
Viola & Jones 얼굴 검출 알고리즘의 성능 분석

오정수 · 허훈

부경대학교

Performance Analysis of Viola & Jones Face Detection Algorithm

Hoon Heo · Jeong-su Oh

Department of Display Engineering, Pukyong National University

E-mail : ojs@pknu.ac.kr

요 약

Viola와 Jones의 객체 검출 알고리즘은 대표적인 얼굴 검출 알고리즘이다. 알고리즘은 얼굴 표현을 위해 하르-유사 특징들을 사용하고 분류를 위해 약분류기들의 선형 조합인 강분류기들로 구성된 cascade-Adaboost 알고리즘을 사용하고 있다. 이 알고리즘은 구현을 위해 몇 개의 변수 설정을 요구하고 설정된 값들이 알고리즘 성능에 영향을 준다. 본 논문은 알고리즘에 설정되는 변수에 따른 얼굴 검출 성능을 분석한다.

ABSTRACT

Viola and Jones object detection algorithm is a representative face detection algorithm. The algorithm uses Haar-like features for face expression and uses a cascade-Adaboost algorithm consisting of strong classifiers, a linear combination of weak classifiers for classification. This algorithm requires several parameter settings for its implementation and the set values affect its performance. This paper analyzes face detection performance according to the parameters set in the algorithm.

키워드

face detection, object detection, Haar-like feature, cascade-Adaboost classification

I. 서 론

영상에서 얼굴 영역을 검출하는 얼굴 검출은 얼굴인식 시스템, 출입통제 시스템, IT 융합 시스템 등에서 핵심적으로 활용되는 기술이다. 현재 얼굴 검출을 위해 사용되는 대표적인 알고리즘은 비올라(Viola)와 존스(Jones)이 발표한 객체 검출 알고리즘(V&J object detection algorithm)이다. V&J 객체 검출 알고리즘은 얼굴을 포함한 눈, 코, 입 등의 객체들을 선택적으로 검출할 수 있으나 본 논문에서는 얼굴 영역을 대상 객체로 하고 있다. 얼굴 검출을 위한 특징으로 주로 하르-유사 특징(Haar-like feature)들을 사용하고, 얼굴과 비얼굴을 구분하기 위해 단일 특징을 이용한 유효 약분류기(weak classifier)들의 선형 조합으로 강분류기(strong classifier)를 구성하는 AdaBoost 알고리즘을 사용하고 있다. 얼굴 표현을 위한 하르-유사 특징들은 원형 특징을 근거로 수 많은 특징들로 확장되어 다양한 인종, 얼굴 크기나 위치에 관계없이 얼굴 모양을 잘 표현해 주고 있다. V&J

객체 검출 알고리즘은 매우 우수한 성능을 보이거나 설정하는 변수들에 의해 다른 성능을 보인다. 그래서 본 논문은 변수에 따른 V&J 객체 검출 알고리즘의 얼굴 검출 성능을 비교 분석한다.

II. V&J 객체 추적 알고리즘

V&J 객체 검출 알고리즘에서 얼굴 검출을 위한 하르-유사 특징들은 하르-유사 특징 원형들을 토대로 검색창(sub-window)에 포함될 수 있는 가능한 모든 특징들로 확대되어 구성된다. V&J 객체 검출 알고리즘은 그림 1과 같은 4개의 하르-유사 특징 원형들만 사용하였으나 이후 다양한 하르-유사 특징 원형들이 추가된 알고리즘들이 발표되었다. V&J 객체 검출 알고리즘에서 특징 값은 하르-유사 특징을 구성하는 흰 영역과 검은 영역에 대응되는 얼굴 영역의 밝기 합의 차이이다. 특징 값을 위해 검색창 내의 특징들간, 화소 단위로 이동되는 검색창간에 중복된 영역의 밝기 합 계산

이 반복되면서 방대한 계산량이 요구되고 계산량 감소를 위해 적분 영상이 사용된다.

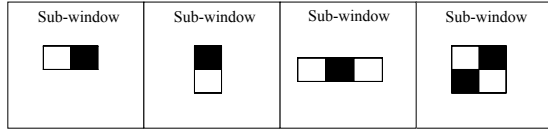


그림 1. 하르-유사 특징 원형들

적분 영상(I)은 객체 검출을 위한 입력 영상(i)의 시작점 $(0,0)$ 와 임의의 점 (x,y) 로 구성되는 사각형 내 화소들의 밝기 합으로 표현된 영상이다. 그림 2의 적분 영상에서 한 화소 I_1 혹은 $I(x,y)$ 는 영역 A 내 화소들의 밝기 합으로 식 (1)과 같이 계산된다. 적분 영상을 이용한 하르-유사 특징 값 계산은 영상의 특정 영역의 밝기 합을 계산하는 것으로 그림 2의 D영역을 예로 들면 식 (2)와 같이 계산된다. 따라서 적분 영상을 생성하는데 다소의 계산량의 추가되지만 적분 영상이 생성되면 특징 값을 위한 특정 영역 내 화소들의 밝기 합 계산은 영역 크기에 관계없이 3 회의 +/-연산으로 쉽게 계산된다.

$$I(x,y) = \sum_{l=0}^x \sum_{m=0}^y i(l,m) \quad (1)$$

$$I_D = I_4 - I_2 - I_3 + I_1 \quad (2)$$

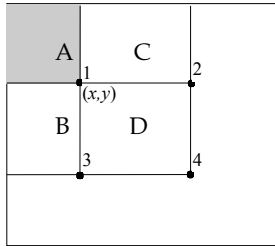


그림 2. 적분 영상

V&J 객체 검출 알고리즘의 Adaboost 알고리즘에서 약분류기(h)는 하르-유사 특징을 이용해 수식 (3)과 같이 구성한다.

$$h(x,f,p,\theta) = \begin{cases} 1, & \text{if } pf(x) < p\theta \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

여기서 x 는 검색창의 영상, f 는 하르-유사 특징 값, p 는 부등식 방향을 위한 극성, θ 는 분류를 위한 임계값이다. 그리고 강분류기는 선택된 일부 약분류기들이 선형적으로 결합되어 식 (4)와 같이 구성된다.

$$C(x) = \begin{cases} 1, & \sum_{t=1}^T \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^T \alpha_t \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

여기서 T 는 약분류기의 수이고, α 는 약분류기의 반복 학습 과정에서 생성된 가중치이다. Adaboost 분류기도 계산 속도를 높이기 위해 분류기를 다단계(cascade)로 구성해 적용하고 있다.

Cascade 분류기는 그림 3과 같이 다단계의 분류기들로 구성된다. 각 단계의 분류기는 다수의 약분류기(WC_x)로 구성된 작은 강분류기(SC_x)이다. 앞 단계 분류기의 긍정적인 결과에 대해서만 다음 단계 분류기를 동작시킴으로 검출 속도를 높일 수 있다. 즉 앞 단계에서 적은 수의 약분류기(특징)로 분류된 비얼굴에 대해 나머지 많은 수의 약분류기를 수행할 필요가 없어 계산량을 줄일 수 있다. Cascade 분류기는 앞 단계에서 잘못된 긍정률(false positive rate)를 어느 정도 감소해야 하나 뒤 단계들에서 단계적으로 잘못된 검출된 객체들을 제거해 나간다. 최종 단계를 거치면 높은 검출율의 고속 분류기가 완성된다.

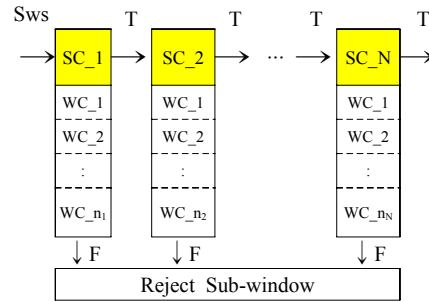


그림 3. Cascade 분류기

III. V&J 객체 검출 알고리즘의 중요 변수

V&J 객체 검출 알고리즘은 Visual Studio나 Matlab 등에서 쉽게 사용할 수 있게 라이브러리로 개발되었고 알고리즘의 MinSize, MaxSize, ScaleFactor, MergeThreshold, ROI는 중요 변수이다. 본 장에서는 얼굴 검출을 위한 V&J 객체 검출 알고리즘의 중요 변수들을 기술한다.

본 논문의 V&J 객체 검출 알고리즘은 하르-유사 특징들을 이용한 얼굴 검출을 목적으로 하고 있으므로 분류 모델을 FrontalFaceCART로 설정하였다. 이 모델은 정면 얼굴 검출을 위한 분류와 회기 트리 분석(classification and regression tree analysis, CART)에 기반한 분류기이다. 얼굴 특징 표현을 위해 하르-유사 특징들을 사용하고 학습 영상은 정면 얼굴 영상으로 크기는 [20, 20]이다.

변수 MinSize와 MaxSize는 각각 영상에서 검출할 수 있는 객체의 최소와 최대 크기를 화소 단위의 [높이, 너비]로 지정한다. 지정될 수 있는 값은 학습 영상의 크기 이상이고 객체 검출 대상 입력 영상의 크기 이하이어야 한다. 값이 지정되지 않은 MinSize와 MaxSize는 각각 학습 영상의 크기와 객체 검출 대상 입력 영상의 크기를 사용한다. MinSize는 MaxSize 이하이어야 한다. 검출

하고자 하는 객체의 크기를 알고 있다면 두 변수를 적절히 지정하므로 검출을 위한 계산 시간을 줄일 수 있다. 그러나 지정된 크기 영역 밖의 객체가 존재한다면 객체를 검출할 수 없다.

변수 ScaleFactor는 다양한 크기의 객체를 검출하기 위해 영상 크기를 조정하는 변수로 1.0001이상의 값이 지정되어야 하고 기본 값으로 1.1이 지정된다. 이 변수에 의해 검색창의 크기를 MinSize에서 MaxSize로 단계적으로 증가시킨다. 변수 값이 작을수록 고려되는 스케일이 많아지므로 다양한 크기의 객체를 검출할 수 있으나 계산량이 많아지고 중복검출 객체들이 발생할 수 있다. 변수 값이 클수록 고려되는 스케일이 적어지므로 계산량이 적어지나 미검출 객체가 발생할 수 있다.

변수 MergeThreshold는 유효 객체를 위한 동일 객체의 수를 지정한다. 지정된 스케일에서 동일 객체에 대해 주변에서 여러 객체들 검출될 수 있다. 이때 검출된 객체 수가 지정된 값 이상일 때 객체로 인정되고 이들은 하나의 객체로 병합된다. 변수 값은 0이상이고 기본 값으로 4가 지정되어 있다. 객체 병합에 의해 동일한 객체에 대해 여러 객체가 검출되는 것을 방지할 수 있고 지정된 값에 의해 오검출 객체도 방지해 줄 수 있다. 변수 값이 작을수록 객체가 중복 검출되거나 비객체가 검출될 수 있고, 변수 값이 클수록 비객체가 제거될 수 있으나 객체도 제거될 수 있다.

ROI 변수는 False 혹은 True가 지정되며 기본 값으로 False이나 True로 지정되면 객체 검출이 지정된 관심 영역에서만 수행된다. ROI 기능을 사용하기 위해서는 객체 검출에 앞서 관심 영역을 지정하는 선행 처리가 이루어져야 한다.

IV. 실험 결과 및 검토

얼굴 검출을 위한 V&J 객체 검출 알고리즘의 변수에 따른 성능 평가는 표 1의 영상들을 대상으로 수행된다. 영상들은 다양한 해상도와 다양한 수의 얼굴을 포함하고 있다. R과 N은 각각 영상 해상도와 유효 얼굴 수이다.

표 2는 V&J 객체 검출 알고리즘에 설정된 MinSize와 MaxSize의 영향을 보여주고 있다. Fmin과 Fmax는 사전 얼굴 검출 실험을 통해 얻은 영상 내 얼굴의 최소와 최대 크기이다. 얼굴 검출 성능에는 차이가 거의 없었고 수행 시간에서 차이를 보였다. MaxSize보다 MinSize의 영향이 컸고 최소 얼굴 크기와 최대 얼굴 크기를 제한하였을 때 계산 시간이 30~50% 감소된다. G2는 수많은 작은 거짓 얼굴들에 의해 불필요한 계산이 많이 요구되는 특별한 영상이다.

표 3과 그림 4는 V&J 객체 검출 알고리즘에 설정된 ScaleFactor(SF)의 영향을 보여주고 있다. 나머지 변수는 기본 값으로 설정되었고, DOB, ROB, MOB, WOB, Ti는 각각 검출된 얼굴, 유효 얼굴, 미검출 얼굴, 오검출 혹은 중복 검출 얼굴의 수,

표 1. 실험 영상들



표 2. MinSize와 MaxSize에 따른 수행 시간[sec]

Images	Fmin	Fmax	ROI	[Fmin]	[Fmax]	[Fmin][Fmax]
G1	45	55	0.18	0.150	0.162	0.128
				-17.9	-11.3	-30.0
G2	84	107	1.87	0.412	1.736	0.307
				-78.0	-7.2	-83.6
G3	37	69	0.63	0.422	0.565	0.358
				-32.5	-9.7	-42.8
G4	36	47	0.32	0.229	0.255	0.170
				-28.4	-20.3	-47.0
G5	43	80	0.95	0.565	0.848	0.477
				-40.5	-10.6	-49.8

표 3. ScaleFactor에 따른 얼굴 검출

Images	SF	DOB	ROB	MOB	WOB	Ti
G1	1.001	24	7	0	17	10.18
	1.005	19	7	0	12	1.48
	1.01	16	7	0	9	0.81
	1.1	8	7	0	1	0.28
G2	1.001	389	19	0	370	253.43
	1.005	221	19	0	202	33.12
	1.01	151	19	0	132	15.73
	1.1	37	19	0	18	1.77
G3	1.001	149	66	0	83	423.23
	1.005	94	66	0	28	23.36
	1.01	85	66	0	19	8.05
	1.1	70	66	0	4	0.69
G4	1.001	48	46	20	2	0.35
	1.001	127	18	0	109	35.14
	1.005	36	18	0	18	4.30
	1.01	33	18	0	15	0.38
G5	1.1	17	16	2	1	0.28
	1.4	7	6	12	1	0.13
	1.001	322	34	1	288	329.61
	1.005	156	34	1	122	24.25
G5	1.01	114	34	1	80	9.74
	1.1	42	34	1	8	1.01
	1.4	34	28	7	6	0.48

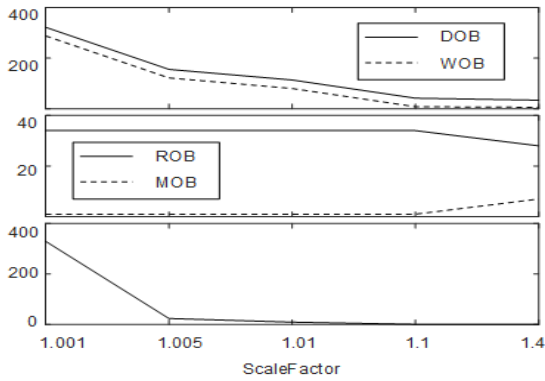


그림 4. G5에서 ScaleFactor에 따라 얼굴 검출

표 4. MergeThreshold에 따른 얼굴 검출

Images	MT	DOB	ROB	MOB	WOB	Ti
G1	1	12	7	0	5	0.27
	2	10	7	0	3	0.27
	4	8	7	0	1	0.27
	8	7	7	0	0	0.27
	16	7	7	0	0	0.27
G2	1	139	19	0	120	2.04
	2	63	19	0	44	1.90
	4	37	19	0	18	1.87
	8	22	19	0	3	1.88
	16	19	19	0	0	1.83
G3	1	88	66	0	22	0.71
	2	71	66	0	5	0.72
	4	70	66	0	4	0.69
	8	64	64	2	0	0.72
	16	60	60	4	0	0.71
G4	1	27	17	1	10	0.41
	2	20	17	1	3	0.41
	4	17	16	2	1	0.41
	8	14	14	4	0	0.41
	16	11	11	7	0	0.41
G5	1	116	34	1	82	1.11
	2	65	34	1	31	1.03
	4	42	34	1	8	1.01
	8	35	33	2	2	1.03
	13	32	32	3	0	1.02



그림 5. G5에서 검출된 얼굴 영상

그리고 수행 시간을 의미한다. 작은 SF에서는 미검출되는 얼굴은 없으나 오검출 및 중복 검출 얼굴이 많아져 얼굴 검출 성능을 떨어뜨리고 있고 계산량이 기본 값에 비해 수백 배로 늘어나고 있

다. 반면 큰 SF에서는 오검출된 얼굴이 발생하고 있다. 기본 값에서 가장 좋은 결과를 보이나 영상에 따라 오검출 및 중복 검출 얼굴이 많이 발생하고 있다.

표 4는 V&J 객체 검출 알고리즘에 설정된 MergeThreshold(MT)의 영향을 보여주고 있다. 나머지 변수는 기본 값으로 설정되었다. MT 역시 작은 값에서 미검출되는 얼굴은 없으나 오검출 및 중복 검출 얼굴이 많아져 얼굴 검출 성능을 떨어뜨리고 있고 큰 값에서 오검출된 얼굴이 발생하고 있다. 한편 수행 시간은 MT의 영향을 받지 않고 있다.

그림 5는 변수를 기본 값으로 설정해 얼굴을 검출한 예이다. 진한 화살표는 다른 사람에 의해 거의 가려져 검출되지 않은 얼굴이고 연한 화살표는 중복 혹은 잘못 검출된 얼굴이다.

V. 결 론

본 논문은 변수에 따른 V&J 객체 검출 알고리즘의 얼굴 검출 성능을 평가하였다. 변수 MinSize와 MaxSize는 특별한 영상이 아니면 1[sec]미만으로 영향을 받지 않고 있고, 변수 ScaleFactor와 변수 MergeThreshold는 기본 값으로 설정될 때 가장 좋은 성능이라고 판단되는 결과를 보여주고 있다 그러나 최적의 변수에서도 미검출되는 얼굴이 존재하고 오검출 및 중복 검출되는 얼굴이 존재하는 문제를 갖고 있고 이들에 대한 처리가 필요할 것 같다.

참고문헌

- [1] Viola, Paul and Michael J. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features", *Proc. of the 2001 IEEE Computer Society Conference on CVPR*, vol. 1, pp.511-518, 2001.
- [2] S. Zafeiriou, C. Zhang, and Z. Zhang. A survey on face detection in the wild: past, present and future. *CVIU*, vol. 138, pp.1-24, 2015.
- [3] https://kr.mathworks.com/help/vision/ref/vision_cascadeobjectdetector-system-object.html
- [4] https://docs.opencv.org/3.0-beta/modules/objdetect/doc/cascade_classification.html.