

ი. ქართველიშვილი

ორიგინალი და უცნობი ხელნაწერი სიმბოლოების

ბაზისური მოდელები

რეზიუმე

ნაშრომში წარმოდგენილია ორიგინალი და უცნობი ხელნაწერი სიმბოლოებისთვის ბაზისური მოდელების შექმნა, რომელიც წარმოადგენს ხელნაწერი სიმბოლოების ამოცნობის ერთ-ერთ სტადიას.

მოდელების შექმნის პროცესები ჩამოყალიბებულია ეტაპების სახით, სადაც თითოეული მათგანი დახასიათებულია თავისი ფუნქციონალური დანიშნულებებით.

საკვანძო სიტყვები: ეტალონური, ბაზისური, წონითი.

1. შესავალი:

ორიგინალი და უცნობი სიმბოლოების ბაზისებისთვის, კლასტერიზაციის პროცესის დასრულების შემდეგ თითოეული ჯგუფისათვის მოიძებნება ეტალონური აღწერა, რომელსაც ვუწოდოთ ბაზისური მოდელი. ბაზისური მოდელის შესაქმნელად, საჭიროა განვიხილოთ სწავლების პროცესი, როგორც მრავალბიჯიანი პროცესი. თითოეული ჯგუფისათვის საჭიროა მოხდეს ზღურბლური მნიშვნელობის შერჩევა და შემდეგ საუკეთესო გამოსახულებების დაჯგუფება ქვეჯგუფებში. ამ მოდელის ფარგლებში იქმნება შუალედური რგოლი, სადაც შეიძლება მოიძებნოს გამოსახულებები, რომლებსაც შეუძლიათ გამოიწვიონ დიდი რაოდენობის შეცდომები ამოცნობის პროცესში, ვიდრე ფაქტობრივი ამოცნობის დროს. დასკვნითი ბიჯი არის ის, რომ ამოგდებული იქნას ცუდი გამოსახულებები, ე.ი. მოხდეს შუალედური სამაგალითო გამოსახულებების რგოლის გაწმენდა და გამოსახულებების განცალკევება ორ კლასად.

სწორად კლასიფიცირებული გამოსახულება მიიღება მაშინ, როცა მიღებული მანძილი ნაკლები იქნება სხვა კლასის ზღურბლზე, ვიდრე იმ კლასისა, რომელიც განხილვის პროცესშია.

2. პირითადი ნაწილი:

ამოცნობის ზღურბლების შეფასების დროს ყოველთვის მთავარ საკითხად დგება შემდეგი: ამოგდებული იქნას ცუდი გამოსახულებები ეტალონური გამოსახულებების ფუძიდან. ამისათვის გამოითვლება სწორედ კლასიფიცირებული გამოსახულებების რაოდენობა და დაშვებული შეცდომების რაოდენობა. ეტალონი გამოსადეგია, თუ მისი საშუალებით შესაძლებელია უცნობი სიმბოლოების ამოცნობა. თუ ეტალონური გამოსახულებათა ნაკრები შემთხვევით შეიცავს ცუდ ეტალონებს, მაშინ ეს უარყოფითად მოქმედებს ზღურბლებზე, ამცირებს მათ და პროცესი მთავრდება ამოცნობის დაბალი საიმედოობით.

პირველადი შერჩევის პროცესი მდგომარეობს შემდეგში. ამოვარჩიეთ გამოსახულებათა მოცემული რაოდენობა ამოცნობით და ვუწოდეთ მას ეტალონი.

ერთის მხრივ, ორი პრობლემა მაინც რჩება. პირველი არის ის, რომ ამორჩეული ეტალონების ნაკრები შეიძლება არ შეიცავდეს წარმოდგენილი ხელნაწერების ყველა შესაძლებელ სტილს. ეს პრობლემა შეიძლება ჩათვლილიყო ადვილად გადასაწყვეტ საკმედ, თუ უბრალოდ გამოყენებული იქნებოდა დიდი რაოდენობის ეტალონები და ჩავთვლიდით, რომ ეს მოდელები იქნებოდა ეტალონური. მაგრამ ამას მიყვავართ მეორე პრობლემა, რომელიც იმაში გამოიხატება, რომ ეტალონური ნაკრების ზომა შეზღუდულია. ხეშემთ ადნიშნული პრობლემების გადასაწყვეტად, გამოყენებული იქნა ჯგუფის მახასიათებელი სიდიდე. თითოეული ჯგუფიდან განისაზღვრა პარამეტრი: ჯგუფის საშუალო არითმეტიკული, რომელიც ახასიათებს ამ ჯგუფს და რომლის გარშემოც ხდება ჯგუფის ელემენტების კონცენტრაცია.

ხელნაწერი სიმბოლოების ჯგუფები ხასიათდებიან გამოსახულების სიმძიმის ცენტრიდან ეტალონური წერტილებამდე მანძილების საშუალებით, რომელიც წარმოადგენს გარკვეული ნიშან-თვისების მიხედვით განსაზღვრულ n მოცულობის D სიდიდეს და აერთიანებს იმავე ნიშან-თვისების შესაბამისად n_1, n_2, \dots, n_k მოცულობებიან k რაოდენობის ემპირიულ ჯგუფს, სადაც k წარმოადგენს ეტალონური წერტილების რაოდენობას. მიიღება n მოცულობის ორი სახის სიმრავლე, რომლებიც k რაოდენობის ჯგუფად დაყოფის საფუძველზე, შეიძლება წარმოდგენილი იქნას შემდეგი სახით:

$$\overbrace{d_1, d_1, \dots, d_1}^{n_1} \quad \overbrace{d_2, d_2, \dots, d_2}^{n_2} \quad \overbrace{d_k, d_k, \dots, d_k}^{n_k} \quad (1)$$

სადაც n_i არის d_i მანძილების რაოდენობა, $i = \overline{1, k}$

$$d_1^{(1)}, d_2^{(1)}, \dots, d_{n_1}^{(1)}, \quad d_1^{(2)}, d_2^{(2)}, \dots, d_{n_2}^{(2)}, \quad d_1^{(k)}, d_2^{(k)}, \dots, d_{n_k}^{(k)} \quad (2)$$

სადაც $\{d_j^{(i)}\}$ არის ჯგუფის i -ური სიმბოლოს სიმძიმის ცენტრიდან ეტალონურ წერტილებამდე მანძილების ერთობლიობა და $j = \overline{1, n_i}$.

ორივე რიგისთვის შესაბამისი სისწორეები იქნება: n_1, n_2, \dots, n_k .

სწორადობები: $\frac{n_1}{n}, \frac{n_2}{n}, \dots, \frac{n_k}{n}$ და მოცულობა: $n_1 + n_2 + \dots + n_k = n$.

(2) რიგისთვის განისაზღვრება $\overline{D}_1, \overline{D}_2, \dots, \overline{D}_k$ კერძო საშუალო არითმეტიკულები:

$$\overline{D}_1 = \frac{d_1^{(1)} + d_2^{(1)} + \dots + d_{n_1}^{(1)}}{n_1}; \quad \overline{D}_2 = \frac{d_1^{(2)} + d_2^{(2)} + \dots + d_{n_2}^{(2)}}{n_2}; \quad \dots \quad \overline{D}_k = \frac{d_1^{(k)} + d_2^{(k)} + \dots + d_{n_k}^{(k)}}{n_k}. \quad (3)$$

საიდანაც:

$$d_1^{(1)} + d_2^{(1)} + \dots + d_{n_1}^{(1)} = \overline{D}_1 \cdot n_1, \quad d_1^{(2)} + d_2^{(2)} + \dots + d_{n_2}^{(2)} = \overline{D}_2 \cdot n_2, \quad \dots \\ \dots \quad d_1^{(k)} + d_2^{(k)} + \dots + d_{n_k}^{(k)} = \overline{D}_k \cdot n_k; \quad (4)$$

მაშინ (2) რიგის საშუალო იქნება:

$$\overline{D} = \frac{d_1^{(1)} + d_2^{(1)} + \dots + d_{n_1}^{(1)} + d_1^{(2)} + d_2^{(2)} + \dots + d_{n_2}^{(2)} + \dots + d_1^{(k)} + d_2^{(k)} + \dots + d_{n_k}^{(k)}}{n_1 + n_2 + \dots + n_k} \quad (5)$$

რომელშიაც კერძო ჯამების მაგიერ, მათი მნიშვნელობების ჩასმისას მიიღება:

$$\overline{D} = \frac{\overline{D}_1 \cdot n_1 + \overline{D}_2 \cdot n_2 + \dots + \overline{D}_k \cdot n_k}{n_1 + n_2 + \dots + n_k} = \frac{1}{n} \sum_{i=1}^k \overline{D}_i \cdot n_i \quad (6)$$

(6)-ით განსაზღვრული სიდიდე წარმოადგენს D სიდიდის ემპირიული (სტატისტიკური) მნიშვნელობების \overline{D} ემპირიული (სტატისტიკურ) წონით საშუალოს, რომელიც გამოითვლება ორივე ბაზის თითოეული ჯგუფისათვის და წარმოადგენს კლასტერის მახასიათებელ ნიშანს, რომლის მიხედვითაც შესაძლებელი იქნება ორიგინალი და უცნობი სიმბოლოების შედარება ჯგუფების მიხედვით.

3. დასკვნა:

ორიგინალი და უცნობი ხელნაწერი სიმბოლოებისათვის ბაზისური მოდელების შექმნა, რომელიც წარმოადგენს ხელნაწერი სიმბოლოების ამოცნობის ერთ-ერთ სტადიას, აუცილებელია იმისთვის, რომ შედარებისა და ამოცნობის პროცესი წარმატებით იქნას გადაჭრილი.

4. ლიტერატურა:

1. P. Scattolin. Recognition of handwritten numerals using elastic matching. The Department of Computer Science. November 1995.
2. ვერულავა ო. ხუროძე რ. ამოცნობის სისტემების თეორიის საფუძვლები. სტუ, 2001 წ.
3. ქართველიშვილი ი. ხელნაწერი სიმბოლოების წინასწარი კომპიუტერული დამუშავების სტადია. ტექნიკური უნივერსიტეტის სამეცნიერო შრომების კრებული №4(437) 2001 წ.

И. КАРТВЕЛИШВИЛИ

БАЗИСНЫЕ МОДЕЛИ ОРИГИНАЛЬНЫХ И НЕИЗВЕСТНЫХ РУКОПИСНЫХ СИМВОЛОВ

Резюме

Рассмотрен создать базисные модели оригинальных и неизвестных рукописных символов, который является один из стадий по распознаванию рукописных символов. Процессы создание моделей представлены по этапам. Все этапы охарактеризованы по функциональному назначению.

I. KARTVELISHVILI

BASIS MODELS OF ORIGINAL AND UNKNOWN HANDWRITTEN SYMBOLS

Resume

Work describes the creation of basis models for original and unknown handwritten symbols, which represents one of the stages of handwritten symbols decoding.

The processes of model design are established in stages and each of them is described by their functions.