

論文97-34C-9-7

HMM-Net 분류기의 학습

(On Learning of HMM-Net Classifiers)

金商雲*, 吳璣煥*

(Sang-Woon Kim and Soo-Hwan Oh)

요 약

HMM-Net은 신경망 구조로 은닉 마르코프 모델(HMM)을 구현한 것으로, HMM이 갖고 있는 시계열 모델링 능력과 신경망이 갖고 있는 우수한 변별력을 결합한 것이다. HMM-Net 분류기를 학습하는 학습기 준에는 최우도(ML), 최대 상호정보(MMI), 최소 자승오차(MMSE) 등이 있다. 이 논문에서는 세가지 학습 기준에 기울기 강하법을 적용하는 방법으로 HMM-Net 분류기를 학습하여 성능을 비교한 실험상의 연구 결과를 보고한다. /영/부터 /구/까지의 고립 숫자음을 대상으로 실험한 결과, 이진패턴으로 실험하였을 경우에는 MMSE가 우수하였고, 퍼지패턴으로 하였을 경우에는 MMI가 높은 인식률을 나타내었다.

Abstract

The HMM-Net is an architecture for a neural network that implements a hidden Markov model(HMM). The architecture is developed for the purpose of combining the classification power of neural networks with the time-domain modeling capability of HMMs. Criteria which are used for learning HMM-Net classifiers are maximum likelihood(ML), maximum mutual information (MMI), and minimization of mean squared error(MMSE). In this paper we report the results of an experimental study comparing the performance of HMM-Net classifiers trained by the gradient descent algorithm with the above criteria. Experimental results for the isolated numbers from /young/ to /koo/ show that in the binary inputs the performance of MMSE is better than the others, while in the fuzzy inputs the performance of MMI is better than the others.

I. 서 론

은닉 마르코프 모델(hidden Markov model; HMM)은 시계열 패턴의 통계적인 성질을 잘 표현할 수 있기 때문에 음성인식에 널리 이용되고 있다^[1]. 그러나 HMM은 최우도(maximum likelihood; ML) 학습기준을 사용하기 때문에 변별력이 약하다는 단점이 있고, 최대 상호정보(maximum mutual information; MMI)의 학습기준을 적용할 경우에는 계산이

복잡해지고 많은 가정을 해야 한다는 문제점이 있다^[2]. 한편 신경망의 연결 모델은 통계적인 데이터를 다루는 데는 적합하지만 연속적인 데이터 취급이 용이하지 않다. 이러한 문제를 해결하기 위한 한가지 방법으로 시계열 패턴의 시간적 특성을 잘 나타내는 HMM을 보다 변별력이 있는 신경망 구조로 구현하는 방법인 HMM-Net(hidden Markov model network)이 제안되었다^[3].

HMM-Net은 신경망 구조이기 때문에 벡터 양자화 등의 전처리 과정을 분류기에 통합하여 패턴을 입력할 때 연속값을 관측을 할 수 있다. 또한 HMM-Net은 기울기 강하법으로 학습할 수 있기 때문에 신경망의 학습에서 많이 이용하는 최소 자승오차(minimization

* 正會員, 明知大學校 컴퓨터工學科

(Dept. of Computer Engineering, Myong-Ji University)

接受日字:1997年1月22日, 수정완료일:1997年8月29日

of mean squared error; MMSE)의 학습 기준을 이용할 수 있다^[3]. 그러나 지금까지 이들 학습기준에 대한 학습식을 유도하여 실제의 데이터를 대상으로 학습기준별 HMM-Net 분류기의 학습특성을 고찰한 연구는 찾기 어려웠다. 따라서 이 논문에서는 ML, MMI, MMSE의 학습기준에 기울기 강하법을 적용하여 학습하는 이론식을 유도한 후, 이를 이용하여 실제의 음성 데이터를 대상으로 HMM-Net 분류기를 학습하여 인식성능을 비교한 실험결과를 보고한다.

우선, II절에서는 HMM의 파라미터를 학습하는 전향 알고리즘과 후향 알고리즘^[1]을 신경망으로 구현한 HMM-Net의 구조에 대하여 설명하고, III절에서는 기울기 강하법으로 HMM-Net 분류기를 학습하기 위한 학습식의 도출과정에 대하여 기술한다. 그리고 IV절에서는 고립된 숫자 음성 데이터에서 추출한 선형 예측계수(LPC)를 k-평균 알고리즘^[1]으로 벡터 양자화한 데이터를 대상으로 HMM-Net 분류기를 학습한 후 변별력의 차이를 고찰한다. 또한, 입력을 이진으로 하였을 경우와 연속으로 하였을 경우의 인식률 차이를 비교하기 위하여 퍼지 최근방 표준 알고리즘^[4]으로 관측기호의 퍼지값을 구하여 학습한 후 변별력의 차이를 고찰한다. 끝으로 V절에서 결론을 맺는다.

II. HMM-Net 분류기

HMM을 이용하여 시계열 패턴을 인식한다고 하는 것은 클래스별로 HMM을 준비한 다음 관측된 출력 기호열이 어느 HMM으로부터 출력되는 확률이 높은 것인지를 결정하는 문제이다. 즉 $\pi_i = P(s(1)=q_i)$: i 상태의 초기 확률값; $a_{ij} = P(s(t+1)=q_j | s(t)=q_i)$: i 상태에서 j 상태로 천이하는 확률값; $b_{jk} = P(v_{y_t}=v_k | s(t)=q_j)$: j 상태에서 k 번째 심볼 v_k 를 출력할 확률값; q_i : 상태 $i, i=1,2,\dots,N$; v_k : 관측기호 $k, k=1,2,\dots,M$; $s(t)$: 시간 t 에서의 상태, $t=1,2,\dots,T$; v_{y_t} : 시간 t 에서의 관측기호, $y_t=1,2,\dots,M$; N : 상태의 수; M : 관측기호의 수; T : 관측열의 길이라고 할 때 한 HMM인 λ_M 이 어떤 관측열 $O=v_{y_1}, v_{y_2}, \dots, v_{y_T}$ 를 출력할 확률 $\lambda_M(y_1^T)$ 은

$$\lambda_M(y_1^T) = P_M(y_1^T) = \sum_{i=1}^N \tilde{\alpha}_i(t) \beta_i(t) \quad (1)$$

과 같다^[1]. 여기서, $\alpha_i(t)$ 와 $\beta_i(t)$ 는 각각 전향 알고리즘과 후향 알고리즘^[1]으로 계산하며, 이 때 언더플로우를 방지하기 위하여

$$\tilde{\alpha}_i(t) = \frac{\sum_{j=1}^N \tilde{\alpha}_j(t-1) a_{ji} b_{ij}}{\sum_{j=1}^N \sum_{l=1}^N \tilde{\alpha}_j(t-1) a_{jl} b_{lj}}, \quad \tilde{\alpha}_i(0) = \pi, \quad (2a)$$

$$\tilde{\beta}_i(t) = \frac{\sum_{j=1}^N a_{ji} b_{ij} \tilde{\beta}_j(t+1)}{\sum_{j=1}^N \sum_{l=1}^N \tilde{\alpha}_j(t-1) a_{jl} b_{lj}}, \quad \tilde{\beta}_i(T) = 1 \quad (2b)$$

와 같이 정규화를 수행한다. 또한 출력확률 $\lambda_M(y_1^T)$ 은

$$\begin{aligned} A_M(y_1^T) &= \log P_M(y_1^T) \\ &= A_M(y_1^{T-1}) + \log \sum_{i=1}^N \tilde{\alpha}_i(t-1) a_{ji} b_{ij} \end{aligned} \quad (3)$$

와 같이 로그를 취한 확률 $A_M(y_1^T)$ 로서, 재귀적으로 구한다.

HMM-Net은 식(3)의 계산 과정을 신경망으로 구현한 것으로서, HMM이 파라미터를 학습하기 위하여 ML기준의 Baum-Welch 알고리즘^[1]을 이용하는 반면에 신경망 구조인 HMM-Net은 기울기 강하법을 사용할 수 있다. 식(3)에서 시간 1에서 t 까지의 관측열에 대한 출력확률 $A_M(y_1^T)$ 를 계산하기 위한 재귀 신경망 구조는 그림 1과 같다.

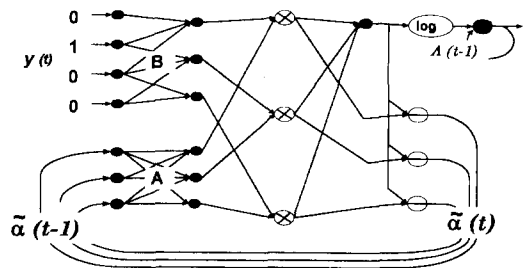


그림 1. HMM-Net 구조^[3]
Fig. 1. Structure of HMM-Net.

그림에서 ●는 각 링크의 값을 더하는 노드이고 ⊗는 두 입력을 서로 곱하는 노드이며 ⊖는 하나의 입력을 다른 입력으로 나누는 노드이다. 그리고 A와 B는 HMM의 상태천이확률과 출력기호확률에 대응하는 연결 가중치이고 A, B 이외의 모든 연결 가중치 값은 1이다. 또한 $y(t)$ 는 M 차원 이진 입력벡터로서 t 시간의 관측기호 v_{y_t} 에 해당하는 노드만 1이고

다른 노드는 모두 0의 값을 갖는다.

HMM-Net 모델 K 개를 이용하여

$$k^* = \arg \max_k \{X_k\}, \quad k=1, 2, \dots, K \quad (4)$$

의 식별규칙으로 K 개의 클래스를 식별하는 HMM-Net 분류기는 그림 2와 같다.

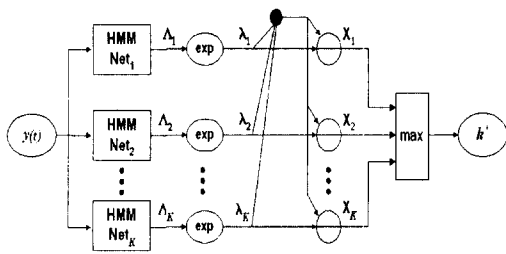


그림 2. HMM-Net 분류기 구조
Fig. 2. Structure of HMM-Net Classifier.

여기서 모든 연결 가중치의 값은 1이고, K 개의 HMM-Net 유닛에서 나온 출력은 전체의 합이 1이 되도록 정규화 된다.

III. HMM-Net 분류기 학습

HMM-Net 분류기의 학습방법에는 어떤 학습기준을 이용하는가에 따라 여러 가지 방법이 있다. 이 절에서는 ML, MMI, MMSE를 이용하여 기준함수 G 를 정의하여 놓고 학습 데이터 $\{Y_1, \dots, Y_{N_r}\}$ 를 이용하여 G 가 최대가 되도록 파라미터값을 바꾸어 가는 기울기 강하법의 학습식을 유도한다.

우선, $\lambda_k(n)$ 과 $\Lambda_k(n)$ 을 k 번째 모델에서 학습기호열 Y_n 에 대한 확률과 로그 확률이라고 할 때, ML, MMI, MMSE의 기준함수를 각각

$$\begin{aligned} G^{ML} &= \log(P(Y_1, \dots, Y_{N_r} | M_1, \dots, M_K)) \\ &= \sum_n \log P(Y_n | M_{C_n}) \\ &= \sum_n \Lambda_{C_n}(n) \end{aligned} \quad (5a)$$

$$\begin{aligned} G^{MMI} &= \sum_n \{ \log \lambda_{C_n}(n) - \log(\frac{1}{K} \sum_{k=1}^K \lambda_k(n)) \} \\ &= \sum_n \{ \log X_{C_n}(n) + \log K \} \end{aligned} \quad (5b)$$

$$G^{MMSE} = -\frac{1}{2} \sum_n \sum_k (X_k(n) - \delta_{k,C_n})^2 \quad (5c)$$

와 같이 정의한다^[3]. 여기서, M_k 는 클래스 k 에 대

한 HMM-Net 유닛이고, C_n 은 데이터 Y_n 의 출력이며, $\delta_{i,j}$ 는 크로네커 델타함수이다.

기울기 강하법은 기준함수 G 의 극점을 찾기 위하여 각각의 파라미터로 G 를 미분하여 $\partial G / \partial = 0$ 을 만족하는 파라미터를 결정하는 방법으로, 파라미터 a_{ij} 에 대한 G 의 변화량은

$$\frac{\partial G}{\partial a_{ij}} = \sum_{n=1}^{N_r} \sum_{k=1}^K \phi_{n,k} \frac{\partial \Lambda_k(n)}{\partial a_{ij}} \quad (6)$$

로 구한다^[3]. 여기서 $\phi_{n,k} = \partial G / \partial \Lambda_k(n)$ 는 가중치 함수로서, 각 학습방법에 따라

$$\phi_{n,k}^{ML} = \delta_{k,C_n} \quad (7a)$$

$$\phi_{n,k}^{MMI} = \delta_{k,C_n} - X_k(n) \quad (7b)$$

$$\phi_{n,k}^{MMSE} = X_{C_n}(n)(\delta_{k,C_n} - X_k(n)) - \sum_{k=1}^K X_k^2(n)(\delta_{k,k} - X_k(n)) \quad (7c)$$

과 같다. 따라서, HMM-Net 분류기를 학습하기 위해서는 식(6)에서 $\partial \Lambda_k(n) / \partial a_{ij}$ 를 계산하여야 한다. 이를 위하여 그림 1.에서 노드 \ominus 를 통과하여 정규화 되기 전의 값(노드 \otimes 의 출력)을

$$\hat{a}_j(t) = \sum_i \tilde{a}_i(t-1) a_{ij} b_{jy} \quad (8)$$

이라고 하면, 로그를 취하기 직전의 값(노드 \bullet 의 출력)은

$$\hat{a}(t) = \sum_j \hat{a}_j(t) \quad (9)$$

이 되고, 따라서 $t=1$ 부터 $t=T$ 까지의 입력에 대한 출력확률 $\Lambda_k(y_1^T)$ 는

$$\begin{aligned} \Lambda_k(y_1^1) &= \log \hat{a}(1) \\ &= \log \left(\sum_{i=1}^N \sum_{j=1}^N \tilde{a}_i(t-1) a_{ij} b_{jy} \right) \end{aligned} \quad (10a)$$

$$\Lambda_k(y_1^2) = \Lambda_k(y_1^1) + \log \hat{a}(2) \quad (10b)$$

$$\begin{aligned} \Lambda_k(y_1^T) &= \Lambda_k(y_1^{T-1}) + \log \hat{a}(T) \\ &= \log \hat{a}(1) + \log \hat{a}(2) + \dots + \log \hat{a}(T) \end{aligned} \quad (10c)$$

이 된다. 따라서,

$$\begin{aligned} \frac{\partial \Lambda_k(n)}{\partial a_{ij}} &= \frac{\partial \Lambda_k(y_1^T)}{\partial a_{ij}} \\ &= \frac{1}{\hat{a}(1)} \frac{\partial \hat{a}(1)}{\partial a_{ij}} + \dots + \frac{1}{\hat{a}(T)} \frac{\partial \hat{a}(T)}{\partial a_{ij}} \end{aligned} \quad (11)$$

이 되고, 여기서 $\partial \hat{\alpha}_i(t) / \partial a_{ij}$ 는 식(9)를 이용하면

$$\begin{aligned} \frac{\partial \hat{\alpha}_i(t)}{\partial a_{ij}} &= \sum_p (\hat{\alpha}_p(t-1) \frac{\partial a_{pj} b_{ij}}{\partial a_{ij}} + \frac{\partial \hat{\alpha}_p(t-1)}{\partial a_{ij}} a_{pj} b_{ij}) \quad (12) \\ &= \hat{\alpha}_i(t-1) b_{ij} + \sum_p \frac{\partial \hat{\alpha}_p(t-1)}{\partial a_{ij}} a_{pj} b_{ij} \end{aligned}$$

이 된다. 여기서 $\partial \hat{\alpha}_p(t-1) / \partial a_{ij}$ 는 식(2)와 식(8), (9)로부터

$$\hat{\alpha}_i(t) = \frac{\sum_{j=1}^N \hat{\alpha}_i(t-1) a_{ij} b_{ij}}{\sum_{j=1}^N \sum_{k=1}^N \hat{\alpha}_i(t-1) a_{ij} b_{ij}} = \frac{\hat{\alpha}_i(t)}{\hat{\alpha}_i(t)} \quad (13)$$

이므로

$$\frac{\partial \hat{\alpha}_i(t)}{\partial a_{ij}} = \left(\frac{\partial \hat{\alpha}_p(t)}{\partial a_{ij}} \hat{\alpha}_i(t) - \hat{\alpha}_p(t) \frac{\partial \hat{\alpha}_i(t)}{\partial a_{ij}} \right) / \left(-\frac{\partial \hat{\alpha}_i(t)}{\partial a_{ij}} \right)^2 \quad (14)$$

이 된다. 즉, 그림 1.의 노드 \otimes 와 노드 \bullet 의 출력값인 $\hat{\alpha}_p(t-1)$ 와 $\hat{\alpha}_i(t-1)$ 를 식(14)에 대입하여 $t-1$ 시간에서의 a_{ij} 변화에 따른 $\hat{\alpha}_p(t-1)$ 의 변화량을 구한다. 이를 식(12)에 대입하여 t 시간에서의 노드 \otimes 의 출력값 변화 $\partial \hat{\alpha}_i(t) / \partial a_{ij}$ 를 구하고, 다시 식(9)로부터 노드 \bullet 의 출력값 변화 $\partial \hat{\alpha}(t) / \partial a_{ij}$ 를 구한다. 이러한 과정을 재귀적으로 수행하여 시간 $t=1$ 부터 T 까지의 a_{ij} 변화에 따른 $\hat{\alpha}(t)$ 의 변화량 $\partial \hat{\alpha}(t) / \partial a_{ij}$ 를 구한다. 다음, 이를 식(11)에 대입하여 $\partial \Lambda_k(n) / \partial a_{ij}$ 를 계산한다. 같은 방법으로 $\hat{\alpha}(t)$ 의 b_{ij} 에 대한 변화량 $\partial \hat{\alpha}(t) / \partial b_{ij}$ 도 구할 수 있으며, 이를 이용하여 $\partial \Lambda_k(n) / \partial b_{ij}$ 를 계산한다.

한편, 입력이 그림 1.에서와 같은 이진값이 아니라 모든 입력 노드를 통하여 연속값으로 들어올 경우 $v_{m(t)}$ 를 t 시간에 m 번째 노드로 들어온 입력이라고 하면 식(12)의 $\partial \hat{\alpha}_i(t) / \partial a_{ij}$ 는

$$\begin{aligned} \frac{\partial \hat{\alpha}_i(t)}{\partial a_{ij}} &= \sum_p \sum_m (\hat{\alpha}_p(t-1) \frac{\partial a_{pj} b_{im} v_{m(t)}}{\partial a_{ij}} + \frac{\partial \hat{\alpha}_p(t-1)}{\partial a_{ij}} a_{pj} b_{im} v_{m(t)}) \\ &= \sum_m \hat{\alpha}_i(t-1) b_{im} v_{m(t)} + \sum_p \sum_m \frac{\partial \hat{\alpha}_p(t-1)}{\partial a_{ij}} a_{pj} b_{im} v_{m(t)} \quad (15) \end{aligned}$$

이 된다. 같은 방법으로 $\hat{\alpha}(t)$ 의 b_{ij} 에 대한 변화량 $\partial \hat{\alpha}_i(t) / \partial b_{ij}$ 도 구할 수 있다. 이상과 같은 방법으로 HMM-Net을 구성하는 각 파라미터 변화에 대한 기준함수의 변화량을 구한 다음, 이 변화량이 최소가 되

도록 파라미터 값을 갱신하는 처리를 반복하여 HMM-Net 분류기를 학습한다.

IV. 실험 및 결과 고찰

1. 실험 방법

이 절에서는 HMM-Net 분류기를 ML, MMI, MMSE의 학습기준에 따라 학습한 후 인식성능을 비교하는 실험방법 및 실험결과에 대하여 고찰한다. 우선, HMM-Net 분류기를 III.절에서 유도한 학습식으로 학습하여 학습기준별로 성능을 비교하고, Baum-Welch법으로 학습한 HMM 분류기와도 비교한다. 먼저 이진값을 사용한 실험결과를 고찰하고 다시 퍼지값을 사용한 경우를 고찰한다.

실험 데이터는 전자통신연구소에서 배포한 음성 데이터베이스^[5]의 전화 음성자료 중에서 /영/, /일/, /이/, /삼/, /사/, /오/, /육/, /칠/, /팔/, /구/의 고립 숫자음으로 모두 9명의 남자가 각기 4번씩 발음한 360개의 데이터를 이용하였다. 이 데이터는 저역통과 필터를 이용하여 3.4KHz로 필터링하고 8KHz로 샘플링한 후 16비트로 양자화 한 데이터이다. 이 데이터에서 앞과 뒤의 소리가 나지 않는 묵음 부분을 제거한 후 10차 LPC계수를 특징으로 추출하여 이용하였다.

실험방법은 실험 데이터에 대하여 전표본 학습법(resubstitution; R법), 분할 학습법(holdout; H법), 상호 교정법(cross-validation; CV법)^[1]으로 실험하였다. 이렇게 R법, H법, CV법으로 실험한 이유는 제한된 학습 데이터에 기인한 한계점을 극복하여 학습방법을 정확히 평가하기 위해서다. 즉, R법은 성능을 실제보다 낙관적으로 평가하는 방법으로 알려져 있고 H법은 실제보다 비관적으로 평가하는 방법으로, 그리고 CV법은 두 방법을 중화시킬 수 있는 방법으로 알려져 있기 때문이다.

R법은 모든 데이터를 학습과 테스트에 이용하는 방법으로, 전체 숫자음 데이터를 대상으로 분류기를 학습한 후 다시 이 데이터를 대상으로 테스트하였다. 그리고 H법에서는 4회 발음한 데이터 중에서 처음 2회는 학습 데이터로 나머지 2회는 테스트 데이터로 사용하였으며, 학습 데이터와 테스트 데이터를 바꿔 다시 실험하였다. 또한 CV법에서는 학습 데이터로 3회의 데이터를 사용하고 나머지 1회의 데이터로 테스트하는 방법으로 4가지 경우의 수를 모두 테스트하였다.

실험을 위한 전처리 과정은 다음과 같다. 먼저, A/D 변환된 음성 데이터에서 10차 LPC계수를 추출한 후, LPC계수를 다시 k-평균 알고리즘으로 12개의 군집으로 클러스터링하여 그림 1.의 y(t)로 입력하였다. 이때 이진 데이터로 만들기 위해서는 최소 거리에 있는 군집 중심의 기호를 1로 하는 방법을 사용하였고, 퍼지 입력 데이터로 만들기 위해서는 식(16)의 퍼지 최근접 표준(fuzzy nearest prototype) 알고리즘^[4]을 사용하였다.

$$u_i(x) = \frac{1 / \|x - Z_i\|^{2/(m-1)}}{\sum_{j=1}^C (1 / \|x - Z_j\|^{2/(m-1)})}, \quad i=1, 2, \dots, C \quad (16)$$

여기서, $u_i(x)$ 는 미지패턴 x 가 클래스 i 에 속하는 정도를 나타내는 소속함수이며, C 는 클래스의 수, z_i 는 클래스 i 의 표준패턴, m 은 실험상수이다. 이 논문의 실험에서 x 는 LPC계수에 해당하고, $u_i(x)$ 는 LPC계수의 관측기호 v_i 에 해당한다. 그리고 C 는 관측기호의 수 M 에 해당하며, z_i 는 k-평균 알고리즘으로 구한 군집중심에 해당한다.

그리고 HMM-Net 분류기 학습의 학습률에 해당하는 식(7a), (7b), (7c)는 학습기준 G 를 계산하는 방법에 따라 서로 다르기 때문에 경험적인 방법으로 결정하였다. 즉, 학습률은 ML과 MMI에서는 0.01을, MMSE에서는 0.02를 사용하였고, 학습횟수는 최고 500회까지로 하였으며 G 값이 떨어지지 않고 진동하여도 학습을 종료시켰다. 식(16)의 실험상수 m 은 1.1로 하였다.

2. 결과 고찰

HMM-Net 분류기의 실험 결과를 고찰하기 위하여, 우선 분류기를 구성하는 HMM-Net 출력의 정규화를 고찰하고, 이진입력과 퍼지입력에 대한 인식성능을 고찰하며, 끝으로 기준함수와 학습방법과의 관계에 대하여 고찰한다.

우선, HMM-Net의 정규화를 고찰한다. HMM분류기의 경우 Baum-Welch 알고리즘으로 학습하기 때문에 파라미터 값들은 확률값으로서 합이 1이고 음수가 되지 않는 성질이 계속 유지되지만, HMM-Net 분류기의 경우 기울기 강하법을 사용하기 때문에 파라미터 값이 양수로서 합이 1이 된다는 보장이 없다. 따라서 HMM-Net 분류기에서 파라미터를 조정할 때 마다 합이 1이 되고 음수가 되지 않도록 정규화를 시

켜주어야 한다. 예를 들어, /칠/의 이진 및 퍼지 입력에 대하여 ML, MMI, MMSE의 학습기준별 각 클래스에 해당하는 HMM-Net 유닛의 출력값은 표 1.과 같다.

표 1. 숫자음 /칠/ 에 대한 HMM-Net 유닛의 출력값

Table 1. Outputs of HMM-Net units for number /chil/.

유닛 출력	이진입력의 경우			퍼지입력의 경우		
	ML	MMI	MMSE	ML	MMI	MMSE
X ₀	0.000	0.062	0.045	0.000	0.032	0.034
X ₁	0.031	0.073	0.083	0.035	0.069	0.049
X ₂	0.008	0.064	0.048	0.037	0.069	0.068
X ₃	0.020	0.049	0.060	0.027	0.073	0.082
X ₄	0.021	0.043	0.030	0.042	0.010	0.025
X ₅	0.000	0.001	0.007	0.000	0.001	0.005
X ₆	0.000	0.065	0.089	0.001	0.028	0.024
X ₇	0.711	0.569	0.558	0.798	0.624	0.640
X ₈	0.210	0.075	0.077	0.062	0.092	0.075
X ₉	0.000	0.000	0.003	0.000	0.000	0.000
Σ	1.001	1.001	1.000	1.002	0.997	1.002

표 1은 숫자음 /칠/의 이진입력에 대하여 ML기준을 적용하였을 경우, /칠/에 해당하는 유닛의 출력 X₇ 값은 0.711이고 다른 유닛의 출력 X_i ($i \neq 7$) 은 0.000 부터 0.031까지의 값으로서, 입력된 /칠/의 변별정도를 나타낸다. 그리고 유닛의 출력 합계 Σ는 각 학습기 준별로 값이 0.997, 1.000, 1.001, 1.002 로서 정규화 수행정도를 나타낸다. 또한 이진입력의 경우와 퍼지입력의 경우 정규화를 비교하여 나타내고 있다.

다음으로, 이진입력에 대한 HMM-Net 분류기의 ML, MMI, MMSE 학습기준별 인식률과 HMM 분류기의 인식률은 표 2와 같다.

표 2에서는 ML학습기준을 제외하고는 HMM-Net 분류기가 HMM 분류기의 성능보다 우수함을 나타낸다. 또 같은 HMM-Net 분류기 중에서는 MMSE기준의 인식률이 가장 양호함을 나타내고 있다.

R법 실험의 경우 MMSE로 학습하였을 경우가 다른 학습기준 보다 성능이 양호함을 보여준다. 여기서 ML의 경우 HMM-Net 분류기의 인식률이 HMM보다 저조한 것은 HMM-Net의 학습시 가중치를 확률

값으로 보고 매회 학습시 정규화 하는 반면 가중치 변화량은 정규화를 하지 않았기 때문인 것으로 보인다. 반면에 MMI와 MMSE의 경우에는 학습패턴이 속하지 않는 다른 클래스도 고려하여 학습하였기 때문에 ML보다 인식률이 우수한 것으로 보인다.

표 2. 이진입력에 대한 HMM-Net 분류기의 인식률

Table 2. Classification rates of HMM-Net for binary inputs.

실험 방법	HMM-Net			HMM
	ML	MMI	MMSE	ML
R법	81.4(2.9)	89.2(3.0)	91.1(3.2)	84.4(1.7)
H법	72.2(4.5)	77.7(8.5)	73.8(10.0)	76.4(3.5)
CV법	78.6(4.2)	81.1(4.7)	79.7(5.5)	76.7(6.4)

인식률 % (표준편차)

H법과 CV법 실험의 경우에 MMSE가 MMI보다 저조한 인식률을 보이고 또 인식률의 편차가 커 견고성이 떨어지는 것으로 나타났다. 이것은 MMSE가 학습이 잘 이루어 질 때에는 MMI보다 좋은 결과를 보이지만 그렇지 않을 때에는 최적 학습이 이루어지지 않아 현저하게 낮은 인식률을 보인 예가 있었기 때문이다. 따라서 MMI로 학습한 가중치를 초기값으로 이용하여 다시 MMSE로 학습하였을 경우 MMI보다 인식률이 양호하였다. 또한 표 2.는 ML은 인식률에서는 떨어지지만 다른 방법들보다 견고성이 우수하다는 점을 나타내고 있다.

다음으로, 퍼지값을 사용했을 경우의 인식결과는 표 3과 같다.

표 3. 퍼지입력에 대한 HMM-Net 분류기의 인식률

Table 3. Classification rates of HMM-Net for fuzzy inputs.

실험 방법	HMM-Net		
	ML	MMI	MMSE
R법	79.6(3.6)	88.6(3.3)	91.0(3.0)
H법	75.8(3.1)	82.1(3.5)	79.1(7.8)
CV법	79.2(6.2)	83.9(3.4)	79.7(3.4)

인식률 % (표준편차)

표 3에서는 R법을 제외하고는 퍼지값을 입력으로 사용할 경우 이진입력을 사용하는 경우보다 나은 인식률을 보여주고 있다. 학습에 참여한 패턴만으로 테스트하는 R법의 경우는 퍼지값을 사용한 경우에 비슷하게나 오히려 좋지않은 인식률을 보이고 있으나, 학습에 참여하지 않은 패턴으로 테스트를 하는 H법과 CV법에 있어서는 뚜렷하게 나은 성능을 보여주고 있다.

이러한 특성은 다음과 같은 사실에 기인한 것으로 사료된다. 즉 이진값의 관측기호를 사용하여 학습할 경우 LPC계수와 같은 벡터를 양자화 할 때 특징공간 상에서 가까운 거리에 위치한 벡터라 할지라도 다른 관측기호로 양자화되면 분류기에서는 전혀 다른 기호로 취급되므로 실제 학습패턴과 비슷한 패턴이라하더라도 올바르게 인식되기 어려워진다. 그러나 퍼지값을 사용할 경우에는 서로 다른 기호로 양자화된 벡터라 하더라도 가까운 거리의 패턴들은 각 기호들이 적당한 퍼지값을 갖게되기 때문에 더 유연하게 인식할 수 있게 된다.

끝으로, 기준함수와 학습방법과의 관계를 고찰한다. 학습을 종료하였을 때 기준함수는 최대값을 유지해야 한다. 따라서 양호한 학습방법은 최소의 학습으로 최대의 기준함수 값을 탐색할 수 있는 방법이 된다. R법 실험에서 500회로 학습을 종료하였을 때 학습방법별로 기준함수 값을 계산한 결과는 표 4와 같다.

표 4. 학습방법에 따른 기준함수 값

Table 4. Values of criterion function G.

학습방법	기준함수 G값			인식률
	G^{ML}	G^{MMI}	G^{MMSE}	
ML	-15.667	1.900	-0.1236	0.814
MMI	-21.927	1.957	-0.0846	0.892
MMSE	-22.017	1.917	-0.0765	0.911

표 4의 첫째 칸이 의미하는 내용은 다음과 같다. 즉 기준함수를 G^{ML} 로 하여 ML로 학습하였을 경우 학습 종료시점에서의 기준함수 값은 -15.667이 되고, MMI로 학습하였을 경우는 -21.927, MMSE의 경우에는 -22.017이 됨을 나타낸다. 따라서 기준함수를 G^{ML} 로 하였을 경우에는 ML법이 다른 학습방법보다 좋은 방법이 된다. 또한 둘째 칸과 셋째 칸은 기준함수 G^{MMI} 에는 MMI가, 그리고 G^{MMSE} 에는 MMSE가 가장 적합한 학습방법임을 나타낸다.

V. 결 론

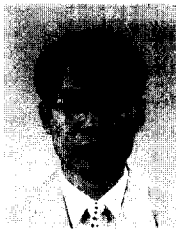
HMM-Net은 신경망 구조로 은닉 마르코프 모델(HMM)을 구현한 것으로, HMM이 갖고 있는 시계열 모델링 능력과 신경망이 갖고 있는 우수한 변별력을 결합한 것이다. 이 논문에서는 최우도(ML)와 최대 상호정보(MMI), 최소 자승오차(MMSE) 등의 학습기준에 기울기 강하법을 적용하여 HMM-Net 분류기를 학습하는 방법을 고찰하였다. 또한 고립 숫자음을 대상으로 분류기의 입력을 이진으로 하였을 경우와 퍼지값을 사용하였을 경우의 인식률을 비교하였다.

실험 결과, 이진입력의 경우 신경망에서 많이 사용되는 MMSE 학습기준을 사용하였을 때 좋은 인식률을 나타내었으며, 퍼지값을 입력으로 사용하였을 경우에는 MMI가 좋은 인식률을 나타내었다. 퍼지입력의 경우 MMSE로 학습할 때 효율적으로 학습되지 않는 경우도 발생하였으며, 이는 가중치의 왜곡으로 인하여 HMM-Net이 최적으로 학습되지 못했기 때문으로 여겨진다. 따라서 앞으로 MMSE로 HMM-Net을 최적으로 학습하는 방법과 보다 인식률이 좋아지도록 퍼지 입력값을 결정하는 방법이 연구되어야 한다.

참 고 문 헌

- [1] 김상운, 패턴인식 입문, 홍릉과학출판사, pp. 137-152, 1995.
- [2] Chris J. Wellekens, "Improved hidden Markov models for speech recognition through neural network learning", in *From statistics to Neural Networks*, V. Cherkassky, et al. edit, Springer-Verlag, pp. 226-242, 1993
- [3] Les T. Niles and Harvey F. Silverman, "Combining hidden Markov model and neural network classifiers", *Proceedings of Inter. Conf. Acoust., Speech, Signal Process*, pp. 417-420, April 1990
- [4] James M. Keller, Michael R. Gray and James A. Givens, "A fuzzy k-nearest neighbor algorithm", *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 15, no. 4, pp. 580-585, 1985
- [5] 한국전자통신연구소 음성언어연구실, ETRI 음성 DB 배포 설명서, 1995.

저 자 소 개



金 商 雲(正會員)

1956년 3월 13일생. 1978년 2월 한국항공대학교 통신정보공학과(공학사). 1980년 2월 연세대학교 대학원 전자공학과(공학석사). 1988년 2월 연세대학교 대학원 전자공학과(공학박사). 1984년 4월 ~ 1989년 3월 한국방송대학교 전자계산학과 조교수. 1992년 12월 ~ 1993년 12월 일본 홋카이도대학 정보공학과 Post-Doc. 1989년 4월 ~ 현재 명지대학교 컴퓨터공학과 부교수. 주관심분야는 패턴인식 및 학습, 미디어공학



吳 珣 煥(正會員)

1972년 3월 22일생. 1995년 2월 명지대학교 전자계산학과(공학사). 1997년 2월 명지대학교 대학원 컴퓨터공학과(공학석사). 1997년 3월 명지대학교 대학원 박사과정 입학. 주관심분야는 패턴인식 및 학습, 미디어공학

기본 요소분류기를 이용한 가상현실용 실시간 동적 손 제스처 인식 시스템의 구현에 관한 연구

(On-line dynamic hand gesture recognition system for virtual reality using elementary component classifiers)

金鍾成*, 李瓊洙**, 宋慶俊*, 卞增男***

(Jong-Sung Kim, Chan-su Lee, Kyung-Joon Song, and Zeungnam Bien)

요 약

본 논문에서는 가상 현실에서의 동적 손 제스처를 인식하는 시스템의 구현에 관하여 상술한다. 동적 손 제스처는 컴퓨터와 제스처를 사용하는 사람과의 상호 연결 수단이다. 사람의 손과 손가락은 사람마다 같은 동작이라도 변화가 다양하며 같은 동작을 반복해서 할 때에도 다른 데이터를 얻게되는 시간에따른 변화도 존재한다. 또한, 손가락의 외형 및 물리적 구조가 사람마다 다르기 때문에 다른 두사람에의해 만들어진 같은 손 모양도 일반적인 센싱장비에의해 측정될 때 다른 측정값을 나타낸다. 본 논문에서는 동적 손 제스처를 퍼지 이론을 도입한 특징 해석에의해 기본 요소인 손의 운동을 분류하고 퍼지 최대-최소 신경망을 적용하여 손의 모양을 분류함으로써 전체 손 제스처를 인식하는 시스템을 제안한다.

Abstract

This paper presents a system which recognizes dynamic hand gestures for virtual reality(VR). A dynamic hand gesture is a method of communication for a computer and human who uses gestures, especially both hands and fingers. Since the human hands and fingers are not the same in physical dimension, the same form of a gesture produced by two persons with their hands may not have the same numerical values which are obtained through electronic sensors. In this paper, we apply a fuzzy min-max neural network and feature analysis method using fuzzy logic for on-line pattern recognition.

I. 서 론

인간의 손은 주변의 환경과 상호 교류하기 위한 가장 유용하고 다양한 수단의 하나로써 특히, 중

래에 키보드 또는 마우스를 이용하던 사람과 컴퓨터사이의 상호연결과정에 사람이 두손으로 표현하는 자유로운 몸짓을 직접 이용하려는 시도가 추구되고 있다. 손 운동 자체를 인식하는 분야의 연구는 크레인 조종을 위한 수신호 인식, 지휘자의 지휘동작 인식, 수화(sign language) 인식 등을 대상으로 미국, 일본 등지에서 활발히 추진되는 중이다^[1-5].

특히, 가상현실(virtual reality) 연구분야에서 사용하는 원격 조작 시스템에서 기존의 3차원 제어 장치는 상당한 자유도를 가진 장비를 제어하기 위해 많은 전문성을 요구한다. 이와 같은 일들을 효

* 正會員, 韓國 電子通信 研究院 人工知能研究室 (Artificial Intelligent Sect., ETRI)

** 正會員, 시스템 工學 研究所 假想現實 研究室 (VR Lab., SERI)

*** 正會員, 韓國 科學技術員 電氣 및 電子工學科 (Dept. of Elec. Eng., KAIST)

接受日: 1996年12月4日, 수정완료일: 1997年8月18日

과적으로 하기 위해 사용자와 기계가 3차원 정보를 교환하기 위한 새로운 입력장비가 필요하게 되었고 결국 사람의 손 제스처에 상당한 관심이 집중되었다^[6]. 즉, 인간의 손은 실제 세계(real world)와 교감하는 유용한 툴(tool)로써 뿐만 아니라 가상환경(virtual environment)에 직접 손의 사용을 가능하게 하였다.

사람이 손으로 표현하는 자세(gesture)는 손가락의 공간상의 고정된 모양으로 의미를 나타내는 정적인 자세(static posture)와, 손과 손가락의 공간적·시간적 움직임에 의해 의미를 나타내는 동적인 자세(dynamic gesture)로 구분된다.

본 논문에서는 사용자가 가상환경에서 특정한 손 제스처 명령을 수행하면 그 결과를 실시간으로 인식하는, 장갑장치에 기반을 둔 인식 시스템을 구현하고자 한다. 전체 시스템의 실시간성을 고려하여 본 논문에서는 손의 움직임을 감지하는 장비로 Virtual Technologies사의 CyberGloveTM 시스템을^[7] 사용하였다.

정적인 손 자세를 인식할 때 사람마다 손의 크기 및 골격구조가 다르고 동적인 손의 운동의 경우에도 사람마다, 시간에 따라 매번 손의 운동 속도, 운동 크기 등이 다르므로 손의 자세를 전자적으로 감지한다는 측면에서 데이터 집합의 변화 정도는 매우 심하다. 동적 손 제스처의 인식에 관한 연구는 미국 수화(American Sign Language)를 대상으로 몇 가지 연구가 있었다. Fels^[3]는 미국 수화에서 선택한 몇 가지 손 제스처를 인식하기 위해 역전파 신경망을 이용하였는데, 동작의 시작과 끝의 명확한 구분이 필요하며, 특히 많은 학습 시간을 필요로 하는 시스템이다. 인공 신경망에서의 과도한 학습시간을 줄이기 위해 Waldron^[5]은 자기구성 Map(Self Organizing Map)에 의한 재배열법을 제안하였다. 이 방법은 초기 집단의 인식을 위해 약간의 학습시간은 필요하지만 새로운 집단이 추가되었을 때의 과도한 학습시간을 줄일 수 있는 장점이 있지만 전체 인식율의 저하가 발생한다.

이와 같은 어려움을 극복하기 위하여 본 논문에서는 퍼지 이론을 이용한 손의 운동 방향 분류기와 P. Simpson^[8]이 제안한 퍼지 최대-최소 신경망을 이용하여 손 모양 분류기를 개발하였다. 이

패턴 분류기는 새로운 집단(class)을 학습할 수 있고 기존 집단의 정보를 없애지 않으면서 온라인으로 빠르게 집단을 분류할 수 있었다. 이와같은 두 개의 패턴 분류기를 결합하여 실시간으로 동적 손 제스처를 인식할 수 있으며 시스템의 큰 변화 없이 인식 집단을 확장할 수 있다.

II장에서는 본 논문의 인식 대상인 가상 현실에서의 가상교육용 손 제스처 명령들에 대한 해석 및 인식 알고리즘을 설명하고, III장에서 인식 알고리즘을 포함한 손 제스처 인식 시스템의 구성을 설명한다. IV장에서는 실제 실험을 통하여 전체 시스템의 운영 결과를 보이며, 마지막으로 V장에서 결론을 맺는다.

II. 가상현실에서의 손 제스처 인식

자연스러운 일상생활에서 사용하는 손 제스처(손짓)는 대부분의 경우 양손을 사용하지만 본 논문에서는 단지 한 손에 의한 손 제스처를 다룬다. 가상 사회에서의 가장 대표적인 서비스의 하나인 가상교육(virtual education) 환경에서 사용자가 컴퓨터에게 명령을 내리거나 물체를 움직이기 위해 반드시 존재해야 하는 요소가 손 제스처 인식이다. 예를 들어, 가상공간상의 물체(Avatar)를 조정하는 역할을 사용자의 손 제스처 명령에 의해 수행하게 된다.

가장 사용자에게 친숙한 인터페이스는 모든 사람이 이해하는 비언어적인(non-verbal) 통신 수단을 도입하는 것이지만 3차원 공간에서 움직이는 영상의 고속 처리는 설명 비디오 카메라를 사용하더라도 매우 어렵다. 본 논문에서는 손 제스처의 데이터를 감지하는 장치로 최근에 가상현실 분야에서 주요한 입력 장치로 사용되고 있는 CyberGloveTM^[7] 시스템을 이용하였는데, 이는 각 손가락 관절의 굽힘 정도를 실시간으로 측정할 수 있다. 또한 장갑의 손등에는 손의 3차원 공간상의 자세(x, y, z, yaw, pitch, roll)를 측정하는 Polhemus센서가 부착되어 있다. 모두 13개의 손에 대한 정보(손가락 굽힘, 손의 위치)가 1/15초마다 측정되어지고 실시간 그래픽과 함께 인식된 결과가 컴퓨터에 표시된다. 본 논문의 좌표축은 CyberGloveTM 시스템과 관련된 것으로 수평방향의 좌,

우 운동은 x축, 수직방향의 상, 하 운동은 y축, 그리고 z축은 앞, 뒤 방향의 운동으로 정의되어 있다.

일반적으로 손의 움직임에 의해 표현되는 동작은 매우 다양하지만 이러한 많은 표현 가능한 동작을 체계적으로 정리한 것이 한글 수화(Korean Sign Language)라고 할 수 있다. 본 논문에서는 가상 현실용 손 제스처로 정의된 10가지 명령에 대한 동작들을 정의하기 위해 한글 표준사전^[9]을 참고하였다. 그림 1은 본 논문에서 인식하고자 하는 손 제스처에 대한 10가지 명령을 보인 것이다. 여기서 '정지'는 가상 공간상의 제어 대상 물체(Avata)를 멈추게 하는 것이고, '추적'은 현재의 손의 위치를 가상 공간상에서 똑같이 움직이도록 하는 것을 의미하며, '잡기'는 가상 공간상의 물체를 잡는 명령이다. 이상의 3가지 동작은 사용자가 임의의 방향으로 움직일 수 있도록 하였다. 본 논문에서는 사용자가 명령을 내릴 때의 편의를 위해 손의 회전은 고려하지 않고 있으며 손 제스처는 다음과 같이 표현된다.

$$G_k = (x_k, y_k, z_k, t_k, F_{hi}), 1 \leq k \leq K: \text{time step}[1/15\text{sec.}]$$

x_k, y_k, z_k : 손의 위치 정보
 t_k : 동작의 시작부터의 시간 정보
 F_{hi} : flex angles of right hand, $i=1, \dots, 10$.

(1)

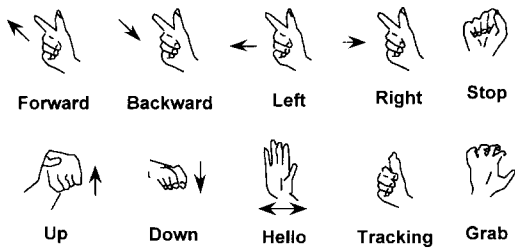


그림 1. 10가지 손 제스처 명령 집단
Fig. 1. 10 gestures for VR.

서론에서 설명한 기존의 동적 손 제스처 인식 시스템은 손 제스처 전체의 요소를 입력으로 사용한 단일의 분류기를 사용하였지만, 본 연구에서는 동적 손 제스처를 두 개의 기본적인 요소(basic elements)로 분류하였는데, 즉, 손의 운동과 손의 모양에 의해 손 제스처를 구성할 수 있다. 본 논문에서는 복잡한 손 제스처를 손 운동 분류기(classifier)와 손 모양 분류기로 구분하여 두 개의

인식 결과를 결합하여 동적 손 제스처 인식기를 구성하고자 한다. 이와 같은 방법으로 전체 인식 시스템을 구성할 경우 새로운 손 제스처를 인식기에 추가할 경우 인식 동작의 확장이 쉽게 가능하다. 그림 1을 인식기 측면에서 분석한 후 기본적인 손 운동 방향으로 그림 2와 같은 7가지의 방향 집단을 얻었고 손의 모양은 그림 3과 같이 모두 6 종류의 집단이 존재한다. 앞에서 설명한 바대로 손의 모양은 본 시스템의 특성상 회전에 관계없이 같은 집단으로 분류하게 된다.

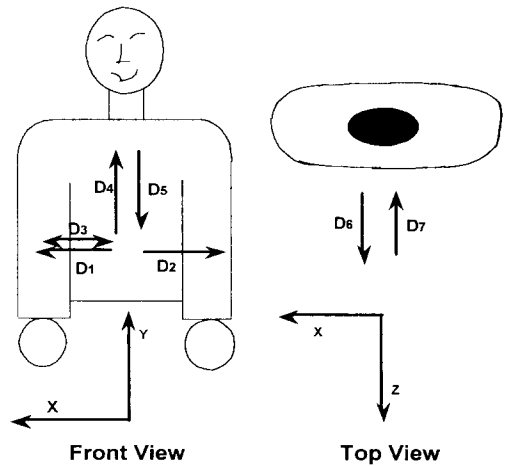


그림 2. 7가지의 기본 방향 집단
Fig. 2. 7 Basic direction classes.

손 운동은 3차원 공간에서 이루어지기 때문에 사람마다 같은 동작이라도 변화가 다양하며 같은 동작을 반복해서 할 때에도 다른 데이터를 얻게되는 등 시간에 따른 변화도 함께 존재한다. 또한, 손가락의 외형 및 물리적 구조가 사람마다 다르기 때문에 다른 두사람에 의해 만들어진 같은 손 모양도 일반적인 센싱장비에 의해 측정될 때 다른 측정값을 나타낸다. 그림 4(a)는 두사람(*와 'o'로 표시, *와 'o'로 표시된 것은 샘플된 데이터)에 의해 센싱된 상향 운동(D_4)의 궤적을 보인 것인데, 각각의 크기(운동 궤적의 길이)나 운동의 속력(샘플링 사이의 거리)이 다름을 알 수 있다. 그림 4(b)는 각기 다른 두 종류의 손의 모양(P_4, P_5)에 대한 센싱 데이터를 나타낸 것인데 각 손가락마다 2개씩의 관절(첫번째, 두번째)에 대한 굽힘 정보를 0과 1사이로 정규화(normalization)하여 표시한 것

이다.

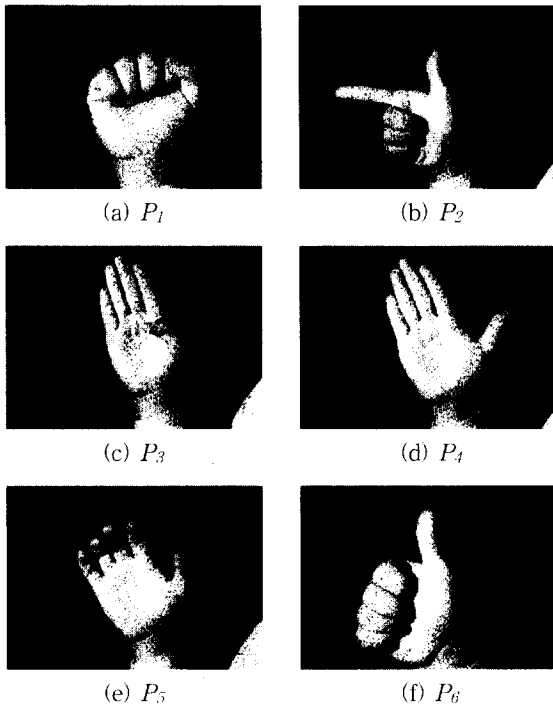


그림 3. 6가지의 기본 손 모양
Fig. 3. 6 Basic hand postures.

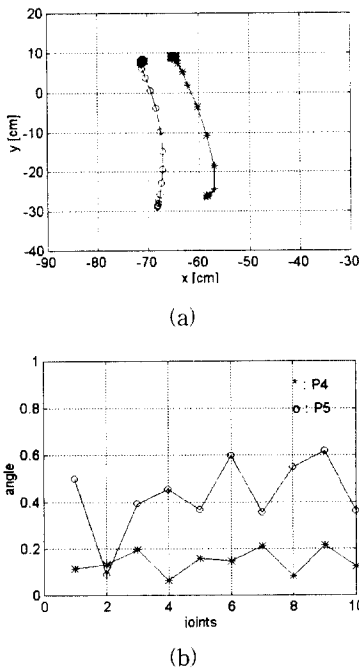


그림 4. 손 제스처 데이터의 다양성
Fig. 4. Variations of dynamic hand gestures.

이와 같이 다양한 데이터의 특성을 나타내고 있는 동적 손 제스처의 인식을 위해 본 논문에서는 6가지의 손 모양은 P. K. Simpson이 제안한 퍼지 최대-최소 신경망 분류기^[8]를 사용하고 7가지 손 운동의 인식은 퍼지 이론을 이용한 특징 해석 방법(feature analysis method)에 의해 이루어진다.

1. 퍼지 이론을 이용한 특징해석법(Feature Analysis Method)에 의한 손 운동의 분류

각 특징들은 다음의 조건에 의해 선정되었다. 즉, 특징치는 입력 데이터에 대해 정해진 시간(실 시간) 내에 계산 가능해야하고 인식뿐만 아니라 운동의 해석에도 사용 가능해야 한다. 또한 모든 운동 형태(7가지)를 분리할 수 있을 만큼의 특징들이 존재해야한다. 여기에 사용된 특징들은 가장 최근의 상대 위치, 손 운동의 누적된 운동 거리, 운동 방향 변화이다. 다음에서 이들 특징에 대해서 상술한다.

(1) 동작 끝 위치(relative Ending Position)

동작 시작점을 기준으로한 동작 끝의 상대적 위치를 나타낸다.

$$EP_{x,y,z} = R_{x,y,z}(k) - R_{x,y,z}(0) \quad (2)$$

$R_{x,y,z}(0)$: 동작 시작위치
 k : time step(1/15sec.)

(2) 누적 운동 거리(Cumulative length)

손 제스처의 운동 시작이후 매 순간의 각 축에 대한 각각의 이동 거리를 누적한 값을 나타낸다.

$$CL_{x,y,z}(k) = CL_{x,y,z}(k-1) + L_{x,y,z}(k) \quad (3)$$

$L_{x,y,z}(k) = |R_{x,y,z}(k) - R_{x,y,z}(k-1)|$

(3) 방향 변화(Change of Direction) 횟수

왕복 운동의 경우(D_3)를 고려하여 x축에 대한 방향 변화의 횟수를 표시한다.

$$CD_x = CD_x + f(x)$$

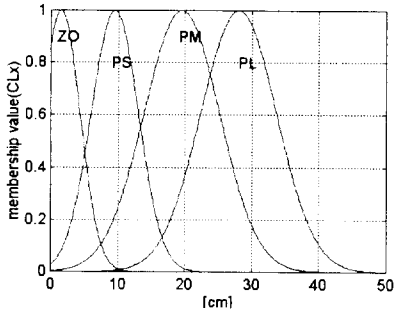
$$f(x) = \begin{cases} 1, & \text{if } [R_x(k) - R_x(k-1)] \times [R_x(k-1) - R_x(k-2)] < 0 \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

이상의 특징들을 이용하여 7개의 기본 방향을 분류하기 위해 퍼지논리를 이용한 분류기를 설계하였다. 즉, 각 손 방향에 대하여 반복 실험을 통해서 특징들을 언어적 변수로 표현하는 규칙을 만들고 이들에 대해 Mamdani의 Max-Min 추론^[10]을 하도록 하였다.

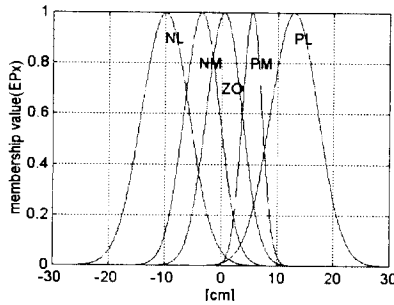
$$A \text{ AND } B = \{(x, \mu_{A \text{ AND } B}(x)) | \mu_{A \text{ AND } B}(x) = \min\{\mu_A(x), \mu_B(x)\}, x \in X\}$$

$$A \text{ OR } B = \{(x, \mu_{A \text{ OR } B}(x)) | \mu_{A \text{ OR } B}(x) = \max\{\mu_A(x), \mu_B(x)\}, x \in X\}$$

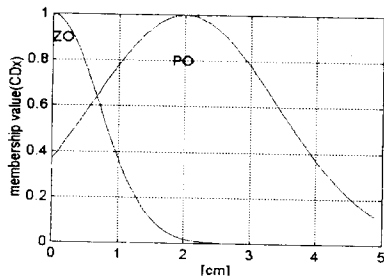
이렇게 하여 구해진 특징 벡터의 영역 구분을 통하여 소속 함수를 구성하였다. 그림 5는 x축에 대한 소속 함수를 표시한 것으로서 실제 구현에 있어서 계산 속도와 각 동작의 범위 영역을 정하기 용이한 종 모양의 소속 함수를 사용하였다.



(a)



(b)



(c)

그림 5. 소속함수의 구성
Fig. 5. Fuzzy Membership Function.

퍼지 패턴 클래스는 앞에서 정의된 AND, OR 연산을 이용하여 구현한다. 실제 이 연산을 사용하여 정의된 패턴 클래스의 예를 살펴보면

D_1 (수평 우향운동)의 경우

언어적 기술: 수평 방향의 오른쪽 방향으로의 운동이 있으며 수직 방향이나 전후 방향의 운동이 거의 발생하지 않으며 시작점을 기준으로 수평 우측 방향에서 동작이 끝난다 그리고 동작 중에 방향 변화는 일어나지 않는다.

퍼지 명제 논리식에 의한 표현은

$$(CL_x = PM \text{ or } CL_x = PS) \text{ and } (CL_y = ZO) \text{ and } (EP_x = PM \text{ or } EP_x = PL) \text{ and } CD_x = ZO \text{ and } CL_z = ZO$$

이고 다른 손 운동에 대해서도 이와 같은 방법으로 작성할 수 있다.

이와 같이 하여 D_1 에서부터 D_7 까지에 대한 방향 집단의 퍼지 규칙을 구현 할 수 있다. 표 1은 이와 같이 작성된 클래스를 규칙 테이블로 보여 주고 있다.

표 1. Fuzzy Rule Table
Table 1. Fuzzy Rule Table.

운동 방향	규 칙
D_1	$(CL_x = PM \text{ or } CL_x = PS) \text{ and } (CL_y = ZO) \text{ and } (EP_x = PM \text{ or } EP_x = PL) \text{ and } CD_x = ZO \text{ and } CL_z = ZO$
D_2	$(CL_x = PM \text{ or } CL_x = PS) \text{ and } (CL_y = ZO) \text{ and } (EP_x = NM \text{ or } EP_x = NL) \text{ and } CD_x = ZO \text{ and } CL_z = ZO$
D_3	$(CL_x = PM \text{ or } CL_x = PL) \text{ and } (CL_y = ZO \text{ or } CL_y = PS) \text{ and } (EP_x = ZO \text{ or } EP_x = PL \text{ or } EP_x = NL) \text{ and } CD_x = PO$
D_4	$CL_x = ZO \text{ and } (CL_y = PS \text{ or } CL_y = PM) \text{ and } (EP_x = PM \text{ or } EP_x = PL) \text{ and } CL_z = ZC$
D_5	$CL_x = ZO \text{ and } (CL_y = PS \text{ or } CL_y = PM) \text{ and } (EP_x = NM \text{ or } EP_x = NL) \text{ and } CL_z = ZO$
D_6	$(CL_x = PM \text{ or } CL_x = PS) \text{ and } (CL_y = ZO \text{ or } CL_y = PS) \text{ and } (EP_x = PM \text{ or } EP_x = PL) \text{ and } CL_z = ZC$
D_7	$(CL_x = PM \text{ or } CL_x = PS) \text{ and } (CL_y = ZO \text{ or } CL_y = PS) \text{ and } (EP_x = NM \text{ or } EP_x = NL) \text{ and } CL_z = ZC$

2. 퍼지 최대-최소 신경망 분류기에 의한 손 모양의 분류

앞에서 설명한 것처럼 사람의 손은 손가락의 굽힘에서 다양한 변화가 존재하고 사람들마다 다양한 크기, 모양의 손가락 구조를 가진다.

본 논문에서는 그림 3에 보여진 6가지의 기본 손 모양을 P. K. Simpson의 퍼지 최대-최소 신경망 분류기^[8]를 이용하여 실시간으로 분류하고자 한다. 사용된 소속 함수는 다음과 같이 표시할 수 있고 그림 6은 전체 분류기의 구조를 나타낸 것이다.

$$b_i(F, V, W_i) = \frac{1}{10} \sum_{j=1}^{10} [1 - f(F_j - w_{ij}, \gamma) - f(v_{ij} - F_j, \gamma)]$$

$$f(x, \gamma) = \begin{cases} 1, & xy > 1 \\ xy, & 0 \leq xy \leq 1 \\ 0, & xy < 0 \end{cases} \quad (5)$$

여기서, $F_j, j=1,3,\dots,9$ 는 손가락의 첫 번째 관절의 굽힘 정도이고 $F_j, j=2,4,\dots,10$ 은 두 번째 관절의 굽힘 정도를 나타내고 있다. γ 는 소속함수의 민감도(sensitivity)를 나타내는 것으로서 γ 가 커지면 소속함수는 점점 크리슌(crisp)하게 되므로 본 시스템에서는 첫 번째 관절의 γ 가 두 번째 관절의 γ 보다 작게 하였다. 그림 6에서 입력으로는 각 손가락의 첫 번째와 두 번째 관절의 굽힘정도($0^\circ - 90^\circ$)를 정규화(normalization, $[0,1]$)하여 사용하고 있으며, 이 분류기의 출력은 6가지 집단에 대한 각각의 소속정도($[0,1]$)를 나타내고 역치(δ , threshold)를 초과하는 최대값을 가진 집단이 현재 입력된 손 모양의 소속집단이 된다.

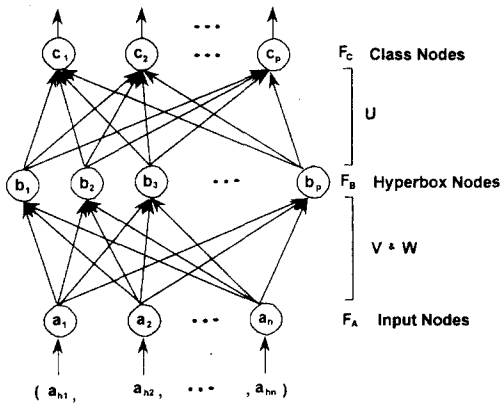


그림 6. 퍼지 최대-최소 신경망의 구조^[8]
Fig. 6. The structure of Fuzzy Min-Max Neural Network^[8].

III. 손 제스처 명령의 인식

사용자(signer)가 손 제스처 명령을 시작함에 따라 CyberGlove™로부터 많은 데이터가 시스템에 입력되게 되고 위에서 설명한 특징 해석법은 이러한 데이터를 특징을 포함하는 작은 데이터 집단으로 만들고 기본적인 7가지 방향을 인식하기 위해 사용되어진다. 모든 동작의 시작은 3차원 공간상의 임의의 위치에서 이루어 질 수 있으며 특징 추출과정에서 미소한 움직임은 무시되기 때문에 가상 현실공간에서 작업할 때 생기는 사용자의 약간의 손 떨림, 작은 손의 움직임 등은 무시할 수 있게 된다. 본 논문에서는 이러한 미소 움직임을 반복실험을 통하여 1인치 이하의 움직임으로 설

정하였다. 일단 미소 범위 이상의 동작이 시작되면 특징 추출과정이 진행되고 이후 동작의 정지가 생기면 비로소 퍼지 추론이 진행되고 현재의 손의 움직임에 대한 소속함수 값을 얻게 된다. 그림 1에 있는 '정지', '추적', '잡기' 명령의 경우는 손의 방향에 상관없이 움직임만을 확인할 필요가 있는 명령이므로 7가지의 방향에 대한 정보가 의미가 없다. 일단 현재의 동작에서 방향이 정해지면 이때 손의 모양에 대한 정보가 앞에서 설명한 퍼지 최대-최소 신경망 분류기에 입력되어 6가지 손 모양 중의 하나로 분류가 된다. 이 두개의 결과를 결합함으로써 온라인으로 동적 손 제스처 동작을 인식하게 된다. 그림 7에서 전체 인식 알고리즘의 구성을 보여준다.

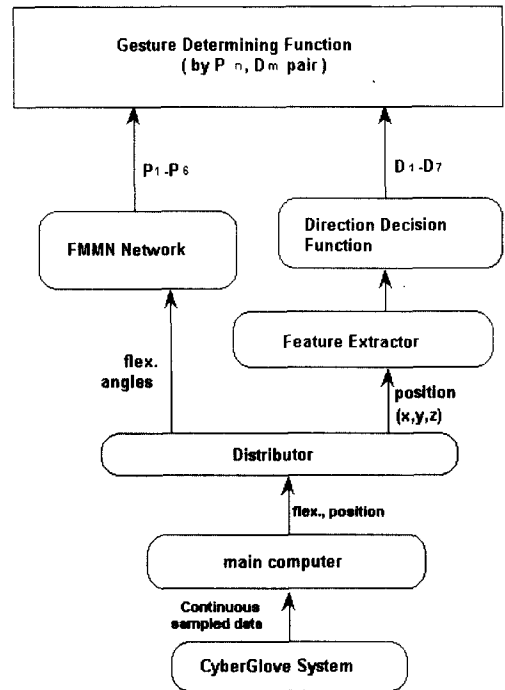


그림 7. 손 제스처 인식 알고리즘의 구성
Fig. 7. A block diagram of overall system.

IV. 실험결과

그림 8은 본 논문의 실험을 위한 가상 현실용 손 제스처 인식 시스템의 구성을 나타낸 것이고 그림 9는 실제 모습을 보인다. 전체 시스템은 주(main) 컴퓨터로 INDIGO2 워크스테이션(workstation)을 사용하였고 CyberGlove™는 38,400

baud rates로 RS232 통신을 통해 1/15초(sampling time) 마다 주 컴퓨터에 13종류의 데이터(10개의 손 굽힘, 3개의 위치)를 보낸다. 전체 시스템은 매 샘플링 시간마다 온라인으로 동작된다.

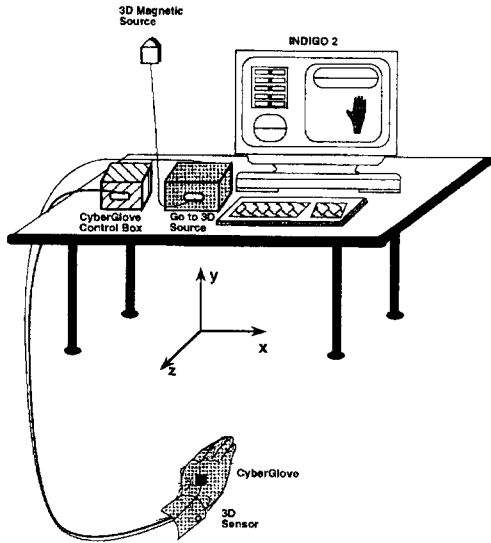


그림 8. 손 제스처 인식 시스템의 구성
Fig. 8. The configuration of the gesture recognition system.

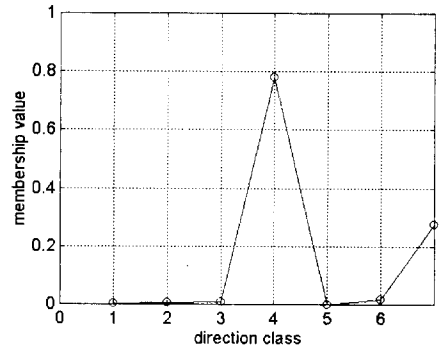


그림 9. 손 제스처 인식 시스템의 외관
Fig. 9. The appearance of the gesture recognition system.

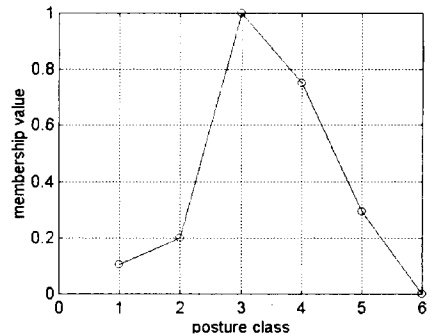
전체 인식시스템은 크게 4가지의 단계로 나누어진다. 첫째는 CyberGlove™를 착용한 사용자가 동작의 시작 지점을 설정한다. 이때, CyberGlove™로부터의 데이터가 특징 데이터를 변환된다. 두번째는 7가지의 기본 방향이 퍼지 이론을 이용한 특징 해석법에 의해 분리되고 세번째에서 6가지의 손모양에대한 분류가 이

루어진다. 마지막으로 동적 손 제스처의 결과가 위 2개의 분류결과를 결합함으로써 음성 및 그래픽으로 표현된다.

그림 10은 엄지손가락을 굽히고 손바닥을 위로 향하면서 손을 위쪽으로 움직이는 'UP'이라는 동작에 대한 인식 결과를 보인 것이다. 그림 10(a)는 손의 방향에 대한 특징 데이터를 이용한 퍼지 추론의 결과를 나타낸 것이다. 3장에서 설명한 퍼지 추론에 방법에 의해 4번째 방향집단이 최대의 소속함수값을 표시하고 방향이 'D₄'로 정해지고 곧이어 이때의 손가락 정보가 FMMNN에 입력되어 그림 10(b)에 보여진 결과를 얻는다. 그림 10(b)에서 수평축에 있는 숫자는 그림 3에 있는 상응하는 손 모양의 숫자와 관계된다. FMMNN에서 사용된 변수는 $\theta = 0.2$, $\gamma_1 = 4.0$, $\gamma_2 = 8.0$ 이고 $\delta = 0.90$ 이다. 그림 10(b)에서 보는 바와 같이 3번째 집단(P_3)이 역치(δ)를 초과하는 최대값을 가지고 따라서 두 결과를 이용하여 이 동작이 'UP'임을 정할 수 있다.



(a)



(b)

그림 10. 실험 결과 'UP'
Fig. 10. Experimental result on gesture 'UP'.

10가지의 손 제스처 명령에 대하여 5명의 다른 실험자를 대상으로 10회의 반복 실험을 수행하였고 97.6%의 평균 인식율을 얻었다. 표2는 각 제스처 명령에 대한 인식율 결과를 나타낸 것이다. 손 제스처 명령 '오른쪽으로' 경우는 오른쪽방향(D_1)을 뒤쪽방향(D_7)으로 잘못 인식한 경우가 대부분이었으며 사용자가 우측 뒤쪽으로 움직인 경우가 많았다. 또한 '추적' 명령의 경우는 P_6 (그림 3 참조)을 P_1 으로 잘못 인식한 경우이다. 이와 같이 인식 오류는 주로 이용자의 손 제스처 동작 속지의 미숙, 동작 중에 발생한 동작 미숙, 그리고 비주기적인 센서의 오동작 등에서도 기인하고 있다.

표 2. 손 제스처의 인식율

Table 2. Recognition rate of hand gestures.

명령	앞으로	뒤로	좌로	우로	정지	위로	아래로	인사	추적	잡기	계(%)
방향분류기 (오차 %)	0	2	2	8	0	0	0	6	0	0	18
손 모양 분류기 (오차 %)	0	0	0	0	0	2	0	0	4	0	06
전체 인식 성공률 (%)	100	98	98	92	100	98	100	94	96	100	97.6

V. 결론 및 추후과제

본 논문은 컴퓨터와 인간의 상호 접속에서 가장 중요한 매개체가 되고 있는 손 제스처의 인식을 위해 복잡한 사람의 손 제스처 중에서 가장 체계적인 한글 수화에서 선택한 10가지 손 제스처를 대상으로 인식 대상의 확장이 용이한 기본 요소(손의 방향, 손의 모양) 분류기에 의한 인식 알고리즘을 개발하였다. 본 논문에서는 퍼지 이론을 이용한 특징 해석법에 의해 기본 7가지 방향을 실시간으로 분류하고 퍼지 최대-최소 신경망을 이용하여 기본 6가지 손 모양을 분류하여 온라인으로 손 제스처를 인식하는 시스템을 구현하였다.

본 연구는 서론에서 언급한 일반적인 손 제스처의 인식 시스템(수화 인식, 수신호 등) 구현에도 쉽게 응용 가능하며, 앞으로 인식가능 동작의 확대를 위한 기본 요소의 확장 및 정적인 동작과의 결합을 통해 연속적으로 인식할 수 있는 시스템 및 양손에 의한 손 제스처의 확장 시스템을 구현하고자 한다. 또한 인간의 인식방법과 유사한 시

각장치를 이용한(vision based)인식방법에 대한 연구도 본 연구의 결과와 결합하여 진행할 것이다.

참고 문헌

- [1] David J. Sturman, *Whole-Hand Input*, Doctoral Dissertation of MIT, Feb. 1992
- [2] D. Morris, *Manwatching : A Field Guide to Human Behaviour*, Grafton, 1978
- [3] S. Sidney Fels and G. E. Hinton, "Glove-Talk: A Neural Network Interface Between a Data-Glove and a Speech Synthesizer," *IEEE Trans. Neural Networks*, vol. 4, no. 1, pp. 2-8, Jan. 1993
- [4] T. Yamaguchi and et al., "Japanese sign language recognition system using information infrastructure," in *Proc. IEEE Int. Conf. on Fuzzy System, FUZZ-IEEE'95*, pp. 65-66, Mar. 1995
- [5] M. B. Waldron and S. Kim, "Isolated American sign language recognition system for deaf persons," *IEEE Trans. on Rehabilitation Engineering*, vol. 3, pp. 261-271, Sept. 1995
- [6] D. Kessler, L. Hodges, et. al, "Evaluation of the cybergloveTM as a whole hand input device," *ACM Trans. on CHI*, vol. 2, pp. 263-283, Dec. 1995
- [7] Virtual Technologies, Inc., *CyberGloveTM User's Manual*, CA, 1995
- [8] P. Simpson, "Fuzzy Min-Max Neural Networks-Part 1 : Classification," *IEEE Trans. Neural Networks*, vol. 3, pp. 776-786, Sep. 1992
- [9] 김 승국, *표준 수화사전*, 오성 출판사, 1993.
- [10] 이광형, 오길록, *퍼지 이론 및 응용*, 홍릉 과학 출판사, 1991.
- [11] 한국 청각 장애인 복지회 편저, *사랑의 수화 교실*, 수협사, 1995.
- [12] J. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, New York: Plenum Press, 1981
- [13] D. Sturman and D. Zeltzer, "A Survey of Glove-based Input," *IEEE Computer Graphics and Applications*, vol. 14, no. 1, pp. 30-39, Jan. 1994

저 자 소 개

金 鍾 成(正會員) 第 34卷 C編 第 2號 參照

1997년 2월 ~ 현재 한국 전자통신연구원 멀티미디어 연구부 인공지능연구실 선임 연구원. 주관심 분야는 패턴 인식, Machine Vision, Intelligent System(Fuzzy, 인공 신경망), 가상 현실, 재활 공학(Rehabilitation Engineering) 등임



宋 慶 俊(正會員)

1956년 1월 7일생, 1982년 2월 명지대학교 공과대학 전자공학과 졸업(학사). 1984년 2월 명지대학교 공과대학 전자공학과 졸업(석사). 1985년 1월 ~ 현재 한국 전자통신연구원 인공지능연구실 선임연구원. 주관심분야는 분산 가상현실, 휴먼 인터페이스, 멀티미디어 등임



李 瓊 洙(正會員)

1970년 2월 9일생, 1995년 2월 연세대학교 전자공학과 졸업(학사). 1997년 2월 한국 과학기술원 전기 및 전자공학과 졸업(석사). 1997년 3월 ~ 현재 시스템 공학연구소 가상현실 연구실 연구원. 주관심 분야는 제스처 인식, 휴먼 컴퓨터 인터페이스, 가상현실, 지능 시스템 등임

卞 增 男(正會員) 第 30卷 B編 第 10號 參照

현재 한국 과학기술원 전기 및 전자공학과 교수

論文97-34C-9-9

곡선 조각의 군집화에 의한 둥근 물체의 효과적인 인식

(An Efficient Recognition of Round Objects Using the Curve Segment Grouping)

成 孝 慶 * , 崔 興 文 *

(Hyo-Kyung Sung and Heung-Moon Choi)

요 약

본 논문에서는 곡선조각들의 군집화에 의해 부분적으로 가려진 둥근 물체를 인식할 수 있는 효과적인 방법을 제안하였다. 영상으로부터 내부 곡률과 국소 대비를 이용하여 곡선조각들을 분할하였으며, 이 과정에서 일부 가린 물체 및 가려진 물체의 테두리들을 각각 다른 곡선조각으로 분할하였다. 분할된 곡선조각들 중 일정한 내부 곡률을 가지는 조각들을 조각간 곡률에 근거한 군집화 알고리즘을 이용하여 동일한 둥근 테두리로 군집화하였다. 군집화된 테두리의 1차 및 2차 모멘트를 이용하여 둥근 테두리에 가장 근접한 타원의 인수를 추정함으로써 화소단위로 처리하는 종래의 방법보다 적은 계산량 및 메모리로 일부 가려진 둥근 물체까지도 인식 가능케 하였다. 사과 및 호박이 있는 실 영상에 대한 실험을 통하여 제안한 방법이 둥근 물체 인식에 효과적임을 확인하였다.

Abstract

Based on the curve segment grouping, an efficient recognition of round objects from partially occluded round boundaries is proposed. Curve segments are extracted from an image using a criterion based on the intra-segment curvature and local contrast. During the curve segment extraction, The boundaries of partially occluding and occluded objects are segmented to different curve segments. The extracted segments of constant intra-segment curvature are grouped into a round boundary by the proposed grouping algorithm using inter-segment curvature which gives the relationships among the curve segments of the same round boundary. The 1st and the 2nd order moments are used for the parameter estimation of the best fitted ellipse with round boundary, and then recognition is performed based on the estimated parameters. The proposed scheme processes in segment unit and is more efficient in computational complexity and memory requirements those that of the conventional scheme which processed in pixel units. Experimental results show that the proposed technique is very efficient in recognizing the round objects from the real images with apples and pumpkins.

I. 서 론

컴퓨터 비전을 이용하는 자동화 또는 이와 유사

* 正會員, 慶北大 電子電氣 工學部

(The School of Electronics and Electrical Engineering, Kyungpook National University)

接受日字:1997年2月12日, 수정완료일:1997年9月1日

한 응용분야에서 우리가 관심을 갖는 인공 물체 (man-made objects)나 자연적인 물체(natural objects) 들 중에는 둥근 테두리를 가지고 있는 것들이 많으며, 이러한 둥근 물체의 인식 및 추출된 물체 위치 등은 자동조정(automatic manipulation) 또는 물체의 선별이나 등급을 위한 자동 검사(automatic inspection)등에 널리 이용되고 있다^[8,9].

둥근 물체를 추출하는 데는 Hough 변환을 많이 이용하고 있다^[1-3]. Hough 변환은 영상의 전역적인 성질을 사용하기 때문에 가려짐(occlusion)과 잡음에 강하다. 그러나 이 방법은 추출하려는 변수의 수가 증가함에 따라 메모리 요구가 지수적으로 증가하기 때문에 두세 개 이하의 적은 변수로 정의될 수 있는 곡선에 효과적이다^[2,3]. 만약 추출하려는 대상으로 큰 물체와 조금 작은 물체가 공존할 경우에는, 누적공간(accumulator space)에서 첨두치(peak value)를 선택하는 방법에 따라 작은 물체에 대한 변수는 무시(target masking)되어 버리기도 한다. 따라서 메모리 요구량이 적으면서도 가려진 둥근 물체들을 효과적으로 추출할 수 있는 방안이 요구된다. 한편 Jose 등^[4,5] 등은 산업용 부품의 검사를 위하여 추출된 테두리들의 반복적인 분리 및 병합(split-and-merge)방법과 최적의 다각형 근사화 방법을 혼합하여 2단계로 2차원 곡선들을 추출하였으나, 물체가 부분적으로 가려질 경우 물체별로 올바른 곡선조각의 분리가 쉽지 않다.

본 논문은 내부 곡률 및 국소 대비를 사용하여 추출된 곡선조각(curve segment)들의 군집화를 통하여 부분적으로 가려진 둥근 물체를 인식할 수 있는 효과적인 방법을 제안하였다. 먼저, 처리 시간을 줄이기 위하여 내부 곡률과 국소 대비를 이용하여 곡선조각들을 추출하였으며, 가려짐을 처리하기 위한 곡선조각의 분기점(break-point)를 설정하였다.

둥근 물체를 구성하는 곡선조각들을 연결하기 위하여 일정한 곡률을 가지는 조각들을 조각간 곡률을 사용하여 동일한 둥근 테두리로 군집화하였으며, 군집화된 테두리의 1차 및 2차 모멘트를 이용하여 둥근 테두리에 가장 적합한 타원 변수를 추정함으로써 둥근 물체를 인식하였다.

제안된 방법에서는 기존 방법들과는 달리 곡선조각 단위로 처리하므로 계산량 및 메모리량이 적으며 가려진 둥근 물체인식도 가능하다. 이를 계산복잡도 및 메모리 요구량 등을 기존 방법과 비교하고, 둥근 과일 및 채소가 있는 영상에 대하여 인식 실험하고 그 결과를 검토·고찰하였다.

II. 둥근 물체의 인식

본 논문에서는 한정된 이웃에서 국소적으로 분리한 곡선조각내에서 국소적으로 구하는 곡률을 내부 곡률

(intra-segment curvature)이라 하고, 이웃의 범위를 확장하여 곡선조각간의 관계로부터 구한 곡률을 조각간 곡률(intersegment curvature)이라 정의하였다.

제안한 둥근 물체의 인식에서는 그림 1에서 보는 바와 같이 관심영역의 내부곡률과 국소대비를 사용하여 곡선조각을 추출하고, 추출된 곡선조각 중 비슷한 곡률을 가지는 곡선조각들을 조각간 곡률에 따라 군집화하여 인수를 추정하고 둥근 물체를 인식하도록 하였다.

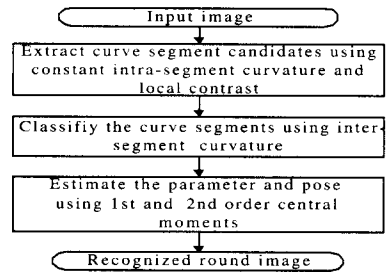


그림 1. 제안한 둥근 물체 인식 흐름도

Fig. 1. Flow chart of the proposed recognition for round objects.

1. 곡선조각 추출

둥근 물체의 테두리 구조를 구하기 위하여 먼저 내부곡률과 국소 대비를 이용하여 곡선조각을 추출하였다.

(1) 내부곡률

이산 신호(discrete signal)에 내재된 미분 가능한 영상잡음의 영향을 피하기 위하여 영상의 명암도 미분을 이용^[5-7] 하지 않고 추출된 곡선조각에 삼각함수를 이용하여 그림 2와 같이 곡률을 정의하였다.

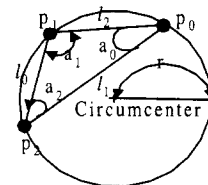


그림 2. 삼각함수를 이용한 곡률의 정의

Fig. 2. Definition of the curvature based on trigonometric function.

그림 2에서 화소 P_0 , P_1 및 P_2 가 만드는 삼각형의 세 내각 α_0 , α_1 및 α_2 와 세 변의 길이 l_0 , l_1 및 l_2 그리고 이 삼각형의 외접원의 반지름을 r 이라 할 때,

이들 사이에는 다음과 같은 관계가 성립한다.

$$\frac{l_0}{\sin \alpha_0} = \frac{l_1}{\sin \alpha_1} = \frac{l_2}{\sin \alpha_2} = 2r \quad (1)$$

이 때, 가운데 화소 P₁에서의 곡률 $k_{(P_1)}$ 은 다음과 같이 구할 수 있다.

$$k_{(P_1)} = \frac{2\sin \alpha_1}{l_1} \quad (2)$$

여기서 곡률 $k_{(P_1)}$ 은 화소 P₀와 P₂사이의 각도 $\angle P_0P_1P_2$ 가 0 또는 π 일 때 가장 작은 값을 가지며, 그 사이의 각도가 $\pi/2$ 일 때 가장 큰 값을 가진다. 또한 정현 함수(sine function)는 저주파 통과 필터의 특성을 가지므로 고주파 영상 잡음은 제거된다. 둥근 테두리는 곡률의 변화가 거의 없으므로, 곡선조각 위의 한 점 s_p 에서의 곡률 $k(s_p)$ 은 일정곡률조건 $dk(s_p)/ds \approx 0$ 을 만족한다. 그러나, 곡률 계산시 정현 함수를 사용하여 고주파 영상 잡음을 억제하여도 여전히 저주파 영상 잡음은 존재하기 때문에 $dk(s)/ds=0$ 근처에서 진동한다. 따라서 여유도(tolerance margin) $\epsilon_k > 0$ 에 대하여 일정 곡률 기준을 다음과 같이 설정하였다. 즉,

$$-\epsilon_k \leq \frac{dk(s)}{ds} \leq \epsilon_k \quad (3)$$

만약 테두리상의 점이 식 (3)을 만족하면 일정한 곡률을 갖는 조각으로, 그렇지 않으면 일정 곡률을 가지지 않는 곡선조각으로 분류하였다.

(2) 국소 대비

이웃의 범위가 σ 인 가우스 필터^[9]를 통과한 입력 영상의 에지 맵 g_σ 에 대하여 기존의 화소 P₋₁, P₀에 연결될 수 있는 화소는 다음과 같이 정의한 국소 대비 c_i 를 최대화 하는 화소로 결정한다.

$$c_i = \max_i \{ g_i^2 - \sum_{j=1}^{n-1} g_j^2 \}, \quad (4)$$

$$i = 1, \dots, n^2 - 2n + 1$$

여기서 i 는 화소 위치이고, g_i, g_j 는 각각 i 와 j 번째 화소의 밝기차이며, n 은 $2\sigma + 1$ 로 설정된다.

그림 3은 $\sigma=1$ 인 이웃에 대한 곡선조각 G_c 와 배경 G_B 에 대한 그림이다.

그림 3(a)는 에지 맵 g_σ 를 나타내며, 그림 3(b)는 화소 P₋₁, P₀, 및 P₁로 연결되는 곡선조각 G_{c1} 과 그 외

P₂, P₃, P₄, P₅, P₆ 및 P₇에 의해 형성되는 배경 G_{B1} 과의 대비가 비교적 낮은 경우를 도시한 것이고,

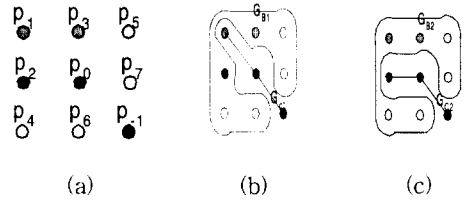


그림 3. $\sigma=1$ 인 이웃에 대한 테두리 G_c 와 배경군 G_B : (a) 밝기차 분포, (b) 테두리 곡선조각 후보, (c) 곡선조각

Fig. 3. The background G_B and the curve segment G_c for a neighborhood with $\sigma=1$: A distribution of grey-level difference (a), boundary curve segment candidate (b), and curve segment (c).

그림 3(c)는 화소 P₋₁, P₀, P₂로 연결되는 곡선조각 G_{c2} 와 화소 P₁, P₃, P₄, P₅, P₆, 및 P₇이 만드는 배경 G_{B2} 와의 대비가 높은 경우를 나타낸다. 이때 식 (3)과 식 (4)를 만족하는 연결화소를 추출하면, 그림 3(c)와 같이 곡선조각이 연결 추출된다. 그리고 일정 곡률 조건을 만족하지 않는 연결 화소 중 가장 높은 대비와 두 번째 대비의 차이가

$$\frac{C_{k(max)} - C_{k(second)}}{C_{k(max)}} \geq th_{break} \quad (5)$$

의 조건을 만족하는 화소점 P₀를 곡선조각의 분기점으로 등록한다. 이렇게 함으로써 분기점에 연결된 또 다른 곡선조각 추출시 별도의 처리 없이 시작점을 선택할 수 있다.

(3) 곡선조각 추출

둥근 물체의 곡선조각 추출을 위해 그림 4(a)와 같이 8 방향 체계로 화소를 검색한다. 이때, 현재의 화소 p_k 까지 추출된 곡선조각에 대해 연결될 수 있는 화소는 식 (2)와 (3)으로 정의된 내부 곡률조건을 만족하는 연결방향을 먼저 검색한 후 식 (4) 및 (5)로 주어진 국소 대비조건을 이용하여 곡선조각을 추출한다.

즉 그림 4(b)와 같이 곡선조각이 서쪽(왼쪽)에서 현재 화소인 중심화소로 연결되는 '경우, 기존의 8방향 왼쪽화소 우선 검색법은 ④, ②, ①, ③, 그리고 ⑤의 순서로 검색하며 만약 이들 중 이어질 화소가 없으면 ①의 순서로 연결화소를 검색한다. 그러나 제안된 방법에서는 연결 방향에 따라 ①, ②, ③, ④, 그리고 ⑤

를 순차적으로 연결방향을 우선 검색하도록 함으로 써 연결방향이 변화가 적은 둥근 테두리의 성질을 이용하여 검색시간을 줄일 수 있도록 하였다. 또한 내부 곡률조건에 의하여 ⑥과 ⑦방향은 검색에서 제외되므로 각 방향마다 최대 5방향만 검색하면 된다.

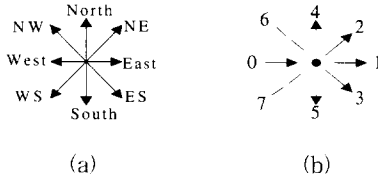


그림 4. 곡선조각 추출 검색 우선 순위: (a) 검색 방향, (b) 곡선조각이 동쪽을 향할 경우의 검색 순서
 Fig. 4. The directional priority of the proposed boundary-segment searching: (a) searching direction, and (b) searching sequence when the segment direction is east.

이 때, 영상의 왼쪽에서 오른쪽으로, 위에서 아래로 순차 주사하면서 처음으로 만나는 영역의 화소를 곡선조각 검색의 시작점으로 하고, 시계방향으로 반복적으로 곡선조각을 연결해 나간다.

2. 곡선조각의 군집화

하나의 동일한 둥근 테두리를 찾기 위하여

$$k(s_i, s_j) = k(s_i) \{ |k(p_i) - k(p_j)| \leq \epsilon_c \} \quad (6)$$

로 정의되는 조각간의 곡률을 이용하여 추출된 곡선조각들을 군집화하였다. 여기서, $k(p_i)$ 는 추출된 곡선조각 s_i 상의 한 점 p_i 에서 구한 내부곡률을, ϵ_c 는 두 조각을 동일한 테두리의 일부로 묶기 위한 최대 허용 곡률차(maximum allowable curvature difference)이다.

곡선조각 추출과정에서 만들어진 1부터 n_s 개의 초기 조각들 s_i 을, k_s 개의 비슷한 곡률을 갖는 테두리 $R_s = \{s_i, i = 1, \dots, k_s\}$ 로 군집화하는 군집화 알고리즘은 다음과 같다.

1. $R_s = \{s_i, i = 1, \dots, n_s\}$ 로 테두리 집단을 초기화한다.
2. $k(s_i) \approx 0$ 인 직선 조각을 검색하여 집단 R_s 에서 제거한다.
3. 곡률에 비하여 곡선조각의 길이가 $k(s_i) \cdot l(s_i) \approx 0$ 로 짧은 조각을 집단 R_s 에서 제거한다.
4. 조각간 곡률 $k(s_i, s_j)$ 를 구한다.

5. $k(s_i, s_j) \leq \epsilon_c$ 이면

s_i, s_j 를 집단 R_s 에서 제거하고, s_i, s_j 를 통합한 새로운 그룹을 R_s 에 추가한 후 단계 2를 다시 수행한다. 그렇지 않으면 끝낸다.

3. 인수 추정 및 둥근 물체 인식

군집화된 둥근 테두리에 가장 유사한 타원의 인수를 추정하기 위하여 Hu의 모멘트(Hu's moments)^{9, 10} 중 1, 2차 중심 모멘트를 이용하여 타원의 기울기, 크기 및 타원정도를 추출하였다.

2차원 입력 $f(x,y)$ 에 대한 $p+q$ 차 2차원 모멘트 m_{pq} 및 변위 이동에 불변한 특징을 갖는 중심 모멘트 μ_{pq} 는 각각

$$m_{pq} = \sum_x \sum_y x^p y^q f(x, y) \quad (7)$$

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q f(x, y) \quad (8)$$

와 같이 정의된다. 여기서 \bar{x}, \bar{y} 는 영상의 중심을 나타낸다.

그리고 2차 중심 모멘트벡터 $\begin{bmatrix} \mu_{02} & \mu_{11} \\ \mu_{11} & \mu_{20} \end{bmatrix}$ 의 고유치 a, b 는 타원의 장축 및 단축 길이를 나타내며

$$a = \frac{\mu_{20} + \mu_{02} + \sqrt{(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2}}{2} \quad (9)$$

$$b = \frac{\mu_{20} + \mu_{02} - \sqrt{(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2}}{2}$$

와 같이 구해지며, 타원 정도(eccentricity) e 는

$$e = \sqrt{\frac{b}{a}} = \sqrt{\frac{\text{minor axis}}{\text{major axis}}} \quad (10)$$

로 구해진다. 이때 타원의 기울어진 방향 θ 는 다음과 같다.

$$\theta = \frac{1}{2} \tan^{-1} \left(\frac{2\mu_{11}}{\mu_{20} - \mu_{02}} \right) + n \left(\frac{\pi}{2} \right) \quad (11)$$

따라서 기울기, 장축 단축길이 및 타원 정도를 추정하여 \bar{x} 및 \bar{y} 를 중심으로 하는 둥근 물체를 인식할 수 있다.

제안된 방법에서는 화소 단위가 아닌 곡선조각단위로 처리하기 때문에 메모리 요구량은 영상 전체의 크기가 아닌 입력 영상에 존재하는 곡선조각을 구성하는 화소 수에 비례한다. 또한 표 1에 정리한 바와 같이 제안된 방법의 계산량은 곡선조각을 추출하기 위해 처리되어야 할 곡선조각상의 화소 수에 비례한다.

연결 영상을 반복적인 분리/병합 방법과 다각형 근사화방법을 적용하여 둥근 물체를 추출한 결과를 그림 (e)에 나타내었다.

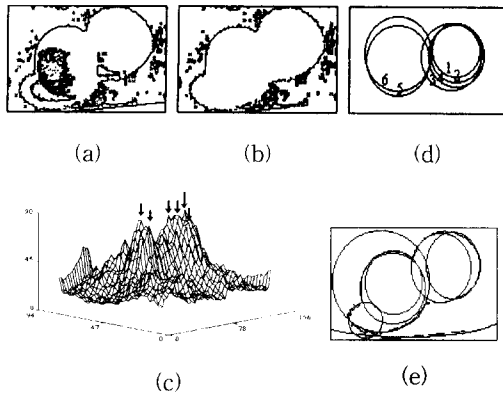


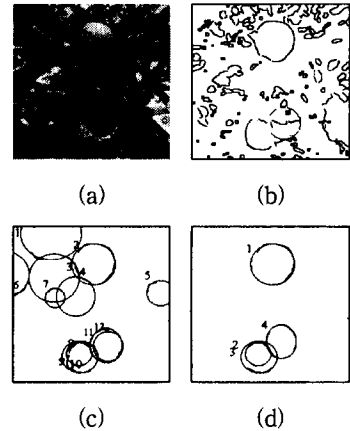
그림 6. 그림 5(a)에 Hough변환을 적용한 원 추출: (a) 에지 영상, (b) 에지 연결 영상, (c) 중심위치에 대한 Hough 공간, (d) Hough방법으로 추출된 원 및 (e) 다각 근사화기법에 근거한 반복적 분리 병합을 이용한 둥근 테두리 추출

Fig. 6. Circle extraction using Hough transform: (a) edge image, (b) linked broken edge image, (c) Hough parameter space for center position, (d) extracted circles using Hough transform, and (e) extracted circles using iterative split and merge based on polygon approximation.

그림 6(a)는 그림 5(a)에 대한 에지 영상을, 그림 6(b)는 에지를 연결한 결과 영상을, 그리고 그림 6(c)는 Hough변환의 결과인 인수의 누적분포를 나타낸다. 그림 6(d)는 Hough 공간의 누적 분포를 바탕으로 추출한 결과로서 추출된 원 1, 2, 3 및 4의 순으로 높은 누적값을 나타내었으며, 원 5 및 6의 누적치는 현저히 떨어졌다. 즉, 누적값이 적은 원을 추출할 경우 누적값이 큰 원은 동일한 원에 대하여 여러 개의 원이 비슷한 위치에 겹쳐서 존재하는 것으로 추출된다. 한편 그림 6(e)와 같이 추출된 윤곽선을 분리 및 병합하여 곡선조각으로 분리한 후 다각 근사화^[4]를 행하여 둥근 물체를 추출하여 본 결과 다각근사화 과정에서 하나의 물체가 여러개의 물체로 분리되었다. 그러나 제안된 방법에서는 추출된 곡선조각간의 곡률을 검사하여 테두리를 구성하는 곡선조각들이 비록 분리되어 있더라도 효과적으로 곡선조각을 군집화하여 오 추출없이 둥근 테두리를 추출할 수 있었다.

과일의 수확 및 선별과정에서 만날 수 있는 과일

나무 영상 및 원형이 아닌 타원의 형태를 가지는 호박 영상에 대한 실험 결과를 각각 그림 7 및 그림 8의 (a), (b), (c) 및 (d)에 도시하였으며, 인식 결과와 추정된 인수의 값 분포를 그림 7(e) 및 8(e)에 도시하였다.



	Major axis	Minor axis	Perimeter	Roundness
Mean	57.9635264078	39.2932561238	229.968456014	509.770592053
Min	47.7054634094	29.2256050109	214.613778686	407.743195758
Max	70.3764266967	53.3657569885	238.779708862	631.1101787090
Std Dev	9.3799657748	10.2533969799	10.733686817	092201747604

(e)

그림 7. 일부가 가려진 과일을 가진 나무 영상으로부터 원형 물체 인식: (a) 과일 나무 영상 (b) 곡소 대비 및 내부곡률을 이용해 추출한 곡선 조각, (c) 조각간 곡률을 이용하여 추출한 테두리 집합, (d) 인수 추정결과에 대한 원형들 및 (e) 추정된 인수분포

Fig. 7. Recognition of circular objects from tree image with partially occluded circular fruits: (a) image of fruit tree, (b) result of the extracted curve segments, (c) boundary groups, (d) result of the object-parameter estimation, and (e) parameter distribution.

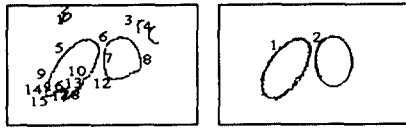
그림 7 (c)의 원 1, 2, 4, 5, 6, 7, 9 및 11은 군집화 과정에서 직선 및 곡률에 대한 곡선조각 길이 검증과정에서 버려지므로 최종인식과정에서 무시된다.

그림 7 및 8 에서 보인 바와 같이 복잡한 영상에 대해서도 제안된 방법은 효과적으로 둥근 물체를 인식할 수 있었다. 또한 군집화 과정에서 짧은 곡선조각에 대한 정보가 사라지지만 최종 인식결과에는 큰 영향을 미치지 않았다.



(a)

(b)



(c) (d)

	Major axis	Minor axis	Perimeter	Roundness
Mean	57.1987533569	29.6945571899	199.184453328	542856832345
Min	49.48056793210	24.68177232200	50.987806227	382219480110
Max	70.24719238280	34.88319778440	229.3554534910	740290343761
Std Dev	9.27799618056	4.16557351052	34.4389785235	149479578144

(e)

그림 8. 호박 영상으로부터 둥근 물체인식: (a) 호박 영상 (b) 국소 대비 및 내부곡률을 이용해 추출한 곡선 조각, (c) 조각간 곡률을 이용하여 추출한 테두리 집합, (d) 인수 추정 결과 (e) 추정된 인수의 분포.

Fig. 8. Recognition of round objects from pumpkin image in outdoor scene: (a) Pumpkin image, (b) extracted curve segments, (c) boundary group, and (d) result of the object parameter estimation, and (e) parameter distribution.

IV. 결론

본 논문은 내부 곡률 및 국소 대비를 사용하여 추출된 곡선 조각들의 군집화를 통하여 부분적으로 가려진 둥근 물체를 인식할 수 있는 효과적인 방법을 제안하였다. 그림자 및 다른 물체 등으로 일부 가려진 둥근 물체의 추출에 제안된 방법을 적용하여 실험한 결과, 화소 단위가 아닌 곡선 조각 단위로 둥근 물체를 추출하기 때문에 전체 화소 크기가 아닌 곡선 조각을 이루는 화소수 정도의 메모리와 곡선 조각의 수에 비례하는 계산량으로 둥근 물체를 인식할 수 있었으며, 추출할 둥근 물체의 크기에 상관없이 둥근 물체의 크기 및 기울어진 방향 등도 추출할 수 있었다.

참 고 문 헌

[1] R. O. Duda and P. E. Hart, "Use of the Hough transformation to detect lines and curves in pictures," *Communication of the ACM*, vol 15, pp. 204-208, 1972

[2] H. K. Tuen, J. Illingworth, and J. Kittler, "Detecting partially occluded ellipses using the Hough transform," *Image Vision Computing*, vol. 7, no. 1, pp. 31-37, 1989

[3] H. K. Yuen, J. Princen, J. Illingworth, and J. Kittler, "Comparative study of Hough transform methods for circle finding," *Image Vision Computing*, vol. 8, no. 1, pp. 71-77, 1990

[4] D. M. Wuescher and K. L. Boyer, "Robust contour decomposition using a constant curvature criterion," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 13, no. 1, pp. 41-51, Jan. 1991

[5] F. Mokhtarian and A. Mackworth, "Scale-based description and recognition of planar curves and two-dimensional shapes," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 8, no. 1, pp. 34-43, 1986

[6] J. A. Ventura and J. M. Chen, "Segmentation of two-dimensional curve contours," *Pattern Recognitions*, vol. 25, no. 10, pp.1129-1140, 1992

[7] M. K. Hu, "Visual pattern recognition by moment invariants," *IRE Transaction on Information Theory*, pp. 179-187, Feb. 1962

[8] T. S. Newman and A. K. Jain, "A survey of automated visual inspection," *Computer Vision and Image Understanding*, vol. 61, no. 2, pp. 231-262, 1994

[9] A. Ishii, "Trend of Inspection and Shape Measurement Using Image Processing," *Journal of the Japan Society for Precision Engineering*, vol. 61, no. 5 pp. 603-604, 1995

[10] E. R. Davis, *Machine Vision: Theory, Algorithm, and Practicalities*, Academic Press, London, 1990

저 자 소 개