Cepat Obyek Deteksi menggunakan Didorong Cascade Sederhana

fitur

Paul Viola

viola@merl.com

Mitsubishi Electric Research Labs

201 Broadway, 8 FL

Cambridge, MA 02139

Abstrak

*Makalah ini menjelaskan pendekatan pembelajaran mesin untuk deteksi objek visual yang mampu memproses gambar yang sangat pesat dan mencapai tingkat deteksi yang tinggi. pekerjaan ini dibedakan oleh tiga kontribusi utama. Yang pertama adalah pengenalan representasi citra baru yang disebut “Integral Image” yang memungkinkan fitur yang digunakan oleh detektor dan dihitung dengan komputasi yang sangat cepat. Yang kedua adalah algoritma pembelajaran, berdasarkan AdaBoost, yang memilih sejumlah kecil fitur visual kritis dari satu set yang lebih besar dan hasil yang sangat e fi sien dikelompokkan ers [5]. Kontribusi ketiga adalah metode untuk menggabungkan ers dikelompokkan semakin lebih kompleks dalam “cascade” yang memungkinkan daerah latar belakang gambar yang akan dengan cepat dibuang sementara menghabiskan lebih perhitungan di daerah objek-seperti yang menjanjikan. cascade dapat dilihat sebagai objek spesifik mekanisme fokus-of-perhatian yang tidak seperti pendekatan sebelumnya memberikan jaminan statistik yang dibuang daerah tidak mungkin mengandung objek yang menarik. Dalam domain dari deteksi wajah sistem menghasilkan tingkat deteksi sebanding dengan yang terbaik sistem sebelumnya. Digunakan dalam aplikasi real-time, detektor berjalan pada 15 frame per detik tanpa menggunakan differencing gambar atau deteksi warna kulit.*

Michael Jones

michael.jones@compaq.com

Lab Compaq Cambridge Penelitian

Satu Cambridge Centre

Cambridge, MA 02142

frame rate yang tinggi. Sistem kami mencapai frame rate yang tinggi bekerja hanya dengan informasi yang ada dalam gambar skala abu-abu tunggal. Sumber-sumber informasi alternatif juga dapat diintegrasikan dengan sistem kami untuk mencapai frame rate yang lebih tinggi.

Ada tiga kontribusi utama kerangka deteksi obyek kita. Kami akan memperkenalkan masing-masing gagasan secara singkat di bawah ini dan kemudian menjelaskan secara rinci dalam bagian berikutnya.

Kontribusi pertama dari makalah ini adalah representasi gambar baru disebut *image terpisahkan* yang memungkinkan untuk evaluasi fitur yang sangat cepat. Sebagian didorong oleh karya Papageorgiou et al. sistem deteksi kami tidak bekerja secara langsung dengan intensitas citra [9]. Seperti penulis ini kita menggunakan satu set fitur yang mengingatkan fungsi Dasar Haar (meskipun kami juga akan menggunakan lters fi terkait yang lebih kompleks daripada Haar lters fi). Dalam rangka untuk menghitung fitur ini sangat pesat di banyak sisik kami memperkenalkan citra representasi terpisahkan untuk gambar. Gambar terpisahkan dapat dihitung dari gambar menggunakan beberapa operasi per pixel. Setelah dihitung, salah satu dari fitur Harr-seperti dapat dihitung pada setiap skala atau lokasi di *konstan* waktu.

1. Perkenalan

Makalah ini menyatukan algoritma baru dan wawasan untuk membangun kerangka kerja untuk deteksi objek yang kuat dan sangat cepat. Kerangka ini menunjukkan pada, dan sebagian dimotivasi oleh, tugas deteksi wajah. Untuk itu kami telah membangun sistem deteksi wajah frontal yang mencapai deteksi dan palsu tingkat positif yang setara dengan hasil terbaik yang diterbitkan [14, 11, 13, 10, 1]. Sistem deteksi wajah ini paling jelas dibedakan dari pendekatan sebelumnya dalam kemampuannya untuk mendeteksi wajah sangat pesat. Beroperasi pada 384 oleh 288 piksel gambar, wajah yang terdeteksi pada 15 frame per detik pada 700 MHz Intel Pentium III konvensional. Dalam sistem deteksi wajah lainnya, informasi tambahan, seperti perbedaan gambar dalam urutan video, atau warna pixel pada citra berwarna, telah digunakan untuk mencapai

Kontribusi kedua dari makalah ini adalah metode untuk membangun fi er klasi dengan memilih sejumlah kecil fitur penting menggunakan AdaBoost [5]. Dalam gambar apapun subwindow jumlah total fitur Harr-seperti sangat besar, jauh lebih besar dari jumlah piksel. Dalam rangka untuk memastikan cepat klasifikasi, proses pembelajaran harus mengecualikan sebagian besar fitur yang tersedia, dan fokus pada satu set kecil fitur penting. Termotivasi oleh karya Tieu dan Viola, seleksi fitur dicapai melalui modi fi kasi sederhana dari prosedur AdaBoost: pelajar lemah dibatasi sehingga setiap lemah dikelompokkan er kembali dapat bergantung pada hanya fitur tunggal [15]. Akibatnya setiap tahap dari proses boosting, yang memilih lemah dikelompokkan er baru, dapat dilihat sebagai proses seleksi fitur.

Kontribusi utama ketiga dari makalah ini adalah metode untuk menggabungkan berturut-turut ers fi klasifikasi yang lebih kompleks dalam struktur kaskade yang secara dramatis meningkatkan kecepatan

ISBN 0-7695-1272-0 / 01 $ 10,00 (C) 2001 IEEE

detektor dengan memfokuskan perhatian pada daerah yang menjanjikan gambar. Gagasan di balik fokus pendekatan perhatian adalah bahwa hal itu sering mungkin untuk cepat menentukan di mana dalam foto obyek mungkin terjadi [16, 7, 1].

Lebih pengolahan kompleks hanya diperuntukkan bagi daerah-daerah yang menjanjikan. Ukuran kunci dari pendekatan semacam itu adalah “false negative” tingkat proses attentional. Ini harus menjadi kasus bahwa semua atau hampir semua, contoh objek yang dipilih oleh attentional fi lter.

Kami akan menjelaskan proses untuk melatih sebuah fi sien dikelompokkan er sangat sederhana dan ef yang dapat digunakan sebagai “diawasi” Fokus operator perhatian. Istilah diawasi mengacu pada fakta bahwa operator attentional dilatih untuk mendeteksi contoh dari kelas tertentu. Dalam domain dari deteksi wajah adalah mungkin untuk mencapai kurang dari 1% negatif palsu dan 40% positif palsu menggunakan fi er klasifikasi dibangun dari dua fitur Harr-seperti. Efek dari fi lter ini adalah untuk mengurangi oleh lebih dari satu setengah jumlah lokasi di mana fi detektor nal harus dievaluasi.

Mereka sub-jendela yang tidak ditolak oleh awal dikelompokkan er diproses oleh urutan ers fi klasifikasi, masing-masing sedikit lebih kompleks daripada yang terakhir. Jika setiap fi klasi er menolak sub-jendela, tidak ada proses lebih lanjut dilakukan. Struktur proses deteksi mengalir pada dasarnya bahwa dari pohon keputusan merosot, dan karena itu terkait dengan pekerjaan Geman dan rekan [1, 3].

Sebuah detektor wajah yang sangat cepat akan memiliki aplikasi praktis yang luas. Ini termasuk antarmuka pengguna, database gambar, dan

telekonferensi. Dalam aplikasi di mana

cepat frame-suku yang tidak diperlukan, sistem kami akan memungkinkan untuk signifikan tambahan pasca-pengolahan dan analisis. Selain itu sistem kami dapat diimplementasikan pada berbagai perangkat daya rendah kecil, termasuk tangan-helds dan prosesor tertanam. Di lab kami, kami telah menerapkan ini detektor wajah pada genggam Compaq iPaq dan telah mencapai deteksi pada dua frame per detik (perangkat ini memiliki daya rendah 200 MIPS *Lengan kuat* prosesor yang tidak memiliki terapung titik hardware).

Sisa kertas menggambarkan kontribusi kami dan sejumlah hasil eksperimen, termasuk penjelasan rinci tentang metodologi eksperimental kami. Diskusi pekerjaan terkait erat berlangsung pada akhir setiap bagian.

2. Fitur

Kami deteksi obyek prosedur dikelompokkan es gambar berdasarkan nilai fitur sederhana. Ada banyak motivasi untuk menggunakan fitur daripada piksel langsung. Alasan yang paling umum adalah bahwa fitur dapat bertindak untuk mengkodekan ad-hoc pengetahuan domain yang sulit untuk belajar menggunakan kuantitas fi nite data pelatihan. Untuk sistem ini juga ada motivasi penting kedua untuk fitur: sistem fitur berbasis beroperasi jauh lebih cepat daripada sistem berbasis pixel.

fitur sederhana yang digunakan mengingatkan Haar fungsi dasar yang telah digunakan oleh Papageorgiou et al. [9].

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | B |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| SEBUAH | | |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

|  |  |
| --- | --- |
| C | D |

Gambar 1: Contoh persegi panjang fitur yang ditampilkan relatif terhadap jendela deteksi melampirkan. Jumlah dari piksel yang terletak di dalam persegi panjang putih dikurangkan dari jumlah piksel dalam persegi panjang abu-abu. fitur dua persegi panjang yang ditunjukkan dalam (A) dan (B). Gambar (C) menunjukkan fitur tiga persegi panjang, dan (D) fitur empat persegi panjang.

Lebih khusus lagi, kita menggunakan tiga jenis fitur. Nilai dari *Fitur dua persegi* *panjang* adalah perbedaan antara jumlah dari piksel dalam dua wilayah persegipanjang. Daerah memiliki ukuran dan bentuk yang sama dan secara horizontal atau vertikal yang berdekatan (lihat Gambar 1). SEBUAH *Fitur tiga persegi* *panjang*

menghitung jumlah dalam dua persegi panjang luar dikurangkan dari jumlah dalam persegi panjang pusat. akhirnya *Fitur empat persegi panjang* menghitung perbedaan antara pasangan diagonal dari persegi panjang.

Mengingat bahwa resolusi dasar dari detektor adalah 24x24, set fitur lengkap persegi panjang yang cukup besar, lebih

180.000. Perhatikan bahwa tidak seperti dasar Haar, set fitur persegi panjang adalah overcomplete 1.

2.1. Integral Gambar

Fitur persegi panjang dapat dihitung dengan sangat cepat menggunakan representasi perantara untuk gambar yang kita sebut gambar terpisahkan. 2 Gambar terpisahkan di lokasi berisi jumlah dari piksel di atas dan di sebelah kiri

inklusif:

dimana adalah gambar integral dan adalah origi- yang

image nal. Menggunakan pasangan berikut kekambuhan:

(1)

(2)

(dimana adalah baris jumlah kumulatif,

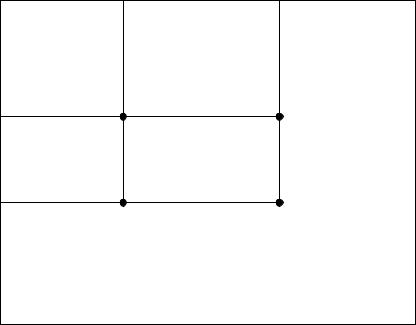
dan gambar terpisahkan dapat dihitung dalam

salah satu melewati gambar asli.

1 Sebuah dasar yang lengkap tidak memiliki ketergantungan linier antara elemen-elemen basis dan memiliki jumlah elemen yang sama sebagai ruang gambar, dalam hal ini 576. The set lengkap 180.000 ribu fitur ini berkali-kali selama-lengkap.

2 Ada hubungan dekat dengan “menyimpulkan tabel daerah” seperti yang digunakan dalam grafis [2]. Kami memilih nama yang berbeda di sini untuk menekankan penggunaannya untuk analisis gambar, bukan untuk pemetaan tekstur.

ISBN 0-7695-1272-0 / 01 $ 10,00 (C) 2001 IEEE



SEBUAH B

* 2
* D

|  |  |
| --- | --- |
| 3 | 4 |

Gambar 2: Jumlah piksel dalam persegi panjang dapat dihitung dengan empat referensi array yang. Nilai dari gambar yang tidak terpisahkan di lokasi 1 adalah jumlah piksel di persegi panjang

Nilai di lokasi 2 adalah di lokasi 3 adalah

dan di lokasi 4 adalah Jumlah dalam bisa

dihitung sebagai

Menggunakan gambar terpisahkan setiap jumlah persegi panjang dapat dihitung dalam empat referensi array (lihat Gambar 2). Jelas perbedaan antara dua jumlah persegi panjang dapat dihitung dalam delapan referensi. Karena fitur dua persegi panjang didefinisikan di atas melibatkan jumlah persegi panjang yang berdekatan mereka dapat dihitung dalam enam referensi array, delapan dalam kasus fitur tiga persegi panjang, dan sembilan untuk fitur empat persegi panjang.

2.2. Diskusi fitur

fitur persegi panjang yang agak primitif bila dibandingkan dengan alternatif seperti lters fi steerable [4, 6]. lters fi steerable, dan keluarga mereka, sangat baik untuk analisis rinci batas-batas, kompresi gambar, dan analisis tekstur. Dalam fitur persegi panjang Sebaliknya, sementara sensitif terhadap kehadiran tepi, bar, dan struktur gambar sederhana lainnya, cukup kasar. Tidak seperti lters fi steerable satu-satunya orientasi tersedia adalah vertikal, horisontal, dan diagonal. Set fitur persegi panjang yang namun memberikan citra representasi kaya yang mendukung pembelajaran yang efektif. Dalam hubungannya dengan gambar integral, efisiensi dari fitur persegi panjang set memberikan kompensasi yang cukup untuk fleksibilitas mereka terbatas.

dikelompokkan er mendekati nol secara eksponensial dalam jumlah putaran. Lebih penting lagi sejumlah hasil yang kemudian terbukti tentang kinerja generalisasi [12]. Patut diketahui bahwa kinerja generalisasi berhubungan dengan margin satu contoh, dan bahwa AdaBoost mencapai margin yang besar dengan cepat.

Ingat bahwa ada lebih dari 180.000 fitur persegi panjang yang terkait dengan setiap gambar sub-window, jumlah yang jauh lebih besar dari jumlah piksel. Meskipun setiap fitur dapat dihitung sangat e fi sien, komputasi set lengkap mahal. Hipotesis kami, yang ditanggung oleh percobaan, adalah bahwa jumlah yang sangat kecil dari fitur ini dapat dikombinasikan untuk membentuk efektif dikelompokkan er. Tantangan utama adalah untuk mendapati fitur ini.

Untuk mendukung tujuan ini, algoritma pembelajaran yang lemah dirancang untuk memilih fitur persegi panjang tunggal yang terbaik memisahkan contoh positif dan negatif (ini mirip dengan pendekatan [15] dalam domain pengambilan database gambar). Untuk masing-masing fitur, pelajar lemah menentukan ambang klasifikasi fungsi fi kasi optimal, sehingga jumlah minimum contoh adalah misclassi fi ed. Sebuah lemah dikelompokkan er

dengan demikian terdiri dari fitur , Ambang batas

dan polaritas menunjukkan arah ketidaksamaan

tanda:

jika

jika tidak

Sini adalah sub-window 24x24 pixel dari suatu gambar. Lihat Gambar 3 untuk ringkasan dari proses meningkatkan.

Dalam prakteknya ada fitur tunggal dapat melakukan klasifikasi tugas fi kasi dengan kesalahan rendah. Fitur yang dipilih dalam putaran awal proses meningkatkan memiliki tingkat kesalahan antara 0,1 dan 0,3. Fitur yang dipilih dalam putaran kemudian, sebagai tugas menjadi lebih sulit, yield tingkat kesalahan antara 0,4 dan 0,5.

3.1. Diskusi belajar

Banyak prosedur seleksi fitur umum telah diusulkan (lihat bab 8 dari [17] untuk review). Aplikasi nal fi kami menuntut pendekatan yang sangat agresif yang akan membuang sebagian besar fitur. Untuk pengakuan yang sama problemPapageorgiou et al. mengusulkan skema untuk seleksi fitur berdasarkan fitur varian [9]. Mereka menunjukkan hasil yang baik memilih 37 fitur dari total 1734 fitur.

3. Belajar Klasi Fungsi fi kasi

Mengingat satu set fitur dan satu set pelatihan gambar positif dan negatif, sejumlah pendekatan pembelajaran mesin dapat digunakan untuk belajar fungsi fi kasi klasifikasi. Dalam sistem kami varian dari AdaBoost digunakan *kedua* untuk memilih satu set kecil fitur *dan* melatih dikelompokkan er [5]. Dalam bentuk aslinya, yang AdaBoost belajar algoritma yang digunakan untuk meningkatkan kinerja fi kasi klasifikasi dari (kadang-kadang disebut lemah) algoritma pembelajaran sederhana. Ada sejumlah jaminan formal yang disediakan oleh prosedur pembelajaran AdaBoost. Freund dan Schapire membuktikan bahwa kesalahan pelatihan yang kuat

Roth et al. mengusulkan proses seleksi fitur berdasarkan menampi eksponensial perceptron aturan belajar [10]. TheWinnow proses belajar konvergen ke solusi di mana banyak dari bobot ini adalah nol. Namun demikian jumlah yang sangat besar dari fitur dipertahankan (mungkin beberapa ratus atau ribu).

3.2. Hasil belajar

Sementara rincian tentang pelatihan dan kinerja sistem fi nal disajikan dalam Bagian 5, beberapa hasil sederhana jasa diskusi. percobaan awal menunjukkan bahwa

ISBN 0-7695-1272-0 / 01 $ 10,00 (C) 2001 IEEE

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Mengingat contoh gambar ( |  |  |  |  | dimana |  |
| untuk contoh negatif dan positif masing-masing. | | | | | |  |
| inisialisasi bobot |  |  |  | untuk | dilakukan masing |  |
|  |  |  |

tively, di mana dan adalah jumlah negatif dan positif masing-masing.

Untuk

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1. | Menormalkan bobot, | |  |  |  |  |  |  |  |
|  | maka |  | |  |  |  |  |  |  |
|  | adalah distribusi probabilitas. | | | |  |  |  |  |
| 2. | Untuk setiap fitur, | | melatih dikelompokkan er | | |  | yang |  |  |
|  | dibatasi untuk menggunakan fitur tunggal. Kesalahan | | | | | | |  |  |
|  | dievaluasi sehubungan dengan | | | | | . | |  |  |
| 3. Pilih dikelompokkan er, | | |  | . dengan kesalahan terendah | | | | . |  |
| 4. | Update bobot: | |  |  |  |  |  |  |  |
|  | dimana | jika misalnya | | adalah diklasifikasikan cor- | | | |  |  |
|  | tidak tepat, | sebaliknya, dan | |  |  |  | . |  |  |
|  |  |  |  |  |  |

The fi nal kuat dikelompokkan er adalah:

jika tidak

dimana

Gambar 3: Algoritma AdaBoost untuk klasifikasi pembelajaran fi er. Setiap putaran meningkatkan menyeleksi satu fitur dari

180.000 fitur potensial.

frontal wajah dikelompokkan er dibangun dari 200 fitur menghasilkan tingkat deteksi 95% dengan tingkat positif palsu dari 1 di

14084. Hasil ini menarik, tapi tidak mencukupi untuk banyak tugas dunia nyata. Dalam hal perhitungan, ini dikelompokkan er mungkin lebih cepat daripada sistem yang diterbitkan lainnya, membutuhkan 0,7 detik untuk memindai gambar 384 oleh 288 piksel. Sayangnya, teknik yang paling mudah untuk meningkatkan kinerja deteksi, menambahkan fitur ke dikelompokkan er, langsung meningkatkan waktu komputasi.

Untuk tugas deteksi wajah, fitur persegi panjang awal dipilih oleh AdaBoost bermakna dan mudah ditafsirkan. Fitur pertama yang dipilih tampaknya fokus pada properti yang wilayah mata sering lebih gelap dari daerah hidung dan pipi (lihat Gambar 4). Fitur ini relatif besar dibandingkan dengan sub-jendela deteksi, dan harus agak sensitif terhadap ukuran dan lokasi dari wajah. Fitur kedua yang dipilih bergantung pada properti bahwa mata adalah lebih gelap dari jembatan hidung.

Gambar 4: Yang pertama dan fitur kedua yang dipilih oleh AdaBoost. Dua fitur yang ditampilkan di baris atas dan kemudian terhampar di wajah pelatihan yang khas di baris bawah. Yang pertama featuremeasures perbedaan intensitas antara wilayah mata dan daerah di pipi atas. Fitur yang mengkapitalisasi pada pengamatan bahwa wilayah mata sering gelap dari pipi. Fitur kedua membandingkan intensitas di daerah mata dengan intensitas melintasi jembatan dari hidung.



4. attentional Cascade

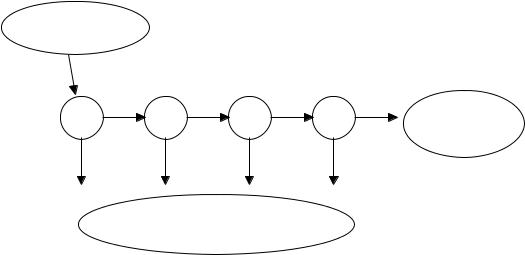
Bagian ini menjelaskan algoritma untuk membangun riam ers fi klasifikasi yang mencapai kinerja deteksi peningkatan sementara secara radikal mengurangi waktu komputasi. Patut diketahui bahwa lebih kecil, dan karena itu lebih efisien, meningkatkan klasifikasi ers fi dapat dibangun yang menolak banyak sub-jendela negatif sementara mendeteksi kasus hampir semua positif (yaitu ambang meningkatkan dikelompokkan er dapat disesuaikan sehingga tingkat negatif palsu adalah mendekati nol). Sederhana ers fi klasifikasi digunakan untuk menolak sebagian besar subwindows sebelum ers fi klasifikasi yang lebih kompleks dipanggil untuk mencapai tingkat positif palsu rendah.

Bentuk keseluruhan dari proses deteksi adalah bahwa dari pohon keputusan merosot, apa yang kita sebut “cascade” (lihat Gambar 5). Hasil positif dari pertama dikelompokkan er memicu evaluasi dari fi er klasifikasi kedua yang juga telah disesuaikan untuk mencapai tingkat deteksi yang sangat tinggi. Hasil positif dari kedua dikelompokkan er memicu fi klasifikasi ketiga er, dan sebagainya. Sebuah hasil negatif pada setiap titik mengarah ke penolakan langsung dari sub-jendela.

Tahapan dalam kaskade dibangun dengan melatih ers fi klasifikasi menggunakan AdaBoost dan kemudian menyesuaikan ambang untuk meminimalkan negatif palsu. Perhatikan bahwa default ambang AdaBoost dirancang untuk menghasilkan tingkat kesalahan yang rendah pada data pelatihan. Secara umum batas bawah menghasilkan tingkat deteksi yang lebih tinggi dan tingkat positif palsu tinggi.

Sebagai contoh sebuah pertama tahap dikelompokkan er baik dapat dibangun dari dua fitur klasifikasi yang kuat fi er dengan mengurangi ambang untuk meminimalkan negatif palsu. Diukur terhadap satu set pelatihan validasi, ambang batas dapat disesuaikan untuk mendeteksi 100% dari wajah dengan tingkat positif palsu dari 40%. Lihat

ISBN 0-7695-1272-0 / 01 $ 10,00 (C) 2001 IEEE



*Semua Sub-jendela*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| T | T | T |  | T |
| 1 | 2 | 3 | 4 | *Proses lebih* |
|  |  |  |  | *lanjut* |

F F F F

*Tolak Sub-jendela*

Gambar 5: Gambaran skematis dari kaskade deteksi. Serangkaian ers fi klasifikasi yang diterapkan untuk setiap sub-jendela. Awal dikelompokkan er menghilangkan sejumlah besar contoh negatif dengan sangat sedikit pengolahan. lapisan berikutnya menghilangkan negatif tambahan namun memerlukan perhitungan tambahan. Setelah beberapa tahapan pengolahan jumlah sub-jendela telah berkurang secara radikal. pengolahan lebih lanjut dapat mengambil bentuk apapun seperti tahap tambahan cascade (seperti dalam sistem deteksi kami) atau sistem deteksi alternatif.

Gambar 4 untuk penjelasan dari dua fitur yang digunakan dalam fi er klasifikasi.

Perhitungan dua fitur dikelompokkan er berjumlah sekitar 60 instruksi

mikroprosesor. Sepertinya sulit untuk membayangkan bahwa setiap sederhana fi lter bisa mencapai tingkat penolakan yang lebih tinggi. Sebagai perbandingan, memindai template gambar sederhana, atau lapisan perceptron tunggal, akan membutuhkan setidaknya 20 kali lebih banyak operasi per sub-jendela.

Struktur kaskade merefleksikan fakta bahwa dalam setiap gambar tunggal mayoritas dari subwindows negatif. Dengan demikian, kaskade mencoba untuk menolak sebanyak negatif mungkin pada tahap sedini mungkin. Sementara contoh positif akan memicu evaluasi setiap fi er klasi di kaskade, ini adalah sebuah acara sangat jarang.

Sama seperti pohon keputusan, selanjutnya ers fi klasi dilatih menggunakan contoh-contoh yang melewati semua tahapan sebelumnya. Akibatnya, kedua dikelompokkan er menghadapi tugas fi kultus lebih dif dari yang pertama. Contoh-contoh yang membuatnya melalui tahap pertama adalah “sulit” dari contoh yang khas. Contoh-contoh fi kultus lebih dif dihadapi oleh ers fi klasifikasi yang lebih dalam mendorong seluruh penerima operasi karakteristik (ROC) kurva ke bawah. Pada tingkat deteksi yang diberikan, lebih ers fi klasi memiliki tarif positif Sejalan tinggi palsu.

4.1. Pelatihan Cascade dari Klasi ers fi

Proses pelatihan berjenjang melibatkan dua jenis pengorbanan.

Dalam kebanyakan kasus klasifikasi ers fi dengan lebih banyak fitur akan mencapai tingkat deteksi lebih tinggi dan tingkat positif palsu lebih rendah. Pada saat yang sama dikelompokkan ers dengan lebih banyak fitur requiremore waktu untuk menghitung. Pada prinsipnya orang bisa mendefinisikan kerangka kerja optimasi di mana: i) jumlah tahap fi er klasifikasi,

1. jumlah fitur dalam setiap tahap, dan iii) ambang setiap tahap, diperdagangkan off untuk meminimalkan jumlah yang diharapkan dari fitur dievaluasi. Sayangnya fi nd-

ing optimal ini adalah masalah fi kultus sangat dif.

Dalam prakteknya kerangka kerja yang sangat sederhana digunakan untuk menghasilkan dikelompokkan er efektif yang sangat e fi sien. Setiap tahap dalam kaskade mengurangi tingkat positif palsu dan mengurangi tingkat deteksi. Target dipilih untuk pengurangan minimum di positif palsu dan penurunan maksimum dalam deteksi. Setiap tahap dilatih dengan menambahkan fitur sampai tingkat deteksi target dan positif palsu terpenuhi (tarif ini ditentukan dengan menguji detektor pada set validasi). Tahapan ditambahkan sampai target keseluruhan untuk positif palsu dan deteksi tingkat terpenuhi.

4.2. Detektor Cascade Diskusi

Kaskade deteksi wajah lengkap memiliki 38 tahap dengan lebih dari 6000 fitur. Namun demikian hasil struktur kaskade di kali deteksi rata-rata cepat. Pada dataset fi kultus dif, mengandung 507 wajah dan 75 juta sub-jendela, wajah dideteksi menggunakan rata-rata 10 evaluasi fitur per subwindow.

Sebagai perbandingan, sistem ini adalah sekitar 15 kali lebih cepat dari sebuah implementasi dari sistem deteksi dibangun oleh Rowley et al. 3 [ 11]

Sebuah gagasan yang mirip dengan cascade muncul dalam sistem deteksi wajah dijelaskan oleh Rowley et al. di mana dua jaringan deteksi digunakan [11]. Rowley et al. menggunakan jaringan yang lebih cepat namun kurang akurat untuk prescreen gambar untuk fi nd daerah calon jaringan lebih lambat lebih akurat. Meskipun sulit untuk menentukan dengan tepat, tampak bahwa Rowley et al. Dua sistem jaringan wajah adalah detektor wajah yang ada tercepat. 4

Struktur proses deteksi mengalir pada dasarnya bahwa dari pohon keputusan merosot, dan karena itu terkait dengan karya Amit dan Geman [1]. Tidak seperti teknik yang menggunakan detektor yang tetap, Amit dan Geman mengusulkan titik pandang alternatif mana yang tidak biasa co-kejadian fitur gambar sederhana yang digunakan untuk memicu evaluasi dari proses deteksi yang lebih kompleks. Dengan cara ini proses deteksi penuh tidak perlu dievaluasi di banyak lokasi gambar potensial dan timbangan badan. Sementara wawasan dasar ini sangat berharga, dalam pelaksanaannya perlu untuk terlebih dahulu mengevaluasi beberapa detektor fitur di setiap lokasi. Fitur-fitur ini kemudian dikelompokkan ke fi nd biasa co-kejadian. Dalam prakteknya, karena bentuk detektor dan fitur yang menggunakan sangat yang efisien, yang diamortisasi biaya mengevaluasi detektor kami di *setiap skala* *dan lokasi* jauh lebih cepat daripada perintisan dan pengelompokan tepiseluruh gambar.

Dalam karya terbaru Fleuret dan Geman telah disajikan teknik deteksi wajah yang bergantung pada “rantai” dari tes dalam rangka untuk menandakan kehadiran wajah di skala tertentu dan

3 Henry Rowley sangat anggun disediakan kami dengan implementasi sistem deteksi nya untuk perbandingan langsung. hasil yang dilaporkan menentang sistemnya tercepat. Hal ini sulit untuk menentukan dari literatur yang diterbitkan, tetapi detektor Rowley-Baluja-Kanade secara luas dianggap sebagai sistem deteksi tercepat dan telah banyak diuji pada masalah di dunia nyata.

4 detektor diterbitkan lain telah baik diabaikan untuk membahas kinerja dalam detail, atau tidak pernah diterbitkan deteksi dan tingkat positif palsu pada set fi pelatihan kultus besar dan dif.

ISBN 0-7695-1272-0 / 01 $ 10,00 (C) 2001 IEEE



Gambar 6: Contoh gambar wajah tegak frontal digunakan untuk pelatihan.

Lokasi [3]. Sifat gambar diukur dengan Fleuret dan Geman, disjunctions dari tepi skala fi ne, cukup berbeda dari fitur persegi panjang yang sederhana, ada di semua skala, dan agak ditafsirkan. Dua pendekatan juga berbeda secara radikal dalam filsafat pembelajaran mereka. Motivasi untuk proses pembelajaran Fleuret dan Geman adalah estimasi kepadatan dan kepadatan diskriminasi, sedangkan detektor kami adalah murni diskriminatif. Akhirnya tingkat positif palsu pendekatan Fleuret dan Geman ini tampaknya lebih tinggi dari pendekatan sebelumnya seperti Rowley et al. dan pendekatan ini. Sayangnya kertas tidak melaporkan hasil kuantitatif semacam ini. Contoh termasuk setiap gambar memiliki antara 2 dan 10 positif palsu.

5 Hasil

Sebuah 38 lapisan mengalir dikelompokkan er dilatih untuk mendeteksi wajah tegak frontal. Untuk melatih detektor, satu set wajah dan gambar pelatihan nonface digunakan. Training set wajah terdiri dari 4916 tangan berlabel wajah skala dan disesuaikan dengan resolusi dasar 24 oleh 24 piksel. Wajah-wajah yang diambil dari gambar-download saat merangkak acak world wide web. Beberapa contoh wajah yang khas ditunjukkan pada Gambar 6. subwindows non-wajah yang digunakan untuk melatih detektor berasal dari 9544 gambar yang manual diperiksa dan ditemukan tidak berisi permukaan. Ada sekitar 350 juta subwindows dalam gambar non-wajah tersebut.

Jumlah fitur dalam pertama lima lapisan detektor adalah 1, 10, 25, 25 dan 50 fitur masing-masing. Lapisan yang tersisa memiliki semakin lebih banyak fitur. Jumlah fitur di semua lapisan adalah 6061.

Setiap fi er klasi di kaskade dilatih dengan wajah pelatihan 4916 (ditambah gambar cermin vertikal mereka untuk total

dari 9832 wajah pelatihan) dan 10.000 non-face sub-jendela (juga dari ukuran 24 dengan 24 piksel) menggunakan prosedur pelatihan AdaBoost. Untuk awal satu fitur dikelompokkan er, contoh pelatihan nonface dikumpulkan dengan memilih acak sub-jendela dari satu set 9544 gambar yang tidak mengandung wajah. contoh non-wajah yang digunakan untuk melatih lapisan berikutnya diperoleh dengan memindai kaskade parsial seluruh gambar non-wajah dan mengumpulkan positif palsu. Maksimal 10.000 non-face sub-jendela tersebut dikumpulkan untuk setiap lapisan.

Kecepatan Detector Akhir

Kecepatan detektor mengalir secara langsung berhubungan dengan jumlah fitur dievaluasi per scan sub-jendela. Dievaluasi MIT + CMU test set [11], rata-rata 10 fitur dari total 6061 dievaluasi per sub-jendela. Hal ini dimungkinkan karena sebagian besar sub-jendela ditolak oleh pertama atau lapisan kedua di kaskade. Pada prosesor 700 Mhz Pentium III, detektor wajah dapat memproses gambar 384 dengan 288 pixel di sekitar 0,067 detik (menggunakan skala mulai dari 1,25 dan ukuran langkah 1,5 dijelaskan di bawah). Ini kira-kira 15 kali lebih cepat daripada detektor RowleyBaluja-Kanade [11] dan sekitar 600 kali lebih cepat daripada detektor Schneiderman-Kanade [13].

Pengolahan citra

Semua contoh sub-jendela yang digunakan untuk pelatihan adalah varian dinormalisasi untuk meminimalkan efek dari kondisi pencahayaan yang berbeda. Normalisasi itu perlu selama deteksi juga. Varians dari sebuah sub-jendela gambar dapat dihitung dengan cepat menggunakan sepasang gambar terpisahkan. ingat bahwa

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  | dimana | adalah standar |  |
| deviasi, |  |  |  |  |
| adalah mean, dan adalah nilai pixel dalam sub-jendela. Mean | | | |  |

dari sub-jendela dapat dihitung dengan menggunakan gambar terpisahkan. Jumlah pixel kuadrat dihitung dengan menggunakan gambar yang tidak terpisahkan dari gambar kuadrat (yaitu dua gambar terpisahkan digunakan dalam proses scanning). Selama pemindaian efek normalisasi gambar dapat dicapai melalui pos-mengalikan nilai fitur daripada pra-mengalikan piksel.

Memindai Detector

Detektor fi nal dipindai di gambar di berbagai skala dan lokasi. Scaling dicapai dengan skala detektor itu sendiri, bukan skala gambar. Proses ini masuk akal karena fitur dapat dievaluasi pada setiap skala dengan biaya yang sama. Hasil yang baik diperoleh dengan menggunakan satu set timbangan faktor 1,25 terpisah.

Detektor ini juga dipindai di lokasi. lokasi berikutnya diperoleh dengan menggeser jendela beberapa jumlah piksel

* Proses pergeseran ini dipengaruhi oleh skala detektor: jika skala saat ini jendela digeser oleh

di mana [] adalah operasi pembulatan. Pilihan

mempengaruhi baik kecepatan detektor sebagai

serta akurasi. Hasil kami menyajikan adalah untuk =

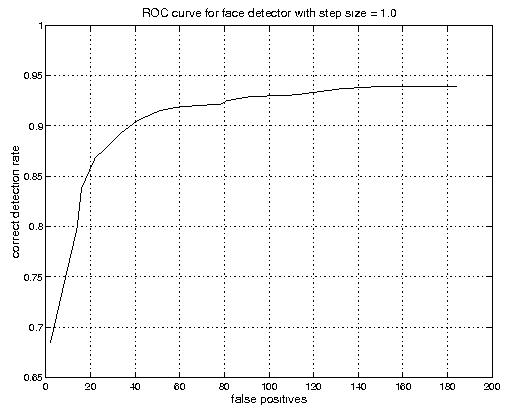
Kita bisa mencapai fi signifikan tidak bisa speedup dengan menetapkan =

dengan hanya sedikit penurunan akurasi.

ISBN 0-7695-1272-0 / 01 $ 10,00 (C) 2001 IEEE

Tabel 1: tarif Deteksi untuk berbagai jumlah positif palsu pada MIT + CMU test set yang berisi 130 gambar dan 507 wajah.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | deteksi palsu |  |  |  |  |  |  |  |  |  |  |
| Detektor |  | 10 | 31 | | 50 | 65 | 78 | 95 |  | 167 |  |
|  |  |  |  |
|  |  | |  |  | |  |  |  |  |  |  |
| Viola-Jones |  | 76,1% 88,4% | | 91,4% 92,0% 92,1% | |  |  | 92,9% 93,9% | |  |  |
| Viola-Jones (voting) |  | 81,1% 89,7% | | 92,1% 93,1% 93,1% | |  |  | 93,2% 93,7% | |  |  |
| Rowley-Baluja-Kanade |  | 83,2% 86,0% | | - |  | - | - | 89,2% 90,1% | |  |  |
| Schneiderman-Kanade |  | - | - |  | - | 94,4% - |  | - |  | - |  |
| Roth-Yang-Ahuja |  | - | - |  | - | - | (94,8%) | - |  | - |  |



Integrasi Beberapa Deteksi

Karena detektor fi nal tidak sensitif terhadap perubahan kecil dalam terjemahan dan skala, beberapa pendeteksian biasanya akan terjadi di sekitar wajah masing-masing dalam sebuah gambar yang dipindai. Hal yang sama sering berlaku dari beberapa jenis positif palsu. Dalam prakteknya sering masuk akal untuk kembali suatu fi nal deteksi per wajah. Menuju akhir ini sangat berguna untuk postprocess terdeteksi sub-jendela dalam rangka untuk menggabungkan deteksi tumpang tindih dalam deteksi tunggal.

Dalam eksperimen ini deteksi digabungkan dengan cara yang sangat sederhana. Himpunan deteksi yang pertama dipartisi menjadi himpunan bagian menguraikan. Dua deteksi berada di bagian yang sama jika daerah bounding mereka tumpang tindih. Setiap partisi menghasilkan deteksi fi nal tunggal. Sudut-sudut wilayah bounding fi nal adalah rata-rata dari sudut semua deteksi di set.

Percobaan pada Dunia Nyata Test Set

Kami menguji sistem kami pada MIT + CMU tes frontal wajah set [11]. set ini terdiri dari 130 gambar dengan 507 wajah frontal berlabel. AROC kurva menunjukkan kinerja detektor kami pada set tes ini ditunjukkan pada Gambar 7. Untuk membuat kurva ROC ambang fi nal lapisan dikelompokkan er disesuaikan dariuntuk + Menyesuaikan ambang untuk +

akan menghasilkan tingkat deteksi 0,0 dan tingkat positif palsu dari 0,0. Menyesuaikan ambang batas untuk Namun, meningkatkan baik tingkat deteksi dan tingkat positif palsu, tetapi hanya sampai titik tertentu. Tingkat tidak bisa lebih tinggi dari tingkat kaskade deteksi minus lapisan fi nal. Akibatnya, ambang batassetara dengan menghapus lapisan itu. Lebih meningkatkan deteksi dan tingkat positif palsu membutuhkan penurunan ambang berikutnya dikelompokkan er di kaskade. Dengan demikian, dalam rangka untuk membangun kurva ROC lengkap, klasifikasi lapisan fi er dihapus. Kami menggunakan *jumlah* positif palsu yang bertentangan dengan *menilai* positif palsu untuk x-sumbu kurva ROC untuk memfasilitasi perbandingan dengan sistem lain. Untuk menghitung tingkat positif palsu, hanya membagi dengan jumlah total sub-jendela dipindai. Dalam percobaan kami, jumlah sub-jendela dipindai adalah 75.081.800.

Sayangnya, sebagian besar hasil yang dipublikasikan sebelumnya pada deteksi wajah hanya termasuk rezim operasi tunggal (yaitu satu titik pada kurva ROC). Untuk membuat perbandingan dengan detektor kita lebih mudah, kami telah terdaftar tingkat deteksi kami untuk tingkat positif palsu dilaporkan oleh sistem lain. Tabel 1

Gambar 7: ROC kurva untuk detektor wajah kita di MIT + CMU test set. detektor dijalankan menggunakan ukuran langkah 1,0 dan mulai skala 1,0 (75.081.800 sub-jendela scan).

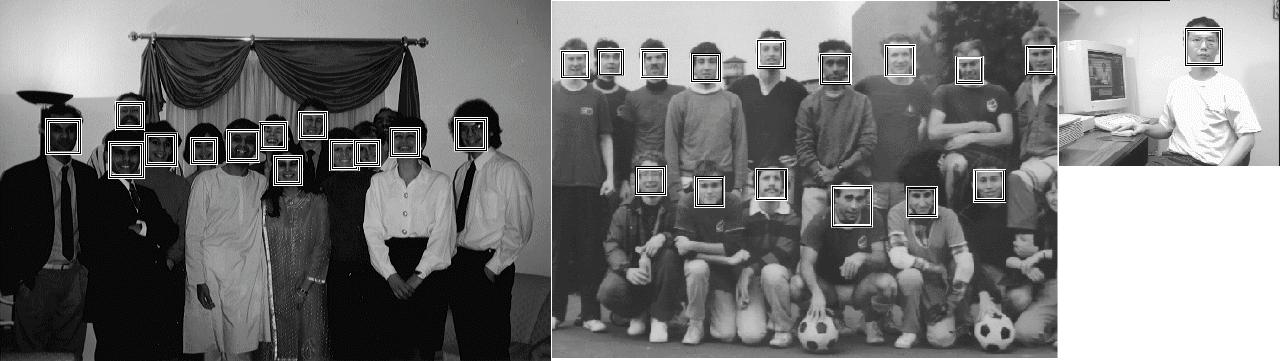
daftar tingkat deteksi untuk berbagai jumlah deteksi palsu untuk sistem kami serta sistem diterbitkan lainnya. Untuk hasil Rowley-Baluja-Kanade [11], sejumlah versi yang berbeda dari detektor mereka diuji menghasilkan sejumlah hasil yang berbeda mereka semua terdaftar di bawah judul yang sama. Untuk detektor Roth-Yang-Ahuja [10], mereka melaporkan hasil mereka pada tes MIT + CMU mengatur minus 5 gambar yang mengandung garis yang ditarik wajah dihapus.

Gambar 8 menunjukkan output dari detektor wajah kita pada beberapa citra uji dari MIT + CMU test set.

Sebuah skema voting sederhana untuk lebih meningkatkan hasil

Dalam tabel 1 kami juga menunjukkan hasil dari menjalankan tiga detektor (38 lapisan yang dijelaskan di atas ditambah dua detektor sama dilatih) dan keluaran suara mayoritas dari tiga detektor. Hal ini meningkatkan tingkat deteksi serta menghilangkan lebih banyak positif palsu. perbaikan akan lebih besar jika detektor yang lebih mandiri. Korelasi kesalahan mereka menghasilkan perbaikan moderat atas tunggal terbaik detektor.

ISBN 0-7695-1272-0 / 01 $ 10,00 (C) 2001 IEEE



Gambar 8: Output dari detektor wajah kita pada sejumlah citra uji dari MIT + CMU test set.

6 Kesimpulan

ters dan rotasi invarian. Di *Prosiding Konferensi IEEE pada Computer Vision* *dan Pengenalan Pola,* 1994.

Kami telah menyajikan pendekatan untuk deteksi obyek yang meminimalkan waktu komputasi sementara mencapai akurasi deteksi tinggi. Pendekatan ini digunakan untuk membangun deteksi wajah systemwhich adalah sekitar 15 kali lebih cepat daripada pendekatan sebelumnya.

Makalah ini menyatukan algoritma baru, representasi, dan wawasan yang cukup generik dan mungkin memiliki aplikasi yang lebih luas dalam visi komputer dan pengolahan citra.

Akhirnya makalah ini menyajikan satu set percobaan rinci pada dataset deteksi fi kultus wajah dif yang telah dipelajari secara luas. dataset ini termasuk wajah di bawah rentang yang sangat luas dari kondisi termasuk: pencahayaan, skala, pose, dan variasi kamera. Percobaan pada dataset yang besar dan kompleks seperti yang sulit dan memakan waktu. Namun demikian sistem yang bekerja di bawah kondisi ini tidak mungkin menjadi rapuh atau terbatas pada satu set kondisi. Lebih penting kesimpulan yang diambil dari dataset ini tidak mungkin artefak eksperimental.

Referensi

1. Y. Amit, D. Geman, dan K. Wilder. induksi bersama bentuk fitur dan klasifikasi pohon ers fi 1997.

[2] F. Crow. tabel dijumlahkan-daerah untuk pemetaan tekstur. Di *Prosiding SIGGRAPH,* Volume 18 (3), halaman 207-212,

1984.

[7] L. Itti, C. Koch, dan E. Niebur. Sebuah model saliency berbasis

perhatian visual untuk analisis adegan cepat. *IEEE Patt. Anal. Mach. Intell.,* 20 (11): 1254-1259, November 1998. [8] Edgar Osuna, Robert Freund, dan Federico Girosi. Latihan

mendukung mesin vektor: aplikasi untuk face detection. Di *Prosiding* *Konferensi IEEE pada Computer Vision dan Pengenalan Pola,* 1997.

[9] C. Papageorgiou, M. Oren, dan T. Poggio. A frame- umum

bekerja untuk deteksi obyek. Di *Konferensi Internasional tentang Computer* *Vision,* 1998.

1. D. Roth, M. Yang, dan N. Ahuja. Sebuah detektor wajah snowbased.

Di *Informasi Neural Pengolahan 12,* 2000.

1. H. Rowley, S. Baluja, dan T. Kanade. Neural jaringan berbasis

deteksi wajah. Di *IEEE Patt. Anal. Mach. Intell.,* Volume 20, halaman 22-38, 1998.

[12] Robert E. Schapire, Yoav Freund, Peter Bartlett, dan

Wee Sun Lee. Meningkatkan margin: Sebuah penjelasan baru untuk efektivitas metode pemungutan suara. Di *Prosiding Konferensi Internasional* *keempatbelas pada Machine Learning,*

1997.

[13] H. Schneiderman dan T. Kanade. Sebuah metode statistik untuk 3D

deteksi obyek diterapkan untuk wajah dan mobil. Di *Konferensi Internasional* *tentang Computer Vision,* 2000.

[14] K. Sung dan T. Poggio. berbasis contoh pembelajaran untuk view-

deteksi wajah berdasarkan. Di *IEEE Patt. Anal. Mach. Intell.,* Volume 20, halaman

39-51, 1998. [15] K. Tieu dan P. Viola. Meningkatkan pengambilan gambar. Di *Memproses-*

1. F. Fleuret dan D. Geman. Kasar-to-fi ne deteksi wajah. *Int.* *J. Computer Vision,* 2001.
2. William T. Freeman dan Edward H. Adelson. Desain

dan menggunakan lters fi steerable. *Transaksi IEEE pada Pola Analisis dan* *Mesin Intelijen,* 13 (9): 891-906, 1991. [5] Yoav Freund dan Robert E. Schapire.Sebuah keputusan-teori

generalisasi dari on-line belajar dan aplikasi untuk meningkatkan. Di *Komputasi* *Teori Belajar: Eurocolt '95,*

halaman 23-37. Springer-Verlag, 1995.

1. H. Greenspan, S. Belongie, R. Gooodman, P. Perona, S. Rak-kotoran, dan C. Anderson. Overcomplete dapat dikendalikan piramida fi l-

*ings Konferensi IEEE pada Computer Vision dan Pengenalan Pola,* 2000.

[16] JK Tsotsos, SM Culhane, Wyk Wai, YH Lai, N. Davis,

dan F. Nu fl o. Pemodelan visual perhatian melalui tala selektif. *Arti fi cial* *Intelligence Journal,* 78 (1-2): 507-545, Oktober 1995. [17] Andrew Webb. *Pola statistik Pengakuan.* Oxford Uni-

hayati Press, New York, 1999.

ISBN 0-7695-1272-0 / 01 $ 10,00 (C) 2001 IEEE