov chain technique is applied on the sparse feature vector. Experimental results on a photographic image database demonstrated the efficiency and effectiveness of the proposed framework. Even though we are still in the process of exploring the set theoretic framework for information analysis and prediction, we feel that the work presented in this paper would contribute a new chapter to the image and related research fields.

Acknowledgment

The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

References

[1] A. Smeulder, M. Worring, S. Santini, A. Gupta, and R. Jain, “Content-based image retrieval at the end of the early Years,” IEEE Trans. on Pattern Anal. And Machine Intell., vol. 22, pp. 1349–1380, 2000.

[2] Y. Rui, T. S. Huang, and S. F. Chang, “Image retrieval: current techniques, promising directions and open issues,” Journal of Visual Communication and Image Representation, vol. l0 (4), pp. 39–62, 1999.

[3] L. Kotoulas and I. Andreadis, “Parallel local histogram comparison hardware architecture for content-based image retrieval,” Journal of Intelligent and Robotic Systems, vol. 39, no. 3, pp. 333–343, Mar. 2004.

[4] M. Flicker, et al., “Query by Image and Video Content: The QBIC System,” IEEE Computers, vol. 25, no. 9, pp. 23–32, Sep. 1995.

[5] J. R. Smith and S-Fu Chang, “VisualSEEK: A Fully Automated Content-Based Image Query System,” Proc. of ACM Multimedia 96, 1996.

[6] A. Pentland, R. W. Picard, and S. Sclaroff, “Photobook: Content-Based Manipulation of Image Databases,” International Journal of Computer Vision, 1996.

[7] D. E. Ilea and P. F. Whelan, “CTex-an adaptive unsupervised segmentation algorithm based on color-texture coherence,” IEEE Trans. on Image Processing, vol. 17, no. 10, pp. 1926–1939, 2008.

[8] S.-L. Lee, S.-J. Chun, D-H. Kim, J.-H. Lee and C.-W. Chung, “Similarity Search for Multidimensional Data Sequences,” In Proc. of IEEE 16th Intl. Conf. on Data Engineering (ICDE’00), San Diego, CA, USA, pp. 599–608, 2000.

[9] Y. Liu, “Automatic Range Image Registration in the Markov Chain,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 1, pp. 12–29, 2010.

[10] M. Pietikainen T. Ojala and T. Maenpaa, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, no. 7, pp. 971–987, 2002.

[11] D. Lowe, “Distinctive Image Features from Scale-invariant Keypoints,” Intl. J. of Computer Vision, vol. 60, no. 2, pp. 91–110, 2004.

[12] E. Hadjidemetriou, M.D. Grossberg, S.K. Nayar, “Multiresolution histograms and their use for recognition,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 26, no. 7, pp. 831–847, 2004.

[13] S. Birchfield and S. Rangarajan, “Spatiograms Versus Histograms for Region based Tracking,” In Proc. of IEEE Conf. on CVPR, San Diego, CA, USA, vol. 2, pp. 1158–1163, 2005.

[14] A. Vedaldi and S. Soatto, “Relaxed Matching Kernels for Robust Image Comparison,” In Proc. of IEEE Conf. on CVPR, Anchorage, AK, USA, pp. 1–8, 2008.

[15] Thi Thi Zin, Pyke Tin, T. Toriu and H. Hama, “Dominant Color Embedded Markov Chain Model for Object Image Retrieval,” In Proc. of the 5th Intl. Conf. on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP-2009), Kyoto, Japan, pp. 186–189, 2009.

[16] Thi Thi Zin and H. Hama, A Method Using Morphology and Histogram for Object-based Retrieval in Image and Video Databases, Int. J. of Computer Science and Network Security, vol. 7, no. 9, pp. 123–129, 2007.

[17] Thi Thi Zin, Pyke Tin and H. Hama, “Object Image Retrieval in Deformable Shapes Using Morphological Operations Based on Dominant Color Regions,” Digital Image Computing: Techniques and Applications, Canberra, Australia, pp. 548–555, 2008.

[18] X. Li, S. Chen, M. Shyu, S. Li, and B. Furht, “A Novel Hierarchical Approach to Image Retrieval Using Color and Spatial Information,” in Proc. IEEE Pacific Rim Conf. on Multimedia, pp.175–182, 2002.

[19] D. Xu, Y. Wang, and J. An, “Applying a new spatial color histogram in mean-shift based tracking algorithm,” in Proc. of the Image and Vision Computing New Zealand (IVCNZ05), University of Otago, Dunedin, New Zealand, 2005.

[20] S.Z. Li, Markov Random Field Modeling in Image Analysis (Advances in Pattern Recognition), Springer-Verlag, Third Edition, 362 pages, 2009.

[21] Thi Thi Zin, Pyke Tin, T. Toriu and H. Hama, “A Novel Way of Computing Multimedia Information Similarities,” ICIC Express Letters (Part B: Applications): An International Journal of Research and Surveys, vol. 1, no. 1, pp. 27–32, Sep. 2010.