

სკანირებადი ინფორმაციის კომბინირებადი დამუშავების ალგორითმი

იოსებ ქართველიშვილი, ავთანდილ ჯლანტიაშვილი
საქართველოს ტექნიკური უნივერსიტეტი

რეზიუმე

ნაშრომში წარმოდგენილია სკანირებადი ინფორმაციის კომბინირებადი დამუშავების ალგორითმი. თითოეული ალგორითმული ბლოკი გათვალისწინებულია ავტომატიზებული ამოცნობის გარკვეული ფუნქციის შესასრულებლად. ალგორითმული ბლოკები დეტალიზებულია ისეთი სახით, რომ შემდგომში შესაძლებელი და გაადვილებული იყოს პროგრამირების პროცესი.

საკვანძო სიტყვები: სკანირება. ალგორითმი. ამოცნობა. დაპროგრამება.

1. შესავალი

ავტომატიზებული ამოცნობის სისტემის მეთოდების საფუძველზე საჭიროა განისაზღვროს და ჩამოყალიბებული იქნეს ხელნაწერი სიმბოლოების ავტომატიზებული ამოცნობის ალგორითმები. თითოეული ალგორითმისთვის დადგენილია შემაგალი და გამომავალი მონაცემები, ჩამოყალიბებულია ორიგინალი და უცნობი სიმბოლოების მონაცემთა მასივები და ბაზები, რომელთა საფუძველზეც ჩატარებულია ალგორითმით გათვალისწინებული ხელნაწერი სიმბოლოების ამოცნობის პროცესები.

თითოეული ალგორითმული ბლოკი გათვალისწინებული უნდა იყოს ავტომატიზებული ამოცნობის გარკვეული ფუნქციის შესასრულებლად. ამოცნობის სისტემის ალგორითმიზაციის ეტაპზე საჭიროა მოვანდინოთ თითოეული ალგორითმული ბლოკის ისეთი სახით დეტალიზაცია, რომ შესაძლებელი და გაადვილებული იყოს პროგრამირების პროცესი.

2. ძირითადი ნაწილი

ხელნაწერი მონაცემების კომბინირებადი დამუშავების ალგორითმის მიხედვით პირველ რიგში ხდება ორიგინალი და უცნობი ხელნაწერი ტექსტების სკანირება და შესაბამისად მონაცემთა ბაზების ფორმირება. განვიხილოთ სკანირებადი ინფორმაციის კომბინირებადი დამუშავების ალგორითმი (ნახ.1).

ალგორითმის მუშაობის საწყის ეტაპზე ხდება ორიგინალი და უცნობი ხელნაწერი სიმბოლოების სკანირება და სისტემაში შეტანა. შემდეგ მოწმდება პირობა, თუ ხელნაწერი მონაცემი ორიგინალია, იქმნება ორიგინალი მონაცემების მასივი, თუ არა – უცნობი მონაცემების მასივი.

შემდეგ ალგორითმულ ბლოკში პროგრამული კოდის საშუალებით ხდება გრაფიკული ინფორმაციის გადაყვანა ორობით სისტემაში (ეს პროცესი სრულდება ორივე მასივის მონაცემებისათვის და ინფორმაცია წარმოდგენილი იქნება პიქსელების სახით).

აუცილებელია მთლიანი გამოსახულების დასუფთავება, სადაც ხორციელდება სკანირებული გამოსახულებისთვის არადაზიანებადი ელემენტების მოცილება.

სისტემაში შემოდის დამატებითი ცვლადები, რათა მოხდეს გამოსახულებაში თეთრი და შავი პიქსელების რაოდენობების დაფიქსირება;

გამოსახულების დასუფთავება ხორციელდება ყველა „1“-იანისთვის (შავი არე), მოწმდება მის გარშემო მდებარე 8 მეზობელი პიქსელი და რაოდენობები შესაბამისად ფიქსირდება დამატებით ცვლადებში. შემდეგ მოწმდება პირობა, თუ „1“-იანების რაოდენობა ნაკლებია „0“-იანების რაოდენობაზე მაშინ შავი პიქსელი გარდაიქმნება თეთრ პიქსელად.

შემდგომ ეტაპზე მიმდინარეობს დასუფთავებული გამოსახულების სეგმენტაცია სტრიქონებად, ფიქსირდება სეგმენტირებული სტრიქონის დასაწყისი და დასასრული, სკანირებულ ინფორმაციაში განისაზღვრება სტრიქონების რაოდენობა, რათა დადგინდეს ციკლის მიმდინარეობის პერიოდი და ხდება სეგმენტირებული სტრიქონების მასივის ჩამოყალიბება.

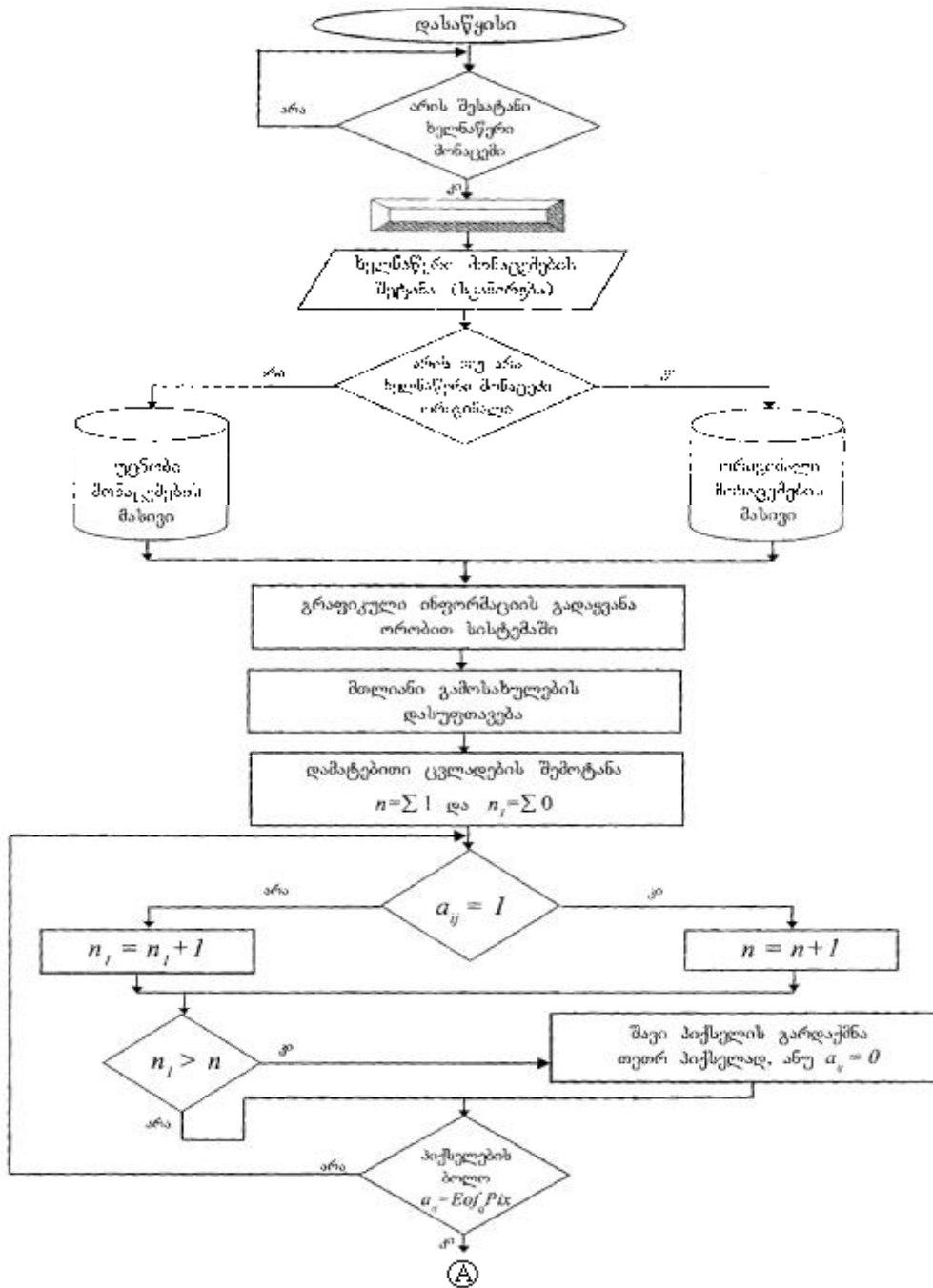
ალგორითმის შემდგომ ბლოკში ხდება სტრიქონებად წარმოდგენილი მასივის სეგმენტაცია სიტყვებად, ფიქსირდება სეგმენტირებული სიტყვის დასაწყისი და დასასრული, თითოეულ სეგმენტირებულ სტრიქონში განისაზღვრება სიტყვების რაოდენობა, რათა დადგინდეს ციკლის მიმდინარეობის პერიოდი და ჩამოყალიბდება სეგმენტირებული სიტყვების მასივი.

შემდგომ ეტაპზე ხდება სიტყვებად სეგმენტირებული მასივის ელემენტების წარმოდგენა სიმბოლოების სახით. მატრიცის სახით წარმოდგენილი გამოსახულების თითოეული სტრიქონის ელემენტებისათვის ფიქსირდება შესაბამისი კოორდინატები.

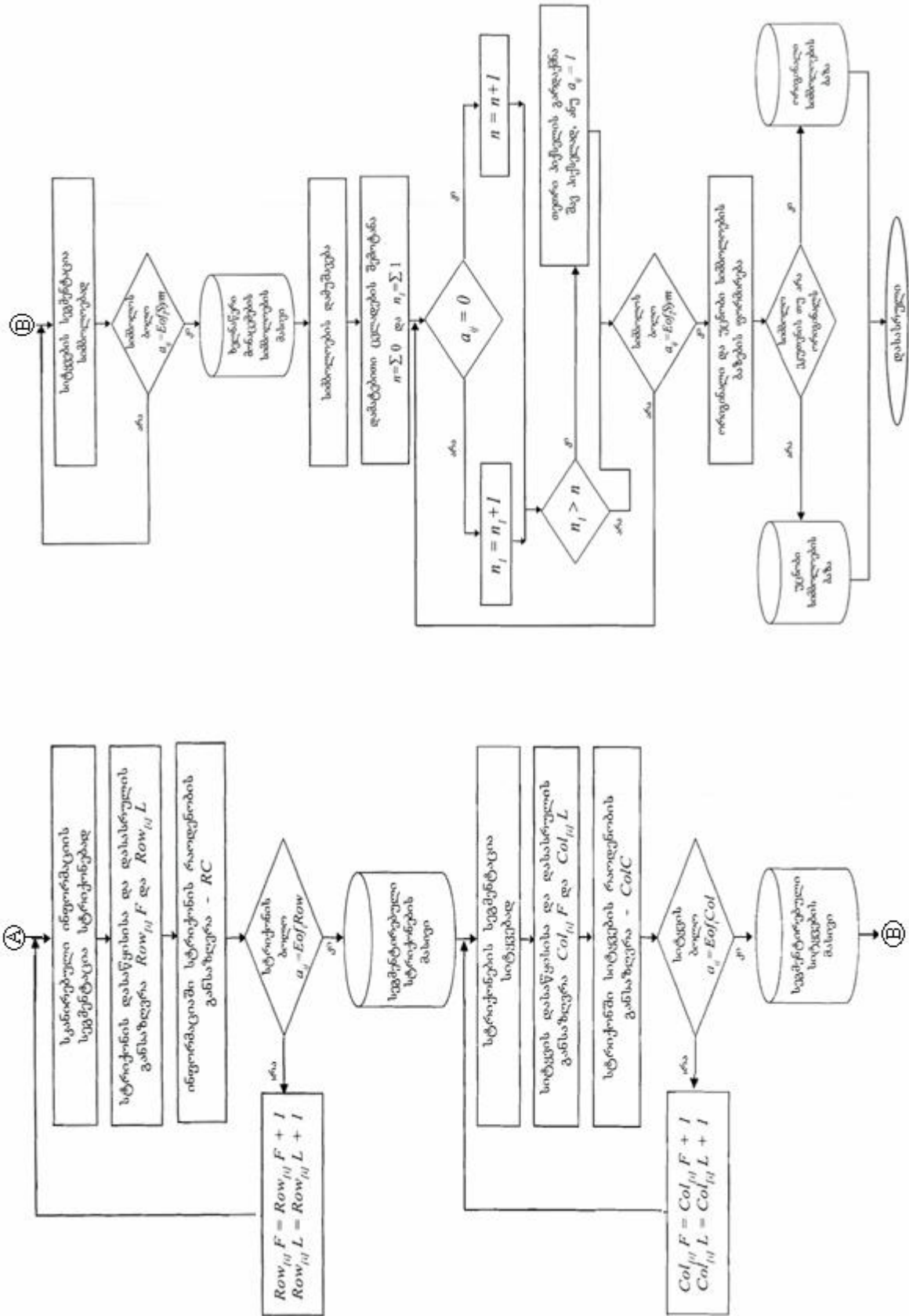
დაწყებული პირველი ელემენტიდან მატრიცის ბოლო ელემენტის ჩათვლით, მოწმდება ელემენტის გარშემო მდებარე 8 მეზობელი პიქსელი და პროგრამული კოდით გათვალისწინებული წესების თანახმად ფიქსირდება სიმბოლოების დასაწყისი და დასასრული. მთელი ეს პროცესი ციკლშია მოქცეული და სრულდება მატრიცის ბოლო ელემენტამდე.

სიტყვების სიმბოლოებად სეგმენტაციის შემდეგ ფორმირდება ხელნაწერი სიმბოლოების მასივი, სადაც განსაზღვრულია სიმბოლოების საზღვრები და სკანირებულ ინფორმაციაში სიმბოლოების რაოდენობა.

შემდგომ აუცილებელია ხელნაწერი მონაცემების სიმბოლოების მასივის თითოეული სიმბოლოს დამუშავება. ეს პუნქტი ანალოგიურია მთლიანი გამოსახულების დასუფთავებისა. პროგრამული კოდით გათვალისწინებული წესების დახმარებით გამოსახულება, რომელიც შედგენილია პიქსელებისაგან, იწმინდება ზედმეტი ნიშნებისაგან, ან ხდება მისი შევსება თუ ეს საჭიროა და სიმბოლო ღებულობს საჭირო სახეს.



ნახ.1. კომბინირებული დაბეჭდვის ალგორითმი (დასაწყისი)



ნახ.1. კომბინირებული დაბუშავების ალგორითმი (გაგრძელება)

ალგორითმის ბოლოს მოწმდება პირობა, თუ აღნიშნული სიმბოლოები ეკუთვნის ორიგინალს, მაშინ ხდება ორიგინალი სიმბოლოების ბაზის ფორმირება, თუ არა – უცნობი სიმბოლოების ბაზის ფორმირება, რაც საჭიროა ამოცნობის პროცესის შემდგომი ეტაპისათვის.

3. დასკვნა

ჩამოყალიბებულია სკანირებული ინფორმაციის კომბინირებული დამუშავების ალგორითმი და თითოეული ალგორითმული ბლოკი წარმოდგენილია და დეტალიზებულია ისეთი სახით, რომ შემდგომში გაადვილებული იყოს სისტემის პრაქტიკული რეალიზაცია.

ლიტერატურა

1. ქართველიშვილი ი. სკანირებული ინფორმაციის დამუშავების კომბინირებული მეთოდი. სამეცნიერო ჟურნალი „ინტელექტი“ №1(15), 2003 წ.
2. P. Scattolin. Recognition of handwritten numerals using elastic matching. The Department of Computer Science. November 1995.
3. ვერულავა თ. ხუროძე რ. ამომცნობი სისტემების თეორიის საფუძვლები. სტუ, 2001 წ.

АЛГОРИТМ ПО КОМБИНИРОВАННОЙ ПЕРЕРАБОТКИ СКАНИРОВАННОЙ ИНФОРМАЦИИ

Картвелишвили И., Джлантиашвили А.
Грузинский Технический Университет

Резюме

Представлен алгоритм по комбинированной обработке сканированной информации. Каждый из алгоритмных блоков учитывает автоматизированное распознавание по выполнению некоторых функций и каждый блок детализирован так, чтобы в дальнейшем был возможен и упрощен процесс программирования.

ALGORITHM FOR COMBINED PROCESSING OF SCANNED INFORMATION

Kartvelishvili Iisipf, Djlantiasvili A.
Georgian Technical University

Summary

The thesis presents the algorithm for combined processing of scanned information. Each algorithmic block is designed to carry out certain function for automatic identification. Each algorithmic block is presented in such a way that it makes possible and simplifies the programming process.