

# 시공간 입체를 이용한 등장인물 얼굴요약

박재희, 김휘용, 김성대

한국과학기술원 전자전산학과 전기및전자공학전공

## Face-Summarization using Spatio-Temporal Volume

Jae-Hee Park, Hui-Yong Kim, and Seong-Dae Kim

Div. of EE, Dept. of EECS, KAIST

E-mail : ysccllo@sdvision.kaist.ac.kr

### Abstract

본 논문에서는 동영상 내 얼굴요약을 통하여 동영상의 접근성을 향상시키고자 하는 방법을 제안한다. 얼굴요약이란 동영상에 등장한 각 사람들을 한 장씩의 얼굴영상으로 요약하는 것을 말한다. 제안하는 얼굴요약 방법은 크게 얼굴그룹생성과 대표얼굴선정의 두 과정으로 이루어진다.

동영상에서의 얼굴그룹이란 한 사람의 얼굴영상들의 집합을 의미한다. 본 논문에서는 살색화소의 시공간에서의 연속성(spatio-temporal connectivity) 및 얼굴검출기법을 이용하여 얼굴영상들을 사람에 따라 그룹화한다. 대표얼굴이란 얼굴그룹에서 그 사람을 알아보는데 가장 적당한 얼굴영상이다. 본 논문에서는 크고 정면인 얼굴을 대표얼굴로 선정하는 방법을 제안한다.

실험결과에서는 제안한 기법을 이용하여 등장인물의 등퇴장이 빈번하게 발생할 경우에도 동영상을 얼굴영상들로 요약할 수 있음을 보인다.

### I. 서론

동영상의 사용이 증가함에 따라 동영상의 접근성 향상에 대한 요구가 증대되고 있다. 이러한 접근성을 향상시키는 방법으로 특정 목적에 맞는 핵심정보를 동영상 내에서 추출하여 출력 또는 저장하는 과정인 동영상 요약의 생각이 있다. 이 때 동영상 요약의 대상이 되는 핵심정보로 동영상 내 등장 인물의 얼굴을 생각할 수 있는데 얼굴은 다양한 분야에서 중요한 정보를 제공한다.

그 대표적인 예로 화상감시를 생각할 수 있다. 화상감시 환경에서 저장된 영상은 일반적으로 장시간에 걸쳐 촬영되기 때문에, 사용자가 직접 그 내용을 확인

하기 위해서는 많은 시간이 소요된다. 따라서 효율적으로 감시 내용을 확인하기 위해서는 화상감시의 주 대상인 인물에 대한 정보를 자동으로 요약하는 기능이 필요하다. 이 때 그 요약방법의 하나로 등장인물 당 한 장씩의 얼굴 영상으로 요약하는 방법(face-summarization)은 그 내용을 작은 용량의 얼굴영상만으로 확인할 수 있다는 점에서 효율적이라 할 수 있다. 이러한 배경 하에 본 논문을 통하여 동영상 요약의 한 방법으로 등장인물 얼굴요약 기법에 대해 제안하고자 한다.

본 논문의 2 장에서는 시공간 입체를 이용하여 한 사람의 얼굴영상들의 집합으로 구성된 얼굴그룹을 생성하는 기법을 제안한다. 3 장에서는 구성된 얼굴그룹 내에서 대표얼굴을 선정하는 방법을 제안하고 4 장에서 모의 감시영상에 대하여 그 성능을 검증한다.

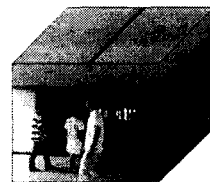


그림 1 시공간입체

### II. 얼굴그룹생성

동영상에서의 얼굴그룹이란 한 사람의 얼굴영상들의 집합을 의미한다. 본 논문에서는 기존의 얼굴검출 및 추적기법[1][2][3] 이용의 어려움을 극복하기 위하여 그림 1과 같은 시공간 입체를 이용하여 얼굴그룹을 생성한다.

#### 2.1 살색화소추출 및 살색입체구성

모든 입력영상 내 화소들을 살색화소와 비살색화소로 구분한다. 그리고 살색화소와 비살색화소로 구분된

이진영상들을 시간축 상에 나열하여 그림 2에 회색으로 나타낸 바와 같이 살색입체를 구성한다.

본 논문에서는 기존의 살색영역 임계화 기법[4]을 적용하여, 중간값 여과기(median filter)를 통과한 입력영상에서 살색화소들을 검출하였다. 이를 시간축에 나열하여 생성된 시공간 살색입체는 평면에서의 8 방향 연결성(connectivity) 정의를 공간으로 확장한 26 방향 연결성을 바탕으로 연결성을 가지는 하나 이상의 집속(cluster)으로 집속화(clustering)할 수 있다.

이 때 각각의 집속결과는 하나(그림 3(a)) 혹은 겹침이 있는 소수의 살색물체(그림 3(b))가 시공간 상에서 만들어 내는 살색입체이며 이를 고립살색입체라 한다.

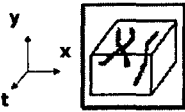


그림 2 살색입체

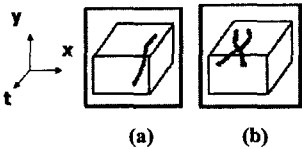


그림 3 고립살색입체

### 2.2 고립살색입체 수정

앞절에서 화소단위의 임계화 방법을 통하여 살색화소들을 획득하고, 살색입체를 얻을 수 있었다. 그러나 조명의 방향에 따라 얼굴에서 빛이 반사되거나 자기 가림(self-shading)이 생기거나 또는 영상에 잡음이 많은 경우에는 임계화를 통한 살색화소 검출결과는 실제 사람이 살색이라고 정의하는 부분을 모두 포함하지 못하는 경우가 많다. 본 논문에서는 얼굴그룹 구성의 정확도를 높이기 위하여 고립살색입체를 수정하였다.



그림 4 XT 단면생성

얼굴내부에 많은 내부 에지(edge)가 존재하거나 배경의 색상이 살색과 유사한 경우 일반적인 입력영상(시공간 입체의 XY 단면)상에서의 영역 수정 기법만으로는 비교적 정확한 결과를 얻기 힘들다. 본 논문에서는 이러한 입력영상 단위의 영역수정의 한계를 극복하기 위하여, 시공간 입체 내에서 다른 시점의 입력영상 정보

를 이용하고자 한다. 그 방법으로 3 차원 XYT 입체 상에서의 영역수정[5]을 생각할 수 있으나, 그 경우 많은 연산량이 필요하다. 본 논문에서는 시공간 입체를 입력영상의 세로축, 가로축에 수직인 평면으로 잘랐을 때의 단면인 XT, YT 단면상에서 초기살색영역을 수정하였다. XT, YT 단면을 통하여, 입력영상 내 물체의 전후 시점의 정보를 한 2 차원 영상 내에서 확인할 수 있다.

본 논문에서는 사람의 경우 가로 움직임이 많기 때문에 XT 단면상에서 X 방향 경계를 이용하여 1 차 수정 후 YT 단면상에서 Y 방향 경계를 이용하여 새로운 살색영역을 얻었는데 XT 단면 수정을 중심으로 구체적인 살색영역 수정방법을 제시하겠다.

XT 단면에서의 살색영역수정을 위하여 본 논문에서는 XT 단면의 살색영역 경계 위의 한 점  $p = (x, y, t)$  에서의 에너지  $E^{XT}(x, y, t)$  를 식(1)과 같이 정의하고 이를 최소화하는 방향으로 고립살색입체를 수정한다. 일반적으로 동적 외곽선(active contour)계열에서 정의하는 에너지와 유사하며, 에너지 최소화 과정을 통해 외곽선의 점은 주위의 외곽선 위의 점들을 부드럽게 이어주는 방향으로 ( $E_{position}^{XT}$ ), 시공간 입체내의 삼차원 에지(edge)의 크기가 큰 점 위에 놓일 수 있도록( $E_{volume}^{XT}$ ) 접근한다. 이는 외곽선의 위치가 잡음 등에 의해 심하게 변화하는 것을 막아주며, 영상 내의 경계선과 일치시키는 의미를 가진다.

$$E^{XT}(x, y, t) = E_{position}^{XT}(x, y, t) + E_{volume}^{XT}(x, y, t)$$

$$E_{position}^{XT}(x, y, t) = \alpha \times (X(y, t+1) - 2X(y, t) + X(y, t-1))^2 + \beta \times (X(y+1, t) - 2X(y, t) + X(y-1, t))^2 \quad (1)$$

$$E_{volume}^{XT}(x, y, t) = \gamma \times |\nabla_{XT} I(x, y, t)|^2 (\gamma < 0)$$

$(X(y_0, t_0) : x - \text{position of neighbor boundary at } y = y_0 \& t = t_0)$

구체적인 수정방법은 다음과 같다. 우선 시공간 입체에서 살색화소를 하나이상 포함하는 XT 단면들을 모두 추출하고, 추출된 각 XT 단면 상의 살색영역의 외곽선 위의 각 점에 대하여 적절한 크기의 X 방향 탐색영역에서  $E^{XT}(x, y, t)$  가 최소인 점을 찾고 그 점을 새로운 경계로 선정하여 수정결과를 갱신한다. 이를 더 이상 수정할 점이 없거나 미리 지정한 최대 반복회수를 넘어설 때까지 반복하여 최종수정영역을 얻는다.

### 2.3 동종살색입체 구성

얼굴, 팔, 다리 등 하나의 살색물체가 만들어 내는 시공간 입체를 동종살색입체라 정의한다. 앞서 구성한 고립살색입체는 하나 혹은 영상 내에서 겹침이 있는 소수의 살색물체가 만들어 내는 것으로 생각할 수 있고,

이를 겹침구간을 중심으로 분할하면 동종살색입체를 얻을 수 있다. 본 논문에서는 고립살색입체에서 다음과정을 거쳐 동종살색입체를 얻는다.

첫째, 그림 5(a) 같은 고립살색입체의 XT 평면 투영 영상을 얻고, 그림 5(b)와 같이 2 개 이상의 물체가 겹쳐있는 구간(겹침구간)을 찾는다.

둘째, 각 겹침구간을 중심으로 그림 5(c)와 같이 겹침구간 전(①,②)과 후(③,④)의 입체로 구분한 후 겹침 전과 후의 입체 하나씩( $V_1, V_2$ )에 대하여 식 (2)를 이용하여 유사도( $S$ )를 측정한다.

$$S(V_1, V_2) = \exp\{-WD^2\} + \frac{\lambda}{ID+1} \exp\{-PD^2\}$$

$$WD(\text{width difference}) = |\text{width}_1 - \text{width}_2|$$

$$ID(\text{inclination difference}) = |\text{Incl}_1 - \text{Incl}_2|$$

$$PD = \text{Position Diff of prediction point.}$$

여기서  $WD$  는 두 입체의 겹침구간 근처에서  $x$  방향 너비의 차이를 나타내며,  $ID$  는 기울기 차이이다. 또한 겹침구간에 인접한 부분에서의 기울기와 위치로 예측한 겹침구간 내 중간시점에서의 예상위치를 구하고 두 입체의 예상위치 차이를  $PD$  로 나타내었다. 이 때, 기울기 차이( $ID$ )는 얼마나 두 물체가 비슷한 속도로 움직이는지를 나타내는 것으로 본 논문에서는 예상위치 차이의 신뢰도로 사용되었다. 제시한 유사도는 두 입체의 크기, 위치, 속도가 얼마나 비슷한가를 나타내는 척도이다.

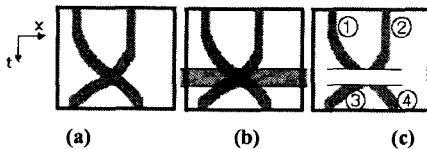


그림 5 동종살색입체 구성

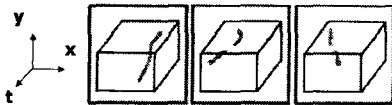


그림 6 동종살색입체

셋째, 임계값 이상의 유사도를 갖는 후보입체 쌍에 대하여, 가장 유사도가 큰 쌍을 우선으로 입체들을 연결한다. (①-④, ②-③) 이를 모든 겹침구간에 대하여 반복한다.

겹침구간이 없는 고립살색입체 및 위 과정을 통하여 분할 후 연결된 고립살색입체의 부분입체들은 하나

의 동종살색입체를 구성하게 된다.

## 2.4. 얼굴여부 판단

동종살색입체를 만들어 내는 살색물체가 얼굴인지 아닌지를 판단한다. 만약 얼굴일 경우, 그 동종살색입체는 하나의 얼굴그룹에 대응되며, 얼굴그룹을 구성하는 얼굴영상들의 시공간 내 위치정보를 제공한다.

본 논문에서는 동종살색입체의 XY 단면을 기존의 얼굴검출기[4]를 이용하여 얼굴과 비얼굴영상으로 구분하고 동종살색입체 내에서 하나 이상의 얼굴이 검출될 경우 그 동종살색입체는 얼굴로부터 구성되었다고 판단하였다.

## III. 대표얼굴선정

대표얼굴이란 한 사람에 대한 얼굴영상 중 그 사람을 알아보는데 가장 적합한 얼굴영상을 의미한다. 본 논문에서는 구성된 각 얼굴그룹에 대하여 정면이고 크기가 큰 얼굴 한 장씩을 선택하여 대표얼굴로 선정하였다.

대표얼굴 선정을 위한 기준으로 영상을 포함하는 최소직사각형의 면적( $Area_{mask}$ )과 정면인 정도( $frontness$ )를 바탕으로 사람이 인지하는 정도를 정량화한 대표성( $keyness$ )이란 정량화된 척도를 식 (3)과 같이 제안한다.

$$keyness = frontness \times \log(Area_{mask} + 1)$$

$frontness$  는 식(4)과 같이 직사각형 내 살색화소수의 비율  $R$  (skin color Ratio)과 좌우 대칭인 정도  $S$  (degree of Symmetry)의 곱으로 나타내었다. 이 때  $R$  과  $S$  두 값 모두 일정 수준 이상인 얼굴을 찾기 위하여 두 값을 곱하였다

$$frontness = R \times S$$

$$\text{where, } R = \frac{Pixels_{Skin}}{Area_{mask}} \quad S = \frac{Edge_{Sym.}}{Edge_{Mask}}$$

살색화소수의 비율  $R$  과 좌우 대칭인 정도  $S$  는 최소직사각형의 면적  $Area_{mask}$ , 직사각형 내 살색화소수  $Pixels_{Skin}$ , 직사각형 내 에지(Edge) 수  $Edge_{Mask}$ , 직사각형 내 좌우 대칭 에지 수의 최대값  $Edge_{Sym.}$  을 이용하여 식 (5)와 같이 계산하였다. 이때, 좌우 대칭 에지 수의 최대값  $Edge_{Sym.}$  는 주어진 직사각형 내의 각각의 세로축에 대하여 서로 대칭인 위치에 존재하는 에지 화소의 총 개수를 계산한 후, 그 개수 중 가장 큰 값으로 정의된다.

이를 이용하여 각 얼굴그룹 내 영상들 중 가장 대

표성 값이 큰 영상을 대표얼굴로 선정하였다.

#### IV. 실험결과

##### 4.1 얼굴그룹생성 결과

본 논문에서 제안한 얼굴그룹화 성능을 검증하기 위하여, 모의 감시영상에 대하여, 얼굴그룹화 결과를 살펴보았다. 실험결과 그림 10에서 제시한 실험영상과 같이 비교적 복잡한 움직임이 있는 경우에도 그림 7과 같이 한 사람의 얼굴 또는 뒷모습 영상들로 각각의 얼굴 그룹을 성공적으로 만들어냄을 확인할 수 있었다.

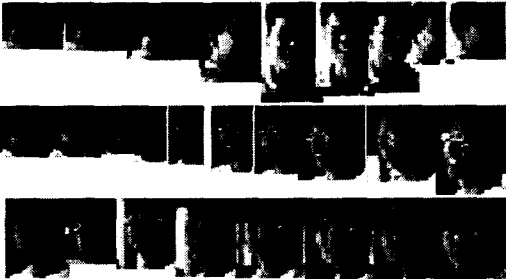
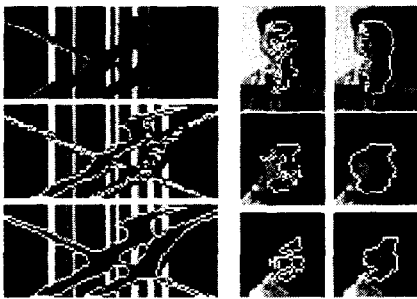


그림 7 얼굴그룹생성결과

##### 4.2 살색입체 수정결과

2.2 절에서 제시한 기법을 이용하여 시공간 입체의 XT 단면 상에서 살색영역을 수정하였다.



(a)XT 단면 (b) 살색영역변화

그림 8 XT 단면수정결과

그림 8(a)는 그림 10에 제시한 영상열로 구성된 XT 단면(上)과 단면 내 초기검출 살색영역(中) 및 수정결과(下)이다. 그림 8(b)는 수정 전(左)과 XT(YT)평면에서의 수정 후(右) 변화된 XY 평면 살색영역이다. 제안한 기법을 이용하여 실제 살색영역과 가까운 살색영역을 얻을 수 있음을 확인할 수 있었다.

##### 4.3 대표얼굴선정결과

그림 7과 같이 구성된 얼굴그룹들에 대하여 대표얼굴을 선정한 결과는 그림 9와 같다. 제안한 식 (3)을 이용하여 비교적 크기가 크고 정면인 얼굴을 찾을 수 있었다.



그림 9 대표얼굴

#### V. 결론

본 논문에서는 동영상 요약의 한 기법으로 시공간 입체를 이용한 등장인물 얼굴요약 시스템을 제안하였다. 각 등장인물의 얼굴영상들을 시공간 내 살색화소의 연속성을 기반으로 하나의 얼굴그룹으로 생성하였으며, 각 얼굴그룹에 대하여 그 크기가 크고 비교적 정면인 얼굴을 찾기 위하여 대표성 개념을 제안하고 이를 이용 대표얼굴을 선정하였다. 또한 살색영역 자동검출결과와 수동지정과의 차이를 XT-YT 단면 상에서의 영역확장을 통하여 보완하였다. 그리고 XT 투영영상에서 살색영역 분할과정을 통하여 동영상 내에서 발생할 수 있는 살색영역 겹침 현상에 의한 얼굴 그룹화 실패문제를 해결하였다.

#### 참고문헌

- [1] Y. Li, A. Goshtasby, and O. Garcia, "Detecting and Tracking Human Faces in Videos," *Proc. 15<sup>th</sup> Int' l Conf. Pattern Recognition*, pp. 807-810, 2000.
- [2] E. Ioutas, C. Nikou, and I. Pitas, "An Information Theoretic Approach to Joint Probabilistic Face Detection and Tracking," *Proc. Int' l Conf. Image Processing*, pp. 505-508, 2002.
- [3] S. McKenna and S. Gong, "Tracking Faces," *Proc. 2<sup>nd</sup> Int' l Conf. Automatic Face and Gesture Recognition*, pp. 271-276, 1996.
- [4] C. Garcia and G. Tziritas, "Face Detection Using Quantized Skin Color Regions Merging and Wavelet Packet Decomposition," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 1, no. 3, pp.264-277, Sep. 1999
- [5] J. Hall, D. Greenhill, and G. A. Jones, "Segmenting Film Sequences using Active Surfaces," *Proc. IEEE Int' l Conf. Image Processing*, pp. 751-754, 1997.

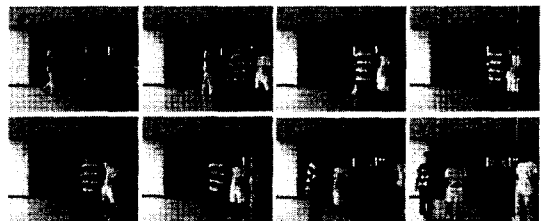


그림 10 실험영상