# A Novel Image Recognition Method Based on Feature-Extraction Vector Scheme

*Masao Hiramoto*†, *Takahiro Ogawa*‡ *and Miki Haseyama*‡

† Panasonic Information Systems Co., Ltd (E-mail: hiramoto.masao@jp.panasonic.com)
6-18, Miyahara 4-Chome, Yodogawa-ku, Osaka City, Japan
‡ School of Engineering, Hokkaido University (E-mail: {ogawa,mikich}@media.eng.hokudai.ac.jp)
Kita 13, Nishi 8-Chome, Kita-ku, Sapporo City, Japan

## ABSTRACT

This paper introduces a method for recognizing images using a new approach to expressing images as vectors. Using this expression method, an image is constructed from 2 types of vectors – vectors indicating positions and vectors denoting intensity gradients for those positions. When investigating the amount of difference between two images, similarities are evaluated by calculating voting densities in the image space, using the vectors making up the sample image in relation to the vectors expressing the reference image. The expression proposed is invariant to image rotation and, by changing the resolution hierarchically, recognition using this expression is also adaptable to perspective and detail. Using this method, we carried out experimentation recognizing representative images from various fields and the results show that the method is effective in discriminating between them.

## 1. INTRODUCTION

Feature extraction from images is a key technology for recognizing and classifying images in multimedia databases. Many feature extraction methods have been developed, and the precision of image recognition has greatly improved. For example, approaches carrying out statistical analysis on wavelet-coefficients [4], [5], and those further including color components in the analysis [1] have been proposed. Other approaches divide the image into several segments and use the characteristics of those segments [2], [3], [6]. These approaches can be basically categorized as color-based, texture-based and segment-based. However, because the image is not constructed from geometrical elements, these methods experience problems when applied to images which have undergone transformations in geometry or resolution due to image processing.

In this paper, we propose an image recognition technique using a new approach to express images in terms of vectors. The purpose of our proposal is to illustrate the effectiveness of the expression proposed and prove that images can be recognized and retrieved hierarchically, even if they have undergone image processing such as filtering or rotation. We present the results of our experiments, and show that several representative images can be accurately distinguished.

## 2. VECTOR EXPRESSION OF IMAGES

### 2.1. Image Data

In general, the image is expressed in terms of an image intensity Z and 2-D coordinates X and Y. The image can therefore be thought of as a 3-D surface expressed in terms of X, Y and Z. From this, we regard an image as being expressed in terms of planar regions where ($\partial Z/\partial X$) and ($\partial Z/\partial Y$) are constant only, and define the image as being constructed from vectors ***Vi*** (i = 1, 2, …, k; where k is the number of vectors), as shown below. In the following, ***Vi*** is referred to as the intensity gradient vector.

$$\boldsymbol{Vi} = (\partial Z/\partial X,\ \ \partial Z/\partial Y)$$

However, when X, Y and Z are on a plane, the following condition is necessary: $\partial^2 Z/\partial X \partial Y = 0$

In order to determine the position of ***Vi***, an origin is set on the X-Y plane. For example, the origin could be the center of the image. The vector from the origin to the position of ***Vi*** is expressed as ***Ri*** (i = 1, 2, …, k; where k is the number of vectors). In the following, ***Ri*** is referred to as the position vector. Ultimately, we will express the image in terms of intensity gradient vectors ***Vi*** and position vectors ***Ri***. The relation between ***Vi*** and ***Ri*** is shown in Fig. 1(a).

### 2.2. Recognition Data

Image recognition data is created using the image data ***Vi*** and ***Ri*** described above. Recognition data is constructed from two types of vectors: ***Vi*** and one more type, ***Ui***, as shown below:

$$\theta i = \cos^{-1}\left(\frac{-ViRi}{|Vi||Ri|}\right),\ \boldsymbol{Ui} = (|\boldsymbol{Ri}|\cos\theta i,\ |\boldsymbol{Ri}|\sin\theta i)$$

0-7803-8554-3/04/$20.00 ©2004 IEEE.

***Ui*** is a vector denoting the start point (origin) of ***Ri*** based on ***Vi***. This is referred to as the voting vector. However, adjustment is carried out for ***Vi*** and ***Ui*** to ensure that no overlapping data exists. The relation between ***Vi*** and ***Ui*** is shown in Fig. 1(b).

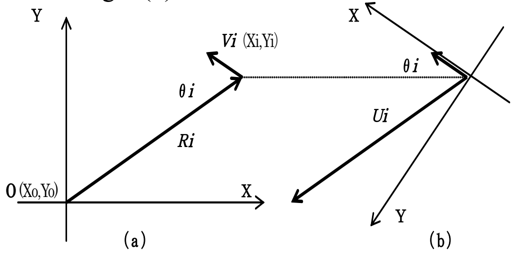



Fig. 1 Relation between ***Vi***, ***Ri*** and ***Ui***

## 3. PREPARATIONS FOR IMAGE RECOGNITION

Intensity gradient vectors $\boldsymbol{Vi}^s$ and position vectors $\boldsymbol{Ri}^s$ (i = 1, 2, …, N; where N is the total number of vectors) are prepared as a recognition sample image, and intensity gradient vectors $\boldsymbol{Vj}^r$ and voting vectors $\boldsymbol{Uj}^r$ (j = 1, 2, …, M; where M is the number of vectors) are prepared as a reference image for recognition. In addition, 2-D memory, sufficiently large to cover the size of the two images being compared, is prepared for image recognition. This is referred to as a voting memory.

## 4. PRINCIPLES OF PROPOSED IMAGE RECOGNITION

First, to fit the sample image in the voting memory, position alignment is carried out and the origin of the sample image is established. Next, $|\boldsymbol{Vi}^s|$ and $|\boldsymbol{Vj}^r|$ are compared for all points. If $|\boldsymbol{Vi}^s|$ and $|\boldsymbol{Vj}^r|$ are identical, one vote is added at the position in the voting memory denoted by $\boldsymbol{Uj}^r$, based on $\boldsymbol{Vi}^s$ from the position denoted by $\boldsymbol{Ri}^s$. If the reference image and sample image are identical, the point with most votes is the origin, and the number of votes is N. Our recognition method is based on this idea. The number of votes cast at the origin point of the sample image is checked and if the value is greater than a predetermined threshold, the sample image and the reference image are judged as being identical. Fig. 2 shows ***Vi***, ***Ri*** and ***Ui*** when the sample image and reference image are identical (N=M=5). Fig. 3 shows the vote state for that case.

## 5. PROPOSED IMAGE RECOGNITION METHOD

### 5.1. Data Creation

We use wavelet decomposition for creation of image data and recognition data. That is, the intensity gradient vector is (HL, LH) when HH=0. The calculation of HL, LH and

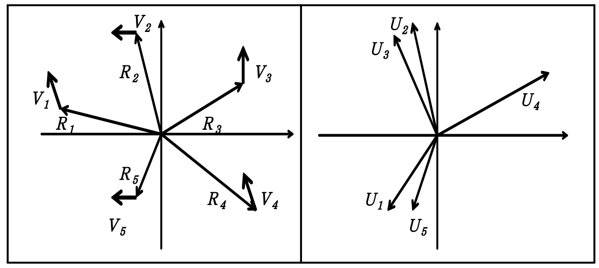



Fig. 2 ***Vi***, ***Ri***, and ***Ui*** (M=N= 5)

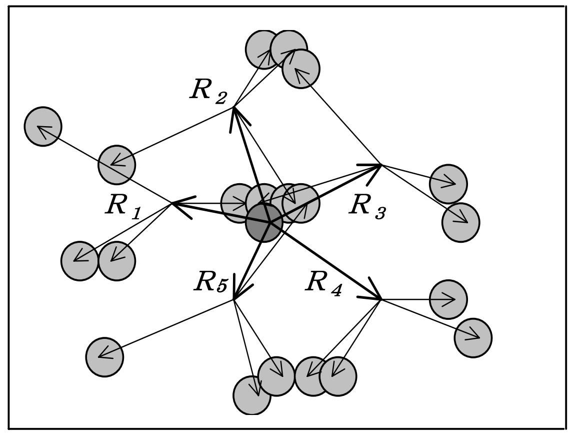
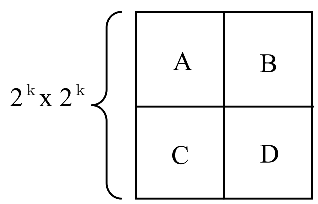



Fig. 3 Vote State (M=N= 5)
[ ○ : Point with votes , ● : Point with most votes]




A,B,C,D=∑f (x+i, y+j)
HL= (A-B+C-D)/($2^k$ x $2^k$)
LH= (A+B-C-D)/($2^k$ x $2^k$)
HH= (A-B-C+D)/($2^k$ x $2^k$)
[ f : intensity of an image]

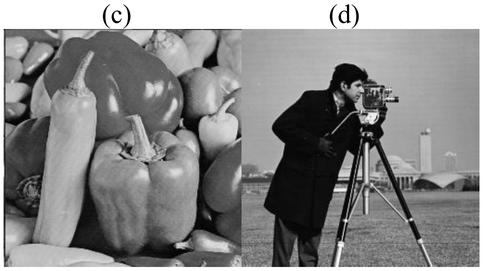
Fig. 4 Window ($2^k$ x $2^k$ Pixels) of Calculation

HH is shown in Fig. 4. The actual data creation process is described below.

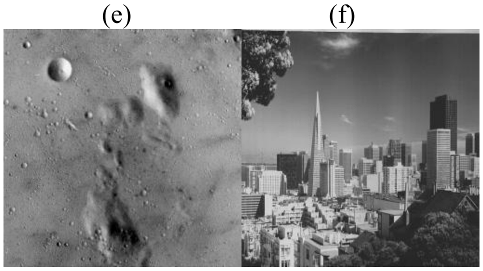
(1) Establish the region for recognition in one of the images
(2) Establish the window ($2^k$ x $2^k$) for calculation of wavelet decomposition
Calculation: $\sum f(xi+I, yi+J)/(2^k \times 2^k)$
X Range: I = $-2^{(k-1)} \sim -1, 0 \sim 2^{(k-1)}-1$
Y Range: J = $-2^{(k-1)} \sim -1, 0 \sim 2^{(k-1)}-1$
(3) Set (Xi, Yi) in the recognition area and carry out wavelet decomposition. If HH ≒ 0 (the wavelet decomposition window is a plane), calculate HL, LH.
(4) Set the intensity gradient as ***Vi***=(HL, LH) and save the relevant position (Xi, Yi).
(5) Shift the wavelet decomposition window one pixel within the recognition area and repeat steps (3) - (4).

(a) (b)

(c) (d)



(e) (f)



(g) (h)

Fig. 5 Images (256 x 256 pixels, 8 bits/pixel)

(6) Calculate the center point for the intensity gradient vectors $\boldsymbol{Vi}$ ($V_1$, $V_2$, …, $V_N$ [where N is the total number of $\boldsymbol{Vi}$]) and set that point to the origin O (Xo, Yo).
(7) Calculate all position vectors $\boldsymbol{Ri}$ from (Xo, Yo) and (Xi, Yi). Align $\boldsymbol{Ri}$ with $\boldsymbol{Vi}$ and set it as the image data.
(8) Calculate all values for $\boldsymbol{Ui}$ from $\boldsymbol{Vi}$ and $\boldsymbol{Ri}$.
(9) Select $\boldsymbol{Ui}$ and $\boldsymbol{Vi}$, ensuring that there is no overlapping, and set them as the recognition data.

### 5.2. Image Recognition

Before comparing two images, first their similarities have to be defined. In this paper, similarity is defined as described below:

$$\text{Similarity}=\frac{\text{Maximum Number of Votes at the Origin of a Sample Image}}{\text{M x N}}$$

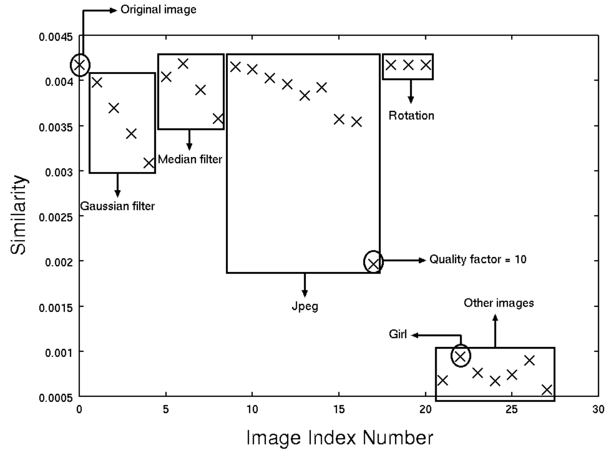



Fig. 6 Similarity of images

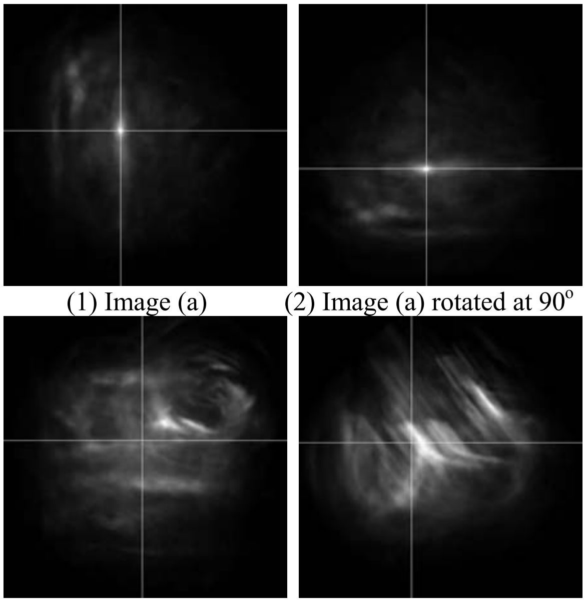
Where M and N are the number of intensity gradient vectors for each image.
The recognition process when comparing the sample image and the reference image is as follows:
(1) Prepare $\boldsymbol{Vi}^s$ and $\boldsymbol{Ri}^s$ (i = 1, 2, ..., M) for the sample image, $\boldsymbol{Vj}^r$ and $\boldsymbol{Uj}^r$ (j = 1, 2, ..., N) for the reference image and the voting memory.
(2) Compare $|\boldsymbol{Vi}^s|$ with $|\boldsymbol{Vj}^r|$ and if they are identical, add 1 vote at the position in voting memory denoted by $\boldsymbol{Uj}^r$, based on $\boldsymbol{Vi}^s$ from the position denoted by $\boldsymbol{Ri}^s$.
(3) Repeat (2) for all instances of $\boldsymbol{Vi}^s$ and $\boldsymbol{Vj}^r$ (M x N times).
(4) Calculate Similarity.
(5) If the similarity is above a predetermined threshold, judge the images as identical.

## 6. RESULTS OF EXPERIMENTATION

In order to verify the high performance of the method proposed in this paper, we present here the results of experimentation applying the method to the recognition of some actual images. In this experiment, 8 types of image were used (see Fig. 5), and 20 variations of image 5(a) were prepared.
The 19 variations other than the base image are shown below:

- Gaussian filter: 4 types (radius = 3, 5, 7 and 9 pixels)
- Median filter: 4 types (window size = 3, 5, 7 and 9 pixels)
- JPEG compression: 9 types (level = 10-90 at intervals of 10)
- Rotation: 3 types (angle = 90, 180, 270 degrees)

The image sizes are all 256 x 256 pixels (8 bits/pixel), and the following test conditions were imposed:



(1) Image (a) (2) Image (a) rotated at 90°

(3) Image (b) (4) Image (c)

Fig. 7 Voting states

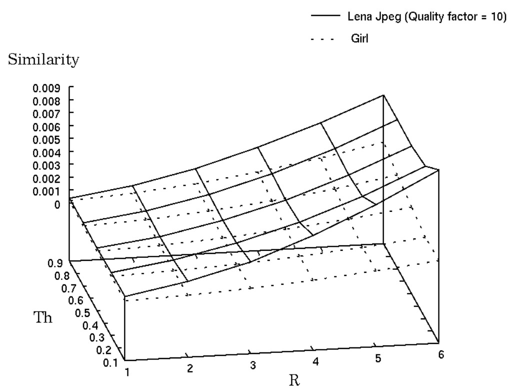



Fig. 8. Changes in similarity due to Th and R

- Wavelet decompression window size is $2^k$ x $2^k$: k = 4, 5 and 6
- Wavelet decompression window is judged as being one a plane (Th>|HH|): Th = 0.1, 0.3, 0.5, 0.7, 0.9
- Voting range at origin [(2R+1) x (2R+1)]: R = 1, 2, 3, 4, 5 and 6

As the first experiment, we measured the similarity of the image in Fig. 5(a) with the other 27 images. A typical result is shown in Fig. 6 (k=5, Th=0.1, R=1). Fig. 6 shows that no difference was detected between image (a) and its variants, but a large difference can be seen with the other test images. Using the recognition data from image (a), the voting results for image (a) itself, image (a) rotated to 90 degrees, image (b) and image (c) are shown in Fig. 7. In this figure, the origin points of each image are shown by a cross, and the brighter the area, the greater the number of votes.

In the second experiment, we investigated the similarity between image (a) [Lena: JPEG image with compression level 10] and image (c) [Girl] while altering the values for Th and R. The first image is the variant of image (a) with the lowest similarity score in the first experiment, and the second is the image type other than image (a) with the highest similarity score. The results of the experiment are shown in Fig. 8.

The results show that even when the values for Th and R are changed, there is still a difference in the similarity score for both pictures when compared to image (a). The approach proposed makes a clear distinction between images which are similar, and images which are not.

## 7. CONCLUSION

In this paper we proposed a new approach to image recognition. Using this method, an image and variants of it which have undergone image processing (such as filtering, JPEG compression or rotation) can be recognized as the same image, and a clear distinction can be made between that image and the other images.

## 8. REFERENCES

[1] Jia Li and James Z.Wang, "Automatic Linguistic Indexing of Pictures by a Statistical Modeling Approach," *IEEE Trans. Pattern Analysis and Machine Intelligence,* Vol. 25, No. 9, pp. 1075-1088, Sept. 2003.

[2] Yixin Chen and James Z.Wang, "A Region-Based Fuzzy Feature Matching Approach to Content-Based Image Retrieval," *IEEE Trans. Pattern Analysis and Machine Intelligence,* Vol. 24, No. 9, pp. 1252-1267, Sept. 2002.

[3] C.Carson, S.Belongie, H.Greenspan, and J.Malik, "Blobworld: Image Segmentation Using Expectation-Maximization and Its Application to Image Querying," *IEEE Trans. Pattern Analysis and Machine Intelligence,* Vol. 24, No. 8, pp. 1026-1038, Aug. 2002.

[4] Minh N.Do and Martin Vetterli, "Rotation Invariant Texture Characterization and Retrieval Using Steerable Wavelet-Domain Hidden Markov Models," *IEEE Trans. Multimedia,* Vol. 4, No. 4, pp. 517-527, Dec. 2002.

[5] M. K. Mandal, T. Aboulnasr, and S.Panchanathan, "Fast Wavelet Histogram Techniques for Image Indexing," *Computer Vision and Image Understanding,* Vol. 75, Nos. 1/2, July/August, pp. 99-110, 2002.

[6] Daniel L. Swets and Juyang Weng, "Hierarchical Discriminant Analysis for Image Retrieval," *IEEE Trans. Pattern Analysis and Machine Intelligence,* Vol. 21, No. 5, pp. 386-401, May 1999.