

Relevance feedback based on n-tuplewise comparison and the ELECTRE methodology and an application in content-based image retrieval

Pawel Rotter

Published online: 23 February 2013

© Springer Science+Business Media New York 2013

Abstract In this article we propose a method for information retrieval based on relational Multi-Criteria Decision Making. We assume that a user cannot define precise search criteria so that these criteria must be found based on the user's assessment of several sample alternatives ('alternatives' here are database records, e.g. images). This situation is common in Content-based Image Retrieval, where it is easier for a user to indicate relevant images than to describe a proper query, especially in formal language. The proposed algorithm for the elicitation of criteria is based on ELECTRE III—a method originally designed for ranking a set of alternatives according to defined criteria. In our algorithm, however, the direction of reasoning is reversed: we start with several sample alternatives that have been assigned a rank by the user and then we select criteria that are compatible (in the sense of ELECTRE methodology) with the user's preferences expressed on a sample set. Then, having determined the user's criteria, we apply classical ELECTRE III to retrieve the relevant solutions from the database. We implemented the method in Matlab and tested it on the Microsoft Cambridge Image Database.

Keywords Multiple criteria analysis · Relational MCDM · ELECTRE III · Preference elicitation · Content-based Image Retrieval

1 Introduction

Over the last two decades Content-Based Image Retrieval (CBIR) has become especially important because rapidly developing multimedia technologies have yielded a large number of databases containing graphic documents. The level of difficulty for visual information retrieval is much greater than for text-based search or for applications where textual information and meta-data are used along with the visual content [16].

P. Rotter (✉)
AGH University of Science and Technology, al. Mickiewicza 30, 30-059 Krakow, Poland
e-mail: rotter@agh.edu.pl
URL: <http://home.agh.edu.pl/~rotter>

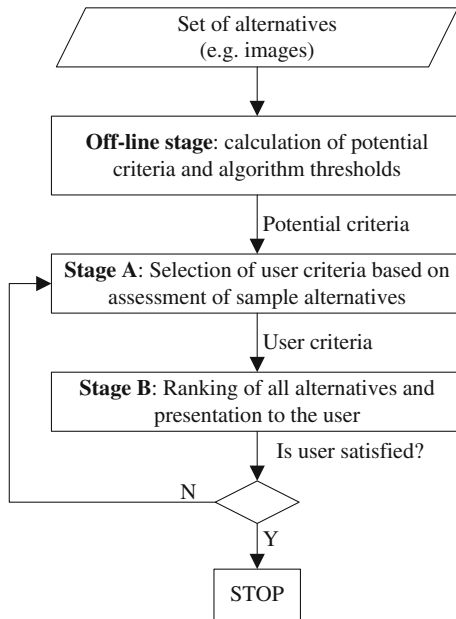
Early CBIR methods, developed mostly in 1990s but still used in many applications today, are based on the concept of *virtual query*—a point in feature space that is assumed to correspond to the image the user is looking for. Images in the database are ranked based on their distance from the virtual query in the feature space [5]. One of the main problems of CBIR is the gulf between low-level features that can be faultlessly extracted from images and the high-level concepts used by humans—the so-called ‘semantic gap.’ Addressing the semantic gap has been the target of much successful research in the last decade [21]. The semantic approach involves decomposing an image into objects and image descriptions based on the type of objects present, their features and spatial relations [17]. Images in the database are automatically annotated with high-level descriptors and are compared to the user query at the level of these descriptors [23]. In recent years there has been considerable progress in the recognition of high-level semantics and the performance of methods for the automatic semantic classification and annotation of images is good enough to make them practical, but still only in some limited fields of application [22].

Another way to tackle the semantic gap, instead of proposing high-level descriptors, is the development of advanced methods for the elicitation of the user’s semantic-based preferences as a function of non-semantic, low-level or middle-level descriptors. In these methods, local descriptors, used for example within the bag-of-visual-words framework, are generally more efficient than global features. In a method of this kind proposed by El Sayad et al. [1], a SURF descriptor is combined with an edge-context descriptor and a novel weighting scheme to substantially improve performance. Another example of successful, non-semantic retrieval are the descriptors used in 3D non-rigid shape recognition, such as those based on the geodesic distance [20]. These descriptors do not belong to high-level semantics but they can be efficient, when used together with adequate learning and classification methods [8]. In such a case, user preferences can then be approximated in descriptor space and a wide range of machine learning techniques can be applied [11], such as support vector machine [25] or RBF neural networks [13].

Apart from the high complexity of the matching stage, a significant barrier to the development of visual search systems is run into at the first stage of the search process: the elicitation of user preferences from relevance feedback, which is the problem addressed in this article. We propose an approach that exploits a large number of high-level image descriptors (referred to as *potential criteria*) and uses MultiCriteria Decision Making (MCDM) methodology to select those criteria that are consistent with user preferences (stage A in Fig. 1). This subset of potential criteria is called *user criteria*. In stage B the user criteria are used to find relevant images. A similar approach was proposed in [14], where the proposed method was based on the concept of reference sets [18, 19] and in [12] where the algorithm was based on the Analytic Hierarchy Process (AHP).

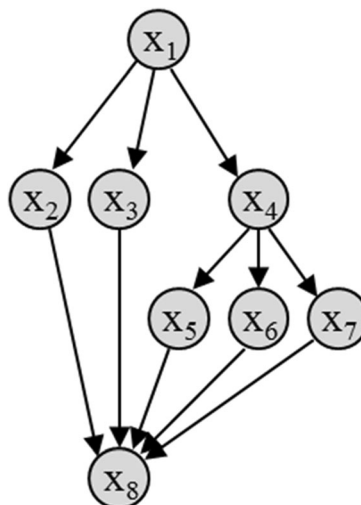
An important feature of our approach, which makes it different from other methods, is the general form used to define user search preferences. User input to the algorithm is any set of relations between sample images, represented by a graph of relations. In addition, the proposed method does not require as much user effort as the methodology based on AHP, which we proposed in [12], where the user is required to compare images pairwise. We use partially ordered sets for representing user criteria—as shown in Fig. 2, the method allows for incomparability of alternatives, e.g. the user does not need to define whether x_3 is more or less relevant to x_5 (although he/she may define it by adding an edge between x_3 and x_5).

Fig. 1 Scheme of the interactive information retrieval proposed in the article. Our method for user criteria elicitation applied on a sample set of alternatives is followed by the ranking all alternatives



The method for elicitation of user criteria proposed in this paper is based on the relational approach to Multicriteria Decision Making (MCDM), which exploits the relationships between available solutions [15]. In contrast to the functional approach, there is no assumption of a utility function, so there is no scalarization process. In this respect the relational approach can be regarded as more general than the functional approach. The concept is

Fig. 2 In our method the user provides a ranking of sample alternatives in form of graph



based on ELECTRE—a family of methods proposed in the 1960s¹ that are still widely used for the selection of optimal solutions and rankings [7]. As stated above, our algorithm for the selection of criteria is analogous to ELECTRE III but the direction of reasoning is reversed: the input to our algorithm are ranks defined by the user for several samples and the output are user criteria that are in concordance with the user's ranks. We will refer to this algorithm as *backward ELECTRE* in this article—see Fig. 3.

After finding the user criteria we apply ELECTRE III (the original, 'forward' version) in order to rank all the alternatives in the database, and then present the user with those that have the highest rankings. If the alternatives found in one iteration do not fully correspond to the user's preferences, they can serve as a starting point for the next iteration.

The main contribution of this article is therefore a new method for finding hidden user criteria from the ranks assigned by the user to sample alternatives and its application in content-based image retrieval with relevance feedback. The general scheme of the method is shown in Fig. 1 and the details of the algorithm are presented in Section 2. In Section 3 we present an application of the proposed algorithm to Content-based Image Retrieval (CBIR). In Section 4 we describe the experimental results.

2 Criteria elicitation and information retrieval based on ELECTRE methodology

In this section we present a method based on the relational MCDM approach for the selection of user search criteria from a large set of the pre-defined features. In our previous work [12] we proposed a method for the elicitation of user preferences and information retrieval based on pairwise comparison, or, more precisely, on the Analytic Hierarchy Process (AHP). The methods proposed in [12] and in this paper are based on a similar concept. There are well-established methods for ranking alternatives, such as AHP or ELECTRE, which are based on the importance of criteria given as an input. In contrast, the first step in information retrieval based on relevance feedback requires the elicitation of these criteria through user assessment of several sample alternatives (which can be expressed by ranking). There is, therefore, a need for an opposite process that can answer the question: if we use a standard method for multicriteria ranking, which set of criteria would lead to the same or at least a similar ranking of sample alternatives as the ranks assigned by the user?

From the point of view of the user, the main difference between the approach based on AHP and the one presented here is the number of alternatives that need to be compared. In AHP, alternatives are compared in pairs, so in order to provide a complete comparison in a set of K alternatives, the user needs to assess the relative relevance of alternatives for each of $K(K-1)/2$ pairs. This method ensures data redundancy, which is useful for checking the consistency of user preferences, but providing a large amount of data is not convenient for the user.

The problem of automatic elicitation of ELECTRE parameters from assignment examples was addressed in [9]. The authors proposed a parameter elicitation method for *sorting problem* and implemented it as so-called ELECTRE-TRI assistant [10]. In contrast, we use ELECTRE III as a basic algorithm. Consequently, the set of sample relations should be provided by the user as a ranking in the same form as the output ranking of ELECTRE III, where the incomparability of some alternatives is acceptable (Fig. 2). This form of

¹ The first ELECTRE method (ELECTRE I) was published in [14]

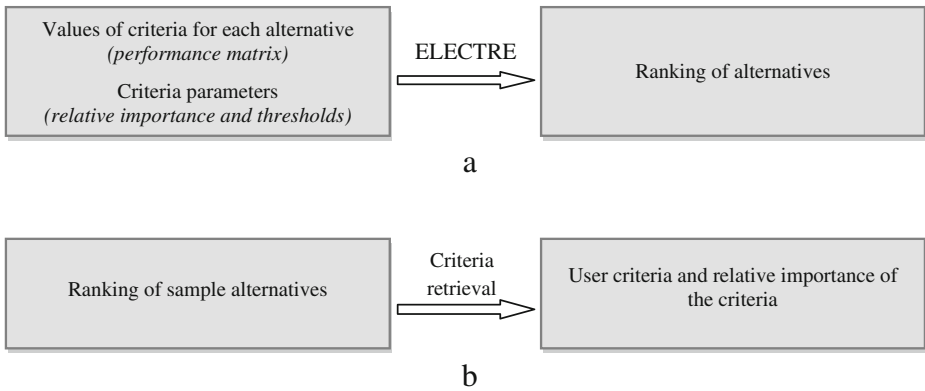


Fig. 3 Input and output data in ELECTRE (a) and the proposed algorithm—backward ELECTRE (b)

expressing user preferences is more general than using utility values: the user may leave some relations undefined (while if utility values are assigned to alternatives, they implicitly define relations for all the pairs). The details of the proposed approach are discussed in Section 2.2. Unlike Mousseau and Slowinski [9] we are not proposing a method for finding a complete set of ELECTRE parameters. Instead we focus on calculating those parameters that define user preferences in order to retrieve relevant information from the database. We assume that there is a large set of *potential criteria* and the algorithm should select a fraction of this set that is compatible with actual *user criteria*. We seek only those parameters of ELECTRE that are directly related to user preferences—means the set of criteria and their relative importance, while we do not provide solution how to set other parameters of ELECTRE.

2.1 The ELECTRE III method

ELECTRE III was proposed about 40 years ago, but is still probably the most popular ranking method, besides AHP, that exploits the relational approach—which means that there is no assumption that the utility function that aggregates all the criteria exists. Although the input to the ELECTRE algorithm includes criteria weights, they are interpreted in a different way than is the case in the functional approach: i.e. they are not used as multipliers to build the weighted sum of criteria, they reflect the voting power of the criteria instead, so they are also referred to as *relative importance* of criteria.

The method belongs to the group of outranking methods, i.e. it is based on outranking relation $S: x_1 S x_2$ means that x_1 is at least as good as x_2 .

Three other types of relations are used as well as outranking relations:

- strong preference: for the criterion f_j , the alternative x_1 is strongly preferred to x_2 ($x_1 P_j x_2$) iff $f_j(a) - f_j(b) > p_j$, where p_j is the preference threshold for the criterion f_j ;
- weak preference: for the criterion f_j , the alternative x_1 is weakly preferred to x_2 ($x_1 Q_j x_2$) iff $q_j < f_j(a) - f_j(b) \leq p_j$, where q_j is the indifference threshold;
- indifference: for the criterion f_j , the alternative x_1 is indifferent to x_2 ($x_1 I_j x_2$) iff $|f_j(a) - f_j(b)| \leq q_j$.

The relation $x_1 S x_2$ holds if either $x_1 P x_2$ or $x_1 Q x_2$ or $x_1 I x_2$.

Apart from the thresholds p_j and q_j , there is a *veto threshold* v_j associated with each criterion, which reflects the influence of this criterion on the final result when it is discordant with other criteria. If for the j -th criterion the alternative x_1 is preferred to x_2 and the difference in values of this criterion exceeds v_j , the assertion $x_2 \mathbf{S} x_1$ is rejected regardless of other criteria.

Below, we present a short description of the method.

In the first step, the concordance index $c_j(x_1, x_2)$ and discordance index $d_j(x_1, x_2)$ are calculated, which measure the strength of the assertion $x_1 \mathbf{S}_j x_2$:

- $c_j(x_1, x_2) = 1$ if $x_1 \mathbf{S}_j x_2$; $c_j(x_1, x_2) = 0$ if $x_2 \mathbf{P}_j x_1$, otherwise $c_j(x_1, x_2)$ varies linearly between 0 and 1:

$$c_j(x_1, x_2) = \begin{cases} 1 & \text{if } f_j(x_1) + q_j \geq f_j(x_2) \\ 0 & \text{if } f_j(x_1) + p_j \leq f_j(x_2) \\ \frac{p_j + f_j(x_1) - f_j(x_2)}{p_j - q_j} & \text{otherwise} \end{cases} \tag{1}$$

- $d_j(x_1, x_2) = 0$ if $x_2 \mathbf{P}_j x_1$; $d_j(x_1, x_2) = 1$ if $f_j(x_1) + v_j \leq f_j(x_2)$, otherwise $d_j(x_1, x_2)$ varies linearly between 0 and 1:

$$d_j(x_1, x_2) = \begin{cases} 0 & \text{if } f_j(x_1) + p_j \geq f_j(x_2) \\ 1 & \text{if } f_j(x_1) + v_j \leq f_j(x_2) \\ \frac{f_j(x_2) - f_j(x_1) - p_j}{v_j - p_j} & \text{otherwise} \end{cases} \tag{2}$$

The joint concordance matrix, for all criteria, is calculated as:

$$C(x_1, x_2) = \frac{\sum_j w_j c_j(x_1, x_2)}{\sum_j w_j} \tag{3}$$

and the outranking matrix is given as:

$$S(x_1, x_2) = \begin{cases} C(x_1, x_2) & \text{if } d_j(x_1, x_2) \leq C(x_1, x_2) \quad \forall j \\ C(x_1, x_2) \prod_{j \in J(x_1, x_2)} \frac{1 - d_j(x_1, x_2)}{1 - C(x_1, x_2)} & \text{otherwise} \end{cases} \tag{4}$$

where $J(x_1, x_2)$ is the set of criteria such that $d_j(x_1, x_2) > C(x_1, x_2)$.

The matrix S is then binarized with a threshold Θ , which depends on the value of the largest element of S :

$$T(x_1, x_2) = \begin{cases} 1 & \text{if } S(x_1, x_2) > \Theta \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

Based on the matrix T , a ranking of alternatives is created using the so-called distillation process. For a detailed description of the method see e.g. [3], chapter 3.2.2. The input and output data for the algorithm are summarized in Fig. 4.

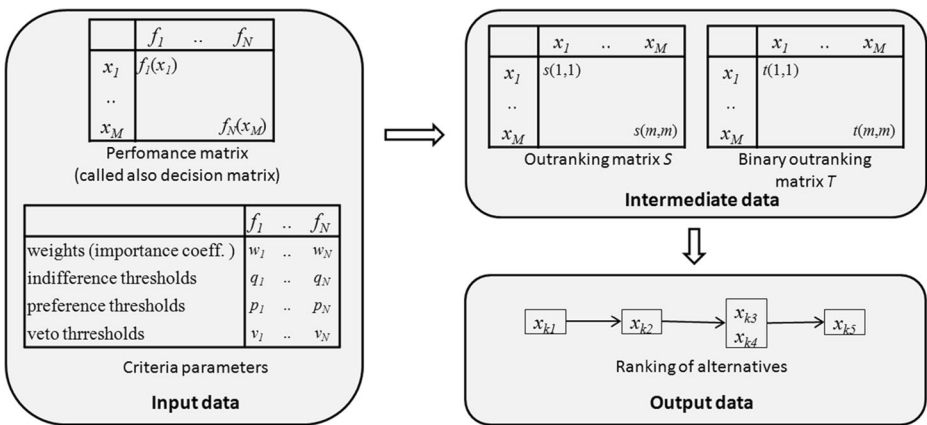


Fig. 4 Input and output data in ELECTRE III

2.2 Proposed method for criteria elicitation and information retrieval

The general scheme of the proposed method is presented in Fig. 1 and a detailed diagram in Fig. 5. The retrieval algorithm uses some data that is prepared off-line and saved as part of the database. The method is, therefore, composed of 3 stages:

- **Off-line stage: calculation of potential criteria values and ELECTRE thresholds**
 Calculation of potential criteria for all alternatives in the database and ELECTRE thresholds (veto threshold, preference threshold and indifference threshold). In some applications, the value of the criteria can be given as problem input data but, in an application for visual search for example, the criteria are calculated by processing of database images. ELECTRE thresholds for criteria can be established by the user or calculated based on statistics of criteria values.
- **Stage A: Selection of user criteria**
 Selection of those criteria from the pre-defined criteria set that are consistent with user preferences, in sense of ELECTRE methodology. The stage is analogous to ELECTRE, but the direction of reasoning is reversed: the set of ranks that define the relationships between alternatives is the input, while the set of criteria and their relative importance is the output. We refer to this stage as *backward ELECTRE*.
- **Stage B: Ranking alternatives**
 Calculation of the ranking of all alternatives in the database with *ELECTRE III*, using the criteria found in the first stage. In contrast to the proposed backward method, we refer to the classical ELECTRE algorithm as *forward ELECTRE*. The top alternatives are presented to the user.

An interactive loop allows the user to refine a query using intermediate results. We use the following notation in our description of the algorithm:

- X - set of all alternatives in the database
- X_s - subset of alternatives presented to the user

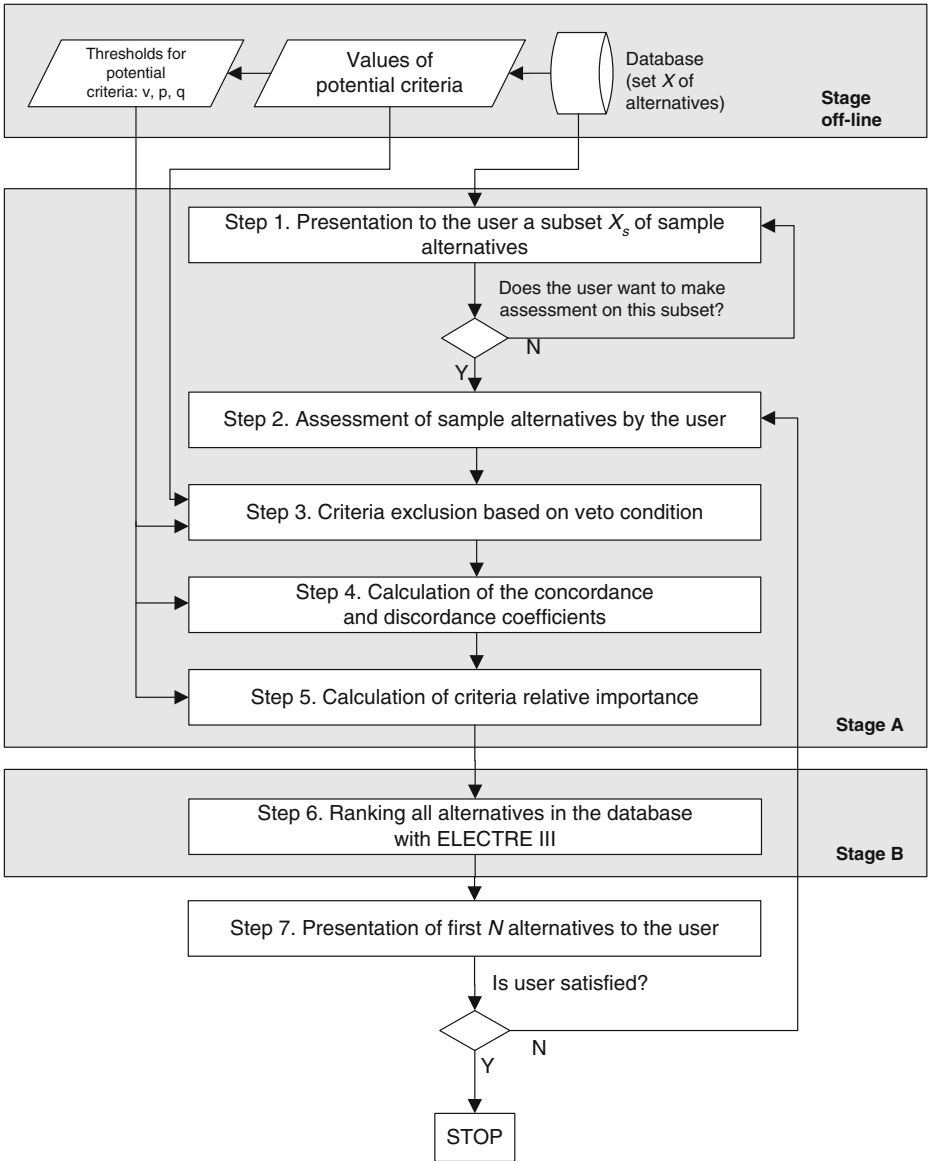


Fig. 5 Diagram of the proposed algorithm

- F - set of all potential criteria
- F_s - set of user criteria (subset of F calculated by backward ELECTRE)
- q_j - indifference threshold for j -th criterion: if the criteria values of two alternatives differ by less than q_j , they are considered indifferent to the user
- p_j - preference threshold: if the criteria values of two alternatives differ by less than p_j , they are considered indifferent to the user or one is weakly preferred; otherwise one is strongly preferred to the other
- v_j - veto threshold

- v_j - veto threshold for backward ELECTRE algorithm: it is used for the exclusion of j -th criterion from the set of user criteria if, for *any* pair of sample images, j -th criterion is too inconsistent with a relation defined by the user
- r_j - range of j -th criterion (the difference between maximum and minimum value of the j -th criterion over all alternatives)
- w_j - relative importance of j -th criterion
- c_j, d_j - concordance and discordance matrices for j -th criterion (calculated as in the original ELECTRE III algorithm)

The algorithm (stages A and B) can be described in the following steps:

Step 1. Presentation of sample alternatives to the user

From the set X of alternatives in the database a subset X_s containing N alternatives, is randomly chosen and presented to the user. If the user cannot find any relevant images among them, another set of N alternatives is presented.

Step 2. Defining user preferences on sample set of alternatives

We propose a combination of two ways of defining user preferences:

- By assigning a value Q to each alternative (for example: 0—irrelevant, 5—very relevant). This is a common way of defining user preferences in CBIR systems [14]. The values of Q define the preference relation $x_1 \mathbf{P} x_2 \Leftrightarrow Q(x_1) > Q(x_2)$.
- By defining a graph of sample preferences. The user ranks sample alternatives by sketching a graph of preference relations, which has the same structure as the ranking graph produced by the ELECTRE method—like the graph in Fig. 2. An arrow from alternative x_m to x_n means that $x_m \mathbf{P} x_n$.

Approach (a) is compatible with the functional rather than the relational approach to decision making. Using a graph of sample preferences is more general. The preference structure defined by method (a) can always be converted to a preference graph but not vice versa. Approach (b) allows the user to define relations for selected pairs of alternatives only, while approach (a) enforces an implicit definition of all relations. In the example in Figure 2 the user may not be willing to define relations between x_3 and x_4 or between x_3 and x_5 .

Approach (a) is, however, quicker and sometimes more convenient, at least for a rough definition of user preferences. Therefore, in our algorithm, the user may choose first to approximate his/her preferences using method (a) and then refine them by adding/removing relations in the corresponding preference graph (see Fig. 6).

If the step has been performed in previous iterations with different sample alternatives, the graph provided by the user is merged with the graph defined previously.

Step 3. Criteria exclusion based on the veto condition

For each potential criterion $f_j \in F$ if for *any* pair of alternatives $(x_m, x_n) \in X_s^2$, $m \neq n$, such that $x_j \mathbf{P} x_2$, it holds that:

$$f_j(x_m) < f_j(x_n) - v_j, \quad (6)$$

the criterion f_j is excluded from the set F_s regardless of the values for other alternatives.

Inequality (6) means that the j -th criterion for the pair of alternatives (x_m, x_n) is too inconsistent (exceeds the veto threshold) with the relation defined by the user to be considered as a user criterion.

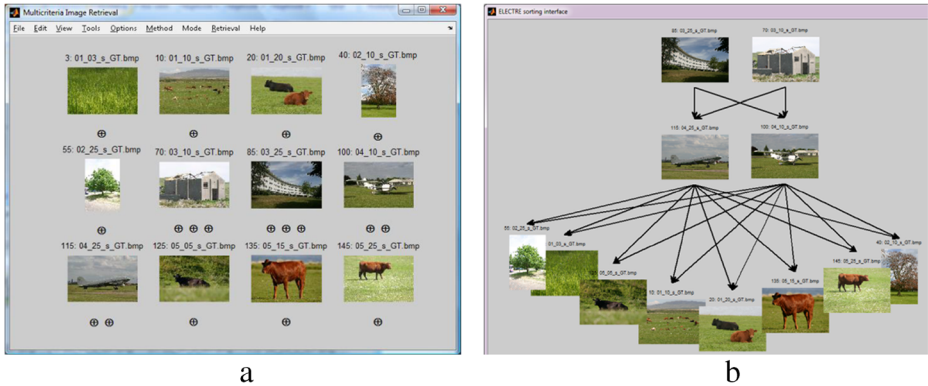


Fig. 6 **a** Starting set of images with preferences defined by the user by assigning ranks; the symbols under the pictures describe the ranks assigned by the user by clicking the image: \downarrow —non-relevant, \oplus —neutral, $\oplus\oplus$ —relevant, $\oplus\oplus\oplus$ —very relevant. **b** the corresponding graph of relations of relations automatically generated from ranks. Our prototype software allows for interactive modification of this graph

Step 4. Calculation of the concordance and discordance coefficients

The concordance and discordance coefficients $c_j(x_m, x_n)$ and $d_j(x_m, x_n)$ are calculated according to formulas (1) and (2) for all criteria that were not excluded in Step 3, for each pair of sample alternatives that were ranked by the user.

Step 5. Assignment of relative importance of the user criteria

For the criteria that were not excluded in step 3, a criteria's relative importance is calculated according to the formula:

$$w_j = \sum_{m,n : x_m \mathbf{P} x_n} c_j(x_m, x_n). \quad (7)$$

Relative importance coefficients are then normalized to [0, 1].

Heuristic equation (7) means that the importance of the j -th criterion is high if the relation between the values of this criterion (expressed in coefficients c_j) is compatible with relations between sample alternatives, which were defined by the user in Step 2.

The experiments described in the next section confirm that the heuristic rule proposed in this paper allows for the efficient retrieval of user criteria. Nevertheless, there is no guarantee that the original (forward) ELECTRE III algorithm run on sample data presented to the user would lead to exactly the same ranking as defined by the user. Some steps of ELECTRE cannot be reversed, especially binarization (equation (5)).

Note that the criteria relative importance are later used according to the principles of the relative approach. They should not, therefore, be interpreted as the weights in the sense of the weighted sum scalarisation method, which is typical in the functional (based on utility function) approach.

Step 6. Ranking of alternatives—forward ELECTRE

The set of user criteria F_s with relative importance coefficients w_j is used as the input for the ELECTRE III algorithm in order to rank all alternatives in the database.

Step 7. Presentation of the results

N highest ranked alternatives are presented to the user. If the user is not satisfied, the presented alternatives can be used as a sample set for the next iteration, which starts from Step 2.

3 Application to content-based image retrieval

In this section we describe how the methodology presented in Section 2 can be applied to Content-based Image Retrieval. Classical CBIR systems mostly use low-level descriptors. After updating the feature vector of a query based on interaction with the user, various classification methods (such as Bayesian rule, SVM and neural networks) are used to retrieve relevant images from the database. See [4] for a review of these methods. The term ‘query’ in systems such as these is usually used not for an explicit inquiry but for the point in descriptor space that represents an ‘ideal’ but non-existent image sought by the user [6].

In contrast, we use high-level features that can be considered as potential user criteria so that MCDM methods can be applied. By ‘query’ we mean the set of user criteria: the query is not explicitly expressed by the user and our goal is to arrive at it from a user’s assessment of sample images. We assume that the classification of objects has been performed, i.e. each pixel is assigned to the category (class) of object that it represents. For small databases, e.g. in some e-commerce applications, the classification of objects can be assisted by humans in order to avoid the errors that are still relatively common in fully automatic systems. There are also methods dedicated to specific tasks that operate on specific types of images with generally much lower error rates. Examples are biometric identification systems (retrieval is based on specific features of faces, fingerprints, etc.) or niche applications such as the atlas of fish species that we proposed in [14]. For databases that are not limited to a specific class of images the automatic classification of objects remains a difficult task, though a lot of research is being done in this field and many successful methods are presented every year at the PASCAL Visual Object Classes Challenge [2]. The problem of the automatic detection and classification of objects in images is not addressed in this paper. For our tests we used the Microsoft Research Cambridge Image Database version 2 (MSRC-v2), in which each pixel is assigned to one of 23 categories of object.² The database contains 591 images and several samples are shown in Fig. 7. Screenshots come from the application that we have developed in Matlab 7.1 to test our method.

We set the number of sample images at $N=12$, which is enough to ensure sufficient diversity among the samples. Setting N too high can make comparison difficult, especially when the set of sample images does not fit on the screen or the images become too small to be properly judged by the user. If the initial set of sample images is too small and it does not contain relevant images, another sample set is presented.³ We calculated ELECTRE parameters for each criterion in relation to the criterion range. We led a series of experiments focused on the correct elicitation of user criteria to find the coefficients, which determine the relationship between the range r_j of j -th criterion and the corresponding ELECTRE parameters. After manual tuning we achieved good results for the values: $q_j=0.04r_j$, $p_j=0.08r_j$, $v_j=0.5r_j$, $v_j=0.1r_j$. Parameter Θ is set to 0.85λ , where λ is the largest element of the outranking matrix S . The method does not require very precise tuning of coefficient values, we observed that variations of the coefficient values up to 20–30 % do not influence noticeably the method performance. Since only a fraction of pre-defined features are used as criteria for ranking database images, we can define a large set of *potential* criteria, which refer to the features of objects in the images. For each category of objects we consider 8 criteria:

² The database is available for non-commercial purposes at: <http://research.microsoft.com/en-us/projects/objectclassrecognition/default.htm>

³ In our interface N is fixed but the proposed method can be used with a larger set of sample images and the value of N does not influence significantly the computational complexity of the algorithm.

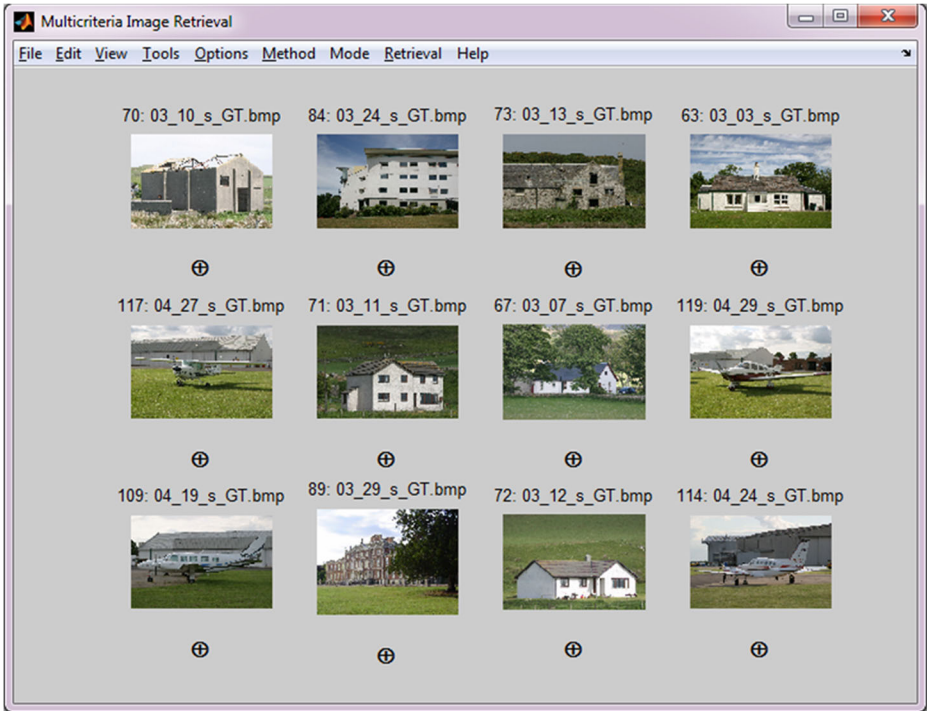


Fig. 7 Images retrieved in the first iteration based on preference structure shown in Fig. 6

- Total area of object divided by area of image
- Horizontal spread ($x_{\max}-x_{\min}$) of the union of all patches (segments) of the object in relation to the horizontal resolution of the image
- Vertical spread ($y_{\max}-y_{\min}$) of the union of all patches in relation to the vertical resolution of the image
- Aspect ratio (vertical spread divided by horizontal spread) of the union of all patches
- Horizontal spread of the largest patch in relation to the horizontal resolution of the image
- Vertical spread of the largest patch in relation to the vertical resolution of the image
- Aspect ratio (vertical spread divided by horizontal spread) of the largest patch
- Number of patches of the object in the image

Each of the above criteria also has a reversed version, i.e. any monotonically decreasing function of the original criterion. There are 23 labelled classes of objects in the database, so the total number of potential criteria is $23 \cdot 8 \cdot 2 = 368$. There are usually several objects in a single image, and for most images there are between 32 (for two objects in the image) and 64 (4 objects) criteria with a value between 0 and the maximum. For the rest of the criteria, when the corresponding object is not present in the image, the value is either 0 for non-reverse criteria or maximum over all images in the database for reverse criteria.

Some criteria are present in a large number of images but are rarely used by the user, for example those related to the presence, size and shape of the sky or grass. In the example presented in Fig. 8 the user indicated that two images with airplane were relevant, and both of them contain large areas of sky and grass. The algorithm,

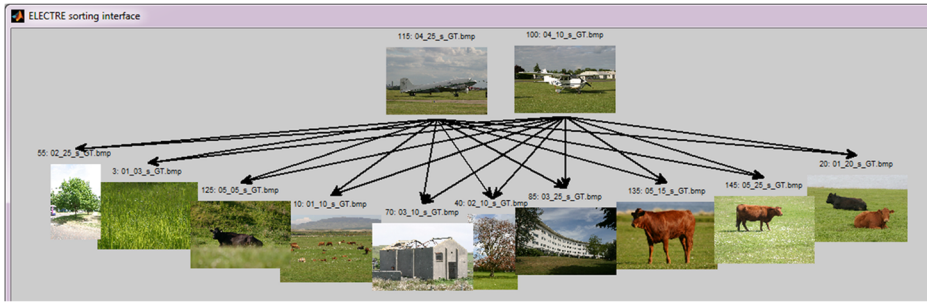


Fig. 8 The user marked two airplane images as the most relevant but because both of them contained more sky than other images presented to the user, the area of the sky become one of the retrieval criteria

therefore, found many more criteria than those that resulted from the actual user preferences (the criteria list is shown in Fig. 10) and some images were consequently highly ranked only because they contained large areas of sky (Figs. 9 and 10).

We therefore introduced an option to decrease the relative importance of these criteria by a constant factor or to zero. The elimination of these criteria or (depending on the application) a reduction on their importance, generally improves performance. The retrieval results for the example from Fig. 8 after this modification are shown in Fig. 11.

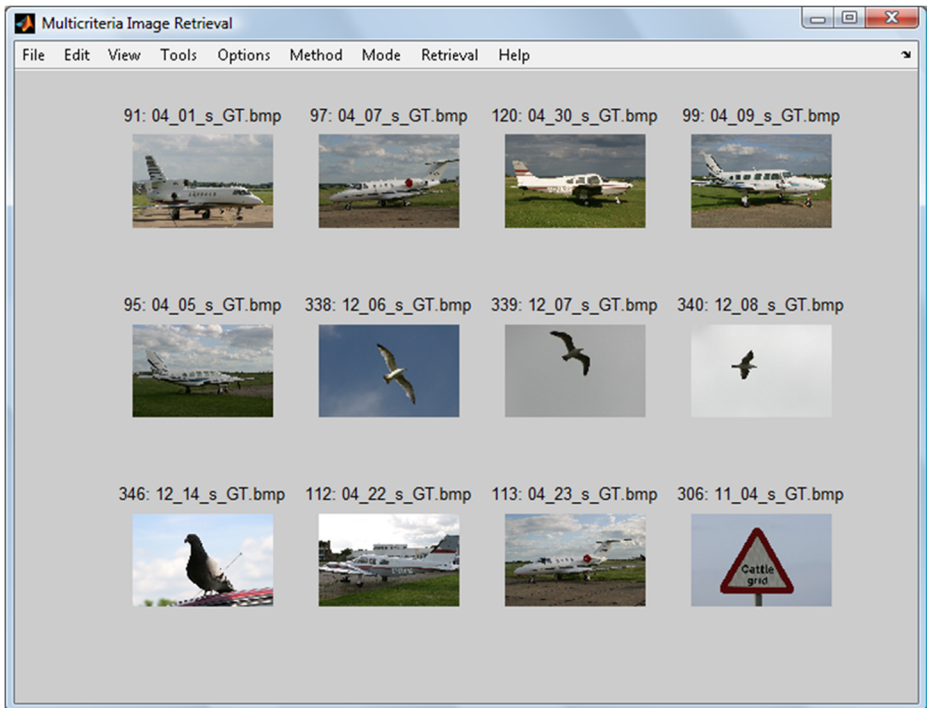


Fig. 9 First 12 images retrieved by the algorithm based on user preferences shown in Fig. 8

```

22 criteria were found:
Criterion 7: w=0.4 Object: building. Feature: aspect ratio - union of all patches
Criterion 13: w=0.4 Object: building. Feature: aspect ratio of the biggest patch
Criterion 15: w=0.4 Object: building. Feature: number of patches
Criterion 19: w=0.1 Object: grass. Feature: left-right spread (union of all patches)
Criterion 23: w=0.1 Object: grass. Feature: aspect ratio - union of all patches
Criterion 25: w=0.2 Object: grass. Feature: left-right spread of the biggest patch
Criterion 29: w=0.05 Object: grass. Feature: aspect ratio of the biggest patch
Criterion 97: w=1 Object: sky. Feature: total area of object / area of image
Criterion 99: w=0.6 Object: sky. Feature: left-right spread (union of all patches)
Criterion 101: w=0.7 Object: sky. Feature: top-bottom spread (union of all patches)
Criterion 103: w=0.65 Object: sky. Feature: aspect ratio - union of all patches
Criterion 105: w=0.8 Object: sky. Feature: left-right spread of the biggest patch
Criterion 107: w=0.8 Object: sky. Feature: top-bottom spread of the biggest patch
Criterion 109: w=0.5 Object: sky. Feature: aspect ratio of the biggest patch
Criterion 129: w=1 Object: airplane. Feature: total area of object / area of image
Criterion 131: w=1 Object: airplane. Feature: left-right spread (union of all patches)
Criterion 133: w=1 Object: airplane. Feature: top-bottom spread (union of all patches)
Criterion 135: w=1 Object: airplane. Feature: aspect ratio - union of all patches
Criterion 137: w=1 Object: airplane. Feature: left-right spread of the biggest patch
Criterion 139: w=1 Object: airplane. Feature: top-bottom spread of the biggest patch
Criterion 141: w=1 Object: airplane. Feature: aspect ratio of the biggest patch
Criterion 143: w=1 Object: airplane. Feature: number of patches

```

Fig. 10 Criteria found by reverse ELECTRE algorithm based on user preferences from Fig. 8

4 Performance and efficiency

As noted in [24], evaluating the performance of CBIR systems is much more complex than is the case for text-based information retrieval. Standard performance measures for information retrieval systems, such as precision and recall, are also applied in CBIR but they cannot be considered fully objective; all performance measures are based on user assessment of satisfaction, which varies depending on the user, task and application.

We used precision-recall charts to estimate the performance of our method for several different queries. We allowed two iterations for all. When the initial set of sample images did not allow the user to assign diversified ranks (all the sample images were similarly relevant/irrelevant to the user's query), another initial set of randomly chosen images was presented to the user.⁴ We tested the method using 5 queries. Ten users participated in our experiments—colleagues of the author and the author himself. There were 8 scientists (6 man and 2 woman) and two IT male technical specialists among the participants, age between 33 and 42. All participants were proficient in IT technology, 6 of them had some experience in image processing and one (the author) in CBIR. Every query was judged by all the users and the results were averaged. The users were looking for images containing:

- Q1: airplanes and buildings
- Q2: animals: sheep, cat or dog on a road
- Q3: large area of sea with a hilly coastline
- Q4: close-up of a sheep
- Q5: slim trees

⁴ In CBIR systems with large databases the *page zero problem* is reduced by applying textual search to obtain the initial set of images: Grigorova, A. and F. D. Natale (2006).

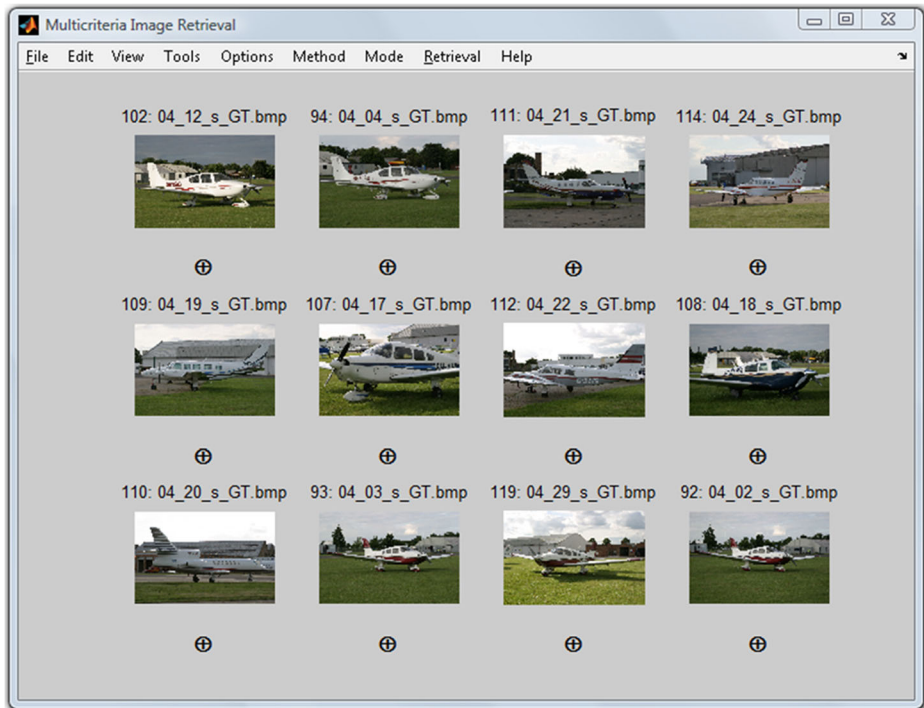


Fig. 11 Results after criteria related to sky and grass are automatically eliminated (criteria: 19, 23, 25, 29, 97, 99, 101, 103, 105, 107, 109 in Fig. 10)

The users were not provided with any additional information about the queries.

The results are presented in Fig. 12.

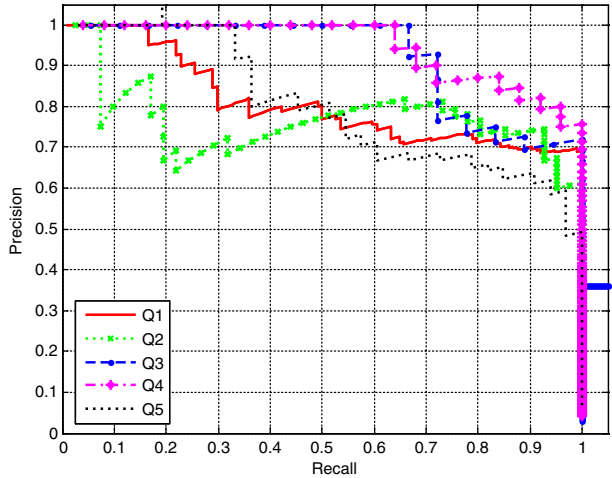
The score is highly dependent on the complexity of the user query. The high performance for query Q4 results from the fact that, in this case, only one criterion was important: the relative size of the area classified as ‘sheep,’ while in all the other cases the user preferences required combinations of several basic criteria.⁵ In query Q2, which achieved the lowest performance, the number of criteria involved is the largest.

In Fig. 13 we present a comparison of the average performance of our method and those of several existing algorithms. All methods were tested on the same input features (potential criteria mentioned in Section 3). Reversed criteria, i.e. monotonically decreasing functions of the original criteria were excluded for the methods were they would not work correctly (algorithm based on virtual query and RFB neural networks).

The comparison with the algorithm proposed in our previous paper [12], based on the Analytic Hierarchy Process, is particularly interesting, because both methods exploit the MCDM approach. Apart from the area of high recall values (above 0.75), which is usually not very important for the user, both methods demonstrate

⁵ The actual number of criteria found by the algorithm is greater, e.g. when looking for large objects, found criteria are: area, width, height, etc.

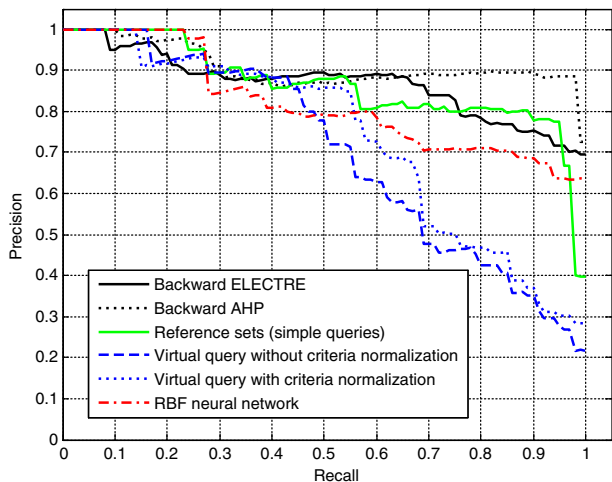
Fig. 12 Precision vs. recall for 5 different queries: Q1—airplanes and buildings; Q2: animals: sheep, cat or dog on a road; Q3: large area of sea with a hilly coastline; Q4: close-up of a sheep; Q5: slim trees



similar performance. The AHP-based method slightly outperforms the method based on ELECTRE. However, the main advantage of the ELECTRE-based method is not its precision-recall score but the compatibility with the user interface that allows for quicker and more convenient expression of the user preferences than in the case of the method based on AHP. The method based on reference sets [14] also performs relatively well, but an important limitation of this method is that it is not designed for complex queries, especially where a user query contains alternative of several criteria. The plot for reference sets in Fig. 13 is limited to queries Q4 and Q5. The performance of the other methods, based on virtual query is similar to the proposed method for small recall values (below 0.4), but our method performs significantly better for higher recall values.

The last method compared is Radial Basis Function (RBF) neural network with one RBF layer and linear output layer. The number of RBF neurons is adjusted so as to

Fig. 13 Average precision vs. recall for the proposed algorithm compared to several existing methods



keep the mean square error below the threshold (so it may differ for different queries) and the number of inputs to each RBF neuron is equal to the number of potential criteria. RBF neural network performs better than the algorithms based on virtual query, especially for high recall values but it is outperformed by all three methods based on MCDM.

Regarding the efficiency of the method, a large part of the necessary computations can be done off-line and only need be done once for a database:

- database preparation, such as the calculation of image features, image segmentation, calculation of object descriptors;
- calculation of potential criteria for all images in the database;
- calculation of criteria range and thresholds: q_j, p_j, v_j, \cup_j ;
- calculation of concordance and discordance matrices.

The amount of memory used by the algorithm is determined by the size of the concordance and discordance matrices. Each of them has $n_A^2 n_C$ elements, where n_A is the number of images in the database and n_C is the number of potential criteria. In our case $n_A=591$ and $n_C=368$, so both matrices together occupied 245 MB,⁶ which allows for the software to be run on a standard PC without the need for memory swapping. Storing concordance and discordance matrices in memory allows more calculations to be done off-line, but it is not necessary. For larger databases:

- in stage A (Fig. 1) concordance and discordance coefficients can only be calculated for the set of sample images presented to the user,
- in stage B the values $c_j(x,y)$ and $d_j(x,y)$ can be calculated when needed. Then the biggest variable stored in memory is binary matrix T , which occupies n_A^2 bits of memory. This may become a limitation if the database size exceeds tens of thousands of images.

5 Conclusion and future work

In this article we have proposed a method for the elicitation of user search criteria based on the relational MCDM methodology. The method is analogous to the ELECTRE algorithm, but the direction of reasoning is reversed: user criteria and their relative importance are found based on user assessment of sample alternatives. We have demonstrated that our approach can be an effective tool for the multicriteria retrieval of information, in particular for Content-based Image Retrieval.

Our goal was to develop a prototype that would demonstrate and test our methodology rather than to create a complete CBIR system. In order to develop an application that could operate on any database of images without preliminary processing, automatic object classification would have to be integrated with the algorithm. The set of potential criteria that we used could be enlarged and include, for example, the location of objects in the image and relationships between objects. These extensions could be developed within the methodological framework proposed in this article.

⁶ We used 1 byte per element (uint8 type, with scaling between [0..1] and [0..255])

Acknowledgments This work was supported by the Polish Ministry of Science and Higher Education under SIMPOZ project, no. 0128/R/100/2010/12. We thank to our colleagues who participated in experiments and to anonymous reviewers for many valuable comments and suggestions.

References

1. El Sayad I, Martinet J, Urruty T, Djeraba C (2012) Toward a higher-level visual representation for content-based image retrieval. *Multimed Tool Appl* 60:455–482
2. Everingham M, Gool LV, Williams CKI, Winn J, Zisserman A (2010) The Pascal Visual Object Classes (VOC) Challenge. *Int J Comput Vis* 88(2):303–338
3. Figueira J, Mousseau V, Roy B (2005) ELECTRE methods. In: Figueira J, Greco S, Ehrgott M (eds) *Multiple criteria decision analysis: State of the art surveys*. Springer, Boston, pp 133–162
4. Heesch D, Rüger S (2007) Interaction models and relevance feedback in image retrieval. In: Zhang Y-J (ed) *Semantic-based visual information retrieval*. IRM Press, Hershey, London, pp 160–186
5. Lew MS (2001) *Principles of visual information retrieval*. Springer, London
6. Lew MS, Sebe N, Djeraba C, Jain R (2006) Content-based multimedia information retrieval—state of art and challenges. *ACM Trans Multimed Comput Commun Appl* 2(1):1–19
7. Li HF, Wang JJ (2007) An improved ranking method for ELECTRE III. In: Zhou H (ed) *International Conference on Wireless Communication, Networking and Mobile Computing (WiCOM '07)*, Shanghai. pp 6659–6662. doi:10.1109/WICOM.2007.1634
8. Lian Z, Godil A, Bustos B, Daoudi M, Hermans J (2013) A comparison of methods for non-rigid 3D shape retrieval. *Pattern Recogn* 46(1):449–461
9. Mousseau V, Slowinski R (1998) Inferring an ELECTRE TRI Model from assignment examples. *J Glob Optim* 12(2):157–174
10. Mousseau V, Slowinski R, Zielniewicz P (2000) A user-oriented implementation of the ELECTRE-TRI method integrating preference elicitation support. *Comput Oper Res* 26(7–8):757–777
11. Muneesawang P, Guan L (2006) *Multimedia database retrieval: A human-centered approach*. Springer Series on Signals and Communication Technology. Springer
12. Rotter P (2012) Multimedia information retrieval based on pairwise comparison and its application to visual search. *Multimed Tool Appl* 60(3):573–587. doi:10.1007/s11042-011-0828-8
13. Rotter P, Skulimowski AMJ (2008) A new approach to interactive visual search with RBF networks based on preference modelling. In: Rutkowski L et al. (eds) *LNAI 5097*. Springer-Verlag Berlin Heidelberg, pp 861–873
14. Rotter P, Skulimowski AMJ (2009) Preference extraction in image retrieval. In: Ma Z (ed) *Artificial intelligence for maximizing content based image retrieval*. Idea Group Inc, Harshley, pp 235–260
15. Roy B, Vanderpooten D (1997) An overview on The European school of MCDA: emergence, basic features and current works. *Eur J Oper Res* 99(1):26–27
16. Schroff F, Criminisi A, Zissermann A (2011) Harvesting image databases from the web. *IEEE Trans Pattern Anal Mach Intell* 33(4):754–766
17. Shen X, Wang Z, Cho S-Y (2008) Mining user hidden semantics from image content for image retrieval. *J Vis Commun Image Represent* 19(3):145–164
18. Skulimowski AMJ (1996) *Decision support systems based on reference sets*. AGH Publishers, Krakow
19. Skulimowski AMJ (2011) Freedom of choice and creativity in multicriteria decision making. *Lect Notes Comput Sci* 6746:190–203
20. Smeets D, Fabry T, Hermans J, Vandermeulen D, Suetens P (2010) Inelastic deformation invariant modal representation for non-rigid 3D object recognition. In: Perales FJ, Fisher RB (eds) *VI Conference on Articulated Motion and Deformable Objects (AMDO)*, Springer, Andratx, Mallorca, Spain. pp 162–171
21. Tao D, Xu D, Li X (eds) (2009) *Semantic mining technologies for multimedia databases*. IGI Global, Hershey, London
22. Tian Y, Jiang S, Huang T, Gao W (2009) Semantic classification and annotation of images. In: Tao D, Xu D, Li X (eds) *Semantic mining technologies for multimedia databases*. IGI Global, Hershey, London, pp 350–377
23. Vasconcelos N (2007) From pixels to semantic spaces: advances in content-based image retrieval. *Computer* 40(7):20–26
24. Vogel J, Schiele B (2006) Performance evaluation and optimization for content-based image retrieval. *Pattern Recogn* 39:897–909
25. Wong WT, Hsu SH (2006) Application of SVM and ANN for image retrieval. *Eur J Oper Res* 173:938–950



Pawel Rotter received a PhD degree from the Faculty of Electrical Engineering, AGH-University of Science and Technology in Kraków where he currently holds the post of assistant professor in the Automatics Department. In the past he has worked at the Artificial Intelligence and Data Analysis Laboratory of the Aristotle University of Thessaloniki (2001–02), at the Cracow University of Technology (2002–04) and at the Institute for Prospective Technological Studies of the EC Joint Research Centre in Seville (2005–08). He was engaged in 14 research projects and is the author or co-author of over 30 publications and 10 technical reports, several chapters in world-range monographs and a series of popular science articles. His main research interests are computer vision and multicriteria optimization.