

# 선택적 주의 기법 기반의 영상의 기대효과 자동생성

윤종철<sup>o</sup>이인권<sup>†</sup>

연세대학교 컴퓨터과학과

media19@cs.yonsei.ac.kr<sup>o</sup>, iklee@yonsei.ac.kr<sup>†</sup>

## Perception based video anticipation generation

Jong-Chul Yoon

In-Kwon Lee

Dept. of Computer Science, Yonsei University

### 요약

기대효과란 행동이 시작하기 전에 반대방향으로 일어나는 행동을 위한 준비단계로 주로 이차원 애니메이션에서 행동을 강조하기 위해 사용되어 왔다. 본 논문은 영상을 매개로 하여 인식기반의 기대효과를 자동 생성함으로써 시청자의 주의를 임의의 정보에 집중시키는 방법을 제안한다. 이미지 기반의 집중점을 찾는 문제와 영상 속의 움직임을 찾는 방법을 바탕으로 집중성 강화 알고리즘을 역으로 풀어 기대효과를 생성하는 방법을 제시한다. 주요점에 정보를 반대로 약화시키는 기대효과를 통해 주요점을 더욱 역동적으로 강조 할 수 있다. 본 논문의 알고리즘을 통해 광고 등의 정보 전달을 위한 영상 또는 역동적 표현이 필요한 영상의 보정이 가능해진다.

### Abstract

Anticipation effect has been used as a traditional skill to enhance the dynamic motion of the traditional 2D animation. Basically, anticipation means the action of opposite direction which performs before the real action step. In this paper, we propose the perception-based video anticipation method to guide a user's visual attention to the important region. Using the image based attention map, we calculate the visual attention region and then combine this map with temporal saliency of video. We apply the anticipation effect in these saliency regions using the blur kernel. Using our method, we can generate the dynamic video motion which has attentive guidance.

키워드: 기대효과, 선택적주의 기법, 비디오 프로세싱

Keywords: anticipation, perception based graphics, video processing

## 1. 서론

정보전달에 있어서, 비디오와 이미지는 정보 전달자와 수용자 양쪽 모두에게 가장 간단하고 효율적인 매개체이다. 하지만 광고와 같이 정보전달을 목적으로 한 영상에서 강조하고자 하는 정보를 정확히 전달하기 위해선 많은 경험과 기술을 필요로 한다. 일반적으로 인간의 시각적 인식능력은 한계가 있어서 집중을 하고 있는 점을 기준으로 점점 외곽으로 갈수록 인식능력이 떨어질 수 밖에 없다. 때문에 수용자가 잘못된 위치

에 집중을 하게 되면 정보전달의 효율성이 떨어진다. 보다 정확한 정보전달을 위해선 숙련된 기술자에 의한 화면구도 및 색 분할, 그리고 움직임 제한 등의 방법이 필요하다. 본 논문은 이러한 작업들을 보정하기 위한 영상기반의 기대효과를 자동으로 이끌어내는 알고리즘을 제안한다.

기대효과란 일반적으로 이차원 애니메이션에서 많이 사용된 강조 기법이다. 보편적으로 이차원 애니메이션에서의 모션은 기대 효과(Anticipation)와 행동(Action), 그리고 반작용(Reaction)으로 이루어진다. 기대효과는 행동을 위한 준비

단계이고 반작용은 행동으로 인한 결과를 나타낸다. 그림 1에서 보듯이 기대효과는 화살이 앞으로 나가는 행동에 필요한 추가 행동, 즉 활을 당기는 행동에 해당한다. 이런 기대효과는 주요 행동이 있기 전에 발생하므로 수용자의 관심을 끌어 다음 행동을 예상할 수 있게 만든다 [1]. 만약 모션에 있어 기대효과가 없으면 모션의 역동성과 집중성이 떨어질 수밖에 없다. 우리는 기대효과를 비디오의 집중점을 기반으로 표현하여, 영상의 집중성과 역동성을 높이는 방법을 제안하고자 한다. 기존의 여러 연구에선 이미지의 집중점을 찾는 CSF (Contrast sensitivity function) 함수를 사용하여, 외부영역에 대한 정보를 줄여주는 LOD 기법들이 사용되었다 [2,3,4,5]. 하지만 이런 결과들은 사용자의 주의를 임의적으로 강제하는 경향이 있어서 자연스러운 영상표현을 헤칠 가능성이 있다. 따라서 임의적인 주의 강제가 아닌, 기대효과를 통한 사용자 예측을 통한 집중성 강화를 통해 좀 더 나은 결과를 기대할 수 있다.

영상에서 집중점을 찾는 문제는 크게 두 가지로 나눌 수 있다. 우선 각각의 이미지에서의 집중점을 명도의 차를 기준으로 한 이미지 분석을 통해 각각의 이미지의 집중점을 찾아낼 수 있다. 또한 영상에서 주된 물체의 움직임 정도도 고려되어야 한다. 우리는 Ma와 Zhang [6] 이 제안한 명도를 통한 가우시안 거리(Gaussian Distance)를 통해 이미지의 특징점 맵을 찾아낸 뒤, 주요 물체의 움직임을 찾아내기 위해 에고 모션(Ego-motion)을 분석하여 전체적인 영상의 집중점을 찾아내려 한다.

우리는 얻어진 집중점을 바탕으로 하여 블러에 의한 기대효과를 생성하려 한다. 이미지 또는 영상촬영에 있어 초점심도 (Depth of field)는 영상의 질을 높임과 동시에 주요한 물체를 부각시키는 장점을 가진다. 우리는 이러한 초점심도를 이용하여 그것에 역효과를 주는 형식으로 기대효과를 생성하였다.

본 논문의 구성은 다음과 같다. 2장에선 본 논문에서 사용할 영상의 집중점을 찾는 알고리즘을 할 것이고 3장과 4장에선 집중점을 바탕으로 영상에 기대효과를 추가하는 과정을 설명하겠다. 마지막으로 5장에서는 실험의 결과를 보이고 6장에서 결론을 맺도록 하겠다.

## 2. 영상의 집중점 찾기 알고리즘

영상에서 집중점을 찾는 문제는 크게 이미지 기반과 모션기반으로 나눌 수 있다. 본 논문에선 두 가지 집중도에 대한 계산을 분리한 후 하나의 식으로 합쳐서 표현하도록 하겠다.

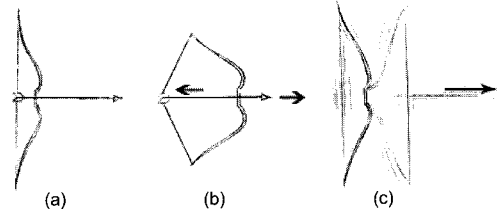


그림 1: 이차원 애니메이션에서의 모션 분석 (a) 행동 (b) 기대효과 (c) 반작용

### 2.1 이미지 기반의 집중점 찾기

하나의 이미지에서 각 픽셀당 집중이 되는 정도를 특징점 맵(Saliency map)이라 한다. 특징점 맵을 분석하여 가장 집중이 되는 점 또는 영역을 구할 수 있으며, 이것의 응용으로 이미지의 압축이나 지도와 같은 큰 이미지의 이해를 돕기 위한 도구로 많이 사용되고 있다. Itti [7] 가 제안한 알고리즘이 대표적인 집중점을 찾는 방법이다. Itti는 이미지의 색상차와 명도차 그리고 기저의 차를 바탕으로 한 특징점 맵의 제안을 통해 이미지 상의 집중도를 계산하였다. 하지만 Y.F. Ma의 실험을 통해 이미지 상에서 색상이나 모양 등은 집중도 명도차에 비해 집중도에 영향을 많이 끼치지 않는다는 것이 증명되었다 [5]. 따라서 우리는 명도차를 기반으로 한 가우시안 거리를 통해 특징점 맵을 구성하겠다. 가우시안 거리란 가우시안 피라미드를 통해 특정 단계간의 차이 즉 낮은 주기를 가지는 데이터와 높은 주기를 가지는 데이터의 차이이라 볼 수 있다.  $G_n$ 을 가우시안 피라미드의 단계라 보면 가우시안 거리는 아래와 같은 식으로 나타낸다.

$$G_{distance} = |G_n - G_{n+\Delta}|, \quad (1)$$

여기서  $n$ 은 사용자가 입력한 특정 단계의 값이고  $\Delta$ 는 거리를 구하고자 하는 단계의 차이를 뜻한다. 명도차 기반의 특징점 맵을 위해서 우리는 기존의 이미지의 컬러 공간을 LUV공간으로 이동시킨 후 위의 거리를 구하였다. 그림 2를 보면 구해진 특징점 맵을 확인할 수 있다.

### 2.2 모션기반의 집중점 찾기

앞에서 구한 특징점 맵은 단일 이미지에 대한 고려이므로, 우리는 영상의 집중점을 찾기 위해 모션을 분석하는 방법을 추가하였다. 영상에서의 모션이란 카메라의 모션과 실제 움직이는 물체의 모션으로 나눌 수 있다. 집중도에 관여하는 요소는 카메라의 모션에 따른 배경의 변화가 아닌 실제 물체의 움직임의 양이므로 두 모션을 분리해내는 에고 모션 보정 과정을 사용하겠다.

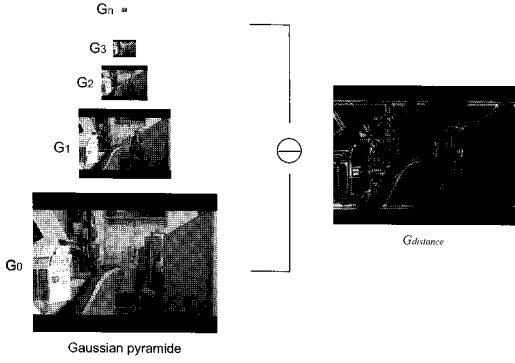


그림 2: 가우시안 피라미드를 이용한 특징점 맵 찾기: 특정 단계의 가우시안 레벨의 차이를 구하면 특징점 맵인  $G_{distance}$ 를 구할 수 있다.

영상을 다중이미지로 가정하면, 각 프레임간의 거리맵이 모션이 일어나 발생했는지 나타낸다. 하지만, 카메라가 이동하는 상태에서 촬영된 영상이라면 전체 이미지에서 거리맵이 생겨날 것이다. 따라서 카메라 자체가 이동하여 생기는 배경의 변화인  $I_{inlier}$ 를 분리한 뒤, 각 프레임  $t$ 의 카메라의 이동  $T_t$ 를 사용하여  $I_{outlier}$  차이맵을 구해내어야 한다. 이것은 아래와 같은 식으로 나타낼 수 있다.

$$O_{distance} = |I_{i,j}^t - I_{T_t(i,j)}^{t-1}| \quad (2)$$

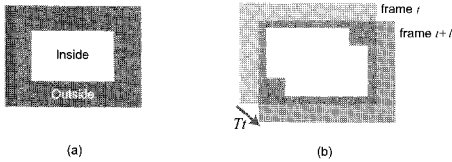


그림 3: 배경의 움직임을 통해 카메라의 움직임 정도를 얻는다. (a) 최적화 문제를 풀때 outside 영역만을 계산한다. (b) 겹쳐지는 영역의 차가 가장 작은  $T_t$ 를 구한다.

실제 각 프레임의 카메라의 움직임을 나타내는  $T_t$ 는 이동 뿐 아니라 회전 또한 포함해야 한다. 하지만 본 연구에선 정확한 카메라 이동 보다는 움직이는 오브젝트의 정도만을 필요로 하므로 간단한 픽셀 기반의 방법을 사용하도록 하겠다. [8]와 [11]의 방법을 기초로 하여 다음과 같은 최적화 과정을 통해  $T_t$ 를 구할 수 있다.  $N$ 이 겹쳐지는 영역의 픽셀 수라 할 경우, 목적함수는 다음과 같이 나타낼 수 있다.

$$\text{Minimizing} \sum_{outside} \frac{(I_{i,j}^t - I_{T_t(i,j)}^{t-1})^2}{N} \quad (3)$$

위와 같은 겹치는 영역의 픽셀값의 차이를 최소화하는 문제를 풀어, 연속되는 두 프레임 간의 차이가 최소화되는  $T_t$ 를 구할 수 있다. 즉 배경의 움직임이 최소화 되는  $T_t$ 값을 통해 실제 움직이는 물체의 움직임 정도를 맵으로 나타낼 수 있는 것이다. 위 식에서 Outside란 배경이 있을 확률이 큰 영역, 즉 이미지의 외곽 부분을 뜻한다(그림 3 (a)). 우리는 Outside의 크기를 이미지 중심을 기준으로 1/4영역을 제외한 영역으로 정의하였다. 식 3의 최적화 문제를 풀기위해선 각 프레임마다 모든 겹쳐지는 경우의 수를 다 풀어보아야 한다. 이럴 경우 최적화 문제를 풀기 위해선 각 프레임마다 모든 겹쳐지는 경우의 수를 다 풀어보아야 한다. 이럴 경우 계산량이 너무 늘어나기 때문에 처음 시작  $t_0$ 일 때만 모든 경우의 수를 계산하고 나머지  $t_i$ 일 경우  $t_{i-1}$ 프레임에서 계산된  $T_t$ 를 기준으로 5픽셀 정도의 불력만을 계산하였다. 카메라의 움직임은 연속되기 때문에 간략화된 식으로도 좋은 결과를 얻을 수 있었다.

이미지의 외곽 부분에 움직이는 오브젝트가 있을 경우를 대비하여, 위에서 구해진  $T_t$ 를 통해 얻어진  $O_{distance}$  값이 미리 정해진 에리 값  $\epsilon$  이상이 된 부분의 픽셀은 제외하고 다시 한번 같은 최소화 문제를 풀어 비교적 정확한  $T_t$ 를 계산해 낼 수 있다. 즉 아래와 같은 분류 과정을 통해 남겨진  $I_{inlier}$  픽셀만을 가지고 식 3의 최소화 문제를 다시 풀게 된다.

$$\begin{cases} I_{i,j}^t \in I_{inlier}, & \text{if } O_{distance}(i,j) \leq \epsilon \\ I_{i,j}^t \in I_{outlier}, & \text{otherwise.} \end{cases}$$

이 과정을 통해 계산된 모션의 차이맵은 그림 4에 나와있다.

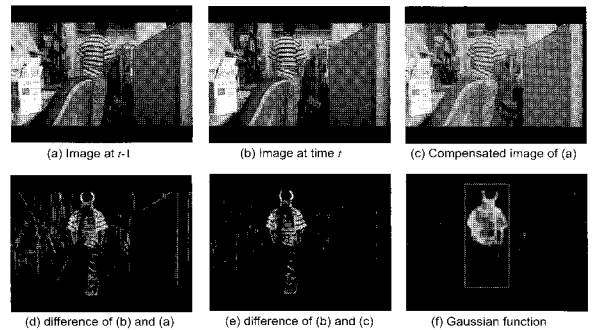


그림 4: 모션의 차이맵 : (a)와 (b)가 이웃하는 두 개의 프레임일 경우,  $T_t$ 를 통한 보정이 없으면 (d)와 같이 배경부분에 많은 노이즈가 발생한다. 따라서  $T_t$ 를 통해 (c)와 같이 보정 후 차이맵을 구하면 비교적 정확한 모션맵 (e)를 얻을 수 있다. 좀더 부드러운 차이맵을 위하여 가우시안 함수를 사용한 (f)를 사용하였다.

### 3. 집중점 기반의 기대효과 생성

본 절에선 앞에서 구한 이미지와 모션기반의 집중도를 통해 영상의 기대효과를 생성하려 한다. 본 논문에선 영상에서의 행동을 집중점의 강조라는 가정하에 시작했기 때문에 기대효과는 집중점의 약화가 된다. 집중점의 약화를 위해서 Reddy [4]가 제안한 인식 기반의 식을 통해 블러링의 커널의 크기를 조정하고, 이웃 점과의 명도 차를 제어 하려 한다.

#### 3.1 인식기반 집중도 모델

Reddy [2]에 따르면 사람의 집중하는 정도는 집중 포인트로부터 얼마나 떨어져 있는가에 따라 달라진다. 집중도가 낮아지는 정도는 아래와 같은 시각 정확도 식으로 나타낼 수 있다.

$$M(e) = 1/(1 + 0.29e + 0.000012e^3). \quad (4)$$

여기서  $e$ 값은 집중점과 현재 픽셀 사이의 사이 각을 뜻하므로, 이 값을 현재 우리가 구한 특징점 맵과 모션 차이 맵으로 대체하여 현재 픽셀이 얼마만큼의 집중력을 가지는 지 고려하고자 한다. 따라서 앞에서 구한 두 가지 맵을 통해  $e$ 를 나타내면 아래와 같다.

$$e_{i,j} = \frac{I_{size} * (1 - D(i, j))}{2 * P_f}. \quad (5)$$

$$D(i, j) = w_0 G_{distance}(i, j) + w_1 O_{distance}(i, j),$$

여기서  $P_f$ 는 픽셀의 주기, 즉 이미지의 크기를 시야각 FOV값으로 나누어 준 값이고  $I_{size}$ 는 이미지의 크기이다. 앞에서 구한 두 차이맵의 합인  $D(i, j)$ 는 상수  $w_i$ 를 통해 정규화 된다. 2로 나누는 이유는, 이미지의 중심에 집중점이 있을 때 가장 외곽이 가장 낮은 집중성을 가진다는 가정을 통해 두 차이맵의 값이 중심점에서의 거리를 나타내게 한 것이기 때문이다. 여기서 얻어진 시각 정확도 값을 통해 기대효과를 주는 과정을 설명하도록 하겠다.

#### 3.2 블러링을 통한 기대효과 표현

기존의 연구에서 이미지 영역의 불필요한 부분의 집중성을 떨어뜨리기 위해 블러링을 가하는 방법이 제안되었다 [4]. 이러한 블러링을 사용하면 초점심도와 같은 효과를 얻을 수 있으며, 이를 사용하여 주요한 부분을 강조할 수 있다. 식 (4)에서 계산된  $M(e)$ 의 값을 사용하여 인식 기반의 블러링 커널을 다음과 같이 정의할 수 있다:

$$\kappa = \frac{1}{\alpha M(e)} \quad (6)$$

본 논문에선 모션 분석에 따른 기대효과, 행동, 반작용 중 행동을 집중성을 가지는 물체의 움직임으로 가정하였기 때문에, 모션이 시작되는 위치에서 블러링을 역으로 취하여 기대효과를 표현하고자 한다. 모션이 탐지 되지 않은 상태, 즉 정적인 상태에선 시각 정확성을 통해 커널의 크기를 정의하고 모션이 탐지되는 순간부터 커널의 크기를 역으로 정하는 기대효과를 넣어주어야 한다. 기대효과가 시작되는 지점과 끝나는 지점의 연속성을 위해 우리는 식 5의  $(1 - D(i, j))$ 를 아래와 같은 식으로 대체 하고자 한다.

$$f(t)(1 - D(i, j)) + (1 - f(t))(D(i, j)). \quad (7)$$

where

$$\begin{cases} f(t) = 1 & t \leq t_{start} \text{ or } t_{end} \leq t \\ f(t) = (t - \frac{t_{start} + t_{end}}{2})^2 & t_{start} < t < t_{end}, \end{cases} \quad (8)$$

여기서  $t_{start}$ 는 기대효과가 시작되는 위치를 뜻하고,  $t_{end}$ 는 기대효과가 끝나는 위치를 뜻한다. 기대효과란 행동이 시작되기 전에 일어나는 과정이기 때문에,  $t_{start}$ 는 모션의 차이맵인  $O_{distance}$ 가 갑자기 커지기 시작하는 시점으로 정하였다. 또한  $t_{end}$ 는 현재 영상의 남은 영역에서의 비율, 즉  $(t_{start} - t_n)/\Psi$ 를 사용하여 구한다. 여기서  $\Psi$ 란 모션이 존재하는 영역을 뜻한다(그림 4(e)의 사각형). 즉, 모션 영역이 작으면 기대효과를 좀 더 길게 주어 더 쉽게 인식하게 하고, 모션 영역이 크면 기대효과를 좀 더 짧게 주어 집중도에 방해를 최소화 하였다. 간단한 이차 함수 형태인  $f(t)$ 를 통해 현재 프레임이 기대효과 지점에 있을때는 시각 정확성을 위한 차이맵이 역으로 계산되어 집중성이 떨어지는 부분은 더 명확해 지고 집중성이 많은 부분에 더 많은 블러링이 들어가게 된다.

## 4. 결과

그림 5는 기대효과가 추가된 영상과 기존 영상간의 차이를 보여준다 (a) 그림을 비교하면, 모션이 없는 정적인 상태에선 집중점 기반의 블러링을 통해 필요 없는 부분의 집중력이 떨어짐을 확인 할 수 있다. 가까이 있는 박스의 길 표면은 멀리 있는 다른 사물이 흐릿해짐에 따라 좀 더 집중력을 가짐을 알 수 있다. (b)의 경우 문이 열리는 모션이 시작됨에 따라 기대효과가 시작되는 지점이다. (b)에서 (c)로 진행할수록, 기대효과를 제어하는  $f(t)$ 가 줄어들게 되어 주위 사물이 뚜렷해지고 모션이 일어나는 부분이 흐릿해지는 기대효과가 일어난다. (c)에서 (d)로 오면서 기대효과는 점점 감소하게 되고 (e)의 경우는 기대효과는 사라지고 원래 행동인 주의점 강조가 다시 나타나게 된다. (e)의 그림을 보면 알 수 있듯이, 모션이 일어나는 부분과 집중성을 가지는 부분은 뚜렷해지고 그렇지 않은 부분은 흐릿해 짐으로써 영상에 집중력이 더 강해짐을 알 수 있다.

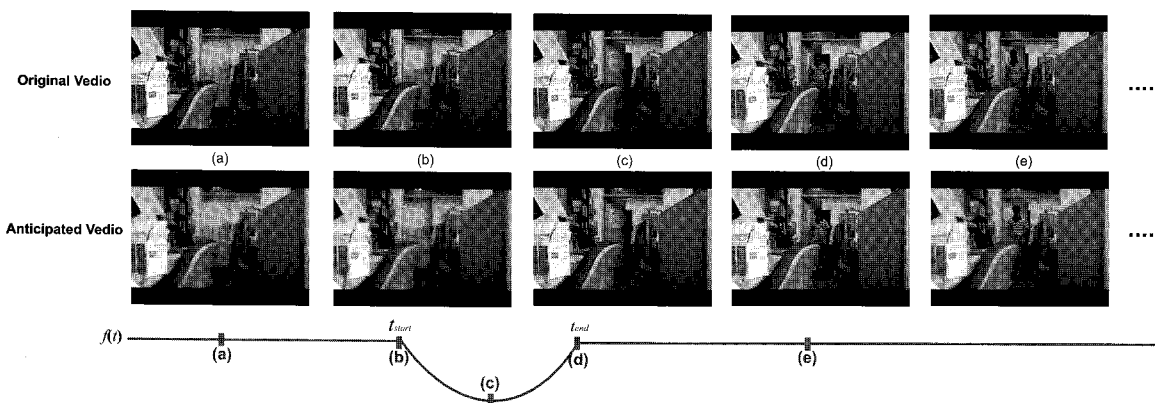


그림 5: 영상에서의 기대효과와 행동: (a) 모션이 없는 상태 (b) 모션이 시작되는 지점  $t_{start}$  (c) 기대효과가 가장 큰 지점 (d) 기대효과가 끝나는 지점  $t_{end}$  (e) 행동 상태.

그림 6는 영상에서 정적인 상태와 기대효과 그리고 행동의 과정을 이미지로 나타내었다. 정적인 상태 (a)에선 특징점 맵을 기반으로 필요 없는 부분이 블러링이 되었고, 모션이 일어나는 (b)에서는 주위 배경이 뚜렷해지고 실제 강조되어야 할 박스와 손이 흐릿해진다. 마지막으로 행동과정인 (c)에선 주위배경이 흐릿해지고 실제 강조하고자 하는 손과 박스가 뚜렷해 짐을 볼 수 있다.

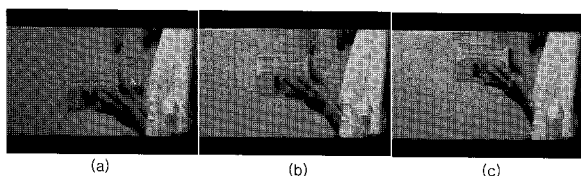


그림 6: 기대효과와 행동: (a) 정적인 상태 (b) 기대효과 (c) 행동

## 5. 결론

우리는 지금까지 영상의 집중점을 이용한 선택적 주의 기법을 통해 기대효과를 자동으로 생성하는 방법을 제안하였다. 기대효과의 생성을 통한 집중성 강화뿐만 아니라, 실제 행동 부분 즉 기대효과가 없는 부분에서도 특징점 맵과 모션맵을 기반으로 한 강조를 통해 정보 제공자가 자신이 전달하고자 하는 내용을 좀 더 쉽게 전달할 수 있다. 현재의 결과는 보완점이 많이 남아있어서 실험적 결과만 보여주었다. 만약 보완점들이 보충된다면 좀 더 나은 결과를 낼 수 있을 것이다. 우선 본 논문에선 기대효과를 주는 방법으로 블러링 효과만을 사용했다. 하지만 단순한 블러링 효과 보다는 좀 더 다양한 효과를 통해 기대효과를 높일 수 있을 것이라 예상된다. 또한

HDRI(High Dynamic Range Image)로의 확장을 통해 집중력을 더욱 강화시킬 방안을 고려 중이다. 또한 모션맵을 찾기 위해 우리는 픽셀단위의 계산을 통한 방법을 사용하였다. 하지만 이 방법은 효율성에 있어서 특징점 기반 방법에 비해 떨어진 것이다. 따라서 Particle filter [10] 와 같은 특징점 기반 모션 추적 방법을 사용하여 좀 더 효율적이고 정확한 결과를 낼 수 있을 것이다. 마지막으로, 이미지 기반의 제어이다 보니 프레임간의 연계성이 떨어질 위험성이 있다. 이 부분에 대한 고려가 필요 할 것이다.

## Acknowledgements

이 논문은 2005년도 한국학술진흥재단의 지원에 의하여 연구되었음 (KRF-2005-D00699).

## 참고 문헌

- [1] J.J. Choi, D.S. Kim and I.K. Lee, Anticipation for Facial Animation, In proceeding of Computer Animation and Social Agents(CASA),2004
- [2] M. Reddy, Perceptually Modulated Level of Detail for Virtual Environments, Doctoral thesis, Dept of Computer Science, University of Edinburgh, UK, 1997.
- [3] H. Yee, S. Pattanaik and D.P. Greenberg, Spatiotemporal Sensitivity and Visual Attention for Efficient Rendering of Dynamic Environments, ACM Transection of Graphics, 2001
- [4] M. Reddy, Perceptually Optimized 3D Graphics, IEEE Computer Graphics and Applications, 21(5):68-75,2001.
- [5] Y.F. Ma and H.J. Zhang , Contrast-based Image Attention Analysis by Using Fuzzy Growing , In proceeding of 11th

ACM international conference on Multimedia , pp.374-381, 2003.

- [6] A. Aubel, R. Boulic and D. Thalmann, , Real-time display of virtual humans: Level of details and impostors , IEEE Transactions on Circuits and Systems for Video Technology Special Issue on 3D Video Technology , 10(2):207-217 , 2000.
- [7] L. Itti, C. Koch and E. Niebur, A Model of Saliency-Based Visual Attention for Rapid Scene Analysis, IEEE Trans. on Pattern Analysis and Machine Intelligence,20(11):1254-1259, 1998.
- [8] B.Y. Jung and G.S. Sukhatme, Detecting Moving Objects using a Single Camera on a Mobile Robot in an Outdoor Environment, In proceeding of 8th Conference on Intelligent Autonomous System, pp. 980-987, 2004
- [9] M. Irani, R. Rousso and S. Peleg, Recovery of ego-motion using image stabilization, In proceeding of the IEEE Computer Vision and Pattern Recognition, pp. 454-460, 1994
- [10] C. Hue, J.L. Cadre, and P. Perez, A particle filter to track multiple objects, In IEEE Workshop on Multi-Object Tracking, pp. 61-68, 2001
- [11] D.J Lan, Y.F. Ma and H.J Zhang , A Novel Motion-Based Representation for Video Mining, In proceeding of the IEEE International conference of Multimedia and Expo, 2003