

논문 2006-43CI-5-5

# 형상 정보와 모션 정보 융합을 통한 움직이는 물체 인식

## ( Moving Object Classification through Fusion of Shape and Motion Information )

김정호\*, 고한석\*\*

( Jungho Kim and Hanseok Ko )

## 요약

기존의 인식 방법은 물체에 대한 형상 정보 또는 움직임을 특징으로 한 단일 인식기를 사용한다. 하지만, 기존의 단일 특징 기반의 단일 인식기를 사용하는 방법의 인식 성능은 물체의 영역에 대한 정확한 검출에 크게 의존하는 단점을 가진다. 본 논문에서는 이러한 기존 인식방법의 단점을 해결하고, 인식의 신뢰성을 높이기 위해서 세 가지 인식기에 의한 각 결과를 Bayesian을 이용하여 융합하는 새로운 인식 방법을 제안한다. 첫 번째 인식기는 푸리에 묘사자로부터 얻은 형상 정보를 특징으로 한 신경망을 사용하고, 두 번째 인식기는 형상 정보에 대한 기울기를 바탕으로 한 통계적인 방법을 사용한다. 또한, 세 번째 인식기는 검출된 물체의 일정 부분의 움직임에 대한 모션 정보를 특징으로 하여 인식한다. 본 논문의 실험결과에서 제안한 결과 융합방법은 기존의 Majority Voting과 Weight Average Score 방법에 비해서 더 우수한 인식 성능을 보여준다.

## Abstract

Conventional classification method uses a single classifier based on shape or motion feature. However this method exhibits a weakness if naively used since the classification performance is highly sensitive to the accuracy of moving region to be detected. The detection accuracy, in turn, depends on the condition of the image background. In this paper, we propose to resolve the drawback and thus strengthen the classification reliability by employing a Bayesian decision fusion and by optimally combining the decisions of three classifiers. The first classifier is based on shape information obtained from Fourier descriptors while the second is based on the shape information obtained from image gradients. The third classifier uses motion information. Our experimental results on the classification performance of human and vehicle with a static camera in various directions confirm a significant improvement and indicate the superiority of the proposed decision fusion method compared to the conventional Majority Voting and Weight Average Score approaches.

**Keywords :** Classification, Fusion, Shape Information, Motion Information

## I. 서론

컴퓨터 사용의 증가와 기술이 발전함에 따라 최근 산업 및 주요 기간시설에 경계 및 감시를 위한 감시카메라를 이용하여 움직이는 물체를 스스로 인지 및 추적하는 지능형 감시/경계 로봇시스템 개발이 활발히 진행 중이다<sup>[1]</sup>. 이러한 시스템에서 움직이는 물체에 대한 검출과 인식은 중요시되고 있다. 대부분의 영상 감시/경

계 시스템은 배경영상으로부터 움직임 영역을 검출하는 과정이 필요하다. 이러한 움직임 검출과정은 이후, 물체의 인식이나 모션분석과 같은 과정에 많은 영향을 미치게 됨으로써 매우 중요하다. 기존의 움직임 검출 알고리즘으로는 프레임과 프레임 사이의 시간차를 이용한 시간차(Temporal Differencing)방법, 배경영상과 현재 영상의 차이를 이용한 배경차(Background Subtraction)방법, 연속되는 프레임에서 움직임이 있는 영역을 검출하기 위한 시간차에 의한 움직임의 방향 및 속도벡터를 이용하는 광학적 유동(Optical Flow)방법 등이 있다<sup>[1,2,5]</sup>. 시간차(Temporal Differencing) 방법은 동적인 환경에 잘 적응하는 장점이 있는 반면, 일반적으로 정확한 움직임 영역 검출이 힘들다는 단점을 가지고 있다.

\* 학생회원, \*\* 정회원 고려대학교 전자컴퓨터공학과  
(Dept. of Electronics & Computer Engineering,  
Korea University)

※ 본 연구는 산업자원부의 위탁과제 연구 지원에 의해 수행되었습니다.(과제번호 : 10012805 )

접수일자: 2005년12월8일, 수정완료일: 2006년9월12일

그리고 배경차(Background Subtraction) 방법은 가장 완벽한 움직임 영역을 검출할 수는 있지만, 조명이나 외부 환경으로 인한 동적인 장면 변화에 너무 민감한 단점을 가지고 있다. 또한, 광학적 유동(Optical Flow) 방법은 카메라 움직임에도 독립적으로 움직임 영역을 검출할 수 있는 장점이 있지만, 계산이 너무 복잡하여 실시간 알고리즘 적용에는 부적절한 단점이 있다. 따라서 본 논문에서는 현재 실시간 영상 감시/경계 응용분야에서 많은 연구가 이루어지고 있는 적응형 배경차(Adaptive Background Subtraction) 방법을 사용한다.<sup>[4]</sup> 이 방법의 핵심은 감시/경계지역 내에서 가능한 모든 변화에 적용할 수 있도록 배경영상을 지속적으로 갱신함으로써 동적인 환경에서도 강인하게 움직임 영역을 검출하는 것이다. 움직이는 영역을 검출한 후, 검출된 물체에 대한 인식과정을 수행한다. 기존의 인식방법으로는 형상 정보와 모션정보를 바탕으로 하는 두 가지 접근법이 있다. 형상 정보를 이용하는 방법으로는 VSAM(Visual Surveillance and Monitoring), Lipton et al, Kuno et al<sup>[1,9,10]</sup> 등이 있으며, 모션 정보를 이용하는 방법으로는 Cutler et al, Lipton et al<sup>[6,8]</sup> 등이 있다. 그리고 최적의 인식성능 향상을 위한 융합 기법으로는 데이터 융합과 결과 융합이 있다<sup>[6,7]</sup>. 본 논문에서는 물체의 푸리에 묘사자와 기울기를 이용한 형상 정보와 특징점 변화량을 이용하는 모션 정보를 사용하여 다중 인식기를 구성한 후 인식한 결과를 인식기의 신뢰성에 의해 결정된 가중치를 적용하여 융합하는 Bayesian 결과 융합방법을 제안한다. 이러한 방법은 배경상태에 따라 검출된 움직임 영역의 정확도에 의존하는 기존의 인식방법의 단점을 해결할 뿐만 아니라 동시에 인식성능을

개선하는 장점이 있다. 그림 1은 본 논문에서 제안한 인식 시스템의 전반적인 구성도이며, 크게 3단계로 구성된다. 첫 번째 단계는 적응형 배경차(Adaptive Background Subtraction) 방법을 사용한 움직이는 물체 검출, 두 번째 단계는 검출된 물체의 형상 정보와 모션 정보를 특징으로 사용하여 물체를 인식하는 과정, 마지막으로 다중 인식기로부터 얻은 인식결과를 융합하는 과정이다.

위와 같은 과정에 따라 본 논문의 구성은 다음과 같다. II장에서는 움직이는 물체 검출, III장은 형상 정보와 모션 정보를 이용한 물체 인식방법, IV장은 결과 융합에 사용된 방법에 대해서 설명한다. 또한, V장에서는 다수투표(Majority Voting)방법과 가중치 평균점수(Weight Average Score)에 의한 결과 융합방법과의 비교를 통해 제안한 방법의 인식성능을 보이고, 마지막으로 VI장에서는 결론을 맺는다.

## II. 움직이는 물체 검출

일반적인 환경에서 바람에 의한 장면변화 및 조명변화로 인하여 얻어진 초기 배경이미지를 지속적으로 사용할 수가 없다. 따라서 본 논문에서는 이러한 단점을 보완하면서 보다 정확한 움직임 영역을 검출하기 위해서 적응형 배경차(Adaptive Background Subtraction) 방법을 사용한다<sup>[4]</sup>.

### 1. 적응형 배경차

#### (Adaptive Background Subtraction)

이 방법은 2단계의 과정으로 구성된다. 첫 번째 단계는 아무런 움직이는 물체가 없는 경우 Harmonic 성분을 제거하면서 배경이미지 생성을 위해서 IIR 필터를 적용한다. 아래의 식은 이를 이용하여 갱신되는 배경을 나타낸다.

$$B_{n+1} = \alpha I_n + (1 - \alpha) B_n \quad (1)$$

여기에서,  $B_n$ 은 시간  $t = n$ 에서 배경영상,  $B_{n+1}$ 은 갱신된 배경영상,  $I_n$ 은  $t = n$ 에서 입력영상,  $\alpha$ 는 필터의 계수를 나타낸다. 이 계수는 배경영상과 입력영상에 적절한 가중치를 부여하여 움직이는 물체를 배경에서 제외하는 역할을 한다. 본 논문에서는 실험적으로 구한 0.05값을 적용하였다. 두 번째 단계는 갱신된 배경에 움직이는 물체가 검출되었을 때, 배경영상과 입력영상

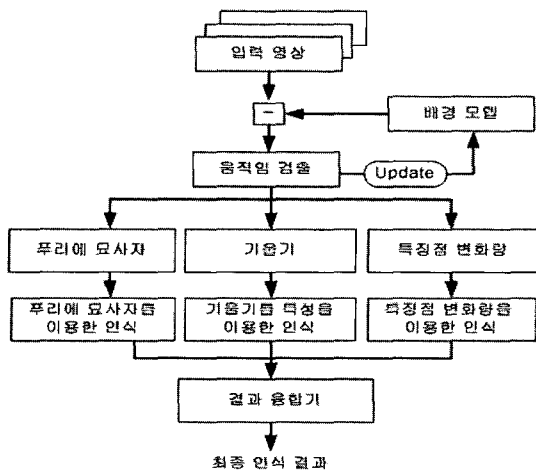


그림 1. 인식 시스템 구성도  
Fig. 1. Overview of the overall system.

과의 차 영상을 구한 후, 움직이는 물체만을 구하기 위하여 이진화 시키는 과정이다. 아래의 식 (2)와 (3)을 이를 나타낸다.

$$F(D_n) = \begin{cases} 1 \text{ (Moving)} & \text{if } D_n > \tau \\ 0 \text{ (Static)} & \text{Otherwise} \end{cases} \quad (2)$$

$$\tau = \text{mean}[MC(D_n, N)] + \sigma(D_n) \quad (3)$$

위의 식에서  $D_n$ 은 입력영상과 배경영상간의 차 영상을 나타내고,  $\tau$ 는 이진영상을 검출하기 위한 임계값을 나타낸다. 또한,  $F(D_n)$ 은 차 영상에 의한 이진영상,  $MC(D_n, N)$ 은 차 영상  $D_n$ 에  $N \times N$  사이즈의 윈도우를 사용하여 얻은 평균을 나타내고,  $\sigma(D_n)$ 은 차영상의 표준편차를 나타낸다. 위의 식(2)와 (3)에 의해서 얻어진 결과를 식 (4)에 적용하여 움직임 픽셀이라면 다음 배경영상의 픽셀값을  $\alpha$  필터를 적용한 후의 값으로 갱신하고, 움직임이 없는 픽셀은 다음 배경영상의 픽셀값을 이전 배경영상의 픽셀값을 그대로 유지한다.

$$B_{n+1} = \begin{cases} \alpha I_n + (1 - \alpha) B_n & \text{if } F(D_n) = 0 \\ B_n & \text{Otherwise} \end{cases} \quad (4)$$

## 2. 잡음제거

이렇게 구해진 이진영상에는 많은 잡음들이 섞여 있다. 이러한 잡음 제거를 위해 침식과 팽창을 두 번 적용하는 형태학적 영상처리 기법을 사용한다.

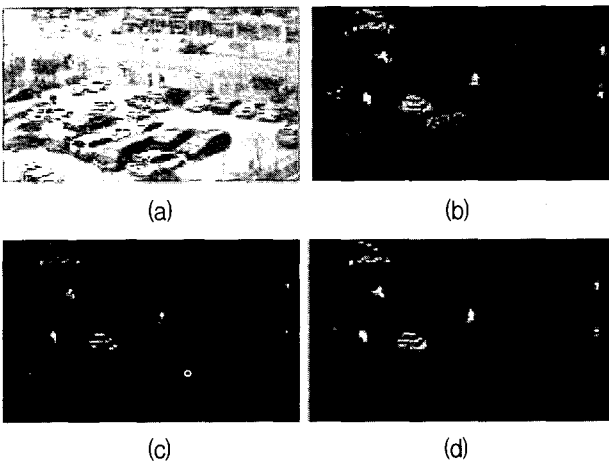


그림 2. 잡음 제거를 위한 전처리 : (a) 입력영상, (b) 차 영상, (c) 팽창, (d) 침식

Fig. 2. Preprocessing process for noise reduction : (a) input image, (b) difference image, (c) dilation, (d) erosion.

$$M_n = (((F(D_n) \ominus A) \oplus A) \oplus A) \ominus A \quad (5)$$

위의 식(5)에서,  $M_n$ 은 잡음이 제거된 결과영상,  $F(D_n)$ 은 잡음이 섞인 이진영상을 나타내고, 또한  $\oplus A$ 는 팽창,  $\ominus A$ 는 침식을 나타낸다. 침식과 팽창 필터의 사이즈는  $3 \times 3$ 을 적용하였다. 그림 2는 위의 식(5)에 의해서 계산된 결과를 보여준다.

## 3. 움직임 영역 분할/경계(Boundary)영상 검출

잡음제거 과정을 거친 이진영상으로부터 움직임 영역에 대한 labeling(분류표시)을 통해서 물체 영역만을 분할한다. 분할 된 이진영상으로부터 더욱 정확한 물체의 형상을 얻기 위해서 아래 식(6)과 같이 분할된 이진영상과 단위원 필터와 회선(convolution)연산을 한다. 이 필터 사이즈는 실험에 의해서  $3 \times 3$ 을 적용하였다.

$$SM_n(x, y) = SM_n(x, y) * C(x, y) \quad (6)$$

$$C(x, y) = (x - i)^2 + (y - j)^2 = r^2$$

여기에서,  $SM_n$ 은 분할된 물체 영역을 나타내고,  $C$ 는 단위원 필터를 나타낸다. 또한,  $(x, y)$ 는 영상에 대한 좌표를 나타내고,  $(i, j)$ 는 단위원 필터의 중심좌표를 나타내며,  $r$ 은 단위원 필터의 반지름을 나타낸다. 이를 이용하여 형상에 대한 경계(boundary)영상을 구한다. 아래 그림 3은 위의 과정을 보여준다.

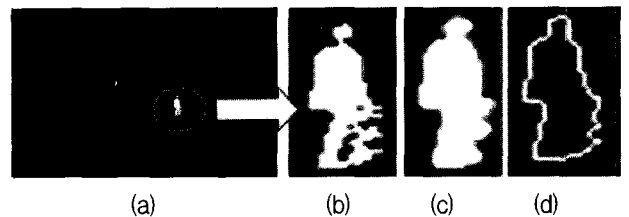


그림 3. 움직임 영역 분할 및 경계 영상 검출 :

(a) 이진 영상, (b) 물체 영역 분할, (c) 원 필터와 이진 영상의 컨벌루션 결과, (d) 경계 영상

Fig. 3. Moving region segmentation and boundary image extraction process: (a) binary image, (b) moving region segmentation, (c) convolution result with circle filter, (d) boundary image.

## III. 형상 정보와 모션 정보를 이용한 인식 방법

### 1. 형상 정보를 이용한 인식 방법

본 논문에서는 형상정보에 대한 특징으로 Fourier descriptor(푸리에 묘사자)와 기울기 값을 사용하여 인

식한다. 푸리에 묘사자의 사용은 패턴 인식이나 이미지 분석 등에 널리 쓰이는 방법으로 여러 가지 변화(이동, 회전 및 크기)에 불변하는 장점이 있고, 기울기 값 또한 실험을 통해서 여러 가지 변화에 불변하는 특성을 갖는다.

가. Fourier descriptor(푸리에 묘사자)를 특징으로 한 신경망 인식방법

형상정보의 특징인 푸리에 묘사자는 다음의 과정에 의해서 구한다.

(1) 움직임 영역에 대한 중심점을 구한다.

$$x_c = \frac{1}{K} \sum_{i=1}^K x_i, \quad (7)$$

$$y_c = \frac{1}{K} \sum_{i=1}^K y_i$$

위의 식에서  $x_i$ 는  $i$ 번째 경계에 대한 픽셀의  $x$ 좌표,  $y_i$ 는  $i$ 번째 경계에 대한 픽셀의  $y$ 좌표,  $K$ 는 영역의 경계에 대한 총 픽셀 수,  $x_c, y_c$ 는 영역의 경계의 픽셀에 대한 좌표의  $x$ 와  $y$ 에 대한 평균 중심 좌표를 나타낸다.

(2) 움직임 영역의 중심점으로부터 경계(Boundary) 각 픽셀까지의 거리함수를 구한다.

$$r(k) = ((x(k) - x_c)^2 + (y(k) - y_c)^2)^{1/2} \quad (8)$$

여기에서,  $k = 0, 1, \dots, K-1$ ,  $K$ 는 영역의 경계에 대한 픽셀의 개수를 나타내고,  $r$ 은 중심점으로부터 경계까지의 길이를 나타낸다.

(3) 거리함수를 이산 푸리에 변환(DFT) 시킨다.

$$a(n) = \frac{1}{K} \sum_{k=0}^{K-1} r(k) e^{-j2\pi nk/N} \quad (9)$$

여기에서  $n = 0, 1, \dots, K-1$ ,  $a(n)$ 은 중심으로부터 경계에 까지 계산한 거리함수에 대한 푸리에 묘사자를 나타낸다.

(4) 위에서 구한 푸리에 묘사자는 거리함수  $r(k)$ 처럼 물체의 형상에 대한 같은 정보를 포함하고 있다. 하지만, 이러한 푸리에 묘사자를 물체 인식을 위한 특징으로 사용하기 위해서는 아래 식 (10)의 전처리 과정이

필요하다.

$$FD_{feature} = \left[ \frac{|a(n)|}{|a(0)|} \right], n = 1, 2, \dots, N/2 \quad (10)$$

이러한 전처리 과정을 거친 푸리에 묘사자의 성질은 형상의 위치를 나타내는 중심점으로부터의 거리를 사용하기 때문에 거리함수  $r(k)$ 는 이미 이동에 대해 불변한 성질을 갖게 되고, 회전은 계수의 위상에 의해 코드화되는데, DC-성분인  $a(0)$ 를 무시하고, 단지  $a(n)$ 의 크기만을 사용함으로써 회전에 불변한 성질을 갖게 된다. 또한 모든  $a(n)$ 을  $a(0)$ 의 크기로 나누어 줌으로써 크기에 불변한 성질을 갖게 된다.

(5) 푸리에 묘사자의 저주파 성분은 물체의 전반적인 형상을 결정하고, 고주파 성분은 상세한 부분을 나타내기 때문에 차원을 줄이기 위해서 고주파 성분을 제거하고 처음 30개의 저주파 성분만을 물체 인식을 위한 특징으로 사용한다. 아래의 그림 4는 이를 보여준다.

위에서 구한 30개의 푸리에 묘사자를 신경망의 특징으로 사용하여 훈련 및 인식을 수행하는 방법을 사용하였다. 신경망의 구조는 특징마다 하나의 뉴런을 갖는 입력층, 각각 7개의 뉴런을 갖는 두 개의 은닉층, 2개의 뉴런을 갖는 출력층으로 구성되어 있다. 또한 뉴런을 위해서는 Sigmoidal 함수를 사용하였고, 계층적 학습을 위해서 역전파(Back-Propagation) 알고리즘을 사용하였다. 훈련 데이터는 사람과 차량에 대해 각각 600개를 사용하였다.

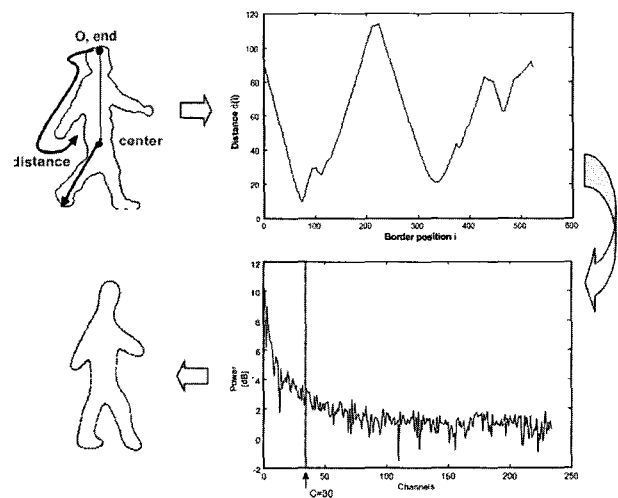


그림 4. 푸리에 묘사자 계산  
Fig. 4. Calculation of fourier descriptors.

나. 기울기(Gradient)를 이용한 인식방법

위의 II장에서 구한 물체영역의 경계 영상으로부터 밝기 값의 불연속을 검출하기 위해서 1차 미분을 이용한다. 2차원 함수  $N(x,y)$ 의 기울기는 아래의 벡터로 정의된다.

$$\nabla f = \begin{bmatrix} G_x \\ G_y \end{bmatrix} = \begin{bmatrix} \frac{\partial N(x,y)}{\partial x} \\ \frac{\partial N(x,y)}{\partial y} \end{bmatrix} \quad (11)$$

이러한 기울기는 일정한 밝기를 가진 영역에서는 0이며, 화소 값이 변하는 영역에서는 밝기의 변화 정도에 따라 비례한다. 기울기 벡터의 기본적인 성질은 좌표  $(x,y)$ 에서  $N$ 의 최대 변화율 방향을 가리킨다는 것이다. 이러한 최대 변화율이 일어나는 각도는 다음과 같이 구한다.

$$\tan \theta = \frac{\frac{\partial N(x,y)}{\partial y}}{\frac{\partial N(x,y)}{\partial x}} = \frac{\Delta y}{\Delta x} \quad (12)$$

$$\theta = \arctan\left(\frac{\Delta y}{\Delta x}\right)$$

위의 식(12)에서 각 경계영역에서 구하여진 각 픽셀에 대한  $\theta$ 값을 이용하여 수직에 가까운 성분이 많은지 수평 수평에 가까운 성분이 많은지를 계산한다.(식 13,14)

$$\theta_{ver} = \begin{cases} \text{경계픽셀수증가, } |\theta| > \text{수직성분에대한임계값} \\ \text{경계픽셀수유지, } \textit{else} \end{cases} \quad (13)$$

$$\theta_{hor} = \begin{cases} \text{경계픽셀수증가, } |\theta| < \text{수평성분에대한임계값} \\ \text{경계픽셀수유지, } \textit{else} \end{cases} \quad (14)$$

위의 식에서  $\theta_{ver}$ 과  $\theta_{hor}$ 는 각각 임계값에 의해서 계산된 총 픽셀수를 나타낸다. 이렇게 구한 수평, 수직 성분을 이용하여 아래의 식(15)를 이용하여 비를 구한 후, 사람과 차량에 대한 인식을 한다. 이와 같이 물체의 경계에 대한 기울기를 특징으로 사용하기 때문에 위치 및 크기에 불변한 성질을 갖으며, 또한 그림 5의 다양한 각도에서의 사람과 차량에 대한 실험결과를 통해서 회전에 대해서도 어느 정도 불변한 성질을 갖는다. 따라서 이러한 형상 정보에 대한 기울기를 물체 인식을 위한 특징으로 사용한다.

$$\theta_{feature} = \begin{cases} 0(car), & \text{if } \theta_{ver}/\theta_{hor} < \textit{threshold} \\ 1(human), & \textit{otherwise} \end{cases} \quad (15)$$

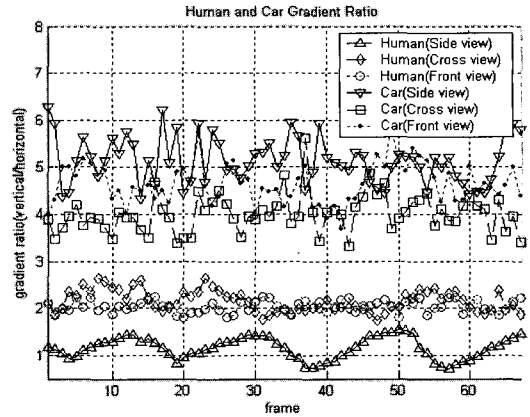


그림 5. 다양한 각도의 사람과 차량에 대한 기울기의 수평 성분 대 수직 성분 비  
Fig. 5. Horizontal versus vertical elements ratio of gradients of human and car in various directions.

2. 모션 정보를 이용한 인식 방법

일반적으로 사람과 차량의 모션을 판단했을 때 사람은 주기적으로 모션이 변하는 성질을 보이는 반면, 차량은 변하지 않는 성질을 가진다. 이러한 성질은 움직이는 물체를 인식하는데 중요한 근거가 된다. 본 논문에서는 모션 분석을 위해서 물체에 대한 골격을 특징으로 추출하여 시간에 대하여 골격이 주기성을 갖는지 비주기성을 갖는지 분석하여 인식한다.

가. 물체의 골격화

본 논문에서는 움직임 영역의 경계(Boundary)에서 최 외곽 지점들을 간단하면서도 강인하게 검출할 수 있는 “스타 골격 방법”을 사용한다.<sup>[3]</sup> 이러한 골격은 다음의 과정을 거쳐 구한다.

- (1) III장에서의 움직임 영역에 대한 중심점과 중심점으로부터 경계의 각 픽셀까지의 거리함수를 구한다.
- (2) 경계의 잡음제거를 위해서 거리함수에 대한 평탄화 과정을 한다.
- (3) 위의 그림 에서,  $d(k)$ 는 거리함수를 나타내고,  $H(u)$ 는 평활화 필터를 나타낸다. 또한,  $\hat{d}(k)$ 는 거리함수의 지역적 최대(Local Maxima)지점을 움직임 영역의 최외곽 지점으로 간주하고, 이 지점의 Zero-Crossing되

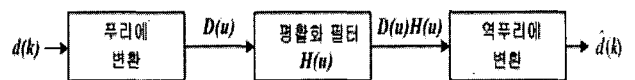


그림 6. 거리 함수의 평탄화 과정  
Fig. 6. Smoothness process of distance function.

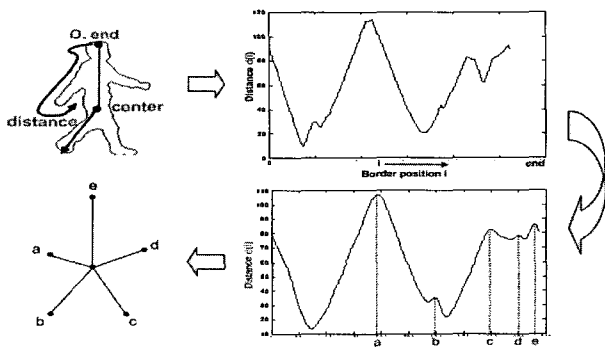


그림 7. 최외각 지점의 선정 및 "스타"형 골격 형성  
 Fig. 7. Selection of external point and "star" skeleton creation.

는 지점을 찾음으로써 구한다.

$$\hat{\delta}(k) = \hat{d}(k) - \hat{d}(k-1) \quad (16)$$

(4) 마지막으로 움직임 영역의 중심점과 최 외곽 지점을 연결함으로써 "스타" 골격을 구할 수가 있다. 아래의 그림 7은 스타 골격 구성을 보여준다.

나. 골격의 특징점 결정

골격의 특징지점으로 본 논문에서는 사람의 경우는 다리에 해당하는 왼쪽 최 하단 부분을 차량의 경우는 앞, 뒤 범퍼에 해당하는 왼쪽 최 하단 부분으로 결정하였다. 이러한 부분은 다른 부분에 비해 변화가 가장 큰 지점이다. 만약  $X$ 가 최 외곽 지점들의 집합이라면, 왼쪽 최 하단 지점은 아래 조건에 의해서 구할 수가 있다.

$$X = \{(x_i, y_i)\}, i = 1, 2, \dots, n$$

$$(l_x, l_y) = (x_i, y_i) = \min_{y_i < y_c} x_i \quad (17)$$

위의 식(17)에서,  $(x_i, y_i)$ 는 최외각 지점  $i$ 번째에서의 위치 좌표를 나타낸다. 또한,  $(l_x, l_y)$ 는 임계값에 의해서 선택된 특징점에 대한 좌표를 나타낸다. 아래의 그림 8은 이를 보여준다.

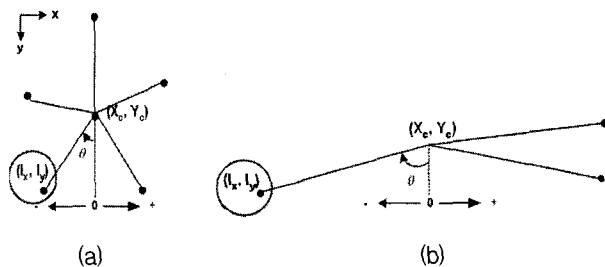


그림 8. 물체별 특징 지점 : (a) 사람, (b) 차량  
 Fig. 8. Feature point : (a) human, (b) car.

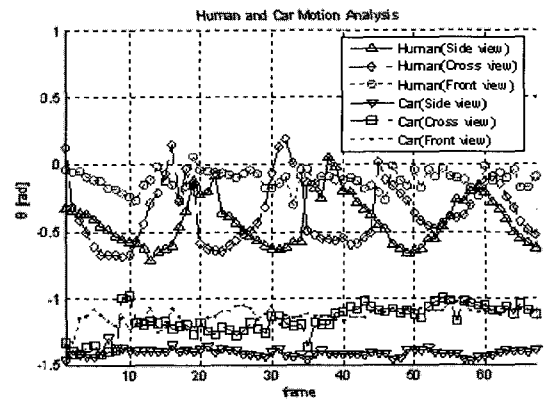
다. 모션분석

모션 분석을 위해서 골격의 특징지점  $(l_x, l_y)$ 과 중심점  $(x_c, y_c)$ 의 수직하는 선과 이루는 각  $\theta$ 를 구한다. 아래의 식은 이를 나타낸다.

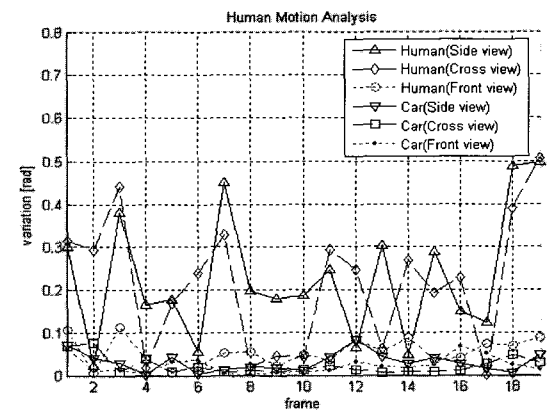
$$\theta = \arctan \frac{l_y - y_c}{l_x - x_c} \quad (18)$$

그림 9 (a)에서 사람의 경우에는 일정한 주기로 각  $\theta$ 가 변하는 반면 차량의 경우에는 각  $\theta$ 가 거의 일정하게 유지됨을 알 수가 있다. 또한, 이러한 분석을 바탕으로 그림 9 (b)에서는 3 프레임 간 각  $\theta$ 의 변화량을 측정된 결과를 보여준다.

위의 그림 9 (b)에서 정면으로 움직이는 사람의 일부 구간을 제외하고는 각  $\theta$ 의 변화량이 차량과 확연히 구



(a)



(b)

그림 9. 물체의 중심선과 특징 지점의 각도( $\theta$ ) 변화 : (a) 연속적인 프레임에서의( $\theta$ ) 변화, (b) 3 프레임간 각도 ( $\theta$ )의 변화

Fig. 9. Variation of angle( $\theta$ ) for centroid and feature point variation : (a) variation of angle( $\theta$ ) in continuous frames (b) variation of angle( $\theta$ ) during three frames.

분류를 알 수가 있다. 따라서 이러한 각  $\theta$ 의 변화량을 물체 인식을 위한 특징으로 사용한다. 아래의 식(19)은 이를 특징으로 하여 인식하는 식을 나타낸다.

$$\theta_{feature} = \begin{cases} 0(car), & \text{if } |\theta_{max} - \theta_{min}| < threshold \\ 1(human) & \text{otherwise} \end{cases} \quad (19)$$

#### IV. 형상 정보와 모션 정보를 이용한 인식 방법

##### 가. Bayesian Decision Fusion(BDF)

BDF는 형상 정보와 모션 정보를 특징으로 사용한 세 가지 인식기에 의한 결과를 인식기의 신뢰성에 의해 결정된 가중치를 적용하여 융합하는 방법이다. 결과 융합구조는 그림 10에서 보는바와 같이 세 가지의 인식기와 하나의 결과 융합기로 구성된다. 여기에서 세 가지 인식기는 각각 푸리에 묘사자에 의한 인식결과, 기울기에 의한 인식결과, 모션정보 변화에 의한 인식을 나타낸다.

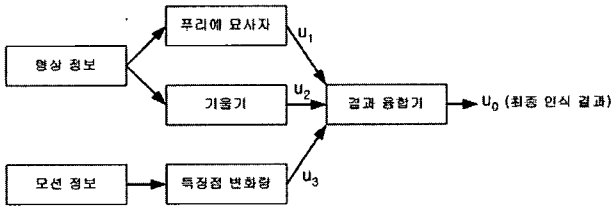


그림 10. 인식 결과 융합 구조

Fig. 10. Structure of decision fusion.

##### 나. 융합 규칙

결과 융합기로 입력되는 데이터는 각 인식기의 지역적결과  $u_i$  ( $i = 1, 2, 3$ )는 인식 결과 순서에 따라 이진 변수로서 표현된다.

$$u_i = \begin{cases} 0, & \text{if classifier } i \text{ decides } H_0(car) \\ 1, & \text{if classifier } i \text{ decides } H_1(human) \end{cases} \quad (20)$$

이러한 지역적 결과는 결과 융합기에 의해서 전역적 결과를 얻는다. 식 (21)은 이를 나타낸다.

$$u_0 = \begin{cases} 0, & \text{if } H_0 \text{ is decided,} \\ 1, & \text{otherwise} \end{cases} \quad (21)$$

각 인식기의 False Alarm( $P_{Fi}$ ), Miss( $P_{Mi}$ ), Classification 확률( $P_{Ci}$ )은 다음과 같이 정의한다.

$$\begin{aligned} P_{Fi} &= P(u_i = 1|H_0) \\ P_{Mi} &= P(u_i = 0|H_1) \\ P_{Ci} &= P(u_i = 1|H_1) \end{aligned} \quad (22)$$

이러한 Bayesian 결과 융합기의 목적은 Average Cost를 최소화하기 위한 융합규칙을 찾는 것이다. 최적의 융합규칙은 다음과 같은 Likelihood Ratio Test를 통해서 얻어진다.

$$\frac{P(u_1, u_2, u_3|H_1)}{P(u_1, u_2, u_3|H_0)} \begin{matrix} u_0 = 1 \\ > \frac{P_0(C_{10} - C_{00})}{P_1(C_{01} - C_{11})} \equiv \eta \\ < \\ u_0 = 0 \end{matrix} \quad (23)$$

여기서  $C_{ij}$ 는  $H_j$ 가 존재하는데 전역적 결과는  $H_i$ 를 선택했을 시 발생하는 Cost를 말한다. 세 가지의 인식기로부터 얻은 지역적 인식결과는 서로 다른 쪽의 인식 결과에 영향을 끼치지 않기 때문에, 식 (23)의 왼쪽 항은 다음과 같이 표현될 수 있다.

$$\begin{aligned} \frac{P(u_1, u_2, u_3|H_1)}{P(u_1, u_2, u_3|H_0)} &= \prod_{i=1}^3 \frac{P(u_i|H_1)}{P(u_i|H_0)} \\ &= \prod_{S_j} \frac{P(u_i = 1|H_1)}{P(u_i = 1|H_0)} \prod_{S_j} \frac{P(u_i = 0|H_1)}{P(u_i = 0|H_0)} \\ &= \prod_{S_j} \frac{1 - P_{Mi}}{P_{Fi}} \prod_{S_j} \frac{P_{Mi}}{1 - P_{Fi}} \end{aligned} \quad (24)$$

여기서  $S_j$ 는 모든 지역적 결과들의 집합으로  $j = 0, 1$ 이다. 식 (23)에 식 (24)를 대입한 후, 양변에 대수(log)를 취하면 다음과 같이 표현될 수 있다.

$$\sum_{S_j} \log \frac{1 - P_{Mi}}{P_{Fi}} + \sum_{S_0} \log \frac{P_{Mi}}{1 - P_{Fi}} \begin{matrix} u_0 = 1 \\ > \log \eta \\ < \\ u_0 = 0 \end{matrix} \quad (25)$$

식 (25)은 또한 다음과 같이 표현될 수 있다.

$$\sum_{i=1}^3 \left[ u_i \log \frac{1 - P_{Mi}}{P_{Fi}} + (1 - u_i) \log \frac{P_{Mi}}{1 - P_{Fi}} \right] \begin{matrix} u_0 = 1 \\ > \log \eta \\ < \\ u_0 = 0 \end{matrix} \quad (26)$$

식 (26)에 따라 최종적인 전역적 결과는 각 인식기로부터 입력되는 지역적 결과와 가중치의 곱에 대한 합을 임계값과 비교함으로써 얻어질 수가 있다. 여기서 가중치는 3장의 각 인식기에 대한 검증 결과로 얻은 False Alarm( $P_{Fi}$ ), Miss( $P_{Mi}$ )을 이용하여 구할 수가 있으며, 이러한 가중치는 각 인식기의 신뢰성을 나타낸다. 또한 임계값은 사전 확률과 Cost들에 의해서 결정된다. 본 논문에서 사전 확률은 어떤 대상을 인식할지 모르기 때문에 1/2값을 주었으며, Cost는  $C_{10} = 1$ ,  $C_{01} = 0.9$ 로 주었다.

V. 실험 결과

가. 푸리에 묘사자를 특징으로 한 신경망 인식

(1) 훈련 영상에 대한 실험

사람과 차량을 인식하기 위한 최적의 임계값 설정을 위해서 임계값을 변화시켜 가면서 Equal Error Rate를 이루는 지점을 찾아본 결과 Total Error Rate가 최소가 되는 즉, False Acceptance Rate와 False Rejection Rate가 만나는 지점이 임계값 0.7 이라는 것을 확인할 수 있었다. (그림 11)

(2) 검증

사람과 차량의 정면, 측면 및 대각선 방향에서 획득한 각각 200개씩의 총 1200개의 데이터를 사용하여 표 1과 같은 결과를 얻었다.

나. 기울기에 의한 수직/수평 성분의 비를 이용한 인식

(1) 훈련 영상에 대한 실험

사람과 차량을 인식하기 위한 최적의 임계값 설정을 위해서 임계값을 변화시켜 가면서 Equal Error Rate를 이루는 지점을 찾아본 결과 Total Error Rate가 최소가 되는 즉, False Acceptance Rate와 False Rejection Rate가 만나는 지점이 임계값 3.2 라는 것을 확인할 수 있었다. 그림 12는 이를 보여준다.

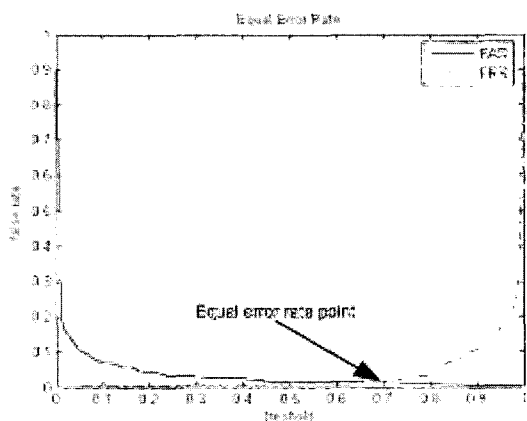


그림 11. 푸리에 묘사자를 이용한 인식결과 동일 오류율 지점

Fig. 11. Equal error rate point.

표 1. 푸리에 묘사자를 이용한 인식 결과

Table 1. Classification result using fourier descriptors.

False alarm(%)	Miss(%)	Classification(%)
3.0	4.0	96

(2) 검증

사람과 차량의 정면, 측면 및 대각선 방향에서 획득한 각각 200개씩의 총 1200개의 데이터를 사용하여 표 2와 같은 결과를 얻었다.

다. 모션정보의 스타골격을 이용한 인식

(1) 훈련 영상에 대한 실험

사람과 차량을 인식하기 위한 최적의 임계값 설정을 위해서 임계값을 변화시켜 가면서 Equal Error Rate를 이루는 지점을 찾아본 결과 Total Error Rate가 최소가 되는 즉, False Acceptance Rate와 False Rejection Rate가 만나는 지점이 임계값 0.06 라는 것을 확인할 수 있었다. 그림 13은 이를 보여준다.

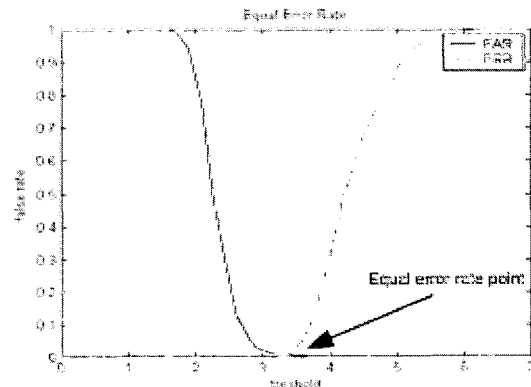


그림 12. 수평/수직 기울기비에 의한 인식 결과 동일 오류율 지점  
Fig. 12. Equal error rate point.

표 2. 기울기 성분비를 이용한 인식 결과

Table 2. Classification result using image gradients.

False alarm(%)	Miss(%)	Classification(%)
2.5	3.5	96.5

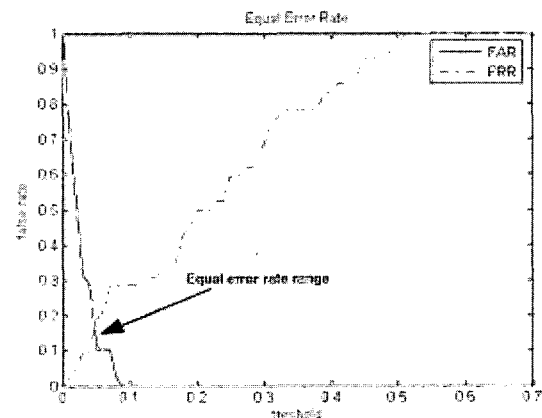


그림 13. 움직임 정보를 이용한 인식 결과 동일 오류율 지점  
Fig. 13. Equal error rate point.



표 3. 움직임 정보를 이용한 인식 결과

Table 3. Classification result using motion information.

False alarm(%)	Miss(%)	Classification(%)
5.5	6.5	93.5

표 4. 융합 방법에 따른 인식 결과

Table 4. Classification result about each fusion method.

융합방법	FRR(%)	FAR(%)	인식률(%)
MV	2.7	2.5	97.3
WAS	2.5	2.0	97.5
BDF	1.5	1.3	98.5

## (2) 검증

사람과 차량의 정면, 측면 및 대각선 방향에서 획득한 각각 200개씩의 총 1200개의 데이터를 사용하여 표 3과 같은 결과를 얻었다.

라. 세 가지 지역적 인식기를 융합한 전역적 인식 제안한 방법의 실험에서는 형상 정보에 대한 푸리에 묘사자를 특징으로 하는 신경망의 훈련과 기울기를 특징으로 하는 통계적 분석을 위해서 다양한 종류와 각도(정면, 측면 및 대각선 방향)에서 획득한 사람과 차량의 영상을 사용하였고, 모션 분석을 위해서 다양한 각도(정면, 측면 및 대각선 방향)에서 얻은 연속적인 사람과 차량의 영상을 사용하였다. 테스트 영상은 사람과 차량의 정면, 측면 및 대각선 방향의 연속적인 영상에서 획득한 각각 200개씩의 총 1200개의 데이터를 사용하였다. 본 논문이 제안하는 융합방법은 MV나 WAS와 같은 결과 융합 방법과의 비교 실험을 위해서 같은 테스트 영상을 사용하였다. 표 4는 FRR(False Rejection Rate)과 FAR(False Acceptance Rate)로써 인식률 평가 결과를 보여준다. 표 4의 결과를 보면, MV 방법을 적용한 경우 전체 인식률이 97.3%, WAS 방법은 97.5%, 제안한 Bayesian 결과 융합 방법은 98.5%를 얻어 제안한 방법의 전체 인식률이 1-1.2%로써 다소 높다. 따라서 실험 결과를 통해 제안한 Bayesian 결과 융합방법은 MV나 WAS 방법 보다 인식 성능이 우수함을 보여준다.

## VI. 결론 및 향후계획

본 논문에서는 형상 정보(푸리에 묘사자와 기울기)와 모션 정보(특징점 변화량)를 각각 특징으로 사용한 세 가지 인식기에 의한 결과를 인식기의 신뢰성에 의해 결정된 가중치를 적용하여 융합하는 Bayesian 결과 융합방법

을 제안하였다. 제안한 방법은 기존의 형상 정보나 모션 정보만을 사용한 단일 인식기에 의한 방법과 MV나 WAS의 결과 융합방법 보다 더 높은 인식성능을 얻을 수가 있었으며, 또한 기존 방법의 단점인 인식성능이 배경 영상의 상태에 따라 검출된 움직임 영역의 정확도에 의존하는 단점을 상당히 해결할 수가 있었다. 향후에는 사람의 포복하는 자세 및 앉은 자세에 대해서 인식률을 높일 수 있는 인식 알고리즘에 대해서 연구할 계획이다.

## 참고 문헌

- [1] A. J. Lipton, H. Fujiyosi and R. S. Patil, "Moving target classification and tracking from real-time video," Proc. of IEEE Workshop, on Applications of Computer Vision, pp. 8-14, 1998.
- [2] C. Stauffer and W. Grimson, "Adaptive background mixture models for real-time tracking," Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, Vol. 2, pp. 246-252, 1999.
- [3] H. Fujiyoshi. and A. J. Lipton "Real-time human motion analysis by image skeletonization", IEICE Trans, on Info & Systems, Vol. E87-D, No. 1, Jan 2004.
- [4] I. Harigaoglu, D. Harwood and L. S. Davis, "A real-time system for detecting and tracking people," Proc. of Int. Conf. on Face and Gesture Recognition, pp. 222-227, April 1998.
- [5] J. Barron, D. Fleet and S. Beauchemin, "Performance of optical flow techniques," Int. J. Comput. Vis., Vol. 12, No. 1, pp. 42-77, 1994.
- [6] Li. X. R., Zhu, Y., Wang, J., Han and C. "Optimal linear estimation fusion-Part I: Unified fusion rules," IEEE Trans. Information Theory. Vol. 49, No. 9, Sep 2003.
- [7] M. M. Kokar and J. A. Tomasik "Data vs. decision fusion in the category theory framework," Proc. 2nd Int. Conf. on Information Fusion, 2001.
- [8] R. Cutler and L. S. Davis, "Robust real-time periodic motion detection, analysis, and applications," IEEE Trans. Pattern Anal. Machine Intell., Vol. 22, pp. 781-796, Aug 2000.
- [9] R. T. Collins, A. J. Lipton and T. Kanade, "Introduction to the special section on video surveillance," IEEE Trans. Pattern Anal. Machine Intell., Vol. 22, pp. 745-746, Aug 2000.
- [10] Y. Kuno, T. Watanabe, Y. Shimosakoda and S. Nakagawa, "Automated detection of human for visual surveillance system," Proc. of Int. Conf. on Pattern Recognition, pp. 865-869, 1996.

저 자 소 개



김 정 호(학생회원)  
 1996년 해군사관학교  
 전자공학과 학사.  
 2006년 고려대학교  
 전자컴퓨터공학과 석사.  
 2006년~현재 해군 제1함대  
 안동함 기관장

<주관심분야 : 영상신호처리, 지능형 경계 감시,  
 패턴 인식, 데이터 융합>



고 한 석(정회원)  
 1982년 미국 카네기멜론대학교  
 전기공학 학사.  
 1986년 미국 메릴랜드대학교  
 시스템공학 석사.  
 1988년 미국 존스홉킨스대학교  
 전기공학 석사.

1992년 미국 카톨릭대학교 전기공학 박사.  
 1995년~현재 고려대학교 전자컴퓨터공학과 교수  
 <주관심분야 : 영상 및 음성 신호처리, 패턴 인식,  
 데이터 융합>