

## 얼굴인식에서 정확도 향상을 위한 SURF 특징점에서의 Gabor 기술어 추출

\*김지은, \*\*조혜정, \*\*\*정광수, \*\*\*\*오승준  
 광운대학교 방통융합 플랫폼 센터  
 \*jjeun@media.kw.ac.kr

### Gabor descriptors extraction in the SURF feature point for improvement accuracy in face recognition

\*Kim, Ji Eun \*\*Cho, Hye Jeong \*\*\*Chung, Kwang-Sue \*\*\*\*Oh, Seung-Jun  
 BnC Convergence Platform Center, Kwangwoon University

#### 요 약

본 논문에서는 대표적인 특징점 추출 알고리즘인 SURF (Speeded Up Robust Features)와 얼굴인식에서 널리 쓰이는 Gabor 기술어를 이용한 얼굴 인식 방법을 소개한다. SURF 기반 영상인식 방법은 특징점을 찾고 해당 특징점에서 기술어를 추출한 후, 정합을 수행한다. 본 논문에서는 SURF 를 통해 추출한 특징점에서 Gabor 웨이블릿 변환을 사용해 기술어를 추출하는 얼굴인식 방법을 제안한다. 잘 알려진 ORL 데이터베이스에서의 실험에서 제안한 방법이 기존 SURF 기반의 얼굴 인식 방법에 비해 더 높은 얼굴 인식 성능을 보여줄 뿐 아니라 정합시간을 포함한 처리 속도면에서도 더 우수한 성능을 보였다. 이러한 실험 결과를 통하여 제안하는 방법이 SURF 보다 얼굴 인식에 적합함을 확인할 수 있었다.

#### 1. 서론

얼굴인식은 다양한 응용이 가능한 주제이지만 해결하기 어려운 문제이기 때문에 지난 10년간 많은 관심을 받았다[1]. 데이터 집합의 통계적 특성을 분석하여 적절한 차원의 특징을 추출하는 부분공간분석 방법인 Eigenface 방법과 Fisherface 방법, Gabor 웨이블릿 변환(wavelet transform)을 이용한 방법은 잘 알려진 얼굴 인식을 위한 접근 방법들이다[2-3]. 하지만 이러한 전역적 특징 추출 방법은 사용하는 벡터의 차원 수가 높기 때문에 연산 시간의 증가 뿐 아니라 인식률을 감소시키는 차원의 저주 (Curse of dimensionality)와 같은 문제를 유발하기도 한다. 이를 해결하기 위해, 벡터의 차원을 축소시키는 PCA(Principal Component Analysis), LDA (Linear Discriminant Analysis) 및 ICA (Independent Component Analysis) 등의 방법이 제안되었다[4-5].

반면 영상의 객체 인식 분야에서 주로 사용되는 지역적 특징 추출 방법은 일정 지역에 해당하는 화소를 이용하기 때문에 벡터의 차원 수는 전역적 특징 추출 방법 보다 낮아진다. 따라서 연산시간이나 차원의 저주와 같은 문제에 상대적으로 영향을 덜 받는다. 하지만 해당 화소를 선택하고 그에 따른 기술어를 추출함에 있어서 기술어는 영상의 주요 부분을 충분히 잘 표현할 수 있도록 하는 구별성(distinctness)이 보장되어야 하며, 동시에 환경 변화에 강인해야 한다[6].

대표적인 지역적 접근 방법으로는 SIFT (Scale Invariant Feature Transform)와 SURF(Speeded Up Robust Feature)

등이 있으며 이들은 물체인식, 위치인식 등에 사용되고 있다. 최근에는 얼굴 인식과 관련된 연구에도 응용되는 추세이다 [7-8].

일반적으로 SURF 는 물체를 인식하는데 응용되어 왔다. 서로 다른 물체의 경우에는 형태 및 구조에 많은 차이가 있어 특징점 간에 분별력이 있다. 반면에, 서로 다른 사람의 경우에는 얼굴의 형태 및 구조가 유사해 물체의 경우보다 분별력이 떨어지게 된다. 따라서 SURF 를 얼굴인식에 응용함에 있어 얼굴 인식의 정확도를 높이는 방법이 연구되어야 한다. 또한 대용량의 데이터베이스를 이용하기 때문에 얼굴인식 수행시간도 매우 중요하게 고려되어야 할 부분이다.

SURF 의 미리 결정된 스케일과 간단한 방법의 기술어 추출 보다는, 다양한 스케일과 여러 방향 성분을 고려한 Gabor 웨이블릿 변환을 이용해 추출된 기술어가 더 강인하다. 본 논문에서는 좀 더 강인한 기술어를 추출하기 위해서 SURF 를 이용해 특징점을 찾고 해당 특징점에서 Gabor 웨이블릿 변환을 이용해 기술어를 추출하여 정합하는 얼굴 인식 방법을 제안한다.

본 논문의 구성은 다음과 같다. 2 장에서는 영상에서 기술어를 추출하기 위해 기존에 사용된 SURF 기반 기술어 추출 방법과 Gabor 웨이블릿 변환을 이용한 기술어 추출 방법에 대해 설명하고, 3 장에서는 제안하는 방법에 대하여 설명한다. 이어서 4 장에서는 실험 환경을 제시하고 실험을 통해 제안하는 방법을 입증한다. 마지막으로, 5 장에서 본 논문에 대한 결론을 맺는다.

## 2. 기존의 기술어 추출 방법

### 2.1 SURF 기반 기술어 추출

특징점 추출 알고리즘인 SURF 는 크게 특징점 추출과 특징점 서술로 나누어진다. 먼저, 특징점 추출에서는 근사 헤시안 행렬(Approximated hessian matrix)의 행렬식(determinant)을 이용하여 특징점이 될 수 있는 코너의 여부를 판단한다. 이때 코너는 행렬식의 값이 0 보다 큰 경우이며, 극값(maximum) 이라고도 한다.

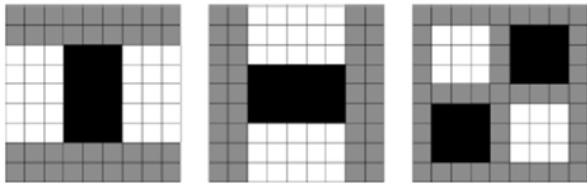


그림 1. 근사 헤시안 행렬의 필터 ( $D_{xx}$   $D_{yy}$   $D_{xy}$ )

근사 헤시안 행렬을 구성하는  $D$  는 가우시안 커널(gaussian kernel)의 2 차 편도 함수(second order partial derivative)의 근사식을 의미하며, 각각의 근사식은 그림 1 과 같이 필터의 형태로 영상에 컨벌루션(convolution) 된다. 이때 필터의 흑색 영역은 -2, 백색영역은 1, 그리고 회색 영역은 0 의 가중치가 적용된다. 필터들은 스케일이 일정하게 증가되기 때문에 다양한 스케일의 코너가 검출될 수 있다.

특징점 서술에서는 추출된 특징점의 위치 주변의 화소 값을 이용하여 특징점의 방향을 계산한다. 이는  $x$  방향과  $y$  방향으로의 1 차 도함수에 해당하는 하르 웨이블릿 필터(haar wavelet filter)로 컨벌루션 한 결과인 기울기의 크기(magnitude)와 방향을 이용한다. 근사 헤시안 행렬의 필터와는 다르게 흑색 영역은 1 그리고 백색 영역은 -1 의 가중치가 적용된다.

기울기가 계산되는 화소 값들의 범위는 특징점의 스케일에 따라 다르며, 계산된 기울기들은 그림 2 의 사각형에 해당 한다. 그리고 특징점의 위치를 중심으로 하는 부채꼴 모양의 슬라이딩 윈도우(sliding window)가 360° 회전하면서 윈도우 내부의 기울기들을 모두 더하여 다수의 벡터를 생성한다. 이 벡터들 중에서 가장 크기가 큰 벡터의 방향이 특징점의 방향이 된다.

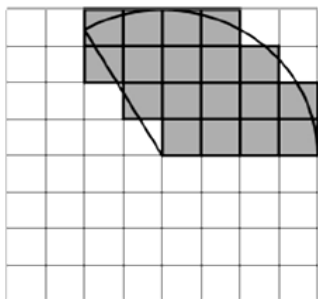


그림 2. 슬라이딩 윈도우를 이용한 특징점의 방향계산

방향이 계산된 특징점은 최종 단계인 벡터 형태의 기술어로 구성된다. 이는 특징점의 위치와 방향을 기준으로, 주변의 화소값들을 이용해 20x20 의 기울기들을 다시 계산하고 5x5 의 기울기 16 개로 분할된다.

5x5 의 기울기는 표 1 과 같이 분류되고 계산된다. 이때  $dx$  는  $x$  방향의 기울기이며  $dy$  는  $y$  방향의 기울기를 의미한다. 즉,  $x$  방향과  $y$  방향에 따라 기울기를 나눈 다음, 모두 합한 값과 절댓값을 취한 후 합한 값을 계산함으로써 총 4 개의 값이 계산된다. 나머지 5x5 기울기에서도 이를 동일하게 적용하면 총 64 개의 값이 계산되며, 이를 이용해 64 차원의 기술어를 구성하게 된다. 이를 본 논문에서는 SURF64 로 정의한다.

표 1. 64 차원 기술어의 기울기 구성

$dx$		$dy$	
$\sum dx$	$\sum  dx $	$\sum dy$	$\sum  dy $

표 2 와 같이, 방향과 절댓값뿐만 아니라 값의 크기도 고려하여 분류하게 되면 총 8 개의 값이 계산된다. 이를 이용하여 구성된 기술어는 128 차원이 된다. 본 논문에서는 이를 SURF128 로 정의한다.

표 2. 128 차원 기술어의 기울기 구성

$dx \geq 0$	$dx < 0$	$dy \geq 0$	$dy < 0$
$\sum dy$	$\sum  dy $	$\sum dx$	$\sum  dx $

SURF 의 특징점 추출 및 서술에서 가장 많이 반복되는 컨벌루션 작업은 화소 값의 합을 계산하는 연산이 주를 이룬다. 때문에 원본영상을 이용하여 계산한 적분영상(integral image)에서 컨벌루션을 수행하면, 필터의 크기와 상관없이 빠른 속도로 계산할 수 있다.

### 2.2 Gabor 웨이블릿 변환을 이용한 기술어 추출

Gabor 웨이블릿 변환에 사용되는 Gabor 필터는 인간의 시각적 인식 체계를 가장 잘 표현하고 있으며, 각각의 신경 세포들의 기능을 한꺼번에 표현할 수 있는 함수로써, 인간 대뇌의 시각 피질을 모델로 설계되었다.

Gabor 필터는 다양한 스케일과 다양한 방향을 갖는다. 스케일이 클 때는 영상 전체의 정보를 잘 나타내고, 스케일이 작을 때는 이미지의 지역적인 정보를 잘 나타낸다. 또한 특정 주파수 성분과 방향성을 가지는 영상 구조가 어디에 속해 있는지의 지역적인 정보를 효과적으로 표현할 수 있는 필터이다.

Gabor 필터는 시간 영역의 해상도 정보와 주파수 영역의 성분을 나타내는 STFT(Short Time Fourier Transform)으로 사인파에 의해 변조된 가우시안 커널 함수의 형태를 가진다. 2 차원 Gabor 필터의 일반적인 형태는 다음과 같다.

$$g(x, y; \lambda, \theta, \phi, \sigma, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \cos\left(2\pi \frac{x'}{\lambda} + \phi\right) \quad (1)$$

여기서  $x, y$  는 Gabor 웨이블릿 커널에서 중심점을 원점으로 하는 위치를 나타내며,  $x' = x \cos \theta + y \sin \theta$  ,  $y' = -x \sin \theta + y \cos \theta$  이고,  $\theta$  는 웨이블릿의 방향,  $\lambda$  는 웨이블릿의 파장,  $\phi$  는 웨이블릿의 이동된 정도, 그리고  $\sigma$  는 가우시안 커널의 크기를 나타낸다. 그림 3 은 Gabor 웨이블릿 커널 집합의 예이다. 가로축이 방향의 변화를 나타내고, 세로축이 스케일의 변화를 나타낸다.

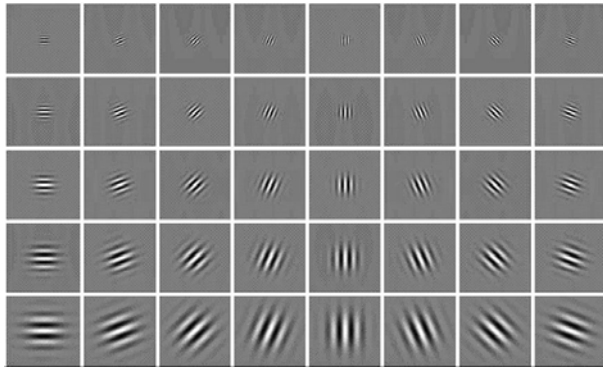


그림 3. Gabor 웨이블릿 커널 집합

이러한 Gabor 웨이블릿 변환은 물질 고유의 특징을 방향성과 크기를 이용하여 표현하는 방식으로 얼굴인식 분야에서 널리 사용되고 있다. Gabor 기술어는 입력되어진 얼굴 영상과 Gabor 필터와의 컨벌루션에 의하여 얻어진다. Gabor 기술어는 입력 영상의 전체 화소에서 구하며 하나의 화소의 기술어는 40 차원이 된다.

### 3. 제안하는 기술어 추출 방법

SURF 기반의 기술어 추출 방법에서는 특징점을 검출할 때 정해진 하나의 스케일에 따라 하나의 방향과 영역의 크기가 결정되게 된다. 결정된 해당 영역의 분할 영역에서 x 방향의 기울기와 y 방향의 기울기 성분을 이용해 기술어를 추출한다.

반면, Gabor 웨이블릿 변환을 이용한 기술어 추출 방법에서는 특징점에서 여러 방향 성분과 스케일의 조합으로 이루어진 커널을 컨벌루션하여 기술어를 추출한다. 따라서 Gabor 웨이블릿 변환 기반의 기술어는 SURF 기반의 기술어보다 더 강한 특성을 갖는다.

이러한 특성을 이용해 본 논문에서는 표 3 과 같이 SURF 의 근사화시킨 헤시안 행렬과 비 최대 억제법(non-maximum suppression)을 이용해 특징점을 검출한 후, Gabor 웨이블릿 변환을 이용해 기술어를 추출하고 정합하는 방법을 제안한다. 제안하는 방법에서는 Gabor 웨이블릿 필터의 크기를 25x25 로 고정하여 사용하며, 이는 실험을 통해 결정된 값이다.

표 3. 제안하는 방법

특징점 검출	기술어 추출	정합
근사화시킨 헤시안 행렬 + 비 최대 억제법	Gabor 웨이블릿 변환	KD 트리 + 유클리디안 거리

얼굴을 인식하는데 있어 인식을 뿐 아니라 수행 시간 또한 중요하게 고려되는 요소이다. 수행시간은 기술어 추출 시간과 정합시간을 합한 것인데, 입력된 영상의 기술어와 데이터베이스의 기술어를 비교하는 정합시간이 상당부분 차지한다. 제안하는 방법은 표 4 와 같이 다른 방법들에 비해 한 영상의 기술어 개수가 적기 때문에 정합시간을 상당히 줄일 수 있고, 그에 따라 수행시간이 감소된다. 표 4 의 SIFT, SURF, 제안하는 방법의 특징점의 수는 데이터베이스에서 각 영상에서 구한 특징점 수의 평균을 낸 값이다.

표 4. 영상의 특징점과 기술어의 수 (92x112 기준)

방법	Gabor	SIFT	SURF128	SURF64	제안하는 방법
한 영상의 특징점의 수	10304	60.92	27.76	27.76	27.76
한 특징점에서 기술어의 수	40	128	128	64	40
한 영상의 기술어의 수	412160	7797.76	3555.84	1777.92	1111.2

### 4. 실험 결과

이 절에서는 제안된 방법의 유용성을 검증하기 위하여 다양한 변화를 지니는 ORL 데이터베이스를 이용한 비교실험을 수행한다. 사용한 데이터는 Cambridge 의 Olivetti Research Laboratory 에서 제작된 ORL 데이터베이스이다. 영상의 크기는 92x112 이고 40 명에 대하여 빛의 위치 변화, 안경 착용 유무, 표정 변화(눈의 변화, 웃음 유무) 등에 따라 각 10 장씩으로 구성되어 있다. 전체 데이터베이스에 일정하게 나타나는 규칙은 없고, 모든 얼굴 영상은 검은 배경을 바탕으로 한다. 안경, 수염 등이 있는 얼굴들이 많이 분포 되어 있고 좌우 회전을 통해 변형된 얼굴 영상도 다소 존재하기 때문에 난이도가 있는 얼굴 데이터베이스라고 할 수 있다[9].



그림 4. ORL 데이터베이스의 예

ORL 데이터베이스에 대하여 기존의 기술어 추출 방법인 Gabor, SIFT, SURF128, SURF64 와 제안하는 방법의 성능을 비교하였다. 실험의 정량적인 평가를 위해서 MPEG-7 검색 성능 평가 지수인 ANMRR(Average Normalized Modified Retrieval Rank)를 이용한다[10].

기존의 다른 방법들과 본 논문에서 제안하는 방법의 성능 비교는 그림 5 와 같다. 동일한 조건에서 Gabor 의 ANMRR 은 0.1517 이고, SURF128 은 0.3157, SURF64 는 0.303 그리고 제안된 방법은 0.233 의 성능을 보인다. 그림 5 를 통해 제안하는 방법의 성능은 Gabor 와 SIFT 에는 못 미치지만 SURF 보다 향상된 것을 확인할 수 있다.

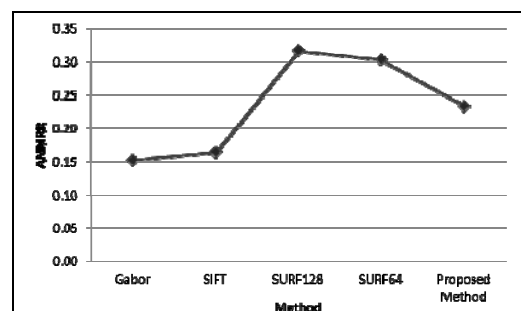


그림 5. 기존 방법들과 제안한 방법의 성능비교

표 5 는 기존 방법들과 제안한 방법의 처리시간을 비교한 것이다. 수행시간은 기술어 추출시간과 정합시간을 더한 시간이다.

표 5. 기존 방법들과 제안된 방법의 소요시간 비교

방 법	Gabor	SIFT	SURF128	SURF64	제안하는 방법
기술어 추출시간(ms)	498.36	21.94	4.11	4.07	186.04
정합시간(ms)	384479.68	3920.77	2276.12	1255.61	879.16
수행시간(ms)	384978.04	3942.73	2279.23	1259.68	1065.22

표 5 를 통해 제안된 방법의 수행시간이 가장 적게 걸리는 것을 확인할 수 있으며, 좋은 성능을 보인 Gabor 와 SIFT 는 수행시간이 제안하는 방법보다 각각 약 360 배, 3 배 더 걸린다. 많은 데이터베이스와 정합해야 하는 점에 있어 오랜 수행시간은 불리할 수 밖에 없다.

따라서 SURF 보다 성능이 좋고, 다른 방법과 비교해 수행시간도 가장 빠르기 때문에 제안하는 방법이 인식률과 수행시간을 동시에 고려할 때 효율적인 방법이라고 볼 수 있다.

## 5. 결론

본 논문에서는 SURF 를 이용해 찾은 특징점에서 Gabor 웨이블릿 변환을 이용해 기술어를 추출하고, 정합하는 얼굴인식 알고리즘을 제안하였다. 제안한 방법의 유용성을 판단하기 위해 ORL 데이터베이스를 이용하여 다양한 변화를 가진 얼굴을 인식하였다. 제안하는 방법은 실험 결과에서 기존 SURF 보다 높은 얼굴인식 성능과 동시에 더 빠른 처리 속도를 보여주었다.

향후에는 사용하는 Gabor 필터를 고정된 25x25 크기가 아닌 특징점의 스케일을 반영한 필터를 사용하는 방법을 연구할 예정이다.

## 감사의 글

본 연구는 지식경제부, 한국산업기술진흥원의 전략 기술 인력양성사업으로 수행된 결과이며, 서울시 산학연 협력사업(SS100006) 지원으로 수행되었음.

## 참 고 문 헌

[1] W. Zhao, R. Chellappa, P.J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," ACM Comput. Surv., vol. 35, No. 4, pp. 399-458, 2003.

[2] M. Turk and A. Pentland, " Eigenfaces for recognition," Journal of cognitive neuroscience, vol. 3, No. 1, pp. 71-86. 1991.

[3]D. Gabor, "Theory of Communication," J. Inst. Elect. Eng., vol. 93, no 3, pp. 429-457, 1946.

[4] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, " Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," IEEE Trans. Pattern Anal. Mach. Intell., vol. 19, no. 7, pp. 711- 720, Jul.1997.

[5] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, " Face recognition by independent component analysis," IEEE Trans. Neural Netw., vol.13, no. 6, pp. 1450- 1464, Jun. 2002.

[6] R. Verschae, J. Ruiz-del-Solar, and M. Correa, " Face recognition in unconstrained environments: a comparative study," Proc. of ECCV Workshop on Faces in Real-Life Images, 2008.

[7] D.G. Lowe, " Distinctive Image Features from Scale-Invariant Keypoints," Int. J. Comput. Vision, vol. 60, No.2, pp. 91-110, 2004.

[8] H. Bay, Beat Fasel, and Luc Van Gool, " Interactive museum guide: Fast and robust recognition of museum objects" , In First international workshop on mobile vision, 2006.

[9] Ferdinando Samaria, Andy Harter. Parameterisation of a Stochastic Model for Human Face Identification. Proceedings of 2nd IEEE Workshop on Applications of Computer Vision, Sarasota FL, December 1994

[10] V. V. Vinod and B. S. Manjunath, " Report on AHG of color and texture," ISO/IEC/JTC1/SC29/WG11, Doc. M5560, Maui, December 1999. 332