

논문 2010-47SP-3-2

컨텍스트 기반의 지능형 영상 감시 시스템 구현에 관한 연구

(Implementation of Intelligent Image Surveillance System based Context)

문 성 룡*, 신 성**

(Sung Ryong Moon and Seong Shin)

요 약

본 논문은 컨텍스트 기반의 지능형 영상 감시 시스템 구현에 관한 연구로써 기존 연구의 시공간적 제약성 및 실시간 처리가 어려운 단점을 보완하여 초당 30 프레임으로 이루어져 있는 저해상도 동영상(320*240)을 대상으로 다양한 환경에서 실시간 처리가 가능한 움직임 검출 및 장면 분석 알고리즘을 제안하고 이를 이용해 동영상 감시 시스템을 구축한다. 먼저 장면 분석을 수행하기 위한 전처리 과정인 움직임 검출 알고리즘에서는 연속된 프레임 중 의미 없는 유사 프레임과 배경을 제거하고 움직임 영역만을 검출하기 위해 웨이블릿 변환과 에지 히스토그램을 이용하여 샷의 경계를 검출한다. 다음으로 키프레임 선정 파라미터에 의해 샷 경계 내 대표 키프레임을 선정하며, 에지 히스토그램 및 수학적 형태론을 이용하여 움직임 영역만을 검출한다. 장면 분석 알고리즘에서는 검출된 객체의 수직 수평 비율과 질량 중심을 통해 재구성된 허프 변환 후의 각도를 이용해 독립 객체 분석을 수행하며, '서다, 걷다, 눕다, 앉다'의 4가지 기본 상황 정보를 정의한다. 또한 각 상황의 연결 상태 추정을 통해 일반 상황 및 위급 상황으로 구성되는 단순 상황 모델을 정의함으로써 장면 분석을 수행하며, 제안된 알고리즘의 실시간 처리 가능성을 확인하기 위해 시스템을 구성한다. 제안된 시스템은 저해상도 영상을 대상으로 인식률 면에서 평균 92.5%의 성능을 보였으며, 처리속도는 프레임 당 평균 0.74초로 실시간 처리가 가능함을 확인하였다.

Abstract

This paper is a study on implementation of intelligent image surveillance system using context information and supplements temporal-spatial constraint, the weak point in which it is hard to process it in real time. In this paper, we propose scene analysis algorithm which can be processed in real time in various environments at low resolution video(320*240) comprised of 30 frames per second. The proposed algorithm gets rid of background and meaningless frame among continuous frames. And, this paper uses wavelet transform and edge histogram to detect shot boundary. Next, representative key-frame in shot boundary is selected by key-frame selection parameter and edge histogram, mathematical morphology are used to detect only motion region. We define each four basic contexts in accordance with angles of feature points by applying vertical and horizontal ratio for the motion region of detected object. These are standing, laying, seating and walking. Finally, we carry out scene analysis by defining simple context model composed with general context and emergency context through estimating each context's connection status and configure a system in order to check real time processing possibility. The proposed system shows the performance of 92.5% in terms of recognition rate for a video of low resolution and processing speed is 0.74 second in average per frame, so that we can check real time processing is possible.

Keywords : Context, HSV color model, Low Resolution Image, Wavelet transform, Edge histogram detector

I. 서 론

* 정회원, ** 학생회원, 원광대학교 전자공학과

(Dept. of electronic Eng. Wonkwang University)

※ 이 논문은 2008학년도 원광대학교 주산학술연구 재단의 지원에 의해서 연구됨

접수일자: 2009년12월29일, 수정완료일: 2010년3월2일

현재 활발히 연구 되고 있는 컨텍스트(context) 기반의 지능형 영상 감시 시스템은 관리자에 의해 수동으로

대상을 감시하는 것이 아니라 상황 정보를 이용하여 자동으로 위급상황을 탐지하여 실시간 대처 가능한 수단으로 만들어 주는 것으로서 하드웨어(hardware) 및 영상처리 기술의 발달로 인해 그 효용이 실제화 되어지고 있다. 특히 최근 들어 여러 사회적 문제들로 인해 영상 보안에 대한 관심이 집중되고 있는 만큼 앞으로 그 수요는 더 늘어날 것이며, 갈수록 늘어나는 보안용 카메라를 관리자가 수동으로 감시 및 관리하는 것은 많은 비용을 초래할 뿐만 아니라, 업무 가중으로 인한 불사사를 초래할 수 있다. 따라서 향후 지능형 동영상 감시 시스템은 영상 보안 분야에 필수 불가결한 요소로 자리매김 할 것이다.

컨텍스트 기반의 지능형 영상 감시 시스템은 장면 분석 기술, 영상의 내용 기반 검색 기술, 영상 내 물체의 특징을 파악하여 특성을 검색하는 기술 등으로 구성할 수 있다. 즉 카메라 영상의 정량화된 데이터를 사용하기 때문에 효율적이고 손쉬운 관리 및 검색이 가능하지만 자동으로 사용자가 원하는 객체를 추출하여 분석하기 어려운 단점을 가지고 있다. 본 논문은 인간의 동작에 대한 기본적인 상황 패턴을 기반으로 독립 객체를 분석하여 위급 상황을 자동 인식하는 지능형 감시 시스템 구축에 관한 것으로 기존 관련 연구는 다음과 같이 크게 3가지로 구분해 볼 수 있다. 먼저 Yacoob^[1], Bregler^[2]와 같이 2차원 또는 3차원의 인체 구조화 방법을 통해 인체 각 부분의 연관성을 분석함으로써 상황을 인식하는 방법이 있다. 이러한 인체 구조화 방식은 정밀도가 높은 반면 설계가 복잡하고, 많은 제약 조건을 가져 실시간 처리가 어려운 단점을 가지고 있다. 다음으로 Bobick^[3]이 제안한 누적 움직임 정보를 이용한 방법과 Blank^[4]처럼 움직임의 외곽 라인, 즉 실루엣 기반의 동작을 이용한 인식 방법이 있다. 이는 각 프레임의 움직임 정보를 하나의 영상으로 만들어 표현함으로써 복잡한 모델링 과정 없이 좋은 인식 결과를 얻을 수 있는 장점이 있으나, 기준이 되는 템플릿과 대상 동영상 간의 높은 시간적·공간적 제약성을 가지며 실루엣의 특성상 외형적 의존도가 높아 주변 환경의 영향을 많이 받는 단점이 있다. 마지막으로는 Efros^[5]와 같이 옵티컬 플로어(optical flow)를 이용해 동작 자체에 대한 영상 분석으로 상황을 인식하는 방법이 있다. 이는 옵티컬 플로어의 정밀도에 의존하게 되어 인식률이 감소할 가능성이 높고, 계산량이 많다는 단점이 있다^[6].

본 논문에서는 기존 연구의 시공간적 제약성 및 실사

간 처리가 어려운 단점을 보완하여 초당 30 프레임으로 이루어져 있는 저해상도 동영상(320*240)을 대상으로 다양한 환경에서 실시간 처리가 가능한 움직임 검출 및 장면 분석 알고리즘을 제안하고 이를 이용해 기본 상황을 정의함으로써 컨텍스트 기반의 동영상 감시 시스템을 구축한다. 먼저 장면 분석을 수행하기 위한 전처리 과정인 움직임 검출 알고리즘에서는 연속된 프레임 중 의미 없는 유사 프레임과 배경을 제거하고 움직임 영역만을 검출하기 위해 웨이블릿(wavelet) 변환과 에지 히스토그램(edge histogram)을 적용하여 샷(shot)의 경계를 검출한다. 다음으로 키프레임(key-frame) 선정 파라미터(parameter)에 의해 샷 경계 내 대표 키프레임을 선정하며, 에지 히스토그램 및 수학적 형태론을 이용하여 움직임 영역만을 검출한다. 다음으로 장면 분석 알고리즘에서는 검출된 객체의 수직 수평 비율과 질량 중심을 통해 재구성된 허프(Hough) 변환 후 각도를 이용해 독립 객체 분석을 수행하며, '서다, 걷다, 눕다, 앉다'의 4가지 기본 상황 정보를 정의한다. 또한 각 상황의 연결 상태 추정을 통해 일반 상황 및 위급 상황으로 구성된 단순 상황 모델을 정의함으로써 장면 분석을 수행한다. 최종적으로 제안된 알고리즘의 실시간 처리 가능성을 확인하기 위해 시스템을 구성하고, 4가지 기본 상황 및 단순 상황 모델에 대한 모의 실험 결과를 비교하여 성능을 평가하고 실시간 처리 가능성을 확인한다.

II. 움직임 검출 알고리즘

본 논문에서는 장면 분석을 수행하기 위한 전처리 과정으로 움직임 검출 알고리즘을 제안한다. 이는 연속된 프레임 중 의미 없는 유사 프레임과 배경을 제거하고 움직임 영역만을 검출하기 위한 것으로 크게 네 단계로 나누어 볼 수 있다. 첫 번째 단계에서는 입력된 영상을 공간컬러모델을 이용하여 HSI 컬러모델로 변환하며, 두 번째 단계에서는 이전 단계의 변환된 I(intensity) 공간에 웨이블릿 변환과 에지 히스토그램을 적용하여 샷의 경계를 검출한다. 세 번째 단계에서는 키프레임 선정 파라미터를 이용해 검출된 샷 경계 내 대표 키프레임을 추출하며, 최종적으로 에지 히스토그램 및 수학적 형태론을 이용하여 움직임 영역만을 검출한다.

1. 공간컬러모델 및 HSV 컬러모델

공간 컬러 모델(space color model)은 일반적으로 받

아들여지는 표준적인 방법으로, 좌표 시스템의 하나이며 각각의 색을 한 점으로 나타내는 좌표시스템의 부분 집합이다. 컬러모델의 종류로는 RGB, CMY, YCbCr, YPbPr, HSI, HMMD 등이 있으며, 이 중 HSI 컬러모델은 영상을 가장 근본적으로 구별지어주는 색상, 채도, 명도의 3가지 특징을 포함하고 있다. 본 논문은 RGB 컬러모델을 HSI 컬러모델로 변환하며, 히스토그램 연산 및 에지 검출이 용이한 I(intensity, 명도) 공간에 알고리즘을 적용한다. RGB 컬러모델에서 HSI 컬러모델로의 변환식은 식 (1)과 같다^[7].

$$\begin{aligned}
 V &= \frac{1}{3}(R + G + B) \\
 S &= 1 - \frac{3}{(R + G + B)}[\min(R, G, B)] \\
 H &= \cos^{-1} \left[\frac{\frac{1}{2}[(R - G) + (R - B)]}{\sqrt{(R - G)^2 + (R - B)(G - B)}} \right]
 \end{aligned} \tag{1}$$

2. 샷 경계 검출

동영상은 독립된 개별 프레임의 연결에 의해 만들어진다. 여기서 동영상 자체는 비구조화 된 데이터 집합으로 보다 손쉬운 접근을 위해 동영상을 의미있는 데이터의 집합으로 구분지어 줄 필요가 있다. 일반적으로 동영상에 장면전환이 이루어지는 부분을 컷(cut)이라고 하며, 하나의 카메라로 촬영되어진 객체의 움직임과 관련된 일련의 연속된 프레임 집합을 샷이라고 한다. 이때 동영상을 의미있는 데이터의 집합인 샷으로 분할하는 과정을 샷 경계 검출 또는 장면 분할이라고 하는데 대부분의 동영상이 경계부분에서 커다란 변화를 보여주기 때문에 영상처리 기술에 의해 자동으로 분류가 가능하다. 그림 1은 프레임과 샷으로 이루어진 비디오의 구조를 보여준다^[8].

그림 1에서 전체 동영상은 연속된 프레임의 집합으

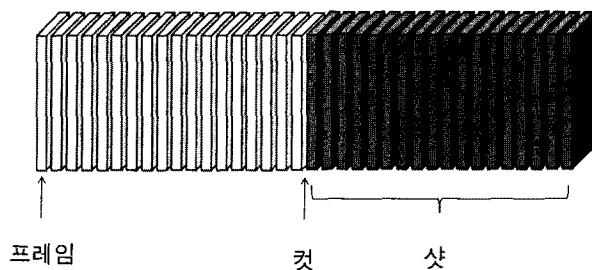


그림 1. 동영상 구조
Fig. 1. Video structure.

로 샷 경계를 제외한 인접한 프레임 사이의 유사도가 매우 높다. 따라서 장면 분석과 같은 응용 분야에 적용할 때 매 프레임을 모두 처리하는 것은 계산량 측면에서 상당히 비효율적이며 프레임간의 차이를 이용하여 샷 경계를 검출하고, 의미있는 데이터 집합으로 나누는 후, 데이터 집합 내 대표 키프레임만을 추출하여 장면 분석을 수행하면 연산량을 줄이고 시스템의 속도를 높일 수 있다.

본 논문에서는 인간의 기본 행동패턴을 앓다, 높다, 걷다, 서다의 4가지 기본 상황으로 정의하며 각각의 상황 변화가 일어나는 지점에서 프레임 내 객체의 이동 방향성 유사도가 적다는 성질을 이용하여, 이 지점을 컷으로 가정하고 샷 경계를 검출한다.

기본적으로 샷 경계는 프레임 간 차이를 측정하고 측정된 차이에 의해 적절한 임계치를 선택함으로써 검출할 수 있다. 프레임 간 차이를 측정하는 방법은 일반적으로 다음의 4가지 방법이 사용되어진다.

첫 번째는 근접 프레임 간 픽셀 차의 합을 측정하는 방법으로 가장 간단하고 기본적인 방식이나 많은 샷 검출 오류의 발생 가능성이 있어 효과적이지 못하다. 두 번째는 근접 프레임 간 그레이 또는 컬러 히스토그램의 차를 측정하는 방법으로 근접 프레임들은 서로 유사한 색상분포와 그레이 레벨을 가진다는 특성을 이용해 히스토그램의 차이를 측정하는 가장 보편적인 방법이다. 이러한 방법은 공간 정보를 포함 할 수 없고, 조명 변화에 취약한 단점을 가지고 있다. 세 번째는 블록기반 비교 방법으로 각 프레임을 블록으로 나누고 근접 프레임 내 일치하는 블록의 픽셀수를 비교하여 변화를 측정하는 방법이다. 이는 객체나 카메라의 변화에 민감한 단점을 가지고 있다. 마지막으로 네 번째는 각 프레임을 8*8 블록으로 나눈 DC이미지의 히스토그램에 대한 카이 제곱(chi-square) 값을 이용하는 방법으로 DCT 변환의 상수 정보를 이용해 점진적인 샷의 변화를 찾을 수 있는 장점이 있으나 정확도가 떨어지는 단점이 있다.

본 논문에서는 실시간 입력되는 동영상을 웨이블릿 변환 후 MPEG-7의 에지 히스토그램 기술자(Edge Histogram Descriptor, EHD)를 이용해 샷 경계를 검출한다. 본 논문에서 제안된 방법은 Wavelet 변환 후 영상의 고유정보를 유지하고 있는 근사공간(LL영역)만을 이용함으로써 처리 시간 측면에서 효율성을 높일 수 있으며, 에지 히스토그램을 이용함으로써 외부 조명의

변화에 강인하고, 사람의 이동과 같은 점진적으로 변화하는 동영상의 샷 검출에 우수한 성능을 보인다.

가. 웨이브렛 변환의 적용

본 논문에서 원영상의 고유정보를 유지하며 데이터의 양의 감소하기 위해 적용한 웨이브렛 변환은 “시간-스케일(time-scale)” 변환으로 다해상도 신호처리가 가능하여 영상 압축 및 다양한 분야에 이용되는 변환 방법이다. 이는 모든 주파수의 국부적 해상도를 나타내기 위해 적절한 방법으로 신호가 연속 변수의 함수라면 식 (2)와 같이 정의된다^[9].

$$\psi_{a,b}(x) = \frac{1}{\sqrt{a}} \psi\left(\frac{x-b}{a}\right) \quad (2)$$

여기서 a 는 웨이블릿의 크기에 관련된 압축계수이며, b 는 시간축 이동에 관련된 전이계수이다. 웨이브렛 변환을 이분 구간으로 샘플링 할 경우 직교 웨이브렛을 구성할 수 있는데 이와 같은 웨이브렛 변환을 이산 웨이브렛 변환(DWT)라한다. 이산 웨이브렛 변환은 이분 주파수 대역으로의 투영을 의미하며 다해상도 분석방법이 제시되면서 다양한 분야에 적용되고 있다. 분해에 사용되는 스케일 함수와 웨이브렛 함수는 식 (3)과 같이 정의할 수 있다.

$$\begin{aligned} \phi(t) &= \frac{1}{\sqrt{2}} \sum_k h(k) \phi\left(\frac{t}{2} - k\right) \\ \psi(t) &= \frac{1}{\sqrt{2}} \sum_k g(k) \psi\left(\frac{t}{2} - k\right) \end{aligned} \quad (3)$$

다해상도 분석은 저역필터와 고역필터를 번갈아 수행하게 되는데, 2 차원 영상에 대한 이산 웨이브렛 변환은 4 개의 주파수 대역(LL, LH, HL, HH)을 얻게 된다. 그 중 LL는 근사 공간으로 영상의 중요정보들을 모두 포함하고 있어 이 영역만을 이용함으로써 영상의 크기를 효과적으로 줄여 적은 연산량으로 샷 검출을 수행할 수 있다. 식 (4)는 본 논문에서 기저 웨이브렛으로 사용한 Haar 웨이브렛으로 속도가 빠르고 설계가 간단한 장점을 가지고 있다.

$$\psi(t) = \begin{cases} 1 & 0 \leq t < 1/2 \\ -1 & 1/2 \leq t < 1 \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

그림 2는 원영상이고, 그림 3은 Haar 웨이브렛을 이용한 변환 결과 영상이다. 근사공간인 LL영역은 저주파

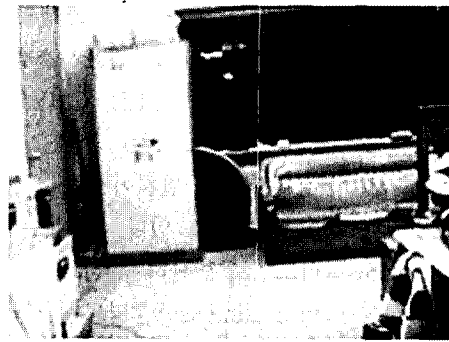


그림 2. 원영상 (320*240)

Fig. 2. Original image (320*240).

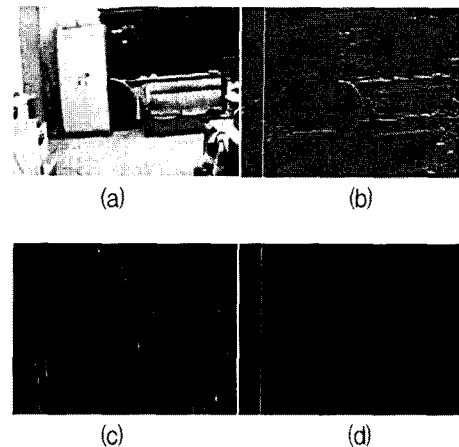


그림 3. 웨이브렛 변환 결과.

(a) LL, (b) LH, (c) HL, (d) HH

Fig. 3. Results of wavelet transform.

(a) LL, (b) LH, (c) HL (d) HH.

영역으로 객체의 군집에 대해 완전한 특성을 보이는 영상의 특성으로 인해 중요정보들을 모두 포함하고 있음을 확인할 수 있다. 이 LL영역에 알고리즘을 적용하게 되면 320*240의 크기를 가지고 있는 영상을 영상의 고유정보를 유지하면서 160*120 영상으로 줄일 수 있다. 이는 이후 적용되는 알고리즘의 연산량을 고려할 때 1/4의 연산량으로 결과에 지장이 없는 동일한 효과를 얻을 수 있음을 의미한다.

나. 에지 히스토그램 기술자에 의한 샷 검출

MPEG-7의 에지 히스토그램 기술자는 영상 내 국부적 또는 전역적 에지 정보를 효과적으로 표현할 수 있는 기술자이다. 그림 4와 같이 원영상을 16개의 부영상으로 분할한 후 다시 여러 개의 영상 블록으로 나누어 그림 5의 수직, 수평, 45도, 135도, 비방향성의 5가지 에지를 마스크 연산하여 각각의 부영상에 대해 5개의 빈(bin)을 갖는 에지 히스토그램을 생성한다. 즉 16개의

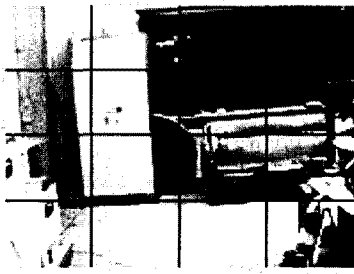


그림 4. 16개 부영상
Fig. 4. Sixteen sub-image.

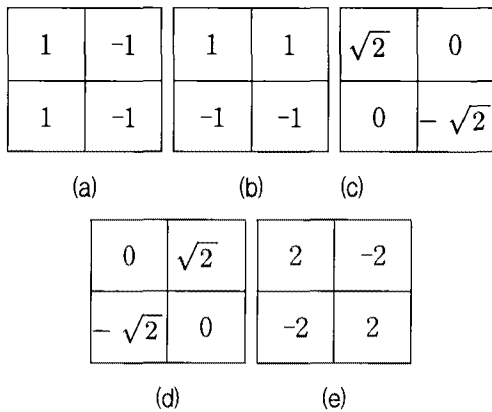


그림 5. 5가지 에지 종류. (a) 수직, (b) 수평, (c) 45도, (d) 135도, (e) 비방향성 에지
Fig. 5. Five kinds of edge. (a) vertical, (b) horizontal, (c) 45°, (d) 135°, (e) undirected edge.

부영상에 대해 5개의 에지 빈이 추출되므로 총 80개의 에지 히스토그램 빈이 추출된다. 각각의 영상 블록은 다시 4개의 부분블록으로 나누어지며 마스크 연산되어 나온 값들 중에서 가장 큰 값을 갖는 에지 성분을 영상 블록의 방향으로 결정한다^[10].

본 논문에서는 16개 부영상의 5가지 에지방향 성분을 나타내는 히스토그램 빈의 총 평균을 구해 사용한다. 따라서 객체의 움직임, 즉 앉다, 눕다, 걷다, 서다가 진행 중일 때에는 비록 위치가 변화하여도 유사한 에지 성분으로 인해 변화가 적어 임계치를 줄일 수 있으며, 상태의 변이가 일어날 때 에지의 성분 변화가 발생하여 행동 패턴에 대한 경계를 검출 할 수 있다. 정의는 식 (5)와 같다.

$$edge \mathbb{X}_{tot} = mean \left[\sum_{i=1}^{16} (edge \mathbb{X}_i) \right] \quad (5)$$

, $\mathbb{X} = 0, 45, 90, 135, non$

그림 6은 영상별 16개의 부영상에 대한 80개의 에지

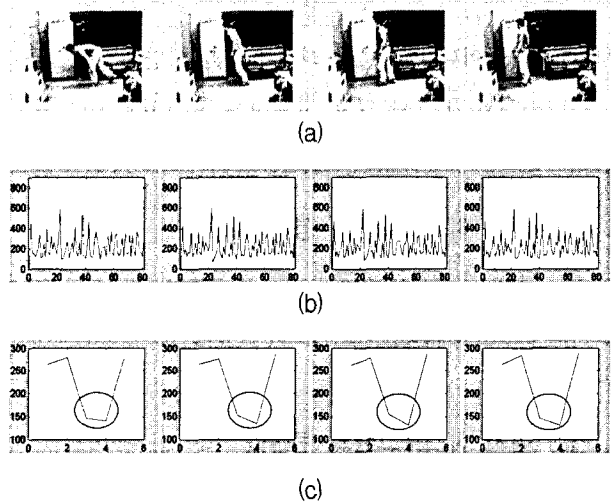


그림 6. 행동 패턴에 의한 에지 성분. (a) 움직임 영상, (b) 80개의 에지 히스토그램, (c) 히스토그램 빈의 평균
Fig. 6. Edge component by behavior patterns. (a) motion images, (b) edge histogram of 80bin, (d) average of histogram bin.

히스토그램 빈과 빈의 평균을 구한 값으로 객체가 동일 방향으로 진행할 경우에는 (a)의 두 번째, 세 번째, 네 번째 영상과 같이 위치가 달라도 에지 히스토그램이 유사함을 확인할 수 있으나 첫 번째 영상처럼 앉다, 눕다, 걷다, 서다의 다른 행동패턴으로 변이가 일어나는 지점에서는 에지 히스토그램의 차이가 큰 것을 확인할 수 있다. 이는 (b), (c)를 비교해보면 구체적으로 확인할 수 있는데 80개의 에지 히스토그램은 움직임에 따른 히스토그램의 변화가 보이나, 빈의 평균을 구한 값은 에지의 방향성 정보에 의해서만 변화하므로 행동패턴의 변이가 일어나는 지점을 잘 표현하는 것을 확인할 수 있다. 따라서 에지 히스토그램의 평균 변화가 임계치보다 클 경우 다른 상황으로 변이가 일어난 것으로 보고 샷의 경계로 가정하여 검출을 수행할 수 있다.

3. 키프레임 추출

키프레임 추출은 구해진 샷 경계 내 객체의 움직임을 가장 잘 표현할 수 있는 대표 키프레임을 찾는 과정이다. 본 시스템은 동영상 정보를 실시간 감시하는 것으로 초당 30프레임을 모두 처리하는 것은 프레임 간 영상의 유사도를 고려하였을 때 상당히 비효율적이며, 따라서 검출된 샷 경계 내 프레임 중 객체의 움직임을 가장 잘 표현할 수 있는 대표 키프레임만을 선정하여 사용한다.

식 (6), (7), (8)는 본 논문에서 사용한 키프레임 선정 파라미터들이다. 다음 세 파라미터의 시간에 대한 국부적 최소치를 구하면 가장 평균적인 움직임을 보이는 키프레임을 검출할 수 있다.

(1) 에지 히스토그램의 평균

$$\mathbb{H}_i = \text{mean} \left[\sum_{j=1}^{16} \text{shot}_i[\text{edge}(\mathbb{X}_j)] \right] \quad (6)$$

(2) 에지 히스토그램의 분산값

$$\mathbb{H}_i = \text{mean} \left[\sum_{j=1}^{16} \text{shot}_i[\text{edge}(\mathbb{X}_j)]^2 \right] - \left[\text{mean} \sum_{j=1}^{16} \text{shot}_i[\text{edge}(\mathbb{X}_j)] \right]^2 \quad (7)$$

(3) 에지 히스토그램 간 분산의 평균

$$H_i = \left\{ \text{mean} \sum_{j=1}^5 \mathbb{H}_j \right\} \quad (8)$$

4. 움직임 추출

다음으로 추출된 키프레임 영역에 대해 배경을 제거하고 움직임 영역만을 추출하기 위해 에지 히스토그램 기술자를 에지 검출에 활용하는 방법을 제안한다. 에지 히스토그램 기술자는 대부분 히스토그램 빈값 측정을 통한 유사도 구별에 사용하는데 에지 히스토그램의 부블록은 가장 강한 성분의 국부적 에지 정보를 포함하고 있기 때문에 이를 바로 에지 추출에 이용하면 특정 방향성 정보가 아닌 전체 영상에 대해 세밀한 에지 추출이 가능하여 차영상에 의한 움직임 검출이 용이하다. 또한 별도의 에지 검출 알고리즘을 적용할 필요가 없이 이전 단계에서 계산된 정보들을 이용해 바로 에지를 검출할 수 있어 계산량 측면에서도 효과적이라 할 수 있다. 제안된 에지 추출 방법은 생성된 에지 히스토그램의 빈 측정값, 즉 그림 5에서 가장 큰 값을 갖는 에지 성분의 마스크 연산 결과값을 부블록 위치의 새로운 화소값으로 삼아 국부적으로 가장 강한 에지 영상을 생성한다. 그림 7은 소벨(Sobel), 프리윗(Prewitt), 로버츠(Roberts) 연산자와 에지 히스토그램을 이용한 에지 검출 결과를 보여준다. (a), (b), (c)의 경우 각각의 마스크

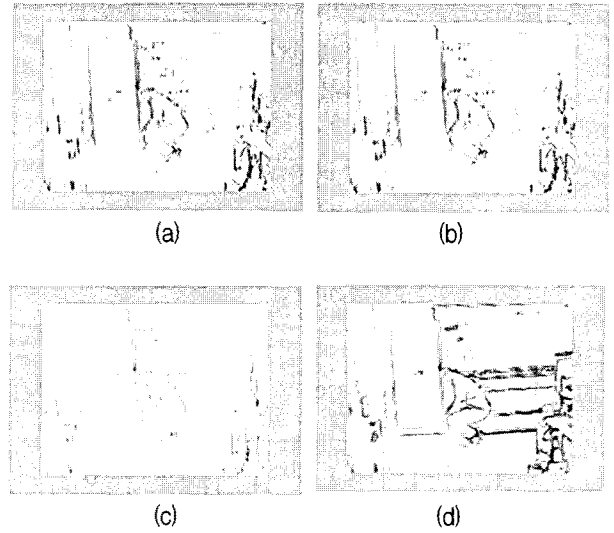


그림 7. 에지 검출 결과. (a) 소벨, (b) 프리윗, (c) 로버츠, (d) 에지 히스토그램
Fig. 7. Edge detection results. (a) sobel, (b) prewitt, (c) Roberts, (d) edge histogram.

특성에 따른 방향성 에지 위주로 검출되는 것을 확인할 수 있으나 (d)의 경우는 국부적으로 가장 강한 에지 성분을 포함하고 있는 에지히스토그램의 특성상 특정 방향성에 구애받지 않고 세밀한 에지 검출이 가능함을 확인할 수 있다.

본 논문에서는 메인 프레임과 키프레임을 에지 히스토그램 기술자를 이용하여 에지 검출하며 검출된 영상의 차영상을 구한다. 여기서 메인프레임은 고정된 카메라의 이동객체가 존재하지 않는 원영상으로 키프레임의 움직임 영역을 검출하기 위한 비교 데이터로 활용된다.

차영상을 구한 후 최종적으로 수학적 형태론을 이용하여 움직임 영역만을 검출하게 된다. 수학적 형태론^[11]은 영역 모양의 표현과 기술에 유용한 경계선, 골격 등을 검출할 때 유용하게 활용되는 방법이다. 형태론의 가장 기본적인 연산으로 팽창 연산은 에지 외곽픽셀을 증가 시키는 역할을 하며 침식 연산은 에지 외곽 픽셀을 감소시키는 역할을 한다. 본 논문에서는 에지의 내부화소를 제거한 후 선분의 끝점(1인접 연결화소)에서 8인접 연결 성분 이전 3 픽셀의 수평 수직 대각의 방향성 정보를 이용해 가장 큰 방향으로 1 픽셀씩 팽창연산을 수행한다. 연산은 진행방향과 진행방향 좌우의 3방향 픽셀을 검색하여 진행하며 추가된 연결선분과 연결선분이 만나면 만나는 지점에서 종료된다. 또한 이전 픽셀이 3 픽셀 이상이 안 되거나 15 픽셀이상 연결 선분이 만나지 않으면 에지는 소거되며, 만약 진행 경로

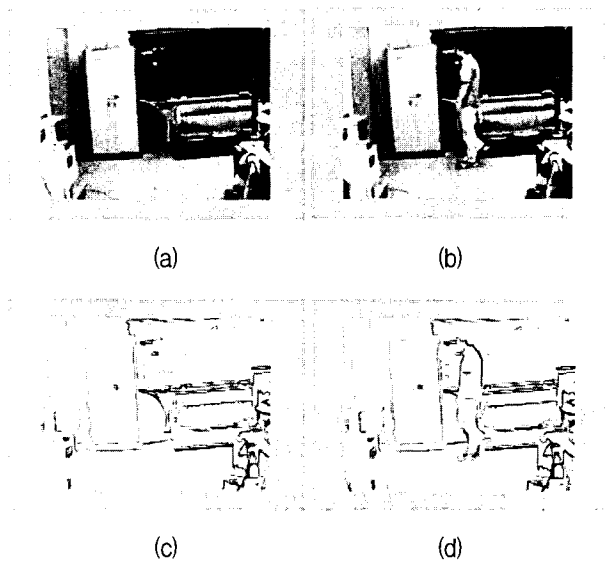


그림 8. 메인프레임과 키프레임의 에지 검출. (a) 메인 프레임, (b) 키프레임, (c), (d) 에지 검출
 Fig. 8. Edge detection of main-frame and key-frame. (a) main-frame, (b) key-frame, (c), (d) 에지 검출.

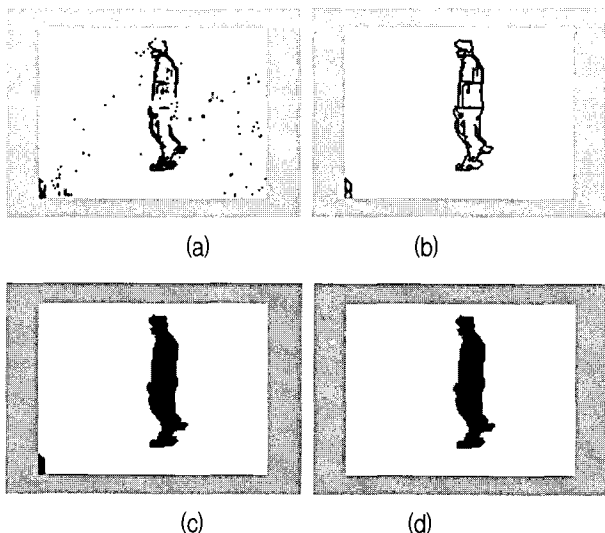


그림 9. 움직임 검출 결과. (a) 차영상, (b) 방향성 팽창 적용, (c) 조건적 팽창 적용 (d) 움직임 검출 영상
 Fig. 9. Motion detection results. (a) difference image, (b) directed dilation, (c) conditional dilation, (d) motion detection image.

상에 픽셀이 존재하면 3픽셀 이하라도 연결 선분으로 지정한다.

식 (9)는 각각 팽창 및 침식 연산에 대한 정의이다.

$$\begin{aligned}
 A \oplus B &= \{z (\hat{B})_z \cap A \neq \phi\} \\
 A \ominus B &= \{z (B)_z \cap A^c = \phi\}
 \end{aligned}
 \tag{9}$$

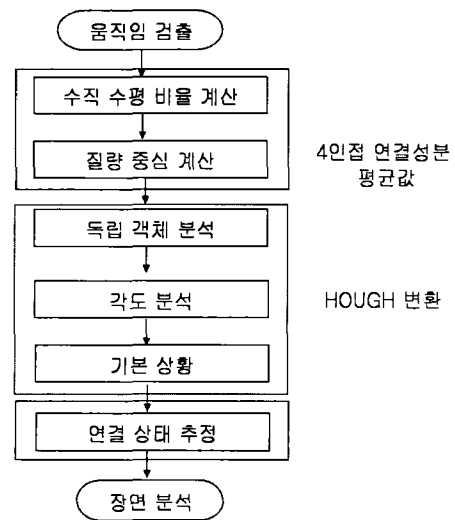


그림 10. 제안된 움직임 검출 알고리즘 블록다이어그램
 Fig. 10. Block diagram of proposed motion detection algorithm.

이때 팽창연산을 에지의 내부로 국한시키는 조건적 팽창 연산을 수행하게 되면 닫힌 에지에 대해 에지 내부를 채우는 효과를 얻을 수 있으며 정의는 식 (10)과 같다. 이후 가장 큰 군집만을 선택함으로써 움직임 영역만을 얻을 수 있다.

$$X_k = (X_{k-1} \oplus B) \cap A^c, k = 1, 2, 3, \dots
 \tag{10}$$

그림 8은 메인 프레임과 키프레임을 에지 히스토그램을 이용하여 에지 검출한 영상이다.

그림 9의 (a), (b), (c), (d)는 각각 차영상, 방향성 팽창 적용, 조건적 팽창 적용 및 움직임 검출 결과 영상이다.

그림 10은 제안된 움직임 검출 알고리즘의 블록다이어그램이다.

III. 장면 분석 알고리즘

본 논문에서는 장면 분석을 위하여 검출된 객체의 움직임 영역에 수직 및 수평 비율에 따른 특징 공간을 정의하며, 각각의 비율은 이진화 영역의 수직 수평 중간점을 연결한 평균 특징점(질량 중심)과 허프 변환 후 각도를 이용한 독립객체 분석을 통해 ‘서다, 걷다, 눕다, 앉다’의 4가지 기본 상황 정보로 재구성된다. 여기서 ‘앉다’의 경우는 의자 등 물체 위에 앉는 것을 제외한 모든 상황을 통칭한다. 최종적으로 각 상황의 연결 상태 추정을 통해 일반 상황 및 위급 상황으로 구별되는

단순 상황 모델을 정의함으로써 장면 분석을 수행한다.

1. 수직 수평 비율 및 질량 중심

먼저 II장에서 검출된 움직임 영역, 즉 키프레임 이진화 영역에 대해 질량의 끝점을 이용 수직 수평 비율을 계산한다. 여기서 수직 수평 비율은 유클리드 거리를 이용하여 실험에 의해 결정된 국부적 임계치인 3:7, 7:3, 5:5에 수렴하게 되며, 이 이진화 영역에 대해 픽셀당 4인접 연결성분의 수직 수평 평균값을 통해 질량의 중심을 계산하게 된다. 그림 11은 질량중심을 통한 좌표축의 중심점을 보여주고 있다.

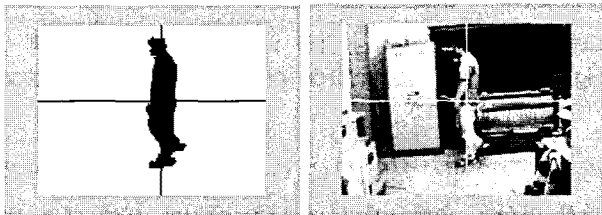


그림 11. 좌표축의 중심점
Fig. 11. Core-point of coordinate axis.

2. 독립 객체 분석

다음으로 질량 중심을 좌표축의 중심으로 삼아 허프 변환을 이용하여 움직임 영역의 가장 긴 직선라인을 검출하여 독립객체를 분석한다.

허프 변환은 직선상의 점들을 파라미터 평면으로 허프 변환을 수행하면, 하나의 점으로 표현할 수 있다는 것이다. 따라서, 같은 직선상의 점들은 허프 변환한 파라미터 평면에서의 어떤 한 점에서 만나게 된다는 것으로 식 11과 같은 기울기-절편과 각-거리 수식으로 표현된다^[12].

$$y_i = ax_i + b \tag{11}$$

$$x \cos \theta + y \sin \theta = \sigma$$

이때 질량 중심축과의 가장 긴 직선라인과의 각도 θ 를 분석하여 '서다, 걷다, 눕다, 앉다'의 4가지 기본 상황이 정의된다.

그림 12는 좌표축 중심점(질량 중심)으로부터의 각도를 보여주고 있으며, 표 1은 수직수평 비율과 각도에 따른 독립객체 분석으로 4가지 기본상황을 나타내고 있다. 여기서 각도는 다양한 움직임 영상을 실험한 결과에 의해 결정되었다.

그림 13은 4가지 독립 객체 분석에 따른 기본 상황의

좌표축 중심점으로부터의 각도

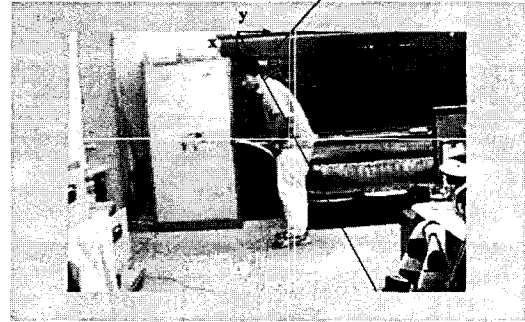


그림 12. 좌표축 중심점에 따른 각도
Fig. 12. Angle using core-point of coordinate axis.

표 1. 독립 객체 분석 및 기본 상황
Table 1. Basic context and independent object analysis.

비율(수평:수직)	각도	기본 상황
3:7 수렴	0-20도	서다
	20도 이상	걷다
7:3 수렴	70-110도	눕다
5:5 수렴	30-150도	앉다

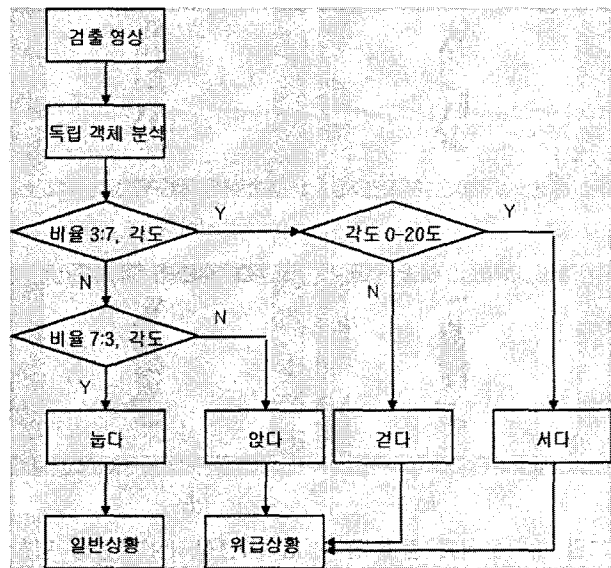


그림 13. 독립객체 분석 과정
Fig. 13. Independent object analysis process.

연산과정을 보여주고 있다.

3. 단순 상황 모델

최종적으로 각각의 프레임별 기본 상황의 연결 상태 추정을 통해 표 2와 같은 규칙을 정하고 일반상황과 위급상황으로 분류되는 단순 상황 모델을 정의함으로써 장면 분석을 수행한다.

표 2. 독립 객체 연결에 의한 단순 상황 모델
Table 2. Simple context model by independent object connection.

이전 상황	이후 상황	추정 상황	단순 상황 모델
서다	서다	정지	일반상황
	걷다	이동	일반상황
	눕다	쓰러짐	위급상황
	앉다	정지	일반상황
걷다	서다	정지	일반상황
	걷다	이동중	일반상황
	눕다	넘어짐	위급상황
	앉다	정지	일반상황
눕다	서다	상황복귀	일반상황
	걷다	상황복귀	일반상황
	눕다	상황복귀	위급상황
	앉다	상황지속	일반상황
앉다	서다	정지	일반상황
	걷다	이동	일반상황
	눕다	쓰러짐	위급상황
	앉다	정지	일반상황

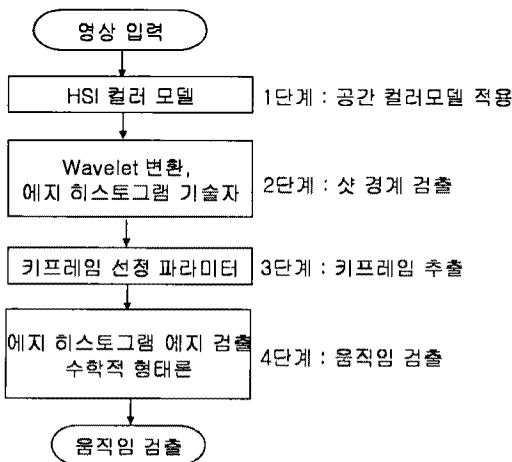


그림 14. 제안된 장면 분석 알고리즘 블록다이어그램
Fig. 14. Block diagram of proposed scene analysis algorithm.

그림 14는 제안된 장면 분석 알고리즘의 블록다이어그램이다.

IV. 시스템 구현

본 논문에서는 제안된 알고리즘의 성능 평가 및 실시간 처리 가능성을 확인하기 위해 소프트웨어와 하드웨어로 구성된 시스템을 구현하였다. 구현된 시스템은 세부적으로 Image sensor인 CCD 카메라, 컨트롤 보드 및 응용 소프트웨어로 구성된다.

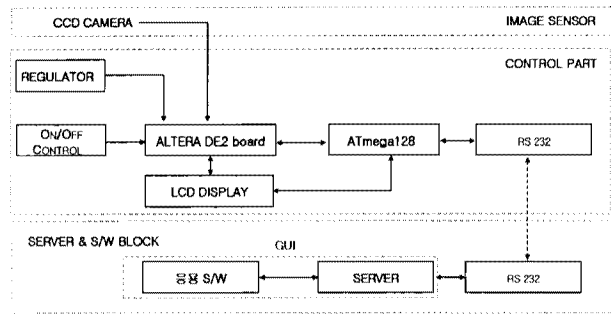


그림 15. 전체 시스템의 블록다이어그램
Fig. 15. Block diagram of overall system.

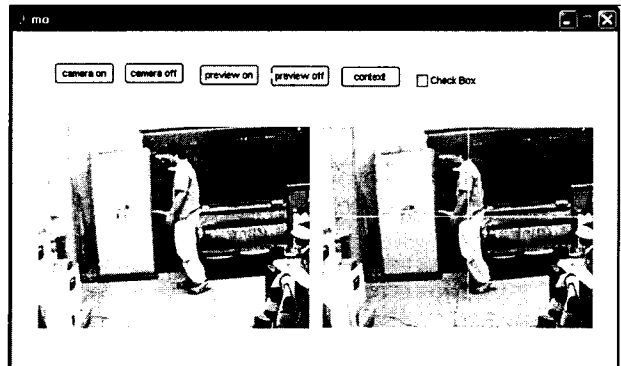


그림 16. 응용 소프트웨어
Fig. 16. Application software.

CCD 카메라를 통해 입력된 영상은 컨트롤 보드와 통신회선(RS 232)을 통해 서버로 전송되어 응용 소프트웨어 내에서 데이터베이스 저장 및 연산이 이루어지며 위급상황이 존재할 경우 소프트웨어 및 LCD 패널을 통해 사용자에게 알려주게 된다.

그림 15는 소프트웨어 상에 구현된 전체적인 시스템의 최종 블록다이어그램이고, 그림 16은 개발된 응용 소프트웨어이다.

V. 모의 실험

본 논문에서 개발한 영상 감시 시스템은 초당 30 프레임으로 이루어져 있는 저해상도 동영상(320*240)을 대상으로 하였으며, Intel Core2 Duo, CPU 3.0Ghz, 2GB RAM 환경에서 Matlab 7.0을 이용하여 실험하였다.

그림 17은 키프레임 검출 영상의 일부분으로 30초 동안(900 프레임) 실험한 결과 총 38프레임으로 초당 평균 1.27프레임이 검출됨을 확인할 수 있었다.

그림 18은 '서다, 앉다, 눕다, 걷다'의 대표적 기본 상황의 움직임 검출 결과 영상이다.

그림 19는 독립객체 분석을 통한 단순상황모델의 4

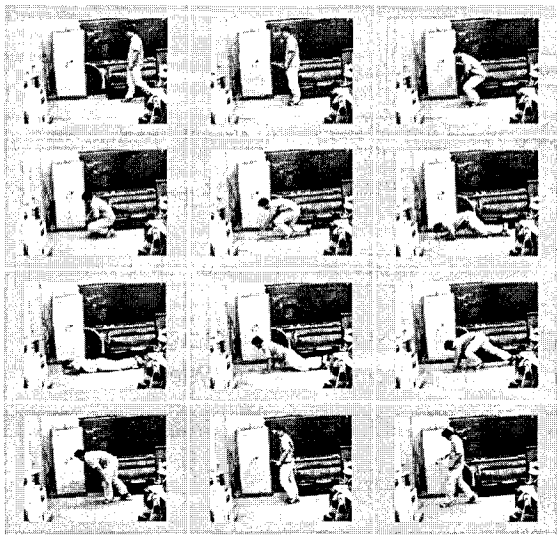


그림 17. 키프레임 검출 결과
Fig. 17. Key-frame detection results.

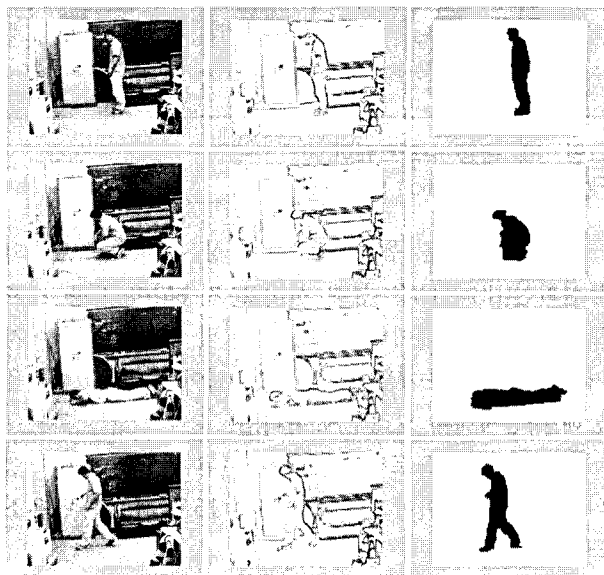


그림 18. 움직임 검출 결과
Fig. 18. Motion detection results.

가지 인식 결과 영상을 응용 소프트웨어 내에서 보여 주고 있다. (c)번의 경우 ‘앉다’ 후 ‘눕다’ 경우로 쓰러짐에 해당되며 위급상황 경보가 발생하는 것을 확인할 수 있다.

그림 20은 배경, 조명, 거리 등이 다른 다양한 환경에서의 위급상황 인식 결과이다.

표 3은 4가지 기본 상황 및 단순 상황 모델에 대한 인식률 비교표이다. ‘서다’와 ‘걷다’의 경우 비슷한 질량 중심과 각도를 가지고 있어 독립객체 분석시 ‘눕다’나 ‘앉다’보다 낮은 인식률을 보이나 최종 결과, 즉 일반상

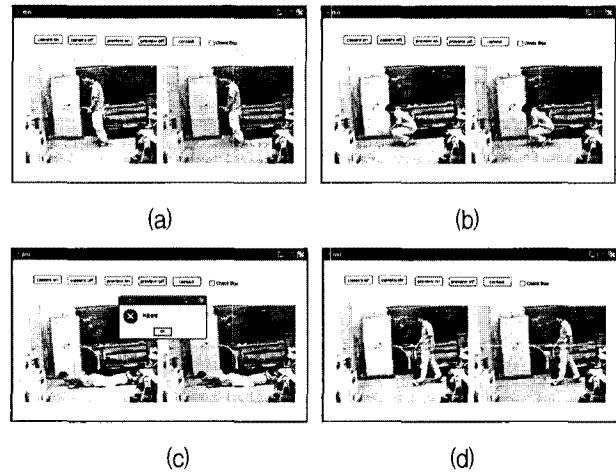


그림 19. 단순상황모델의 4가지 인식결과.
(a) 서다, (b) 앉다, (c) 눕다, (d) 걷다
Fig. 19. 4 Results of simple context model.
(a) stand, (b) sit, (c) lie, (d) walk.

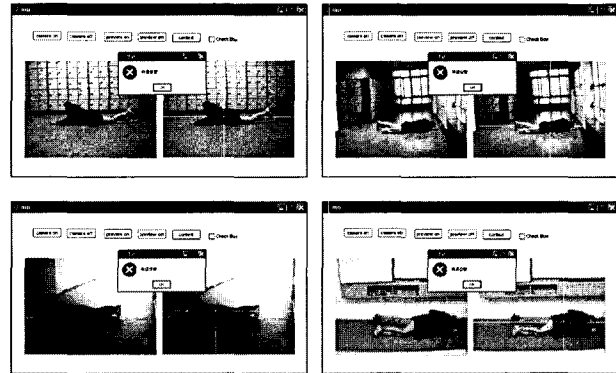


그림 20. 다양한 환경에서의 위급상황 인식
Fig. 20. Emergency context recognition in various environments.

표 3. 기본 상황 및 단순 상황 모델의 인식률 비교
Table 3. Identification rate comparison of basic context and simple context model.

기본 상황	인식률 (%)	단순 상황 모델	인식률 (%)	속도(s)
서다	89.3	일반상황	89.1	0.73
걷다	88.0			0.74
앉다	90.1			0.76
눕다	95.6	위급상황	95.6	0.73
평균	90.75	평균	92.35	0.74

황과 위급상황으로 구분되는 단순 상황 모델의 경우 큰 임계치 변화로 인해 92.5%로 높아짐을 확인할 수 있다. 또한 전반적인 처리속도는 저해상도 영상(320*240)을 대상으로 프레임 당 평균 0.74초로 실시간 처리에 무리가 없음을 확인하였다.

VI. 결 론

본 논문은 컨텍스트 기반의 지능형 영상 감시 시스템 구현에 관한 연구로써 기존 연구가 주변 환경 등의 제약 조건으로 인해 다양한 분야에 적용이 불가능하고, 많은 계산량으로 인한 실시간 처리의 어려움을 보완하여 저해상도 영상을 기반으로 실시간 처리가 가능한 영상 감시 시스템을 구현하였다.

제안된 시스템은 크게 움직임 검출 알고리즘과 장면 분석 알고리즘의 두 부분으로 나눌 수 있는데 먼저 움직임 검출 알고리즘은 연속된 프레임 중 의미 없는 유사 프레임과 배경을 제거하고 움직임 영역만을 검출하기 위한 것으로 크게 네 부분으로 나누어 볼 수 있다. 먼저 입력된 영상을 공간컬러모델을 이용하여 HSI 컬러모델로 변환하며, 이전 단계의 변환된 I(intensity) 공간에 웨이브렛 변환과 에지 히스토그램을 적용하여 샷의 경계를 검출한다. 다음으로 키프레임 선정 파라미터를 이용해 검출된 샷 경계 내 대표 키프레임을 추출하며, 최종적으로 에지 히스토그램 및 수학적 형태론을 이용하여 움직임 영역만을 검출하였다.

장면 분석 알고리즘에서는 검출된 객체의 움직임 영역에 수직 및 수평 비율에 따른 특징 공간을 정의하였으며, 각각의 비율은 이진화 영역의 수직 수평 중간점을 연결한 평균 특징점(질량 중심)과 허프 변환 후 각도를 이용한 독립객체 분석을 통해 '서다, 걷다, 눕다, 앉다'의 4가지 기본 상황 정보로 재구성되었다. 최종적으로 각 상황의 연결 상태 추정을 통해 일반 상황 및 위급 상황으로 구별되는 단순 상황 모델을 정의함으로써 장면 분석을 수행하였으며, 실시간 처리 가능성을 확인하기 위해 시스템을 구성하였다.

제안된 시스템은 저해상도 영상(320*240)을 대상으로 인식률 면에서 표 3과 같이 평균 92.5%의 성능을 확인할 수 있었으며, 전반적인 처리속도는 프레임 당 평균 0.74초로 실시간 처리에 무리가 없음을 확인하였다.

향후 검출 성능을 높일 수 있는 다해상도 영상에서의 프레임 추출 성능 및 여러 사람이 모여있는 집합장소의 보안 감시에 대한 연구가 이루어져야 하며, PTZ 카메라 등의 이동 및 제어가 가능한 카메라를 적용하여 위급상황에 대한 상시 원격 제어 및 보안감시 강화 측면에서의 기능적 연구가 필요할 것으로 사료된다.

참 고 문 헌

- [1] Y. Yacoo and M. J. Black, "Parameterized modeling and recognition of activities," *Computer Vision and Image Understanding* 73(2), pp. 232-247, 1999.
- [2] C. Bregler, "Learning and recognizing human dynamics in video sequences," *CVPR*, June 1997.
- [3] A. Bobick and J. Davis, "The recognition of human movement using temporal templates", *PAMI* 23(3), pp.257-267, 2001.
- [4] M. Blank, L. Gorelick, E. Shechtman, M. Irani and R. Basri, "Actions as Space-Time Shapes," *ICCV* pp. 1395-1402, 2005.
- [5] A. Efros, A. Berg, G. Mori and J. Malik, "Recognizing action at a distance," *ICCV Vol.II*, pp. 726-733, 2003.
- [6] 김태형, 변혜란, "동영상에서 MGH를 이용한 실시간 다수 동작 인식," *Korea Computer Congress 2006*, Vol 33, No 1, pp. 325-327, 2006.
- [7] R. Gonzalez and R. Woods, *Digital Image Processing*, Addison-Wesley, 1992.
- [8] J. S. Kim, "Text Locating from Natural Scene Movie Using Image Intensities," *proc. of the 2005 Eight Int. Conf. on Document Analysis and Recognition*, Vol. 5, pp. 1520-5263, 2005.
- [9] R. M. Rao and A. S. Bopardikar, *Wavelet Transforms : Introduction to Theory and Applications*, Addison- Wesley, 1998.
- [10] B. S. Manjunath, P. Salembier and T. Sikora, *Introduction to MPEG-7 : Multimedia Content Description Interface*, Wiley, 2002.
- [11] I. Pitas, *Digital Image Processing Algorithms and Applications*, Wiley, 2000.
- [12] J. Cha, R. H. Cofer and S. P. Kozaitis, "Extended Hough Transform for Linear Feature Detection," *Pattern Recognition*, Vol. 39, pp. 1034-1043, June 2006.

— 저 자 소 개 —



문 성 룡(정회원)
 1982년 원광대학교 전자공학과
 학사 졸업.
 1986년 전북대학교 전자공학과
 석사 졸업.
 1993년 전북대학교 전자공학과
 박사 졸업

1994년 원광대학교 전자및제어공학부 교수
 2001년 1월~2001년 12월 Sydney Univ. 교환교수
 2009년 1월~2009년 8월 UTS Univ. 교환교수
 <주관심분야 : 신경망 이론, 퍼지, 얼굴 인식,
 디지털 시스템 및 지능 시스템 >



신 성(학생회원)
 2004년 원광대학교 전기전자 및
 정보공학부 학사 졸업
 2006년 원광대학교 전자공학과
 석사 졸업.
 2008년 원광대학교 전자공학과
 박사과정 수료

<주관심분야 : 얼굴인식, 객체인식, 지능 시스템,
 영상보안>