SOM과 PRL을 이용한 고유얼굴 기반의 머리동작 인식방법

이 우 진† · 구 자 영††

요 약

본 논문에서는 머리동작의 인식을 위한 새로운 동작 인식방법을 제안하고 있다. 먼저 각 동작을 이루는 얼굴의 자세들 사이의 높은 상관관계를 이용해서 얼굴 화상들을 주성분 분석함으로써 얼굴 화상을 축약된 벡터 형식으로 표현한다. 그 다음에 이 데이터들을 이용하여 SOM을 무감독 학습시키는데 그 결과는 유사한 자세가 인접한 노드에 반응하도록 각 노드가 학습되는 것이다. 각 모델 동작들을 이루는 자세들에 대해서 주성분 분석과 SOM 분류를 시행하여 그 결과를 데이터 베이스에 저장한다. 미지의 동작을 나타내는 임의의 프레임들이 입력되면 각 프레임에 주성분 분석과 SOM 분류를 함으로써 얻어지는 노드 열을 데이터 베이스에 저장된 모델 프레임들의 노드 열과 비교하는데, 인접하는 자세들 사이의 문맥정보를 이용하는 PRL을 적용함으로써 동작을 분류하는 방법을 제시하고 있다.

A Head Gesture Recognition Method based on Eigenfaces using SOM and PRL

Woo-Jin Lee† · Ja Young Koo††

ABSTRACT

In this paper a new method for head gesture recognition is proposed. At the first stage, face image data are transformed into low dimensional vectors by principal component analysis(PCA), which utilizes the high correlation between face pose images. Then a self organization map(SOM) is trained by the transformed face vectors, in such a that the nodes at similar locations respond to similar poses. A sequence of poses which comprises each model gesture goes through PCA and SOM, and the result is stored in the database. At the recognition stage any sequence of frames goes through the PCA and SOM, and the result is compared with the model gesture stored in the database. To improve robustness of classification, probabilistic relaxation labeling(PRL)is used, which utilizes the contextual information imbedded in the adjacent poses.

1. 서 론

인간이 문자가 아닌 다른 매체를 통해서 컴퓨터와 상호작용 하고자하는 것은 컴퓨터의 출현 이후 지금까지의 오랜 바램이었다. 컴퓨터가 인간처럼 말을 알아듣고 눈으로 봄으로써 세계를 인식한다면 인간과의 보다 자연스러운 상호작용이 가능해 질 것이며, 특히 신체적 장애를 갖은 사람들에게 유용할 것이다. 본 논문에서 다루는 것은 인간과 컴퓨터 사이의 상호작용을 위한 화상 인식의 한 문제로서 단일 화상의 인식이 아닌 일련의 화상으로 표현된 동작의 인식에 관한 것이다. 인식의 대상은 얼굴을 중심으로 한 머리의 동작으로서 전체적인 시스템의 구성은 (그림 1)과 같다.

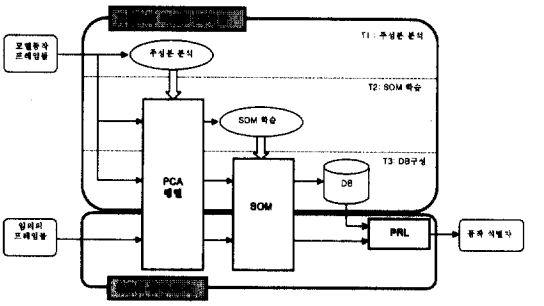
† 준 회 원 : 단국대학교 대학원 전산통계학과
†† 정 회 원 : 단국대학교 전산통계학과 교수
논문접수 : 1999년 11월 26일, 심사완료 : 2000년 3월 9일

지금까지 연구된 동작인식 기법들에는 은닉 마르코프 모델을 이용한 방법[1, 2, 3], 신경망에 근거한 방법[4, 5, 6], 또는 유한상태 기계를 이용한 방법[7, 8, 9] 등이 있다. 이러한 방법들에서 인식의 단위는 동작으로서, 일반적으로 한정된 동작들에 대해 학습을 한 후에 미지의 동작이 입력될 때 훈련된 동작들과의 정합을 통해서 인식결과를 얻는다.

우리는 Kohonen이 그의 음성인식기[10]에서 사용한 방법을 동작인식에 적용하였다. 그는 그의 연구에서 단어단위가 아닌 음소를 인식하도록 SOM(Self Organizing Map)을 학습한 다음에 입력음성을 음소로 분해하여 인식하게 함으로써 활성화되는 노드열의 패턴을 미리 기억된 단어의 모형에 정합하는 방법을 사용하였다. 본 논문에서는 동작의 기본요소인 자세를 인식하도록 SOM을 학습한 다음에 입력동작을 이루는 자세들을 인식함으로 생성되는 노드열의 패턴을 미리 기억된 동작의 모형에 정합하도록 하였다. 이 과정을 통해서 동작의 인식문제가 주어진 기호열과 가장 유사한 기호열을 저장된 기호열들 중에서 찾아내는 비교적 단순한 문제로 치환되었다.

기호열의 비교를 위해서는 Rosenfeld 등이 제안한 PRL(Probabilistic Relaxation Labeling) [11]을 사용하였는데 이는 Walts의 constraint propagation[12] 방식을 확률적으로 확장한 것이다. 입력 기호열 내의 한 기호와 모형 기호열 내의 한 기호의 유사도를 결정함에 있어서, 두 기호 사이의 유사도와 더불어 두 기호와 인접하는 기호들 사이의 관계의 유사도를 함께 고려하여 유사도를 반복적으로 갱신함으로써 입력 기호열과 가장 비슷한 기호열을 찾게 된다.

처리과정은 크게 두 부분으로 나뉘는데 하나는 오프라인 학습과정 및 데이터베이스 구성이며 다른 하나는 온라인 동작 인식이다. 오프라인 학습은 세 단계로 이루어지는데 첫째는 모델 동작 프레임들의 주성분 분석을 통한 PCA(Principal Component Analysis) 행렬을 구하는 것이고, 두 번째 단계는 주성분 분석을 통해서 낮은 차원으로 표현된 프레임들을 이용해서 얼굴 자세의 인식을 위한 SOM 신경망을 학습하는 것이다. 세 번째 단계는 SOM에 의해서 자세가 분류된 모델 동작 프레임들을 데이터 베이스로 저장하는 것이다. 온라인 동작인식 과정에서는, 부류를 알 수 없는 프레임들에 PCA를 적용하고 그 결과를 얼굴 자세들로 학습된 SOM으로 분류하여 각 프레임들에 일차적인 레이블을 부여한다. 그 다음에 각 프레임들 사이의 문맥정보를 이용하는 PRL을 사용하여 반복적으로 레이블들을 갱신하여 최종적인 동작의 분류를 수행한다.



(그림 1) 제안된 동작인식 시스템의 구조

2. PCA를 이용한 자세의 표현

객체를 인식하기 위해서는 그 객체를 표현하는 효과적인 수단이 필요한데, 얼굴 화상의 경우 입력 화소 열을 벡터로 표현한다면 대단히 높은 차원의 벡터가 된다. 예를 들어, 가로 세로가 각기 100 화소라면 10000 차원의 벡터가 됨으로 이후의 처리가 현실적으로 불가능하다. 그러나 얼굴 화상들 사이의 높은 상관관계를 이용해서 주성분 분석을 할 경우 불과 수 십 차원의 벡터만을 사용해도 대단히 적은 오차 범위 안에서 원 화상을 표현할 수 있다[13].

$m \times n$ 크기를 갖는 N개의 얼굴 화상들의 집합이 주어질 때 각 화상은 mn 차원의 일차원 벡터로 간주될 수 있다. 고유얼굴은 식 (1)과 같은 고유값 문제를 풀어서 얻어지는 고유벡터들을 말한다.

$$\Lambda = \Phi^T \Sigma \Phi \tag{1}$$

여기서 Σ는 얼굴 데이터 벡터들의 공분산 행렬이고 Φ는 Σ의 고유벡터 행렬이다. Λ는 고유값의 대각행렬이다. 얼굴 데이터의 경우 인근 화소들 사이에 상관관계가 크므로 고유값들을 크기 순으로 정렬할 때 급격한 감소를 보이고, 따라서 적은 개수의 고유벡터만을 사용하여도 적은 오차범위 내에서 원래의 화상을 표현할 수 있다.

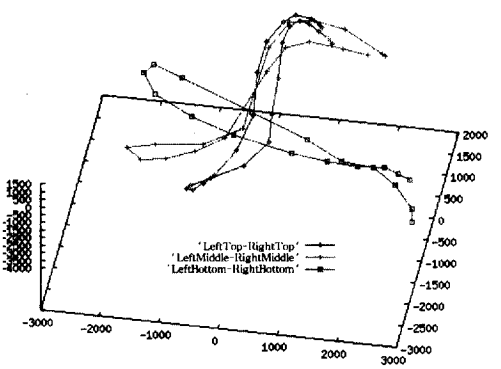
PCA에서 고유값의 크기 순으로 M개의 고유벡터를 택하면 주성분 특징 벡터 $\mathbf{y}$는 식 (2)와 같이 구해진다.

$$\mathbf{y}=\boldsymbol{\Phi}_{\mathrm{M}}^{\mathrm{T}}\ \tilde{\mathbf{x}} \tag{2}$$

여기서 $\tilde{\mathbf{x}}=\mathbf{x}-\bar{\mathbf{x}}$ 인데, $\mathbf{x}$는 얼굴 데이터 벡터이고 $\bar{\mathbf{x}}$는 그 평균값이다. $\boldsymbol{\Phi}_M$은 주성분 고유벡터로 이루어진 $\boldsymbol{\Phi}$의 부분행렬이다. 따라서 주성분 분석의 결과는 적은 정보의 손실만으로 원래의 mn차원의 얼굴 벡터를 M(M<< mn)차원의 벡터로 변환한 것이다.

3. SOM을 이용한 자세의 분류

이제 각 동작들을 구성하는 프레임들을 주성분 분석함으로써 저차원의 표현을 얻게 되었다. 그중 세 개의 동작을 처음 세 개의 주성분 공간에 그려보면 (그림 2)와 같이 각기 구별되는 궤적을 나타냄으로써 동작분류의 가능성을 보인다. 하나의 동작은 일련의 자세들로 이루어지므로 먼저 해야할 일은 주어진 프레임이 어떤 자세에 속하는지를 분류하는 일이다. 본 논문에서는 이러한 자세 분류기를 신경망의 한 종류인 SOM(Self Organizing Map)으로 구성한다. SOM은 (그림 3)과 같은 구조를 갖는 무감독 학습분류기로서 유사한 특징에 반응하는 노드들을 기하학적으로 근접한 공간에 배치하도록 학습된다. 이와 같은 SOM의 특징은 다음 과정인 PRL에서 SOM의 노드 위치가 입력으로 사용되는 점을 고려할 때 오인식률을 줄여주는 바람직한 특성이다.

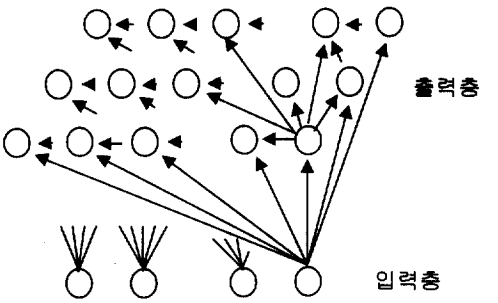


(그림 2) 처음 세 개의 주성분 공간에서의 동작의 궤적들

SOM의 학습과정에서 노드들의 연결 강도는 작은 값의 임의의 수로 초기화되고, 새로운 입력 벡터 $\mathbf{x}$가 들어오면 이와 가장 유사한 입력 연결강도를 갖는 승자 노드를 택하여 그 노드의 연결강도를 식 (3)과 같이 갱신한다[14].

$$\mathbf{w}_i(t+1)=\mathbf{w}_i(t)+\alpha(t)(\mathbf{x}-\mathbf{w}_i(t)) \tag{3}$$

여기서 $\mathbf{w}_i$는 승자 노드의 연결강도를 나타내는 벡터이고, $\alpha(t)$는 학습의 수렴을 위해 시간의 경과에 따라 감소하는 학습계수이다. 이때 SOM에서는 승자 노드뿐만 아니라 승자 노드의 "이웃"으로 정의된 주변의 노드들도 같은 방식으로 학습에 참여함으로써 자기 조직화의 특성을 나타내게 된다.



(그림 3) SOM의 구조

4. PRL을 이용한 동작의 분류

지금까지의 처리 결과, 하나의 입력동작을 구성하는 일련의 프레임들이 주성분 분석된 후, 각 노드가 얼굴의 특정 자세에 반응하도록 훈련된 SOM의 노드 열로 변환되었다. 남은 일은 이 노드 열들이 미리 저장된 동작들의 모델들 가운데 어떤 것과 가장 유사한지를 판단하는 것인데, 본 논문에서는 프레임 사이의 문맥정보를 이용하기 위하여 PRL을 사용한다. 이때 분류의 대상이 되는 객체들의 집합은 입력동작을 나타내는 프레임들의 집합으로서 A로 표기된다.

$$A=\{f_1, f_2, \cdots, f_N\}$$

각 f_k에 부여될 수 있는 레이블의 집합은 $S=\{s_{ij}\}$로서 s_{ij}는 i번째 모델 동작의 j번째 프레임을 나타내는 기호이다. 모델의 수가 m이고 m번째 모델에 속한 프레임의 개수를 z라고 하면 확률적 레이블링(probabilistic labeling)은 각 객체 f_k에 식 (4)와 같은 확률벡터를 할당하는 일이다.

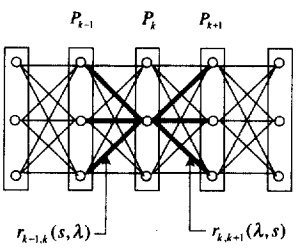
$$P_k=[P(l_k=s_{11})\ \cdots\ P(l_k=s_{mz})]^T \quad (4)$$

여기서 $P(l_k=s_{ij})$는 f_k가 s_{ij}일 확률을 나타내며 l_k는 f_k의 레이블을 나타내는 확률변수이다. PRL에서 각 P_k의 초기값 $P_k^{(0)}$가 주어지고 식 (5)와 같은 갱신규칙에 따라서 각 P_k들이 갱신되어서 반복회수가 충분히 커지면 주변 프레임들의 레이블링으로 부터 지지를 받는 레이블에 대한 확률이 지속적으로 강화되어 f_i에 유일한 레이블을 할당하게 된다.

$$P_k^{(n+1)}(l_k=\lambda)=\frac{P_k^{(n)}(l_k=\lambda)\cdot Q_k^{(n)}(l_k=\lambda)}{\sum_{s\in S}P_k^{(n)}(l_k=s)\cdot Q_k^{(n)}(l_k=s)} \quad (5)$$

여기서 $Q_k^{(n)}(l_k=\lambda)$는 $P_k^{(n)}(l_k=\lambda)$에 대한 이웃으로부터의 지지를 나타내는 것으로서 (그림 4)에 보이는 바와 같이 P_{k-1}과 P_{k+1}로부터의 영향을 받으며 식 (6)과 같이 정의된다.

$$Q_k^{(n)}(l_k=\lambda)=\frac{1}{2}\Big(\sum_{s\in S}r_{k-1,k}(s,\lambda)P_{k-1}^{(n)}(l_{k-1}=s) + \sum_{s\in S}r_{k,k+1}(\lambda,s)P_{k+1}^{(n)}(l_{k+1}=s)\Big) \quad (6)$$



(그림 4) $Q_k(l_k=\lambda)$의 계산을 위한 이웃관계

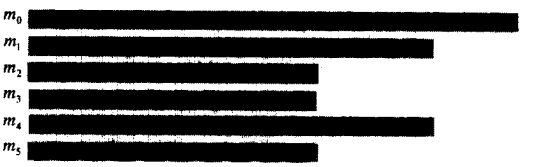
여기서 $r_{k-1,k}(s,\lambda)$는 $k-1$번째 프레임의 레이블이 s이고 k번째 프레임의 레이블이 λ인 레이블링의 적합성(compatibility)이다. 우리는 s와 λ가 다른 모델에 속한 레이블이거나 같은 모델에 속한 레이블이라 할지라도 모델 내의 순서에서 λ가 s를 앞서거나 같으면 $r_{k-1,k}(s,\lambda)=0$으로 정의하고, 그렇지 않은 경우에는 s와 λ의 모델 내에서의 프레임 차이를 d라고 할 때 $r_{k-1,k}(s,\lambda)=\frac{1}{d}$로 정의한다.

f_k에 반응한 SOM의 노드와 s_{ij}에 반응한 노드와의 거리를 d_{kij}라고 할 때 f_k와 s_{ij}의 유사도는 그들 사이의 거리에 반비례함으로 유사도를 $t_{kij}=\frac{1}{d_{kij}}$로 정의하고, 이에 따라서 $P_k(l_k=s_{ij})$의 초기값을 식 (7)과 같이 정의한다.

$$P_k^{(0)}(l_k=s_{ij})=\frac{t_{kij}}{\sum_{ij}t_{kij}} \quad (7)$$

5. 실험결과

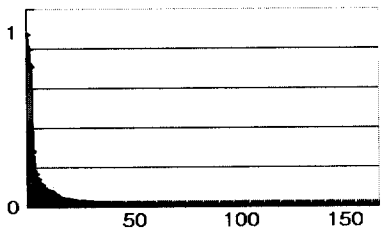
제안된 방법의 실험에는 화상입력을 위해서 SONY CCD-TR705 비디오 카메라와 Silicon Graphics사의 INDY R5000을 사용하였다. 화상은 128×96크기의 256계조화상을 사용하였다. 모델 동작은 (그림 5)에 보이는 바와 같은 6개의 동작을 사용하였다. 이 경우에 얼굴의 자세는 얼굴의 전면이 지향하는 방향벡터에 의해서 기술될 수 있는데 이러한 방향벡터는 얼굴의 상하 각도를 나타내는 앙각과 좌우 회전각을 나타내는 방위각으로 표현될 수 있다. 앞의 세 개는 얼굴의 앙각 세 가지에 대한 좌우 동작이고, 뒤의 세 개는 얼굴의 방위각 세 가지에 대한 상하동작이다. 화상입력 이후의 처리는 Pentium II 400MHz의 PC를 사용하였다.



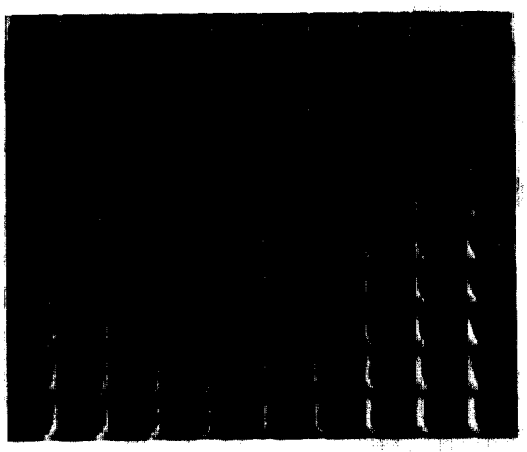
(그림 5) 모델 동작들

128×96크기의 화상들을 주성분 분석하였는데, 얼굴화상들 사이의 높은 상관관계에 의해서 고유값들을 내림차순으로 정렬할 때 (그림 6)에 보이는 바와 같이 급격한 감소를 보인다. 고유값의 크기 순으로 50개의 고유 벡터를 선택하였는데, 이렇게 함으로써 적은 재구성 오차 범위 내에서 데이터의 크기를 대폭 줄일 수 있다. 다음으로 447개의 얼굴화상들을 주성분 분석하여 SOM을 학습시켰다. 학습은 얼굴화상들로부터 10000번의 임의추출을 통해서 시행되었다. 식 (3)의 학습계수는 단조감소 함수로서 $\alpha(t)=0.5*10^{-2t/T}$, $T=10000$을 사용하였다. 그 결과는 (그림 7)에 보이는 바와 같은 자기조직화의 결과를 나타내는데, 이웃하는 노드들이 비슷한 얼굴의 자세에 반응함을 보인다. 여기에 보이는 그림들은 10×10의 SOM 노드들이 학습결과로 갖게된 PCA 계수들로 재구성된 얼굴화상들이다.

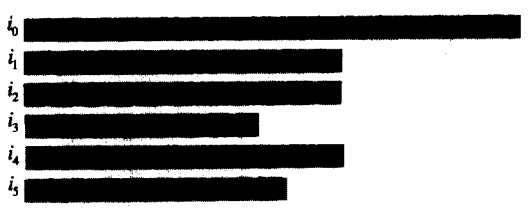
(그림 8)은 각각의 클래스에 속하는 미지의 동작 여섯 개를 보인다. (그림 9)는 PRL의 진행과정을 보여주는 것으로 P_i에 해당하는 열은 f_i에 할당된 확률벡터, 즉 확률적 레이블링을 나타낸다. 그림이 보여주는 바와 같이 초기에는 여러 모델 프레임들에 대해서 0이 아닌 확률값을 보이지만 PRL이 진행함에 따라서 이웃으로부터 강한 지지를 받는 레이블들이 점점 더 강화되어 궁극적으로는 확률값 1로 수렴하는 결정적 레이블링을 얻게되어 각각의 클래스로 인식될 수 있음을 보인다.



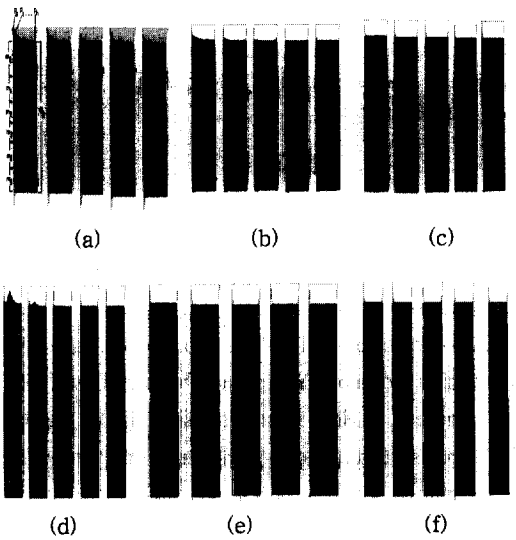
(그림 6) 최대값으로 정규화되고 정렬된 고유값들



(그림 7) 10×10의 SOM에 447개의 얼굴 화상으로 학습한 결과



(그림 8) 여섯 개의 미지의 동작들



(a) i_0에 대한 진행과정. 초기값, 1회 반복, 2회 반복, 3회 반복, 16회 반복 후 결과
(b) i_1에 대한 진행과정. 초기값, 1회 반복, 2회 반복, 3회 반복, 32회 반복 후 결과
(c) i_2에 대한 진행과정. 초기값, 1회 반복, 2회 반복, 3회 반복, 34회 반복 후 결과
(d) i_3에 대한 진행과정. 초기값, 1회 반복, 2회 반복, 3회 반복, 22회 반복 후 결과
(e) i_4에 대한 진행과정. 초기값, 1회 반복, 2회 반복, 3회 반복, 17회 반복 후 결과
(f) i_5에 대한 진행과정. 초기값, 1회 반복, 2회 반복, 3회 반복, 15회 반복 후 결과

(그림 9) 각 클래스에 속한 여섯 개의 미지의 입력 동작들에 대한 PRL에 의한 확률적 레이블의 진행

6. 결 론

본 논문에서는 인접하는 자세들 사이의 문맥정보를 이용하여 동작을 분류하는 분류기를 제안하였다. 먼저 PCA를 통해서 얼굴화상을 축약시켰으며 이 축약된 데이터로 얼굴의 자세들을 분류하는 SOM이 효과적으로 학습됨을 보였다. SOM을 통해서 분류된 임의의 동작에 대한 자세들이 모델 동작들의 자세의 순서와 완전히 일치될 수는 없으므로 인접하는 객체의 레이블들을 주변의 지지도에 의해 갱신해 나가는 PRL을 적용하여 결

정적 레이블을 얻을 수 있음을 보였다. 본 논문에서는 균일한 얼굴크기와 조명을 조건으로 하고 있는데 이러한 조건을 없애는 것이 차후의 연구과제일 것이다.

참 고 문 헌

[1] J. Yamato, J. Ohya, and K Ishi, "Recognizing Human Action in Time-Sequential Images Using Hidden Markov Model," IEEE Computer Society Conference on CVPR, Vol.1, pp.379-385, June 1992.

[2] Carlos Morimoto, Yaser Yacoob, Larry Davis, "Recognition of Head Gestures Using Hidden Markov Models," International Conference on Pattern Recognition, Austria, pp.461-465, 1996.

[3] Andrew D. Wilson, Aaron F. Bobick, "Recognition and Interpretation of Parametric Gesture," M.I.T Media Lab. TR No.421, 1998.

[4] R. J. T. Mrris, L. D. Rubin and H. Tirri, "Neural Network Techniques for Object Orientation Detection: Solution by Optimal Feedforward Network and Learning Vector Quantization Approaches," IEEE Trans. on PAMI, Vol.12, No.11, pp.1107-1114, Nov. 1990.

[5] E. Littmann, A. Drees, and H. Ritter, "Visual gesture-based robot guidance with a modular neural system," NIPS1995(Poster), Denver, Colorado, USA, November 1995.

[6] Enno Littmann, Andrea Drees, and Helge Ritter, "Visual Gesture Recognition by a Modular Neural System," Proceedings ICNN'96(W. v. Seelen, Chr. v.d. Malsburg, eds), Springer 1996.

[7] Andrew Wilson, Aaron Bobick, "Using Configuration States for the Representation and Recognition of Gesture," M.I.T Media Lab. TR No.308, 1995.

[8] Trevor J. Darrell, Irfan A. Essa, Alex P. Pentland, "Task-specific Gesture Analysis in Real-Time using Interpolated Views," M.I.T Media Lab. TR No.364, 1995.

[9] A. F. Bobick and A. D. Wilson, "A State-Based Approach to the Representation and Recognition of Gesture," IEEE Trans. on PAMI, Vol.19, No.12, pp.1325-1337, Dec. 1997.

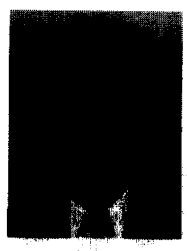
[10] Teuvo Kohonen, "The "Neural" Phonetic Typewriter," IEEE Computer, pp.11-22, Mar. 1988.

[11] A. Rosenfeld, R. Hummel, Zucker, "Scene labeling by relaxation algoritms," IEEE Trans. Syst., Man, Cyben., Vol.SMC-6, pp.420-433, 1976

[12] DL Waltz, 'Understanding line drawings of scenes with shadows,' In The Psychology of Computer Vision, McGraw-Hill, New York, 1957

[13] M. Kirby and L. Sirovich, "Application of the Karhunen-Loéve Procedure for the Characterization of Human Faces," IEEE Trans. on PAMI, Vol.12, No.1, pp.103-107, Jan. 1990.

[14] James A. Freeman, David M. Skapura, 'Neural Networks Algorithms, Applications, and Programming Techniques,' Addison-Wesley Publishing Componay, Inc. 1991.



이 우 진

e-mail : wjlee@ns.dankook.ac.kr

1990년 단국대학교 전자계산학과 졸업(학사)

1994년 단국대학교 대학원 전산통계학과 졸업(이학석사)

1994년~현재 단국대학교 대학원 전산통계학과 박사과정 수료. 단국대학교 멀티미디어산업기술연구소 연구원

관심분야 : 얼굴인식, 동작인식, GIS, 문자인식 등

구 자 영

e-mail : jykoo@dankook.ac.kr

1977년 서울대학교 전자공학과 (공학사)

1980년 한국과학기술원 전기 및 전자공학과(공학 석사)

1986년 한국과학기술원 전기 및 전자공학과(공학박사)

1986년~현재 단국대학교 전산통계학과 교수

1990년~1991년 영국 Imperial College Visiting Academic

관심분야 : 패턴인식, 그래픽스, 영상처리, 얼굴인식