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 (IPSearch) 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 of content with illicit nature (child pornography and other illicit images), or variations of a content with copyright, and so forth. 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).

2. IDENTIFICATION PROBLEM
Let \( \Xi \) be the space of images. We consider a reference image \( \mathbf{R} \in \Xi \) and a set of image operators \( \mathcal{O} = \{ \mathcal{O}_i(\cdot) \}_{i=1,...,O} \) with \( \mathcal{O}_i(\cdot) : \Xi \to \Xi \). The equivalence class of \( \mathbf{R} \) is defined as:

\[
\mathcal{E}_{\mathcal{O}}(\mathbf{R}) = \mathbf{R} \bigcup \{ \mathcal{O}_i(\mathbf{R}) \}_{i=1,...,O}.
\]

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 iden-
tification problem, as defined above, in a practical manner. That
is, the identification problem is tackled by building an equivalence
distance function \( e_R(U) \), quantifying the equivalence between \( U \)
and \( R \). This function depends on the targeted image \( R \) and the
operators set \( O \).

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 
\( O \). 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 \( d_R(I), I \in \Xi \), can be defined as

\[
d_R(I) = [d_1(I, R) \cdots d_f(I, R) \cdots d_F(I, R)]^T
\]

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
\( d_R(\cdot) \) correspond to images in the equivalence class \( E_O(R) \).

The distance vectors corresponding to images that are equiva-
lent to a given reference image \( R \) are concentrated near the origin
in the space defined by \( d_1(\cdot, R), \ldots, d_F(\cdot, R) \), and denoted as 
\( \Omega_R \). Figure 1 illustrates \( \Omega_R \) 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 dis-

tance 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 fre-
quency, 0.05 for the lower center frequency, five scales and six
orientations. 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 follow-
ing, each component of the distance vector is normalized by its
median. The latter is computed using the whole set of equivalent
images.

3.2. Equivalence distance function

Let denotes by \( d_R \left( E_O(R) \right) \) the set of normalized distance vectors
for all the elements in \( E_O(R) \). An image \( U \) is equivalent to \( R \) if 
\( d_R(U) \) is a member of the subspace spanned by \( d_R \left( E_O(R) \right) \). To
to quantify this membership, an orthotope (generalized rectangle)
is built in \( \Omega_R \) containing most of the elements in \( d_R \left( E_O(R) \right) \). The
vertices of the orthotope are the origin \( 0 \) and the following points:

\[
(0, w_2(R), \ldots, 0), \ldots, (0, \ldots, w_F(R))
\]

where \( w_f(R) \geq 0 \) is the orthotope limit associated with \( d_f(\cdot, R) \).

The computation of these limits are detailed in Sec. 3.3.

\( ^{1} \)Refer to [5] for more details about these parameters.

\( ^{2} \)corresponding to \( d_R(R) \)

\( ^{3} \)That is, \( x_m = d_R(I_m) \). For more details on the choice of the \( M \)
equivalent images \( I_m \), refer to Sec. 4.
expressed in the Standard Form:

\[
\begin{align*}
\min_y & \quad f^T y \\
\text{subject to} & \quad Ay \leq b \text{ and } -y \leq 0.
\end{align*}
\]

As an example, the matrix \( A \) and the vectors \( b, f \) and \( y \) are explicitly given (in the case where \( F = 2 \) and \( M = 3 \)):

\[
A = \begin{bmatrix}
1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
\]

\[
b = \begin{bmatrix}
x_1(1) & x_2(1) & x_3(1) & x_1(2) & x_2(2) & x_3(2)
\end{bmatrix}^T
\]

\[
f = \begin{bmatrix}
1 & 1 & C & C & C & C
\end{bmatrix}^T
\]

\[
y = \begin{bmatrix}
w_1 & w_2 & \xi_{11} & \xi_{21} & \xi_{31} & \xi_{12} & \xi_{22} & \xi_{32}
\end{bmatrix}^T
\]

4. EXPERIMENTAL RESULTS

To test the proposed method, a set of four reference images was chosen. They are denoted as \( R_1, \ldots, R_4 \) and depicted in Fig. 3. These images are all 256 gray-level and of size 400 \times 300 pixels.

The image operators are obtained by binary composition of the 12 basic operators in Tab. 1. Hence, the set \( O \) consists of the 118 possible binary compositions\(^4\), and the basic operators. The training data are the 130 distance vectors between the equivalent images in \( E_O(R) \) and \( R_r \) (for \( r = 1, \ldots, 4 \)). The distance vectors between \( R_r \) 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 training data are 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 \( X_r \). 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 \( R_r \), corresponds to the fraction of non-equivalent images whose distance vectors fall inside the \( R_r \) 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 \( R_r \) 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.
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 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 the orthotope are computed using a standard linear programming approach. 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 useful insights into the relative importance of each feature. This will pave the way to a scheme based on automatic feature selections.

6. ACKNOWLEDGMENTS

The first author is partly supported 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