Valerii Filatov¹, Anna Filatova¹, Anatolii Povoroznyuk¹, Shakhin Omarov²
¹National Technical University “Kharkiv Polytechnic Institute”, Kharkiv, Ukraine
²Kharkiv National University of Radio Electronics, Kharkiv, Ukraine

IMAGE CLASSIFIER FOR FAST SEARCH IN LARGE DATABASES

Abstract. Relevance. The avalanche-like growth in the amount of information on the Internet necessitates the development of effective methods for quickly processing such information in information systems. Clustering of news information is carried out by taking into account both the morphological analysis of texts and graphic content. Thus, an urgent task is the clustering of images accompanying textual information on various web resources, including news portals. The subject of study is an image classifier that exhibits low sensitivity to increased information in databases. The purpose of the article is to enhance the efficiency of searching for identical images in databases experiencing a daily influx of 10-12 thousand images, by developing an image classifier. Methods used: mathematical modeling, content-based image retrieval, two-dimensional discrete cosine transform, image processing methods, decision-making methods. The following results were obtained. An image classifier has been developed with low sensitivity to increased database information. The properties of the developed classifier have been analyzed. The experiments demonstrated that clustering information based on images using the developed classifier proved to be sufficiently fast and cost-effective in terms of information volumes and computational power requirements.

Keywords: information systems; content-based image retrieval; image classifier; large databases; two-dimensional discrete cosine transform.

Introduction

Currently, much attention is paid to the collection and processing of information. A significant amount of information is contained in news feeds of electronic media. One of the ways to process such information is to cluster it according to its semantic load [1]. Typically, clustering of news information is carried out taking into account the morphological analysis of texts. It should be taken into account that in news feeds of electronic media, almost every news is accompanied by graphic content (photo materials), so it becomes possible to combine different news articles into clusters not only according to their content but also according to graphic accompanying information (the so-called “visual feature”). Image clustering plays the role of an additional feature in the text clustering algorithm, which sometimes plays an even more significant role than content clustering using keywords or other information features (cities, people, sports, or other characteristics that can be identified in the text and lead to the original word form).

It is important to take into account that the text can contain more than one image, which has both a negative and a positive effect. A negative effect may include the insertion of images that have little relevance to the analyzed texts (for example, company logos). The positive effect is that if the images correctly reflect the text, complex cross-clustering rules can arise that will allow clusters to be combined into digest groups containing the most complete information about the content (for example, an event or incident).

With such clustering, the following hypothesis can be formulated: the more unique the image, the higher the likelihood of “similarity” of the text content. In this case, headings, keywords, and other text information can be used as additional features for clustering.

Thus, the task of clustering images accompanying text information on various sites, including news feeds of electronic media, is relevant.

Literature analysis. Today there are many systems for image recognition. The most popular them are TinEye, Google Similar Images, and AntiDupl.NET. A common disadvantage of these systems is the inability to load an image gallery and create your database (DB) for work [2]. At the same time, there are a large number of image search methods, which differ in varying complexity and efficiency.

In works [3, 4], the authors proposed an invariant method and a model for quickly searching for digital images in data warehouses. However, the proposed model [3] has several disadvantages: the model is considered only for halftone images; The model only takes into account the shape of the histogram, and it has not been proven that different halftone images cannot have histograms of the same shape. The authors also indicated that on average it takes 1.7 seconds to search for one image in a database that contains more than 100 thousand images [4]. This speed is unacceptable in an information system in which about 10-12 thousand images need to be classified per day. This classification may take more than 5 hours.

One of the modern approaches is content-based image retrieval (CBIR), which plays an important role in finding images similar to the query image by extracting visual features [5, 6]. This approach is based on converting images into low-level functions that describe the images being analyzed.

Fig. 1 shows a generalized block diagram of CBIR systems, which includes three stages: feature extraction from the query image, feature selection, and then similarity matching [7, 8]. The CBIR system generates multidimensional feature vectors that are compared with image vectors in the database. According to the method of information processing, the CBIR system is divided into online and offline subsystems, which have the same feature extraction block (Fig. 1, [7]). To compare vectors, the corresponding similarity measures or distance measures are used: Euclidean distance [9, 10], Manhattan distance (city block metric) [11, 12], cosine similarity [13, 14], Mahalanobis distance [15], Hamming distance [16] and others. The resulting similarity score is compared with a threshold value that is pre-determined by the CBIR system.
As a rule, the image characteristics that make up multidimensional feature vectors are divided into global and local [17, 18]. Global features such as color, texture, shape, and spatial information, for example, provide insight into the entire image and enable faster feature extraction and similarity calculations. Although global features do not allow searching for objects in an image, they are well-suited for image classification. Local features are used to search for objects in an image, including in problems of pattern recognition and labeling of objects in an image [19, 20].

Although many CBIR methods have been developed, they all suffer from the disadvantage of the “semantic gap,” which is the gap between low-level features describing images (machine-parsed pixels) and high-level semantics (human perception), resulting in irrelevant searching for “similar” images in large databases, in which the rate of adding information reaches 10-12 thousand images per day.

To achieve this goal, it is necessary to solve the following tasks:
– develop an image classifier with low sensitivity to increased database information.
– carry out a study of the properties of the developed image classifier.

**Development of an image classifier**

The general model of an image classifier can be represented as a tuple
\[ IC = \{I, C, S, R, O\} \],
where \( I = \{I_1, I_2, \ldots, I_p\} \) is the set of images that need to be classified (collection of images); \( C = \{C_1, C_2, \ldots, C_k\} \) is a set of clusters (image classes), while \( C_i \cap C_j = \emptyset \) \( \forall i \neq j \); \( S = \{S_1, S_2, \ldots, S_k\} \) – is a set of image signatures; \( R \subseteq C \times S \) – relationship between clusters and signatures; \( O: I \rightarrow C \) – is a clustering operation, which consists of transforming images, after which either an image \( I_u \in I \) with a signature \( S_i \in S \) belongs to an existing cluster \( C_i \in C \), or a conclusion is made about the need to create a new cluster \( C_{k+1} \in C \) to which this image can be assigned, while one image can be assigned to only one cluster.

By image signature we mean a vector of feature values used for unambiguous image classification.

The relation \( R \) has the following property:
\[ \forall C_i \in C \exists S_j \in S \mid (C_i, S_j) \in R. \]

Based on the proposed model (1), we can conclude that the speed of the classifier is primarily affected by the method of calculating the image signature. The paper proposes to use the two-dimensional discrete cosine transform (2D-DCT) to compute the image signature.

In general, the 2D-DCT
\[ F[u, v] = \frac{2}{N} \cdot \alpha[u] \cos \frac{\pi(2u + 1)u}{2N} \cos \frac{\pi(2v + 1)v}{2N} \]

The analysis showed that existing methods for searching images in databases are of little use for solving the problem of clustering images accompanying text information on various sites, due to their low performance when it is necessary to search in large databases that are constantly updated.

**Problem statement and purpose of the study.** The purpose of the research is to improve the performance of searching for “similar” images in large databases, in which the rate of adding information reaches 10-12 thousand images per day.

Fig. 1. Generalized block diagram of CBIR systems [7]
where \( F[u,v] \) are the coefficients of the 2D-DCT (\( 0 \leq k \leq N - 1 \); \( 0 \leq l \leq M - 1 \)); \( \alpha[u] \), \( \alpha[v] \) are normalizing coefficients calculated using expression (3).

The inverse 2D-DCT can be calculated as follows

\[
A[x,y] = \left( \frac{2}{N} \right) \times \sum_{v=0}^{N-1} \sum_{u=0}^{N-1} \left( \alpha[u] \alpha[v] F[u,v] \times \frac{\cos (\pi (2x + 1)u / 2N) \cos (\pi (2y + 1)v / 2N)}{2} \right),
\]

where \( \alpha[u] \), \( \alpha[v] \) are normalizing coefficients calculated using expression (3).

The 2D-DCT has a number of useful properties that can be used when calculating an image signature. Let's look at these properties.

As a result of calculating the 2D-DCT of a matrix \( A \), the lowest frequencies will be located in the upper left part of the resulting matrix \( F \), and the highest frequencies will be located in the lower right part. Since the amplitude of low-frequency components is usually greater than that of high-frequency components, the upper left part of the display contains relatively large values that carry basic information about the image, and the lower right contains small values that carry information about details. Therefore, to increase the speed of signature calculation, it is advisable to use only a given small number of low-frequency components of the matrix \( F \). If the matrix \( A_m \) is a mirror image of the matrix \( A \) relative to the vertical axis, i.e., \( A_m[x,y] = A[N - 1 - x, y] \), then the 2D-DCT of the mirror image \( A_m \) has the following form

\[
F_m[u,v] = \left( \frac{2}{N} \right) \times \alpha[u] \alpha[v] \times \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} (-1)^x A[x,y] \cos \frac{\pi (2x + 1)u}{2N} \cos \frac{\pi (2y + 1)v}{2N}.
\]

If the matrix \( A_m \) is a mirror image of the matrix \( A \) relative to the horizontal axis, i.e., \( A_m[x,y] = A[N - 1 - y, x] \), then the 2D-DCT of the mirror image \( A_m \) has the following form

\[
F_m[u,v] = \left( \frac{2}{N} \right) \times \alpha[u] \alpha[v] \times \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} (-1)^y A[x,y] \cos \frac{\pi (2x + 1)u}{2N} \cos \frac{\pi (2y + 1)v}{2N}.
\]

Rotate an image on \( 90^\circ \) is equivalent to the simultaneous transposition and mirror reflection of the matrix \( A \) relative to the horizontal axis, i.e., \( A_{w90} = A_m \). Then the 2D-DCT rotated by \( 90^\circ \) the matrix \( A_{w90} \) has the following form

\[
F_{w90}[u,v] = \left( \frac{2}{N} \right) \times \alpha[u] \alpha[v] \times \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} (-1)^y A[y,x] \cos \frac{\pi (2x + 1)u}{2N} \cos \frac{\pi (2y + 1)v}{2N}.
\]

Thus, the property of rotated and mirror images can be formulated as follows

\[
\]

Property (5) is advisable to use to generate signatures that are invariant to rotations and mirror reflections of the image along any of the axes.

There are two different ways to implement the 2D-DCT. The first method is based on the direct calculation of monetary policy coefficients using expressions (2) – (4). However, this method requires significant computational resources, especially for large images. The second method uses a square discrete cosine transform matrix \( T \), which is calculated as follows

\[
T[k,l] = \begin{cases} \frac{1}{\sqrt{N}}, & \text{if } k = 0, 0 \leq l \leq N - 1; \\ \sqrt{2/N} \cos (\pi (2l + 1)k/2N), & \text{if } 1 \leq k \leq N - 1, 0 \leq l \leq N - 1. \end{cases}
\]

Then the 2D-DCT of the matrix \( A \) is calculated as follows

\[
F = T \times A \times T^T.
\]
Taking into account the placement of frequency components in the matrix $F_{(q)}$, we transform each of the matrices $F_{(q)x}, F_{(q)y}, F_{(q)z}, F_{(q)z}$ into the corresponding one-dimensional vectors $\vec{f}_x, \vec{f}_y, \vec{f}_z, \vec{f}_b$ (zigzag scanning of the matrix, starting from the upper left corner).

From each frequency component one can construct the following vector $\vec{f}_i = (f_{(q)}^1[i], f_{(q)}^2[i], f_{(q)}^3[i], f_{(q)}^4[i]),$ \( i = 0, M - 1 \), where $M \ll N^2$, i.e., it is proposed to take into account only the first $M$ quantized frequency components. Then the length of the vector $\vec{f}_i$ is calculated as $|\vec{f}_i| = \sqrt{f_{(q)}^1[i]^2 + f_{(q)}^2[i]^2 + f_{(q)}^3[i]^2 + f_{(q)}^4[i]^2}$. The decision rule of the classifier has the following form

$$ L_n \in \begin{cases} C_{k+1}, & \text{if } \exists \left( C_k, S_j \right): d(S_n, S_j) = \min_{p=0}^{K} d(S_n, S_p) < \varepsilon_k, 1 \leq k \leq K; \\ C_{k+1}, & \text{otherwise.} \end{cases} $$

Study of the properties of the developed classifier

To develop a fast image classification algorithm using the proposed method, let’s consider some of its properties. For experiments, we will take into account all DCT coefficients, i.e., quantization matrices $Q_p, Q_\xi, Q_\zeta, Q_\eta$ are matrices of units.

Characteristics of the data set for experiments: number of images $\approx 804724$; image sizes are different; all images are in color; the number of image clusters is 445999, with 172344 (38.64%) clusters containing 2 images each, 2 clusters containing 1792 (0.22%) and 6828 (0.85%) images, the number of images in the remaining clusters is shown in Fig. 2.

![Fig. 2. Structure of the data set, number of clusters containing: a – from 3 to 20 images; b – from 21 to 40 images; c – more than 40 images](image-url)
The matrix $T$ needs to be calculated once using expression (6). Research has shown that when an image is reduced to $64 \times 64$ pixel sizes (using the cubic interpolation method), the signatures of the original and reduced images do not change (Fig. 3, Table 1). In all experiments, the signature length is $M = 10$. As can be seen from the vector values $\bar{s} = (s_0, s_2, \ldots, s_9)$ in Table 1, the zero value is an order of magnitude greater than the values of all other elements of the vector, therefore the following vector was used as weighting coefficients in formula (10) $\bar{w} = (0.1, 1, 1, 1, 1, 1, 1, 1, 1, 1)$.

Experiments have shown that further reducing the size of the images leads to the fact that the signatures begin to differ slightly (Fig. 3).

Therefore, at the first stage of image preprocessing, the work proposes to bring all images to the same $64 \times 64$ pixel size. At the same time, the signatures retain their ability to identify specific images, for example, the signatures of images No. 1 and No. 2 differ significantly (Table 1) $d(\bar{S}_1, \bar{S}_2) = 103$. 

**Table 1 – Examples of signatures for images of different sizes**

<table>
<thead>
<tr>
<th>No.</th>
<th>Image</th>
<th>Image dimensions, pixels</th>
<th>Unnormalized signature vector $\bar{s} = (s_0, s_2, \ldots, s_9)$ according to (9)</th>
<th>Signature $\bar{S} = (S_0, S_2, \ldots, S_9)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="image1.jpg" alt="Image 1" /></td>
<td>$1000 \times 1500$</td>
<td>(302204, 26798, 26798, 23164, 14941, 23164, 7439, 13837, 13837, 7439)</td>
<td>(47, 42, 42, 36, 23, 36, 12, 22, 22, 12)</td>
</tr>
<tr>
<td>2</td>
<td><img src="image2.jpg" alt="Image 2" /></td>
<td>$64 \times 64$</td>
<td>(1579 2, 1401, 1401, 1210, 782, 1210, 388, 722, 722, 388)</td>
<td>(47, 22, 22, 44, 18, 44, 31, 19, 19, 31)</td>
</tr>
<tr>
<td>3</td>
<td><img src="image3.jpg" alt="Image 3" /></td>
<td>$448 \times 669$</td>
<td>(151201, 7120, 7120, 14308, 5998, 14308, 10071, 6079, 6079, 10071)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td><img src="image4.jpg" alt="Image 4" /></td>
<td>$400 \times 360$</td>
<td>(4888 0, 11665, 11665, 2642, 4948, 2642, 6838, 9627, 9627, 6838)</td>
<td>(20, 47, 47, 11, 20, 11, 28, 39, 39, 28)</td>
</tr>
<tr>
<td>5</td>
<td><img src="image5.jpg" alt="Image 5" /></td>
<td>$64 \times 64$</td>
<td>(824 4, 1969, 1969, 445, 836, 445, 1153, 1625, 1625, 1153)</td>
<td></td>
</tr>
<tr>
<td></td>
<td><img src="image6.jpg" alt="Image 6" /></td>
<td>$400 \times 360$</td>
<td>(1602 3 3, 4984, 4984, 7209, 509, 7209, 3515, 3097, 3097, 3515)</td>
<td>(75, 23, 23, 34, 2, 34, 16, 15, 15, 16)</td>
</tr>
<tr>
<td></td>
<td><img src="image7.jpg" alt="Image 7" /></td>
<td>$64 \times 64$</td>
<td>(2 6998, 835, 835, 1216, 89, 1216, 595, 523, 523, 595)</td>
<td>(33, 50, 50, 40, 12, 40, 14, 7, 7, 14)</td>
</tr>
<tr>
<td></td>
<td><img src="image8.jpg" alt="Image 8" /></td>
<td>$465 \times 620$</td>
<td>(142646, 21929, 21929, 17459, 5352, 17459, 5958, 3118, 3118, 5958)</td>
<td></td>
</tr>
<tr>
<td></td>
<td><img src="image9.jpg" alt="Image 9" /></td>
<td>$64 \times 64$</td>
<td>(17005, 2612, 2612, 2082, 636, 2082, 710, 371, 371, 710)</td>
<td></td>
</tr>
</tbody>
</table>

If in the decision rule (12) $\varepsilon_i = 0$, then completely identical images or images that cannot be visually distinguished will be collected in the same clusters. Fig. 4 shows graphs of the dependence of the number of clusters $K(\varepsilon)$ and the number of image classification errors $CE(\varepsilon)$ on the threshold value $\varepsilon$ (in each of the experiments the threshold value for all clusters $\varepsilon_i$ was set the same).

Analysis of the graphs shown in Fig. 4 showed that increasing the threshold above 40 leads to a sharp increase in image clustering errors, while the number of clusters sharply begins to decrease (i.e., clusters begin to merge).

Image analysis showed that at a threshold, $\varepsilon_i < 10$, similar images fall into one cluster. Examples of similar images are the images in Fig. 5.

In the case of fuzzy clustering, such images can be combined into one cluster using the developed classifier, because, in the future, text content is clustered not only based on images, but also other text parameters, such as news date, headline, keywords, and so on.
Quick searching in large databases

When a new image enters an information system (for example, a news portal), the following basic steps are performed for classification:

1) reducing the image size to $64 \times 64$;
2) calculation of the signature using expression (10) taking into account expressions (6)-(9);
3) classification according to the decision rule (12) taking into account the metric (11);
4) writing a new image to the database indicating an existing cluster, if one is found, or a new cluster, if the image is unique.

To proceed with step 3, the first task is to attempt to find an existing cluster in the database. We will consider the case when in the decision rule (12) for all clusters $\varepsilon_k = 0$.

Modern relational database management systems allow for fast searching through indexed fields. Then, to implement a quick search, it is necessary to perform indexing in the relational database by the field containing the signature. According to the proposed classifier model (1), a signature is a vector containing several values, so it is necessary to convert this vector into one field by which the database will be indexed and subsequently searched.

The work proposes to convert a vector $S = (S_{10}, S_{11}, \ldots, S_{M-1})$ into a string of the following form $S_{10}:S_{11}:\ldots:S_{M-1}$, i.e., write sequentially the elements of the vector, separated by a colon. For example, the value of the string field for image signature No. 1 from Table. 1 will look like 47:42:36:23:12:22:12. This representation makes it possible to record the signature in one field, regardless of the value of $M$ (the number of elements in the vector). Then the number of the cluster $N_{k\text{lasser}}$ with images whose signature $\text{imsig}$ matches the new image signature $\text{newimsig}$ can be obtained by a simple query to the table $\text{news\_images}$:

```
SELECT N_{k\text{lasser}} FROM news\_images
WHERE imsig=\text{newimsig}
```

This search will be completed in the shortest possible time for tables with an index field. Experiments showed that on average the classifier speed was 0.03 sec/image. Thus, when up to 10-12 thousand images are received per day, up to 5-6 minutes are spent on their clustering.

Conclusions

A general model of a content-based image classifier has been developed, which is based on transforming the image to its signature – a vector of feature values used for unambiguous image classification. Based on the proposed model, it is shown that the speed of the classifier is primarily influenced by the method of calculating the image signature. The paper proposes to use a two-dimensional discrete cosine transform to calculate the image signature. Based on a study of the
properties of the two-dimensional discrete cosine transform when analyzing the grayscale and color components of images, respectively (RGB color model), mathematical expressions for calculating the components of a signature that are invariant to rotations and mirror reflections of the image along any of the axes are obtained.

A method has been developed for classifying images by comparing their signatures, using the metric of city blocks as a measure of similarity.

The resulting method is the scientific novelty of this research.

A study of the properties of the developed image classifier was carried out, as a result of which the conversion of all images to the same 64 × 64 pixel size was justified.

At the same time, their signatures retain their ability to identify specific images, and the matrix of DCT coefficients is calculated once for all images, which significantly reduces the complexity of the method.

The experiments showed that clustering information from images turned out to be quite fast and low-cost in terms of information volumes and computing power requirements.

Further research is aimed at finding the optimal parameters of the proposed classifier, as well as studying the possibility of using the classifier for fuzzy clustering, as well as fuzzy search in the database.

REFERENCES


множества в базе данных, N.N. (2012), "Study of the effectiveness of clustering methods in the formation of...

матриц в..., Актуальність...

методов в цифровых изображениях, color, shape and texture descriptors and features", Arabian Journal for Science and Engineering,

методов в цифровых изображениях, color, shape and texture descriptors and features", Arabian Journal for Science and Engineering,

методов в цифровых изображениях, color, shape and texture descriptors and features", Arabian Journal for Science and Engineering,

методов в цифровых изображениях, color, shape and texture descriptors and features", Arabian Journal for Science and Engineering,

методов в цифровых изображениях, color, shape and texture descriptors and features", Arabian Journal for Science and Engineering,

методов в цифровых изображениях, color, shape and texture descriptors and features", Arabian Journal for Science and Engineering,