

BAB II

LANDASAN TEORI

2.1 Penelitian Terkait

Penelitian terkait yang telah dilakukan ini menjadi salah satu acuan untuk melakukan penelitian ini sehingga dapat memperkaya teori yang digunakan dalam mengkaji penelitian yang akan dilakukan. Penelitian terkait dapat dilihat sebagai berikut :

Penelitian yang dilakukan oleh Pratama,dkk (2016) dengan judul “Desain dan Analisis Penerjemah Aksara Sunda dengan Metode *Speeded Up Robust Features* dan *Radial Basis Function Neural Network*” penelitian ini menjelaskan bagaimana Desain dan Analisis dari suatu sistem penerjemah aksara sunda. aksara sunda yang diterjemahkan adalah aksara sunda media cetak dan tulisan tangan, segmentasi, ekstraksi dan klasifikasi. Metode yang digunakan penelitian ini adalah metode *Speeded Up Robust Features* dan *Radial Basis Function Neural Network* .

Selanjutnya, penelitian yang dilakukan oleh Rahmat, dkk (2017) dengan judul “Sistem Pengenalan Aksara Sunda menggunakan Metode *Modified Directionfeature* dan *Learning Vector Quabtization*” penelitian ini menjelaskan bagaimana teknik yang digunakan untuk mengenali pola citra aksara diperlukan teknik ekstraksi ciri dan pengelompokan pola. Dalam prosesnya teknik yang dapat digunakan salah satunya adalah *modified direction feature* (MDF) dimana metode ini merupakan metode ekstraksi fitur yang mentransformasi vektor dari citra biner menjadi sebuah vektor yang mewakili ciri dari arah dan transisi dari suatu pola. Selanjutnya teknik yang digunakan adalah *learning vector quantization* (LVQ) dimana metode ini digunakan sebagai metode pengelompokan atau pengklasifikasi dari suatu citra. Untuk menunjang fungsinya pada pengenalan aksara ini dapat dikombinasikan dengan teknik *text to speech*.

Kemudian, penelitian dari Widiyanto,dkk (2017) dengan judul “Analisis Sistem pengenalan dan keamanan kriptografi hill cipher pada plat nomer kendaraan menggunakan metode *template matching* “ menyimpulkan bahwa

kemajuan teknologi berkaitan erat dengan efisiensi manusia dalam melakukan pekerjaannya, salah satunya adalah pada sistem perparkiran. Semakin tinggi frekuensi kendaraan yang keluar masuk suatu tempat parkir akan memungkinkan komputer untuk menggantikan peran manusia dalam melakukan pencatatan plat nomor kendaraan.

Selanjutnya, penelitian dari Trianto,dkk (2019) dengan judul “Klarifikasi huruf katakana dengan template matching corelation” menyimpulkan bahwa Algoritma Template Matching adalah metode sederhana yang dapat digunakan untuk mengenali pola sebuah citra. Metode Template Matching ini juga sering digunakan untuk mengidentifikasi citra karakter huruf, angka, sidik jari dan aplikasi-aplikasi pencocokan citra lainnya. Pada Project Akhir ini dibuat sebuah aplikasi identifikasi huruf jepang khususnya katakana yang digunakan untuk tujuan memperkenalkan bahasa jepang kepada masyarakat luas.

Penelitian dari Muhamad,dkk (2018) yang berjudul “rancang bangun aplikasi pengenalan huruf aksara sunda berbasis mobile” menyimpulkan bahwa Dalam suatu proses belajar mengajar, dua unsur yang sangat penting adalah metode mengajar dan media pengajaran. Seiring dengan perkembangan zaman, teknologi informasi Android telah berkembang dengan sangat pesat dan telah melekat dengan kehidupan masyarakat. Saat ini android (smartphone) tidak hanya digunakan oleh kalangan menengah atas saja tetapi juga banyak digunakan oleh kalangan menengah bawah. Tidak hanya orang dewasa, anak-anak pun sudah sangat mahir dalam menggunakannya.

Selanjutnya, penelitian dari Setiawan,dkk (2017) dengan judul “implementasi optical character recognition (OCR) pada mesin penerjemah bahasa indonesia ke bahasa inggris” disimpulkan bahwa Tujuan dari penelitian ini adalah membuat *Optical Character Recognition (OCR)* pada mesin penerjemah bahasa Indonesia ke bahasa Inggris dengan menggunakan platform Android untuk penginputan kata penerjemahan. Sistem dibangun menggunakan *Optical Character Recognition (OCR)* dengan library mobile vision agar user tidak perlu lagi mengetikkan teks sebagai inputan kata yang akan di terjemahkan dengan material pengujian menggunakan gambarteks sebanyak lima gambar dan lembar teks sebanyak 30 kalimat.

Kemudian, penelitian yang dilakukan oleh Octaviano,dkk (2019) dengan judul “perancangan aplikasi kamus digitas berbahasa Indonesia-sunda-inggris berbasis android “ penelitian ini menyimpulkan bahwa banyaknya pengguna yang kesulitan dalam menerjemahkan bahasa salah satunya bahasa daerah maupun bahasa luar yaitu bahasa inggris yang penting dalam dunia pembelajaran, oleh karena itu pentingnya mengenali bahasa sunda maupun bahasa inggris yang berguna dalam mengartikan dan memahami sebuah kalimat ke dalam bahasa Indonesia kedalam bahasa sunda maupun bahasa Inggris. Seiring semakin banyaknya pengguna smartphone berbasis android maka dalam mempermudah suatu proses menerjemahkan suatu bahasa guna untuk membantu masyarakat dan pelajar khususnya di daerah jawa barat dengan cara merancang suatu aplikasi android yang memiliki fitur yang dapat memenuhi kebutuhan dan daya tarik tersendiri bagi user. Penelitian ini menghasilkan rancangan aplikasi kamus digital berbahasa indonesia ke dalam bahasa sunda dan kedalam bahasa inggris berbasis android.

Selanjutnya, penelitian yang dilakukan oleh Setiady (2019) dengan judul “efektifitas pemberkasan digital melalui *metode template matching*” Tujuan dari penelitian ini adalah untuk mengetahui efektivitas hasil pemindaian dokumen dengan memanfaatkan teknik pengenalan karakter optik yang dikenal dengan OCR (*Optical Character Recognition*).

Penelitian dari Munjiat (2015) yang berjudul “pengenalan huruf pada citra digital menggunakan algoritma *template matching*”, menjelaskan Perkembangan teknologi yang sudah berkembang pesat saat ini sudah merambahkedalam bidang sistem cerdas. Salah satu penerapan sistem cerdas ini yaitu pengenalan pola huruf pada citra digital yang dapat dikenali oleh komputer. Dengan adanya pengenalan huruf ini agar dapat membantu pengguna untuk mengenali huruf, terutama dibidang pendidikan yang berkaitan dengan pengenalan dasar huruf. Berbeda dengan manusia yang memiliki kecerdasan alami, sebuah komputer harus memiliki algoritma tertentu yang digunakan untuk membentuk kecerdasan yang akan dimilikinya untuk mengenali huruf didalam citra digital.

Selanjutnya, Penelitian dari Amalia,dkk (2020) dengan judul “ pengenalan

aksara sunda menggunakan metode jaringan saraf tiruan backpropagation dan deteksi tepi canny” Penelitian ini bertujuan dalam memudahkan pengenalan aksara sunda yang dapat di kembangkan kegiatan yang berkaitan dengan komputerisasi dan berkembangnya zaman, salah satunya yaitu dengan pengenalan pola citra aksara sunda. Jaringan Saraf Tiruan mampu mengenali pola yang akan digunakan selama pelatihan serta kemampuan jaringan untuk memberikan respon yang benar terhadap pola masukan yang serupa dengan pola yang di pakai selama penelitian.

Kemudian, penelitian yang dilakukan oleh Tamatjita,dkk (2008) yang berjudul “optical character recognition dengan algoritma orchie “Penelitian ini bertujuan untuk Pengembangan OCR diharapkan dapat memperkecil waktu dan tenaga manusia yang dibutuhkan untuk memindahkan teks cetak ke dalam bentuk digital. Meski demikian masih terdapat beberapa hal yang harus diuji pada algoritma-algoritma OCR yang dikembangkan, di antaranya dalam ketepatan pengenalan karakter.

Berdasarkan hasil dari beberapa peneliti terdahulu diatas, banyak yang mengembangkan metode OCR terhadap penerjemah aksara sunda dengan beberapa algoritma. Tetapi, pada penelitian ini penulis membuat matriks penelitian dan membuat *state of the art* (ringkasan penelitian sebelumnya) yang didalamnya terdapat perbedaan dan kesamaan dengan beberapa peneliti sebelumnya sehingga dapat memudahkan penulis dalam menyelesaikan tugas akhir ini, yaitu sebagai berikut :

Tabel. 2.1 Matriks penelitian

No	Peneliti	Metode						Huruf						Algoritma			
1.	Pratama, dkk (2016)		✓					✓									✓
2.	Mahendra (2019)						✓		✓								✓
2.	Rahmat, dkk (2017)				✓			✓									✓
3.	Widianto, dkk (2017)	✓										✓		✓			
4.	Trianto, dkk (2019)	✓								✓				✓			
5.	Octaviano , dkk (2019)							✓						✓			
6.	Setiady, (2019)	✓								✓				✓			
7.	Munjiat (2015)	✓									✓			✓			
8,	Muhamad, dkk (2018)	✓						✓									✓
9.	Setiawan, dkk (2019)	✓										✓					
10.	Amalia, dkk (2020)						✓	✓									
11.	Tamatjita dkk (2008)	✓															✓
12.	Penelitian yang diusulkan	✓												✓			

2.2 State Of The Art

Terdapat beberapa penelitian yang telah dilakukan sebelumnya dalam penelitian menerjemah aksara sunda menggunakan metode OCR dengan pengenalan pola Algoritma *Template Matching Corelation*. Dengan adanya hal tersebut maka perlu adanya studi literatur terhadap penelitian sebelumnya yang memiliki variabel yang sama, hal itu berguna untuk peneliti sebagai acuan dan perbandingan dalam melaksanakan penelitian ini. Berikut *state of the art* yang terdapat pada beberapa jurnal :

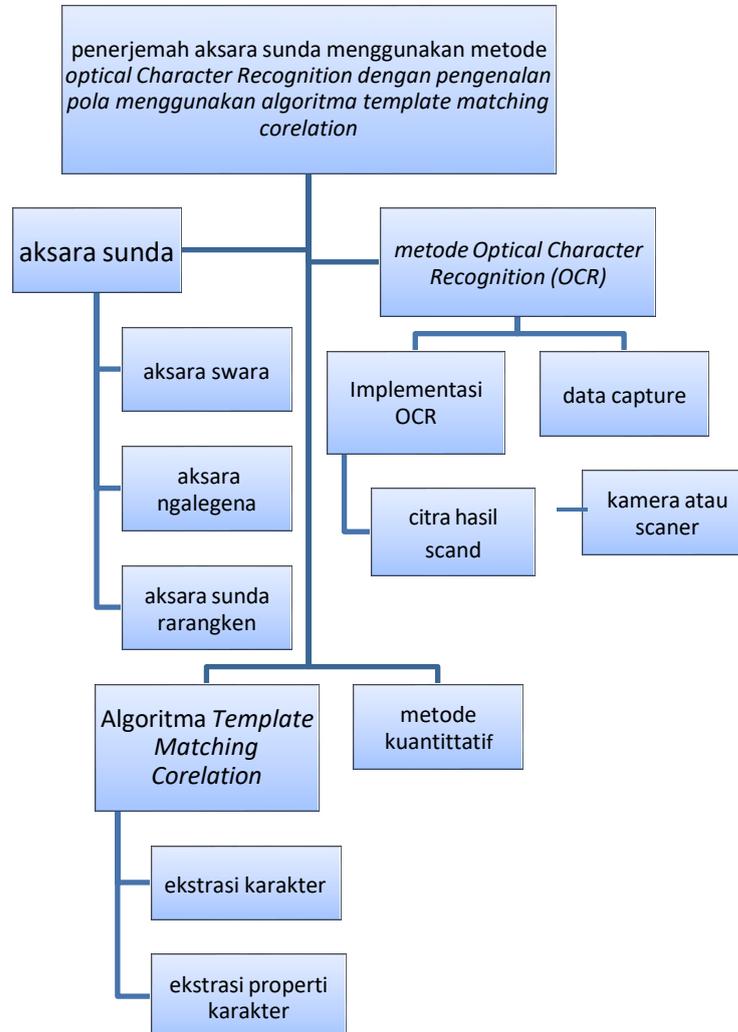
Penelitian pertama dari Setiawan, dkk pada tahun 2017 dengan judul implementasi *optical character recognition* (OCR) pada mesin penerjemah bahasa indonesia ke bahasa inggris. Tujuan dari penelitian ini adalah membuat *Optical Character Recognition* (OCR) pada mesin penerjemah bahasa Indonesia ke bahasa Inggris dengan menggunakan platform Android untuk penginputan kata penerjemahan. Sistem dibangun menggunakan *Optical Character Recognition* (OCR) dengan library mobile vision agar user tidak perlu lagi mengetikkan teks sebagai inputan kata yang akan di terjemahkan dengan material pengujian menggunakan gambarteks sebanyak lima gambar dan lembar teks sebanyak 30 kalimat.

Penelitian selanjutnya dari Tamatjita, dkk pada tahun 2008 dengan judul *optical character recognition* dengan algoritma *orchie*. Penelitian ini bertujuan untuk Pengembangan OCR diharapkan dapat memperkecil waktu dan tenaga manusia yang dibutuhkan untuk memindahkan teks cetak ke dalam bentuk digital. Meski demikian masih terdapat beberapa hal yang harus diuji pada algoritma-algoritma OCR yang dikembangkan, di antaranya dalam ketepatan pengenalan karakter.

Kemudian penelitian dari Setiady pada tahun 2019 dengan judul efektifitas pemberkasan digital melalui metode *template matching correlation*. Tujuan dari penelitian ini adalah untuk mengetahui efektivitas hasil pemindaian dokumen dengan memanfaatkan teknik pengenalan karakter optik yang dikenal dengan OCR (*Optical Character Recognition*).

2.3 Mind Map

Berikut ini adalah gambar diagram *mind map* penelitian



Gambar 2.1 Diagram *mind map*

2.4 Aksara Sunda

Terdapat beberapa aksara sunda, seperti di bawah ini :

2.4.1 Aksara swara

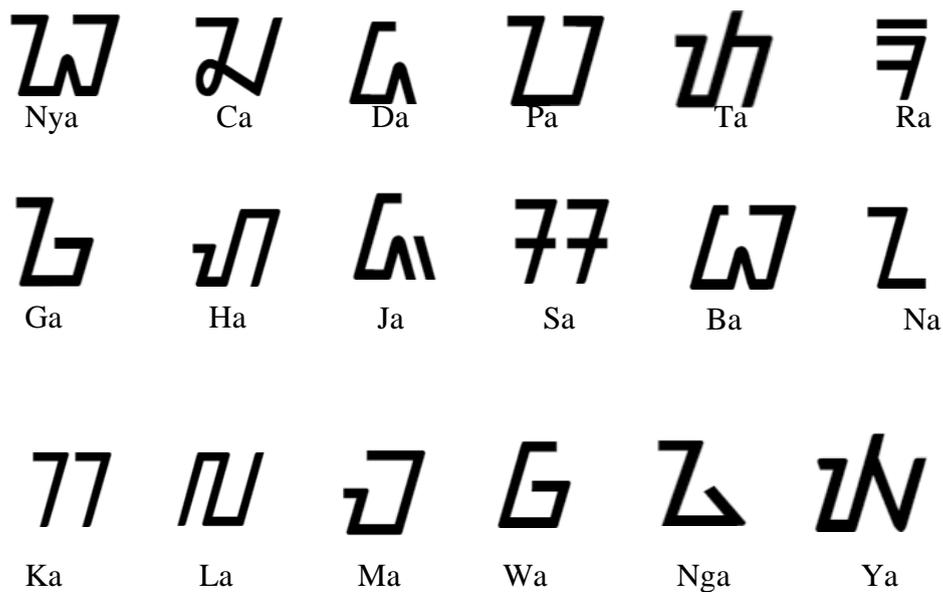
Aksara sunda atau huruf vokal dalam Bahasa sunda baku terdiri dari 7 huruf, lebih banyak dari huruf vokal Bahasa Indonesia baku yang hanya memiliki 5 huruf, berikut gambar Aksara Swara:



Gambar 2.2 aksara swara

2.4.2 Aksara ngalagena

Aksara ngalagena berarti bunyi suatu kata sebagai huruf yang dapat dilihat sebagai huruf konsonan. Aksara ngalagena secara pengucapan diakhiri dengan vokal /a/. Aksara ngalagena kuno (utama) memiliki 18 huruf, berikut gambar Aksara Ngalagena :

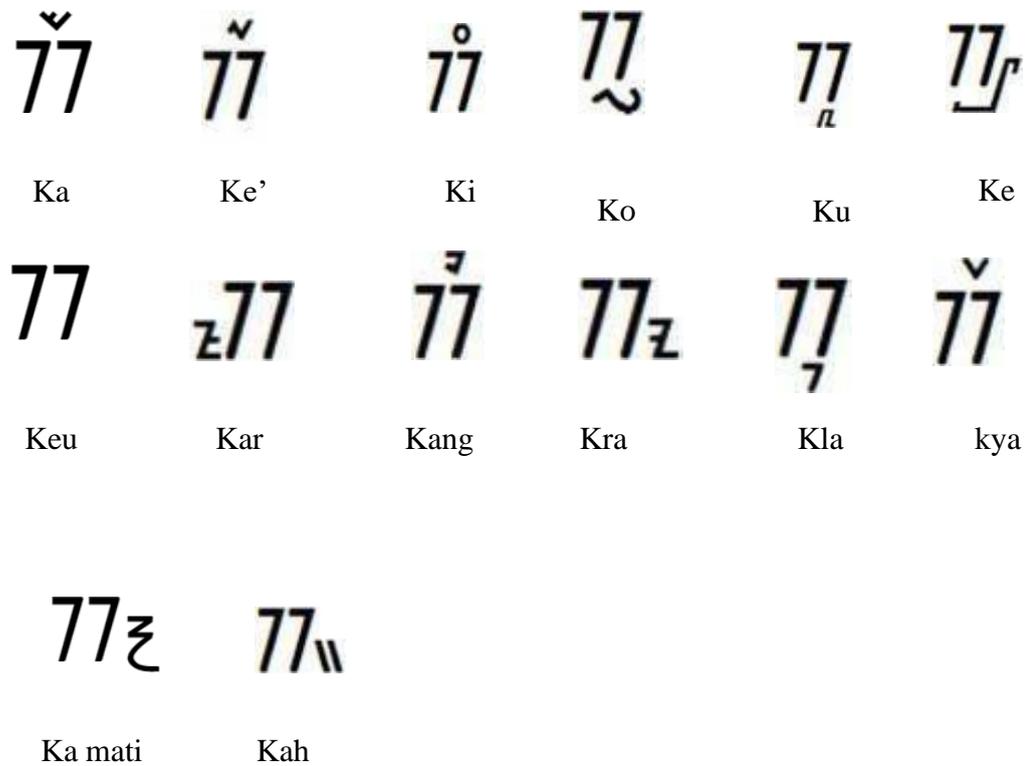


Gambar 2.3 Aksara ngalagena

2.4.3 Aksara sunda rarangken

Berdasarkan letak penulisannya, rarangkén terbagi menjadi 3 :

- Rarangkén diatas huruf
- Rarangkén dibawah huruf
- Rarangken sejajar huruf

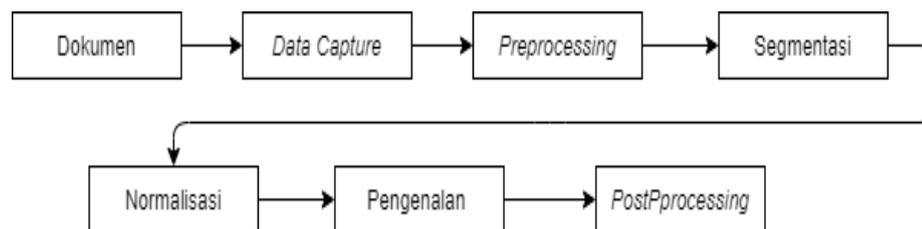


Gambar 2.4 aksara sunda rarangken

2.5 Optical Character Recognition (Ocr)

Optical character recognition OCR adalah sebuah aplikasi komputer yang digunakan untuk mengidentifikasi citra huruf maupun angka untuk di konversi ke dalam bentuk file tulisan (Hartanto,dkk.2015).

Optical character recognition (OCR) proses konversi gambar huruf menjadi karakter ASCII yang dikenali oleh komputer. Gambar huruf yang dimaksud dapat berupa hasil scan dokumen, hasil print-screen halaman web, hasil foto, dan lain-lain (Mochammad,dkk.2014), Secara umum blok diagram kerja OCR dapat dilihat padagambar :



Gambar 2.5 Diagram OCR

2.5.1 Implementasi *Optical Character Recognition* (OCR)

Optical Character Recognition (OCR) merupakan sistem konversi elektronik dari suatu karakter yang diproses secara optik (citra). Umumnya proses pengenalan karakter dapat berupa online ataupun offline. Pengenalan karakter online mengacu pada akuisisi dan pengenalan karakter secara real time. Umumnya pena optik digunakan untuk menulis karakter. Karakter tersebut dikenali pada waktu yang sama ketika menulis dengan pena optik dan kemudian ditampilkan pada komputer. Disamping itu pengenalan karakter offline mengacu pada pengenalan karakter yang ada pada lembaran kertas (dokumen/naskah). Kertas di-scan menggunakan suatu perangkat digital seperti scanner atau kamera, kemudian citra disimpan di komputer. Citra hasil scan (*scanned image*) kemudian digunakan pada proses pengenalan dimana karakter tertentu dikenali (Mehta, 2016).

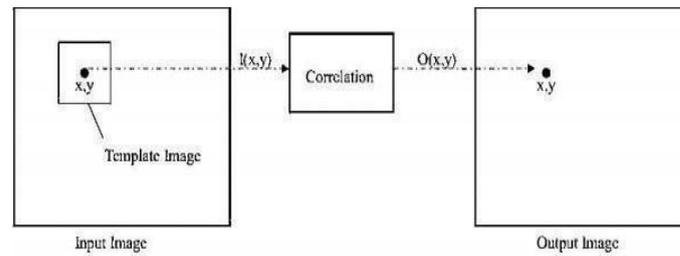
2.5.2 Data capture

Data capture adalah proses konversi dari suatu dokumen menjadi suatu file gambar digital (Bahri & Maliki, 2012). Dapat disimpulkan bahwa data capture adalah perubahan objek dari bentuk dokumen kedalam bentuk digital dengan bantuan sebuah alat seperti kamera ataupun scanner.

2.5.3 Algoritma *template matching*

Template matching merupakan sebuah teknik dalam proses pengolahan citra digital untuk menemukan bagian terkecil dari gambar yang cocok dengan template gambar. Template matching adalah salah satu ide yang dapat digunakan untuk menjelaskan bagaimana otak kita mengenali kembali pola-pola (Laksono, dkk. 2010).

Koefisien *Template Matching Correlation* merupakan jumlah representasi kemiripan antara 2 citra yang berhubungan dengan masing-masing tiap intensitas piksel.



Gambar 2.6 Contoh proses *Template Matching*

Analisis : Proses pencocokan menggerakkan citra acuan (template) ke semua kemungkinan posisi dalam citra uji dan menghitung indeks numerik yang mengidentifikasi seberapa baik template mencocokkan citra pada posisi tersebut.

a. Ekstraksi Karakter

Proses ekstraksi karakter terdapat beberapa tahap yaitu:

1. Pendeteksian garis pemisah dan teks Proses ini membedakan garis pemisah (line break) dan karakter pada input, yang bertujuan menentukan posisi karakter dari baris-baris teks output dan mengambil satu baris teks untuk diproses pada langkah selanjutnya. Caranya dengan menghitung jumlah piksel hitam yang ada pada sebaris piksel horizontal. Beberapa baris piksel yang berdekatan dan sama-sama memiliki jumlah piksel minimum akan dianggap sebuah garis pemisah. Derau mungkin akan mengacaukan proses ini, namun dapat diatasi dengan pemberian nilai toleransi jumlah piksel hitam minimum.
2. Pendeteksian dan pengisolasian komponen karakter Secara prinsip, proses ini mendeteksi secara individual setiap karakter dalam sebuah baris teks, kemudian mengisolasi tiap-tiap karakter menjadi sebuah bagian yang berbentuk segi empat. Cara yang paling sederhana yaitu dengan menerapkan cara untuk mendeteksi garis pemisah, namun dilakukan secara vertikal, mulai dari ujung paling kiri sebuah baris teks. Setelah terbagi dalam tiap-tiap karakter, piksel putih yang redundant dihilangkan sehingga terbentuk isolasi berbentuk segi

empat yang melekat pada ujung-ujung karakter.

b. Ekstraksi Properti Karakter

Setelah proses isolasi dilakukan, akan dilakukan property extraction atau perhitungan nilai-nilai properti yang melekat pada sebuah karakter. Secara keseluruhan terdapat 29 sifat atau properti yang didapatkan dari setiap karakter. Langkah pertama, karakter dibagi menjadi dua puluh lima segmen yang sama besar, dengan penentuan posisi segmen :

0	1	2	3	4
5	6	7	8	9
10	11	12	13	14
15	16	17	18	19
20	21	22	23	24
25	26	27	28	29

Selanjutnya akan dihitung pada tiap-tiap properti dengan rincian sebagai berikut: Properti 0-24 merupakan nilai rasio jumlah piksel hitam dan jumlah piksel putih pada tiap segmen karakter dalam skala keabuan (grayscale). Sebagai contoh segmen kiri atas (0) terdiri dari 4 piksel putih dan 5 piksel hitam, maka properti 0 = $(255+255+255+255+0+0+0+0+0)/9 = 113,333 = 113$. Nilai pecahan yang kemudian didapat dibulatkan; Properti 25 merupakan rasio jumlah piksel hitam dan jumlah piksel putih dalam skala keabuan, untuk separuh bagian atas karakter; Properti 26 merupakan rasio jumlah piksel hitam dan jumlah piksel putih dalam skala keabuan, untuk separuh bagian bawah karakter; Properti 27 merupakan rasio lebar dan tinggi karakter dalam skala 0-255. V; dan Properti 28 merupakan indikator karakter yang terpisah secara vertikal seperti i dan j. Jika tidak terpisah, indikator bernilai 0. Jika terpisah bernilai 255. Rangkaian kedua puluh sembilan properti di atas disebut property.

2.6 Penelitian Terdekat

Dari penelitian terdekat ini akan menjadi salah satu sumber acuan dalam melakukan penelitian sehingga dapat membandingkan metode yang digunakan dalam penelitian yang dilakukan. Penelitian terdekat ini diambil dari beberapa penelitian terdahulu yang paling mendekati dengan penelitian yang akan dilakukan. Penelitian terdekat ini juga akan dijadikan bahan perbandingan dengan penelitian yang akan dilakukan sehingga dapat diketahui perbedaan apa saja yang ada pada penelitian ini.

Dari penelitian (Setiawan, dkk 2017) dan (Tatmajita,2008) ada kemiripan dengan penelitian yang akan dilakukan dimana metode yang digunakannya sama yaitu menggunakan OCR. Perbedaan terdapat dari segi obyek dan algoritma yang diterapkan. (Setiady.2019) memiliki kesamaan yang menggunakan algoritma *template matching correlation*.