

Tema Penelitian : Transportasi dan Infrastruktur

**LAPORAN AKHIR
PENELITIAN DASAR UNHAS
PELAKSANAAN TAHUN 2019**



**SISTEM *COUNTING* PENGUNJUNG UNTUK PENGAMBILAN KEPUTUSAN
RETAILING DI INDUSTRI 4.0**

TIM PENGUSUL

**Dr.Eng. Intan Sari Areni, ST., MT (Ketua) NIDN. 0003027508
Dr. Indrabayu, ST., MT., M.Bus.Sys (Anggota1) NIDN. 0016067507
Anugrayani Bustamin, ST., MT (Anggota2) NIDN. 0001129004**

**FAKULTAS TEKNIK
UNIVERSITAS HASANUDDIN
2019**

**HALAMAN PENGESAHAN
PENELITIAN DASAR UNHAS**

Judul Penelitian : Sistem Counting Pengunjung Untuk Pengambilan Keputusan Retailing di Industri 4.0

Tema Penelitian : Transportasi dan Infrastruktur
Luaran Wajib Penelitian : Artikel ilmiah pada jurnal bereputasi

Ketua Peneliti
a. Nama Lengkap : Dr.Eng. Intan Sari Areni, ST., MT
b. NIDN : 0003027508
c. Jabatan Fungsional : Lektor Kepala
d. Fakultas/Prodi : Teknik/Teknik Elektro
e. Telpon/Faks/E-mail : 08114188802
f. Alamat surel (e-mail) : intan@unhas.ac.id

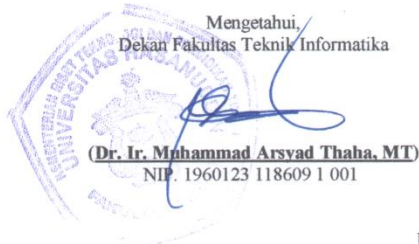
Anggota Peneliti (1)
a. Nama Lengkap : Dr. Indrabayu, ST., M.Bus.Sys., MT
b. NIDN : 0016067507
c. Fakultas/Prodi : Teknik/Teknik Informatika

Anggota Peneliti (2)
a. Nama Lengkap : Anugrayani Bustamin, ST., MT
b. NIDN : 0001129004
c. Fakultas/Prodi : Teknik/Teknik Informatika

Lama Penelitian : 10 bulan
Biaya Penelitian diusulkan : Rp. 118.950.000,-
Biaya Penelitian yang diterima : Rp. 103.000.000,-

Gowa, 24 Oktober 2019

Mengetahui,
Dekan Fakultas Teknik Informatika



(Dr. Ir. Muhammad Arsvad Thaha, MT)
NIP. 1960123 118609 1 001

Ketua Peneliti,



(Dr.Eng. Intan Sari Areni, ST., MT)
NIP. 19750203 200012 2 002

Menyetujui,
Ketua Lembaga Penelitian dan Pengabdian Masyarakat UNHAS,

(Prof. Dr. Andi Alimuddin, M.Si)
NIP. 196208181987021001

DAFTAR ISI

HALAMAN SAMPUL	i
HALAMAN PENGESAHAN	ii
IDENTITAS DAN URAIAN UMUM.....	iii
DAFTAR ISI	iii
RINGKASAN	iv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Tujuan Penelitian	2
1.3 Urgensi Penelitian	2
BAB II TINJAUAN PUSTAKA	3
2.1 <i>State of The Art</i>	3
2.2 <i>Roadmap</i> Penelitian	5
BAB III METODE PENELITIAN	6
3.1 Desain Sistem dan Tahapan Penelitian	6
3.2 Lokasi Penelitian	7
BAB IV HASIL PENELITIAN	8
REFERENSI	12

RINGKASAN

Perkembangan industri ritel di masa kini berkembang dengan sangat pesat. Salah satu hal yang dapat menjadi landasan industri ritel agar tetap bertahan adalah mendapatkan aspek perkembangan strategi *marketing* dengan memperhitungkan jumlah pengunjung yang ada. Penelitian sebelumnya sudah mencoba melakukan pendeteksian *gender* dengan fitur yang ada pada wajah namun hanya sebatas penggunaan media gambar. Penelitian ini berfokus pada pengembangan sistem yang aplikatif dengan menggunakan masukan berupa video. Sistem ini terdiri dari tiga tahapan utama, yaitu deteksi wajah, ekstraksi fitur dan klasifikasi. Algoritma-algoritma yang rencananya digunakan pada penelitian ini berdasarkan *state of the art* adalah algoritma Viola Jones, algoritma *Shearlet Transform*, dan algoritma *Support Vector Machine*.

Hasil rata-rata akurasi sistem penghitung jumlah pengunjung berdasarkan *gender* perempuan dengan menggunakan metode Viola-Jones sebagai pendeteksi wajah, metode Gabor Filter 2-D untuk ekstraksi fitur global wajah, dan metode SVM untuk proses klasifikasi memiliki total akurasi 96,52%. Adanya kesalahan sistem terjadi karena metode Gabor Filter 2-D bergantung pada tingkat kecerahan *frame*. Perbedaan tingkat kecerahan yang signifikan dapat menyebabkan kesalahan klasifikasi oleh sistem. Hasil-hasil penelitian yang diperoleh telah dituliskan dalam bentuk *manuscript* yang telah di *submit* pada *Journal of Engineering Science and Technology Review* (terindeks Scopus-Q2).

Kata kunci: Deteksi wajah, *gender*, industri 4.0, industri ritel, kecerdasan buatan.

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Masyarakat perkotaan kini dimanjakan oleh kehadiran berbagai pusat perbelanjaan. Bahkan lokasinya kadang-kadang di satu kawasan. Kondisi ini sangat menguntungkan karena masyarakat tinggal memilih gerai mana yang akan dimasukinya. Ritel merupakan mata rantai yang penting dalam proses distribusi barang dan merupakan mata rantai terakhir dalam suatu proses distribusi. Melalui ritel, suatu produk dapat bertemu langsung dengan penggunanya. Tetapi karena memiliki banyak fungsi dalam hal pemenuhan kebutuhan masyarakat, maka minat pembeli pun juga sangat tinggi, sedangkan perusahaan tetap perlu untuk melihat secara jelas perkembangan *market* yang ada pada perusahaan tersebut (Soliha, 2008).

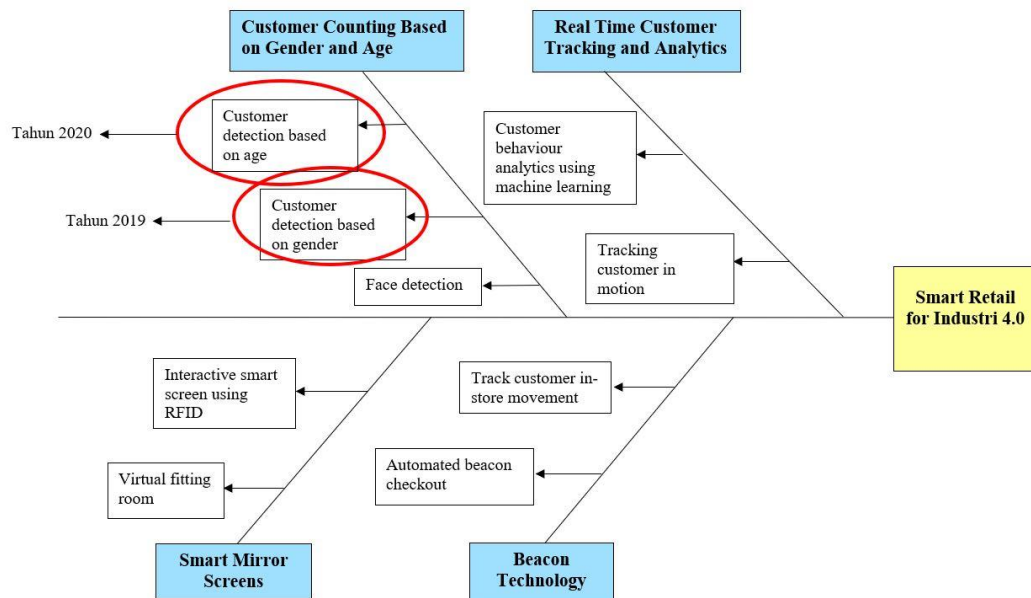
Tingginya perkembangan pasar modern akan menimbulkan permasalahan dalam suatu persaingan dan dengan keadaan tersebut menuntut para peritel untuk menyusun strategi-strategi secara inovatif. Bagi peritel untuk dapat mempertahankan serta merebut pangsa pasar di pasar modern. Diperlukannya cara pengelolaan yang tepat agar memenuhi keinginan konsumen di lingkungan bisnis ritel karena kepuasan konsumen merupakan tolok ukur peritel bagi kelangsungan perusahaan. Hal tersebut penting karena akan memberikan keuntungan yang sangat besar pada perusahaan yang bergerak dibidang ritel (Wahyu Purnama, 2016).

Salah satu hal yang dapat mendukung agar peritel dalam bersaing di pangsa modern saat ini adalah dengan mengembang *strategic marketing* yang tepat. Dalam bidang kajian *strategic marketing*, pertumbuhan pengunjung merupakan indikator utama dari pengembangan kinerja. Meningkatnya jumlah pengunjung mengindikasikan adanya kinerja pemasaran yang baik, dan tentunya sangat berpengaruh positif bagi kelangsungan hidup perusahaan (Soliha, 2008). Salah satu indikator untuk mengindikasikan adanya kinerja pasar yang baik melalui perspektif pelanggan, seperti *market share* di berbagai lokasi/wilayah, kenyamanan yang didapat oleh pengunjung, *brand image*, dan jumlah pengunjung yang datang. Dengan mengetahui perspektif tersebut, seperti jumlah dan detail pengunjung berdasarkan *gendernya* yang datang pada perusahaan tersebut, maka akan dapat lebih memaksimalkan perspektif pelanggan ini (Dr. Zaroni, 2015).

Banyaknya jumlah pengunjung yang datang pada waktu waktu tertentu menyulitkan untuk melakukan pendataan dengan cara yang mudah. Oleh karena itu, dibutuhkan suatu

sistem untuk melakukan perhitungan jumlah pengunjung yang ada berdasarkan *gender* dan usia.

Penelitian ini bertujuan untuk membangun sebuah sistem untuk menghitung jumlah pengunjung dalam pengambilan keputusan *retailing* di industri 4.0. Sistem yang diusulkan diharapkan mampu menjadi solusi bagi perusahaan yang bergerak dibidang ritel untuk dapat mengembangkan *strategic marketing* yang tepat. Adapun *fishbone* penelitian ditunjukkan pada Gambar 1 berikut.



Gambar 1. Fishbone Penelitian

1.2 Tujuan Penelitian

Penelitian ini bertujuan untuk:

1. Membuat sistem yang mampu mendeteksi wajah.
2. Membuat sistem yang dapat menghitung jumlah wajah pengunjung berdasarkan *gender* dan umur (dewasa dan anak-anak).

1.3 Urgensi Penelitian

Urgensi penelitian ini adalah:

1. Dapat menjadi alat bantu dalam pengambilan keputusan.
2. Mendukung program 100 smart city khususnya smart retailing.
3. Mendukung program making Indonesia 4.0.

BAB 2

TINJAUAN PUSTAKA

2.1 *State of The Art*

Penelitian terkait mengenai deteksi wajah dengan menggunakan deteksi citra telah dilakukan oleh beberapa peneliti dengan menggunakan metode serta teknologi yang berbeda. Déniz dkk pada penelitiannya melakukan dua percobaan pada gambar berukuran 58x50 *pixel*. Hasil yang diperoleh memperlihatkan bahwa untuk menghasilkan deteksi yang baik, maka digunakan data sampel seragam dengan menggunakan fitur HOG (*Histogram of Oriented Gradients*). Sedangkan untuk menghilangkan data yang berlebih, digunakan reduksi dimensional pada representasi HOG (*Histogram of Oriented Gradients*). Dapat dilihat juga bahwa kombinasi hasil fitur ekstraksi HOG (*Histogram of Oriented Gradients*) akan lebih maksimal jika memilih satu patokan ukuran terbaik, sehingga dapat memaksimalkan performa deteksi sebesar 13% (O. Déniz, 2011).

Sedangkan Ghorbani dkk melakukan percobaan pada gambar 58x50 *pixel*, dan rata-rata hasil klasifikasi HOG (*Histogram of Oriented Gradients*) sebesar 0,75 hingga 0,95, sedangkan klasifikasi LBP (*Local Binary Pattern*) sebesar 0,7 hingga 0,85. Penelitian ini mendapatkan hasil bahwa untuk mengimbangi kesalahan dalam deteksi fitur wajah, berupa oklusi, pose dan perubahan iluminasi, mengekstrak HOG (*Histogram of Oriented Gradients*) descriptor dari *grid* yang ada. Perpaduan antara HOG (*Histogram of Oriented Gradients*) dan LBP (*Local Binary Pattern*) memungkinkan memaksimalkan struktur penting untuk pengenalan wajah, serta menghapuskan fitur yang tidak relevan untuk mengurangi *overfitting* pada proses pengenalan wajah (Ghorbani, 2016).

Penelitian selanjutnya oleh Cerna dkk melakukan implementasi dan percobaan dengan menggunakan SVM (*Support Vector Machine*) *classifier* dengan menggunakan 2385 wajah dan 7025 sampel bukan wajah, lalu membaginya menjadi data *training* (positif dan negatif) sebanyak 30%, dengan ukuran *codebook* sebesar 10 melatih sebanyak 6588 sampel tes sebesar 99,71% dan sebanyak 2822 data *training* menghasilkan akurasi 100%, sedangkan ukuran *codebook* sebesar 100 akurasi menurun hingga 84,51% dengan data tes 1911 sedangkan menurun 84,71% dengan data latih sebanyak 7499. Dari eksperimen, didapatkan bahwa Metode *Elastic Bunch Graph Matching* dapat digunakan untuk mengekstrak fitur wajah yang penting (mata, hidung, dan lain-lainnya) dan dari hal itu akan didapatkan

descriptor HOG (*Histogram of Oriented Gradients*) tanpa menggunakan gambar keseluruhan (Cerna, 2013).

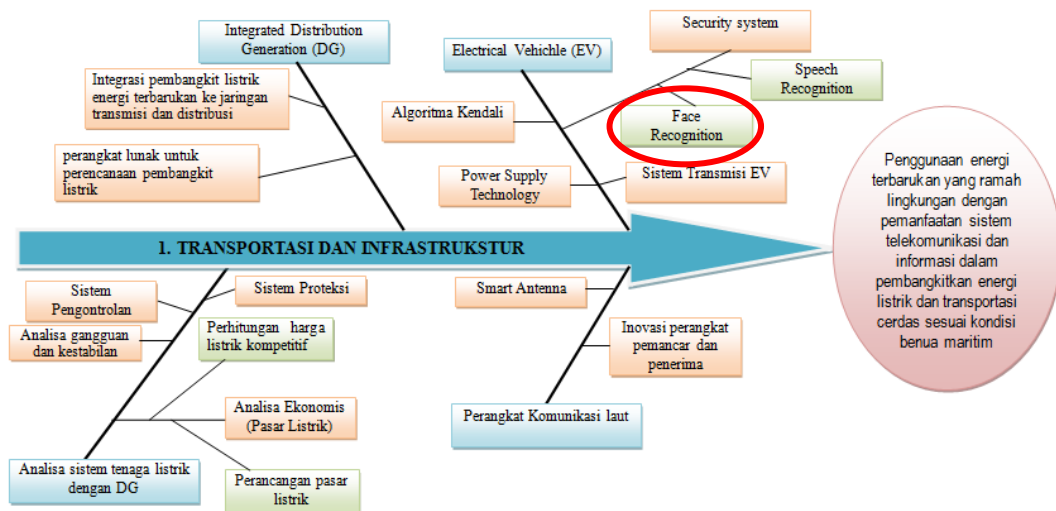
Dey dkk pada penelitiannya menyatakan bahwa jenis kelamin dapat diketahui melalui informasi yang berasal dari fitur lokal (rambut wajah, alir dan tekstur kulit) serta hubungan antara fitur dan struktur wajah. Umumnya, wajah laki-laki memiliki fitur hidung, alis, dagu dan rahang amunis daripada wajah wanita. Wajah perempuan, di sisi lain, memiliki pipi yang lebih menonjol daripada wajah laki-laki. Selain itu, penyimpangan terbesar ditemukan di daerah hidung dan dagu. Deteksi wajah hanya memerlukan 20 ms pada 128 x 128 *pixel* dengan akurasi sistem mencapai 78,91% dengan menggunakan DCT (*Discrete Cosine Transformation*) dan LBP (*Local Binary Pattern*) (Dey, 2013). Sedangkan Rekha dan Kurian melakukan pengembangan deteksi wajah dengan menggunakan descriptor HOG. Dengan menggunakan ukuran gambar 256x256 *pixel* yang dideteksi melalui 8x8 *pixel* dan dapat mendeteksi wajah dengan orientasi wajah $+90^0$ to -90^0 serta kecepatan deteksi berdasarkan jumlah wajah yang ada pada tampilan video, yang akan muncul dengan maksimum keterlambatan sampai 5 ms (Rekha dan Kurian, 2014).

Dadi dan Pillutla melakukan percobaan perbandingan hasil deteksi berdasarkan metode fitur HOG dan klasifikasi SVM dengan algoritma yang menggunakan fitur standar eigen dengan algoritma PCA. Hasil menunjukkan bahwa algoritma HOG mendapatkan pengembangan akurasi deteksi sebesar 8,75% pada *database*. Algoritma juga memverifikasi pada tujuh data set wajah lainnya, dimana terdapat 40 set *database* wajah dan setiap setnya memiliki 10 *database* wajah. HOG memiliki akurasi yang baik pada data *FERET Database*, *Yale Database*, *Yale Face Database*, *BioID*, dan *FEI*, sedangkan PCA memiliki akurasi yang lebih pada *Georgia Tech Database* (Dadi dan Pillutla, 2016). Sedangkan penelitian Azzopardi dkk menggunakan metode gabungan berupa LBP, SVM dan HOG dapat menghasilkan akurasi sebesar 92.6%, yang diekstrak dari *database FERET*. Fitur yang paling sering digunakan adalah titik penting wajah (mata, kontur wajah, dan hidung) (Azzopardi, 2016).

Berdasarkan *state of the art*, beberapa penelitian sebelumnya hanya mendeteksi wajah dan mengklasifikasi *gender* saja tetapi tidak melakukan perhitungan dan pengenalan umur. Oleh karena itu, penelitian yang diusulkan ini bertujuan untuk membuat sistem yang dapat mendeteksi wajah dan mengenali *gender* serta usia untuk memudahkan perhitungan jumlah pengunjung yang ada.

2.2 Roadmap Penelitian

Penelitian yang dilakukan ini terkait dengan bidang kajian *Face Recognition* yang merupakan bagian dari bidang unggulan **Transportasi dan Infrastruktur** sejalan dengan Roadmap Program Studi Teknik Elektro seperti diperlihatkan pada Gambar 2 berikut.



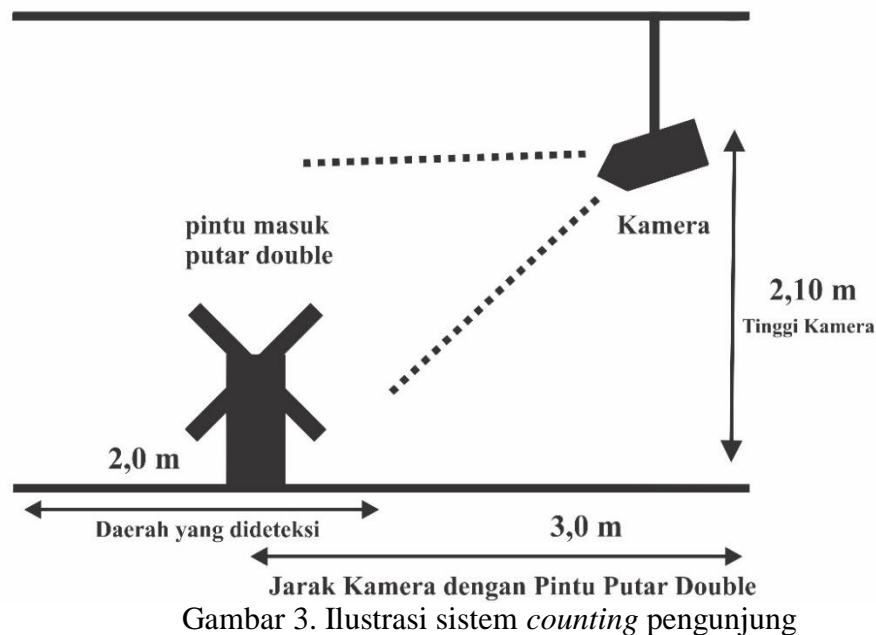
Gambar 2. Roadmap Penelitian Program Studi Teknik Elektro Unhas Terkait bidang Transportasi dan Infrastruktur

BAB 3

METODE PENELITIAN

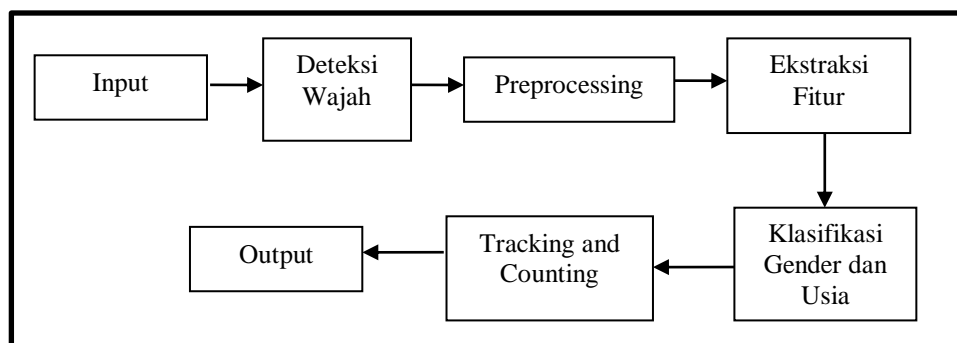
3.1 Desain Sistem dan Tahapan Penelitian

Gambaran umum penelitian ini yaitu pengambilan data video dilakukan dengan mensimulasikan keadaan pada perusahaan retail. Kamera yang terpasang pada jalur masuk pengunjung akan mendeteksi wajah pengunjung, kemudian melakukan proses klasifikasi *gender* dan usia serta perhitungan jumlah pengunjung. *Output* dari sistem ini berupa *database* yang berisikan informasi hasil perhitungan jumlah pengunjung berdasarkan hari, waktu eksekusi dan jumlah laki-laki dan perempuan serta usia dan nama *file* video yang diolah. Ilustrasi desain sistem ditunjukkan pada Gambar 3 berikut.



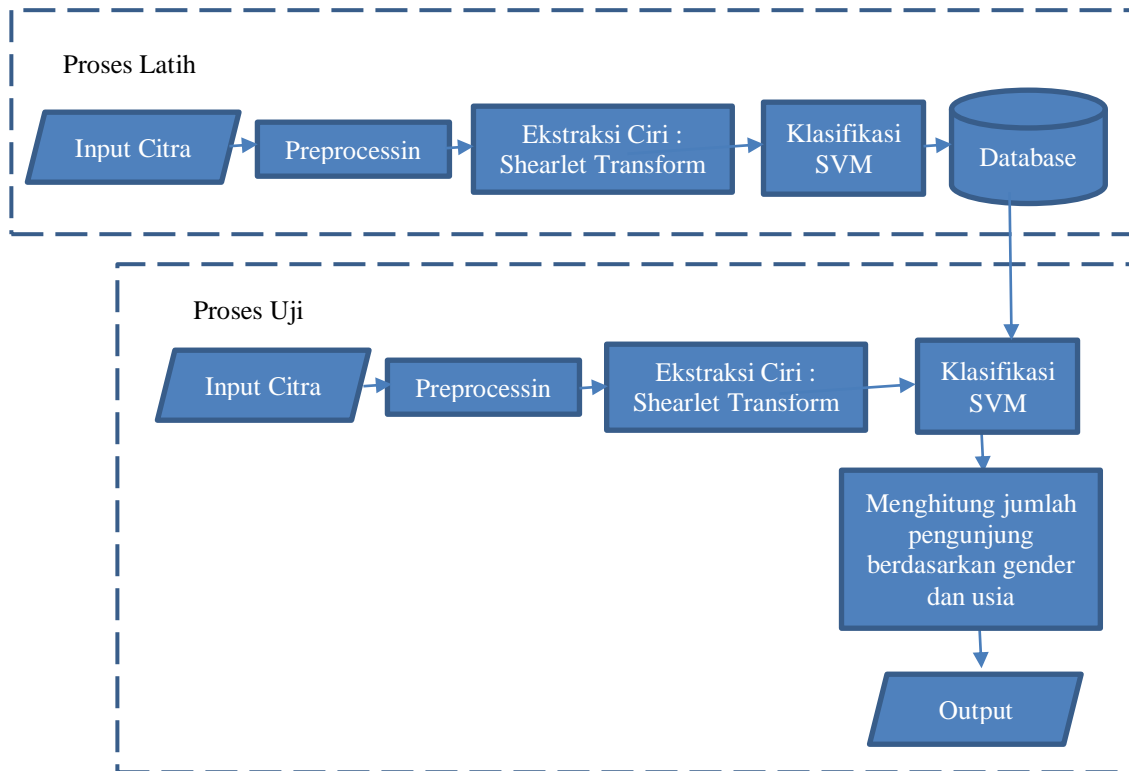
Gambar 3. Ilustrasi sistem *counting* pengunjung

Sedangkan tahapan penelitian yang akan dilakukan diilustrasikan dalam blok diagram pada Gambar 4 berikut.



Gambar 4. Blok diagram sistem *counting* pengunjung

Dari Gambar 5 dapat dilihat desain sistem terbagi menjadi dua bagian utama, yaitu proses latih dan proses uji. Pada proses latih, sistem dilatih untuk mengenali data latih dengan kelompok klasifikasi yang telah ditentukan, sehingga pada proses uji, nantinya data yang masuk akan disesuaikan dengan data yang telah dilatih pada proses uji.



Gambar 5. Desain sistem *counting* pengunjung

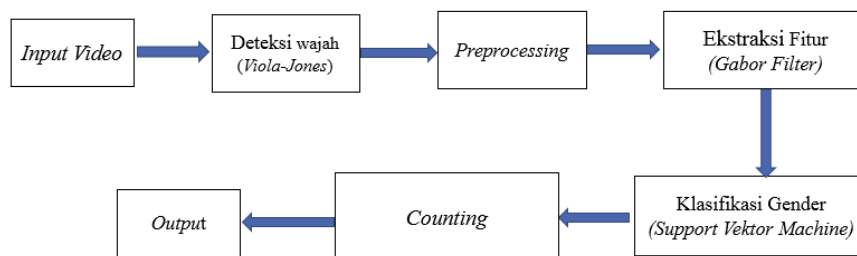
3.2 Lokasi Penelitian

Penelitian ini akan dilaksanakan di laboratorium Telekomunikasi Radio dan Microwave Departemen Teknik Elektro dan Laboratorium Kecerdasan Buatan Departemen Teknik Informatika Universitas Hasanuddin.

BAB 4

HASIL YANG DICAPAI

Tahapan sistem untuk menghitung jumlah pengunjung berdasarkan gender perempuan sebagaimana ditunjukkan pada Gambar 6.



Gambar 6. Blok diagram perancangan sistem

Berdasarkan Gambar 6, proses-proses untuk membuat sistem penghitung jumlah pengunjung perempuan dapat dijelaskan sebagai berikut.

4.1 Input Video

Input sistem merupakan hasil dari pengambilan data berupa rekaman video dengan menggunakan *ip camera* yang disimpan dalam format file video *.avi*, beresolusi 720×1280 *pixel*, dan berjenis RGB. Data yang telah di-*input* kemudian diolah untuk menciptakan data yang seragam yaitu dengan mengubah resolusi data menjadi 480×640 *pixel* agar pemrosesan dapat menjadi lebih cepat. Input video yang diambil berjumlah 6 *file* video dengan jumlah objek sebanyak 89 orang, yang dimana terdapat 41 objek perempuan dan 48 objek bukan perempuan.

4.2 Deteksi Wajah

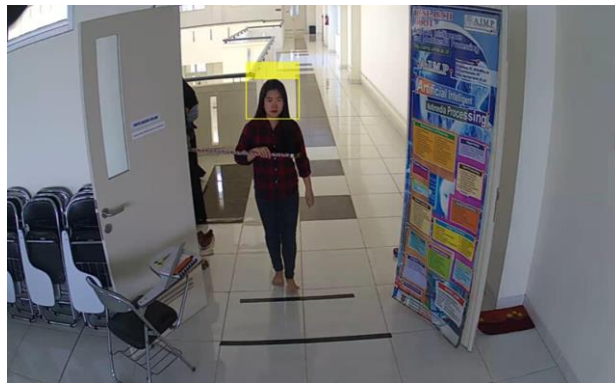
Setelah *input* dari rekaman video didapatkan, kemudian dilakukan proses deteksi wajah. Pendeteksian wajah menggunakan metode Viola-Jones fitur *Haar-Cascade Classifier*. Pada penelitian ini dilakukan uji nilai *threshold* menggunakan nilai 4, 8, dan 10. Pengujian nilai *threshold* dilakukan untuk memperkecil kemungkinan sistem mendeteksi fitur mirip wajah pada benda bukan wajah. Dengan menetapkan nilai *threshold* maka hanya fitur atau sekitarnya dengan minimal 10 kali deteksi yang akan diklasifikasikan sebagai *strong classifier*. Hasil uji coba nilai *threshold* 4, 8, dan 10 dapat dilihat pada Tabel 1.

Tabel 1. Hasil uji coba nilai *threshold*

Video	Jumlah <i>frame</i>	<i>Threshold</i>					
		4		8		10	
		TP	FP	TP	FP	TP	FP
1	2238	669	15	572	0	572	0
2	1918	752	44	632	0	636	0
3	3659	1367	10	1137	4	1137	0
4	1126	451	4	283	0	283	0
5	3545	1356	18	1260	10	1130	0
6	3430	1163	26	1014	18	918	0

Dari hasil pengujian *threshold* diperoleh bahwa nilai *threshold* terbaik adalah 10. Hal ini terlihat dari nilai *false positive* yang rendah.

Contoh keluaran *frame* dari deteksi wajah ditunjukkan pada Gambar 7.



Gambar 7. Contoh *frame* hasil deteksi wajah

4.3 Preprocessing

Tahap *preprocessing* bertujuan untuk mempersiapkan citra sebelum diproses lebih lanjut agar mendapatkan hasil yang maksimal. Pada tahap *preprocessing* ini dilakukan *resize* yang bertujuan agar data seluruh wajah dapat di filter oleh Gabor Filter 2-D. Proses ini akan mengubah ukuran gambar menjadi 101x111 piksel. Setelah *resize* citra wajah dilakukan, proses selanjutnya adalah mengubah citra yang berupa RGB menjadi citra *grayscale*. Gambar 8 menunjukkan contoh keluaran *frame* wajah setelah melalui tahap *preprocessing*.



Gambar 8. Hasil *preprocessing*

4.4 Ekstraksi Fitur

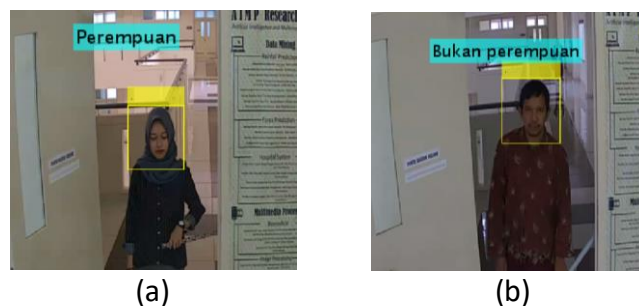
Ekstraksi fitur bertujuan untuk mendapatkan informasi-informasi penting dari sebuah citra digital. Fitur yang diperlukan dapat berupa fitur lokal wajah seperti hidung, mata, mulut, dan lainnya, maupun fitur global wajah (seluruh bagian wajah). Pada penelitian ini menggunakan fitur global wajah. Proses ekstraksi fitur dilakukan dengan menggunakan metode Gabor Filter 2-D.

4.5 Klasifikasi Gender

Klasifikasi *gender* dilakukan dengan melatih data yang telah diekstraksi fiturnya. Hasil pelatihan akan digunakan untuk proses pengujian. Tahap klasifikasi bertujuan untuk melakukan pengenalan fitur wajah berdasarkan data hasil ekstraksi ciri. Metode klasifikasi yang digunakan pada penelitian ini adalah *Support Vector Machine* (SVM). Pada penelitian ini digunakan fungsi kernel *linear*.

4.6 Counting

Proses *counting* dilakukan dengan cara menghitung jumlah hasil kelas prediksi dari setiap *frame* pada satu objek yang sama. Kemudian hasil akumulasi dari setiap kelas prediksi tersebut akan dibandingkan antara kelas satu dan kelas dua. Apabila jumlah prediksi kelas satu lebih besar atau sama dengan jumlah prediksi kelas dua maka jumlah pengunjung perempuan akan bertambah 1. Tetapi apabila jumlah prediksi kelas dua lebih besar dari jumlah prediksi kelas satu maka jumlah pengunjung bukan perempuan akan bertambah 1. Hasil counting dapat dilihat pada Gambar 9.



Gambar 9. Contoh *frame* hasil klasifikasi (a) perempuan;
(b) bukan perempuan

Pada penelitian ini, pengujian sistem menggunakan 6 file video dengan total durasi 527 detik, total frame sebanyak 15916 frame, serta jumlah objek sebanyak 89 orang. Untuk menghitung akurasi sistem, penelitian ini menggunakan aturan *confusion matrix*. Parameter pengukuran akurasi sistem adalah nilai λ yang merupakan panjang gelombang dari faktor *sinusoidal*.

Hasil Perhitungan akurasi deteksi dengan nilai $\lambda = 8$ dapat dilihat pada Tabel 2.

Tabel 2. Hasil deteksi data

Video	Objek	TP	FP	TN	FN	
A	1	√				
	2			√		
	3	√				
	4			√		
	5	√				
	6	√				
	7	√				
	8	√				
	9	√				
	10	√				
	11				√	
	12					×
B	1			√		
	2	√				
	3			√		
	4	√				
	5			√		
	6	√				
	7	√				
	8	√				
	9			√		
	10	√				
	11			√		
	12	√				
	13			√		
C	1				×	
	2				×	
	3			√		
	4			√		
	5			√		
	6			√		
	7			√		

	8			√	
	9			√	
	10	√			
	11	√			
	12	√			
	13	√			
	14	√			
	15	√			
D	16			√	
	1	√			
	2	√			
	3	√			
	4	√			
	5	√			
E	6	√			
	1			√	
	2			√	
	3	√			
	4			√	
	5			√	
	6	√			
	7	√			
	8			√	
	9	√			
	10	√			
	11			√	
	12	√			
	13			√	
	14	√			
	15	√			
	16			√	
	17			√	
	18	√			
	19			√	
	20			√	
	21	√			
22	√				
F	1			√	
	2			√	
	3	√			
	4			√	
	5			√	
	6			√	
	7	√			
	8			√	

	9		√	
	10		√	
	11		√	
	12		√	
	13	√		
	14		√	
	15		√	
	16	√		
	17		√	
	18		√	
	19		√	
	20		√	

Berdasarkan hasil uji coba pada 6 file video testing dengan jumlah objek sebanyak 89 orang, yang dimana terdapat 41 objek perempuan dan 48 objek bukan perempuan. Didapatkan hasil bahwa 41 objek perempuan terdeteksi benar sebagai perempuan dan 45 diantara 48 objek laki-laki terdeteksi sebagai bukan perempuan, sedangkan 3 lainnya terdeteksi sebagai perempuan. Berikut adalah hasil perhitungan akurasi deteksi.

$$\begin{aligned}
 \text{Akurasi} &= \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \\
 &= \frac{42 + 44}{42 + 44 + 0 + 3} \times 100\% \\
 &= \frac{86}{89} \times 100\% \\
 &= 0,9662 \times 100\% \\
 &= 96,52\%
 \end{aligned}$$

Hasil dari perhitungan akurasi kinerja sistem penghitung jumlah pengunjung berdasarkan gender perempuan dengan mengambil frame pertama dan setiap 10 frame berikutnya dapat dilihat pada Tabel 3.

Tabel 3. Hasil perhitungan akurasi dengan nilai $\lambda=8$

Video	Durasi	Jumlah Frame	Time Process (sec)	TP	FP	TN	FN	Akurasi (%)
A	74	2238	332,11	8	0	3	1	91,67
B	63	1918	303,28	7	0	6	0	100
C	121	3659	594,97	6	0	8	2	87,50
D	37	1126	179,94	6	0	0	0	100
E	118	3545	631,23	10	0	12	0	100
F	114	3430	578,06	4	0	16	0	100
Total Akurasi (%)								96,52

Berdasarkan Tabel 3 dapat dianalisa bahwa metode Viola-Jones + Gabor Filter 2-D + SVM untuk menghitung jumlah pengunjung berdasarkan *gender* perempuan memiliki rata-rata akurasi sebesar 96,52%.

Hasil rata-rata akurasi sistem dengan perbandingan nilai λ disajikan pada Gambar 10.



Gambar 10. Akurasi berdasarkan nilai λ

Berdasarkan Gambar 10, dapat dilihat bahwa akurasi sistem penghitung jumlah pengunjung berdasarkan *gender* perempuan dengan menggunakan metode Viola-Jones sebagai pendeteksi wajah, metode Gabor Filter 2-D untuk ekstraksi fitur global wajah, dan

metode SVM untuk proses klasifikasi mencapai angka tertinggi ketika diuji dengan nilai $\lambda=8$ yaitu sebesar 96,52%, sedangkan akurasi terendah diperoleh menggunakan nilai $\lambda=2$, yaitu sebesar 46,10%.

Berdasarkan analisis parameter λ , semakin tinggi nilai λ yang diberikan maka hasil segmentasi tekstur menjadi semakin cerah dan *blur*. Sedangkan semakin rendah nilai λ yang diberikan maka kondisi citra akan semakin sulit dikenali karena kontras yang rendah sehingga hasil ekstraksi menjadi tidak jelas, seperti pada nilai λ 2 dan 3. Selain itu, penyebab lainnya adalah metode Gabor Filter 2-D bergantung pada tingkat kecerahan *frame*. Perbedaan tingkat kecerahan yang signifikan dapat menyebabkan kesalahan klasifikasi oleh sistem. Contoh kesalahan klasifikasi oleh sistem dapat dilihat pada Gambar 11.



Gambar 11. Kesalahan pengklasifikasian oleh sistem.

REFERENSI

- C. Dr. Zaroni, 2015, "Menilai Kinerja Perusahaan".
- E. K. Dey, M. Khan, and H. Ali, 2013, "Computer Vision-Based Gender Detection from Facial Image," vol. 3, no. 8, pp. 428–433.
- E. Soliha, 2008, "Analisis Industri Ritel di Indonesia," *J. Bisnis dan Ekon.*, vol. 15, no. 2, pp. 128–143.
- G. Azzopardi, A. G. B, and M. Vento, 2016, "Gender Recognition from Face Images Using a Fusion of SVM Classifiers," vol. 2, pp. 533–538.
- H. S. Dadi and G. K. Mohan Pillutla, 2016, "Improved Face Recognition Rate Using HOG Features and SVM Classifier," *IOSR J. Electron. Commun. Eng.*, vol. 11, no. 4, pp. 34–44.
- L. R. Cerna dkk, 2013, "Face Detection: Histogram of Oriented Gradients and Bag of Feature Method," *Int. Conf. Image Process.*, pp. 657–661.
- M. Ghorbani, A. T. Targhi, and M. M. Dehshibi, 2016, "HOG and LBP: Towards a robust face recognition system," *10th Int. Conf. Digit. Inf. Manag. ICDIM 2015*, no. Icdim, pp. 138–141.
- N. Rekha and M. Z. Kurian, 2014, "2014_Face Detection in Real Time Based on HOG_IJAR CET-VOL-3-ISSUE-4-1345-1352," vol. 3, no. 4, pp. 1345–1352.
- O. Déniz, G. Bueno, J. Salido, and F. De La Torre, 2011, "Face recognition using Histograms of Oriented Gradients," *Pattern Recognit. Lett.*, vol. 32, no. 12, pp. 1598–1603.
- Wahyu Purnama, 2016, "Pengaruh Bauran Eceran (*Retailing Mix*) Terhadap Kepuasan Konsumen Pada Swalayan Rizki Di Pasir Pengaraian Kabupaten Rokan Hulu," Universitas Islam Negeri Sultan Syarif Kasim Riau.

Lampiran 1 : Draft Manuscript



EMaTtech Journals Editorial Manager

Journal of Engineering Science and Technology Review

Hello, INTAN SARI ARENI | Logged in as: Author/Reviewer

Assigned papers

ID	Assigned on	Original title	Section/Type	Paper	Review round	Response on round 1	Decision on round 1
----	-------------	----------------	--------------	-------	--------------	---------------------	---------------------

You have 0 assigned papers.

Submitted papers

	ID	Submitted on	Original title	Section/Type	Paper	Supplementary file	Review round	Status
Delete	3016	25/07/2019	Gender-Based Customer Counting System Using Computer Vision for Retail Stores	Research Article	jestr_sub_1564059470.doc (../uploads/jestr_sub_1564059470.doc)	jestr_cov_1564059470.doc (../uploads/jestr_cov_1564059470.doc)	1	Assigned to reviewers

You have 1 submitted paper.

[Submit paper \(submit.php\)](#)

Automatic Customer Counting System in Retail Stores Based on Female Gender

Intan Sari Areni¹, Tiwi Nur Safitri², Indrabayu³

¹Department of Electrical Engineering, Universitas Hasanuddin, Gowa, Indonesia

^{2,3}Department of Informatics, Universitas Hasanuddin, Gowa, Indonesia

Received xxxxx; Accepted xxxxxxx

Abstract

The development of a modern retail business is gradually getting faster, increasing the level of competition among retailers. The retailers are changing their business strategies in order to acquire new customers, maintain customer loyalty, and improve customer service. One way to indicate the good market performance of a retail shop is to know the number and details of visitors based on gender. The number of visitors in retail shops can be observed by installing CCTV camera. By utilizing computer vision, the data from the CCTV camera can be used to automatically calculate the number of visitors based on gender. This study designed a system to count the number of female visitors through video data. Data acquisition process involved 89 people, 41 women, and 48 men. The data are stored in .AVI video file format with a resolution of 720×1280 pixels. This system can be divided into three main stages, which are face detection using the Viola-Jones method, feature extraction using Gabor Filter 2D, and classification using Support Vector Machine (SVM) method. The result of the study showed that the system can count the number of female and non-female visitors with an accuracy rate of 96.52%.

Keywords: Computer vision, female customer counting, female face detection, gabor filter 2D, support vector machine, viola-Jones.

1. Introduction

The development of modern technology has an impact on life which is becoming more practical, fast, and economical. Along with the development of these technologies, the existence of modern retail business has become more prominent because of the shift in spending patterns where people tend to favor modern retail such as minimarket or supermarket. The community becomes more fond of shopping in the modern retail market where people can get practicality and convenience in shopping [1].

The development of the modern retail business that continues to increase has intensified the level of competition among the retailers. The competition forces the retailers to change their business strategies in order to win the competition and increase market share by maintaining and attract the customer [2].

A method to indicate good market performance is to observe the business through customer perspectives, such as the market share in various locations or regions, the comfort gained by the customers, the brand image, and the number of customers visit every day. By knowing these aspects, the company will be able to maximize the customer's perspective [3].

Counting the number of people who visit a store or a shopping center gives essential information on whether the retail space performs to its full potential. This information is useful for finding out how many visitors are attracted and identifying the busiest days and the

peak hours, give the managers the ability to make better business decisions by identifying and analyzing trends through the given systematic statistics information, optimize their sales performance, improve sales efficiency, and assign the right number of employees to give the best customer service [4].

Out of the many aspects that can influence a customer's decision, one of the major factors is gender. Men and women approach shopping with different motives, perspectives, and considerations, where women have a propensity to have higher hedonic shopping orientation than men [5]. In contrast to men who shop out of necessity, women go on shopping to purchase both essential and discretionary goods, to relax, and to socialize.

CCTV cameras have been widely used by retail companies to find out the number of customer visits each day. Data recorded by CCTV camera can be used to calculate the number of customers based on gender. However, the manual calculation requires a lot of time and effort. Therefore, retail companies need a system that can automatically count and detect customer's faces via video data from CCTV cameras by utilizing computer vision, specifically a system that focuses on calculating the number of female customers.

Many studies on face detection and recognition using computer vision approaches have been carried out so far. Kumrawat and Chawla proposed a human face detection and recognition system using a combination of skin grabbing, Gabor filter and PCA. The system is

tested with different background and lighting condition and yielded up to 100% accuracy [6]. Dey et al built a system that can automatically detect human face and gender from input images. This research is has experimented on a database of more than 4000 facial images in order to evaluate the performance of the proposed system. The average accuracy rate achieved by the system is more than 78% [7]. Wulansari et al proposed a gender identification system based on facial images using the artificial neural network (ANN). Facial features were segmented into the eye area, right eye area, left eye area, and mouth area. These segments were converted into vectors and used as input for ANN using the Backpropagation method. Identification results from 60 trained data yielded 100% accuracy, while the identification results from 60 data that have never been trained before produced 82% accuracy [8].

This research focuses to build a system to automatically count female visitor in a retail shop using CCTV camera.

2. Methodology

In this study, data retrieval is done by simulating the real situation in a retail company The CCTV camera is placed in front of the entrance door with a distance of 3.0 m from the turnstile entrance and placed on a metal pole with a height of 2.10 m above the floor with 54.5° camera angle. This prescribed angle is meant to allow the camera to capture objects clearly. The illustration of the data acquisition process can be seen in Figure 1.

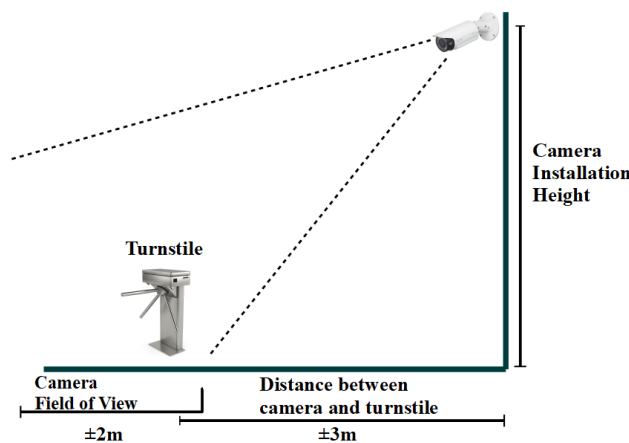


Figure 1. Illustration of data acquisition process

Data retrieval is done from the front corner of the entrance so that the object approaches the camera. Generally, system design comprises of several stages as shown in Figure 2.

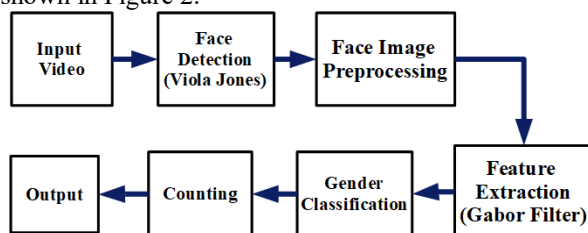


Figure 2. Block diagram of the system

Based on Figure 2, the stages to build a female visitor counting system can be explained as follows.

2.1. Input Video

The initial step to build the system is to prepare input data in video file extension. The input videos are obtained from the data acquisition process and taken using a CCTV camera. The data are stored in .AVI video file format with a resolution of 720×1280 pixels and a duration of 78 seconds with a frame rate of 30 fps. Then, the input video will be extracted to get the RGB frames. An example of a frame taken from the input video is shown in Figure 3.



Figure 3. Example of RGB frame

2.2. Face Detection

RGB frames are then used in the face detection process. Face detection uses the Viola-Jones method.

The Viola-Jones method uses a box-shaped sliding window that will scan the image to see its pixel values. The Haar-like features will be detected in the box. The shapes of Haar features can be seen in Figure 4 [9].

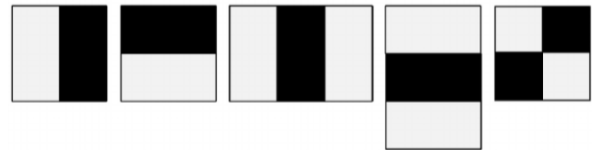


Figure 4. Haar-like Feature

Figure 4 shows the types of rectangular Haar-like features. The types of Haar-like features depending on the number of adjacent rectangles. The first type is the edge feature, which consisted of two squares and represented by the first and the second images in Figure 4. The third and the fourth comprise of three squares and represent line feature, and the last represents the diagonal feature. The value of the feature can be calculated by subtracting the pixels of the black area by pixels in the white area. These following equations can be used to get feature values according to the number of squares [10].

Two Squares : $W - B$

Three Squares: $W1 + W2 - B$

Four Squares : $(W1 + W2) - (B1 + B2)$

where B and W are the pixels value of black area and white area, respectively.

After the sliding window ends, the size of the sliding window will be reduced (resize) and it will re-scan the image. The process will continue until the size of the sliding window cannot be reduced so that many features will be obtained [11].

This study uses a threshold value to minimize the possibility of the system detecting face-like features on objects rather than faces. The threshold value used is 10. The output frame of the face detection system is shown in Figure 5.



Figure 5. Example of face detection process

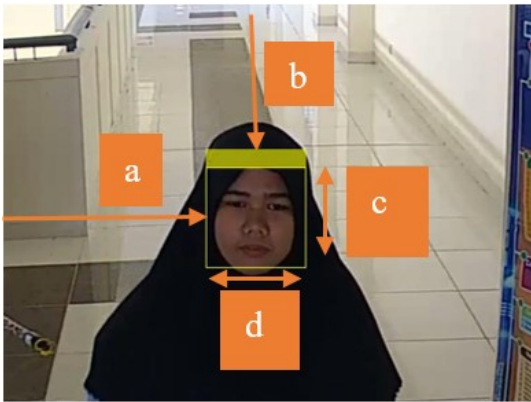


Figure 6. Bounding Box

The output of the Viola-Jones Method is a bounding box indicating the location of the face and face area. The next process is to crop and cut the face area based on the size of the detected bounding box. To get additional features around the face such as headscarves and hairstyles, the bounding box size is slightly enlarged. Figure 6 shows the bounding box section to be enlarged. Equation [12] can be used to increase the size of a bounding box.

$$\begin{aligned} a' &= a - 17 \\ b' &= b - 22 \\ c' &= c + 27 \\ d' &= d + 35 \end{aligned}$$

Where:

- a is the distance from the leftmost border of the frame to the bounding box's left margin.
- b is the distance from the topmost border of the frame to the top margin of the bounding box.
- c is the length of the bounding box.
- d is the width of the bounding box.

Figure 7 shows example of the output frame obtained from cropping process.



Figure 7. Example of the cropped image

2.3. Preprocessing

The image obtained from the face detection process varies in size. Therefore, in the preprocessing stage, image resizing is carried out to equalize face size. In this study, all face data is resized to prescribed size, that is 101×111 pixels. This value is taken based on the average image results from the face detection stage.

The next process is to convert the image from RGB to grayscale. This process is done by calculating the average value of the red, green and blue channels. Then the average value is used as the grayscale value in each pixel.

To calculate the grayscale value of each pixel can be done using equation (1)[13]:

$$Gs = 0.299R + 0.587G + 0.114B \quad (1)$$

Where :

- Gs = grayscale value
- R = red channel value
- G = green channel value
- B = blue channel value

The example of the grayscale images is shown in Figure 8.



Figure 8. Example of the grayscale image

2.4. Feature Extraction

Feature extraction aims to get important information from a digital image. The features can be local facial features (nose, eyes, mouth, etc.) or global facial features (all parts of the face). In this study, global features of the face are used. The feature extraction process is carried out using the Gabor Filter 2-D method [14]. The Gabor Filter 2-D value can be calculated using equation (4):

$$\theta(x, y, \lambda, \psi) = \left(2\pi \frac{x'}{\lambda} + \psi \right) \quad (2)$$

$$\alpha(x, y, \sigma, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \quad (3)$$

$$g = \alpha \cos \theta \quad (4)$$

The parameters used are $\lambda = 8$, $\theta = 0^\circ$, $\psi = 1$, and $\gamma = 0.5$. The result of the feature extraction can be seen in Figure 9.



Figure 9. Result of the feature extraction process

2.5. Classification

The classification process aims to analyze facial features based on data from feature extraction. The classification method used in this study is Support Vector Machine (SVM).

The first stage is to divide the data into two groups, namely training data and testing data. In this study, training data consist of 8 objects, where each object has 35 frames obtained from the face detection result. The training data is classified into two classes. Class 1 includes 4 positive data (female) and class 2 includes 4 negative data (not female). The illustration of training data can be seen in Figure 10.

Class	Sample Images
1	
2	

Figure 10. Example of positive and negative data

SVM utilizes a scoring function to determine classification groups according to the class given. SVM is a classification method using binary methods (two classes). If the input score for a certain class is positive, the input will be classified as the corresponding class. Otherwise, if the score is negative, the input will be classified as the opposite class instead.

By giving -1 label for the first class (female) and 1 for the second class (not female) then the predicted score value for all test data in the female class can be calculated using equation (5) [15]:

$$f(x) = \left(\frac{x}{s}\right)' \beta + b \quad (5)$$

$$T = \begin{cases} \text{Positive, } f(x) < 0 \\ \text{Negative, } f(x) > 0 \end{cases} \quad (6)$$

Where :

- $f(x)$ = predicted score value
- x = feature value
- s = SVM scale kernel
- β = weight
- b = bias

2.6. Counting

After the classification process is carried out, the next stage is to count female and non-female visitors. The counting process is performed by calculating the number of frames showing each person, then observing which class has the highest number of frames for each person. If the number of frames predicted in class 1 is greater than or equal to the number of frames predicted in class 2, the counter for female visitors will increase by 1. Otherwise, if the number of frames predicted in class 2 predictions is greater than the number of frames predicted in class 1, the counter for non-female visitors will increase by 1. These counters are shown on both the left and right upper corners of each frame. The examples are shown in Figure 11.



(a)



(b)

Figure 11. Visitor counting result for (a) female, (b) non-female

4. Discussion

In this study, testing data consist of 6 videos with a total duration of 527 seconds and 15916 frames. Data acquisition process involves 89 people, 41 women, and 48 men, as the test objects. To calculate system accuracy, this study uses the confusion matrix rule. Basically, the confusion matrix contains information on the comparison between classification results carried out by the system with actual classification results [16]. Example of a confusion matrix can be seen in Table 1.

Table 1. Confusion matrix

	Female	Male
Female	True Positive (TP)	False Positive (FP)
Male	False Negative (FN)	True Negative (TN)

Table 1 can be explained as follows:

- **True Positive (TP)**, is the number of positive objects correctly classified as female.
- **False Positive (FP)**, is the number of negative objects falsely classified as female.
- **False Negative (FN)**, is the number of positive objects falsely classified as non-female.
- **True Negative (TN)**, is the number of negative objects correctly classified as non-female.

To calculate the accuracy of female visitor counting system the following equation is used:

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad (4)$$

The parameter for measuring system accuracy is the value of λ which is the wavelength of the sinusoidal factor. To get the highest accuracy, the system was experimented with different values of λ . The experiment showed the best system performance is obtained using $\lambda = 8$. The result can be seen in Table 2.

Table 2. System result using $\lambda = 8$

Video	Duration	Number of Frames	Processing Time (sec)	TP	FP	TN	FN	Accuracy (%)
A	74	2238	332.11	8	0	3	1	91.67
B	63	1918	303.28	7	0	6	0	100
C	121	3659	594.97	6	0	8	2	87.5
D	37	1126	179.94	6	0	0	0	100
E	118	3545	631.23	10	0	12	0	100
F	114	3430	578.06	4	0	16	0	100
Overall Accuracy (%)								96.52

It can be seen in Table 2 that in Video A there is one positive object that is not counted as female, and in Video C there are two positive objects that are counted as female.

The obtained result is affected by the value of λ parameter. The higher the value of λ , the brighter and more blurry the result of texture segmentation becomes. If the value of λ is too low, the condition of the image will be increasingly difficult to recognize because of the low contrast, making the extraction results unclear. In addition, the Gabor Filter 2-D method depends on the brightness of the frames. Significant differences in brightness levels can cause system misclassification. Examples of system error can be seen in Figure 11.

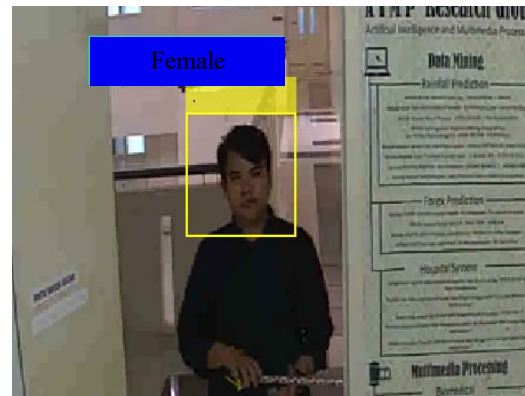


Figure 11. System classification error

Table 1 shows that the combination of Viola-Jones, Gabor Filter 2-D, and SVM method for calculating the number of female visitors has an average accuracy of 96.52%.

5. Conclusion

The combination of Viola-Jones Method, Gabor Filter 2-D, and SVM can be used to detect and count the visitors in a retail shop based on gender. The result shows that from 6 testing video data with a total duration of 527 seconds and total frame of 15916 frames, with 89 people as the test objects, the system can count the number of female and non-female visitors with an accuracy of 96.52%.

References

- [1] Pricilia Adji, "The Influence of Retail Mix on Purchase Decisions of Petra Uk Students at Circle K Siwalankerto Surabaya," vol. 1 no.2, pp. 1–10, 2013 [Indonesian].
- [2] Kasmiruddin, "THE ANALYSIS OF MODERN RETAIL BUSINESS COMPETITION STRENGTH IN PEKANBARU (AS A FORMULATION OF COMPETITIVE STRATEGY)," *J. Apl. Bisnis*, vol. 4 No.1, pp. 10–20, Oktober 2013 [Indonesian].
- [3] Dr. Zaroni, CISCIP., "Assessing Company Performance," *Artikel Supplychain*, Supply Chain Indonesia, Mei-2015 [Indonesian].
- [4] Mavenstat, *People Counting*, Mavenstat, n.d. [online] Available: <http://mavenstat.com/> [Accessed: 30/03/2019]
- [5] M. Wahyuddin, A.A. Setyawan, S.P. Nugroho, "Shopping behavior among urban women", *Mediterranean Journal of Social Sciences*, vol. 8, no. 1, pp. 306-311, January 2017.
- [6] A. Kumrawat, M.P.S. Chawla, "Real time face detection and recognition using hybrid method," *Journal of Control and Instrumentation Engineering*, vol. 3, issue 3, pp. 1-10, 2017.
- [7] E.K. Dey, M. Khan, M.H. Ali, "Computer vision-based gender detection from facial image," *International Journal of Advanced Computer Science*, vol. 3, no. 8, pp. 428-433, Aug. 2013.
- [8] D. Wulansari, E.C. Djamal, R. Elyas, "Gender identification from face image using edge detection and backpropagation," *Seminar Nasional Aplikasi Teknologi Informasi (SNATi)*, Aug. 2017.
- [9] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, Kauai, HI, USA, 2001, vol. 1, pp. I-511-I-518.
- [10] M. Yogi Septian and Fitriyani, "FACE DETECTION USING VIOLA JONES METHOD IN PROCESSING UNIT FACE DETECTION GRAPHICS USING VIOLA JONES," *Telkom Univ.*, 2014 [Indonesian].
- [11] Siti Khumaerah Mufti, "DROWSINESS RECOGNITION SYSTEMS ON CAR DRIVERS BASED ON ANDROID," *Univ. Hasanuddin*, 2019 [Indonesian].
- [12] Rina Anriani Tacok, "RECOGNITION OF HUMAN FACE FEATURES FOR DETECTION OF DROWSING DRIVERS," *Univ. HASANUDDIN MAKASSAR*, p. 2017 [Indonesian].
- [13] Muh. Rahmatullah, "Identification of Cattle (Cow) based on Android-based Muzzle Pattern," *Univ. Hasanuddin*, Feb. 2018 [Indonesian].
- [14] John G. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," *J. Opt. Soc. Am. A*, vol. 2, no. 7, pp. 1160–1169, 1985.
- [15] John C. Platt, "Probabilistic outputs for support vectors machines and comparisons to regularized likelihood methods," *Adv. Large Margin Classif. MIT Press*, pp. 61–74, 1999.
- [16] E. Prasetyo, "Digital Image Processing and Its Application using matlab," *Andi Yogyakarta.*, 2011 [Indonesia].