

# Image Recognition Using Kullback-Leibler Information Discrimination

Andrey Savchenko,

National Research University Higher School of Economics, B. Pecherskaya St. 25/12,  
603155 Nizhniy Novgorod, Russian Federation  
[avsavchenko@hse.ru](mailto:avsavchenko@hse.ru)

**Abstract.** The problem of automatic image recognition based on the minimum information discrimination principle is formulated and solved. Color histograms comparison in the Kullback–Leibler information metric is proposed. It's combined with method of directed enumeration alternatives as opposed to complete enumeration of competing hypotheses. Results of an experimental study of the Kullback–Leibler discrimination in the problem of face recognition with a large database are presented. It is shown that the proposed algorithm is characterized by increased accuracy and reliability of image recognition.

**Keywords:** Image recognition, method of directed enumeration alternatives, Kullback–Leibler information discrimination

## 1 Introduction

The well-known challenging problem of the image recognition [1, 2] is processing of large image databases [3, 4, 5]. Traditional pattern recognition methods [6] based on exhaustive search can't be implemented in real-time applications. For this case a method of directed enumeration alternatives (NDEA) [3] has been proposed as opposed to the traditional method of complete enumeration of competing hypotheses [6]. MDEA can be exploited with different metrics, but the efficiency of that method highly depends on the quality of applied metric in terms of given image database [3]. Hence the choice of metric to compare images becomes quite significant [8].

One of the most perspective methods of image recognition is based on image color histogram comparison [9], [10]. The histogram-based methods are very suitable for color image recognition, because such methods are unaffected by geometrical characteristics of the images, such as translation and rotation [10]. It's shown [11] that histogram-based methods could provide quality comparable with special methods based on specific information (i.e. methods for faces recognition [12]).

In this paper, we propose a novel histogram-based method which applies minimum discrimination information (MDI) principle [13]. It is known to be an effective tool for solving various problems of pattern recognition. Meanwhile, its capabilities have not been fully exploited. In particular, almost no studies have addressed the advantages of the MDI principle over traditional methods and approaches in problems of automatic image recognition, especially, for half-tone images as one of the most

complex cases in the theory and practice of pattern recognition [2, 3]. The present paper seeks to fill this gap.

The rest of the paper is organized as follows: Section 1 introduces image recognition problem and the usage of minimum discrimination information principle and Kullback-Leibler discrimination [14] to compare color histograms. In Section 3, we present metric properties of proposed decision statistics. In Section 4, we introduce new modification of MDEA [3], which can be used to reduce the amount of calculations needed for recognition. In Section 5, we present the experimental results and analyze the proposed method in the acute problem of face images recognition. Finally, concluding comments are presented in Section 6.

## 2. Minimum Information Discrimination Criterion

Let a set of  $R > 1$  half-tone images  $X_r = \left\| x_{uv}^r \right\|$ ,  $(u = \overline{1, U}, v = \overline{1, V})$  be specified. Here

$U$  and  $V$  are the image height and width, respectively,  $x_{uv}^r \in \{0, 1, \dots, x_{\max}\}$  is the intensity of an image point with coordinates  $(u, v)$ ;  $r$  is the reference number ( $r = 1, \dots, R$ ), and  $x_{\max}$  is the maximum intensity. It is assumed that the templates  $X_r$  define some classes of images, for example, as a method of protection against noise. Furthermore, 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. It is required to assign a new input image  $X = \left\| x_{uv} \right\|$  to one of the  $R$  classes.

The procedures for constructing decision rules for the problem are developed in accordance with some deterministic or statistical approach. The deterministic approach is currently the most widely used. This approach seeks to determine a certain distance (measure of similarity) between any pairs of objects. For image recognition, one often uses the criterion based on the standard metric  $l_1$ :

$$\rho_1(X / X_r) = \frac{1}{U \cdot V} \sum_{u=1}^U \sum_{v=1}^V \left| x_{uv} - x_{uv}^r \right| \rightarrow \min \quad (1)$$

However, this approach does not always provide satisfactory results. The first reason is well-known [15] variability of visual patterns. The second reason is the presence of noise in the input image  $X$ , such as unknown intensity of light sources or simply random distortions of some points of the image. In the deterministic approach, the indicated problems are usually solved by adding new images to database, which, in turn, leads to a dramatic increase in its size. The above-mentioned difficulties are overcome by invoking a statistical approach [7, 8].

Let's consider a random variable - template  $X_r$  color of a point. It's distribution

$$H_r = \left[ h_1^r, h_2^r, \dots, h_{x_{\max}}^r \right] \text{ could be evaluated based on matrix } \left\| x_{uv}^r \right\|$$

$$h_x^r = \frac{1}{U \cdot V} \sum_{u=1}^U \sum_{v=1}^V \delta(x_{uv} - x) \quad (2)$$

Here  $\delta(x) = \begin{cases} 1, & x=0, \\ 0, & \text{otherwise} \end{cases}$  -discrete Dirac delta-function.  $H_r$  is often called “color histogram” [10], [11] of image  $X_r$ . The same procedure for color histogram  $H$  definition is applied for the input image  $X$ . Color histograms’ comparison has widely been using in the image recognition problem [9], [10]. The common way to compare histograms is “merged histogram method” [10]

$$\sum_{x=1}^{x_{\max}} \min\{h_x^r, h_x\} \rightarrow \max_r \quad (3)$$

Unfortunately, images with the same visual information, but with shifted color intensity, may significantly degrade if the conventional method of direct comparison of histograms is used. Actually, if the input image  $X$  is one of the images  $X_r$  from database but all pixels are decolorized, its color histogram  $H$  will be quite different from  $H_r$ . The common approach for solving the problem is shifting histograms [11] after their evaluation using (2). This procedure may cause the increase of decision time. Hence in this paper we use normalization of images’ intensities [16] which could be done really efficient and doesn’t influence recognition’s average time

According to statistical approach, recognizing image  $X$  is supposed to be a received signal of noisy communication channel where transmitted signal is one of  $\{X_r\}$  image. Hence the problem is to minimize the mutual-entropy (or Kullback-Leibler information discrimination) [13] between color distributions of template image and recognizing image.

Based on this approach we assume that histogram  $H_r$  stands for the distribution of discrete certain random variable – image color. This interpretation seems justified considering the common properties of discrete distribution are valid for  $H_r$ :

$\sum_{x=1}^{x_{\max}} h_x^r = 1$  (normalization condition) and  $h_x^r \geq 0, x = \overline{1, x_{\max}}$  (regularization condition [13]).

Based on such image probability model, it is required to verify  $R$  hypotheses on the distribution  $H_r, r = \overline{1, R}$  of the input image signal  $X$ . It’s well-known [8], that the optimal decision of the problem of statistical check of hypotheses about distribution of discrete stochastic variable in Bayesian terms is equivalent to the minimum discrimination information principle and the optimal decision rule

$$\rho_{KL}(X/X_r) = \sum_{x=1}^{x_{\max}} h_x \ln(h_x/h_x^r) \quad (4)$$

Here the statistic  $\rho_{KL}(X/X_r)$  defines the Kullback–Leibler information discrimination [14] between the observed image  $X$  and its  $r$ th template from set  $\{X_r\}$ .

Thus, the image recognition procedure in this case involves a multichannel processing scheme in which the number of channels  $R$  is given by images' database size. Decision making is based on the statistic minimum criterion from expressions (1) or (3) for traditional image recognition methods [1] or from expression (4) when using proposed criterion and the Kullback-Leibler discrimination [3], [14].

### 3. Metric Properties of Information Discrimination Decision Statistics

We consider the most important and difficult case  $R \gg 1$ , where the image recognition problem is solved with a template images' set containing hundreds and thousands of images. For the specified conditions, practical implementation of the optimal decision rule (4) by the  $R$ -channel processing scheme encounters the obvious problem of its computational complexity and even feasibility, especially considering the labor-consuming procedure of image equalization in accordance with multiple parameters: size, color, project view, etc. The present work seeks to develop methods other than complete enumeration of reference images' set to solve the above problem. We first note the metric properties of the Kullback-Leibler discrimination decision statistic  $\rho_{KL}(X/X_r) \geq 0$ , which is equal to zero only in the ideal case of coincident input and template signals. Therefore, we first transform the minimum discrimination information criterion (4) to a simplified form, suitable for practical implementation [3]:

$$W_\nu(X): \rho_{KL}(X/X_\nu) < \rho_0 = const \quad (5)$$

Here  $\rho_0$  is the threshold for the admissible information discrimination on the set of similarly-named images due to their known variability. The value of this threshold is easy to find experimentally by fixation of the beta error probability.

$$\beta = P\{\rho_{KL}(X/X_\nu) < \rho_0 \mid \bar{w}_\nu\}$$

It's known [13], [17], that if recognizing image distribution is the same as for template  $X_\nu$ , then  $2UV \cdot \rho_{KL}(X/X_\nu)$  is distributed according to the chi-square distribution with  $x_{\max} - 1$  degrees of freedom. For other classes  $X_r, r \neq \nu$ , random variable  $2UV \cdot \rho_{KL}(X/X_\nu)$  is distributed according to the noncentral chi-square

distribution with  $x_{\max} - 1$  degrees of freedom and noncentral parameter  $\lambda = 2UV \cdot \rho_{KL}(X_r / X_v)$ . Thus, in practice ( $U \cdot V \gg 1$ ), threshold could be determined from the following expression

$$\beta = \frac{1}{R} \sum_{r=1}^R H\left(\rho_0 - \min_{v \neq r} \rho_{KL}(X_r / X_v)\right) \quad (6)$$

where  $H(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}$  is a Heaviside step function.

In fact, expression (5) defines the termination condition for the enumeration procedure using the MDI criterion (4). Thus, in decision-making process based on the minimum discrimination information principle (4), instead of looking through all templates, one needs to calculate the value of Kullback-Leibler divergence only until it becomes smaller than a certain threshold level. It is easy to see that this circumstance should reduce the amount of enumeration by 50% in average. A natural development of this idea is the MDEA described below, which fully exploits the metric properties of the Kullback-Leibler decision statistic (4).

#### 4. Method of Directed Enumeration Alternatives

Following the general computation scheme (2), (4), we reduce the image  $X$  recognition problem to a check of the first  $N$  variants  $X_1, \dots, X_N$  from the specified database  $\{X_r\}$  subject to the condition  $N \ll R$ . If, at least, one of them, namely  $X_v, v \leq N$ , meets the stopping requirement (5), the enumeration of the optimal solution by the minimum information discrimination criterion (4) will end with it. However, it can generally be assumed that none of the first  $N$  alternatives passes the check (5) in the first step. 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 (5) is satisfied. There is also another, more rational, method to solve the problem in question.

Following the definition of information discrimination (4), we generate an  $(R \times R)$  matrix  $P = \|\rho_{ij}\|$  of values

$$\rho_{ij} = \rho_{KL}(X_i / X_j), \quad i, j \leq R$$

This computationally complex operation needs to be made only once: in the preliminary computation step and for each concrete set of alternatives.

Let us arrange the images of the first control sample  $X_1, \dots, X_N$  in decreasing order of their information discriminations

$$\rho_{KL}(X / X_n), n = \overline{1, N}$$

As a result, we have an ordered (ranged) sequence of template images

$$\{X_{i_j}\} = \{X_{i_1}, X_{i_2}, \dots, X_{i_N}\}, i_j \leq N$$

The corresponding sequence  $\{\rho_j\}$  of their discriminations  $\rho_j = \rho_{KL}(X / X_{i_j}), i_j \leq N$ , is a monotonically decreasing dependence  $\rho_1 \geq \rho_2 \geq \dots \geq \rho_N \geq 0$

This procedure is used to obtain a number of data  $\{X_{i_j}\}$  ranged by the MDI criterion (4), and ultimately find the first local optimum  $X_{i_N}$ . The first computation step ends with this. In the second step, for the distinguished template image  $X_{i_N}$  from the matrix P, we find the set of  $M < R$  images  $X_{i_N}^{(M)} = \{X_{i_{N+1}}, \dots, X_{i_{N+M}}\}, i_j \leq R$  that are separated from the image  $X_{i_N}$  by the distance (4) not exceeding the threshold value  $\rho_N = \rho(X / X_N)$ :

$$\left( \forall X_i \notin X_{i_N}^{(M)} \right) \left( \forall X_j \in X_{i_N}^{(M)} \right) \Delta\rho(X_i) \geq \Delta\rho(X_j) \quad (7)$$

Here

$$\Delta\rho(X_j) = \left| \rho(X_j / X_{i_N}) - \rho_{i_N} \right| \quad (8)$$

is the deviation of the information discrimination  $\rho(X_j / X_{i_N})$  relative to the discrimination between the pair of images  $X$  and  $X_{i_N}$ . To this set we add one more  $(M+1)$ -th element  $X_{i_{N+M+1}}$  that did not fall in the control sample in the previous computation step. This brings some randomness to the search procedure as a method of attaining a global optimum in a finite number of steps (computation steps). For the analysis we obtain the second control sample of template images

$$\{X_{i_1}, \dots, X_{i_N}, \dots, X_{i_{N+M+1}}\}, i_j \leq R$$

Next, all computations of the first step are repeated cyclically until, in some  $K$ th step, an element  $X^* = X_{i_N}$  satisfies the termination condition:

$$\rho_{KL}(X / X^*) < \rho_0 \quad (9)$$

At this moment, the input image is within the set of the control points of the last computation step. In this case, a decision is made in favour of the closest pattern  $X^*$  or, at worst, after enumeration of all alternatives from the set  $\{X_r\}$  but in the absence of a solution from (9), the conclusion is drawn that the input image  $X$  cannot be assigned to any class from template images' set and that it is necessary to switch to the decision feedback mode.

Generally, there may be a considerable gain in the total number  $N + M \cdot K \leq R$  of checks carried out according to (7) compared to the size of the used set of alternatives. It's explained by the fact that probability  $p$  of desired image  $X^*$  containing in  $X_{i_N}^{(M)}$ , usually exceeds the probability of belonging  $X^*$  to  $M$  alternatives for random choice

$$p = P\left\{X^* \in X_{i_N}^{(M)}\right\} \gg p_0 = \frac{M}{R} \quad (10)$$

This is the effect of the directed enumeration. The difference in the steps' count  $K$  can be explained by the depending probability  $p$  from applied metric and properties of input image and give images' set.

Thus, the system of expressions (2), (4)–(9) defines the proposed MDEA modification in the image recognition problem. This modification differs from the original method [3] in the way of evaluation  $X_{i_N}^{(M)}$  set (6). In the initial version discrimination extrapolation based on autoregressive model [20] was used causing increase in the amount of calculations.

## 5. Experimental Results

The experiment deals with the problem of face images recognition. We use a real large database of photographs of people [21]. The photos were preliminary processed to detect faces using OpenCV library. Then detected faces were spited into 16 (4x4) parts for information discrimination computation. Such fragmentation is used to take into account heterogeneous illumination of images. The discrimination between images was calculated as a sum of discriminations (4) between these parts.

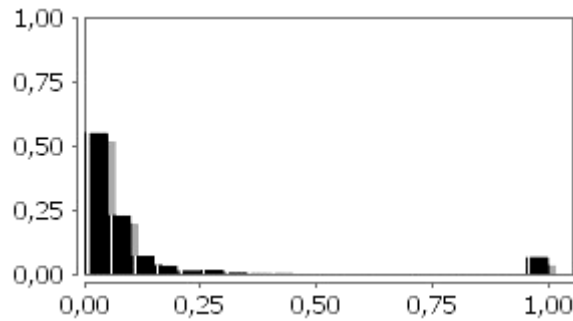
The 6000 photographs of 400 different people were selected as templates  $R=900$  of the most different images using standard clusterization [7]. Thus, the number of

elements involved in search procedures at the next stages of the algorithms was decreased in 5 times.

In addition, 1000 more photographs of the same people were used in the test. These test images were modified by reducing the intensity of all of their points (blacking) to demonstrate the illumination's influence. In each case, the problem of image  $X$  recognition was solved.

In the first case the metric (1) was used and the following method parameters were chosen:  $N=9$  and  $M=64$ . The threshold  $\rho_0 = 30$  was chosen experimentally. Here the

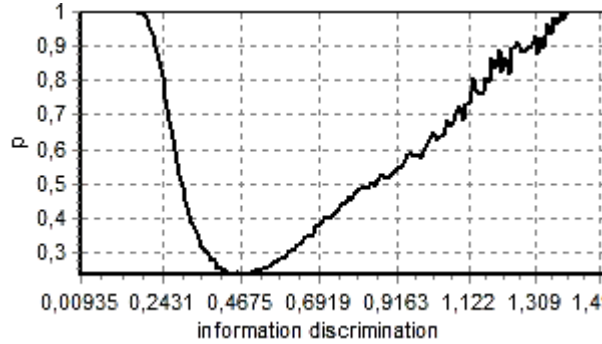
exact solution  $X^*$  (the same person photograph as from input image  $X$ ) was obtained in 90.5% of the cases. For each solution, it was required to check, in average, 51% of the total size of images database  $R$ .



**Fig. 1.** Histogram of the Number of Checks per Template Images' Database Size

In the second case illuminated test images and the Kullback–Leibler metric (4) were used. Threshold  $\rho_0 = 0.019$  was determined from (5) by fixation beta-error probability to 5%. Using the MDEA (2), (4)–(9) with the parameters  $N$ ,  $M$  from the previous experiment, we obtained an average number of checks equal to 13% of  $R$ . A histogram of the number of checks carried out by MDEA for this case is shown in Fig. 1. With a probability of 90%, the number of template images to check does not exceed 15% of  $R$ . In this case, condition (4) was not satisfied for any template from the given database for 7.1% of the initial images; therefore, all  $R$  alternatives were checked. In 96% of the cases, the exact solution was obtained.





**Fig. 2.** Dependence of Probability  $p = P\left\{X^* \in X_{i_N}^{(M)}\right\}$  on  $\rho_{KL}\left(X / X_{i_N}\right)$ .

The probability of desired image  $X^*$  containing in  $X_{i_N}^{(M)}$  from (10) for Kullback-Leibler discrimination was equal to  $p=33\%$  which is quite greater then  $p_0 = \frac{M}{R} = \frac{64}{900} = 7\%$ . Dependence of probability  $p$  on the discrimination  $\rho_{KL}\left(X / X_{i_N}\right)$  is shown at Fig.2.

There is an unquestionable advantage in using the Kullback–Leibler discrimination in this problem and this is due to the fact that, in our example, the input images were artificially distorted (blanked). Let us now consider the case where all images are equally illuminated. Use of the metric  $l_1$  from expression (1) and the directed enumeration method with a threshold  $\rho_0 = 15$  gives much better results then it was for the first experiment. The error probability reduces to 6%, and the average number of checks decreases to 21% of  $R$ . However, the Kullback–Leibler metric (3) seems to be more efficient even in that case of equal illumination. Those metric and proposed methods show practically the same result as for the previous experiment. The error probability reduces to 3,2 with 12% of average number of checks.

For comparison, merged histogram method (3) was used. The error probability for the second experiment (non-illuminated images) was estimated to 3,1% with 11% average number of checks according to MDEA. However, in the first experiment with illuminated test images the quality of (3) was much worse – 6,5% error with 20% of average distance (3) calculation.

## 6. Conclusion

The problem of increasing the computation speed has attracted considerable interest of experts in both the theory and practice of image recognition. Despite huge

number of approaches, most of the algorithms compare an input image with each template image, and unavoidably cannot be implemented in real-time mode for large databases. For solution of that problem directed enumeration method [3] may be used. This method is based on an information theoretical approach which uses the metric properties of the Kullback-Leibler decision statistic [3], [13] and possesses wide functional capabilities and high operational properties. The point of fundamental importance in this enumeration method is the termination criterion (4). Even in the most unprofitable case, the method almost halves the amount of computations. The use of the proposed method in the formulation (2)–(8) reduces the computational complexity by 10–20%. If proposed discrimination is combined with clustering [7] of the database, the performance increases in 40-50 times.

It's shown that efficiency of proposed method depends on the quality of applied metric in terms of given database. Thus, the metric choice becomes very important. In this paper, the new histogram-based method using minimum discrimination information principle is proposed for histogram-based image recognition. It's based on the dominant colors in images. The method is very suitable for color image recognition because it is unaffected by geometrical changes in images, such as translation and rotation. However, images with similar visual information but with shifted color intensity may result in a significant degradation in the similarity level, if the conventional histogram intersection method is used. To solve the problem, the histogram method in Kullback-Leibler metric was proposed. Our experimental results show that histogram methods (3), (4) of image recognition have significantly higher recognition effectiveness than the standard methods using  $l_j$  metric for pixels comparison.

Our method of directed enumeration alternatives is more straightforward than traditional recognition methods used with large databases [4]. It doesn't have special requirements or extra restrictions for images to recognize. Meanwhile, the quality of the solution reached by the method is comparable to that obtained by continuous enumeration of given database [12] using special methods of faces recognition. The main advantage is that we reach the same quality with 50-times reduce of the computational complexity using proposed method.

Our further work on image recognition will continue in the following directions:

- investigating more effective halftone image models (for example, models based on gradient direction features),
- finding metric thresholds to increase recognition accuracy,
- presenting experiments with recognition for most popular image databases (FERET, Yale, AT&T, etc.).

## References

1. Rui Y., Huang T., Chang S.F.: Image retrieval: current techniques, promising directions and open issues, *Visual Communication and Image Representation* 10, 39–62 (1999)
2. Flickner M., et al.: Query by image and video content: The QBIC system. *IEEE Computer* 28(9), 23–32 (1995)

3. A. V. Savchenko: Method of directed enumeration of alternatives in the problem of automatic recognition of half-tone images, *Optoelectronics, Instrumentation and Data Processing* 45(3), 83–91 (2009).
4. Jia Z., Amselang L., Gros P.: Content-based image retrieval from a large image database, *Pattern Recognition* 11(5), 1479-1495 (2008).
5. Huet B., Hancock E.R.: Shape recognition from large image libraries by inexact graph matching, *Pattern Recognition Letters* 20(11-13), 1259-1269 (1999).
6. Cover T.M., Hart P.E.: Nearest Neighbor Pattern Classification, *IEEE Trans. Information Theory* 13, 1968, 21-27
7. Theodoridis S., Koutroumbas C.: *Pattern Recognition*, 4th edn. Elsevier Amsterdam (2009)
8. Eickeler S., Jabs M., Rigoll G.: Comparison of Confidence Measures for Face Recognition, Fourth IEEE International Conference on Automatic Face and Gesture Recognition (FG'00), 257-263 (2000),
9. Min R., Cheng H.D.: Effective image retrieval using dominant color descriptor and fuzzy support vector machine, *Pattern Recognition* 42(1), 147-157 (2009).
10. Wong K.M., Cheung C.H., Po L.M.: Dominant Color Image Retrieval using Merged Histogram. Proc. the 2003 Int. Symposium, (2003).
11. Yoo G.-H., Kim B.K., You K.S.: Content-based image retrieval using shifted histogram, ICCS, LNCS 4489, 894–897 (2007).
12. Foon N. H., Jin A. T. B., Ling D. N. C.: Face Recognition Using Wavelet Transform and Non-negative Matrix Factorization, Proc. 7th Australian Joint Conference on Artificial Intelligence, Cairns, 192-202 (2004)
13. Kullback S.: *Information Theory and Statistics*, Dover Pub., New York (1978).
14. Kullback S., Leibler R.A.: On information and sufficiency, *Annals of Mathematical. Statistics* 22, 79–86 (1951).
15. Oppenheim A. V.: *Discrete-time signal processing*, Prentice-Hall, (1989).
16. Zhao W., Chellappa R. ed. *Face Processing: Advanced Modeling and Methods* Elsevier/Academic Press, (2005)
17. Kupperman M.: Further applications of information theory to multivariate analysis and statistical inference, Dissertation, Graduate Council of George Washington University, (1957).
18. Gonsalvesh M., Papa J., Zhang B. etc: A genetic programming framework for content-based image retrieval, *Pattern Recognition* 42(2), 283-292 (2009).
19. Voskoboinikov Yu.E., Litvinov L.A.: Choosing the stopping iteration in iterative algorithms of image and signal reconstruction, *Optoelectronics, Instrumentation and Data Processing* 40(4), 3–9 (2004).
20. Marpl S L (Jr): *Digital Spectral Analysis with Applications*, Englewood Cliffs, N.J.: Prentice-Hall, (1987)
21. Essex Faces database, <http://cswww.essex.ac.uk/mv/allfaces/index.html>