Directed enumeration method in image recognition

A.V. Savchenko*
National Research University Higher School of Economics, 25/12 Bolshaja Pecherskaja Ulitsa, Nizhny Novgorod 603155, Russia

A R T I C L E   I N F O

Article history:
Received 22 August 2011
Received in revised form 8 February 2012
Accepted 13 February 2012
Available online 24 February 2012

Keywords:
Image recognition
Large database
Approximate nearest neighbor classification
Directed enumeration method
Color histogram
Gradient orientation histogram

A B S T R A C T

The article is devoted to the problem of image recognition in real-time applications with a large database containing hundreds of classes. The directed enumeration method as an alternative to exhaustive search is examined. This method has two advantages. First, it could be applied with measures of similarity which do not satisfy metric properties (chi-square distance, Kullback–Leibler information discrimination, etc.). Second, the directed enumeration method increases recognition speed even in the most difficult cases which seem to be very important in practical terms. In these cases many neighbors are located at very similar distances. In this paper we present the results of an experimental study of the directed enumeration method with comparison of color- and gradient-orientation histograms in solving the problem of face recognition with well-known datasets (Essex, FERET). It is shown that the proposed method is characterized by increased computing efficiency of automatic image recognition (3–12 times in comparison with a conventional nearest neighbor classifier).

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Automatic image recognition is an active research topic in computer vision [1,2]. Its applications include biometrics, law enforcement, human–computer interaction systems, etc. Our focus of research is image recognition in the context of large databases [3,4]. We define the database to be large if it contains hundreds of different classes. Conventional classifiers (artificial neural networks, template matching, the Bayes classifier, hidden Markov models, HMM) [2] are usually based on exhaustive search through all classes, hence they often cannot be implemented in real-time applications. Moreover, if the number of classes is large, the amount of model images in the database is not usually enough to train the complex classifier and achieve high accuracy. In worst cases even one sample per class [5] may happen. Thus nearest neighbor algorithms [6] are preferred [3]. As the demand of such applications increases [7], many novel image recognition methods have been proposed in the last several decades. Unfortunately, the existing algorithms cannot be used with non-traditional measures of similarity such as the naive Bayesian classifier or chi-square statistics [2]. Moreover, they show poor results in most practical cases of object recognition where many neighbors are at very similar distances [3]. To overcome these problems we proposed the directed enumeration method [4]. The method was originally applied with color histogram comparison using naive Bayesian classifier to recognize images from synthetic dataset [4]. Meanwhile, its capabilities have not been fully exploited. In particular, almost no studies have addressed the advantages of our method in real image recognition tasks. The present paper seeks to fill this gap. We propose here a new modification of our method and show that it is beneficial in terms of computational speed and memory capacity in face recognition and various feature sets (conventional color- and modern gradient-orientation-histograms).

The rest of the paper is organized as follows: Section 2 reviews some related studies. Section 3 introduces the image recognition problem and desirable metric properties of decision statistics. In Section 4 we present the directed enumeration method. Section 5 describes image features and measures of similarity used in the experimental study in a problem of face recognition. In Section 6, we present the experimental results and analyze the proposed method in the face recognition problem with large databases (Essex [8] and FERET [9] data sets). Section 7 introduces new modification of the directed enumeration method which could be used to reduce necessary amount of additional memory. Finally, concluding comments are presented in Section 8.

2. Related research

In this section, we focus on the pattern recognition literature that deals specifically with prevention of an exhaustive search.

For uniformly distributed (in terms of some measure of similarity) point sets, good expected case performance can be
achieved by algorithms based on simple decompositions of image space into regular grids. Rivest [10] and later Cleary [11] provided analyses of these methods. Bentley, Weide, and Yao [12] also analyzed a grid-based method for distributions satisfying certain bounded-density assumptions. These results were generalized by Friedman, Bentley, and Finkel [13] who showed that really good results are achievable in the expected case through the use of k-d trees. However, even these methods suffer as the image features dimension increases [14]. Lowe states [15] that the best algorithms, such as the k-d tree [13] provide no speedup over exhaustive search for more than about 10 dimensional spaces. And, for instance, widely used SIFT (Scale-Invariant Feature Transform) keypoint descriptor [15] has a 128-dimensional feature vector. Thus Lowe concludes that “no algorithms are known that can identify the exact nearest neighbors of points in high dimensional spaces that are any more efficient than exhaustive search”.

Therefore, approximate algorithms (in the sense that they return the closest neighbor with high probability) were offered to achieve better recognition performance [3, 14]. One of the oldest but widely-used solutions of that problem is based on a statistical approach where images are grouped (clustered) near several centers [2]. One early method [16] uses a hierarchical tree data structure in which internal nodes are the centers of mass (centroids) of the nodes at the next lower level. The point recovered from the first encountered leaf node provides an approximate nearest neighbor in a very short search time, but the accuracy of this decision is relatively low.

Nowadays most part of such approximate algorithms is based on priority search order [3]. It was first examined by Arya and Mount [17], and they provide further study of its computational properties in [14]. This search order requires the use of a heap-based priority queue for efficient determination of the search order. An approximate answer can be obtained at a low cost by cutting off further search after a specific number of the nearest neighbors have been explored. In Lowe's experiment [15], for a database of 100,000 templates, this algorithm provides a speedup over exact nearest neighbor search by 100 times. It results in less than a 5% loss in the number of correct matches. In their implementation, they cut off search after checking the first 200 nearest neighbor candidates therefore achieving good performance.

Unfortunately, this approach cannot guarantee convergence. Also it requires the nearest neighbor be quite closer to input image than the second-nearest neighbor. Thus it cannot be used to exactly solve the most difficult cases in which many neighbors are at very similar distances [3]. However, this problem is quite acute for the most important cases of image recognition, such as face recognition [18, 19].

Face recognition is a challenging task because factors such as pose, illumination, facial expression make it difficult to achieve high accuracy [20]. It is known [21] that face recognition is substantially different from classical pattern recognition problems, such as object recognition. The shapes of the objects are usually different in an object recognition task, while in face recognition one always identifies objects with the same basic shape. This is of utmost difficulty for a face recognition system when one tries to discriminate faces all of which have the same shape. This is of utmost difficulty for a face recognition system substantially different from classical pattern recognition principle [22]. Some of these similarity measures even do not satisfy metric properties (triangle inequality for both measures) so triangle-tree [1] cannot be built. Moreover, Kullback–Leibler discrimination [22] is not a symmetric statistics.

It seems that no one of existing algorithms could be used in such difficult recognition tasks as automatic face recognition with a large database using non-traditional measures of image similarity. Hence, in this paper we investigate our directed enumeration method originally presented in [4]. We prove experimentally that this method could be used for any continuous measure of similarity which satisfies non-negativity, and identity of indiscernibles metric properties.

3. Automatic image recognition

Let a set of $R > 1$ grayscale model images $X_r = \|x_{uv}\|, (u \in \{1,...,U\}, v \in \{1,...,V\})$ be specified. Here $U$ and $V$ are the image height and width, respectively, $x_{uv}$ is an intensity of an image point with coordinates $(u, v)$; $r$ is the reference number ($r = 1, ..., R$), and $x_{\text{max}}$ is the maximum intensity. It is assumed that each image $X_r$ defines class $c(X_r) \in \{1, ..., C\}$, $C \leq R$ is a given amount of classes. The objects belonging to each class have some common features or similar characteristics. The common feature that unites objects in a class is called a pattern. The image recognition problem [2] is to assign an incoming image $X = \|x_{uv}\|$ (referred to a as a query or probe image) to one of the $C$ classes, i.e., to obtain the model image $X_r$ from the database which contains the same object as $X$.

The template matching procedure for constructing decision rules is usually based on the definition of a certain distance (measure of similarity) between any pair of objects and nearest neighbor rule [6]

$$v = \arg \min_r \rho(X/X_r).$$

Thus, the image recognition procedure involves an exhaustive search through $R$ images from the given database.

We consider the most important and difficult case $C > 1$, where the database contains hundreds or even thousands of classes. For the specified conditions, practical implementation of the optimal decision rule (1) encounters the obvious problem of its computational complexity and even feasibility. The present article is an attempt to develop the image recognition method which does not need brute force.

We first require non-negativity and identity of indiscernible metric properties of decision statistic $\rho(X/X_r) \geq 0$, which is equal to zero only in the ideal case of coincident input and template images. If this assumption is correct, we may simplify criterion (1) to a form suitable for practical implementation [4]:

$$\rho(X/X_r) < \rho_0 = \text{const}. \quad (2)$$

Here $\rho_0$ is the threshold for the admissible distance on the set of similarly-named images from one class. The value of this threshold could be found experimentally by fixing the false-accept rate (FAR)

$$\beta = P(\rho(X/X_r) < \rho_0 | c(X) \neq c(X_r)) = \text{const} \ll 1.$$  

Here $c(X)$ stands for class index of input image $X$, and $P(\rho(X/X_r) < \rho_0 | c(X) \neq c(X_r))$ is a probability that the distance between a query image $X$ and a model image from another class does not exceed the threshold $\rho_0$. In practice, $\rho_0$ could be determined by using the
training procedure from the following equation [28]
\[ \sum_{r=1}^{R} H\left(\rho_{0r} - \min_{(v+\alpha r)} \rho(X_i/X_{r})\right) = [\beta - R], \]  
(3)
where \( H(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}\) is a Heaviside step function, \([-\cdot]\) is a ceil function.

In fact, expression (2) defines the approximate nearest neighbor algorithm [3,14]. The solution \( X_r \) is approximate because there is no guarantee that it stands for the closest neighbor (global optima). We could say only with \( 1-\beta \) confidence (3) that \( X_r \) corresponds to the same class as the closest neighbor. However, in practice it is usually enough as the recognition performance is achieved. Really, in decision-making process one needs to calculate the measure \( \rho(X_r/X) \) only until it becomes smaller than a certain threshold \( \rho_0 \) instead of exhaustive search through all model images in the database. This circumstance should definitely reduce the amount of enumeration. A natural development of this idea is the directed enumeration method described in Section 4.

4. Directed enumeration method

Following the general computation scheme (1), (2), we reduce the image \( X \) recognition problem to a check of the \( N=\text{const} \leq R \) first images \( \{X_1, X_2, \ldots, X_N\} \), from the specified database \( \{X\} \). After that we define the image \( X_r \) as the closest (in terms of used measure of similarity \( \rho(\cdot, \cdot) \)) to query image \( X \) already checked database model
\[ \mu = \arg \min_{r \in \{1, \ldots, N\}} \rho(X_r/X). \]  
(4)

If \( X_r \) meets the termination requirement (2), the enumeration will end. However, it can generally be assumed that none of the first \( N \) alternatives meets condition (2). Then, it is possible to check the second group from \( N \) template images within the set \( \{X_r\} \), then the third group, etc., until condition (2) is met. However, there is also another, more rational method to solve the problem in question.

As a preliminary step we generate an \( (R \times R) \) distance matrix
\[ P = \|\rho_{ij}\| \]  
(5)

This computationally complex operation needs to be made only once: in the preliminary computation step and for each concrete database. However, this matrix is always calculated in conventional clustering algorithms [2] used to reduce the database to the set of centroids as a first step in image recognition.

Let us put all images from the set \( \{X_1, X_2, \ldots, X_N\} \) into heap-based priority queue \( Q \). The queue’s comparator function defines the concrete directed enumeration method implementation. In the initial version of the proposed method [4] the discrimination extrapolation based on autoregressive model and the Burg–Levinson algorithm [29] had been used. Later we discovered that this approach causes an increase in the amount of calculations (to evaluate autoregression coefficients). Hence, it is not as efficient as its simplified form [30] in which we arrange images \( \{X_r\} \) in the queue \( Q \) in a decreasing order of their distances \( \rho(X_r/X) \) to a query image \( X \).—“First Template First” (i.e., the same priority queue based on closeness which is used in the Best-Bin First algorithm [3]). So the first element in the queue is the closest checked model image to a probe image \( X \) (in terms of used similarity measure). However, we could say that other strategies (even “Worth Template First”) are also appropriate.

After that we extract the first element \( X_1 \) from priority queue \( Q \). If this image is closer to \( X \) then \( X_1 \)
\[ \rho(X/X_1) < \rho(X/X), \]  
then we assign \( i \) to \( \mu \). Thus we always refer to \( X_\mu \) as the best already checked database image.

Based on the matrix \( P \) (5) we obtain the set of \( M=\text{const} < R \) images \( X^{(M)}_i = \{X_1, \ldots, X_m\}, i \leq R \) that are separated from the image \( X_r \) by the distance not exceeding \( \rho(X_r/X_\mu) \):
\[ (\forall X_\mu \neq X^{(M)}_i) (\forall X_k \in X^{(M)}_i) \Delta \rho(X_\mu) \geq \Delta \rho(X_k) \]  
(6)

Here
\[ \Delta \rho(X_\mu) = |\rho(X_\mu/X) - \rho_{ij}| \]  
(7)

is the deviation of the discrimination \( \rho_{ij} = \rho(X_j/X) \) relative to the discrimination between the pair of images \( X \) and \( X_\mu \). This procedure could be performed more efficiently, if we preliminarily sort the columns of matrix \( P \) in an ascending order so that the binary search could be used to locate \( X^{(M)}_i \).

This step is ended by putting not previously checked images from the set \( X^{(M)}_i \) into queue \( Q \).

If the queue \( Q \) is empty we add one random element \( X_{\text{ran}} \) from the database that has not been checked previously. This action ensures that the proposed algorithm is finite.

Then, all computations of the first step are repeated cyclically until, in some \( k \)th step, an element \( X^* = X_i \) meets the termination condition (2):
\[ \rho(X/X^*) < \rho_0. \]  
(8)

In this case, a decision is made in favor of the closest pattern \( X^* \) or, at worst in the absence of a solution from (8), the conclusion is drawn that the input image \( X \) cannot be assigned to any class from the database and that it is necessary to switch to the decision feedback mode (i.e., ignore the frame in video recognition or use the reject option [31]). Anyway, \( X_\mu \) still refers to the closest to \( X \) database image.

Generally, there may be a considerable gain in the total count \( k = N+M \cdot K \leq R \) of templates checked according to (6) compared to the size of the used database. It is explained by the fact that the probability \( p \) that the desired image \( X^* \) belongs to the \( X^{(M)}_i \) set usually exceeds the probability of \( X^* \) inclusion into randomly chosen \( M \) alternatives
\[ p = P(X^* \in X^{(M)}_i) \geq p_0 = \frac{M}{R}. \]  
(9)

This is the effect of the directed enumeration. The probability \( p \) usually depends on the applied similarity measure and properties of query image and given database.

The following theorems are correct due to the definition of our method.

**Theorem 1.** The proposed method (2)–(8) is finite. It always converges to database template \( X^* \) which is either the nearest neighbor to query \( X \) or meets condition (8).

**Theorem 2.** Let’s suppose that
\[ \forall i, j, k \in \{1, \ldots, R\}, \ j \neq k \quad \rho(X_j/X_\mu) \neq \rho(X_k/X_\mu). \]  
(10)

Then if query image \( X \) is one of the database images, the count of distance \( \rho(\cdot, \cdot) \) calculations according to proposed method does not exceed \( M=\text{const}(R) \).

The assumption (10) is true in the most important practical cases when a continuous measure of similarity \( \rho(\cdot, \cdot) \) is used.
Table 1
The directed enumeration method.

| Data: input image X, database \(X_i\), matrix of database images distances \(P(5)\) |
| Result: image \(X^*\) from the database which satisfies conditions of Theorem 1 |

1. Select \(N=\text{count random variants}(X_1, X_2, ..., X_n)\). Insert them into queue \(Q\).
2. Define image \(X_r\) as a closest to \(X\) from the set\(\{X_1, X_2, ..., X_n\}\) (4)
3. While \(\rho(X(X_r)) > \rho_0\) and \(\text{(count of checked database images is less than } R)\) do
   3.1. If queue \(Q\) is empty
      3.1.1. Insert random database image \(X_{\text{rand}}\) into the queue \(Q\).
      3.1.2. Fill the set \(X_{\text{rand}}\) based on matrix \(P\) and formulas (6) and (7)
   3.4. For each image \(X_i\) in the set \(X_{\text{rand}}\)
      3.4.1. If \(X_i\) has not checked previously (distance \(\rho(X(X_i))\) has not been calculated), then
      3.4.1.1. Insert \(X_i\) into queue \(Q\).
      3.4.1.2. If \(\rho(X(X_i)) < \rho(X(X_r))\), then
      3.4.1.2.1. Define image \(X_r\) as \(X_i\).
   4. Put \(X_r\) into result \(X^*\)

However, such a simple measure as

\[ \rho(X(X_i)) = \begin{cases} 0, & \forall u \in \{1, ..., U\}, v \in \{1, ..., V\} | X_{uv} = X_{ij}^0 \\ 1, & \text{otherwise} \end{cases} \]

cannot be used in our method.

Note that we could follow Beis and Lowe’s approach [3] and stop search after checking several first candidates. However, we do not use this possibility as we always want to get the satisfactory results (Theorem 1). Thus we decided to sacrifice the performance of offered method to get better results. Of course, this solution could cause an exhaustive search in the database if no image meets condition (2).

Thus, expressions (2)–(8) define the proposed directed enumeration method in the image recognition problem. Its key difference from the existing algorithms is the usage of \(P\) distances matrix (5) to choose the next database template to check (6) and (7). \(P\) should be stored as a matrix of pairs \(\{(j_{ik}, \rho_{ik})\}\), \(i = \{1, ..., R\}, k = \{1, ..., K\}\) where \(j_{ik}\) is a permutation of numbers \(\{1, ..., \}\), \(\rho_{ik} = \rho(X_i/X_{ik})\) and \(\rho_{1i} \leq \rho_{2i} \leq \ldots \leq \rho_{ki}\). In such a case the binary search could be used to obtain \(X_{\text{rand}}\) with expressions (6) and (7).

We summarized the described procedure in the following algorithm (Table 1).

5. Automatic face recognition

As it was stated in the introduction, automatic face recognition is one of the most difficult tasks in image recognition. Nearest neighbor face recognition can be factorized into two essential parts [18]: (1) feature extraction from facial images; and (2) similarity measure design. In this section we describe two conventional feature sets: traditional color histograms [1,32] and modern gradient orientation histograms (modification of SIFT method [15]). For the second part we use histogram-based approach [2,15]: a histogram corresponding to a query image is compared to the histograms of all the images stored in the database using some measure of similarity. In this paper we show the application of the directed enumeration method to non-standard measures: Kullback–Leibler information discrimination [25] and chi-square statistics [26]. In both cases we calculate local features by dividing images into a regular grid to provide illumination and outlier robust appearance–based correspondence with some leeway for small spatial deviations due to misalignment. However, to simplify formulas in this section we suppose that the whole image is considered.

5.1. Color histogram comparison with Kullback–Leibler information discrimination

Let’s consider a discrete certain random variable—color of a pixel in the image \(X\). Its distribution \(H_r = \{h_r^{(1)}, h_r^{(2)}, ..., h_r^{(m)}\}\) could be evaluated based on matrix \(\|X_{uv}\|^2\)

\[ h_r^{(1)} = \frac{1}{UV} \sum_{u=1}^{U} \sum_{v=1}^{V} h(x_{uv}) \]  

(11)

Here \(h(x) = \begin{cases} 1, & x = 0, \\ 0, & \text{otherwise} \end{cases} \) —discrete Dirac delta-function. \(H_r\) is often called “color histogram” [32,33] of image \(X\). The same procedure for color histogram \(H_d\) definition is applied for the input image \(X\).

According to the theory information approach, query image \(X\) stands for a received signal of a noisy communication channel, and one image from the database stands for a transmitted signal. Hence, the problem is to minimize the mutual-entropy (or Kullback–Leibler information discrimination) [25] between color distributions of query and model images. Thus it is required to verify \(R\) hypotheses on the distribution \(H_{q_r} = \{h_{q_r}^{(1)}, h_{q_r}^{(2)}, ..., h_{q_r}^{(m)}\}\) of the input image \(X\). It is well-known [25] that the optimal decision of the problem of statistical check of hypotheses about discrete stochastic variable’s distribution in Bayesian terms is equivalent to the minimum discrimination information principle and the optimal decision rule

\[ \rho_{KL}(X/X_i) = \sum_{r=1}^{N} h_r \ln(h_r/h_r^{(c)}). \]  

(12)

Here the statistics \(\rho_{KL}(X/X_i)\) defines the Kullback–Leibler information discrimination between the observed image signal \(X\) and \(r\)th template image from the database \(X_i\).

Color histograms comparison has widely been used in image retrieval [34]. Unfortunately, images with the same visual information, but with shifted color intensity, may significantly degrade if the conventional method of direct histograms comparison is used. Actually, if the input image \(X\) is one of the template images \(X_i\) from the database but all pixels are decolorized, its color histogram \(H\) will be quite different from \(H_i\). Eventually, we decided to use widely used shifting of histograms [35] after their evaluation (11) to reduce illumination influence.

5.2. Gradient orientation histograms comparison with chi-square distance

Our second criterion is based on SIFT algorithm [15]. At first, we evaluate gradient magnitude and angle for each pixel of the image using modified Robert’s definition [1]

\[ m_{uv}^{(r)} = \frac{1}{2} (|X_{uv+1,p+1}^{(r)} - X_{uv,p}^{(r)}| + |X_{uv+1,p-1}^{(r)} - X_{uv,p}^{(r)}|), \]  

(13)

\[ \theta_{uv}^{(r)} = \arctan \frac{X_{uv+1,p+1}^{(r)} - X_{uv,p}^{(r)}}{X_{uv+1,p-1}^{(r)} - X_{uv,p}^{(r)}}. \]  

(14)

Thus, we found that our magnitude definition (13) is better (in terms of face recognition accuracy) than the more conventional definition [15] used in SIFT

\[ m_{uv}^{(r)} = \sqrt{(X_{uv+1,p+1}^{(r)} - X_{uv,p}^{(r)})^2 + (X_{uv+1,p-1}^{(r)} - X_{uv,p}^{(r)})^2}. \]  

We divide the whole orientation definition range \([-\pi; \pi]\) into \(N_o\) regular segment (thus each segment has width \(2\pi/N_o\)). In our experiments we used \(N_o = 8\) — conventional value of segment count [15]. Then we evaluate its histogram \(H_{io} = \{h_{io1}^{(1)}, h_{io2}^{(2)}, ..., h_{iom}^{(m)}\}\) using
relative magnitude as a weight

\[ h_{(r)}^{(i)} = \frac{1}{U \nu} \sum_{u=1}^{U} \sum_{v=1}^{\nu} m_{u,v}^{(i)} + \Delta m \left( H\left( \frac{p_{u,v}}{N_0} \right) - \frac{2i-1-N_0}{N_0} \right) \]

\[ -H\left( \frac{p_{u,v}}{N_0} \right), \quad i \in \{1, \ldots, N_0\}. \]  

(15)

Here again \( H\cdot \cdot \cdot \) is a Heaviside function. \( \Delta m = \text{const} < \frac{1}{4} \) is a summand to include in the histogram points with zero magnitude \( (m_{u,v}^{(i)} = 0) \). We found experimentally that good recognition results could be obtained with \( \Delta m = 0.1 \). And the third element of (15) which is not yet defined earlier is

\[ m^{(i)} = \max_{u=1, \ldots, U; v=1, \ldots, \nu} m_{u,v}^{(i)}. \]  

(16)

After histogram definition (13)–(16) we use nearest neighbor rule with chi-square distance [26] between orientation histograms

\[ \rho_{(r)}(X/X_t) = \frac{N_0}{\sum_{i=1}^{N_0} \frac{(h_{(r)}^{(i)} - h_{(0)}^{(i)})^2}{h_{(r)}^{(i)} + h_{(0)}^{(i)}}. \]  

(17)

Note that we do not use such stages of SIFT as difference of Gaussians (DoG) filter [1] and gradient orientation rotation to the keypoint orientation as both stages significantly decrease (more than 15%) face recognition accuracy in all our experiments.

6. Experimental results

Our experiment deals with the problem of face recognition [17,18]. The images were preliminarily processed with OpenCV library [36] to detect faces. After that the median filter with window size \((3 \times 3)\) was used to remove some noise in detected faces. This window size provides the best recognition accuracy for both FERET and Essex datasets and measures of similarity (12) and (17).

In our earlier experiments [4,28,30] we found that the directed enumeration method parameter \( N \) can get out almost anyway without making significant impact on recognition speed and with evident condition \( N < R \). At the same time the choice of parameter \( M \) is much more important. \( M \) should not be very close to \( 1 \), otherwise the directed enumeration procedure (6) and (7) could miss the nearest neighbor in a couple of iterations. Also \( M \) should be much less then the database size \( R \), otherwise the proposed method will be equivalent to the random search. After several experiments [28,30] the best (in terms of speed) values of parameters were chosen as follows: \( N = 5, M = 32 \).

For color histograms comparison we divided detected faces in a \( 4 \times 4 \) regular grid and \( x_{\text{max}} = 64 \) (i.e., the feature vector size is \( 64 \cdot 4 \cdot 4 = 1024 \)). The final Kullback–Leibler discrimination was calculated as a sum of discriminations between corresponding parts (12). For gradient orientation histograms we discovered that traditional \( 4 \times 4 \) grid is too large to provide good face recognition accuracy. Therefore, we divided faces in \( 144 \) \( (12 \times 12) \) parts (i.e., the feature vector size is \( 8 \cdot 12 \cdot 12 = 1152 \)). Again, final chi-square distance between images was calculated as a sum of distances (17) between the corresponding parts.

As we stated in Section 1, the well-known algorithms – kd-trees and their modification Best-Bin First [3] – cannot be applied with similarity measures which do not satisfy all metric properties. Therefore, we have decided to carry out a comparative analysis of our method with an algorithm which could really be used for calculations reduction, namely – brute force with early termination when condition (2) is met, i.e., the random search [37].

To evaluate the proposed method performance, we measure the number \( k \) of images that must be checked [3] to obtain the model image which meets the termination condition (2). As our method sometimes requires random model \( X_{\text{rand}} \), the number \( k \) may differ even for a fixed query image. Hence, we show the measure of \( k \) variability (in \( \pm \)) together with the average \( k/R \) (in percentage). This variability was estimated by running test of the directed enumeration method 15 times. The recognition accuracy varies much lower as the size of the training set is large However, we also show the variability of the accuracy in the same manner.

This section has the following structure. In Section 6.1 we provide results of face recognition by using the Essex database [8]. In Section 6.2 the most important results are shown for popular FERET [9] dataset. Final discussion and comparison of color and gradient orientation histograms influence on the proposed method efficiency are presented in Section 6.3.

6.1. Essex dataset

The \( R=900 \) images were selected as the training set \( \{X_i\} \) from the 6400 photographs of \( C=395 \) different people using basic sequential clustering algorithm [2]. Other 1500 photos of the same people were used as a test set to evaluate recognition accuracy.

In the first case the Kullback–Leibler color histogram comparison (11) and (12) was used. The exhaustive search recognition rate is 97.8% \((\pm 0.15\%)\).

The threshold \( p_0=0.18 \) for approximate nearest neighbor methods was chosen experimentally on the training set by fixing of \( \text{FAR} \beta=5\% \). Here the exact solution \( X' \) (the same person as from query image \( X \)) was obtained in 97\% \((\pm 1\%)\) of the test cases.

The average \( k/R \) is 44\% \((\pm 2.3\%)\) for a random search – approximate nearest neighbor algorithm implementation with termination condition (2). Condition (2) was not met for any template from the database for 5.7\% of the initial test images; therefore, all \( R \) alternatives were checked. A histogram of \( k/R \) is shown in Fig. 1(a). The total number of cases in this histogram is equal to the size of the training set. This histogram looks practically like a uniform distribution. For example, we could conclude from this figure that the probability that \( k/R \) will be less than 5\% is equal to 7.6\%.

Using proposed directed enumeration method (2)–(8), an average \( k/R \) is equal to 8.0\% \((\pm 1.2\%)\). A histogram of \( k/R \) carried out by algorithm (2)–(8) for this case is shown in Fig. 1(b). With a probability of 87\%, \( k \) does not exceed 5\% of \( R \). The bar 0.95–1.00 in Fig. 1(b) stands for the 5.7\% of test images for which distances to all model images exceed threshold \( p_0 \).

![Fig. 1. Histogram of k/R (Kullback-Leibler color histogram comparison, Essex dataset). (a) Random search and (b) directed enumeration method.](image-url)
average 85% of checks. However, later we will continue to evaluate query images from the training set, while random search requires In other words, our method checks 7% of the database for 90% of nearest neighbor matching) recognition rate is 98.8% (Chi-square gradient orientation histogram comparison, Essex dataset). (a) Random search and (b) directed enumeration method. The exhaustive search (conventional nearest neighbor matching) recognition rate is 98.8% (± 0.1%), i.e., it is higher than simple color histogram (12) comparison. The approximate nearest neighbor performance is not excellent – the recognition rate here is equal to 97.5% (± 1.15%) for a threshold of \( \rho_0 = 0.21 \). However, condition (2) was not met for 21.5% of test images; therefore, all \( R \) alternatives were checked. Average \( k/R \) was 56% (± 3.4%) for random search (Fig. 2(a)) and 23.4% (± 1.8%) for the directed enumeration method (Fig. 2(b)). Again we could see that the performance of proposed method is essentially higher than that of the traditional random search, though the recognition rate is the same.

We could see from both experiments that the efficiency (the average \( k/R \)) of the proposed method is determined, first of all, by the count of test cases \( X \) for which \( \min \rho(X/X_r) \geq \rho_0 \). This is especially true for SIFT-based algorithm (13)–(17). For instance, we show the distance calculation count \( k^* \) which is necessary to find \( X^* \) using the directed enumeration method in Fig. 3. The value of \( k^* \) is equal to \( k \) when condition (8) is met for \( X^* \). Otherwise \( k^* \) could be much less than \( k \) even when the nearest neighbor \( X^* \) is found, the method continues brute force to guarantee convergence (see Theorem 1). Thus \( k^* \) is the minimum count of distance calculation we could achieve for our method if some other, better termination condition is used instead of (2).

Dependencies on threshold value of recognition rate and \( k/R \) for proposed method are shown as a box plots in Figs. 4 and 5. Based on these figures, we could draw obvious conclusion – when \( \rho_0 \) threshold is increased, recognition rate and speed are decreased.

6.2. FERET dataset

In this subsection we briefly discuss application of our method to widely-used FERET dataset. We should denote that it is much more complicated than the Essex database, so, achieved results are not as good. However, we still can demonstrate the efficiency of our method even in such a difficult task.

![Histogram of k/R (Chi-square gradient orientation histogram comparison, Essex dataset).](image)

![Dependency of recognition rate and k/R for directed enumeration method on threshold \( \rho_0 \) (Kullback-Leibler color histogram comparison histogram comparison, Essex dataset).](image)
We populate the database with $R = 1272$ frontal photos of 994 people and use other 1432 photos of the same persons to test recognition accuracy. The nearest neighbor recognition rate is 21% $(\pm 0.3\%)$ for color histogram comparison (12) and (13) and 9.5% $(\pm 0.24\%)$ for gradient-orientation matching (13)–(17). Thus our SIFT method modification (13)–(17) shows quite better results.

As the error rate of brute force is high, we estimate threshold $(2)$ for 10% FAR. The recognition rate of the approximate nearest neighbor method $(2)$ is equal to 24% $(\pm 2.5\%)$ for Kullback–Leibler discrimination and the comparison of color histograms (12) and (13) with threshold $\rho_0 = 0.2$. Condition $(2)$ was not met for 29% of test images. The average $k/R$ for the random search is equal to 61%. Again, the performance of the directed enumeration method of the approximate nearest neighbor is equal to 87% $(\pm 0.3\%)$ for color histogram comparison (12) and (13) and 9.5% $(\pm 1.7\%)$ for the directed enumeration method (Fig. 6(b)). In this case, the 65-percentile of $k/R$ is 0.2.

In the second case the chi-square matching of the gradient orientation histograms (13)–(17) was used. The recognition rate of the approximate nearest neighbor is equal to 87% $(\pm 2.6\%)$ for a threshold of $\rho_0 = 0.13$. Condition $(2)$ was not met for 30% of test images. Average $k/R$ was 62% $(\pm 4.6\%)$ for random search and 35.4% $(\pm 1.7\%)$ for the directed enumeration method (Fig. 6(b)). In this case, the 65-percentile of $k/R$ is 0.09.

The use of the directed enumeration method reduced computational complexity of face recognition by 8% $(\pm 1.2\%)$ for Essex faces database and 35.4% $(\pm 1.7\%)$ for FERET data set in comparison with an exhaustive search of the nearest neighbor algorithm (1). As expected, the accuracy of gradient-orientation histogram comparison (13)–(17) is essentially higher than the recognition rate of the color histograms matching (11) and (12). However, it seems that our method’s performance (average $k/R$) is better for the color histograms comparison (Figs. 1 and 6(a)) than for our SIFT modification (Figs. 2 and 6(b)). The reason of this remarkable fact is discussed in this subsection and summarized in Table 2.

The color histogram matching approach in image recognition is somewhat unusual in that most other modern methods have used feature vectors that are invariant to any model transformations [3]. However, the authors of the article [3] came to the opinion that the inclusion of invariant feature sets to the requirements to the objects seems to be too restrictive. Besis and Lowe’s method uses features that are “partially invariant (i.e., to translation, scaling, and image plane rotation), but vary with out-of-plane rotations”. Our experiments showed that this proposal is correct though we used the color histogram matching [1] which is not invariant even to illumination changes.

As it seems quite important to determine what distance is better to combine with the directed enumeration method, we propose the following procedure. In the training phase after evaluation of FAR (3), false reject rate (FRR) $\alpha$

$$\alpha = P(\rho(X, X_i) \geq \rho_0 | c(X) = c(X_i))$$

is estimated

$$\alpha = \frac{1}{R} \sum_{r=1}^{R} H_{\rho} \min_{c(X)} \min_{c(X') \neq c(X)} \rho(X_i, X_r) - \rho_0.$$  \hspace{1cm} (18)

Here $P(\rho(X, X_i) \geq \rho_0 | c(X) = c(X_i))$ is a probability that the distance between query image $X$ and model image from the class $c(X)$ exceeds threshold $\rho_0$.

Actually the FRR (18) defines the bar 0.95–1.0 in Figs. 1(b), 2(b), and 6. The increase of the FRR leads to the increase of bar 0.95–1 on histogram and, therefore, to the decrease of the directed enumeration method performance.

In the Table 2, we summarize FRR $\alpha$ for previous experiments. We also show the “ideal” value of our method’s distance calculation count $k^*/R$ which does not depend on FRR.

Based on this table, we could draw the evident conclusion – the more false-reject rate, the lower the efficiency of the directed enumeration method. Hence, the performance of our method is lower for FERET dataset than for the Essex database as the FAR for FERET is essentially higher than the FAR for the Essex. And though recognition accuracy of the color histogram comparison is not so good, its reliability (measured by FRR) is quite better, if illumination of test images does not vary extremely. Therefore, we could conclude that though recognition rate is the most important property of applied similarity measure, its reliability also needs to be scored.

### Table 2

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Distance</th>
<th>FRR $\alpha$ (%)</th>
<th>$k/R$ 100% $\pm$ (%)</th>
<th>$k^*/R$ 100% $\pm$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Essex</td>
<td>Kullback–Leibler, color histograms</td>
<td>12.3</td>
<td>8.0% $(\pm 1.2%)$</td>
<td>4.1% $(\pm 1.2%)$</td>
</tr>
<tr>
<td>Essex</td>
<td>Chi-square, gradient orientation histograms</td>
<td>24.8</td>
<td>23.4% $(\pm 1.8%)$</td>
<td>8.4% $(\pm 0.7%)$</td>
</tr>
<tr>
<td>FERET</td>
<td>Kullback–Leibler, color histograms</td>
<td>27.1</td>
<td>30.5% $(\pm 1.9%)$</td>
<td>4.5% $(\pm 0.4%)$</td>
</tr>
<tr>
<td>FERET</td>
<td>Chi-square, gradient orientation histograms</td>
<td>32.6</td>
<td>35.4% $(\pm 1.7%)$</td>
<td>12.6% $(\pm 1.3%)$</td>
</tr>
</tbody>
</table>

6.3. Discussion

The use of the directed enumeration method reduced computational complexity of face recognition by 8% $(\pm 1.2\%)$ for Essex faces database and 35.4% $(\pm 1.7\%)$ for FERET data set in comparison with an exhaustive search of the nearest neighbor algorithm (1).

As it seems quite important to determine what distance is better to combine with the directed enumeration method, we propose the following procedure. In the training phase after evaluation of $\alpha$ (3), false reject rate (FRR) $\alpha$

$$\alpha = P(\rho(X, X_i) \geq \rho_0 | c(X) = c(X_i))$$

is estimated

$$\alpha = \frac{1}{R} \sum_{r=1}^{R} H_{\rho} \min_{c(X)} \min_{c(X') \neq c(X)} \rho(X_i, X_r) - \rho_0.$$  \hspace{1cm} (18)

Here $P(\rho(X, X_i) \geq \rho_0 | c(X) = c(X_i))$ is a probability that the distance between query image $X$ and model image from the class $c(X)$ exceeds threshold $\rho_0$.

Actually the FRR (18) defines the bar 0.95–1.0 in Figs. 1(b), 2(b), and 6. The increase of the FRR leads to the increase of bar 0.95–1 on histogram and, therefore, to the decrease of the directed enumeration method performance.

In the Table 2, we summarize FRR $\alpha$ for previous experiments. We also show the “ideal” value of our method’s distance calculation count $k^*/R$ which does not depend on FRR.

Based on this table, we could draw the evident conclusion – the more false-reject rate, the lower the efficiency of the directed enumeration method. Hence, the performance of our method is lower for FERET dataset than for the Essex database as the FAR for FERET is essentially higher than the FAR for the Essex. And though recognition accuracy of the color histogram comparison is not so good, its reliability (measured by FRR) is quite better, if illumination of test images does not vary extremely. Therefore, we could conclude that though recognition rate is the most important property of applied similarity measure, its reliability also needs to be scored.
7. Directed enumeration method modification

The most important limitation [30] of the described procedure (2)–(8) is its need to store distance matrix \( P \). Thus the directed enumeration method requires \( O(R^2) \) additional memory. The modification of the directed enumeration method which needs to store only the most valuable part of this matrix is proposed in this Section. We guess that the probability \( p \) (9) should depend also on the distance between \( X \) and \( X_i \). We could assume that image \( X_i \) contains valuable information to obtain \( X^* \) if it is closer (or further) to object \( X \), than the majority of other templates from the database.

To test this hypothesis, the dependence of \( p \) on \( \rho_{KL}(X_i|X) \) for the whole Essex dataset was estimated (see Fig. 7).

The probability that desired image \( X^* \) belongs to \( X_i^{(M)} \) is equal to \( p=5\% \). It is quite greater then \( p_0=M/R=32/900=3.6\% \). Based on Fig. 7 one may argue that though the minimum probability \( p \) (0.44 according to this chart) is quite greater than the random search success probability, the most valuable distances \( \rho(X_i|X) \) for a further decrease of computations are concentrated in the “corners” of Fig. 7.

Thus, we propose to store not all matrix \( P \), but only the most \( T \) lower and the most \( T \) higher distances to model images for each image from database. Here \( T = \text{const} < R/2 \) is a parameter of the proposed modification. The ratio \( 2T/R \) determines the decrease of memory usage. However, to satisfy Theorem 2 we store the whole matrix column for the first image \( X_1 \) which we always put into initial set \( \{X_1, X_2 \ldots X_N\} \). This approach causes modification of rule (6) to select set \( X_i^{(M)} \). If the stored part of \( P \) is insufficient to determine \( X_i^{(M)} \) based on (6) and (7) we just skip this step and select randomly a database template until procedure (6) and (7) could be applied. At first sight such procedure could increase \( k \). But the experiment below shows that we achieve practically the same quality if parameter \( T \) was selected properly.

At first, we examine dependency of average \( k/R \) on the parameter \( T \) (see box plots in Fig. 8). We used here the simplest experiment – Essex dataset with Kullback–Leibler information discrimination (11) and (12).

After that, the proposed method’s modification was used and the parameter \( T=50 \) was chosen. A histogram of the count of checks carried out by this modification for Essex dataset and the Kullback–Leibler color histogram matching is shown in Fig. 9. Note that it is not really similar to our previous results (see Fig. 1(b)). For instance, 90-percentage of \( k/R \) is 0.07 for original method (Section 4) and 0.1 for proposed modification.

However, the average count of checks is 9.1\% (± 0.9\%) (1.1\% increase in comparison with original directed enumeration method (2)–(8)) is appropriate in the most applications as we achieved \((1-2 \cdot 0.07/0.9) \cdot 100\% = 89\% \) memory economy. In other words, the proposed modification needs \( 900\cdot 2 \cdot 0.5 \cdot (4+8)/1024 = 1.05 \) Mb of additional RAM (in comparison with 9.27 Mb of RAM for the original implementation).

We summarize the experimental results for the other dataset and distances in Table 3.
Though the count of distance calculations of the proposed modification exceeds (approximately in 2%) the same indicator for the original method, we could conclude based on this table that the proposed modification definitely overcomes the directed enumeration method limitation [30] – requiring the storage of the whole matrix of distances between given models from the database.

8. Conclusion

In this paper we investigated the possibility of a computational speed increase in automatic image recognition in large databases [38]. It is well-known [3] that when a number of classes in the model database is large, the use of high-dimensional features [15] is critical, due to the improved level of discrimination they can provide. It is also evident that finding the nearest neighbor [6] to a query point rapidly becomes inefficient as the dimensionality of the feature space increases. Therefore, conventional nearest neighbors methods [1,2,6] based on exhaustive search cannot be implemented in real-time applications. Other widely-used classifiers [1] (neural networks, HMM, etc.) perform brute force through all classes and also cannot be used to speed up recognition procedure.

Thus approximate nearest neighbor algorithms were offered [3,15]. They are usually based on some sort of tree data structures [13,38] hence requiring high distinctiveness of database classes and thus limiting their potential applications. For example, this requirement is very hard for such essential problem as a face recognition. Another important disadvantage of many approaches is a demand for measure of similarity that satisfies all metric properties. However researchers usually investigate the potential improvement of the recognition accuracy with non-traditional distances [26,28].

In this article we offered a novel directed enumeration method as a solution of both the problems described above. In our experiment (Section 6) we showed that the proposed method can reduce computational complexity of face recognition by 8% for Essex faces database, and by 35% for the FERET data set, in comparison with the conventional brute force method. The measures used in our experiment are not the best in terms of face recognition accuracy (though the achieved results in Section 6 are rather good). We do show that our method can be applied for similarity measures which do not satisfy triangle inequality (Chi-square distance) or even symmetric metric properties (Kullback–Leibler divergence), where there are no alternatives to exhaustive search. It is known that many researchers concentrate on choosing classifiers which are most suitable for their particular applications [22,23,40].

We define the database to be large if the nearest neighbor algorithm with an exhaustive search cannot be implemented in real-time applications, such as face recognition from video where 50 frames should be processed in a second. Thus, in the experiments 900 images from the Essex database and 1272 images from the FERET database are used. The directed enumeration method has practically no limitation to be used with bigger database except the additional memory required to store the part of distance matrix P (see Section 7). For example, our method was originally used to speed up the recognition procedure with the synthetic database of 5000 images [4]. Unfortunately, modern image feature sets used nowadays seem to be not satisfactory for complex image recognition tasks. Even the recognition quality of gradient-orientation histograms comparison is not so good for the FERET dataset. Hence, an increase of the number of classes in the face database usually leads to a significant decrease of nearest neighbor classification accuracy. As we understand, there is no need to speed up the algorithm is its efficiency is poor. That’s why we concentrated on more important practical tasks with hundreds of classes and thousand images in the database.

In comparison with other approximate nearest neighbor methods [14], our algorithm is more straightforward. It does not have special requirements or extra restrictions for images or used measure of similarity. We cannot report of typical speedup of two orders [3] because our method does not search for a fixed number of nearest neighbor candidates (though it could be adapted to stop after the fixed iteration number). However, the quality of the solution reached by the directed enumeration method is comparable to that obtained by the brute force method. The main advantage is that we reach the same quality with more than 10-times reduction in the computational complexity.

Our experiments demonstrated that the efficiency of the proposed method depends on the reliability of distance measure in terms of known image model database. Thus, the distance choice becomes significantly important. We showed a remarkable fact that even if recognition accuracy of gradient orientation histograms comparison (17) is essentially higher than color histograms matching (12), the reliability of criterion (17) is insufficient. Indeed, it was impossible to define the threshold (3) with FRR less than 20% for the Essex data set. At the same time, FRR of color histograms matching (12) achieved 5% for the same fixed FAR.

The main direction for further research in the directed enumeration method could be related to the synthesis of more robust criteria to reduce FRR and, thus, improve our method’s efficiency. Another possible direction of research is a replacement of traditional termination condition (2) to some other criterion. As we showed in Fig. 3 and Table 2 (k*R) column), the directed enumeration method has a potential to improve its performance. As we stated earlier, we do not want to terminate the algorithm after a fixed number of iterations to save its convergence. However it is likely that other termination conditions should be synthesized to reduce FRR. This research may be applicable especially for the widely-used SIFT algorithm [15].

References


Table 3: k/R for the directed enumeration method and its modification (T=50).

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Distance</th>
<th>k/R (original method)</th>
<th>k/R (proposed modification)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Essex</td>
<td>Kullback-Leibler, color histograms</td>
<td>8.0% (± 1.2%)</td>
<td>9.1% (± 0.9%)</td>
</tr>
<tr>
<td>Essex</td>
<td>Chi-square, gradient orientation</td>
<td>23.4% (± 1.8%)</td>
<td>25.9% (± 1.7%)</td>
</tr>
<tr>
<td>FERET</td>
<td>Kullback-Leibler, color histograms</td>
<td>30.5% (± 1.9%)</td>
<td>31.1% (± 1%)</td>
</tr>
<tr>
<td>FERET</td>
<td>Chi-square, gradient orientation</td>
<td>35.4% (± 1.7%)</td>
<td>37% (± 1.2%)</td>
</tr>
</tbody>
</table>


