

## BAB 3

### METODOLOGI PENELITIAN

#### 3.1 Metode Penelitian

Penelitian ini menggunakan jenis penelitian eksperimen, dengan tahapan penelitian sebagai berikut:

1. Pengumpulan Data

Tahap ini dilakukan sebagai langkah awal dari suatu penelitian. Data pada penelitian ini berupa gambar positif dan gambar negatif. Gambar positif adalah gambar yang memuat *nipple region*. Sedangkan gambar negatif adalah gambar yang tidak memuat *nipple region*.

2. *Preprocessing*

Tahap ini adalah tahap untuk mempersiapkan data yang diperoleh dari tahap pengumpulan data sebelum dilanjutkan ke tahap berikutnya.

3. Model yang Diusulkan

Tahap ini akan dibahas metode/model yang akan digunakan untuk penelitian nanti.

4. Eksperimen dan Pengujian Model

Tahap ini akan menjelaskan tahapan penelitian dan teknik pengujian yang akan digunakan

5. Evaluasi dan Validasi Hasil

Tahap ini akan membahas tentang hasil evaluasi dari eksperimen yang telah dilakukan.

#### 3.2 Pengumpulan Data

*Dataset* pada penelitian ini bersifat privat yang meliputi *data training* dan *data testing*. *Dataset* berupa kumpulan citra positif dan citra negatif. Citra positif adalah citra yang memuat objek yang akan dideteksi. Dalam konteks penelitian ini, berarti objek di sini adalah puting payudara yang sudah ditandai dari citra pornografi. Sedangkan citra negatif adalah citra yang sama sekali tidak mengandung objek puting payudara. Citra negatif yang digunakan sebagiannya merupakan kumpulan

citra yang memiliki kesamaan dengan puting payudara baik warna, tekstur dan bentuknya. Seperti: mata, pusar dan lekukan kulit anggota tubuh. Seluruh citra positif dan negatif diambil dari internet. Pada penelitian ini jumlah citra positif yang digunakan sebesar 897 citra, sedangkan citra negatif sebesar 1836 gambar. Citra positif dibagi dua, 731 untuk pelatihan dan 166 untuk pengujian. Citra negatif juga dibagi dua 1756 untuk pelatihan dan 80 untuk pengujian. Sehingga, gambar yang digunakan untuk pengujian tidak sama dengan yang digunakan untuk pelatihan. Puting payudara yang digunakan sebagai *data training* maupun *data testing* memiliki bentuk, ukuran, warna, tekstur, posisi dan pencahayaan yang berbeda-beda. Dikarenakan bentuk payudara wanita sangat variatif.

Permisahan (*cropping*) area puting payudara pada citra pornografi dilakukan dengan menandai *Region of Interest* (ROI) puting tersebut. Untuk menandai area puting tersebut digunakan tools *objectmarker.exe*. Dengan menggunakan *tool* tersebut pengguna cukup menandai area puting payudara dengan menahan mouse (klik kiri), kemudian menggambar persegi panjang yang meliputi area puting payudara. Setelah diperoleh persegi yang menandai area puting payudara dilanjutkan dengan menekan tombol spasi pada *keyboard* untuk menyimpan area yang berhasil ditandai. Dan untuk melanjutkan ke citra selanjutnya adalah dengan menekan enter. Hasil yang diperoleh dari proses penandaan ini adalah koordinat x, y, tinggi dan lebar area puting payudara.

### **3.3 Preprocessing**

Setelah semua citra digital kita dapatkan maka langkah selanjutnya adalah *preprocessing*. Tahap *preprocessing* pada penelitian ini dengan menyiapkan semua *tools* yang dibutuhkan untuk pembuatan sampel positif dan pelatihan data. *Tools* yang digunakan merupakan hasil download [31], dengan nama *tools.zip*. Kemudian diekstrak menjadi folder khusus. Folder khusus ini diberi nama *nipple training* yang didalamnya berisi kumpulan sub folder dan *tools* yang digunakan untuk melakukan proses *haartraining*. Isi dan urutan dari subfolder ini sudah secara *default* mengikuti sesuai dengan kebutuhan untuk proses *haartraining*.

cascade2xml	12/15/2006 3:22 AM	File folder	
temp	12/15/2006 3:43 AM	File folder	
test_recognition	12/14/2006 11:00 ...	File folder	
haarTraining	12/15/2006 4:09 AM	Windows Batch File	1 KB
samples_creation	12/15/2006 4:09 AM	Windows Batch File	1 KB

**Gambar 8** Tools untuk Haartraining

Terdapat tiga buah sub folder didalamnya, yaitu *cascade2xml*, *temp* dan *test\_recognition*. Sub folder *cascade2xml* digunakan untuk menempatkan file-file *stage classifier* yang diperoleh dari proses *haartraining*. Sedangkan sub folder *temp* terdiri dari folder *data*, *positive* dan *negative* seperti gambar berikut:

data	15/12/2006 4:10	File folder	
negative	08/07/2013 22:36	File folder	
positive	15/12/2006 1:16	File folder	
createsamples	26/07/2005 17:35	Application	72 KB
cv.dll	27/02/2003 19:41	Application extens...	1.161 KB
cv097.dll	26/07/2005 17:34	Application extens...	785 KB
cxcore097.dll	26/07/2005 15:49	Application extens...	1.017 KB
haartraining	03/12/2006 6:06	Application	88 KB
highgui.dll	27/02/2003 19:44	Application extens...	500 KB
highgui097.dll	26/07/2005 17:34	Application extens...	393 KB
libguide40.dll	06/11/2004 12:18	Application extens...	184 KB
performance	26/07/2005 17:35	Application	24 KB

**Gambar 9** Isi sub folder *temp*

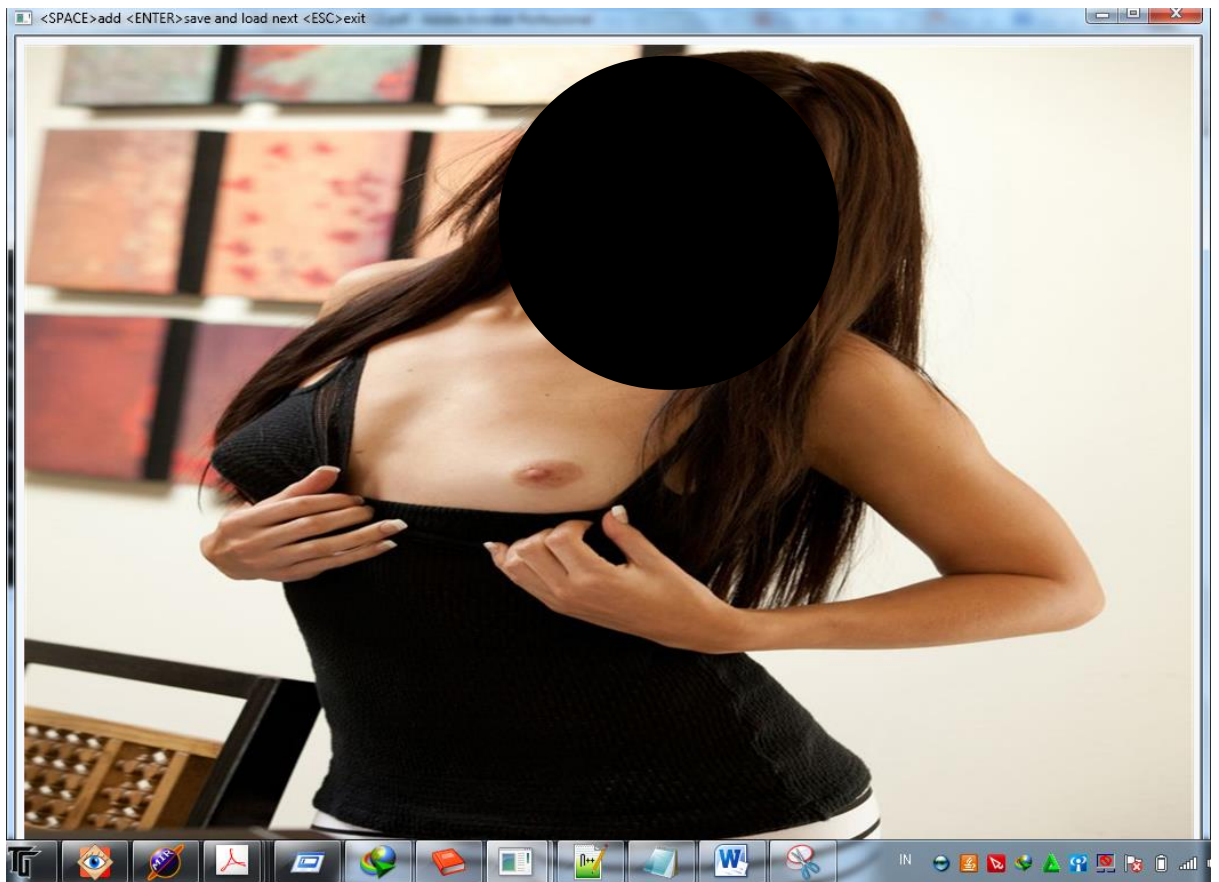
Folder *positive* berisi sub folder *rawdata* yang digunakan untuk meletakkan seluruh citra sampel positif. Folder *negative* berisi seluruh citra sampel negatif. Sedangkan folder *data* berisi file-file *cascade classifier* hasil dari proses *haartraining* sejumlah *stage* yang ditentukan dan juga tempat penyimpanan sampel positif yang telah dikonversi dalam bentuk *vector*.

### 3.3.1 Pembuatan sampel gambar positif untuk pelatihan data

Pada folder *positive* seperti yang tampilkan pada Gambar 9 diatas didalamnya terdapat sub folder *rawdata* dimana pada sub folder tersebut diletakkan semua citra positif. Setelah semua citra positif terkumpul di folder *rawdata*, maka langkah

selanjutnya adalah proses *cropping* atau penandaan ROI dengan *objectmarker.exe*. Cara penggunaanya sangat mudah dan simpel:

1. Jalankan *objectmarker.exe* dengan mengklik dua kali.
2. Setelah itu akan muncul jendela yang menampilkan citra yang tersimpan di folder *rawdata* satu per satu beserta jendela *console* yang menginformasikan koordinat *x*, *y* dan dimensi (lebar dan tinggi) ROI di citra tersebut. Tampilan kedua jendela ditunjukkan oleh Gambar 10 dan Gambar 11.

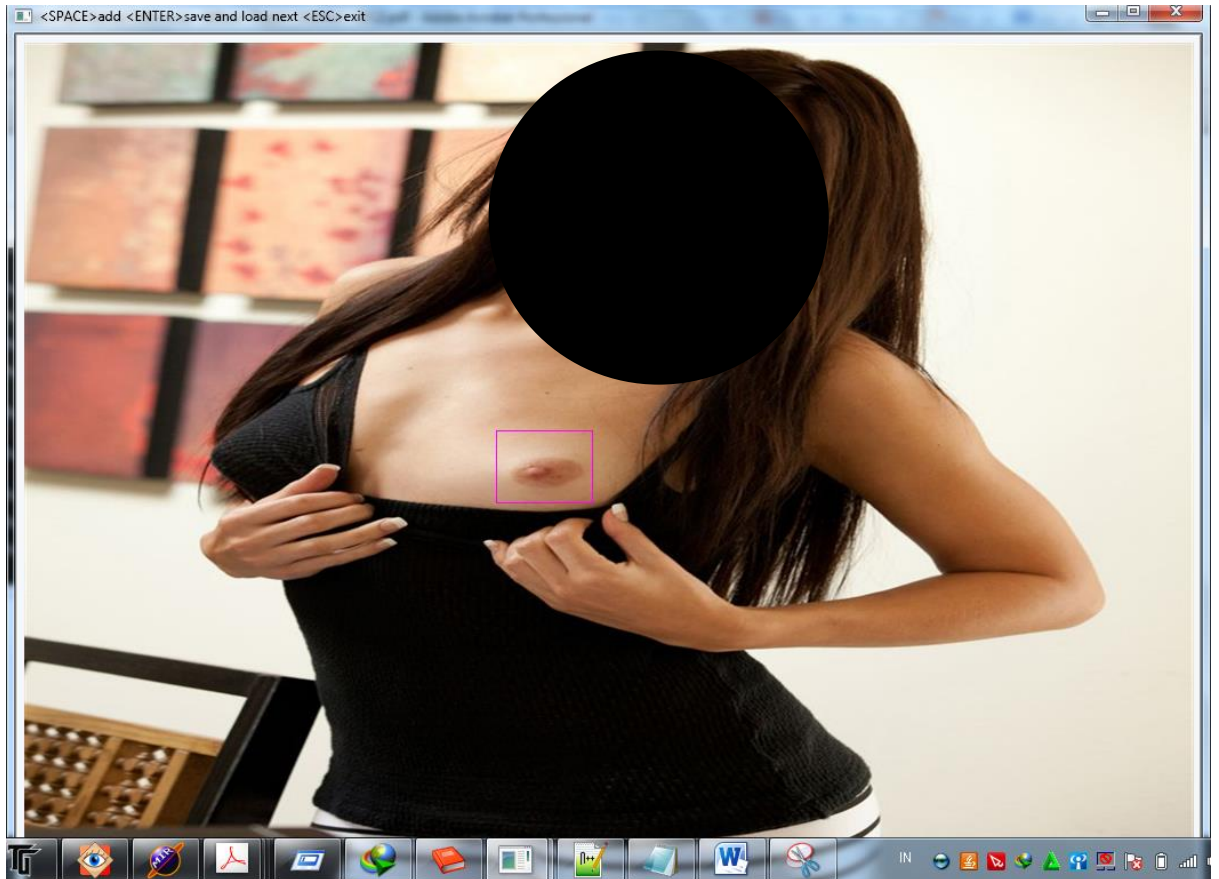


**Gambar 10** Citra yang akan ditandai

```
E:\OpenCV Ugo Rampe\nipple training\temp\positive\objectmarker.exe
2. rect x=612      y=340  width=37  height=31
Porn2161.bmp
1. rect x=584      y=322  width=61  height=41
Porn2162.bmp
1. rect x=74  y=599  width=94  height=64
2. rect x=520     y=611  width=99  height=77
Porn2163.bmp
1. rect x=507     y=313  width=30  height=25
Porn2164.bmp
1. rect x=485     y=674  width=70  height=51
Porn2165.bmp
1. rect x=479     y=309  width=52  height=37
Porn2166.bmp
1. rect x=346     y=376  width=84  height=49
2. rect x=698     y=380  width=54  height=44
Porn2167.bmp
1. rect x=733     y=510  width=92  height=85
Porn2168.bmp
1. rect x=363     y=353  width=84  height=59
Porn2169.bmp
1. rect x=310     y=352  width=72  height=48
Porn217.bmp
1. rect x=705     y=292  width=30  height=29
Porn2170.bmp
```

**Gambar 11** Jendela *console* yang menampilkan informasi koordinat dan dimensi ROI

3. Klik kiri mouse dan tahan, kemudian geser untuk menandai area ROI. Hasil penandaan berupa persegi panjang warna merah. Dan untuk meng-*cancel* penandaan ROI ulangi klik atau klik disembarang lokasi. Jika dalam satu citra ada lebih dari satu ROI, tandai semuanya satu persatu dengan menekan tombol Space keyboard setiap kali satu penandaan selesai dan untuk disimpan kemudian dilanjut ke ROI lain yang masih dalam satu citra yang sama.



**Gambar 12** Penandan ROI

Jendela *console* pada Gambar 11 muncul bersamaan dengan setiap citra yang akan ditandai. Ketika proses penandaan sedang berlangsung tidak boleh komputer dimatikan. Karena akan menghilangkan semua data ROI yang sudah tersimpan dan harus mengulangi lagi dari awal.

4. Untuk berpindah ke citra selanjutnya tekan tombol Enter.
5. Lakukan sampai semua citra pada folder rawdata selesai ditandai. Informasi koordinat dan dimensi ROI tersimpan di file info.txt.

Hasil output penandaan ROI berupa nama citra yang ditandai beserta keterangan koordinat dan dimensi ROI dari citra tersebut. Contoh: `rawdata/Porn995.bmp 1 453 308 65 55`. `Porn995.bmp` adalah nama citra yang ditandai. Angka `1` menandakan jumlah ROI yang ditandai. Angka `453 308` adalah skala koordinat nilai x dan y. Sedangkan angka `65 55` adalah nilai lebar dan tinggi ROI. Jika ROI yang ditandai lebih dari satu maka hasil outputnya `rawdata/Porn1123.bmp 3 421 333 59 40 480`

235 49 40 776 213 55 45 dengan komposisi urutan yang sama. Hasil output penandaan ini disimpan dalam file info.txt, sebelum nantinya diubah ke dalam bentuk vector. Total sampel positif yang dihasilkan dari proses penandaan adalah 1181 sampel citra.

Langkah selanjutnya adalah menciptakan file vector hasil penandaan keseluruhan citra. File vector inilah sebagai bentuk akhir sampel positif untuk proses pelatihan. Informasi file vector ini diperoleh dari file info.txt yang sebelumnya digunakan untuk menyimpan posisi ROI yang berupa jumlah, koordinat dan dimensinya. *Tool* yang digunakan adalah createsamples.exe yang sudah dipaket menjadi satu dalam paket software diatas. Untuk menjalankannya harus melalui Run di Windows. Jalankan createsamples.exe sesuai lokasi *path*-nya. Perintah untuk menjalankan:

```
E:\OpenCV Ugo Rampe\nipple
training\temp\createsamples.exe -info positive/info.txt -
vec data/nipple.vec -num 1181 -w 20 -h 20
```

Keterangan parameter:

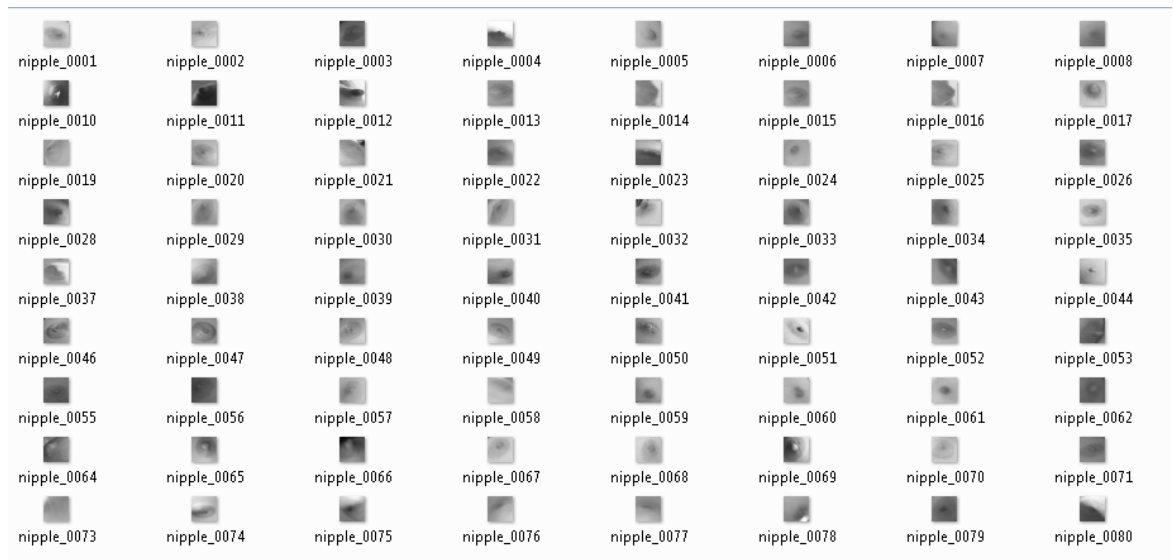
- info positive/info.txt : lokasi file info.txt yang menyimpan informasi seluruh ROI.
- vec data/nipple.vec : output nama file yang berisi sampel positif dalam bentuk vector yang dihasilkan untuk pelatihan.
- num 1181 : jumlah sampel positif. Defaultnya adalah 1000.
- w 20 : lebar sampel positif. Defaultnya adalah 24.
- h 20 : tinggi sampel positif Defaultnya adalah 24.

File yang dihasilkan dari proses pembuatan sampel positif diatas adalah nipple.vec. File nipple.vec ini yang nantinya akan digunakan untuk proses *haartraining*.

File vector dapat dikonversi ke bentuk citra dengan *tool* tambahan vec2img.exe. Untuk mengkonversi file vector ke bentuk citra, ketikkan perintah berikut:

```
E:\OpenCV Ugo Rampe\nipple training\temp\data\Vector to
Image\vec2img.exe nipple.vec image/nipple_%04d.png -w 20
-h 20
```

Pertama kali adalah menjalankan `vec2img.exe` lewat Run Windows dengan target file vector `nipple.vec`. Hasil konversi file vector ke citra Portable Network Graphis (.png) disimpan di folder image dengan nama file `nipple_0001.png` sejumlah total file sampel positif. Hasilnya seperti gambar berikut:



**Gambar 13** Hasil konversi *vector* ke *image*

### 3.3.2 Pembuatan Sampel Citra Untuk Pengujian

Citra positif yang digunakan untuk pengujian disimpan di folder testing dengan berbagai ukuran resolusi piksel. Jumlah citra yang akan digunakan untuk pengujian sebanyak 166 citra yang mengandung objek *nipple* dengan berbagai varian warna, tekstur, bentuk dan kondisi pencahayaan dan 80 citra yang tidak mengandung objek *nipple*. Citra yang tidak mengandung *nipple* digunakan untuk mendapatkan nilai dari *True Negative* (TN).

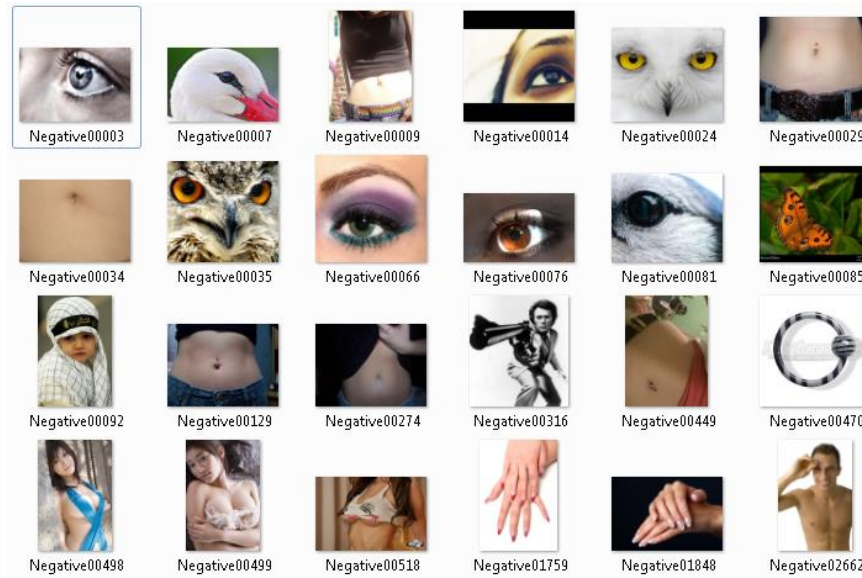




**Gambar 14** Citra pornografi yang akan diuji

### 3.3.3 Pembuatan Sampel Citra Negatif

Pembuatan sampel citra negatif diawali dengan meletakkan seluruh citra negatif di folder Negative. Proses pembuatan sampel negatif tidak sama dengan pembuatan sampel positif yang harus menandai satu per satu area objek. Tapi, cukup dengan menjalankan file `create_list.bat`. Secara otomatis semua nama file citra akan tersimpan di file `infofile.txt`.

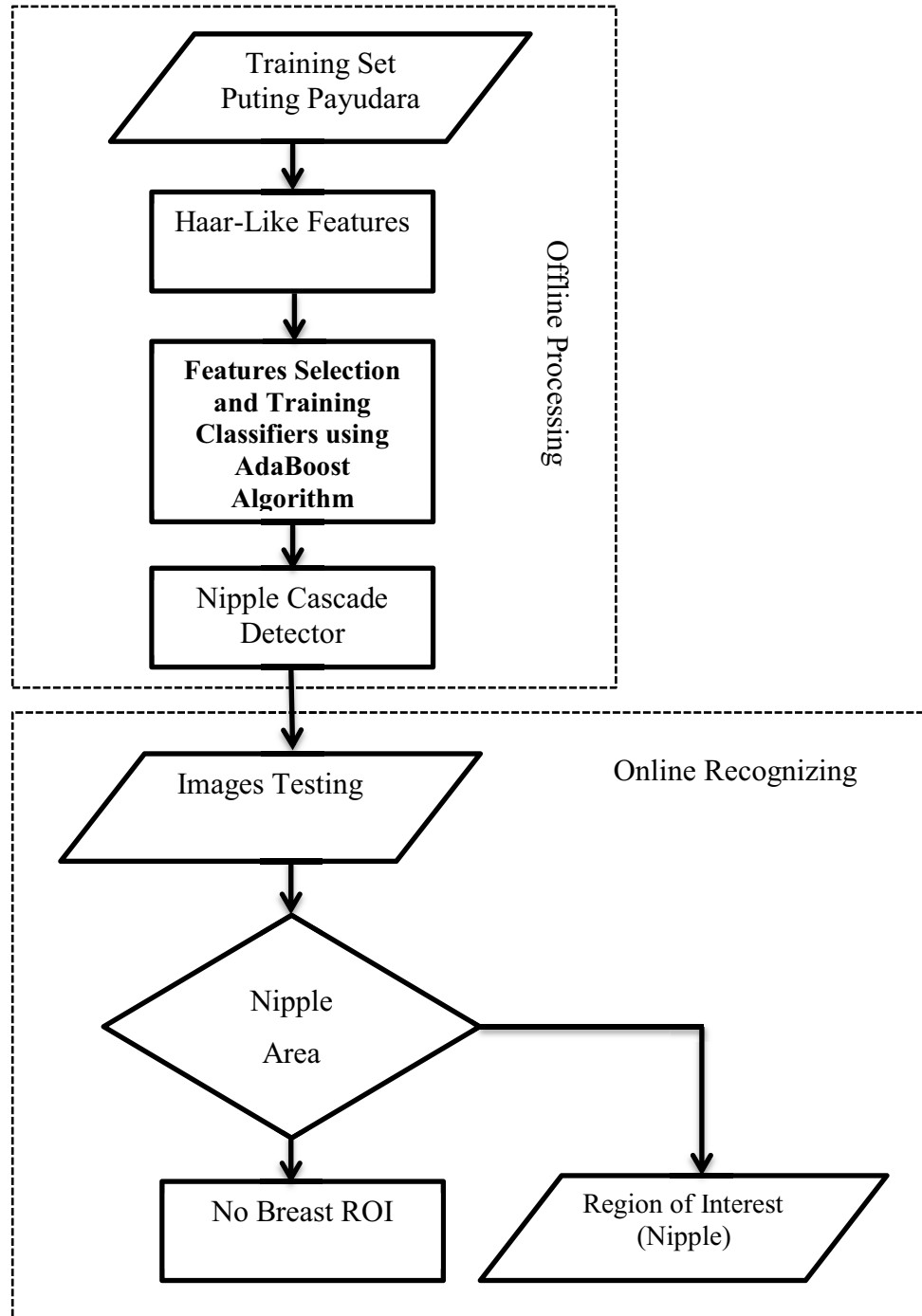


**Gambar 15** Citra negatif yang akan *ditraining*

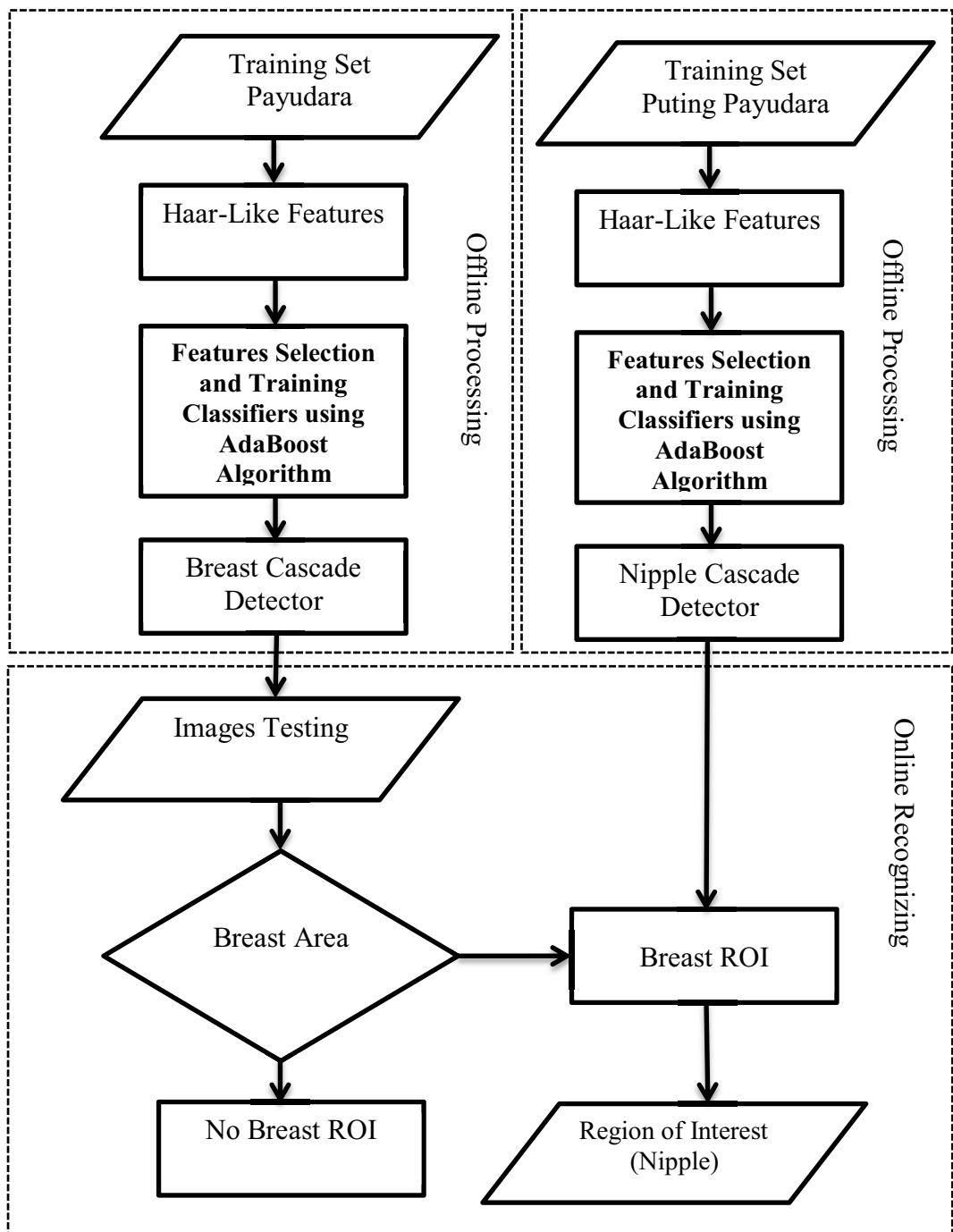
### 3.4 Model Yang Diusulkan

Skema model yang diusulkan ditunjukkan oleh Gambar 17, sedangkan metode single detector ditunjukkan Gambar 16. Langkah pertama adalah melakukan *training* terhadap sampel positif dan negatif dengan ukuran *subwindow* 20x20 untuk menghasilkan *cascade classifiers* atau *cascade detector*. Pada proses training ini menggunakan algoritma AdaBoost untuk menyeleksi fitur-fitur dan melatih *classifiers*. *Classifier* kuat dihasilkan dari penggabungan sekumpulan *classifiers* lemah. Setelah terbentuk *cascade detector* dari proses pelatihan, maka langkah selanjutnya adalah melakukan *online recognizing* dengan dua kali proses deteksi. Deteksi tahap pertama untuk mendapatkan area payudara. *Cascade detector* untuk deteksi payudara diperoleh dari mengunduh [32] dengan nama file *cascadebreast1981.xml*. Sehingga tahapan *training* hanya pada *training set* puting payudara. Setelah berhasil didapatkan area yang dianggap sebagai payudara, dilanjutkan mencari area puting pada area yang didapatkan dari hasil deteksi pertama (payudara). *Cascade detector* terdiri dari dua macam, yaitu *cascade nipple detector* dan *cascade breast detector*.

Tujuan utama dari dua kali pendeteksian adalah untuk mengurangi banyak positif palsu (*false positive*). Umumnya positif palsu yang terjadi pada area mata dan pusar.



**Gambar 16** Skema model metode *single detector*



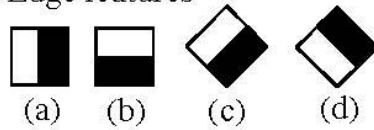
**Gambar 17** Skema model yang diusulkan

1. *Haar-like Features* dan *Integral Image*

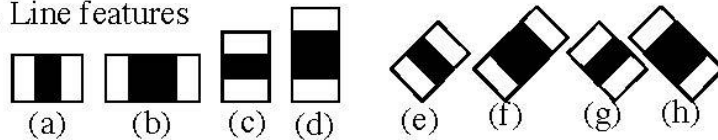
Fitur yang digunakan oleh algoritma Viola and Jones dalam mengenali objek adalah *Haar-like features*. *Haar-like features* yang digunakan pada penelitian

ini adalah *mode ALL* atau semua tipe fitur [23]. Nilai fitur *Haar-like* diperoleh dengan hasil selisih daerah gelap dengan daerah terang.

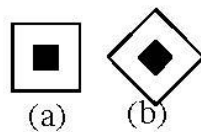
1. Edge features



2. Line features



3. Center-surround features



4. Special diagonal line feature used in [3,4,5]



**Gambar 18** Model *Haar-like Features* yang akan digunakan

Untuk menghitung area tersebut dengan sangat cepat dan radikal Viola and Jones menggunakan sebuah representasi citra baru yang dinamakan *integral image*. Dengan menggunakan *integral image* nilai setiap persegi pada fitur Haar-like bisa diperoleh hanya dengan perhitungan empat referensi *array*, tanpa harus menjumlahkan seluruh nilai piksel pada daerah gelap dan terang. Perhitungan nilai fitur tanpa citra integral:

1	2	2	4	1
3	4	1	5	2
2	3	3	2	4
4	1	5	4	6
6	3	2	1	3

$$\text{Nilai fitur} = |\text{Sum}_{\text{black rectangle (pixel gray level)}} - \text{Sum}_{\text{white rectangle (pixel gray level)}}|$$

$$= |(4+1+3+3+1+5+3+2) - (5+2+2+4+4+6+1+3)|$$

$$= |22 - 27|$$

$$= 5$$

Perhitungan dengan citra integral:

1	3	5	9	10
4	10	13	22	25
6	15	21	32	39
10	20	31	46	59
16	29	42	58	74

$$\begin{aligned} \text{Nilai fitur} &= |Sum_{black\ rectangle} (pixel\ gray\ level) - Sum_{white\ rectangle} (pixel \\ &\quad gray\ level)| \\ &= |[42+1-(16+5)] - [74+5-(42+10)]| \\ &= |22 - 27| \\ &= 5 \end{aligned}$$

## 2. Algoritma AdaBoost

Algoritma AdaBoost digunakan untuk menyeleksi fitur-fitur dan melatih *classifiers*. AdaBoost bertindak menggabungkan beberapa *weak classifier* (pengklasifikasi lemah) untuk membentuk *strong classifier* (pengklasifikasi kuat). Setelah tahap pertama pembelajaran, data-data akan dibobot ulang untuk memberitahu kesalahan klasifikasi *weak classifier* pada tahap sebelumnya. Sehingga, *strong classifier* akhir merupakan bentuk kombinasi berbobot dari *classifiers* lemah yang diikuti dengan sebuah *threshold*. Berikut langkah-langkah algoritmanya [22]:

- Diberikan sampel citra  $(x_1, y_1), \dots, (x_n, y_n)$  dimana  $y_i=0,1$  untuk sampel negatif dan positif berturut-turut.
- Inisialisasikan bobot  $w_{1,i} = \frac{1}{2m}, \frac{1}{2l}$  untuk  $y_i = 0,1$  berturut-turut, dimana  $m$  dan

$l$  adalah jumlah positif dan negatif berturut-turut.

- Untuk  $t = 1, \dots, T$ :

5. Normalisasikan bobot,

$$w_{t,i} \leftarrow \frac{w_{t,i}}{\sum_{j=1}^n w_{t,j}}$$

Sehingga  $w_t$  adalah distribusi probabilitas.

6. Pilih classifier lemah terbaik dengan urutan ke bobot kesalahan

$$\epsilon_t = \min_{f,p,\theta} \sum w_i |h(x_i, f, p, \theta) - y_i|.$$

7. Definisikan  $h_t(x) = h(x, f_t, p_t, \theta_t)$  dimana  $f_t, p_t$ , dan  $\theta_t$  adalah bentuk minimize dari  $\epsilon_t$ .

8. Update bobot:

$$w_{t+1,i} = w_{t,i} \beta_t^{1-e_i}$$

dimana  $e_i = 0$  jika sampel  $x_i$  diklasifikasikan dengan benar,  $e_i = 1$

sebaliknya, dan  $\beta_t = \frac{\epsilon_t}{1-\epsilon_t}$ .

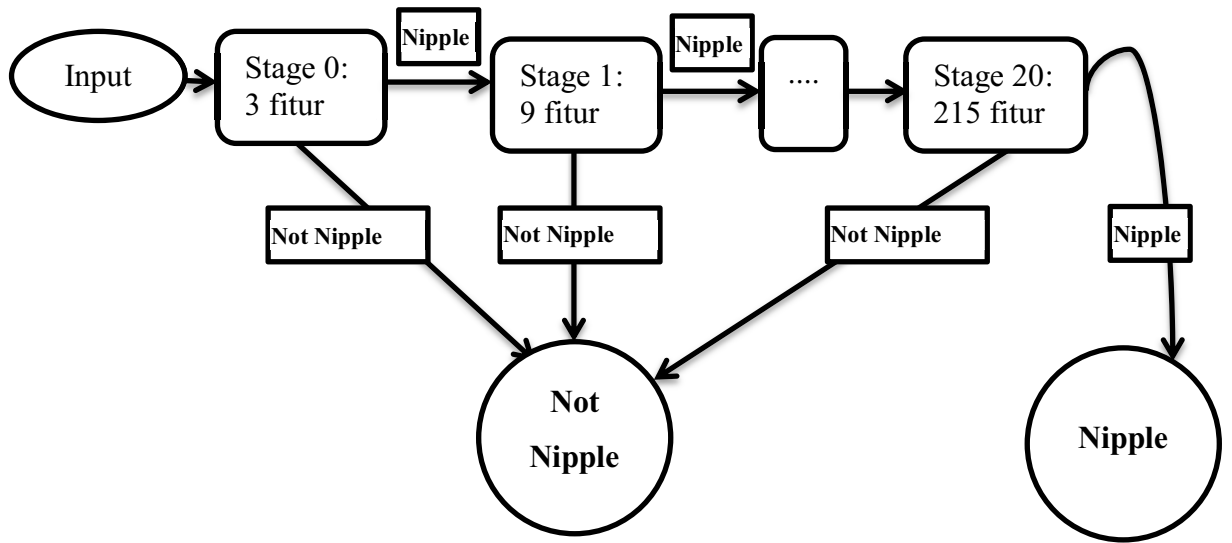
2. Terakhir classifier kuat adalah:

$$C(x) = \begin{cases} 1 & \sum_{t=1}^T \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^T \alpha_t \\ & \text{sebaliknya} \end{cases}$$

dimana  $\alpha_t = \log \frac{1}{\beta_t}$

3. *Cascade classifiers*

*Cascade classifier* berbentuk degenerasi pohon keputusan yang terdiri dari beberapa tahap (*stages*). Setiap tahapan dalam *cascade classifier* akan lebih kompleks dibanding dengan tahap sebelumnya. Setiap *subwindow* yang lolos pada tahap sebelumnya akan diproses, sedangkan yang tidak lolos akan langsung tertolak. Tahap-tahap pada *cascade* dibuat dengan melatih *classifiers* menggunakan AdaBoost.



**Gambar 19** *Nipple cascade classifiers*



Algoritma pelatihan yang digunakan untuk membangun *cascade classifier* adalah seperti berikut:

**Tabel 4** Algoritma untuk membangun *cascade classifier* [25]

1. Pengguna memilih nilai untuk  $f$ , yaitu tingkat maksimum positif palsu dapat diterima per tahap dan  $d$ , yaitu tingkat deteksi minimum yang dapat diterima per *layer*.
2. Pengguna memilih target keseluruhan dari tingkat *false positive*  $F_{\text{target}}$
3. Pos = menset nilai sample positif
4. Neg = menset nilai sample negatif
5.  $F_0 = 1.0, D_0 = 1.0, i = 0$
6.  $i = 0$
7. While  $F_i > F_{\text{target}}$ 
  - A.  $i \leftarrow i + 1$
  - B.  $n_i = 0; F_i = F_{i-1}$
  - C. While  $F_i > f \times F_{i-1}$ 
    - a.  $n_i \leftarrow n_i + 1$
    - b. Gunakan Pos dan Neg untuk melatih classifier dengan fitur  $n_i$  menggunakan adaboost.
    - c. Lakukan evaluasi terhadap classifier *cascade* saat ini untuk menentukan  $F_i, D_i$
    - d. Menurunkan *threshold* untuk classifier ke  $i$  sampai classifier saat ini memiliki tingkat deteksi setidaknya  $d \times D_{i-1}$
  - D.  $N \leftarrow \varphi$
  - E. Jika  $F_{i+1} > F_{\text{target}}$  maka lakukan evaluasi pada *cascade detector* saat ini menjadi bukan puting payudara dan menempatkan setiap pendeteksian palsu ke dalam Neg.

#### 4. Multi Deteksi

Metode yang diusulkan pada penelitian ini adalah multi deteksi algoritma Viola and Jones. Deteksi pertama untuk mendapatkan area payudara. Hal ini bertujuan untuk mengurangi positif palsu terhadap deteksi puting payudara dengan cara mendapatkan kandidat area pasti yang akan dideteksi tahap selanjutnya. Setelah didapatkan area payudara, maka dilanjutkan tahap deteksi yang kedua untuk mendapatkan area puting payudara. Model deteksi tahap kedua sama persis dengan tahap pertama. Sama-sama menggunakan algoritma Viola and Jones berbasis algoritma AdaBoost.

### 3.5 Eksperimen dan Pengujian Model

Eksperimen dibagi ke dalam dua tahap yaitu tahap pelatihan (*training*) dan tahap pengujian (*testing*). Untuk melakukan pelatihan terhadap *training set* digunakan *tool* haartraining.exe yang sudah dipaket dalam software yang digunakan sebelumnya. Hasil akhir dari proses pelatihan adalah *cascade detector* berupa file *Extensible Markup Language* (XML). Setelah dilakukan pelatihan dilanjutkan tahap berikutnya yaitu pengujian. Pengujian dilakukan untuk mengetahui seberapa besar tingkat akurasi dari model yang diusulkan.

#### 3.5.1 Tahap Pelatihan Data dengan Haartraining.exe

Menurut Bradski et al. [29] agar *classifier* dapat bekerja dengan baik diperlukan set sampel positif antara 1.000-10.000 citra dan menghilangkan semua *variance* yang tidak diperlukan oleh data. Oleh karena itu, jumlah sampel positif yang digunakan dalam penelitian ini sebanyak 1181 citra yang sudah dikonversi ke dalam file vector dan sampel negatif sebanyak 1756 citra.

Proses pelatihan *Haar Training* membutuhkan beberapa parameter sebagai berikut:

1. **-data** : digunakan untuk mengatur lokasi penyimpanan hasil *training cascade classifier*.
2. **-vec** : digunakan untuk mengatur lokasi file *vector* (sampel positif).
3. **-bg** : digunakan untuk mengatur lokasi file yang menyimpan informasi sampel negatif.
4. **-npos** : digunakan untuk menentukan jumlah citra sampel positif. Nilai defaultnya 2000.
5. **-nneg** : digunakan untuk menentukan jumlah citra sampel negatif. Nilai defaultnya 2000.
6. **-nstages** : digunakan untuk menentukan jumlah *stage classifier*. Nilai defaultnya 14.
7. **-nsplits** : digunakan untuk menentukan jumlah *classifier* lemah yang digunakan. Jika nilainya 1, maka *classifier* sederhana yang digunakan. Jika nilainya  $\geq 2$  yang digunakan adalah CART classifier dengan jumlah *split internal nodes*. Nilai defaultnya 1.

8. **-mem** : digunakan untuk menentukan memori yang digunakan prekalkulasi. Nilai defaultnya 2000.
9. **-sym** dan **-nonsym** : digunakan untuk menentukan apakah class objek yang digunakan untuk training memiliki *vertical symetry* atau tidak. Nilai defaultnya -sym.
10. **-minihtrate** : minimal hit rate yang diinginkan untuk setiap stage classifier. Nilai defaultnya 0.950000.
11. **-maxfalsealarm** : maksimal *false alarm* yang diinginkan di setiap *stage classifier*. Nilai defaultnya 0.500000.
12. **-mode {BASIC|CORE|ALL}** : digunakan untuk memilih tipe *haar features* yang akan digunakan untuk *training*. Nilai defaultnya BASIC.
13. **-bt {DAB|RAB|LB|GAB}** : digunakan untuk memilih tipe algoritma AdaBoost. Discrete AdaBoost (DAB), Real AdaBoost (RAB), LogitBoost (LB) and Gentle AdaBoost (GAB). Nilai defaultnya Gentle AdaBoost (GAB).
14. **-minpos** : jumlah minimum sampel positif per cluster.
15. **-h** : digunakan untuk menentukan tinggi sampel. Default 24.
16. **-w** : digunakan untuk menentukan lebar sampel. Default 24.

Nilai parameter yang digunakan pada penelitian ini adalah:

```
E:\OpenCV Ugo Rampe\nipple training\temp\haartraining.exe
-data data/cascade -vec data/nipple.vec -bg
negative/infofile.txt -npos 1181 -nneg 1756 -nstages 27 -
mem 1000 -mode ALL -w 20 -h 20 -nonsym
```

Dari proses pelatihan dihasilkan *cascade classifiers* yang tersimpan di folder data/cascade. Untuk jumlah *stage classifiers* dipilih nilai akhir 27 *stages* setelah melalui percobaan *trial and error*. Semakin tinggi *stage classifiers* semakin sedikit jumlah positif palsu dan semakin rendah pula tingkat deteksi. Oleh karena itu, jumlah *stages* sebanyak 27 dikarenakan dianggap paling ideal untuk kasus deteksi puting payudara.

Hasil pelatihan berupa *cascade detector* dikonversi menjadi file .XML dengan *tool* haarconv.exe. XML yang dihasilkan diberi nama cascade\_nipple.xml.

### 3.5.2 Tahap pengujian

Tahap pengujian meliputi dua kali pengujian yaitu pengujian kinerja metode *single detector* dan *multi detector*. Parameter pengukurannya adalah akurasi, *precision rate* dan *recall rate*. Masing-masing metode akan dihitung tingkat akurasi, *precision rate* dan *recall rate* dari total keseluruhan citra pengujian. Kemudian dibandingkan manakah yang memiliki kinerja deteksi lebih baik dalam mendeteksi puting payudara.

Untuk melihat hasil deteksi masing-masing metode dibutuhkan program *prototype* untuk pengujian. Program *prototype* ini dibuat dengan bahasa pemrograman C++ dengan kompiler Code Blocks 12.11 dan *tool computer vision* OpenCV 2.1. Program ini merupakan implementasi dari tahap *online recognizing* metode *single detector* dan *multi detector*. Program mampu mendeteksi objek dengan menjalankan implementasi *cascade classifiers* yang sudah dikonversi dalam bentuk .XML dengan nama *cascadebreast1981.xml* dan *cascade\_nipple.xml*. *Cascadebreast1981.xml* sebagai bentuk implementasi dari *cascade classifiers* untuk deteksi payudara sedangkan *cascade\_nipple.xml* sebagai bentuk implementasi dari *cascade classifiers* untuk deteksi puting payudara.

Parameter kinerja yang diukur pada eksperimen ini adalah:

1. Akurasi

Akurasi merupakan rasio jumlah area puting payudara yang benar terdeteksi sebagai puting (*true positive*) dan area bukan puting payudara yang benar terdeteksi bukan puting payudara (*true negative*) terhadap keseluruhan objek dalam citra yang dideteksi.

Rumus perhitungannya:

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

Keterangan:

- TP (*true positive*) = jumlah area puting payudara yang benar terdeteksi sebagai puting
- TN (*true negative*) = jumlah area bukan puting payudara yang benar terdeteksi bukan puting
- FP (*false positive*) = jumlah area bukan puting payudara yang terdeteksi sebagai puting
- FN (*false negative*) = jumlah area puting payudara yang terdeteksi sebagai bukan puting

## 2. Precision Rate

*Precision rate* adalah rasio jumlah area puting payudara yang benar terdeteksi sebagai puting (*true positive*) terhadap jumlah area puting payudara yang benar terdeteksi sebagai puting (*true positive*) dan jumlah area bukan puting payudara yang salah dideteksi sebagai puting (*false positive*).

Rumus perhitungannya:

$$Precision\ rate = \frac{TP}{TP + FP} \quad (8)$$

Keterangan:

- TP (*true positive*) = jumlah area puting payudara yang benar terdeteksi sebagai puting
- TN (*true negative*) = jumlah area bukan puting payudara yang benar terdeteksi bukan puting
- FP (*false positive*) = jumlah area bukan puting payudara yang terdeteksi sebagai puting
- FN (*false negative*) = jumlah area puting payudara yang terdeteksi sebagai bukan puting

## 3. Recall Rate

*Recall rate* adalah rasio jumlah area puting payudara yang benar terdeteksi sebagai puting (*true positive*) terhadap jumlah area puting payudara yang benar terdeteksi sebagai puting (*true positive*) dan jumlah area puting payudara yang salah dideteksi sebagai bukan puting payudara (*false negative*).

Rumus perhitungannya:

$$\text{Recall rate} = \frac{TP}{TP + FN} \quad (9)$$

Keterangan:

- TP (*true positive*) = jumlah area puting payudara yang benar terdeteksi sebagai puting
- TN (*true negative*) = jumlah area bukan puting payudara yang benar terdeteksi bukan puting
- FP (*false positive*) = jumlah area bukan puting payudara yang terdeteksi sebagai puting
- FN (*false negative*) = jumlah area puting payudara yang terdeteksi sebagai bukan puting

### 3.6 Evaluasi dan Validasi Hasil

Berdasarkan parameter-parameter *true positive*, *true negative*, *false positive* dan *false negative*, maka dapat dihitung akurasi (*accuracy*), *precision* dan *recall rate* dari kedua model. Dari hasil komparasi kedua model tersebut dapat dilihat model manakah yang lebih unggul dalam mendeteksi puting payudara pada citra pornografi.

Salah satu hal yang sangat mempengaruhi ketepatan pendeteksian adalah penentuan *stages classifiers*. Pada penelitian ini ditentukan jumlah *stages* 27 setelah melalui tahapan uji coba *trial and error*. Selain pengaruh nilai *stages classifier*, sebenarnya penentuan ukuran *subwindow* citra pelatihan sangat berpengaruh juga. Tapi, bersifat

*trial and error* juga. Oleh karena itu, pada penelitian ini dibatasi menggunakan *subwindow size* 20x20.

Setelah dilakukan pengujian terhadap keseluruhan citra pengujian, dapat dilihat berapa banyak jumlah *false positive* dan *false negative* pada masing-masing citra dengan kedua metode yang berbeda. Penjabaran lebih lanjut akan diuraikan di Bab IV.