

Penerapan Metode *Optical Character Recognition* (OCR) Untuk Mengambil Data Arsip

Mochammad Rifky Reyvansyah, Edy Setiawan, Rini Indarti, Adianto, Ii Munadhif

Program Studi Teknik Otomasi, Jurusan Teknik Kelistrikan Kapal

Politeknik Perkapalan Negeri Surabaya

Jalan Teknik Kimia, Kampus ITS Sukolilo, Surabaya 60111

E-mail: mrifky@student.ppnns.ac.id, edy_setiawan@ppns.ac.id, riniindarti@ppns.ac.id, adianto@ppns.ac.id, iimunadhif@ppns.ac.id

Abstract— Dalam proses penyimpanan arsip, pendataan adalah tahap sebelum arsip disimpan agar lebih mudah ketika mencarinya. Permintaan dan pengelolaan arsip menuntut ketelitian dan keakuratan, oleh sebab itu banyak perusahaan atau instansi yang penyimpanan arsipnya sulit untuk ditelusuri. Umumnya saat proses pendataan, arsip akan diklasifikasikan berdasarkan jenisnya. Hanya 3 jenis arsip yang digunakan pada penelitian ini yaitu; arsip piutang, hutang dan perizinan. Sedangkan untuk mengklasifikasikannya digunakan OCR (*Optical Character Recognition*). OCR adalah proses dalam mengubah gambar huruf menjadi karakter ASCII (*American Standard Code for Information Interchange*) yang dapat dikenali oleh komputer. Metode tersebut dipilih karena dapat mempermudah serta mempersingkat proses pendataan arsip, yang awalnya menetik menjadi *scan* atau memindai arsip saja. Pendata arsip melalui tiga proses yaitu mengambil citra arsip dengan kamera, mengklasifikasikan jenisnya menggunakan OCR dan memasukan data arsip ke dalam *excel*. Dari hasil penelitian, didapatkan total rata-rata nilai keberhasilan pembacaan OCR sebesar 99.47% dari 3 kriteria sampel pengujian yaitu. Sampel citra sempurna sebesar 100%, citra minim fokus sebesar 98.66% dan citra minim cahaya sebesar 99.75%. Sedangkan tingkat keberhasilan dalam mencari arsip sebesar 95.33%. Sehingga dapat disimpulkan bahwa metode ini memiliki nilai yang baik dalam pendataan arsip.

Kata Kunci— *Optical Character Recognition, Pendataan Arsip, Penelusur Arsip*

I. PENDAHULUAN

Di sebuah perusahaan pasti ada bagian kearsipan mulai dari perusahaan yang bergerak disektor makanan hingga perusahaan yang bergerak di sektor pengiklanan [1]. Sebuah perusahaan atau instansi akan terlihat buruk reputasinya akibat kearsipan yang tidak memadai [2]. Kebutuhan dalam penyediaan arsip secara cepat dan tepat sangat diperlukan karena segalanya memerlukan arsip dan arsip adalah segalanya bagi perusahaan.

Pendataan arsip adalah proses mengklasifikasikan arsip berdasarkan jenis dan isi arsip sebelum disimpan [3]. Selain itu, juga direncanakan penyimpanan dan peletakan arsip [4]. Proses ini umumnya memakan waktu yang lama dan memerlukan ketelitian tinggi karena perlu menetik satu-persatu keterangan didaftar arsip berupa *excel* ataupun *software* lain yang fungsinya sama [5]. Oleh karena itu, metode OCR atau proses mengubah gambar huruf menjadi karakter ASCII (*American Standard Code for Information Interchange*) [6],[7] dapat mengurangi waktu pengetikan. Setidaknya ada 5 macam jenis dalam menyimpan arsip yaitu sistem abjad (*alphabetical*), nomor (*numerical*), tanggal

(*chronological*), wilayah (*geological*), subjek (*subjectical*) [8] setiap macamnya memiliki kelebihan serta kekurangannya masing-masing berkaitan dengan proses pencarian kembali atau pengambilan arsip yang telah disimpan. Pada penelitian kali ini dipilih sistem tanggal untuk kelebihan dan kekurangannya dapat dilihat pada tabel berikut.

Tabel 1. Macam Penyimpanan Arsip

Penyimpanan Sistem Tanggal atau <i>Chronological</i>	
Kelebihan	Kekurangan
<ul style="list-style-type: none">• Cocok untuk pengolah yang kegiatannya berkaitan dengan tanggal jatuh tempo.• Sederhana dan mudah diterapkan karena tanpa klasifikasi.	<ul style="list-style-type: none">• Akan terjadi kesulitan dalam penemuan kembali arsip apabila peminjam menyebutkan perihal arsip.• Orang sering lupa dengan tanggal surat terutama tanggal penyimpanan.• Tidak semua unit pengolahan dalam organisasi itu cocok menetapkan sistem ini.• Agar mudah mengatur letak arsip dalam folder maka pembuatan kode tidak dapat murni 100% tetapi harus ditambahkan dengan kode abjad.

Sumber : [9]

Sesuai tabel 1. penyimpanan tersebut akan disempurnakan lagi dengan menggunakan OCR sebagai metode untuk menggantikan pengetikan ke dalam *excel* [10]. Selain itu, juga memberikan keterangan nama dan perihal dari arsip sehingga pencarian menjadi lebih mudah serta efisien [11].

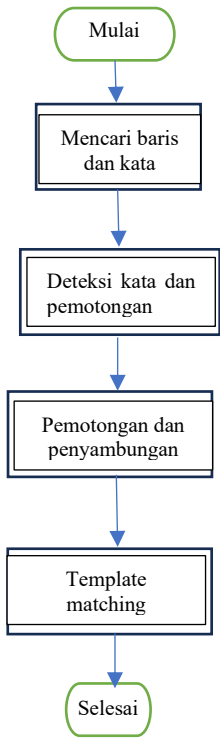
II. BAHAN DAN METODE

Teknik mendapatkan karakter huruf ASCII dari sebuah gambar atau citra dimulai dari *image processing* OCR. dalam prosesnya bisa dilihat pada *flowchart* berikut.

Penjelasan dari tahap *flowchart* gambar 1, sebagai berikut.

A. Pencarian Baris dan Kata

Dari hasil citra yang diambil akan dicari baris dan kata didalamnya. Pencarian baris dilakukan dengan penyaringan gumpalan dan konstruksi garis[12]. Dengan asumsi bahwa analisis tata letak halaman sudah menyediakan wilayah teks dengan ukuran teks yang seragam.



Gambar 1. Proses OCR



Gambar 2. Penandaan Baris Dan Kata

B. Deteksi Kata dan Pemotongan

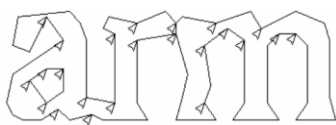
Untuk mencari penataan baku dari kata akan dilakukan pemotong kata menjadi karakter menggunakan *pitch*, dan menonaktifkan *chopper* dan *associator* pada saat ini kata untuk langkah pengenalan kata [13]. Gambar 3. menunjukkan sebuah contoh khas dari kata *fixed-pitch*. Maka jika ada huruf dengan bentuk atau ukuran yang berbeda akan terfilter.



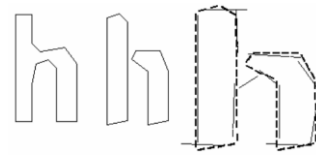
Gambar 3. Pencarian Penataan Baku Dari Sebuah Kata

C. Pemotongan Dan Penyambungan Karakter

Memotong gumpalan citra karakter dengan kemungkinan terbaik dari pengklasifikasi karakter. Kandidat memotong poin ditemukan dari cekungan pendekatan *polygonal* dari garis besar, dan mungkin memiliki cekungan lain seperti simpul berlawanan, atau segmen garis [14]. Hal ini bisa memerlukan hingga 3 pasang *chop point* untuk berhasil memisahkan gabungan karakter.



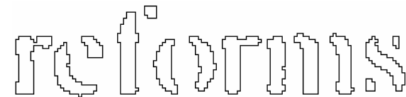
Gambar 4. Pemotongan Dari Gumpalan Citra Karakter



Gambar 5. Penyambungan Karakter Sebelum Diklasifikasikan

D. Template Matching

Template Matching adalah tahap untuk mencocokkan karakter sesuai dengan referensi citra karakter yang telah dipotong sebelumnya menggunakan metode NCC (*Normalize Cross Correlation*) [15]. Dengan metode NCC, persentase kecocokan dengan nilai yang paling tinggi akan menentukan jenis atau model karakter tersebut [16].



Gambar 6. Bentuk Karakter Yang Akan Dicocokkan

Selain *template matching*, metode ini dapat digunakan untuk menentukan kecocokan antara dua sinyal atau lebih dalam suatu interval waktu [17]. Untuk menentukan nilai NCC dapat digunakan persamaan (1) berikut ini.

$$NC = \frac{\sum_i \sum_j W_{ij} W'_{ij}}{\sum_i \sum_j [W'_{ij}]^2} \quad (1)$$

NC : *Normalized Cross Correlation*

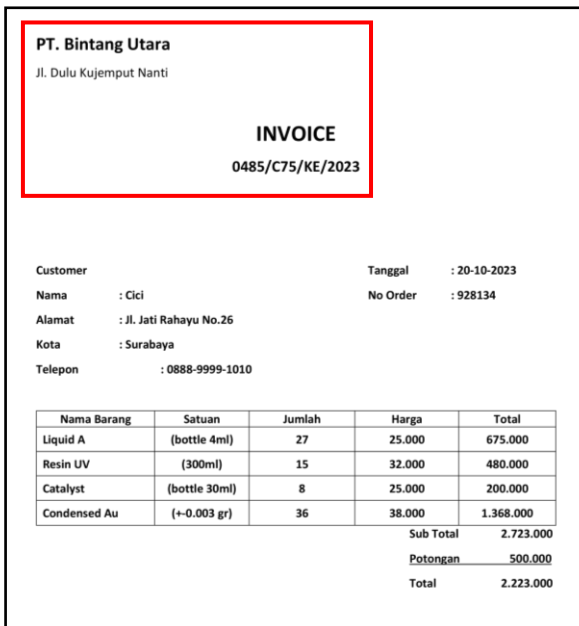
W_{ij} : *Piksel citra referensi*

W'_{ij} : *Piksel citra capture*

i, j : *Elemen citra*

E. Arsip Sampel Uji OCR

Terdapat 3 jenis asip yang digunakan yaitu piutang, hutang, perizinan. Bagian yang nantinya akan diproses OCR dalam kop surat oleh karena itu bagian tersebut harus memiliki *layout* tata letak yang sama agar lebih mudah saat *scanning* [18]. Berikut adalah salah satu sampel kop surat yang digunakan sebagai sampel.



Gambar 7. Bagian yang di-scan Kamera

Bagian yang ditandai kotak merah adalah bagian yang diambil oleh kamera.

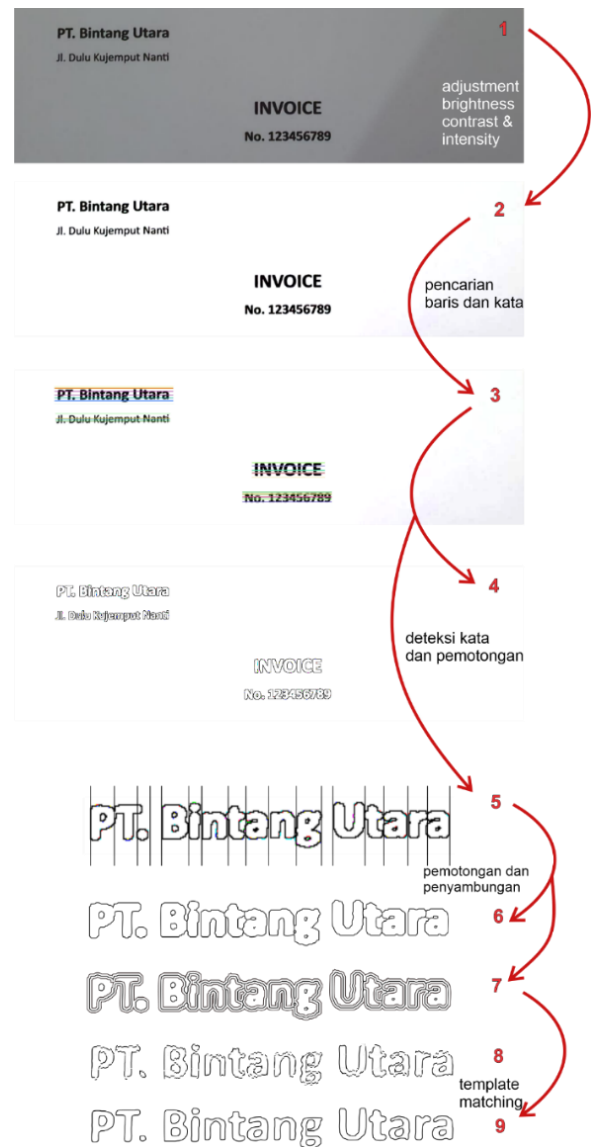
F. Data Arsip

Data arsip merupakan keterangan arsip yang tercantum pada kop surat [19], sesuai dengan yang ditunjukkan pada gambar 7. Data ini berupa nama, alamat perusahaan, jenis dan nomor arsip. Oleh karena itu, pemilihan *layout* kop surat harus rapi dan presisi agar tidak mempersulit proses pembacaan OCR. Dengan mempertimbangkan hal tersebut maka dipilih layout sebagai berikut:

Tabel 2. Layout Kop-surat

Perusahaan	
Alamat	
	Jenis Arsip
	Nomor Arsip

Penataan penulisan atau *layout* tersebut sangat sering dijumpai karena sudah umum digunakan perusahaan dengan berbagai bidang [20]. Selanjutnya merupakan cara mengekstrak karakter ASCII setiap bagian dari gambar melalui serangkaian proses untuk penjabarannya dapat dilihat pada gambar 8, yaitu proses mendapatkan data arsip.



Gambar 8. Proses Mendapatkan Data Arsip

Gambar 8 merupakan tahapan proses OCR untuk mendapatkan data arsip. Proses 1-2 merupakan *adjustment* dari *brightness*, *contrast* dan *intensity*, tahap 3 yaitu pencarian baris dan kata, tahap 4-5 deteksi kata dan pemotongan, tahap 6-7 pemotongan dan penyambungan karakter terakhir tahap 8-9 pencocokan menggunakan *template matching*.

G. Penyaringan Data Arsip

Penyaringan ini adalah proses substitusi, eliminasi serta pemisahan karakter ASCII masing masing data arsip dari hasil OCR yang hanya sebuah paragraf panjang tanpa ada keterangan data seperti jenis, nama perusahaan dan nomor dari arsip itu sendiri. Agar lebih jelasnya dapat dilihat tabel berikut ini.

Tabel 3. Penyaringan Data Arsip

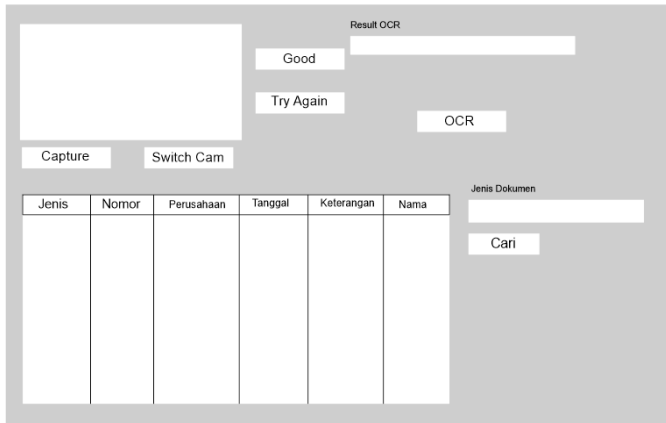
Bentuk Awal	Perubahan	hasil
Perusahaan Alamat Jenis Arsip	Subtitusi (Enter) Menjadi (,)	Perusahaan,,Alamat,,Je nis Arsip,Nomor Arsip,

Nomor Arsip		
Perusahaan,, Alamat,,Jeni s Arsip,Nomo r Arsip,	Substitusi (,) menjadi (,)	Perusahaan,Alamat,Jenis Arsip,Nomor Arsip,
Perusahaan, Alamat,Jenis Arsip,Nomo r Arsip,	Eliminasi (,) terakhir dan pemisahan	(Perusahaan),(Alamat),(Jenis Arsip),(Nomor Arsip)

Dari tabel 3. proses penyaringan data arsip dilakukan dari paragraf hingga menjadi potongan keterangan yang akan dicantumkan sebagai identitas arsip pada *list* pendataan arsip yang berupa *excel*.

H. GUI (Graphical User Interface)

Setelah melalui penyaringan, data arsip berupa nama alamat jenis dan nomor arsip akan dicatat kedalam tabel *list* file *excel* sebagai pendataan arsip. Namun agar lebih mudah mengoprasikan sistem maka digunakan GUI *Tkinter* yang merupakan *framework interface* di-*Python*[21].



Gambar 9. GUI Sistem

Tampilan kamera pada pojok kiri atas lalu berikutnya ada *list* arsip untuk melihat apakah sudah dicatat pada *excel*, kolom pencarian terletak dikanan tengah, hasil OCR ada di kanan atas dan sisanya sebagai berikut.

- Capture* = mengambil gambar dari kamera
- Switch Cam* = mengganti kamera
- Good* = menyimpan kamera
- Try again* = mengambil ulang foto
- OCR* = memperoleh data arsip memasukan data kedalam excel
- Cari* = mencari jenis arsip dalam database

III. HASIL DAN PEMBAHASAN

Terdapat dua pengujian yang dilakukan yaitu pengujian pembacaan OCR dan pengujian pencarian arsip dalam *list* pendataan arsip.

A. Hasil Pengujian OCR

Pengujian ini memiliki 3 macam kriteria uji yaitu pengambilan sampel gambar sempurna, sampel gambar

minim fokus dan minim cahaya. Pertama hasil sampel dengan gambar sempurna dapat dilihat pada tabel 4. Hasil OCR akan berurutan sesuai karakter pada sampel yang sedang diuji.

Tabel 4. Hasil Sampel Gambar Sempurna

NO	Sebenarnya	Hasil OCR	Error
1	PT. Bintang Utara Jl. Dulu Kujemput Nanti INVOICE 0485/C75/KE/2023	Result PT. Bintang Utara Jl. Dulu Kujemput Nanti INVOICE 0485/C75/KE/2023	0%
2	PT. Faster Send Jl. Kelampis Raya INVOICE 0375/B.A6/TP/2023	Result PT. Faster Send Jl. Kelampis Raya INVOICE 0375/B.A6/TP/2023	0%
3	PT. Bintang Utara Jl. Dulu Kujemput Nanti PERIZINAN 0143/M.A2/01/2023	Result PT. Bintang Utara Jl. Dulu Kujemput Nanti PERIZINAN 0143/M.A2/01/2023	0%
4	PT. Bintang Utara Jl. Dulu Kujemput Nanti PENAWARAN 0112/P3/04/2023	Result PT. Bintang Utara Jl. Dulu Kujemput Nanti PENAWARAN 0112/P3/04/2023	0%
5	PT. Bintang Utara Jl. Dulu Kujemput Nanti PAJAK LAHAN INTENS NO.1	Result PT. Bintang Utara Jl. Dulu Kujemput Nanti PAJAK LAHAN INTENS NO.1	0%

Uraian hasil uji sampel sempurna ada pada tabel 5.

Tabel 5. Hasil Sampel Gambar Sempurna

No	Sebenarnya	Hasil OCR	Error
1	PT. Bintang Utara	PT. Bintang Utara	0%
	Jl. Dulu Kujemput Nanti	Jl. Dulu Kujemput Nanti	0%
	INVOICE	INVOICE	0%
	0485/C75/KE/2023	0485/C75/KE/2023	0%
	Rata-rata error		0%
2	PT. Faster Send	PT. Faster Send	0%
	Jl. Kelampis Raya	Jl. Kelampis Raya	0%
	INVOICE	INVOICE	0%
	0375/B.A6/TP/2023	0375/B.A6/TP/2023	0%
	Rata-rata error		0%
3	PT. Bintang Utara	PT. Bintang Utara	0%
	Jl. Dulu Kujemput Nanti	Jl. Dulu Kujemput Nanti	0%
	PERIZINAN	PERIZINAN	0%
	0143/M.A2/01/2023	0143/M.A2/01/2023	0%
	Rata-rata error		0%
4	PT. Bintang Utara	PT. Bintang Utara	0%
	Jl. Dulu Kujemput Nanti	Jl. Dulu Kujemput Nanti	0%
	PENAWARAN	PENAWARAN	0%
	0112/P3/04/2023	0112/P3/04/2023	0%
	Rata-rata error		0%
5	PT. Bintang Utara	PT. Bintang Utara	0%
	Jl. Dulu Kujemput Nanti	Jl. Dulu Kujemput Nanti	0%
	PAJAK	PAJAK	0%
	LAHAN INTENS NO.1	LAHAN INTENS NO.1	0%
	Rata-rata error		0%

Dapat disimpulkan dari tabel 4 dan 5 bahwa metode OCR telah berhasil diterapkan dan memiliki 100% tingkat keberhasilan namun dengan catatan bahwa sampel diambil dengan memaksimalkan kinerja kamera mulai dari penyesuaian tingkat fokus, pencahayaan sampai dengan posisi kertas yang sesuai agar mendapatkan hasil yang terbaik.

Tabel 6. Hasil Sampel Minim Fokus

NO	Sebenarnya	Hasil OCR	Error
1		Result PT. Bintang Utara - Jl. Dulu Kujemput Nanti INVOICE 0485/C75/KE/2023 (fokus kamera rendah)	1.25%
2		Result PT. Faster Send - Jl. Kelampis Raya INVOICE 0375/B.A6/TP/2023 (fokus kamera rendah)	1.65%
3		Result PT. Bintang Utara - Jl. Dulu Kujemput Nanti PERIZINAN 0143M.A201/2023 (terlalu cepat mengambil foto)	1.25%
4		Result PT. Bintang Utara - Jl. Dulu Kujemput Nanti PENAWARAN 0112/P3/04/2023 (terlalu cepat mengambil foto)	1.25%
5		Result PT. Bintang Utara - Jl. Dulu Kujemput Nanti PAJAK LAHAN INTENS NO.1 (fokus kamera rendah)	1.25%

Setelah melakukan pengujian sampel gambar minim fokus, ditemukan kesalahan pembacaan oleh karenanya dengan menggunakan rumus menghitung *error* sebagai berikut;

$$Error \% = \frac{Total\ Incorrect}{Total\ Character} \times 100\% \quad (1)$$

Maka perhitungan tingkat kegagalan dapat diuraikan pada tabel 7.

Tabel 7. Uraian Hasil Pengujian Minim Fokus

No	Sebenarnya	Hasil OCR	Error
1	PT. Bintang Utara	PT. Bintang Utara	0%
	Jl. Dulu Kujemput Nanti	Jl. Dulu Kujemput Nanti	5%
	INVOICE	INVOICE	0%
	0485/C75/KE/2023	0485/C75/KE/2023	0%
	Rata-rata error		1.25%
2	PT. Faster Send	PT. Faster Send	0%
	Jl. Kelampis Raya	Jl. Kelampis Raya	6.66%
	INVOICE	INVOICE	0%
	0375/B.A6/TP/2023	0375/B.A6/TP/2023	0%
	Rata-rata error		1.65%
3	PT. Bintang Utara	PT. Bintang Utara	0%
	Jl. Dulu Kujemput Nanti	Jl. Dulu Kujemput Nanti	5%
	PERIZINAN	PERIZINAN	0%
	0143/M.A2/01/2023	0143/M.A2/01/2023	0%
	Rata-rata error		1.25%
4	PT. Bintang Utara	PT. Bintang Utara	0%
	Jl. Dulu Kujemput Nanti	Jl. Dulu Kujemput Nanti	5%
	PENAWARAN	PENAWARAN	0%
	0112/P3/04/2023	0112/P3/04/2023	0%
	Rata-rata error		1.25%

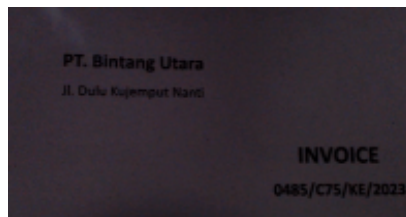
No	Sebenarnya	Hasil OCR	Error
5	PT. Bintang Utara	PT. Bintang Utara	0%
	Jl. Dulu Kujemput Nanti	Jl. Dulu Kujemput Nanti	5%
	PAJAK	PAJAK	0%
	LAHAN INTENS NO.1	LAHAN INTENS NO.1	0%
	Rata-rata error		1.25%

Seperti yang dibuktikan pada tabel 7, yang merupakan penjelasan dari tabel 6, menghasilkan tingkat keberhasilan mencapai 98.66%. Hasil tersebut diperoleh dari total rata-rata pengujian. Selain *settingan* fokus dari lensa kamera yang diturunkan, faktor lain yang mempengaruhi adalah terlalu cepat mengambil foto saat sampel baru saja diletakan sehingga hasil foto dalam keadaan kurang fokus.

Berikut ini merupakan sampel untuk pengujian minim Cahaya, yang diambil dalam rentang 120 lux sampai dengan 1lux.



Gambar 10. Citra Arsip Pada 120 lux



Gambar 11. Citra Arsip Pada 1 lux

Tabel 8. Hasil Sampel Minim Cahaya

No	cahaya (lux)	Sebenarnya	Hasil OCR	Error
1	120	PT. Bintang Utara	PT. Bintang Utara	0%
		Jl. Dulu Kujemput Nanti	Jl. Dulu Kujemput Nanti	0%
		INVOICE	INVOICE	0%
		0485/C75/KE/2023	0485/C75/KE/2023	0%
		Rata-Rata		0%
2	32	PT. Bintang Utara	PT. Bintang Utara	0%
		Jl. Dulu Kujemput Nanti	Jl. Dulu Kujemput Nanti	0%
		INVOICE	INVOICE	0%
		0485/C75/KE/2023	0485/C75/KE/2023	0%
		Rata-Rata		0%
3	24	PT. Bintang Utara	PT. Bintang Utara	0%
		Jl. Dulu Kujemput Nanti	Jl. Dulu Kujemput Nanti	0%
		INVOICE	INVOICE	0%
		0485/C75/KE/2023	0485/C75/KE/2023	0%
		Rata-Rata		0%
4	3	PT. Bintang Utara	PT. Bintang Utara	0%
		Jl. Dulu Kujemput Nanti	Jl. Dulu Kujemput Nanti	0%
		INVOICE	INVOICE	0%

		0485/C75/KE/2023	0485/C75/KE/2023	0%
		Rata-Rata		0%
5	1	PT. Bintang Utara	PT. Bintang Utara	0%
		Jl. Dulu Kujemput Nanti	Jl. Dulu Kujemput Nanti	5%
		INVOICE	INVOICE	0%
		0485/C75/KE/2023	0485/C75/KE/2023	0%
		Rata-Rata		1.25%

Dari tabel 8 dapat dipahami ketika lux sebesar 120 lux, pembacaan OCR menjadi maksimal hingga 3 lux, yaitu keadaan pengambilan citra redup memiliki tingkat keberhasilan pembacaan 100%. Ketika flux cahaya 1 lux pendeteksian OCR berkurang. Sayangnya lux meter yang digunakan tidak dapat mengukur nilai flux cahaya kurang dari 1 lux (<1 lux).

Setelah melakukan tiga pengujian dapat diambil rata rata tingkat keberhasilan dari hasil sampel sempurna, minim fokus dan minim pencahayaan, untuk lebih jelasnya ada pada tabel 9 berikut ini :

Tabel 9. Tingkat Keberhasilan Pengujian OCR

Sesi Uji	Tipe Uji OCR	Success %
Pengujian 1	Citra Sempurna	100
Pengujian 2	Minim Fokus	98.66
Pengujian 3	Minim Pencahayaan	99.75
Total Rata- Rata Keberhasilan		99.47

Pada tabel 9 dibuktikan bahwa pengujian OCR memiliki tingkat keberhasilan sebesar 99.47%.

B. Hasil List dalam Bentuk Excel

List ini langsung ditulis secara otomatis sesuai dengan aturan penyimpanan arsip bersistem tanggal atau *chronological*, Dicatat sesudah melakukan scanning dan memasukan nama orang yang bersangkutan serta perihal dari arsip pada GUI. Namun bila arsip berjenis penawaran atau pajak yang teridentifikasi OCR maka kedua jenis arsip tersebut tidak akan dimasukan kedalam *list* karena *list* ini tidak kerkatian dengan arsip tersebut. Dalam tabel excel telah memiliki *header* perkolomnya untuk membedakan data arsip untuk lebih jelasnya bisa dilihat pada gambar 3.3

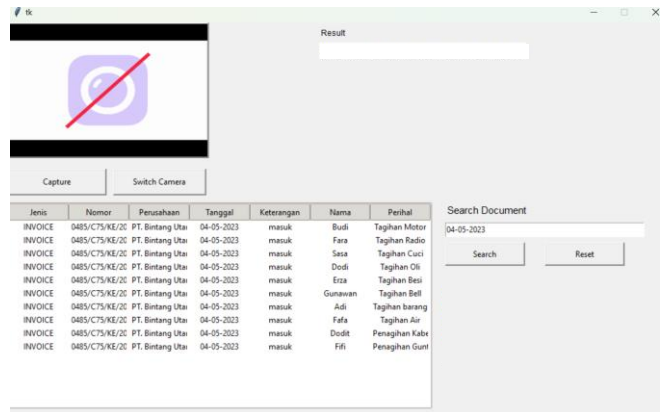
Jenis	Nomor	Perusahaan	Tanggal	Keterangan	Nama	Perihal
INVOICE	0485/C75/KE/2023	PT. Bintang Utara	03-05-2023	masuk	Dani	Tagihan PCB
INVOICE	0375/B.A6/TP/2023	PT. FASTER SEND	03-05-2023	masuk	Rifki	Pembayaran Kirim
PERIZINAN	0143/M.A2/01/2023	PT. Bintang Utara	03-05-2023	masuk	Annas	Izin Tempat
INVOICE	0375/B.A6/TP/2023	PT. FASTER SEND	03-05-2023	masuk	Susi	Tagihan Pengiriman
PERIZINAN	0143/M.A2/01/2023	PT. Bintang Utara	03-05-2023	masuk	Rudi	Izin Alat
INVOICE	0485/C75/KE/2023	PT. Bintang Utara	03-05-2023	masuk	Funi	Tagihan jendela
INVOICE	0485/C75/KE/2023	PT. Bintang Utara	03-05-2023	masuk	Tara	Tagihan Mobil
INVOICE	0485/C75/KE/2023	PT. Bintang Utara	03-05-2023	masuk	Peni	Tagihan Jas
INVOICE	0485/C75/KE/2023	PT. Bintang Utara	03-05-2023	masuk	Fifi	Tagihan Papan
INVOICE	0485/C75/KE/2023	PT. Bintang Utara	03-05-2023	masuk	Rere	Tagihan Motor
INVOICE	0485/C75/KE/2023	PT. Bintang Utara	04-05-2023	masuk	Budi	Tagihan Motor
INVOICE	0485/C75/KE/2023	PT. Bintang Utara	04-05-2023	masuk	Fara	Tagihan Radio
INVOICE	0485/C75/KE/2023	PT. Bintang Utara	04-05-2023	masuk	Sasa	Tagihan Cuci
INVOICE	0485/C75/KE/2023	PT. Bintang Utara	04-05-2023	masuk	Dodi	Tagihan Oli
INVOICE	0485/C75/KE/2023	PT. Bintang Utara	04-05-2023	masuk	Erza	Tagihan Besi
INVOICE	0485/C75/KE/2023	PT. Bintang Utara	04-05-2023	masuk	Gunawan	Tagihan Bell

Gambar 12. List Data Arsip Pada Excel

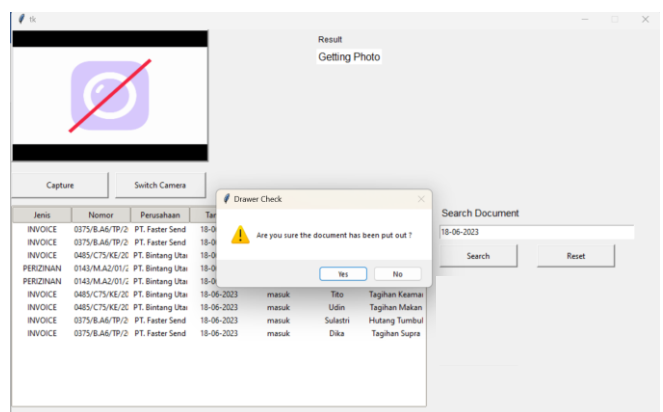
Terdapat 15 variasi tanggal dan terdapat 10 arsip disetiap tanggalnya sebagai data uji. Selanjutnya merupakan pengujian saat pencarian arsip berdasarkan sistem tanggal.

Tabel 10. Hasil Pencarian Arsip

No	Tanggal	Jumlah Arsip		(%)
		Sebenarnya	Ditemukan	
1	03-05-2023	10	10	100
2	04-05-2023	10	10	100
3	06-05-2023	10	10	100
4	07-05-2023	10	8	80
5	16-05-2023	10	10	100
6	23-05-2023	10	9	90
7	28-05-2023	10	10	100
8	01-06-2023	10	9	90
9	05-06-2023	10	8	80
10	09-06-2023	10	10	100
11	11-06-2023	10	10	100
12	13-06-2023	10	10	100
13	18-06-2023	10	9	90
14	19-06-2023	10	10	100
15	20-06-2023	10	10	100
Total rata-rata				95.33



Gambar 13. Tampilan GUI Saat Mencari Arsip



Gambar 14. Warning MessageBox

Hasil tabel 10 membuktikan bahwa pencarian arsip memiliki tingkat keberhasilan mencapai 95.33%. Terdapat arsip yang tidak dapat ditemukan karena data arsip yang tercatat tidak lengkap atau dapat juga dikatakan bahwa data tersebut hilang. Untuk itu GUI akan memberikan *warning message box* seperti pada gambar 14, dengan sistem memberitahu *user* apabila terdapat data yang hilang.

IV. KESIMPULAN

Setelah pengujian OCR didapatkan nilai tingkat keberhasilan dengan sampel sempurna 100%, minim fokus sebesar 98.66%, minim pencahayaan 99.75%, Total rata-rata keberhasilan mencapai 99.47% dengan acuan bahwa orientasi penataan sesuai dengan layout dan jarak kamera terhadap kertas sampel tidak berubah di setiap pengambilan gambar juga peletakan kertas sampel tidak miring terhadap kamera sehingga pembacaan OCR menjadi lebih akurat. Keakuratan tersebut juga mempengaruhi penyaringan data dan klasifikasi jenis arsip. Tingkat fokus lebih mempengaruhi tingkat keberhasilan dibandingkan pencahayaan maka hasil gambar yang tajam akan membuat pembacaan OCR menjadi lebih akurat.

Penelitian ini dapat menjadi rujukan atau referensi penelitian selanjutnya namun terdapat beberapa hal yang harus dipertimbangkan kembali seperti saat pengujian minim pencahayaan. Dari penelitian ditemukan bahwa alat ukur lux meter tidak dapat mengukur nilai lux kurang dari 1 lux hal ini membuat pengujian menjadi kurang lengkap. Nilai *error* baru ditemukan pada kondisi 1 lux. Penambahan sampel juga menjadi acuan dalam penelitian selanjutnya karena pada penelitian ini hanya menggunakan 3 sampel dengan variasi isi arsip. Sudut peletakan kertas terhadap kamera dapat menjadi tipe uji selanjutnya, karena kemiringan tulisan pada gambar kemungkinan besar mempengaruhi pembacaan OCR.

DAFTAR PUSTAKA

- [1] R. Laili *dkk.*, “Analisis Prosedur Penyimpanan Arsip Dalam Menunjang Efektivitas Kerja Pegawai Di Kantor Badan Pengawas Pemilihan Umum (Bawaslu) Kota Palembang,” *Jurnal Ilmiah dan Administrasi Sosial JIADS*, vol. 18, no. 1, hlm. 1–10, 2023.
- [2] L. Adhrianti, “Management Public Relations Melalui Proses Koorientasi Bagi Kepentingan Pengelolaan Arsip Bisnis Jasa Perhotelan (Studi Kasus Upaya Promosi dan Peningkatan Citra Positif Marina Hotel Bengkulu),” *JURNAL KAGANGA*, vol. 6, no. 1, 2022.
- [3] M. D. Trisetiyo dan J. S. Wibowo, “Klasifikasi Surat Menggunakan Metode Naïve Bayes Pada Sistem Informasi Manajemen Surat (Studi Kasus: Bagian Humas Setda Kabupaten Batang),” *Proceeding SENDI U*, hlm. 285–292, Jul 2019.
- [4] A. Anisah, D. Wahyuningsih, E. Helmud, T. Suwanda, P. Romadiana, dan D. Irawan, “Rancang Bangun Sistem Informasi Manajemen Arsip Digital,” *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 10, no. 3, hlm. 419–425, Des 2021, doi: 10.32736/sisfokom.v10i3.1300.
- [5] M. Hamidah dan K. Rukun, “Implementasi Framework YII2 Pada Sistem Electronic Arsip (E-Arsip) Kependudukan Nagari Balah Hilia,” *VOTEKNIKA*, vol. 07, no. 03, 2019, [Daring]. Tersedia pada: <http://ejournal.unp.ac.id/index.php/voteknika/index>
- [6] F. Mohammad, J. Anarase, M. Shingote, dan P. Ghanwat, “Optical character recognition implementation using pattern matching,” *International Journal of Computer Science and Information Technologies*, vol. 5, no. 2, hlm. 2088–2090, 2014.
- [7] A. L. Firdaus, M. S. Kurnia, T. Shafera, dan W. I. Firdaus, “Implementasi Optical Character Recognition (OCR) Pada Masa Pandemi Covid-19,” *JUPITER (Jurnal Penelitian Ilmu dan Teknik Komputer)*, vol. 13, no. 2, hlm. 188–194, 2021.
- [8] M. R. Basya dan D. Puspari, “Analisis Pengelolaan Arsip di Kantor Kecamatan Bungah Kabupaten Gresik,” *Jurnal Pendidikan Administrasi Perkantoran (JPAP)*, vol. 9, no. 2, hlm. 439–453, 2021.
- [9] Kurniawan, “Pengertian Arsip dan 5 Sistem Penyimpanan Arsip Beserta Contohnya,” Sep 2017, Diakses: 17 Juni 2023. [Daring]. Tersedia pada: https://www.academia.edu/38451857/Contoh_Pengertian_Arsip_dan_5_Sistem_Penyimpanan_Arsip_Beserta_Contohnya
- [10] F. Dwima, D. Rahmawati, dan K. Aji, “Rancang Bangun Sistem Keamanan Parkir Berbasis Random Password dan Image Processing Menggunakan Optical Character Recognition (OCR),” *Jurnal Teknik Elektro dan Komputer (TRIAC)*, vol. 04, no. 07, 2017, Diakses: 22 Juni 2023. [Daring]. Tersedia pada: <https://journal.trunojoyo.ac.id/triac/article/view/3259/2431>
- [11] D. Haumahu, A. Soetiksno, dan W. O. Kowey, “Pengaruh Pengelolaan Kearsipan Terhadap Efisiensi Kerja Pegawai Kantor Kecamatan Nusaniwe,” *Jurnal Administrasi Terapan*, vol. 2, no. 1, hlm. 104–115, Mar 2023.
- [12] W. Bian, T. Wakahara, T. Wu, H. Tang, dan J. Lin, “Binarization of Color Character Strings in Scene Images Using Deep Neural Network,” *IEEE 2018 Digital Image Computing*, 2018, doi: doi:10.1109/dicta.2018.8615837.
- [13] R. Smith, “An overview of the Tesseract OCR engine,” dalam *Ninth international conference on document analysis and recognition (ICDAR 2007)*, IEEE, 2007, hlm. 629–633.
- [14] D. Ega Saputra, D. Rahmawati, dan A. Fiqhi Ibadillah, “Pengolahan Citra Digital Dalam Penentuan Panen Jamur Tiram,” *Jurnal Teknik Elektro dan Komputer TRIAC*, vol. 6, no. 1, hlm. 2615–7764, 2019, doi: <https://doi.org/10.21107/triac.v6i1.4356>.
- [15] S. Muharom, “Pengenalan Nomor Ruang Menggunakan Kamera Berbasis OCR Dan Template Matching,” *Jurnal Inform*, vol. 4, no. 1, 2019.
- [16] R. Alfita, A. F. Ibadillah, dan A. Prianto, “Identifikasi Nilai Nominal Uang Kertas Berdasarkan Warna Berbasis Image Processing Menggunakan Metode Template Matching,” *Jurnal Teknik Elektro dan Komputer TRIAC*, vol. 9, no. 1, hlm. 2615–7764, Mei 2022, doi: <https://doi.org/10.21107/triac.v9i1.12487>.
- [17] H. Setiawan, “Implementasi Metode Normalized Cross Correlation Pada Template Matching Untuk Aplikasi Pendeteksi Objek Pada Citra Digital.” November 2014. Diakses: 9 Maret 2023. [Daring]. Tersedia pada:

<https://elib.unikom.ac.id/gdl.php?mod=browse&op=read&id=jbptunikompp-gdl-herisetiaw-35242>

- [18] S. S. Nurhaliza dan L. ETP, “Sistem Pengenalan Karakter Dokumen Secara Otomatis Menggunakan Metode Optical Character Recognition,” *PETIR*, vol. 15, no. 1, hlm. 166–175, Feb 2022, doi: 10.33322/petir.v15i1.1610.
- [19] Y. Zawarnis, “Kesalahan Berbahasa Dalam Surat Dinas Instansi Pemerintahan Di Kabupaten Lampung Selatan,” *Kelasa*, vol. 18, no. 1, hlm. 81–96, 2020.
- [20] E. Alhadi, F. Maretha, H. Setiawan, J. Bustan, J. Administrasi Bisnis, dan P. Negeri Sriwijaya, “Perancangan Surat Bisnis Secara Elektronik Pada Kerupuk-Kemplang dan Pempek Hj. Eva Yunus,” *Seminar Nasional Aplikasi Teknologi untuk Masyarakat SNAPTEKMAS*, vol. 1, no. 2, 2019.
- [21] K. Yuvaraj, G. M. Oorappan, K. K. Megavarthini, M. C. Pravin, R. Adharsh, dan M. A. Kumaran, “Design and development of an application for database maintenance in inventory management system using tkinter and sqlite platform,” dalam *IOP Conference Series: Materials Science and Engineering*, IOP Publishing, 2020, hlm. 012012.