

블록가중치의 최적화를 통해 개선된 LBP기반의 표정인식

박성천*, 구자영**

An Improved LBP-based Facial Expression Recognition through Optimization of Block Weights

Seong Chun Park*, Ja Young Koo**

요약

본 논문에서는 Local Binary Pattern 히스토그램의 템플릿 매칭을 이용한 얼굴 표정 인식에서 인식률을 높이는 방법을 제안한다. 이 방법에서, 주어진 얼굴 영상은 작은 크기의 블록으로 분할되고 각 블록에서 구해진 LBP 히스토그램은 블록 특징으로 사용된다. 입력 영상에서의 블록 특징과 모델의 해당 블록 특징 사이에서 블록 상이도가 계산된다. 주어진 영상과 모델 영상 사이의 영상 상이도는 블록 상이도의 가중 합으로 계산된다. 기존의 방법들에서는 직관에 따른 블록 가중치를 사용하는데 본 논문에서는 블록 가중치를 트레이닝 샘플들로부터 최적화를 통해서 구하는 방법을 제안하고 있다. 실험을 통해서 제안된 방법이 기존의 방법보다 우수함을 보인다.

Abstract

In this paper, a method is proposed that enhances the performance of the facial expression recognition using template matching of Local Binary Pattern(LBP) histogram. In this method, the face image is segmented into blocks, and the LBP histogram is constructed to be used as the feature of the block. Block dissimilarity is calculated between a block of input image and the corresponding block of the model image. Image dissimilarity is defined as the weighted sum of the block dissimilarities. In conventional methods, the block weights are assigned by intuition. In this paper a new method is proposed that optimizes the weights from training samples. An experiment shows the recognition rate is enhanced by the proposed method.

▶ Keyword : 얼굴표정인식(Facial Expression Recognition), Local Binary Pattern(LBP), 템플릿 매칭(Template Matching)

• 제1저자 : 박성천 교신저자 : 구자영
• 투고일 : 2009. 10. 20, 심사일 : 2009. 11. 10, 게재확정일 : 2009. 11. 11.
* 단국대학교 컴퓨터과학전공 박사과정 ** 단국대학교 컴퓨터 학부 교수
※ 본 연구는 2008년 단국대학교 교내연구비에 의해 연구되었음.

I. 서론

인간은 표정을 통해 감정이나 생각을 표현함으로써 타인과 소통한다. 사람의 표정을 기계가 인식하도록 하면 인간과 기계 사이의 소통 방식이 더욱 다양해 질 것이고, 인간 컴퓨터 상호작용(HCI) 및 컴퓨터 애니메이션 등 다양한 분야에서 활용될 수 있을 것이다. 이러한 목적을 따라 지금까지 얼굴 표정의 자동 인식에 관한 많은 연구가 행해져 왔다[1-6,15,16]. 그러나 표정의 복잡성과 다양성으로 인해 표정의 인식은 여전히 어려운 문제로 남아있다.

얼굴 표정의 인식에는 여타의 패턴 인식 문제에서와 마찬가지로 특징의 추출과 분류기 설계라고 하는 두 가지 주요 과제가 포함된다. 성공적 얼굴 표정의 인식을 위해서는 무엇보다도 효과적인 얼굴의 표현방식이 요구된다. 얼굴 특징의 추출 방식은 기하학적 특징 기반의 방법과 외관 기반의 방법으로 대별된다[7]. 기하학적 특징 기반의 방법에서는 눈, 코, 귀, 입, 등의 모양과 위치를 검출해서 얼굴을 표현하는 특징으로 사용한다. 반면에 외관 기반의 방법에서는 특정 기관의 위치에 대한 사전 지식을 사용하지 않고 얼굴의 텍스처 분포를 이용해서 얼굴을 표현한다. 기하학적 특징 기반의 방법에서는 일반적으로 정밀하고 신뢰성 있는 얼굴 기관의 검출을 필요로 하는데, 이것은 실제 상황에서 자세 변경 등 변이에 따라 어려운 문제로 대두된다. 반면에 외관 기반의 방법은 이러한 제약으로부터 비교적 자유롭고 표정을 모델링하는 피부 텍스처 변화를 용이하게 표현할 수 있다.

외관 기반의 특징 가운데 하나인 Gabor-wavelet 표현은 기하학적 특징에 비해 효과적이며 실제계 환경에서 보다 나은 결과를 보인다는 점에서 많은 인식기에서 채택되었다[8]. 그러나 다중 스케일과 다중 방향에 대해서 회선 연산을 통해 Gabor 계수들을 구해야 하는 많은 계산량이 부담이 된다.

Local Binary Pattern(LBP) 기반의 특징 추출은 텍스처 분석 분야에서 먼저 사용되던 방법으로서 조명의 변화에 안정적이고 계산이 빠른 장점을 보임에 따라 최근에 얼굴의 검출과 인식에 활용되었으며[9,10] 얼굴 표정의 인식에도 사용되었다[11]. 이 때 얼굴 영상은 작은 블록으로 분할되며 각 블록에서 LBP 히스토그램이 추출된다. 분류기는 블록 간의 히스토그램의 유사도를 기반으로 전체 영상의 유사도를 계산함으로써 미리 저장된 모델들 중 가장 유사한 모델로 분류하는 기능을 한다. 지금까지 제안된 분류 방법들에서는 블록들 중 표정을 결정하는 중요도에 따라서 눈 코 입 등을 포함하는 블록에는 높은 가중치를 부여하고 그 외의 블록에는 상대적으로

로 낮은 가중치를 부여하는 방법을 사용하였다[9,10,11]. 그러나 이러한 가중치 부여 방법은 직관에 의존한 것일 뿐 합리적 근거를 가지고 있지 못하다.

본 논문에서는 트레이닝 샘플들을 이용해서 블록에 할당되는 가중치를 최적화함으로써 기존의 방법들보다 인식률을 높이는 방법을 제안하고 있다.

2장에서 Local Binary Pattern을 소개하고 3장에 블록 가중치의 최적화를 기술한다. 4장에서는 실험 결과를 보이고 5장에서는 본 연구의 결론을 기술한다.

II. Local Binary Pattern

LBP 연산자는 Ojala 등에 의하여 도입되었다[12]. 그림 1 처럼 화소 주변의 3 by 3 이웃 화소들의 값과 중앙의 값을 비교해서 크면 1, 작으면 0을 할당함으로써 구성되는 8 비트의 수가 중앙 화소에 대한 LBP 값이 된다. 국지적 영역 내에서의 LBP 값들의 히스토그램이 그 영역에 대한 특징 기술자가 된다.

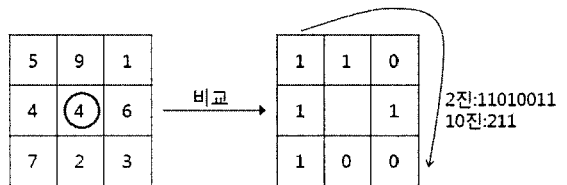


그림 1. LBP 연산자
Fig. 1. LBP Operator

기본적인 LBP 연산자는 3 by 3 마스크를 사용하기 때문에 보다 큰 스케일에서의 구조적 특징을 나타내지 못하므로 이러한 점을 보완하기 위해서 다른 크기의 연산자로 확장되었다[13]. 그림 2에 보이는 바와 같이 하나의 화소를 중심으로 임의의 반경 R 을 갖는 원주 상에서 보간에 의해서 P 개의 점을 등간격으로 샘플링 하도록 확장되었으며 (P, R) 로 표기한다.

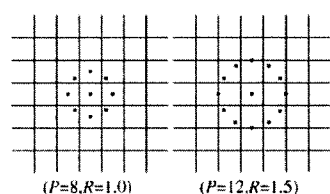


그림 2. 확장된 LBP 의 예들
Fig. 2. Examples of extended LBP.

LBP 연산자는 유니폼 패턴만을 사용하도록 다시 확장될 수 있다. $LBP_{P,R}$ 에 의해 가능한 2^P 개의 패턴들 중 원형으로 고려될 때 0과 1사이의 전환이 2회 이내인 것을 유니폼 패턴이라고 한다[12]. 예를 들어 00000000, 11100011과 00011111은 유니폼 패턴 이다. 유니폼 패턴만을 고려하는 LBP 연산자를 $LBP_{P,R}^{u2}$ 로 표기한다.

이러한 연산에 의해서 산출되는 값들은 국지적인 미세 텍스처를 나타낸다(그림 3)

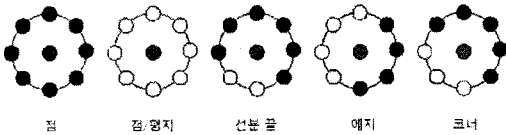


그림 3. LBP에 의해서 표현되는 미세 텍스처
Fig. 3. Texture primitives represented by LBP

LBP 연산자에 의해 레이블된 영상 $f_l(x,y)$ 의 히스토그램 H_i 는 다음과 같이 정의된다.

$$H_i = \sum_{x,y} I(f_l(x,y) = i), \quad i = 0, \dots, n-1 \quad \dots (1)$$

여기서 n 은 LBP에 의하여 산출되는 가능한 레이블의 수이고 $I(A)$ 는 다음과 같은 함수이다.

$$I(A) = \begin{cases} 1 & A \text{ is true} \\ 0 & A \text{ is false} \end{cases} \dots (2)$$

히스토그램은 LBP로 표현되는 미세 텍스처의 분포를 나타낸다. 효과적인 얼굴 표현을 위해서 얼굴 영상은 그림 4 처럼 $m \times n$ 개의 작은 영역, 즉 블록 $R_0, R_1, \dots, R_{m \times n}$ 으로 분할되고 블록 R_j 에서의 히스토그램은 다음과 같다.

$$H_{i,j} = \sum_{x,y} I(f_l(x,y) = i)I((x,y) \in R_j) \dots (3)$$

$i = 0, \dots, n-1, j = 0, \dots, m \times n - 1$

각 화소에 레이블된 LBP는 화소 수준에서의 미세 텍스처를 나타내고, 블록에서의 LBP 히스토그램은 블록 수준에서의 특징 기술자가 되며 모든 블록에서의 히스토그램을 연결시킨 히스토그램은 영상 전체에 대한 특징 기술자로 사용된다(그림 4).

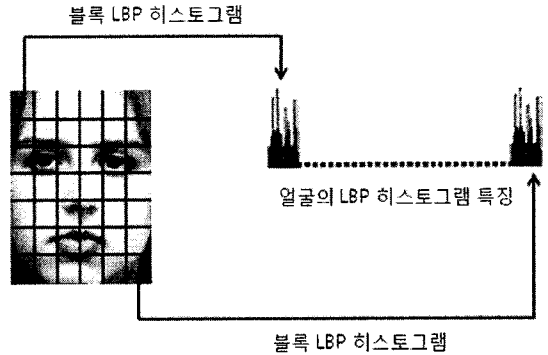


그림 4. 블록 LBP 히스토그램과 그것들이 연결되어 구성된 얼굴 LBP 히스토그램 특징
Fig. 4. The block histograms concatenated to form the face LBP histogram feature.

III. 템플릿 매칭에 의한 표정인식

3.1 템플릿 매칭에 의한 인식

일단 특징이 추출되면 인식 과정은 미리 저장된 모델의 히스토그램과 추출된 입력 영상의 LBP 히스토그램 간의 유사도에 근거해서 분류하는 과정이다. Chi square(χ^2)는 히스토그램의 상이도를 나타내는 통계량으로서 LBP 히스토그램 기반의 인식 방법에서 사용되었다[9,10,11].

먼저 각 표정에 해당하는 트레이닝 샘플들에서 블록별 LBP 히스토그램을 구하고 얼굴 전체에 대해서 연결한 다음 평균을 냄으로써 각 표정에 대한 모델 히스토그램으로 사용한다. 얼굴 영상이 입력되면 다음의 χ^2 값을 모든 모델에 대해서 계산해서 가장 작은 값을 산출하는 클래스에 속한 것으로 분류한다.

$$\chi^2(S, M) = \sum_i \frac{(S_i - M_i)^2}{S_i + M_i} \dots (4)$$

여기서 S 와 M 은 각각 입력 영상과 모델 영상의 LBP 히스토그램이다.

얼굴 영상을 블록으로 분할했을 때 눈, 코, 입을 포함하는 블록이 다른 블록들보다 얼굴 표정을 인식하기 위한 더 많은 정보를 포함한다. 따라서 그림 5 처럼 영역의 중요도에 따라 다른 가중치가 적용되고 위의 수식(4)는 다음 수식으로 수정된다[11].

$$\chi_w^2(S, M) = \sum_{i,j} w_j \frac{(S_{i,j} - M_{i,j})^2}{S_{i,j} + M_{i,j}} \dots\dots\dots (5)$$

여기서 w_j 는 j 번째 블록에 대한 가중치이다. 앞선 논문들에서 이 값들은 그림 5에 보이는 것처럼 영역의 중요도에 따라 0, 1, 2, 4의 가중치를 부여하고 있다(9,10,11). 그러나 이러한 가중치 부여방법은 직관에 의존한 것일 뿐 합리적인 근거를 가지고 있지 못하다. 3.2절에서는 블록의 가중치를 트레이닝 데이터를 이용해서 최적화함으로써 분류기의 인식률을 높이는 방법을 기술하고 있다.

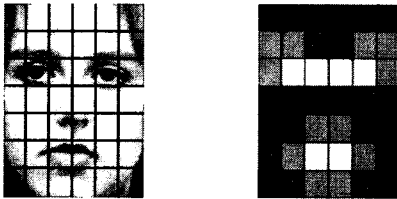


그림 5. 왼쪽 : 블록으로 분할된 얼굴영역. 오른쪽 : 밝기로 표현된 블록 가중치
Fig. 5. Left : A face image divided into blocks. Right : Brightness-coded block weights.

3.2 블록 가중치의 최적화

p 번째 입력 영상과 q 번째 모델 영상의 LBP 히스토그램을 각각 S_p, M_q 라고 할 때 j 번째 블록에서의 두 영상 사이의 상이도는 다음과 같은 χ^2 값으로 정의될 수 있다.

$$\chi_j^2(S_p, M_q) = \sum_i \frac{(S_{p,j,i} - M_{q,j,i})^2}{S_{p,j,i} + M_{q,j,i}} \dots\dots\dots (6)$$

여기서 $S_{p,j,i}$ 는 p 번째 입력영상의 j 번째 블록에서의 i 번째 LBP 히스토그램 값이다. 마찬가지로 $M_{q,j,i}$ 는 q 번째 모델영상의 j 번째 블록에서의 i 번째 LBP 히스토그램 값이다. 입력영상과 모델영상 전체에 대한 LBP 히스토그램의 χ^2 상이도는 다음과 같이 정의된다.

$$\chi^2(S_p, M_q) = \sum_j w_j \sum_i \frac{(S_{p,j,i} - M_{q,j,i})^2}{S_{p,j,i} + M_{q,j,i}}$$

이것은 블록별 상이도의 가중 합을 전체 영상의 상이도로 정의한 것이다. 임의의 q 에 대해 $\chi^2(S_p, M_k) < \chi^2(S_p, M_q)$ 이면 입력 영상은 k 번째 클래스로 분류된다.

본 논문에서는 블록 가중치 w_i 들을 제어변수로 하여 트레이닝 영상과 그 영상이 속한 클래스의 모델 영상 사이의 상이도를 가장 급격하게 줄여주는 방향으로 블록 가중치를 갱신함으로써 최적화된 가중치 값을 탐색한다.

블록 가중치 벡터를 $W = (w_1 \dots w_{m \times n})^T$ 로 정의할 때 우리가 최소화 하려고 하는 목적함수인 $\chi^2(S_p, M_q)$ 의 gradient 벡터는 다음과 같다.

$$g = \nabla \chi^2(S_p, M_q) = \frac{\partial \chi^2(S_p, M_q)}{\partial W} = (\chi_1^2(S_p, M_q), \dots, \chi_{n \times m}^2(S_p, M_q))^T$$

따라서 블록 가중치의 업데이트 룰은 다음과 같다.

$$W = W - \mu \frac{g}{\|g\|}$$

여기서 μ 는 스텝 사이즈를 결정하는 상수이다.

블록 가중치의 초기값 W_0 는 영역의 중요도에 따라 0, 1, 2, 4의 가중치를 부여한 기존 방식에서의 블록 가중치를 사용한다(그림 5). 블록별 가중치 값들을 최적화하는 gradient descent 알고리즘은 다음과 같다.

알고리즘 BLOCK_WEIGHT_OPTIMIZATION

1. 트레이닝 영상 집합 $\{(X_1, y_1), \dots, (X_N, y_N)\}$,
 X_i : 트레이닝 영상
 y_i : X_i 가 속한 클래스
 N : 트레이닝 영상의 수
2. 블록 가중치 벡터를 초기화
 $W = W_0$: 영역의 중요도에 따른 가중치 부여(본문참조)
3. For $k = 1..N$
 (a) 블록 가중치 값들의 크기를 정규화

$$w_j = \frac{w_j}{\sum_s w_s}$$

 (b) $n \times m$ 개의 각 블록에 대해서 트레이닝 영상과

그것이 속한 모델영상의 상이도를 계산.

$$\chi_j^2(S_k, M_k) = \sum_i \frac{(S_{k,j,i} - M_{k,j,i})^2}{S_{k,j,i} + M_{k,j,i}}$$

여기서 S_k 는 X_k 의 LBP 히스토그램이고, M_k 는 클래스 y_k 의 모델영상의 LBP 히스토그램이다.

(c) 블록 가중치 벡터를 업데이트

$$W = W - \mu \frac{g}{\|g\|}$$

4. 스텝 3에서 일어난 상이도의 변화량이 정해진 값보다 적으면 종료. 아니면 스텝 3을 반복한다.

IV. 실험결과

제안된 방법을 실험하기 위하여 본 논문에서는 JAFFE 데이터베이스[14]가 사용되었다. JAFFE 데이터베이스는 10명의 일본인 여성의 표정으로 구성 되어있고, 각각의 사람은 angry, disgust, fear, happiness, normal, sadness, surprise 등, 7개의 표정을 가지며 해상도는 256 x 256이다. 그림 6은 몇 개의 샘플 영상을 보이고 있다. 실험에 사용하는 영상을 64x64로 샘플링하고 leave-one-out 방법을 사용하여 실험을 수행한다. 트레이닝 과정에서 영상을 임의로 선택한 후 나머지 트레이닝 영상으로 가중치를 계산하고 임의로 선택된 영상은 인식을 하는데 사용한다. 위와 같은 방법을 10회 수행하여 전체 인식률과 표정 당 인식률의 평균을 구하여 기존의 방법과 비교한다.

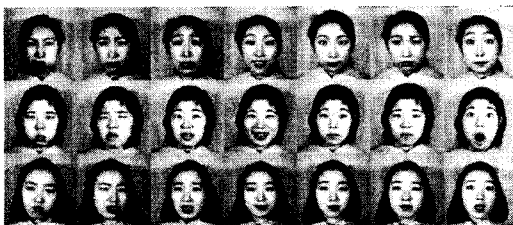


그림 6. JAFFE 데이터베이스의 샘플 영상.
Fig. 6. Sample images of JAFFE Database.

JAFFE 데이터베이스 영상을 대상으로 2회의 실험을 하였다. 알고리즘이 진행됨에 따라 총 상이도가 감소하는 것이 그림 7에 보여진다. 먼저 7개의 표정 모두를 사용한 실험과 다른 표정들과 구분이 명확하지 않은 표정(Fear)을 제외한 6

개의 표정을 사용한 실험을 시행하였다. 7 개의 표정에 대한 실험에서는 제안된 방법의 인식률이 69.1%로 기존의 방법에 의한 65.0% 보다 높았다. 6 개의 표정에 대한 실험에서는 제안된 방법의 인식률이 74.3%로 기존의 방법에 의한 69.6% 보다 높았다. 표 1과 표 2는 각각 얼굴표정 7 개와 6 개에서의 기존의 방법과 제안된 방법의 비교표이다.

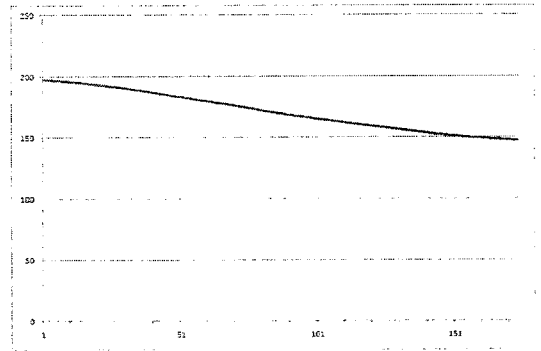


그림 7. 알고리즘 진행에 따른 총 상이도의 변화.
Fig. 7. Change of total dissimilarity through the progress of the algorithm.

표 1. 7개의 표정을 이용한 분류기의 인식률
Table 1. Recognition rates using 7 expressions.

표정	기존의 방법	제안된 방법
Angry	83.6	85.4
Happiness	73.6	76.1
Normal	73.2	73.6
Sadness	46.4	61.4
Surprise	77.1	79.3
Disgust	57.1	60.4
Fear	43.9	47.9
평균	65.0	69.1

표 2. 6개의 표정을 이용한 분류기의 인식률
Table 2. Recognition rates using 6 expressions.

표정	기존의 방법	제안된 방법
Angry	83.2	86.1
Happiness	74.3	77.5
Normal	72.5	77.5
Sadness	52.5	63.6
Surprise	78.2	80.4
Disgust	56.8	61.1
평균	69.6	74.3

V. 결론

본 논문에서는 얼굴 영상을 블록으로 분할하고 각 블록에서 구해진 LBP 히스토그램을 템플릿 매칭 방법으로 비교하는 기존의 얼굴 표정인식 방법을 개선하였다. 기존의 방법에서는 표정에 큰 영향을 미치는 눈, 코, 입과 같은 주요 부분이 포함된 블록의 가중치를 직관에 의해서 높게 결정함으로써 보다 나은 성능을 얻고자 해 왔다. 본 논문에서는 직관 보다는 트레이닝 데이터를 이용한 최적화를 통해서 블록 가중치를 결정함으로써 기존의 방법보다 성능이 상당 부분 개선된 분류기를 구현할 수 있었다.

참고문헌

- [1] Y. Tian, L. Brown, A. Hampapur, S. Pankanti, A. Senior, and R. Bolle, "Real world real time automatic recognition of facial expression," IEEE PETS, Australia, Mar. 2003.
- [2] B. Fasel and J. Luetttin, "Automatic facial expression analysis: a survey," Pattern Recognition, Vol. 36, pp. 259 - 275, 2003.
- [3] M. J. Lyons, J. Budynek, and S. Akamatsu, "Automatic classification of single facial images," IEEE Transactions on Pattern Analysis and Machine Intelligence, 1999.
- [4] I. Cohen, N. Sebe, Garg A., L. Chen, and T. Huang, "Facial expression recognition from video sequences: Temporal and static modeling," CVIU, Vol. 91, pp. 160 - 187, 2003.
- [5] M.S. Bartlett, G. Littlewort, I. Fasel, and R. Movellan, "Real time face detection and facial expression recognition: Development and application to human computer interaction," CVPR Workshop on CVPR for HCI, 2003.
- [6] Y. Tian, "Evaluation of face resolution for expression analysis," IEEE Workshop on Face Processing in Video, 2004.
- [7] Y. Tian, T. Kanade, and J.F. Cohn, "Facial Expression Analysis," Handbook of Face Recognition, Springer, Oct. 2003.
- [8] M. Lyons, S. Akamatsu, M. Kamachi, J. Gyoba, "Coding Facial Expressions with Gabor Wavelets," pp.200, Third IEEE International Conference on Automatic Face and Gesture Recognition (FG'98), 1998.
- [9] T. Ahonen, A. Hadid, and M. Pietikinen, "Face recognition with local binary patterns," ECCV, pp. 469 - 481, 2004.
- [10] A. Hadid, M. Pietikinen, and T. Ahonen, "A discriminative feature space for detecting and recognizing faces," IEEE CVPR, pp. 797 - 804, June 2004.
- [11] C. Shan, S. Gong, and P. W. McOwan, "Robust facial expression recognition using local binary patterns," IEEE ICIP, 2005.
- [12] T. Ojala, M. Pietikinen, and D. Harwood "A comparative study of texture measures with classification based on featured distribution," Pattern Recognition, Vol. 29, No. 1, 1996.
- [13] T Ojala, M Pietikäinen, T Mäenpää "Multiresolution gray scale and rotation invariant texture classification with local binary patterns" IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 971-987, 2002.
- [14] M.J. Lyons, J. Budynek, S. Akamatsu, "Automatic classification of single facial images," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 21, No. 12, pp.757-1362, 1999.
- [15] 정성태 "조명변화에 견고한 얼굴 특징 추출," 한국컴퓨터 정보학회논문지, 제 10권, 제 6호, 1-8쪽, 2005년 12월.
- [16] 김진용, 유재휘 "아바타 통신에서의 얼굴 표정의 생성 방법," 한국컴퓨터정보학회논문지, 제 10권, 제 3호, 55-64 쪽, 2005년 7월.

저자소개



박 성 천
단국대학교 전산통계학과 학사
단국대학교 컴퓨터과학전공 석사
현재 : 단국대학교 컴퓨터과학전공 박사과정
관심분야 : 컴퓨터 비전



구 자 영
서울대학교 전자공학과 학사
한국과학기술원 전기 및 전자공학과 석사
한국과학기술원 전기 및 전자공학과 박사
현재 : 단국대학교 컴퓨터학부 교수
관심분야 : 컴퓨터 비전