

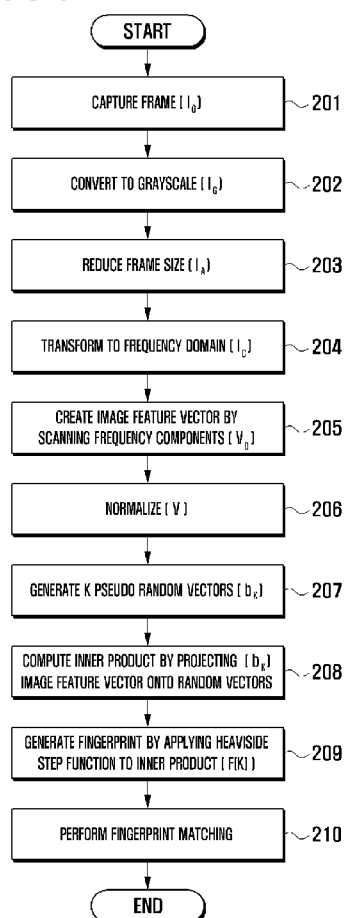


- (51) **International Patent Classification:**
H04N 1/387 (2006.01) *G06T 1/00* (2006.01)
- (21) **International Application Number:** PCT/KR2012/004690
- (22) **International Filing Date:** 14 June 2012 (14.06.2012)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
10-2011-0057628 14 June 2011 (14.06.2011) KR
- (71) **Applicant (for all designated States except US):** **SAM-SUNG ELECTRONICS CO., LTD.** [KR/KR]; 129, Sam-sung-ro, Yeongtong-gu, Suwon-si, Gyeonggi-do 443-742 (KR).
- (72) **Inventors; and**
- (75) **Inventors/Applicants (for US only):** **CHOI, Yoon Hee** [KR/KR]; No. 101-602, Maetan Weve Haneulchae Apt., Maetan-dong, Yeongtong-gu, Suwon-si, Gyeonggi-do 443-370 (KR). **PARK, Hee Seon** [KR/KR]; No. 3-902, Jamwon Hanshin Apt., Jamwon-dong, Seocho-gu, Seoul 137-030 (KR).
- (74) **Agent:** **YOON, Dong Yol;** Yoon & Lee International Patent & Law Firm, 3rd Fl, Ace Highend Tower-S, 226, Gasan Digital 1-ro, Geumcheon-gu, Seoul 153-803 (KR).
- (81) **Designated States (unless otherwise indicated, for every kind of national protection available):** AE, AG, AL, AM,

[Continued on next page]

(54) **Title:** IMAGE PROCESSING METHOD AND APPARATUS

[Fig. 2]



(57) **Abstract:** An image processing method and apparatus extracts unique identifiers directly from images and examines similarities between images using the extracted identifiers, by capturing a frame of an image; reducing the size of the captured frame; transforming the reduced frame to a frequency domain frame; creating an image feature vector by scanning frequency components of the frequency domain frame; computing inner product values by projecting the image feature vector onto random vectors; generating a fingerprint for identifying the captured frame by applying a Heaviside step function to the inner product values; and searching a database for information related to the generated fingerprint and outputting the search results.





AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH,

GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report (Rule 48.2(g))

Description

Title of Invention: IMAGE PROCESSING METHOD AND APPARATUS

Technical Field

- [1] The present invention relates generally to image processing and, more particularly, to an image processing method and apparatus that can extract unique identifiers or fingerprints directly from images and examine similarities between images using the extracted identifiers.

Background Art

- [2] With increased usage of multimedia in recent years, there has been a rise in demand for techniques for multimedia data retrieval and recognition. In examining the similarity between multimedia items, comparing multimedia items in binary form may be impractical since even minor image processing operations may significantly change binary values of the multimedia items. Alternatively, various identifiers may be used to compare multimedia items. Such unique identifiers are referred to as fingerprints, also known as signatures or hash, and several video recognition methods based on various types of fingerprints have been implemented.
- [3] Audio fingerprints have been used in some video recognition methods. However, this method may be unsuitable to silent portions of a video and may take a relatively long time to identify the exact location in time of the audio fingerprint.
- [4] Image fingerprints have been used in video recognition methods as well. In such a method, a frame is captured from a video and a fingerprint is extracted from the captured frame. However, the fingerprint may be ineffective for image matching, where the fingerprint is extracted using color properties of the frame and the color properties of the corresponding frame are changed after image processing. As in existing methods based on image fingerprints, when fingerprints are represented as vectors and the distance between the fingerprint vectors is used for video matching, retrieval efficiency may be lowered in large multidimensional databases.
- [5]

Disclosure of Invention

Technical Problem

- [6] Accordingly, the present invention has been made to solve the above problems occurring in the prior art and the present invention provides an image processing method and apparatus that enable extraction of a fingerprint that is highly resistant to image processing operations and fast retrieval of information matching the fingerprint from a database.

Solution to Problem

- [7] In accordance with an aspect of the present invention, there is provided a method for image processing, including capturing a frame of an image; reducing the size of the captured frame; transforming the reduced frame to a frequency domain frame; creating an image feature vector by scanning frequency components of the frequency domain frame; computing inner product values by projecting the image feature vector onto random vectors; generating a fingerprint for identifying the captured frame by applying a Heaviside step function to the inner product values; and searching a database for information related to the generated fingerprint and outputting the search results.
- [8] In accordance with another aspect of the present invention, there is provided an apparatus for image processing, including a frame capturer capturing a frame of an image; a fingerprint extractor extracting a fingerprint from the captured frame; and a fingerprint matcher searching a database for information related to the fingerprint, wherein the fingerprint extractor reduces the size of the captured frame, transforms the reduced frame to a frequency domain frame, creates an image feature vector by scanning frequency components of the frequency domain frame, computes inner product values by projecting the image feature vector onto random vectors, and generates the fingerprint by applying a Heaviside step function to the inner product values.

Advantageous Effects of Invention

- [9] In a feature of the present invention, it is possible to extract an image fingerprint that is highly resistant to image processing operations and to retrieve information matching the fingerprint from a database in a fast and accurate way.

Brief Description of Drawings

- [10] The above and other aspects, features and advantages of the present invention will be more apparent from the following detailed description taken in conjunction with the accompanying drawings, in which:
- [11] FIG. 1 is a block diagram of an image processing apparatus according to an embodiment of the present invention;
- [12] FIG. 2 is a flowchart of an image processing method according to another embodiment of the present invention;
- [13] FIG. 3 is a diagram illustrating image processing operations in the method of FIG. 2;
- [14] FIG. 4 is a diagram illustrating methods for reducing the image size in the method of FIG. 2;
- [15] FIG. 5 is a diagram illustrating the plots of normalized average matching scores with respect to the compression ratio when original images and their JPEG compressed images are compared;

[16] FIG. 6 is a diagram illustrating the plots of normalized average matching scores with respect to the noise variance when original images and their corrupted images with Gaussian noise are compared; and

[17] FIG. 7 is a diagram illustrating a distribution of the bit error rate for the method obtained from applying the JPEG compression and Gaussian noise.

Mode for the Invention

[18] Hereinafter, exemplary embodiments of the present invention are described in detail with reference to the accompanying drawings. The same reference symbols are used throughout the drawings to refer to the same or like parts. Detailed descriptions of well-known functions and structures incorporated herein may be omitted to avoid obscuring the subject matter of the present invention. Particular terms may be defined to describe the invention in the best manner. Accordingly, the meaning of specific terms or words used in the specification and the claims should not be limited to the literal or commonly employed sense, but should be construed in accordance with the spirit of the invention. The description of the various embodiments is to be construed as exemplary only and does not describe every possible instance of the invention. Therefore, it should be understood that various changes may be made and equivalents may be substituted for elements of the invention.

[19] The image processing apparatus of the present invention is a device having a wired or wireless communication module, and may be any information and communication appliance such as a personal computer, laptop computer, desktop computer, MP3 player, portable multimedia player (PMP), personal digital assistant (PDA), tablet computer, mobile phone, smart phone, smart TV, Internet Protocol TV (IPTV), set-top box, cloud server, or portal site server. The image processing apparatus may include a fingerprint extractor that extracts a fingerprint from an image received from a database server, smart phone, or IPTV. Here, the fingerprint is an identifier specific to an image and is also known as a signature or hash. The image processing apparatus may retrieve images or supplementary information (like an electronic program guide) related to the extracted fingerprint from an image database server. The image processing apparatus may further include a fingerprint matcher that examines similarity between fingerprints and outputs the examination result. The image processing apparatus may display retrieval results and similarity examination results or provide them to an external device. In the description, the image processing apparatus is assumed to act as a server that examines similarity between images.

[20] FIG. 1 is a block diagram of an image processing apparatus 100 according to an embodiment of the present invention.

[21] Referring to FIG. 1, the image processing apparatus 100 may include a first frame

capturer 110, a second frame capturer 120, a fingerprint extractor 130, a fingerprint matcher 140, an image database 150, and a fingerprint database 160.

- [22] The first frame capturer 110 captures a frame of an image to be recognized, which is output from a digital broadcast receiver, IPTV, smart phone, or laptop computer. The second frame capturer 120 captures a frame of a reference image, which is output from a digital broadcast receiver, IPTV, smart phone, or laptop computer. The fingerprint extractor 130 extracts a fingerprint from the frame captured by the first frame capturer 110 and forwards the extracted fingerprint to the fingerprint matcher 140. The fingerprint extractor 130 extracts a fingerprint from the frame captured by the second frame capturer 120 and stores the extracted fingerprint together with reference image information (for example, film information or broadcast channel information) in the fingerprint database 160. The fingerprint extractor 130 may also extract a fingerprint from an image retrieved from the image database 150 and store the extracted fingerprint in the fingerprint database 160. The fingerprint matcher 140 examines similarity between the fingerprint of an image to be recognized and the fingerprint of a reference image. In other words, the fingerprint matcher 140 searches the fingerprint database 160 for image information related to the fingerprint of an image to be recognized. Next, the present invention is described further with focus on the fingerprint extractor 130 and the fingerprint matcher 140 in connection with FIGS. 2 to 7.
- [23] FIG. 2 is a flowchart of an image processing method according to another embodiment of the present invention, and FIG. 3 illustrates image processing operations in the method of FIG. 2.
- [24] Referring to FIG. 2, the frame capturer 110 or 120 captures at least one frame (IO, as indicated by (a) of FIG. 3) from a received image and forwards the captured frame to the fingerprint extractor 130 (201). Here, when the received image is interlace-scanned, the frame capturer 110 or 120 may capture an odd field picture and even field picture from the received image and forward the odd and even field pictures to the fingerprint extractor 130, which then may extract one fingerprint from each field picture. The fingerprint extractor 130 converts the captured frame into a grayscale frame (IG, as indicated by (b) of FIG. 3) (202). Here, step 202 may be skipped. The fingerprint extractor 130 shrinks the captured frame or grayscale frame into a small average image (IA, as indicated by (c) of FIG. 3) of width M and height N (203). Image shrinking is described in detail with reference to FIG. 4.
- [25] FIG. 4 illustrates schemes for image shrinking in the method of FIG. 2.
- [26] As shown in FIG. 4, the fingerprint extractor 130 subdivides the frame into multiple areas. For example, the frame may be subdivided into rows and columns as indicated by (a) of FIG. 4, be subdivided into rows as indicated by (b) of FIG. 4, or be subdivided into oval shapes as indicated by (c) of FIG. 4. The frame may be

subdivided in other ways. Thereafter, the fingerprint extractor 130 selects M*N areas from among the multiple areas. Here, the fingerprint extractor 130 excludes an area in which a caption, logo, advertisement or broadcast channel indicator is to be located in area selection. Finally, the fingerprint extractor 130 computes average values of the individual selected areas. The average values $I_A(i,j)$ can be defined by Equation 1.

[27] [Equation 1]

[28]

$$I_A(i,j) = \frac{1}{|P_k|} \sum_{p \in P_k} I_G(p),$$

[29] *where $k = i \times M + j$, where $i = 0, 1, \dots, M$ and $j = 0, 1, \dots, N$*

[30] Here, $|P_k|$ denotes the number of pixels in the k-th area and $I_G(p)$ denotes

the pixel value at a point p.

[31] Referring back to FIG. 2, the fingerprint extractor 130 transforms the small average image (i.e. shrunk frame $I_A(i,j)$) to a frequency domain frame (IC)(204).

[32] Here, DCT (Discrete Cosine Transform), DFT (Discrete Fourier Transform) or DWT (Discrete Wavelet Transform) may be applied. As DCT is normally used for video coding, usage of two dimensional DCT (2D-DCT) is assumed in the following description.

[33] The fingerprint extractor 130 scans frequency components (coefficients) of the 2D-DCT transformed frame ($I_C = 2DCT(I_A)$), as indicated by (d) of FIG. 3) to create an image feature vector ($V_O = Scan(I_C, L)$) for the captured frame I_0 (205). Here, L denotes the dimensions of the image feature vector (i.e., the number of frequency components). The fingerprint extractor 130 need not scan all the frequency components in IC. For example, as indicated by (e) of FIG. 3, the DC (direct current) component and high-frequency components exceeding a preset threshold value are excluded and only low-frequency components are scanned in a zigzag fashion. This is because the DC component is too sensitive to brightness and high-frequency components exceeding the threshold value may cause signal processing distortion. In other words, low-frequency components not exceeding the threshold value are resistive to various signal processing operations and are not easily distorted. Here, the threshold value may be set by the user. For example, when I_C has 8×8 (=64) entries, the fingerprint extractor 130 may scan 48 frequency components excluding the DC component and high-frequency components to create an image feature vector of 48 dimensions ($V_O = ZigzagScan(I_C, L)$)

[34] The fingerprint extractor 130 normalizes the image feature vector V_o , as indicated by (f) of FIG. 3, so that the mean of V_o becomes 0 and the variance thereof becomes 1 (206). Here, step 206 may be skipped. Normalization may be performed using Equation 2.

[35] [Equation 2]

[36]

$$V = \frac{V_o - \mu_{V_o}}{\sigma_{V_o}}$$

[37] , where μ_{V_o} indicates the mean of $\{V_{o_1}, V_{o_2}, \dots, V_{o_L}\}$ and σ_{V_o}

indicates the standard deviation of $\{V_{o_1}, V_{o_2}, \dots, V_{o_L}\}$.

[38] The fingerprint extractor 130 generates a random vector matrix B having K (for example, 48) random vectors as column vectors (207). Here, the K random vectors may follow a Gaussian distribution with mean of 0 and variance of 1 as indicated by (g) of FIG. 3. The k-th random vector may be obtained using Equation 3.

[39] [Equation 3]

[40]

$$b_k = \text{Rand}(S_k, L)$$

where $k = 0, 1, \dots, K-1$

[41] Here, S_k indicates a seed value and L indicates the dimensions of the pseudo random vector.

[42] The fingerprint extractor 130 computes the inner product value of the normalized image feature vector V and the pseudo random vector b_k by projecting V onto b_k (208). Here, inner product computation is performed once for each random vector, resulting in K inner product values. Projection of the normalized image feature vector V onto random vectors b_1, b_2, b_3 is geometrically illustrated by (h) of FIG. 3.

[43] The fingerprint extractor 130 obtains a fingerprint f for recognizing the captured frame I_o by applying a Heaviside step function to the inner product ($f = F(k)$) (209). Steps 208 and 209 may be represented by Equation 4.

[44] [Equation 4]

[45]

$$f = H(B^T V), \text{ where } H(B^T V) \text{ is a Heaviside step function}$$

[46] Specifically, the Heaviside step function may be defined by Equation 5.

[47] [Equation 5]

[48]

$$H(x) = \begin{cases} 1, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

[49] That is, a Heaviside step function is a function that produces 0 for negative arguments and produces 1 for non-negative arguments. As the Heaviside step function is applied to K inner product values, the obtained fingerprint f is a K-bit binary value. When the captured frame I₀ is a frame of a reference image, the fingerprint extractor 130 stores the obtained fingerprint in the fingerprint database 160. When the captured frame I₀ is a frame of an image to be recognized, the fingerprint extractor 130 forwards the obtained fingerprint to the fingerprint matcher 140.

[50] At step 209, the fingerprint extractor 130 may generate multiple fingerprints for a single frame using Equation 6.

[51] [Equation 6]

[52]
$$f_s = H(B_s^T V), \text{ where } s = 0, 1, \dots, S-1$$

[53] Here, f_s denotes the s-th fingerprint of the frame.

[54] The fingerprint matcher 140 performs fingerprint matching between fingerprints and outputs the matching results (210). The normalized Hamming distance d_H is calculated using Equation 7.

[55] [Equation 7]

[56]
$$d_H(f^q, f^d) = \frac{1}{K} \sum_k |f^q(k) - f^d(k)|$$
, where f^q is a fingerprint for an

image to be recognized and f^d is a fingerprint for an image stored in the database.

[57] After calculation of the Hamming distance between two fingerprints, the fingerprint matcher 140 determines that the two images related respectively to the two fingerprints are different when the Hamming distance is greater than a preset threshold value, and determines that the two images are similar when the Hamming distance is less than or equal to the threshold value. Then, the fingerprint matcher 140 outputs the determination result. For example, assume that f^q is 1111001111₍₂₎, f^d is 1111001110₍₂₎, and the threshold value is 1. As the Hamming distance between the two fingerprints is 1, the fingerprint matcher 140 determines that the two images related respectively to the two fingerprints are the same. As image matching using the Hamming distance (i.e. Equation 7) involves multiple bitwise comparisons, the search time may be long when the fingerprint database is large.

[58] The fingerprint matcher 140 may use a generated integer fingerprint as a key together with indexing techniques employed by existing databases to perform an efficient

search. The fingerprint matcher 140 may perform a constant-time search through direct access to the memory using an integer fingerprint. When S fingerprints are extracted from a single image or video frame as described before, the fingerprint matcher 140 may perform image matching for each fingerprint and combine the matching results. For example, the fingerprint matcher 140 may return as a result an image that has been most frequently matched with the S fingerprints. When the threshold value for matching is set to 1 (bit), the fingerprint matcher 140 may newly generate K fingerprints by modifying one bit of a given fingerprint and perform additional matching using the newly generated fingerprints.

[59] FIGS. 5 and 6 show results of experiments performed using the method of the present invention.

[60] Specifically, FIG. 5 plots normalized average matching scores with respect to the compression ratio when original images and their JPEG compressed images are compared. FIG. 6 plots normalized average matching scores with respect to the noise variance when original images and their corrupted images with Gaussian noise are compared. The experiment was performed using 5000 images that differ in category and size. Hence, the average matching scores are mean values for 5000 images. As indicated by FIGS. 5 and 6, the method of the present invention (labeled “Gaussian Projection”) exhibits the best performance. It is possible to recognize an advertisement currently displayed on the TV screen in real time using the method of the present invention. Based on such matching information, even when the TV screen is used as a monitor of a set-top box, contents of a TV broadcast may be recognized in real time and hence supplementary information or advertisement related to the contents of the TV broadcast may be provided to the viewer.

[61] FIG. 7 illustrates a distribution of the bit error rate obtained from experiments using JPEG compression and Gaussian noise in the case of the method of the present invention.

[62] Referring to FIG. 7, the probability of no bit error is about 90.27. The probability of single bit error is about 7.48, and corresponds to 76.88 percent of the overall probability of bit error. As the probability of one bit error takes a major portion of the overall error probability, when single bit errors are permitted, an accuracy level of 97.75 percent is expected. Additionally, in the experiment using Gaussian noise, the probability of no bit error was about 63.35 and the probability of single bit error was about 25.84. The probability of single bit error corresponds to 70.50 percent of the overall error probability. Hence, when single bit errors are permitted, an accuracy level of 89.19 percent is expected. Since such bit error rates are results of application of incentive image processing operations, significantly lower bit error rates are expected in most image processing applications.

[63] On the basis of the above experimental results, the fingerprint matcher 140 may search the database using a fingerprint obtained by modifying one bit of the original fingerprint. For example, when the original fingerprint is 48 bits, 48 variant fingerprints may be obtained by modifying one bit of the original fingerprint. Hence, when a search using the original fingerprint fails, the fingerprint matcher 140 may perform an additional search using a variant fingerprint.

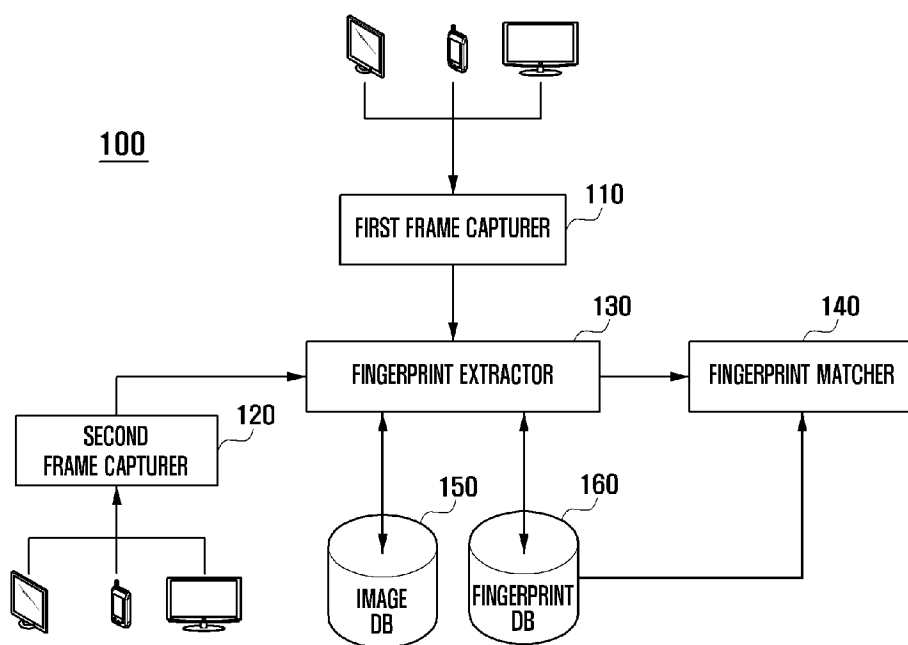
[64] Although exemplary embodiments of the present invention have been described in detail hereinabove, it should be understood that many variations and modifications of the basic inventive concept described herein will still fall within the spirit and scope of the present invention as defined in the appended claims.

Claims

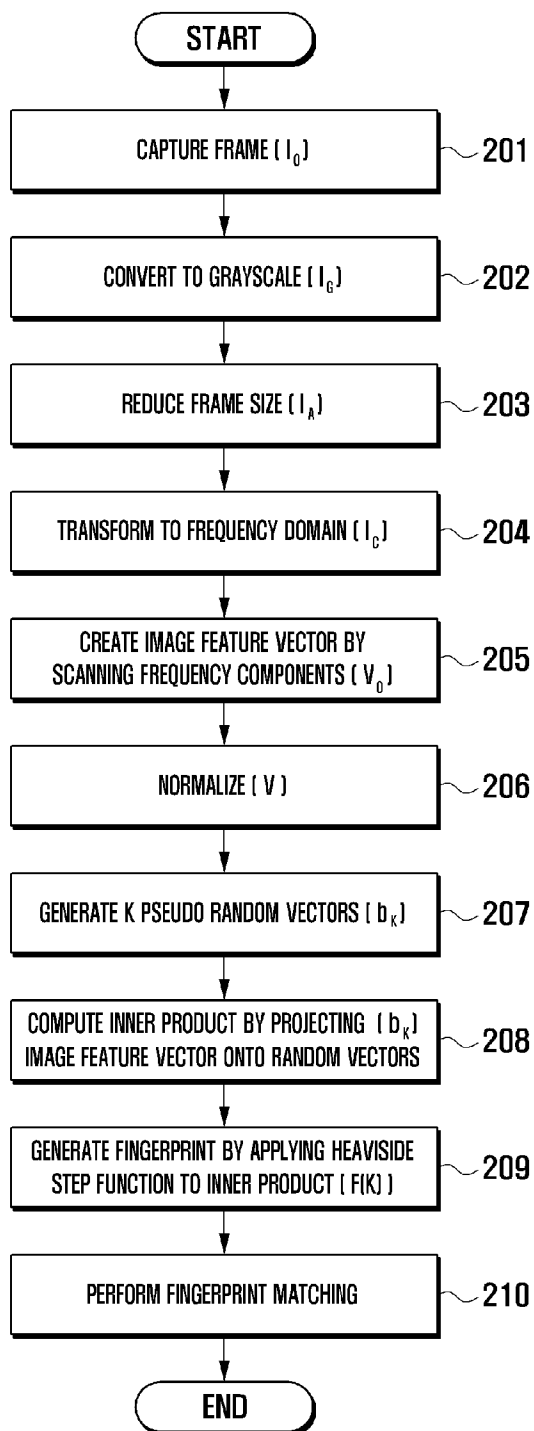
- [Claim 1] A method for image processing, comprising:
capturing a frame of an image;
reducing the size of the captured frame;
transforming the reduced frame to a frequency domain frame;
creating an image feature vector by scanning frequency components of the frequency domain frame;
computing inner product values by projecting the image feature vector onto random vectors;
generating a fingerprint for identifying the captured frame by applying a Heaviside step function to the inner product values; and
searching a database for information related to the generated fingerprint and outputting the search results.
- [Claim 2] The method of claim 1, wherein creating an image feature vector comprises scanning low-frequency components of the frequency domain frame except for a Direct Current (DC) component of the frequency domain frame and high-frequency components of the frequency domain frame exceeding a preset threshold value.
- [Claim 3] The method of claim 2, wherein frequency components of the frequency domain frame are scanned in a zigzag fashion during scanning.
- [Claim 4] The method of claim 2, wherein creating an image feature vector further comprises normalizing the image feature vector.
- [Claim 5] The method of claim 1, wherein creating an image feature vector comprises generating multiple random vectors following a Gaussian distribution.
- [Claim 6] The method of claim 1, wherein reducing the size of the captured frame comprises:
selecting a plurality of areas from the captured frame; and
calculating average pixel values for the individual selected areas.
- [Claim 7] The method of claim 6, wherein selecting a plurality of areas comprises selecting multiple areas excluding a predetermined area.
- [Claim 8] The method of claim 7, wherein the predetermined area excluded from selection is an area in which a caption, logo, advertisement or broadcast channel indicator is located.
- [Claim 9] The method of claim 1, wherein reducing the size of the captured frame comprises converting the captured frame into a grayscale frame and

- reducing the size of the grayscale frame.
- [Claim 10] The method of claim 1, wherein, in transforming the reduced frame, one of Discrete Cosine Transform (DCT), Discrete Fourier Transform (DFT) and Discrete Wavelet Transform (DWT) is applied.
- [Claim 11] The method of claim 1, wherein searching a database for information comprises utilizing a binary search technique to retrieve information related to the fingerprint from the database.
- [Claim 12] The method of claim 1, wherein searching a database for information comprises:
modifying, when no information related to the fingerprint is retrieved, one bit of the fingerprint; and
searching the database for information related to the modified fingerprint.
- [Claim 13] An apparatus for image processing, comprising:
a frame capturer capturing a frame of an image;
a fingerprint extractor extracting a fingerprint from the captured frame;
and
a fingerprint matcher searching a database for information related to the fingerprint,
wherein the fingerprint extractor reduces the size of the captured frame, transforms the reduced frame to a frequency domain frame, creates an image feature vector by scanning frequency components of the frequency domain frame, computes inner product values by projecting the image feature vector onto random vectors, and generates the fingerprint by applying a Heaviside step function to the inner product values.
- [Claim 14] The apparatus of claim 13, wherein the fingerprint extractor scans low-frequency components of the frequency domain frame except for a Direct Current (DC) component of the frequency domain frame and high-frequency components of the frequency domain frame exceeding a preset threshold value.
- [Claim 15] The apparatus of claim 13, wherein the fingerprint extractor selects a plurality of areas from the captured frame and calculates average pixel values for the individual selected areas.

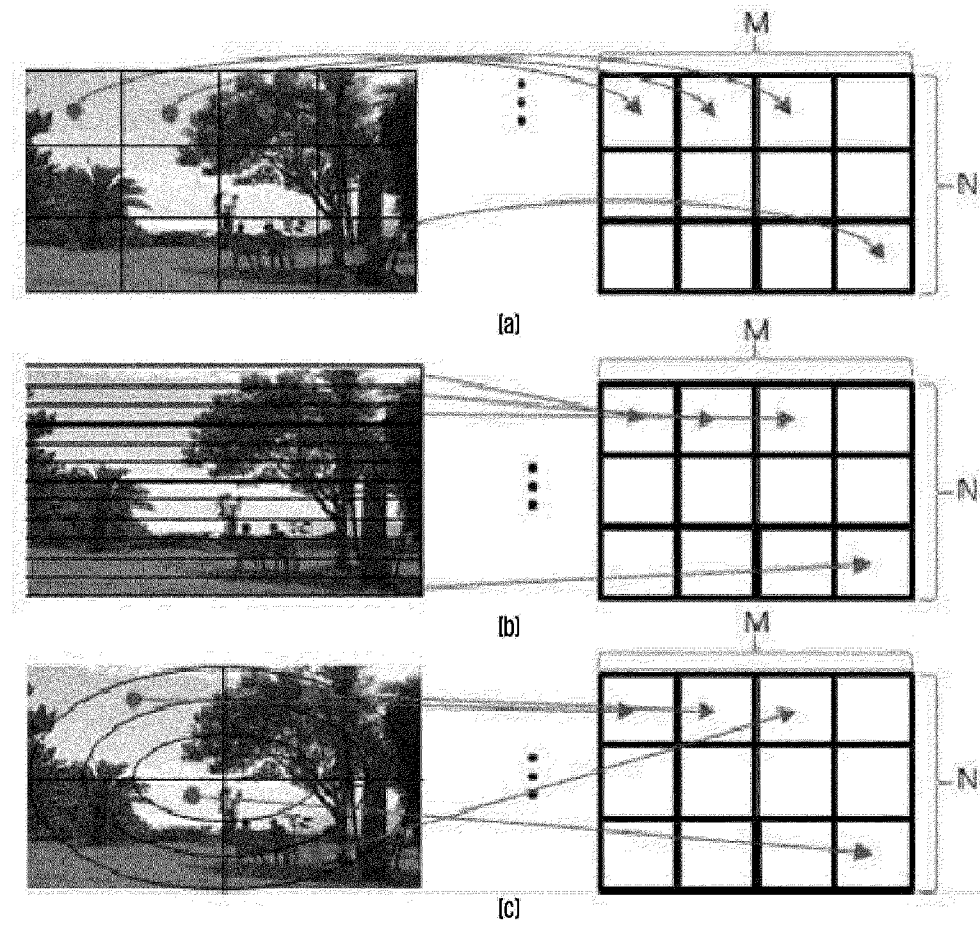
[Fig. 1]



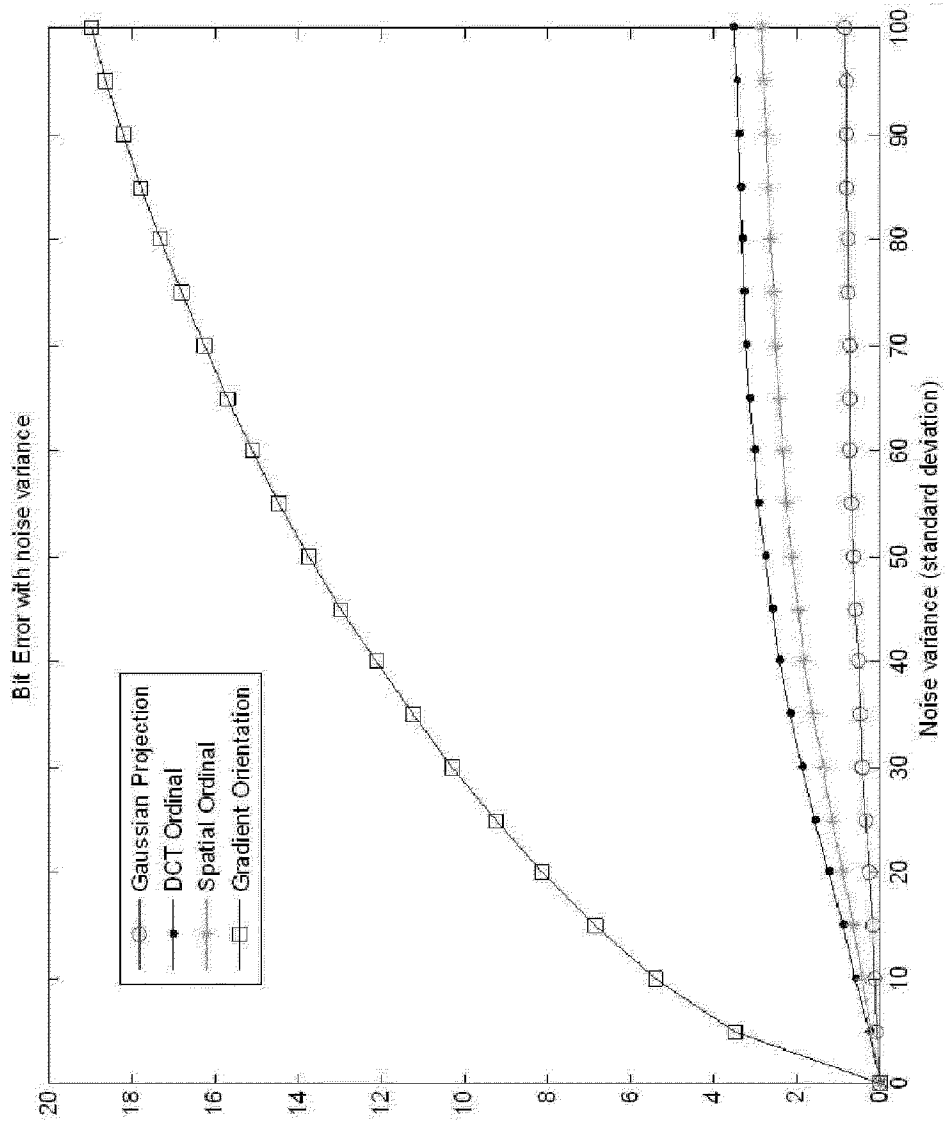
[Fig. 2]



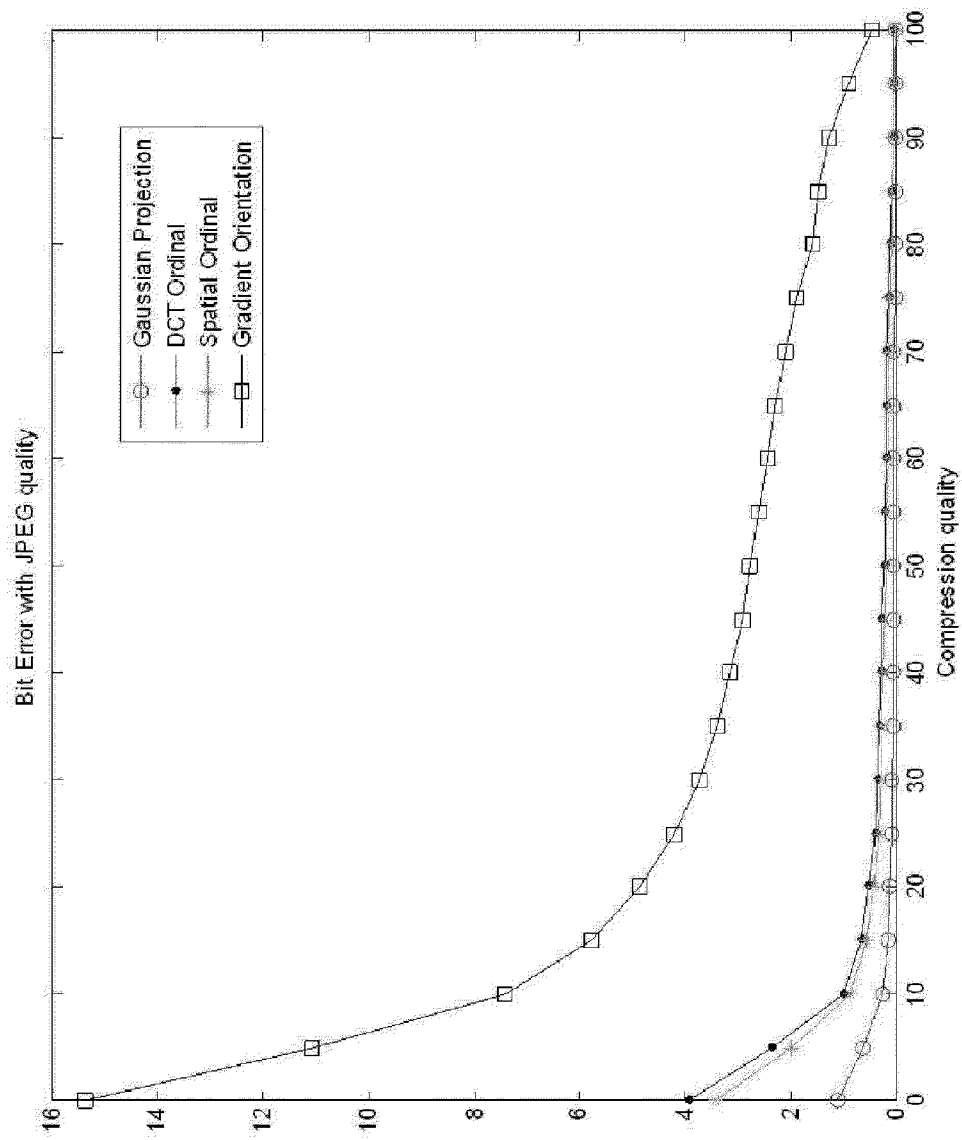
[Fig. 4]



[Fig. 5]



[Fig. 6]



[Fig. 7]

