

## Analisis Teks Bahasa Indonesia Dan Inggris Dari Sebuah Citra Menggunakan Pengolahan Citra Digital

Akhmad Fadjeri<sup>1)</sup>; Atiq Rahmawati<sup>2)</sup>; Esa Restu Fadilah<sup>3)</sup>

<sup>1)2)3)</sup> Teknik Informatika, Universitas Ma'arif Nahdlatul Ulama kebumen

<sup>1)</sup>akhmadfadjeri@umnu.ac.id; <sup>2)</sup>atiqrahmawati46@gmail.com; <sup>3)</sup>esarestufadilah13@gmail.com

### ABSTRACT

*The purpose of this research is to analyze Indonesian text and English text from images using digital image processing. The method used in this research is an experimental method such as: literature review, identification problems, hypotheses, analyzing materials, designing programs, conducting tests and drawing conclusions. The result of this research is to detect the text contained in the image. This research result a digital image processing program in text that is containing from an image that has text in it, using input or input in form of a character image. The image that will be recognized into text with Tesseract OCR and output the text detected by the program. This program can show the level of accuracy to recognize characters and words. It conducted kind of testing on 8 images use several of text characters on it. In this case, the system can recognize characters on the text form, with an average accuracy 96% of the test data taken on images sourced from the internet. According to the testing results, this level of accuracy indicates that this application is quite accurate and efficient in identification the image of certain word or text character.*

**Keywords:** Text, Image, Digital Image Processing

### I. PENDAHULUAN

Pada era dahulu manusia belum bisa menulis dengan sempurna karena belum menggunakan huruf yang seperti sekarang, sehingga butuh waktu untuk dapat bisa menerjemahkan arti dari tulisan tersebut. Menulis merupakan kegiatan yang sangat penting dan bisa memberikan banyak informasi yang didapat. Menulis adalah salah satu bentuk untuk menyampaikan suatu kejadian dimasa lampau agar dapat diketahui oleh generasi penerus. Dengan berjalanya waktu manusia mulai mengenal simbol dan huruf untuk menulis dengan baik.

Manusia berkomunikasi dengan Bahasa Indonesia, Bahasa Inggris dan berbagai bahasa yang mereka bisa dan bahasa daerah mereka masing-masing. Untuk dapat menulis manusia membutuhkan media sebagai sarana, dahulu mereka masih menulis di media seadanya seperti batu, kulit pohon, dinding goa dengan menggunakan alat seadanya. Berkembangnya zaman, manusia mulai beralih dengan teknologi yang ada seperti menulis menggunakan komputer.

Perkembangan komputer sebagai media untuk menulis sangat pesat. Awal mula untuk menginput teks kedalam sebuah komputer, *user* melakukan secara manual dengan mengetik menggunakan keyboard, namun kini terdapat berbagai metode untuk menginput teks kedalam komputer salah satunya *text to*

*text, user* dalam hal ini hanya perlu memasukan gambar yang berisi teks agar terdeteksi pada komputer.

Selain metode input teks dengan suara, saat ini muncul metode baru dimana *user* dapat menginput teks yang berasal dari gambar atau citra kedalam komputer. Program ini menggunakan bahasa pemrograman *Python* dengan menerapkan metode *Image Processing* dengan input berupa gambar yang didapat dari internet atau hasil foto kamera yang didalamnya terdapat kata atau kalimat bertuliskan tangan, digital maupun hasil dari alat cetak.

Penelitian yang dilakukan oleh Munantri dan teman-teman pada tahun 2019 yang berjudul aplikasi pengolahan citra digital untuk identifikasi umur pohon memiliki kesamaan pada pengertian citra dan proses grayscale. Untuk memudahkan proses dipengolahan citra digital [1].

Penelitian yang dilakukan oleh Pambudi dan teman-teman pada tahun 2020 yang berjudul deteksi keaslian uang kertas berdasarkan watermark dengan pengolahan citra digital. Kesamaannya adalah pada pengolahan citra yang menggunakan inputan gambar untuk dilakukan proses [2].

### II. TINJAUAN PUSTAKA

#### 2.1 Citra

Citra adalah gambar yang di dihasilkan dari hasil pengamatan. Menurut Hornby citra adalah

gambar yang terekam kamera. Citra dapat diolah kedalam komputer yang kemudian disebut sebagai citra digital. Citra digital disimpan dalam komputer kedalam sebuah angka-angka atau yang menunjukkan besar intensitas pada masing-masing *pixel* [1].

Citra bisa diartikan sebagai ilustrasi pada suatu objek yang sedang diteliti sebagai hasil rekaman menggunakan alat pemantauan seperti kamera. Citra dapat disimpulkan gambar yang dihasilkan lalu dioalah didalam komputer [3].

## 2.2 Pengolahan Citra

Pengolahan citra adalah metode yang digunakan untuk memperbesar kualitas sebuah citra didalam komputer. Metode ini bertujuan untuk mengolah citra diubah dengan input citra dan menghasilkan *output* citra sesuai keinginan. Pengolahan citra di digolongkan menjadi 3 yaitu citra warna, citra biner dan citra *grayscale*/hitam putih [4].

Pengolahan citra terbagi 2 tahap yaitu citra digital dan citra analog. Citra digital diolah didalam komputer, sedangkan citra analog digunakan untuk pengopian dalam bentuk *print out* [5].

## 2.3 Citra RGB

Citra RGB merupakan warna aditif dengan deskripsi partikel warna merah, hijau dan biru yang ditambahkan secara bersama untuk mengasilkan sebuah warna-warna [6]. Citra RGB representasikan kedalam layer televisi dengan sebuah *pixel*, jika semakin banyak *pixel* yang terkandung maka gambar akan semakin jernih untuk dilihat [7].

## 2.4 Citra Grayscale

Citra *grayscale* adalah citra yang mempunyai satu warna yaitu keabuan. Citra *grayscale* berasal citra warna/RGB yang di proses pada *Color Space* sehingga berubah warna menjadi keabuan. Perubahan citra RGB ke *Grayscale* adalah untuk menyamakan informasi warna pada saat pemrosesan elemen citra. Proses warna abu-abu adalah satu warna yang terdapat pada komposisi warna RGB [8].

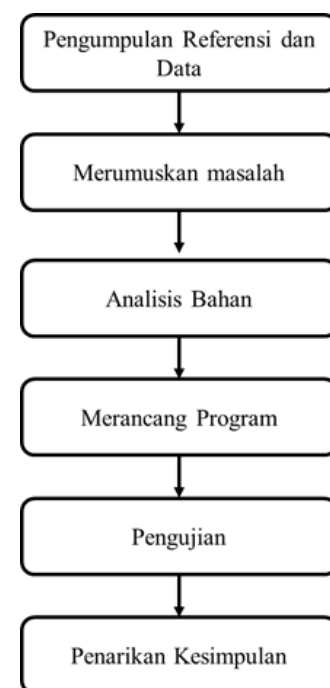
## 2.5 OCR (Optical Character Recognition)

OCR atau pengenalan karakter optik ialah sebuah sistem didalam komputer dimana sistem dapat mengartikan, mengolah teks dengan membaca karakter dari sebuah citra yang berasal dari alat cetak atau tulisan tangan pada format teks yang bisa dibaca oleh komputer [9].

## III. METODE PENELITIAN

Pada penelitian ini menggunakan teknik eksperimen. Teknik ini digunakan untuk pengujian pada sebuah citra terhadap citra lain untuk pengujian bagaimana hubungan sebab akibat antara citra satu dengan lainnya [2].

Dalam pelaksanaan perancangan program, tahap yang dilaksakan penelitian ini dapat dilihat dari bagan berikut.



Gambar 1. Tahap Penelitian

Penjelasan dari Gambar 1 tahapan penelitian ada 6 proses sebagai berikut:

### a. Pengumpulan referensi dan data

Sebelum melaksanakan penelitian terlebih dahulu peneliti mengumpulkan referensi atau data baik berupa jurnal penelitian sebelumnya serta referensi lain yang bersumber dari internet maupun dari pembimbing. Setelah referensi terkumpul yang relevan dengan topik yang akan di bahas pada penelitian, peneliti kemudian mempelajari ketertarikan bahan serta metode metode yang di pakai pana penelitian sebelumnya.

### b. Merumuskan masalah

Untuk lebih memfokuskan topik penelitian, peneliti kemudian membuat rumusan masalah dalam hal ini berkesesuaian dengan topik yang akan di bahas yakni bagaimana mengolah sebuah gambar yang diambil dari internet ataupun hasil foto kamera yang didalamnya terdapat sebuah teks, kemudian mengubahnya menjadi teks melalui sebuah program, dengan memanfaatkan metode pengolahan citra digital.

### c. Analisis bahan

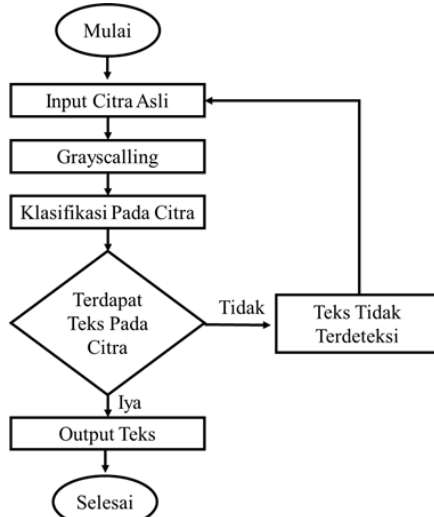
Sebelum membuat sebuah program peneliti mengumpulkan bahan citra yang akan diinput dan dijadikan bahan penelitian, peneliti mengambil bahan citra yang bersumber dari internet dan kamera ponsel. Kemudian mengolahnya agar nantinya dapat diterima oleh program.

Menggunakan bahasa pemrograman *python*, peneliti kemudian mempelajari program yang berkesuaian dengan topik dengan menginstal fungsi-fungsi serta membuat sebuah sintaks codingnya.

### d. Merancang program

Pada tahap ini peneliti membuat sebuah program menggunakan bahasa pemrograman *python*, serta menginstal fungsi-fungsi yang di perlukan. Beberapa sintaks coding yang di pakai memiliki fungsi keterikatan dengan topik penelitian. Sebagai gambaran untuk mengidentifikasi teks yang terdapat pada citra peneliti menggunakan model tesseract OCR (*Optical Character Recognition*) dan menambahkan sintaksnya kedalam bahasa pemrograman *python* [10].

Untuk lebih memahami alur program berikut peneliti paparkan kedalam diagram flowchart alur sistem dalam perancangan program pengolahan citra digital untuk mengidentifikasi teks.



**Gambar 2. Flowchart**

Penjelasan dari flowchart pada Gambar 2 sebagai berikut :

- Langkah pertama adalah input citra kedalam folder direktori kerja serta memanggilnya dengan sintaks coding di program yang sudah di buat di dalam komputer.
- Langkah berikutnya adalah pengolah citra,

kita bisa pastikan citra yang kita input telah dapat diidentifikasi oleh program.

- Kemudian proses pengolahan citra di lakukan di dalam program adalah mengubah warna asli dari yang semula RGB menjadi grayscale. Tujuannya mengurangi informasi yang tidak di perlukan pada citra yang diidentifikasi.
- Program pada nantinya memindai informasi yang di dapat pada citra serta memilih mana yang di butuhkan.
- Selanjutnya adalah program akan mengidentifikasi teks yang ada pada citra, Dengan cara memindai karakter yang terdapat di dalam citra yang di uji.
- Program memindai dan di temukan atau tidak temukan teks pada citra.
- Hasil menjalankan program, menampilkan citra grayscale yang sebagai indikator pengolahan citra digital berhasil berjalan pada program.
- Bila diidentifikasi terdapat teks dalam citra langkah selanjutnya program akan mengubahnya kedalam output teks yang dapat diolah oleh komputer.

### e. Pengujian

Setelah program di buat serta memasukkan bahan dalam hal ini citra kedalam file kerja proram langkah yang utama adalah pengujian. Pada tahap ini merupakan inti dimana menentukan berhasil tidaknya penelitian. Objek yang diidentifikasi berupa citra yang terdapat teks di dalamnya kemudian program mengolahnya menjadi teks

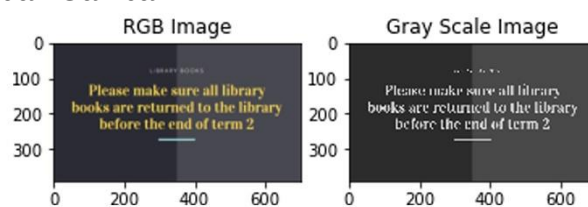
### f. Penarikan kesimpulan

Program dapat dikatakan berhasil tatkala tahap tahap yang telah di rancang, di buat, serta dilakukan sebelumnya berjalan dan menghasilkan output yang sesuai dengan yang diharapkan.

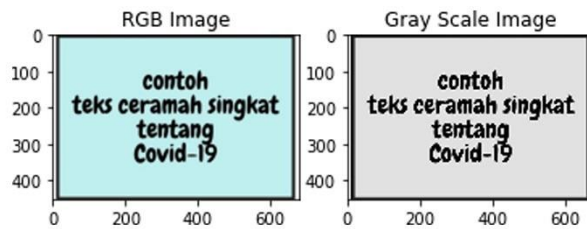
## IV. HASIL DAN PEMBAHASAN

### 4.1 Proses Grayscale

Setelah input citra asli langkah selanjutnya agar lebih dalam mendeteksi sebuah teks citra asli menjadi grayscale seperti pada Gambar 3 dan Gambar 4.



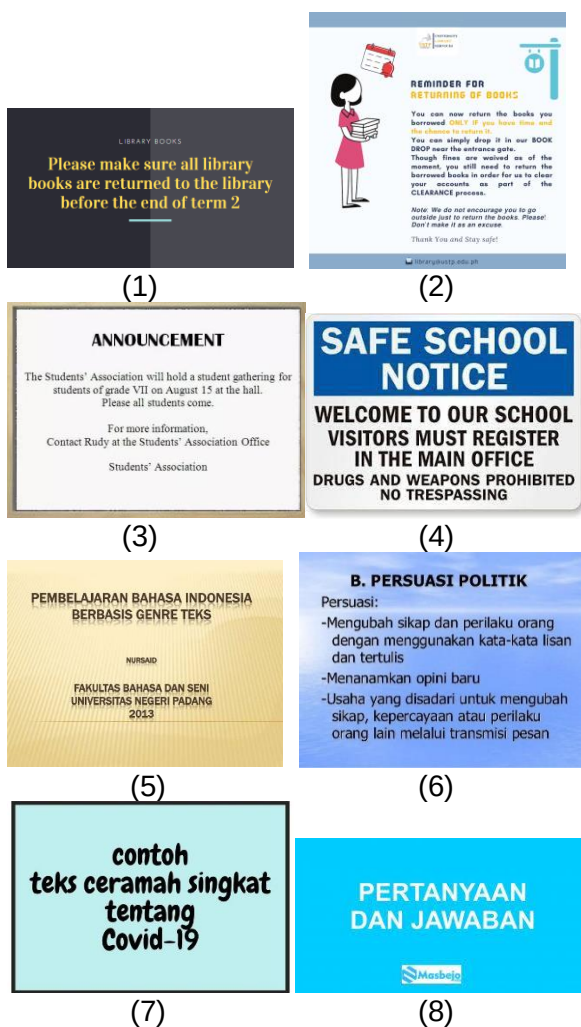
**Gambar 3. Contoh Teks Bahasa Inggris**



Gambar 4. Contoh Teks Bahasa Indonesia

Hasil dari program ini adalah untuk mendeteksi teks yang terdapat pada gambar kemudian mengolahnya menjadi teks. Peneliti menggunakan citra yang diperoleh dari hasil *searching* di internet serta hasil foto kamera.

Citra yang digunakan pada penelitian yaitu ada teks Bahasa Inggris dan Bahasa Indonesia.



Gambar 5. Citra yang digunakan

Gambar 5 diatas berisi sampel citra asli yang digunakan pada penelitian ini, terdiri dari 4 citra Bahasa Inggris dan 4 citra Bahasa Indonesia.

#### 4.2 Hasil Pengujian Perbandingan Teks

Untuk menghitungnya masih sama setelah *output* dari program keluar penelitian menggunakan perhitungan manual di excel menghitung jumlah katanya.

Pencarian jumlah kata yang terdapat pada citra penelitian menggunakan perhitungan manual dengan excel, dengan menggunakan rumus pendekatan umum untuk menghitung jumlah kata pada sel excel seperti pada rumus (1).

$$JK=IF(LEN(TRIM(Q))=0,0,LEN(TRIM(Q))-LEN(SUBSTITUTE(Q," ",""))+1) \quad (1)$$

Keterangan:

JK = Jumlah kata

IF = Untuk membuat perbandingan logis antara nilai dengan perkiraan

LEN = Untuk menghitung jumlah karakter dalam teks

TRIM = Untuk menghapus karakter spasi dari teks kecuali spasi tunggal di antara kata

Q = Teks yang diidentifikasi

SUBSTITUTE = Untuk membuang karakter spasi dari teks

Sementara untuk menghitung persentase keberhasilan program dalam memindai teks di dapat juga dari rumus (2).

$$\text{Persentase} = \frac{\text{Jumlah kata yang terdeteksi}}{\text{Jumlah kata keseluruhan}} \times 100\% \quad (2)$$

Hasil perhitungan jumlah kata antara jumlah kata awal yang dihitung menggunakan Excel dibandingkan dengan yang terdeteksi di sistem diperlihatkan pada Tabel 1.

Tabel 1. Hasil Pengujian

| No               | Jumlah Kata Awal | Jumlah Kata Terdeteksi | Presentase     |
|------------------|------------------|------------------------|----------------|
| 1.               | 19               | 19                     | 100%           |
| 2.               | 93               | 82                     | 88%            |
| 3.               | 36               | 36                     | 100%           |
| 4.               | 20               | 20                     | 100%           |
| 5.               | 15               | 15                     | 100%           |
| 6.               | 33               | 33                     | 100%           |
| 7.               | 7                | 6                      | 85%            |
| 8.               | 3                | 3                      | 100%           |
| <b>Rata-rata</b> | <b>225</b>       | <b>26,875</b>          | <b>96,625%</b> |

Berdasarkan Tabel 1, prosentase keberhasilan mencapai 96,625%.

## V. PENUTUP

Dari hasil penelitian dan pembahasan diatas maka kesimpulannya sebagai berikut :

1. Hasil penelitian program ini adalah pengolahan citra digital menggunakan inputan berupa gambar citra yang didalamnya terdapat karakter teks kemudian mengidentifikasinya serta menjadikan output berupa teks yang dapat langsung diolah oleh komputer.
2. Program ini dibuat dengan menggunakan bahasa pemrograman python. Dalam hal ini menggunakan modul tesseract untuk mengenali kata atau teks pada citra inputan.
3. Program ini menghasilkan keakuratan ketepatan sebesar 96% dari 8 citra yang menjadi sampel. Dalam hal ini sistem dapat mengenali karakter teks yang terdapat pada citra dengan akurasi rata rata sebesar 100% yang bersumber dari internet. Hasil pengujian tingkat keakuratan ini menunjukkan bahwa program ini cukup akurat dan efisien dalam mengenali citra karakter teks.
4. Modul tesseract pada bahasa pemrograman python mampu mengidentifikasi teks yang terpat pada citra serta mengubahnya kedalam bentuk teks yang dapat langsung diolah oleh komputer, namun tidak semua citra mampu mencapai 100% identifikasi oleh program.
5. Terdapat kata pada citra yang tidak mampu di identifikasi oleh program sehingga output teks yang di hasilkan pun kurang sesuai.

## DAFTAR PUSTAKA

- [1] N. Z. Munantri, H. Sofyan, Mangaras Yanu Florestiyanto, "Aplikasi Pengolahan Citra Digital Untuk Identifikasi Umur Pohon," J. T. Informatika, vol. 16, no. 2, pp. 97–104, 2019.
- [2] A. R. Pambudi et al., "Deteksi keaslian uang kertas berdasarkan watermark dengan pengolahan citra digital," Jurnal Informatika Polinema, pp. 69–74, 2020.
- [3] A. R. Pratama, "Klasifikasi Daging Sapi Berdasarkan Ciri Warna Dengan Metode Otsu dan Euclidean Distance," Jurnal Ilmu Komputer dan Teknologi Informasi, vol. 5, no. 1, pp. 26–32, 2020.
- [4] N. F. Arini, A. Ubaidillah, K. A. Wibisono, and M. Ulum, "Identifikasi Embrio dalam Telur Berbasis Image Processing," ELKOM, vol. 1814, pp. 11–19, 2020, doi: 10.32528/elkom.v2i1.3137.
- [5] A. Fadjeri, B. A. Saputra, D. Kusuma, A. Ariyanto, and L. Kurniatin, "Karakteristik Morfologi Tanaman Selada Menggunakan Pengolahan Citra Digital," Jurnal Ilmiah Sinus, Vol. 20, No. 2, pp. 1–12, 2022.
- [6] A. Fadjeri, A. Setyanto, and M. P. Kurniawan, "Pengolahan Citra Digital Untuk Menghitung Ekstraksi Ciri Greenbean Kopi Robusta Dan Arabika (Studi Kasus: Kopi Temanggung)," J. Teknol. Inf. dan Komun., vol. 8, no. 1, pp. 8–13, 2020, doi: 10.30646/tikomsin.v8i1.462.
- [7] Ashari, N. Latif, and A. Astuti, "Pengolahan Citra Digital Untuk Menentukan Bobot Sapi Menggunakan Metode Canny Edge Detection," J. Ilm. Ilmu Komput., vol. 5, no. 1, pp. 1–6, 2019, doi: 10.35329/jiik.v5i1.24.
- [8] Y. N. Nabuasa, "Pengolahan Citra Digital Perbandingan Metode Histogram Equalization Dan Spesification Pada Citra Abu-Abu," J-ICON, vol. 7, no. 1, pp. 87–95, 2019.
- [9] M. M. Effendi, A. Y. Permana, and I. Nawangsih, "Penerapan Metode Ekstraksi Image Ke Text Dengan Optical Character Recognition," no. November, pp. 496–502, 2019, doi: 10.30998/simponi.v0i0.484.
- [10] A. Ramadhani, A. Shalludin "Rancang Bangun Aplikasi Pengarsipan Dokumen Menggunakan Optical Character Recognition Pada Badan Kepegawaian Dan Pengembangan Sumber Daya Manusia Kabupaten Kapuas Berbasis Visual", Jurnal Sains Komputer dan Teknologi Informasi, Vol. 3, No. 1, pp. 76–82, 2020.
- [11] I K. Setia Buana, "Implementasi Aplikasi Speech to Text untuk Memudahkan Wartawan Mencatat Wawancara dengan Python," Jurnal Sistem Dan Informatika (JSI), 14(2), pp. 135–142, 2020, doi: 10.30864/jsi.v14i2.293.