IMAGES IDENTIFICATION BASED ON EQUIVALENCE CLASSES

Y. Maret, G. Garcia and T. Ebrahimi

Ecole Polytechnique Fédérale de Lausanne (EPFL)
Institut de Traitement des Signaux
CH-1015 Lausanne, Switzerland

ABSTRACT

The image identification problem consists in identifying all the equivalent forms of a given reference image. An image is equivalent to the reference image, if the former results from the application of an image operator (or a composition of image operators) to the latter. Depending on the application, different sets of image operators are considered. The equivalence quantification is done in three levels. In the first level, we construct the set of equivalent images which is composed of the reference and its modified versions obtained through the application of image operators. In the second level, visual features are extracted from images in the equivalence set and their distances to the reference image are computed. In the third level, an orthotope (generalized rectangle) is fit to the set of distance vectors corresponding to the equivalent images. The equivalence of an unknown image with respect to a given reference is defined according to whether the corresponding distance vector is inside, or outside, the orthotope. The results of our algorithm are assessed in terms of the false positive and false negative errors computed over different choices of reference images and operators.

1. INTRODUCTION

The problem of search and retrieval of multimedia content is an exciting field of research, which has attracted an increasing attention from both scientific and business communities. The activities in MPEG-7 standardization, and the more recent Still Image Search project within JPEG (JPSearch) are evidences of this growing interest.

In this paper, we describe a particular subset of search and retrieval problem which aims at the identification of all equivalent forms of a given multimedia content. By equivalent, we mean, all instances of a given content, which have been subject to a series of equivalence operators. For instance, an image of Albert Einstein (reference image), and all variants of that particular image, after application of a JPEG compression with different parameters, its zoomed versions, its filtered versions, etc. Such identification system can be of interest in applications in which one is interested in identifying all versions of a same content. Applications include search and retrieval of content with illicit nature (child pornography and other illicite images), or variations of a content with copyright (images made by an artist).

Current methods permitting to identify image variations are mainly based on two approaches: robust *watermarking* and robust *fingerprinting* (or perceptual hashing). In watermarking [2], a signature is embedded in the reference image before broadcasting. A given image is equivalent to the reference image only if the same watermark is present. Watermarking techniques require to modify the reference image, which might be problematic in some cases (for example, when the reference image has already been broadcasted without embedding any watermarks). In fingerprinting, the reference image is analyzed to produce a signature correlated to its content. A given image is equivalent to the reference image only if their signatures are close enough. Fingerprinting techniques often rely on a single feature, for example typical points of the Radon transform [4], log-mapping of the Radon transform [7], or intra-scale variances of the wavelet coefficients [10]. Our method is similar to fingerprinting in the sense that there is no need to modify the reference image. However, contrary to most fingerprinting approach, the present method combines multiple features.

In [3], feature distances such as structure, color and texture are linearly combined to form a unique distance quantifying the *similarity* of two images. The combination weights are empirically estimated; it was also shown that adapting them to the nature of the considered images increased the efficiency of the retrieval system. In this paper, distances are non-linearly combined to define an equivalence distance function, which is specific to each reference image.

2. IDENTIFICATION PROBLEM

Let Ξ be the space of images. We consider a *reference image* $\mathbf{R} \in \Xi$ and a *set of image operators* $\mathcal{O} = \{O_o(\cdot)\}_{o=1,\ldots,O}$ with $O_o(\cdot) : \Xi \rightarrow \Xi$. The *equivalence class* of $\mathbf{R}$ is defined as:

$$\mathcal{E}_{\mathcal{O}}(\mathbf{R}) = \mathbf{R} \bigcup \{O_o(\mathbf{R})\}_{o=1,\ldots,O}. \quad (1)$$

According to this definition, each element in $\mathcal{E}_{\mathcal{O}}(\mathbf{R})$ is *equivalent* to $\mathbf{R}$, with respect to the operators set $\mathcal{O}$. Hence, identifying an image $\mathbf{U}$ as equivalent to $\mathbf{R}$ consists in determining whether $\mathbf{U} \in \mathcal{E}_{\mathcal{O}}(\mathbf{R})$.

The equivalence notion and the choice of operators are application dependent. For example, $\mathcal{O}$ can contain all JPEG compressions operators up to a certain quality factor, all scaling operators in a certain range, and all possible combinations of both compression and scaling operators.

While the equivalence class notion, defined in (1), is useful, its practical implementation is rather difficult. Suppose that $\mathbf{U}$ is a modified version of $\mathbf{R}$, then $\mathbf{U}$ can be identified as $\mathbf{R}$ only if $\mathbf{U} \in \mathcal{E}_{\mathcal{O}}(\mathbf{R})$. Hence, $\mathcal{O}$ should contain the operator transforming $\mathbf{R}$ into $\mathbf{U}$. On the one hand, it is unrealistic to include all possible operators in $\mathcal{O}$. On the other hand, the cardinality $|\mathcal{E}_{\mathcal{O}}(\mathbf{R})|$ is directly proportional to $|\mathcal{O}|$, limiting the size of the latter. In this paper, we propose an efficient method to implement the identification problem, as defined above, in a practical manner. That

is, the identification problem is tackled by building an *equivalence distance function* $e_{\mathbf{R}}(\mathbf{U})$, quantifying the equivalence between $\mathbf{U}$ and $\mathbf{R}$. This function depends on the targeted image $\mathbf{R}$ and the operators set $\mathcal{O}$.

3. PROPOSED METHOD

3.1. Feature Extraction and Distance Vector

Feature extraction is the operation that extracts visual information from a given image. Many visual features can be envisioned: color, texture, shape, etc. For an extensive survey on general features extraction, refer to [6].

The features choice depends on the operators considered in $\mathcal{O}$. For instance, if rotation is considered, it would make sense to choose features that are rotation invariant.

Let F be the number of features used. A *distance vector* $\mathbf{d}_{\mathbf{R}}(\mathbf{I})$, $\mathbf{I} \in \Xi$, can be defined as

$$\mathbf{d}_{\mathbf{R}}(\mathbf{I}) = [d_1(\mathbf{I},\mathbf{R}) \cdots d_f(\mathbf{I},\mathbf{R}) \cdots d_F(\mathbf{I},\mathbf{R})]^T \tag{2}$$

where $d_f(\cdot,\cdot)$ is a distance measure in the space defined by the feature f.

Based on the above definition, the problem of identification of equivalent images amounts to determining which distance vectors $\mathbf{d}_{\mathbf{R}}(\cdot)$ correspond to images in the equivalence class $\mathcal{E}_{\mathcal{O}}(\mathbf{R})$.

The distance vectors corresponding to images that are equivalent to a given reference image $\mathbf{R}$ are concentrated near the origin in the space defined by $d_1(\cdot,\mathbf{R}),\ldots,d_F(\cdot,\mathbf{R})$, and denoted as $\Omega_{\mathbf{R}}$. Figure 1 illustrates $\Omega_{\mathbf{R}}$ in a two-dimensional example using two simple features. The first feature is the gray level histogram. The histogram is quantized into 16-bin. The corresponding distance metric is based on the histogram intersection algorithm [8]. The second feature is the first order statistics of each subband of the Gabor transform. The transform is performed as in [5]. More precisely, the parameters used are 0.75 for the upper center frequency, 0.05 for the lower center frequency, five scales and six orientations[1]. Hence, there is in all 30 subbands corresponding to 30 mean values. The corresponding distance metric is the L_1-norm of the difference between two vectors of mean values. In the following, each component of the distance vector is normalised by its median. The latter is computed using the whole set of equivalent images.

3.2. Equivalence distance function

Let denote by $\mathbf{d}_{\mathbf{R}}\{\mathcal{E}_{\mathcal{O}}(\mathbf{R})\}$ the set of normalized distance vectors for all the elements in $\mathcal{E}_{\mathcal{O}}(\mathbf{R})$. An image $\mathbf{U}$ is equivalent to $\mathbf{R}$ if $\mathbf{d}_{\mathbf{R}}(\mathbf{U})$ is a member of the subspace spanned by $\mathbf{d}_{\mathbf{R}}\{\mathcal{E}_{\mathcal{O}}(\mathbf{R})\}$. To quantify this membership, an orthotope (generalized rectangle) is built in $\Omega_{\mathbf{R}}$ containing most of the elements in $\mathbf{d}_{\mathbf{R}}\{\mathcal{E}_{\mathcal{O}}(\mathbf{R})\}$. The vertices of the orthotope are the origin[2] and the following points:

$$\big(w_1(\mathbf{R}),\ldots,0\big),\ \big(0,w_2(\mathbf{R}),\ldots,0\big),\ \ldots,\ \big(0,\ldots,w_F(\mathbf{R})\big)$$

where $w_f(\mathbf{R}) \geqslant 0$ is the orthotope limit associated with $d_f(\cdot,\mathbf{R})$. The computation of these limits are detailed in Sec. 3.3.

Thus, an equivalence distance function can be defined as:

$$e_{\mathbf{R}}(\mathbf{U}) = 1 - \min_{f=1,\ldots,F}\left(1 - \frac{d_f(\mathbf{U},\mathbf{R})}{w_f(\mathbf{R})}\right). \tag{3}$$

[1] Refer to [5] for more details about the parameters.

[2] corresponding to $\mathbf{d}_{\mathbf{R}}(\mathbf{R})$

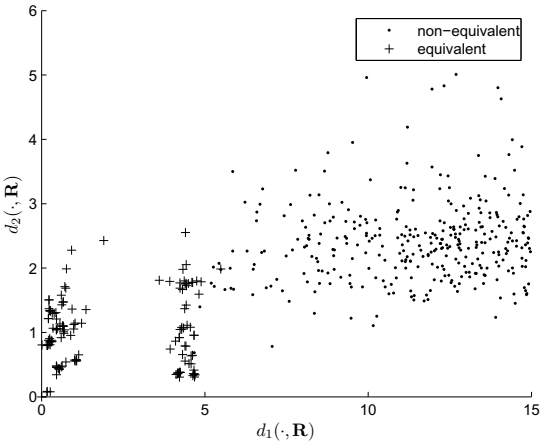


Fig. 1. Illustration of $\Omega_{\mathbf{R}}$ in a two-dimensional space.

When $d_f(\mathbf{U},\mathbf{R})$ is inside the orthotope, or in its border, then $e_{\mathbf{R}}(\mathbf{U}) \leq 1$. Likewise, when $d_f(\mathbf{U},\mathbf{R})$ is outside the orthotope then $e_{\mathbf{R}}(\mathbf{U}) > 1$. For equivalence values larger than 1, we assume that the image $\mathbf{U}$ is not equivalent to the reference image $\mathbf{R}$.

3.3. Orthotope Limits Computation

We consider a training set, denoted as $\mathcal{T}_{\mathbf{R}} = \{\mathbf{T}_1,\ldots,\mathbf{T}_M\}$, containing images that are equivalent to a given reference image $\mathbf{R}$. By computing the distance vectors between the elements in $\mathcal{T}_{\mathbf{R}}$ and $\mathbf{R}$, one obtains the set $\mathcal{X} = \{\mathbf{x}_1,\ldots,\mathbf{x}_M\}$ where $\mathbf{x}_m = \mathbf{d}_{\mathbf{R}}(\mathbf{T}_m)$.

The orthotope limits can be found by solving the following constrained optimization problem:

$$\min_{\mathbf{w},\xi_{mf}} \quad \sum_{f=1}^{F} w_f + C \cdot \sum_{m=1}^{M}\sum_{f=1}^{F} \xi_{mf} \tag{4a}$$

$$\text{subject to} \quad x_m(f) \leq w_f + \xi_{mf} \text{ and } \xi_{mf}, w_f \geq 0 \tag{4b}$$

$$1 \leq m \leq M\ ;\ 1 \leq f \leq F$$

where $\mathbf{w} = \{w_1,\ldots,w_F\}$, $w_f = w_f(\mathbf{R})$, $x_m(f)$ is the f^{th} component of $\mathbf{x}_m$, and ξ_{mf} is a positive slack variable which allows for $\mathbf{x}_m$ to be outside the orthotope [9]. Indeed, as it can be seen in Fig. 2 the orthotope limits should be chosen so as to reject outliers. However, the number of rejections should be limited. To achieve this, the sum of the slack variables is penalized by a positive trade-off constant C, which is determined through cross-validation (see Sec. 4).

The optimization problem (4) can be solved by means of standard linear programming algorithms. Indeed, the problem can be expressed in the *Standard Form*:

$$\min_{\mathbf{y}} \quad \mathbf{f}^T\mathbf{y} \tag{5a}$$

$$\text{subject to} \quad \mathbf{A}\mathbf{y} \leq \mathbf{b} \text{ and } -\mathbf{y} \leq \mathbf{0}. \tag{5b}$$

As an example, the matrix $\mathbf{A}$ and the vectors $\mathbf{b}$, $\mathbf{f}$ and $\mathbf{y}$ are ex-

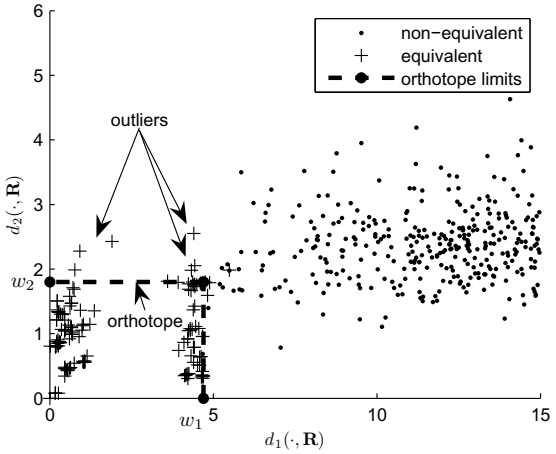
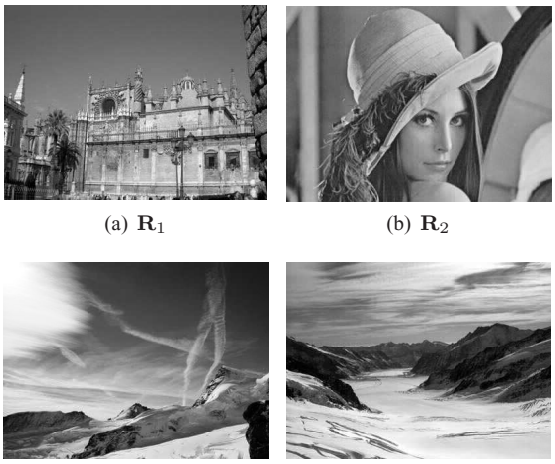


Fig. 2. Outliers rejection.



(a) $\mathbf{R}_1$ (b) $\mathbf{R}_2$

(c) $\mathbf{R}_3$ (d) $\mathbf{R}_4$

Fig. 3. Reference images.

plicitly given (in the case where $F = 2$ and $M = 3$):

$$\mathbf{A} = -\begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{b} = [x_1(1)\ x_2(1)\ x_3(1)\ x_1(2)\ x_2(2)\ x_3(2)]^T$$

$$\mathbf{f} = [1\ 1\ C\ C\ C\ C\ C\ C]^T$$

$$\mathbf{y} = [w_1\ w_2\ \xi_{11}\ \xi_{21}\ \xi_{31}\ \xi_{12}\ \xi_{22}\ \xi_{32}]^T .$$

4. EXPERIMENTAL RESULTS

To test the proposed method, a set of four reference images was chosen. They are denoted as $\mathbf{R}_1, \ldots, \mathbf{R}_4$ and depicted in Fig. 3. These images are all 256 gray-level and of size 400×300 pixels.

Basic operators	**Parameters**
JPEG compression	$Q = 50, 60, 70$
Gaussian noise addition	$\sigma = 2.5, 7.5$
Resizing	scale$= 0.8, 1.2$
Averaging filter	order$= 2, 3$
Gamma correction	$\gamma = 0.8, 1.2$
Horizontal flipping	none

Table 1. Basic operators and their parameters.

The image operators are obtained by binary composition of the 12 basic operators in Tab. 1. Hence, the set $\mathcal{O}$ consists of the 118 possible binary compositions[3], and the basic operators. The set $\mathcal{X}_r$ contains the 130 distance vectors between the equivalent images in $\mathcal{E}_{\mathcal{O}}(\mathbf{R})$ and $\mathbf{R}_r$ (for $r = 1, \ldots, 4$). The distance vectors between $\mathbf{R}_r$ and non-equivalent images are computed as well. The database used for non-equivalent elements contains 536 images including photographs of people, landscapes, and buildings.

In order to find the optimum value for the tradeoff constant C_r in the optimization problem (4), a ten-fold cross-validation procedure is carried out [1]. In this procedure, the set $\mathcal{X}_r$ is subdivided into ten mutually exclusive subsets, and ten runs are carried out. For each run, one set is put aside (validation set), and the orthotope limits are estimated using the remaining sets. The union of the latter is called the training set. For each reference image, its orthotope is derived as explained in Sec. 3.3.

Figure 3 reports the average *false-positive* and *false-negative* errors for each reference image and ten values of C_r sampled in $[0.1, 2]$. The false-positive error, with respect to a reference image $\mathbf{R}_r$, corresponds to the fraction of non-equivalent images whose distance vectors fall inside the $\mathbf{R}_r$ orthotope; it corresponds to non-equivalent images detected as equivalent images by the system. The false-negative error corresponds to the average fraction of equivalent images, in the validation set, falling outside the $\mathbf{R}$ orthotope; it corresponds to equivalent images detected as non-equivalent images by the system. Moreover, the average fraction of equivalent images, in the training set, falling outside the orthotope is also reported (false-negative training).

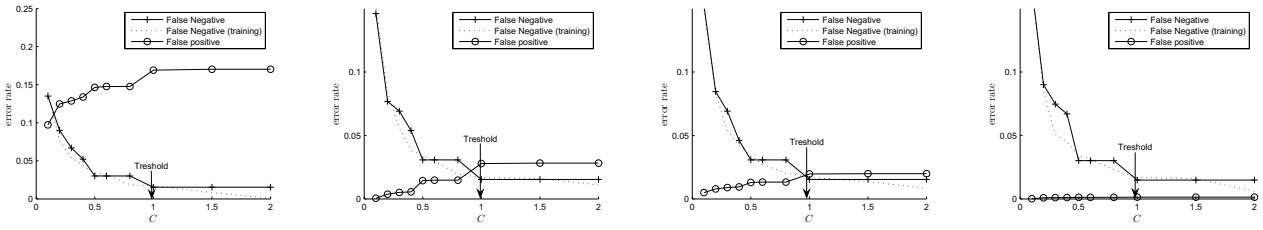
As it can be seen in Fig. 4, the false-negative error decreases with C_r, until it reaches a certain threshold value from which it remains mostly constant. A similar behavior can be observed for the false-negative training. On the contrary, the average false-positive error increases with C_r. Depending on the application, it might be undesirable to have a large false-positive error. In this case, C_r need to be smaller than the threshold.

For illustration purpose, we choose C_1, C_2, C_3 and C_4 to be equal to 0.7 because the false-negative error decreasing rate slows beyond that value. Moreover, it permits to keep a relatively low false-positive rate. The orthotopes limits, for this tradeoff value, are shown in Fig. 5.

5. CONCLUSIONS AND FUTURE WORK

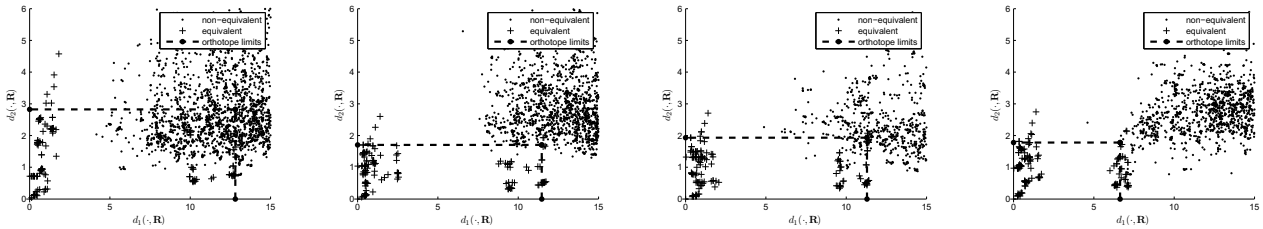
This paper reports an original approach for image identification based on equivalence classes. Equivalence of a reference image is defined as all admissible variations of that image when subjected to a set of operators. The approach is based on the construction of

[3]there is no composition between operators with the same nature. For instance, the composition of two JPEG compression operators with different quality factor is not considered.



(a) $\mathbf{R}_1$ (b) $\mathbf{R}_2$ (c) $\mathbf{R}_3$ (d) $\mathbf{R}_4$

Fig. 4. Average false-positive and false-negative errors.



(a) $\mathbf{R}_1$ (b) $\mathbf{R}_2$ (c) $\mathbf{R}_3$ (d) $\mathbf{R}_4$

Fig. 5. Orthotope limits ($C = 0.7$).

an orthotope in the space of features distance vectors. The limits of the orthotope are computed using a standard linear programming approach. Simulation results to identify four specific images and their 130 equivalent images out of a total of 536 non-equivalent images, and their variations, show promising results.

As a future work, we will take into account the distribution of non-equivalent images to define the equivalence distance function 'shape'. Moreover, a study on the sensitivity of the method with respect to choices of features and distance metrics might provide several usefull insights into the relative importance of each feature. This will pave the way to a scheme based on automatic feature selections.

6. ACKNOWLEDGMENTS

This research was partly funded by the Swiss National Science Foundation – "Multimedia Security", grant number 200021-1018411. The work was partly developed within VISNET, a European Network of Excellence (http://www.visnet-noe.org), funded under the European Commission IST FP6 programme. The authors would like to acknowledge Frederic Dufaux for fruitful discussions and comments.

7. REFERENCES

[1] C. M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.

[2] F. Hartung and M. Kutter. Multimedia watermarking techniques. *Proceedings of the IEEE*, 87(7):1079 – 1107, July 1999.

[3] Q. Iqbal and J. Aggarwal. Combining structure, color and texture for image retrieval: A performance evaluation. In *IEEE Conference on Pattern Recognition*, volume 2, pages 438–443, 2002.

[4] F. Lefbvre, B. Macq, and J.-D. Legat. Rash: Radon soft hash algorithm. In *EURASIP European Signal Processing Conference*, France, September 2002.

[5] B. Manjunath and W. Ma. Texture Features for Browsing and Retrieval of Image Data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8):837–842, August 1996.

[6] Y. Rui, T. Huang, and S. Chang. Image retrieval: current techniques, promising directions and open issues. *Journal of Visual Communication and Image Representation*, 10(4):39–62, 1999.

[7] J. Seo, J. Haitsma, T. Kalker, and C. Yoo. Affine transform resilient image fingerprinting. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Hong Kong, April 2003.

[8] M. Swain and D. Ballard. Indexing via color histograms. In *Computer Vision*, pages 390–393, December 1990.

[9] D. M. Tax and R. P. Duin. Support Vector Data Description. *Machine Learning*, 55:45–66, 2004.

[10] R. Venkatesan, S.-M. Koon, M.-H. Jakubowski, and P. Moulin. Robust image hashing. In *IEEE International Conference on Image Processing*, Vancouver, September 2000.