Journal of Korea Multimedia Society Vol. 20, No. 10, October 2017(pp. 1619-1627)
https://doi.org/10.9717/kmms.2017.20.10.1619

Perceptual Bound-Based Asymmetric Image Hash Matching Method

Jiin Soo Seo[†]

ABSTRACT

Image hashing has been successfully applied for the problems associated with the protection of intellectual property, management of large database and indexation of content. For a reliable hashing system, improving hash matching accuracy is crucial. In order to improve the hash matching performance, we propose an asymmetric hash matching method using the psychovisual threshold, which is the maximum amount of distortion that still allows the human visual system to identity an image. A performance evaluation over sets of image distortions shows that the proposed asymmetric matching method effectively improves the hash matching performance as compared with the conventional Hamming distance.

Key words: Multimedia Fingerprinting, Image Identification, Image Hashing, Asymmetric Matching

1. INTRODUCTION

An image hash is a discriminative and robust summary of an image[1]. Since images often undergo various manipulations during distribution, including compression, enhancement, geometrical distortions, and analog-to-digital conversion, that may preserve the perceptual value, the cryptographic hash functions, which map arbitrary length data to a small and fixed number of bits[2], cannot be employed for image hashing. The aim of image hashing is to make the hashes of the perceptually similar images as close as possible in a metric space. The basic requirements of an image hashing system are robustness, pairwise independence, and database search efficiency[1,3].

The ideal hash function should provide an identifier for an image, even if presented in a different shape or form, by allowing for some modification, but only up to the amount that the human perceptual system cannot discern. As shown in Fig. 1, all image hashing methods implicitly assume that there exists a perceptual bound of each image, which should not be overlapped with that of other images. The perceptual bound is defined as the maximum amount of distortion that the image hash can allow while still providing identity verification. In Fig. 1, the hash h_1' of the distorted image x_1' is within both the human perceptual bound and the Hamming sphere (decision boundary of binary hash matching) of the hash h_1 of the original image x_1. On the other hand, the hash h_2' of the distorted image x_2' is within the perceptual bound but out of the Hamming sphere of the corresponding original hash h_2. This is due to the mismatch between the human perceptual bound and the arbitrary-chosen Hamming sphere which has been conventionally used for hash matching. One way

※ Corresponding Author : Jin Soo Seo, Addres: (25457) 7 Jukhun-gil, Gangneung, Gangwon-do, Korea, TEL : +82-33-640-2428, FAX : +82-33-646-0740, E-mail : jsseo@gwnu.ac.kr
Receipt date : Jul. 4, 2017, Revision date : Aug. 8, 2017
Approval date : Aug. 28, 2017

† Dept. of Electrical Eng., Gangneung-Wonju National University
※ This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Science, ICT & Future Planning(2012012876).

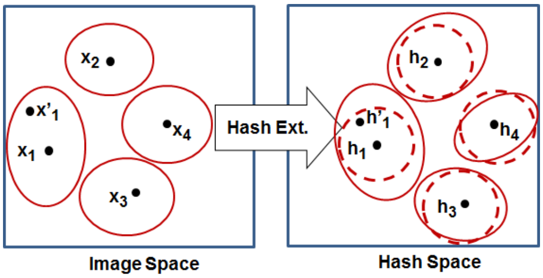

Fig. 1. Image hash and its perceptual bound. The solid ellipsoids around each image x_i and its hash hi represent the *perceptual bound*. The dotted circle around each hash represents the Hamming sphere conventionally used for the Hamming distance.

to overcome this discrepancy is designing the image hash function carefully so that the decision boundary in the hash space resembles the perceptual bound in the image space. However, constructing a perceptually relevant hash function is not a trivial task, since modeling human perception using a condensed representation is an intricate task. It is currently almost impossible, or at least not feasible, to include all the discriminant image information in a series of hash bits. Instead of designing a perceptually relevant hash function, we propose a hash matching method that incorporates the perceptual bound of the hash from the query image.

The perceptual bound of the hash is defined as the maximum amount of distortion that the hashing method can endure while allowing the identity to be determined. This paper assumes that the perceptual bound is proportional to the psychovisual threshold of the query image obtained by using human visual system (HVS) models[4-5]. The psychovisual threshold refers to the amount of pixel values that could be modified without causing a noticeable perceptual difference. As shown in Fig. 1, in most previous hash matching methods, the decision whether the two images are perceptually similar or not has been made by thresholding the Hamming distance between two binary hashes, for which the decision boundary is illustrated as a Hamming sphere (dashed line). By including the perceptual bound, which is represented by a solid circle, in the hash matching, the decision boundary for the hash matching can resemble human perception more closely. However, it is not feasible to store all the psychovisual thresholds of a large number of images in the hash database (DB) because of storage restrictions. This paper proposes a hash matching method that utilizes the psychovisual threshold of the query image in the hash matching. Instead of extracting the binary hash from the query image, we compute the probability that the intermediate hash vector of the query image belongs to each quantization bin by using the perceptual bound, which is referred to as soft quantization binning. We compute the probability that the binary hash in the DB is generated from the soft quantization binning of the query image. This probability value is used for deciding whether the query image corresponds to the binary hash in the DB. As opposed to conventional symmetric matching, which measures the Hamming distance between two binary hashes, the proposed matching method is called asymmetric[6]. Although there have been studies employing asymmetric matching in image retrieval[7-9] and multimedia hashing[6] to improve the matching performance, the proposed method incorporates the HVS model explicitly in deriving the asymmetric dissimilarity. To the best of the author's knowledge, this is the first attempt to use HVS in asymmetric matching. The performance of the proposed matching method was experimentally evaluated using thousands of images having various types of distortion, which verified that the proposed asymmetric matching method is effective for improving the hash matching performance as compared with the conventional Hamming distance.

2. PROPOSED ASYMMETRIC HASH MATCHING METHOD

Functional diagrams of the symmetric and the

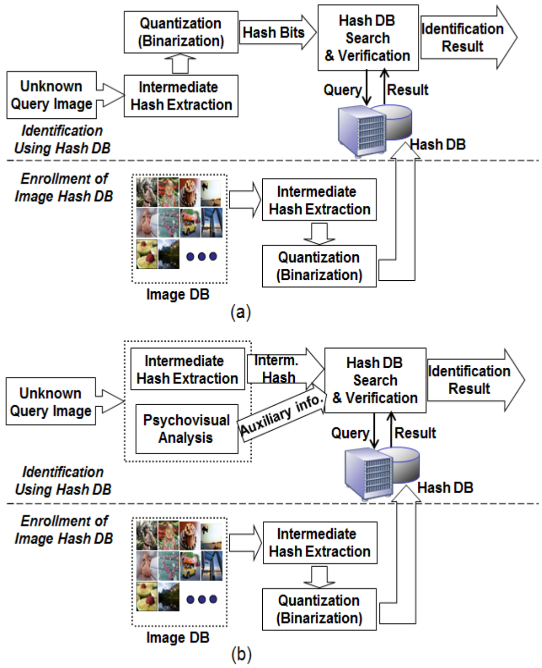

Fig. 2. Overview of the hash DB construction and matching. (a) Symmetric hash matching. (b) Proposed asymmetric hash matching based on the psychovisual analysis.

asymmetric hashing methods are shown in Fig. 2. To construct an image identification system, we first extract hashes from a number of images and store them in a hash DB. The hash extraction process comprises two steps: the intermediate hash extraction and the quantization, also referred to as binarization. In most cases, the intermediate hash of an image is given by a real-valued vector. The hash of an unknown query image is used for identification over the constructed hash DB. The candidates for the query hash are mostly obtained by the nearest neighbor DB search. To these candidates, the hash matching is applied for verification. As shown in Fig. 2(a), conventional symmetric matching compares the hash bits of the query image with the hash bits stored in the DB. It should be mentioned that, in the proposed asymmetric matching method shown in Fig. 2(b), the auxiliary information obtained by the HVS model is utilized along with the intermediate hash of the query in the hash matching. Various properties of the HVS have been extensively studied to find the image information that can be removed without degrading the subjective image quality, the results of which has been used for designing quantization tables in image coding and hiding a watermark signal in an image. The auxiliary information for the asymmetric matching considered in the proposed method is the psychovisual threshold obtained from the HVS. The psychovisual threshold is such that changes in the image below the threshold are not noticeable[4]. By using the psychovisual threshold of the query, we calculate the probability that the binary hash in the DB is generated from the query image. After normalization, the probability value is used in hash matching. The proposed method based on the HVS model is applied to the wavelet-based image hash[10].

2.1 Hash Extraction and Perceptual Bound Estimation

The image hash should be sufficiently concise, typically of an order of hundred bits per image, to allow the hashes of a large number of images to be stored in a DB. As shown in Fig. 2, we first extract an intermediate hash vector, which is mostly real-valued, from an image and then quantize it to a binary string. According to the characteristics of the intermediate hash vector, we need to choose an appropriate quantization method. Using the assumption that the components of the intermediate hash vector are sufficiently independent, each of its components is quantized independently using sign-based binarization [9-11]. Among many hashing methods [9-14], we chose the one based on the Haar wavelet [10]. The extension of the proposed matching method in Section 2.2 to other image hashing methods is quite straightforward. In [15], it was shown that the detail coefficients of the wavelet transform are symmetric around zero and can be modeled as i.i.d. generalized Gaussian random variables. Thus, quantizing wavelet detail coefficients to binary bits by taking their sign yields i.i.d. equiprobable bits, which can be used as a hash to represent an image [10]. According to this observation, we decompose an image up to five levels using the Haar wavelet and take the diagonal detail coefficients as an intermediate hash vector. The intermediate hash vector is quantized by retaining their signs to form a 256-bit hash for an image, as shown in Fig. 3.

The perceptual bound of the intermediate hash vector, which is defined as the maximum amount of distortion that the hashing method can endure, is used in the proposed asymmetric matching. In this paper, the perceptual bound is assumed to be proportional to the psychovisual threshold of the query image. Psychovisual threshold is maximal gray-scale pixel differences that are guaranteed to be imperceptible according to the spatial masking model. We use the spatial masking model in [4], which is also available as a matlab code [16]. Fig. 4 is the psychovisual threshold map of an image using the spatial masking model in [4], where the edges have much greater masking threshold than the near-constant regions of the image. In the case of linear transform methods, such as Haar wavelet [10] and random projection [11] hashing, the intermediate hash vector H_I of an image I can be represented as the matrix product given by $H_I = G^T I$ where the image I is assumed to be an N-dimensional vector for notational convenience and G denotes the transform-coefficient matrix (N by d). A distorted version of the image I is denoted by I'. We assume that the change caused by a distortion is bounded by the psychovisual threshold map PM(I) of the image I, which is also assumed to be a N-dimensional vector, as follows:

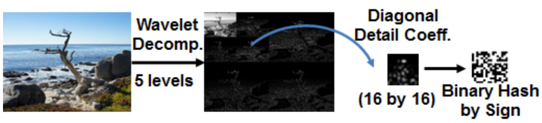

Fig. 3. Overview of the wavelet-based hash extraction.

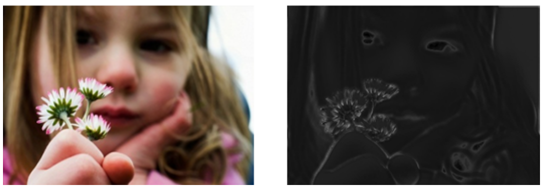

Fig. 4. Psychovisual threshold map of an image using the HVS model in [4].

$$|I' - I| \le \lambda \mathrm{PM}(I) \tag{1}$$

where λ is a positive constant. Then the change of the intermediate hash due to the distortion is given by

$$|H_{I'} - H_I| = |G^T I' - G^T I| \le |G^T||I' - I| \le \lambda |G^T| \mathrm{PM}(I) \tag{2}$$

where $|G|$ denotes the element-wise absolute value of G. The perceptual bound δ_H of the intermediate hash is defined as

$$\delta_H = \lambda |G^T| \mathrm{PM}(I) \tag{3}$$

As in the quality factor in image compression [5], the value of λ determines the distortion level permissible by the hash matching for identifying two hashes, where there is a tradeoff between the robustness and the pairwise independence.

2.2 Asymmetric Matching Based on Soft Quantization Binning Using Perceptual Bound

It is almost impossible, or at least not feasible, to include all the discriminant information of an image in a sequence of hash bits. The remaining discriminant information, which is not included in the hash bits and thus is discarded, is referred to as auxiliary information[6]. Because of storage and memory restrictions, we cannot accommodate all the auxiliary information of a large number of images in a hash DB. Instead of storing them, this paper only utilizes the auxiliary information obtained from the HVS of the query image in hash matching. As shown in Fig. 2, the real-valued intermediate hash vector is first extracted from an image and then quantized to obtain a binary hash. In the asymmetric matching method, instead of directly quantizing the intermediate hash vector of the query, we compute the perceptual bound of the intermediate hash vector, as shown in Fig. 3. The perceptual bound is defined as the amount of distortion that is permissible by the image hashing for identification of two hashes. Using the perceptual bound of the intermediate hash vector, we compute the probability that the intermediate hash vector belongs to each quantization bin, which is referred to as soft quantization binning[14]. From the soft quantization binning, we compute the probability that the binary hash in the DB is generated from the query image. The probability value is used for deciding whether the query image corresponds to the binary hash in the DB.

Let H_Q be the d-dimensional intermediate hash vector of the query image Q. Two-level quantization used is for the Haar-wavelet hashing [10]; The two quantization bins are denoted by R_0 and R_1 whose quantized values are 0 and 1 respectively. Conventionally, symmetric distances, such as Hamming or Euclidean, have been employed for hash matching. Instead of symmetric distances, we propose an asymmetric dissimilarity D_A between H_Q and a quantized hash C_d as one minus the ratio between their probabilities given by

$$D_A(H_Q, C_d) = 1 - \frac{P(H_Q{}^* \in R_d)}{P(H_Q{}^* \in R_q)} \tag{4}$$

where R_q and R_d are the quantization bins corresponding to H_Q and C_d, respectively. In practice, in most image hashing methods [10-14], a scalar quantizer is employed for the hash binarization under the assumption that the elements of the intermediate hash vector are statistically independent. For scalar quantizers, to derive a closed-form representation of D_A, which is important for fast DB verification, we assume that $H_Q{}^*[j]$, the j-th element of the intermediate hash, follows the triangular distribution given by

$$P_{(H_Q^*[j])}(x) = \begin{cases} \delta_Q[j]^{-2}(x - H_Q[j] + \delta_Q[j]) & \text{if } H_Q[j] - \delta_Q[j] \le x \le H_Q[j] \\ \delta_Q[j]^{-2}(H_Q[j] + \delta_Q[j] - x) & \text{if } H_Q[j] \le x \le H_Q[j] + \delta_Q[j] \\ 0 & \text{if } |x - H_Q[j]| \ge \delta_Q[j] \end{cases} \tag{5}$$

where $\delta_Q[j]$ is the perceptual bound for the j-th element of the intermediate hash. In practice, the integration in (4) can be computed easily by using one look-up table, parameterized by $\delta_Q[j]^{-1}(x - H_Q[j] + \delta_Q[j])$, of the cumulative distribution function of the triangular distribution in (5). Fig. 5 illustrates the computation of $P(H_Q{}^* \in R_0)$ and $P(H_Q{}^* \in R_1)$ under the triangular-distribution assumption, where the solid vertical lines represent the quantization boundary.

Conventionally, symmetric distances, such as Hamming or Euclidean, have been employed for hash matching. Instead of symmetric distances, we use the proposed asymmetric dissimilarity, which is given by the normalized probability in (4). With the intermediate hash vector of an image I denoted by H_I and a quantized hash C in DB, the hash matching problem can be formulated as the following hypothesis testing.

- L_0: A quantized hash C is generated from an image Q if the dissimilarity $D_A(H_Q, C)$is below a certain threshold T.
- L_1: A quantized hash C is not generated from an image Q if the dissimilarity $D_A(H_Q, C)$ is above a certain threshold T.

For the selection of threshold T, the false alarm rate P_{FA} and false rejection rate P_{FR} should be considered. The false alarm rate P_{FA} is the probability that different images are declared similar. The false rejection rate P_{FR} is the probability that the images from the same image are declared dissimilar. Both P_{FA} and P_{FR} are considered in

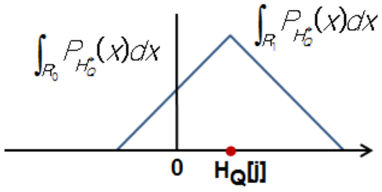

Fig. 5. Probability computation based on the triangular-distribution assumption over a scalar quantizer for the sign-based binarization [7].

evaluating the performance of the proposed matching method in Section 3.

3. EXPERIMENTAL RESULTS

The performance of the proposed asymmetric matching method was evaluated using the image hash DB generated from the first 5000 images of the MIR Flickr-25K image dataset, which include indoor and outdoor scenes, people, vehicles, sporting events, and paintings. After normalizing each test image by taking their luminance component and resizing them to 512 by 512, we extract the quantized and the binary hash by applying the wavelet-based hash extraction method described in Section 2.1. Both the Hamming and the proposed asymmetric distance were computed in hash matching on the hash DB for performance evaluation.

As mentioned in Section 2, the hash matching is formulated as a hypothesis testing in which there are typically two types of error: P_{FA} and P_{FR}. There is a tradeoff between the two probabilities in selecting the threshold T. For a fair comparison of the proposed asymmetric distance with the Hamming distance, which was originally employed for the binary hash, the detection error tradeoff (DET) curve, which plots the P_{FR} versus the P_{FA}, was used. The DET curve is obtained by measuring both error rates while varying the threshold T used in the hash matching. P_{FA} was calculated between all possible pairs of test images in the hash DB. To calculate P_{FR}, we tested the proposed method against the following image processing steps.

- WN: Adding Gaussian white noise with a mean of zero and a standard deviation of 50.
- FT1: Applying the 5 by 5 median filter.
- FT2: Applying the 3 by 3 average filter.
- FT3: Applying the 3 by 3 unsharp contrast enhancement filter H given by

$$H=\begin{bmatrix}-0.1667 & -0.6667 & -0.1667\\ -0.6667 & 4.3333 & -0.6667\\ -0.1667 & -0.6667 & -0.1667\end{bmatrix}$$

- HQ: Histogram Equalization.
- DT: Dithering with MATLAB default parameters.
- FMLR: Applying the frequency mode Laplacian removal (FMLR) [17,18].
- JPG1: JPEG compression of the quality factor 5%.
- JPG2: JPEG compression of the quality factor 75%.

Each test image was subjected sequentially to a set of the selected distortions. We considered six different sets of distortion. By computing the distance between the hashes from the original and the corresponding distorted image using the threshold, P_{FR} was obtained. The resulting DET curve of the considered wavelet-based hashing method is shown in Fig. 6 for the Hamming distance and the proposed asymmetric matching method with different values of λ. The value of λ determines the perceptual bound, that is the amount of distortion allowed by the hashing system, and should be arranged depending on the application scenario. For example, if the hashing is for authentication purposes, λ should be small. For identifying a scanned version of an image, λ should be set at a larger value. However, we found that with a wide range of the λ values the proposed asymmetric dissimilarity D_A performed better than the conventional Hamming distance for all the considered distortions. In particular, the proposed dissimilarity performed more effectively than the Hamming distance in the low P_{FA} region of the DET curve, where most hashing systems operate in practice. Although a noticeable performance gain was observed in the DET curves when the proposed matching method was used, the performance gain was dependent on the type of image hashing methods and the type and amount of distortion. The interplay between the amount of distortion and the

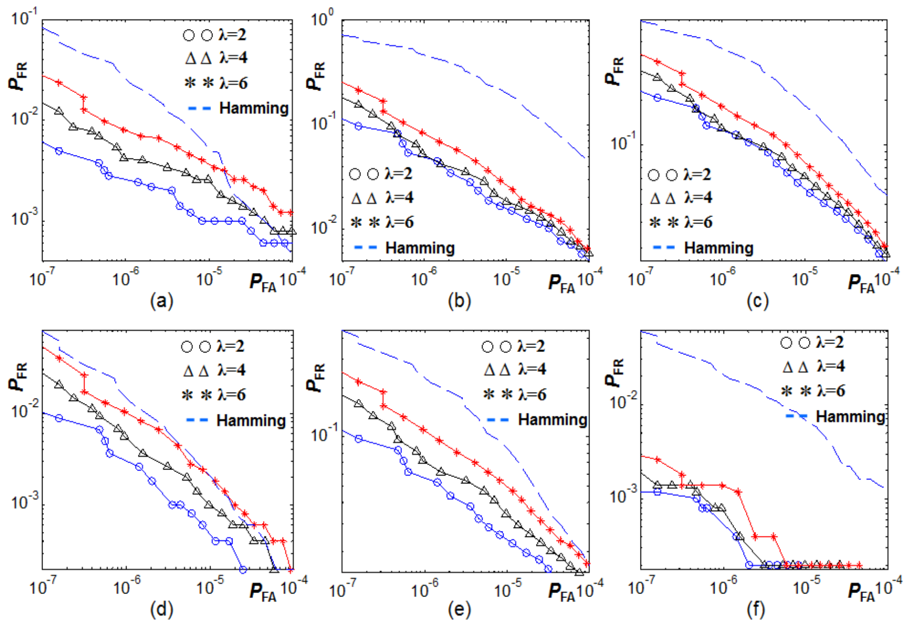

Fig. 6. DET curves of the Haar-wavelet hashing for six sets of distortions. (a) FT1+FT2+FT3+JPG2. (b) JPG1. (c) WN+JPG2. (d) HQ+JPG2. (e) DT+JPG2. (f) FMLR+JPG2.

matching parameter (such as λ and T) need to be addressed. A promising theoretical result is reported at [19] for an additive noise. However, it is difficult to analyze practical situations since there are plenty of image processing steps of those we do not know the exact characteristics.

4. CONCLUSION

In this study, we focused on hash matching and proposed an asymmetric dissimilarity based on the perceptual bound of a query image to improve hash matching performance. By utilizing the perceptual bound of the query image in the hash matching, the proposed matching method is asymmetric, and thus, it does not increase the size of the hash DB and can be applied to the already constructed hash DB without any update or modification. Statistical modeling using the perceptual bound allows us to compute the probability that the quantized hash in the DB was generated from the query image. The normalized probability was used as an asymmetric dissimilarity measure in this study. By incorporating the HVS model explicitly in the derivation of an asymmetric dissimilarity measure, the proposed method is sufficiently general to be utilized for a number of different hashing methods. In this study, we applied the proposed method to wavelet-based image hashing and tested its performance over thousands of images with six different sets of image distortion. The proposed asymmetric dissimilarity was conducive to reducing matching errors as compared with the conventional Hamming distance. This implies that there may be a room for improvement of hashing performance by using human perceptual model. Future work includes a more refined pdf modeling of the perceptual bound and an extension of the proposed method to audio or video hashing.

REFERENCE

[1] T. Kalker, J. Haitsma, and J. Oostveen, "Issues with Digital Watermarking and Perceptual Hashing," *Proceeding of Society of Photo-Optical Instrumentation Engineers, Multi-*

media Systems and Applications IV, pp. 189, 2001.

[2] A. Menezes, P. Oorshot, and S. Vanstone, *Handbook of Applied Cryptography*, Chemical Rubber Company Press, Boca Raton, Florida, 1997.

[3] M. Kim, M. Park, and H. Kim, "Automatic Music Summarization Method by Using the Bit Error Rate of the Audio Fingerprint and a System Thereof," *Journal of Korea Multimedia Society*, Vol. 16, No. 4, pp. 453-463, 2013.

[4] B. Girod, "The Information Theoretical Significance of Spatial and Temporal Masking in Video Signals," *Proceeding of Society of Photo-Optical Instrumentation Engineers, Human Vision, Visual Processing, and Digital Display*, pp. 178-187, 1989.

[5] N. Jayant, J. Johnston, and R. Safranek, "Signal Compression Based on Models of Human Perception," *Proceedings of the IEEE*, Vol. 81, No. 10, pp. 1385-1422, 1993.

[6] J. Seo, "An Asymmetric Matching Method for a Robust Binary Audio Fingerprinting," *IEEE Signal Processing Letters*, Vol. 21, No. 7, pp. 844-847, 2014.

[7] W. Dong, M. Charikar, and K. Li, "Asymmetric Distance Estimation with Sketches for Similarity Search in High-dimensional Spaces," *Proceeding of ACM* SIGIR *Conference on Research and Development in Information Retrieval*, pp. 123-130, 2008.

[8] H. Jegou, M. Douze, and C. Schmid, "Product Quantization for Nearest Neighbor Search," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 33, No. 1, pp. 117-128, 2011.

[9] A. Gordo, F. Perronnin, Y. Gong, and S. Lazebnik, "Asymmetric Distances for Binary Embeddings," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 36, No. 1, pp. 33-47, 2014.

[10] A. Varna, A. Swaminathan, and M. Wu, "A Decision Theoretic Framework for Analyzing Binary Hash-based Content Identification Systems," *Proceeding of ACM Workshop on Digital Rights Management*, pp. 67-76, 2008.

[11] J. Fridrich, "Robust Bit Extraction from Images," *Proceeding of IEEE Conference on Multimedia Computing and Systems*, pp. 536-540, 1999.

[12] R. Venkatesan, S. Koon, M.H. Jakubowski, and P. Moulin, "Robust Image Hashing," *Proceeding of IEEE Conference on Image Processing*, pp. 664-666, 2000.

[13] X. Lv and Z. Wang, "Perceptual Image Hashing Based on Shape Contexts and Local Feature Points," *IEEE Transactions on Information Forensics and Security*, Vol. 7, No. 3, pp. 1081-1093, 2012.

[14] J. Seo and M. Jo, "A Probabilistic Dissimilarity Matching for the DFT-domain Image Hashing," *International Journal of Advanced Culture Technology*, Vol. 5, No. 1, pp. 76-82, 2017.

[15] S. Mallat, "A Theory for Multiresolution Signal Decomposition: the Wavelet Representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 11, No. 7, pp. 674-693, 1989.

[16] J. Fridrich and M. Goljan, Spatial Image Masking Code, http://www.ws.binghamton.edu/fridrich/ masking.html (accessed Oct., 23, 2017).

[17] R. Barnett and D. Pearson, "Frequency Mode LR Attack Operator for Digitally Watermarked Images," *Electronics Letters*, Vol. 34, No. 19, pp. 1837-1839, 1998.

[18] S. Pereira, S. Voloshynovskiy, M. Madueno, S. Marchand-Maillet, and T. Pun, "Second Generation Benchmarking and Application Oriented Evaluation," *Proceeding of Information Hiding Workshop*, pp. 340-353, 2001.

[19] S. Voloshynovskiy, T. Holotyak, O. Koval, F.

Beekhof, and F. Farhadzadeh, "Sign-magnitude Decomposition of Mutual Information with Polarization Effect in Digital Identification," *Proceeding of IEEE Information Theory Workshop*, pp. 385-389, 2011.

Jiin Soo Seo

He received the B.S., M.S., and Ph.D. degrees from Korea Advanced Institute of Science and Technology in 1998, 2000, and 2005 respectively, all in electrical engineering. He was a senior researcher at ETRI from 2006 to 2008. He joined the Department of Electrical Engineering at Gangneung-Wonju National University in 2008. His research interests are audio and image processing, multimedia retrieval, and pattern recognition.