

논문 2015-52-5-12

얼굴 추적에서의 Staggered Multi-Scale LBP를 사용한 선택적인 점진 학습

(Selective Incremental Learning for Face Tracking Using Staggered Multi-Scale LBP)

이 용 겐*, 최 상 일**

(Yonggeol Lee and Sang-Il Choi[©])

요 약

점진 학습은 비교적 높은 얼굴 추적 성능을 보이지만, 환경적인 변화로 인해 추적에 오차가 발생하면 그 이후의 추적에 오차가 전파되어 추적 성능이 감소한다는 단점이 있다. 본 논문에서는, 다양한 변이 조건에서 강인하게 동작할 수 있는 선택적인 점진 학습 방법을 제안한다. 먼저, 개별 프레임에 대해 LBP(Local Binary Pattern) 특징을 추출하여 사용함으로써 조명 변이에 보다 강인하게 동작 할수 있고, Staggered Multi-Scale LBP를 사용하여 점진 학습에 사용할 패치(patch)를 선택하여 이전 프레임에서의 오차가 전파되는 것을 방지하였다. 실험을 통해, 제안한 방법이 조명 변이와 같은 환경적 변이가 존재하는 비디오 영상에 대해서도 기존의 추적 방법들보다 우수한 얼굴 추적 성능을 보이는 것을 확인할 수 있었다.

Abstract

The incremental learning method performs well in face face tracking. However, it has a drawback in that it is sensitive to the tracking error in the previous frame due to the environmental changes. In this paper, we propose a selective incremental learning method to track a face more reliably under various conditions. The proposed method is robust to illumination variation by using the LBP(Local Binary Pattern) features for each individual frame. We select patches to be used in incremental learning by using Staggered Multi-Scale LBP, which prevents the propagation of tracking errors occurred in the previous frame. The experimental results show that the proposed method improves the face tracking performance on the videos with environmental changes such as illumination variation.

Keywords : Incremental Visual Tracking, Selective Incremental Learning, Local Binary Pattern, Staggered-Multi Scale LBP

I. 서 론

컴퓨터 비전 분야에서, 목표로 하는 객체(target

object)를 추적하기 위해 형상 모델(appearance model)을 이용한 많은 얼굴 추적 기법들이 제안되어 왔다^[1~3]. 그러나 실제 환경에서 촬영된 영상에는 조명, 노이즈 그리고 저해상도와 같은 외부(extrinsic) 변이들과 얼굴의 포즈, 크기, 표정, 부분 가림(occlusion)등의 내부 변이들이 존재하는데, 이러한 변이들은 얼굴의 정확한 추적을 어렵게 만든다.

초기 얼굴 추적 연구들은 목표 물체에 대한 고정적인 모델을 사용하는 방법들을^[4~5] 제안하였지만, 목표 객체

* 학생회원, ** 정회원, 단국대학교 컴퓨터학과
(Department of Computer Science and Engineering,
Dankook University)

© Corresponding Author(E-mail: choisi@dankook.ac.kr)

Received ; February 25, 2015 Revised ; April 1, 2015

Accepted ; April 23, 2015

의 변화에 대응하지 못한다는 단점이 있다. 이에 따라 객체의 변화를 반영하는 모델을 구성하기 위해 점진 학습(incremental learning) 기반의 방법들이 제안되었다^[3, 6]. 특히, IVT(Incremental Visual Tracking)^[3] 방법은 기존의 점진 학습 방법으로부터 수집된 영상들의 평균과 고유벡터를 갱신하여 포즈, 크기 그리고 조명 등의 변이에서 비교적 강인하게 동작하였다. 그러나 IVT 방법을 포함한 점진 학습 기반의 추적기들은 다양한 변이에 의해 추적에 오차(error)가 발생할 경우, 오차가 이후 프레임에 전파되어 점차 추적의 정확도가 감소한다는 단점이 있다.

본 논문에서는, 얼굴 추적에서 선택적인 점진 학습 방법을 통해 조명 변이와 같은 환경의 변화에도 강인하게 동작하는 S-IVT(Selective IVT) 방법을 제안한다. 먼저, 조명 변이로 인한 문제를 효과적으로 해결하기 위해, 제안한 방법에서는 연속적인 프레임들에 대해 LBP(Local Binary Pattern)^[7-8] 전처리를 수행하였다. 또한, 영상 전체 영역이 아닌 얼굴의 주변 영역에 대해서만 LBP를 수행함으로써 실시간 처리가 가능하도록 하였다. 두 번째로, 점진 학습에서 다음 프레임의 학습을 위해 추가될 패치를 선택하는 과정에서, 보행자 검출을 위해 제안되었던 SMSLBP(Staggered Multi-Scale LBP)^[9]을 적용하였다. 기존의 IVT 추적기에서는 단순히 이전 프레임에서 구한 고유벡터로부터 계산된 확률 값이 가장 높게 예측된 패치를 다음 프레임의 학습에 포함시켰기 때문에 변이로 인하여 패치가 잘못 예측되었을 경우 이후의 프레임에 오차가 전파(propagation)된다. 그러나 제안한 방법에서는 고유벡터를 이용하여 다수의 후보 패치를 예측하고 SMSLBP를 이용하여 가장 적합한 패치를 선별함으로써 조명 등의 변이가 있는 경우에 대한 추적 성능을 향상 시켰다.

실험 결과, 다양한 변이를 포함한 비디오 데이터베이스에서 얼굴 추적을 수행하였을 때, 제안한 방법이 전반적으로 우수한 얼굴 추적 성능을 나타내는 것을 확인할 수 있었다.

본 논문은 다음과 같이 구성되어 있다. II에서는 점진 학습 기반의 IVT 방법에 대해 살펴보고, III에서는 본 논문에서 제안한 선택적인 점진 학습 기반의 얼굴 추적 방법을 설명한다. IV에서는 다양한 비디오 데이터베이스에 대한 기존의 방법들과 얼굴 추적 성능을 비교하고, V에서는 전체적인 내용을 정리한다.

II. 관련 연구

일반적으로 객체 추적은 markov model로부터 순차적인 입력 영상 패치들의 집합 $\Gamma_t = \{I_1, \dots, I_t\}$ 에 대해 시간 t 에서의 잠재 매개변수 X_t 로부터 객체를 추론한다. 또한 베이즈(Bayes) 이론에 따라 $p(X_t | \Gamma_t)$ 는 다음과 같이 정의된다.

$$p(X_t | \Gamma_t) \propto p(I_t | X_t) \int p(X_t | X_{t-1}) p(X_{t-1} | \Gamma_{t-1}) dX_{t-1} \quad (1)$$

식 (1)에서, $p(X_t | \Gamma_t)$ 는 패치 I_t 를 관측하기 위한 X_t 의 우도비(likelihood)를 측정하는 $p(I_t | X_t)$ (관측 모델)과 이전과 현재 상태 사이의 $p(X_t | X_{t-1})$ (동적 모델)로 구성된다.

1. 점진 학습

목표 객체를 추적하기 위해서는 추적하고자 하는 객체의 움직임을 반영한 형상 모델이 필요하다. 그러나 이미 수집한 영상과 새롭게 수집되는 영상들에 대해 매번 형상 모델을 계산하는 것은 실시간 추적방법으로 적합하지 않다. 이를 위해, 제안된 방법이 SKL(Sequential Karhunen-Loeve)^[10]를 이용한 점진 학습 방법이며, SKL방법은 특이값 분해(SVD)를 사용하여 효율적으로 계산할 수 있다는 장점이 있다. 그러나 기존의 SKL방법은 수집된 모든 영상들의 평균을 반영하지 못한다는 단점이 있어, IVT 방법^[3]에서는 이를 고려한 새로운 SKL 알고리즘을 제안하였다.

먼저, 점진 학습을 위한 d 차원의 데이터는 이전 프레임들로부터 수집한 $t-1$ 개 영상 패치들의 집합 $A = \{I_1, \dots, I_{t-1}\}$, 현재 이후 프레임에서 새롭게 관측 m 개 영상들의 집합 $B = \{I_t, \dots, I_{t+m-1}\}$ 그리고 $\Gamma = [A \ B]$ 로 정의한다. 이에 대한 각각의 평균을 각각 \tilde{I}_A , \tilde{I}_B 그리고 \tilde{I}_Γ 라고 할 때, 공분산 행렬 C_A , C_B 그리고 C_Γ 는 다음과 같다.

$$\begin{aligned} C_A &= \frac{1}{n} (A - \tilde{I}_A)(A - \tilde{I}_A)^T, \\ C_B &= \frac{1}{m} (B - \tilde{I}_B)(B - \tilde{I}_B)^T, \\ C_\Gamma &= C_A + C_B + \frac{nm}{n+m} (\tilde{I}_B - \tilde{I}_A)(\tilde{I}_B - \tilde{I}_A)^T \end{aligned} \quad (2)$$

식 (2)에서, C_T 의 특이값 분해에는 $A - \tilde{I}_A$ 와 $B - \tilde{I}_B$ 에 대한 벡터들의 수평적 결합(concatenation) 이외에도 새로운 벡터항 $\sqrt{\frac{nm}{n+m}}(\tilde{I}_B - \tilde{I}_A)$ 이 추가된다. 특히, \tilde{I}_T 를 계산할 때 기존의 수집된 데이터($t-1$)가 커질수록 새롭게 수집된 데이터가 반영되는 부분(m)이 작아지므로, 망각인자(forgetting factor) f 를 두어 효율적으로 평균을 업데이트 할 수 있다. 새롭게 정의된 \tilde{I}_T 은 다음과 같다.

$$\tilde{I}_T = \frac{fn}{fn+m}\tilde{I}_A + \frac{m}{fn+m}\tilde{I}_B \quad (3)$$

2. 동적 모델

얼굴 추적에서, 현재 프레임 내에서의 영상 패치(I_t)는 다음 프레임에서의 얼굴 패치 후보를 추출하기 위한 사전 정보(prior information)로서, 각각의 영상 패치는 6개의 아핀운동 매개변수(affine motion parameters) $\mathbf{X}_t = (x_t, y_t, \theta_t, s_t, \alpha_t, \phi_t)$ 로부터 생성할 수 있다. 각각의 변수들은 시간 t 에서의 중심 좌표(x, y), 회전 각도(θ), 패치의 크기(s), 가로와 세로의 비율(α), 침도(ϕ)이다. 또한 \mathbf{X}_t 는 미리 정의된 k 개의 파티클(particle)의 집합이며, 아핀 변환(affine transform)으로부터 k 개의 I_t 를 생성하는 데에 사용된다. 특히 각각의 매개변수가 가우시안 분포의 특성을 따른다고 가정하면 \mathbf{X}_{t-1} 와 독립적으로 모델링 될 수 있으며, $\Psi (= \sigma_x^2, \sigma_y^2, \sigma_\theta^2, \sigma_s^2, \sigma_\alpha^2, \sigma_\phi^2)$ 를 매개변수들의 분포에 대한 표준편차 값으로 정의하면, $p(\mathbf{X}_t | \mathbf{X}_{t-1})$ 는 다음과 같이 정의할 수 있다.

$$p(\mathbf{X}_t | \mathbf{X}_{t-1}) = N(\mathbf{X}_t; \mathbf{X}_{t-1}, \Psi) \quad (4)$$

한편, 동적 모델에서 파티클 개수가 많고 표준편차가 큰 경우, 객체에 대한 높은 예측을 할 수 있지만 연산량이 많아지므로 이들 사이에는 trade-off가 발생한다.

3. 관측 모델

동적 모델로부터 예측한 \mathbf{X}_t 로부터 생성된 패치들이 점진 학습에서 계산된 고유벡터 U 와 평균 μ 로부터 확장될 수 있다고 가정하면, 각각의 패치에 대한 확률 값을 계산할 수 있다. 이전 프레임에서 구한 고유벡터들로

구성된 부분 공간으로부터 발생된 패치에 대한 확률(p_d)은 해당 패치와 부분 공간의 중심까지의 거리와 반비례한다고 볼 수 있다. 해당 패치와 부분 공간의 중심까지의 거리(d)는 부분 공간에 투영된 지점과의 거리(d_t)와 투영된 지점과 부분 공간의 중심까지의 거리(d_w)로 표현 될 수 있다. 부분 공간으로부터 발생된 패치에 대한 확률 $p_{d_t}(I_t | \mathbf{X}_t)$ 과 부분 공간에 투영된 샘플의 우도비(likelihood) $p_{d_w}(I_t | \mathbf{X}_t)$ 를 다음과 같이 정의하면,

$$\begin{aligned} p_{d_t}(I_t | \mathbf{X}_t) &= N(I_t; \mu, UU^T + \epsilon I), \\ p_{d_w}(I_t | \mathbf{X}_t) &= N(I_t; \mu, U\Sigma^{-2}U^T) \end{aligned} \quad (5)$$

패치에 대한 우도비는 다음과 같이 표현될 수 있다.

$$\begin{aligned} p(I_t | \mathbf{X}_t) &= p_{d_t}(I_t | \mathbf{X}_t)p_{d_w}(I_t | \mathbf{X}_t) \\ &= N(I_t; \mu, UU^T + \epsilon I)N(I_t; \mu, U\Sigma^{-2}U^T) \end{aligned} \quad (6)$$

식 (5)와 (6)에서, I 는 단위행렬이며 ϵI 는 additive gaussian noise이다.

III. 제안 방법

본 논문에서는, 기존의 점진 학습 기반의 얼굴 추적 방법들의 단점을 보완하고, 얼굴 추적 성능을 향상시키기 위한 선택적인 점진 학습 방법을 제안한다. LBP를 이용하여 영상을 전처리하고 SMSLBP를 이용하여 선택적인 점진 학습을 수행함으로써, 조명 변이에 강인한 얼굴 추적기를 구성한다(그림 1).

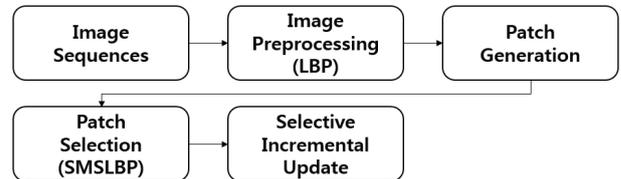


그림 1. 선택적인 점진 학습을 이용한 다이어그램
Fig. 1. Diagram using selective incremental learning.

1. LBP를 이용한 영상의 전처리

입력 영상에 존재하는 조명 변이는 얼굴 추적 성능을 떨어뜨리는 주된 요인 중 하나이다. 특히, IVT 방법은 작은 조명의 변화에 대해서는 비교적 좋은 추적 성능을 보이지만, 입력 영상내의 조명 변이가 커지면 얼굴을

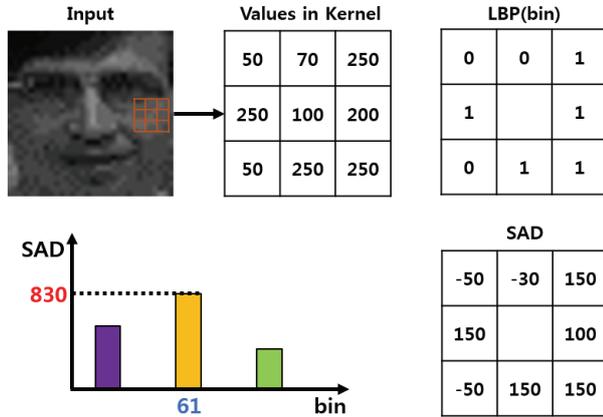


그림 2. 가중치 LBP 히스토그램 생성
Fig. 2. Generation of weighted LBP histogram.

놓치는 경우가 자주 발생한다. [6]에서는 조명 변이의 문제를 해결하기 위해 MCT(Modified Census Transform)^[11~12]를 이용한 MCT+IVT 방법이 제안되었다. MCT는 3×3 크기의 구조화된 커널에서 주변의 8개 픽셀과 중심값을 포함한 9개의 비트를 수평 결합하여 해당 픽셀의 값을 결정하므로 조명 변이에 대해 강인하게 동작한다. 그러나 입력 영상의 해상도가 커질 경우, 연산 비용과 얼굴 추적 성능간의 trade-off가 발생한다.

본 논문에서는, MCT와 비슷한 얼굴 추적 성능을 보이면서 MCT에 비해 연산 비용이 적은 LBP를 사용하여 연속된 프레임들에서 전처리 하였다(그림 2). 영상 내의 지점 (x, y) 의 밝기값을 $I_{(x, y)}$, $I_{(x, y)}$ 를 제외한 8개의 이웃 픽셀들의 밝기값들을 $I_{(x', y')}$ ($x' = x \pm 1, y' = y \pm 1$)로 정의하면 이진 값의 비교 함수(ζ)를 사용하여 다음과 같이 정의할 수 있다.

$$\zeta(I_x, I_y) = \begin{cases} 1 & I_y \geq I_x \\ 0 & otherwise \end{cases} \quad (7)$$

\otimes 를 수평 결합 연산자로 정의하면, 구조화된 kernel $\kappa(x, y)$ 는 다음과 같다.

$$\kappa(x, y) = B2D(\otimes \zeta(I_{(x, y)}, I_{(x', y')})) \quad (8)$$

$\kappa(x, y)$ 는 지점 (x, y) 에 대해 이웃 픽셀의 값을 모두 반영한 LBP 특징 값을 뜻하며 8개의 비트를 이진-십진수 변환 함수 $B2D(\cdot)$ 를 통해 십진수로 표현된다.

한편, LBP를 사용함으로써 MCT에 비해 1비트만큼의 연산 비용을 감소시켰지만, 여전히 영상 전체를 전

처리하는 것은 실시간 얼굴 추적에 적합하지 않다. 이를 해결하기 위해 본 논문에서는 이전 프레임에서 최대 확률값을 가지는 패치의 위치로부터 extend factor ($ef = 2$)를 적용한 영역에 대해서만 전처리를 수행함으로써, 연산 속도를 개선하였다.

2. SMSLBP 분류기 설계

IVT 기반의 추적 방법들은^[3, 6] 점진 학습을 통해 추적하고자 하는 객체의 변화를 추적 과정에 잘 반영하므로, 비교적 높은 성능의 실시간 얼굴 추적을 수행할 수 있다. 그러나 영상 내의 다양한 변이로 인해 한 번 추적에 실패하면 그 결과가 전파되어(propagated) 객체를 놓쳐 버리게 된다. 이는 최대 확률 값을 갖는 패치가 정확하게 분류 되었는지를 판별하는 과정이 없어, 임의의 변이로 인해 추적에 실패한 부정확한 패치들이 점진 업데이트(incremental update)에 사용되기 때문이다. 본 논문에서는 이러한 문제를 해결하기 위해 SMSLBP^[9] 특징을 이용하여 얼굴인지의 여부(face/non-face)를 판별하는 분류기를 설계하였다.

SMSLBP는 보행자 검출을 위해 제안된 방법으로서 실제 환경에서 강인하게 동작하는 texture 기반의 특징을 추출한다. 그림 2에서, SMSLBP는 먼저 weighted LBP histogram을 생성하는데, 크기 3×3 커널에 대해 기존의 LBP와 같이 수평 결합 연산을 통하여 비트 스트링(bit string)을 생성하고 식 (8)을 이용하여 0 ~ 255의 십진수를 생성하고, 그에 대한 히스토그램의 각 bin에 SAD(Sum of Absolute Difference) 값을 누적한다.

그런 다음, 다양한 gray information을 얻기 위해 기준 픽셀을 중심으로 지름 R_n 에 해당하는 이웃 픽셀로부터 특징을 생성하며, 이웃 픽셀의 좌표는 다음을 따른다.

$$(R_n \cos \theta, R_n \sin \theta), \theta = \left(\frac{2\pi}{P} * (p + \frac{n-1}{N}) \right) \quad (9)$$

여기서, 총 8개의 이웃 픽셀을 참조할 때, $P(= 8)$ 와 $p(= \{1, \dots, 8\})$ 는 각각 참조하려는 이웃 픽셀의 개수와 이웃 픽셀의 순서를 나타낸다. $N(= 3)$ 과 $n(= \{1, 2, 3\})$ 은 $R_1 = 1, R_2 = 3, R_3 = 5$ 에서의 전체 스케일 수와 스케일 인덱스를 나타낸다. 식 (9)로부터 3가지 스케일에 대해 각각 8개의 이웃 픽셀들을 참조

할 수 있으며, 특히 각각의 스케일로부터 참조되는 좌표는 모두 방향이 다르므로 총 24개의 어긋난 방향의 이웃 픽셀들로부터 특징을 추출한다.

세 번째로, HOG^[13]와 같이 50%의 overlapping이 존재하는 16×16 크기의 block을 사용하여 고차원의 SMSLBP 특징 값을 생성하고, 추출된 특징들에 대해 SVM(Support Vector Machine)^[14]을 사용하여 학습한다.

얼굴/비(非)얼굴(face/non-face) 분류기를 설계하기 위해 FERET^[15], AR^[16], CMU-PIE^[17], BioID^[18] 그리고 Final 데이터베이스^[19]를 사용하였다. 사용된 얼굴 데이터베이스들은 다양한 변이를 포함하고 있으므로 조명, 표정, 부분 가림 그리고 포즈에 대응할 수 있다. 데이터베이스의 정렬된 얼굴 패치들은 32x32크기로 정렬되어 positive 샘플로 사용되었고, negative 샘플들은 INRIA 데이터베이스^[13]로부터 배경 정보에 해당하는 패치를 무작위로(randomly) 채취하였다. SVM 학습을 위해 사용된 positive와 negative 샘플들의 개수는 각각 25,474개와 25,578개이며(그림 3), 단일 패치로부터 6,912차원의 특징 벡터를 추출하였다.

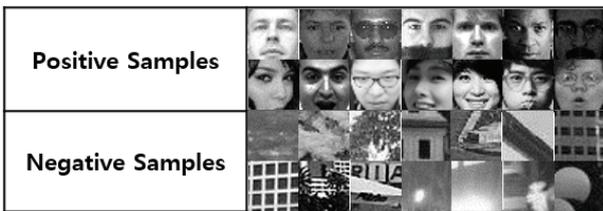


그림 3. 분류기 학습을 위해 사용된 샘플
Fig. 3. Samples for learning of classifier.

3. 선택적 점진 업데이트

매 프레임마다 점진 학습을 통해 객체의 특성을 반영한 고유벡터를 추출하는 것은 연산 양의 문제가 발생하여 실시간 추적을 위해서는 적합하지 않다. 따라서 IVT 방법에서는 일정 기간 동안에 매번 확률적으로 가장 객체에 근접한 패치를 수집하는 batch를 두고 batch 자체를 점진 학습의 데이터로 사용함으로써 연산 비용을 해결하였다. 그러나 영상 프레임에 추적을 방해하는 변이가 발생한 경우에는 수집된 batch로부터 점진 학습을 수행하더라도 이로부터 구성된 형상 모델이 객체의 특성을 제대로 반영하지 못하여 결국 객체 추적 성능이 떨어지게 된다. 따라서 본 논문에서는 정확하게 추적된 패치를 선택하는 점진 학습 방법을 제안한다. 선택적인

점진 학습을 위해 다음과 같은 가정을 한다.

- 1) 외부의 변이가 존재하는 환경으로부터 얼굴 추적을 수행할 경우, 형상 모델로부터 확률 값이 가장 높은 패치가 가장 정확한 추적 결과를 보장하지 못한다. 이 경우 확률 값이 가장 높은 패치들로 점진 업데이트를 수행할 경우, 형상 모델의 성능이 떨어지게 된다.
- 2) 내부 변이가 존재할 경우, 여러 개의 후보 패치들 중 상대적으로 확률 값이 낮은 패치에서도 얼굴에 근접한 패치가 존재할 수 있다.
- 3) 연속적인 프레임에서 객체의 움직임과 함께 변이들도 프레임에서 연속적으로 나타난다. 따라서 프레임 간 변이들도 상관관계(correlation)를 가지므로, 변이들로 인해 정확한 추적이 수행되지 않을 경우에는 이를 정확한 객체 추적에 활용할 수 있다.

이러한 가정들로부터, 제안한 방법에서는 점진 학습에 사용할 패치를 선택하기 위해, 후보 패치들에 대해 SMSLBP 특징을 추출하고 학습된 SVM를 통해 정확한 데이터를 선택한다.

그림 4에서, 추적 결과가 될 수 있는 모든 패치들을 확률 값을 기준으로 정렬한 후, cascade 구조^[20-21]를 따라 확률 값이 가장 높은 패치부터 얼굴인지 여부를 판단하고, 얼굴이라고 판별될 경우의 해당 패치를 점진 학습을 위한 batch 데이터로 저장한다. 만약, 특정 프레임에서 모든 후보 패치가 얼굴이 아닐 경우에는 연속된 프레임 간 후보 패치들의 상관관계를 고려하여 기존에 저장되어 있던 batch를 초기화 시킨다. 이를 통해, 객체의 움직임을 반영하되, 변이로 인해 부정확한 패치들의 오차가 다음 프레임으로 전파되는 것을 방지하는 형상 모델을 구성할 수 있다.

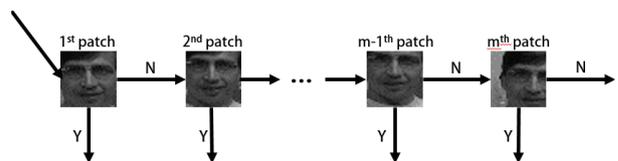


그림 4. Cascade 구조의 얼굴 패치 선택
Fig. 4. Selection of face patch in cascade.

IV. 실험

제안한 방법(S-IVT)이 다양한 변이 조건에서도 얼굴

을 잘 추적하는지를 평가하기 위해, Dudek과 Trellis70 비디오 영상^[3]에 대해 IVT 방법과 MCT+IVT방법을 비교해 보았다.

Dudek은 480×720 해상도의 총 573개 프레임들로 구성되어 있으며, 다양한 표정(#224, #233, #398), 포즈(#39, #183, #573), 프레임 아웃(#487) 그리고 부분 가림(#104)등의 변이들이 포함되어 있다. Trellis70은 240×320 해상도의 총 501개 프레임들로 구성되어 있으며, 캐스트 그림자(cast shadow)를 포함한 다양한 조명 변이들이 프레임 전반에 다수 존재하며, 포즈(#, #308, #343, #361) 그리고 프레임 아웃(#206)등의 변이들이 포함되어 있다. 추적 성능의 평가를 위해 Dudek에 대한 실험에서는 제공된 두 눈, 코 그리고 입에 대한 7개의 실측좌표(ground-truth)를 사용하였으며 실측좌표를 제공하지 않는 Trellis70에 대해서는 자체적으로 실측좌표 값을 설정하여 사용하였다. 추적 성능의 정확도는 실측좌표와 추정된 좌표들 사이의 RMS(Root Mean Square) 오차로 평가하였다. 각 비디오에 대한 실험을 위해 k , f , batch 크기는 각각 600개, 5, 0.95를 사용하였으며, $\Psi = [9, 9, 0.05, 0.05, 0.005, 0.001]$ 의 표준편차를 지정하였다.

그림 5는 Dudek와 Trellis70 비디오에 대한 평균 RMS 결과를 보여준다. Dudek 비디오에 대한 결과(그림 5 (a))에서, MCT+IVT는 얼굴에 부분 가림이 발생한 프레임(#105)이후에는 추적 성능이 현저히 떨어지는 것을 볼 수 있다(그림 6의 첫 번째 행). S-IVT는 프레임 아웃된 프레임(#487) 이후로 추적 성능이 감소하였으나(그림 6의 두 번째 행), 그 이전 프레임까지는 평균

RMS 값이 5.7174로 IVT(7.0885) 보다 우수한 추적 성능을 나타냈다.

Trellis70에 대한 결과(그림 5(b))에서, 조명 변이에

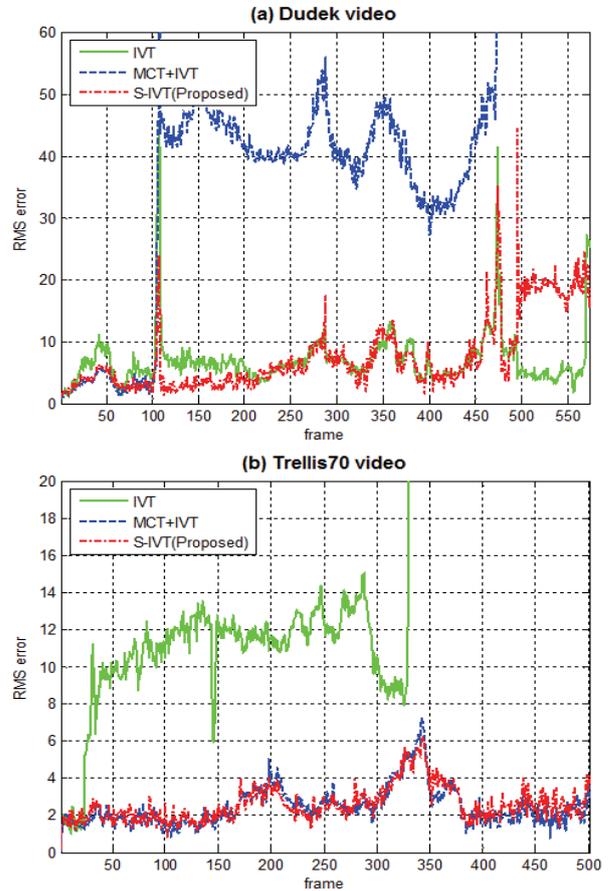


그림 5. 실측 좌표와 추정된 좌표들 사이의 RMS
 Fig. 5. RMS between ground-truth and estimated coordinates.

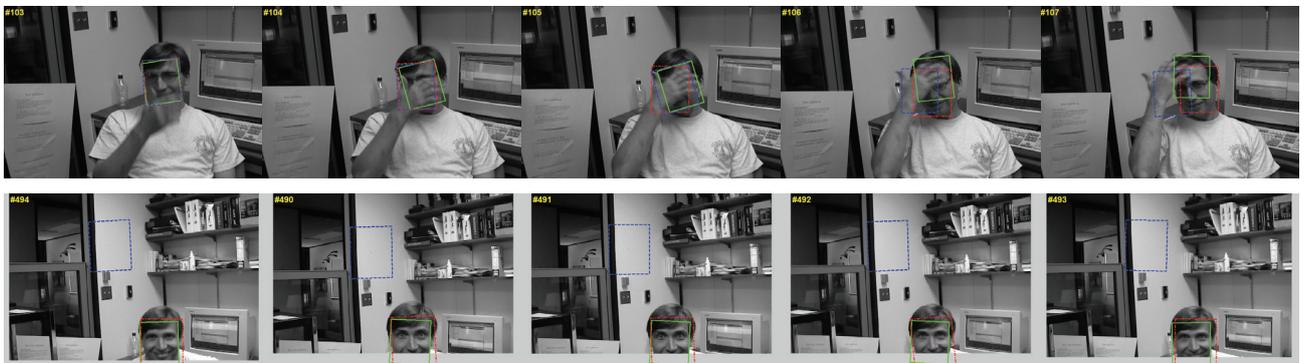


그림 6. Dudek 영상에서의 개별 얼굴 추적 알고리즘 성능 비교
 (빨간색 윈도우(-) : S-IVT, 녹색 윈도우(-) : IVT, 파란색 윈도우(-) : MCT-IVT)
 Fig. 6. Comparison of face tracking algorithm performance on Dudek video (red window(-) : S-IVT, green window(-) : IVT, blue window(-) : MCT-IVT).

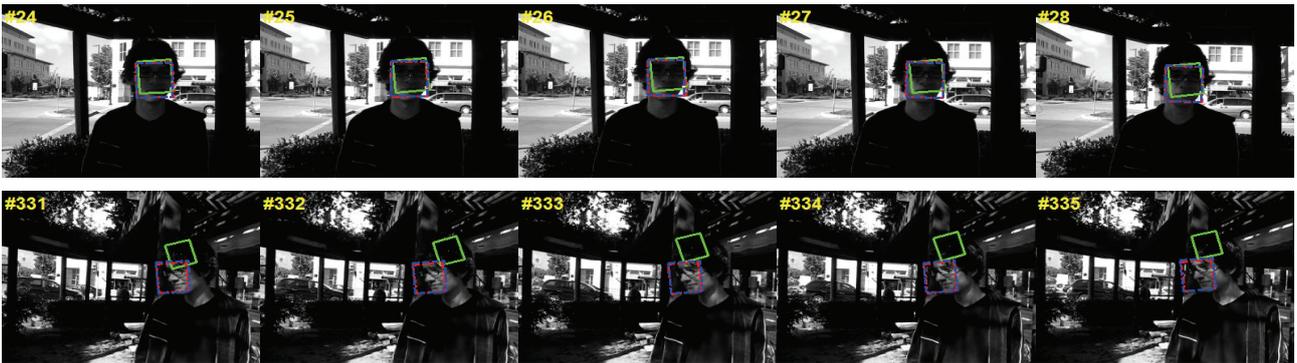


그림 7. Trellis70 영상에서의 개별 얼굴 추적 알고리즘 성능 비교

(빨간색 윈도우(-) : S-IVT, 녹색 윈도우(-) : IVT, 파란색 윈도우(-) : MCT+IVT)

Fig. 7. Comparison of face tracking algorithm performance on Trellis70 video (red window(-) : S-IVT, green window(-) : IVT, blue window(-) : MCT+IVT).

강인한 MCT와 LBP를 사용한 S-IVT와 MCT+IVT 모두 좋은 추적 성능을 보였다. 반면, IVT는 얼굴에 캐스트 그림자가 발생한 프레임(#24)에서부터 추적 성능이 급격히 감소하였으며(그림 7의 첫 번째 행), 캐스트 그림자와 포즈 변이가 동시에 발생한 프레임(#332)부터는 얼굴 추적에 실패하였다(그림 7의 두 번째 행).

실험 결과로부터, S-IVT, IVT 그리고 MCT+IVT는 공통적으로 표정 변이에는 비교적 우수한 성능을 나타내었다. 그러나 IVT와 MCT+IVT는 각각 캐스트 그림자를 포함한 다양한 조명 변이와 부분 가림 변이들로부터 추적 성능이 감소하거나 추적에 실패한 반면, S-IVT는 다양한 변이에 대해 전반적으로 강인하게 얼굴을 추적하는 것을 확인할 수 있다.

V. 결 론

얼굴 추적을 위해 얼굴의 움직임을 반영한 형상 모델을 구성하기 위해 다양한 방법들이 제안되었으나 실제 환경에서 존재하는 다양한 변이들은 보다 강인한 얼굴 추적기를 구현하기 위해 해결해야 할 문제이다.

기존에 제안된 추적 방법들 중, IVT 방법은 평균을 고려한 점진 학습 방법을 사용하여 얼굴의 움직임 뿐만 아니라 표정, 조명 그리고 포즈 등의 변화에도 비교적 잘 동작한다. 그러나 IVT 방법은 캐스트 그림자(cast shadow)를 포함한 다양한 조명 변이에 취약하며, 조명 변이 문제를 해결하기 위해 제안된 MCT+IVT 방법은 부분 가림 및 포즈 변이에 취약점을 보였다. 또한, 기존의 점진 학습 기반의 방법들은 어떤 프레임에

서 추적에 실패했을 경우, 그 추적 결과가 이 후의 프레임에 영향을 주어 점차 추적의 정확성이 감소하는 단점이 있었다.

본 논문에서는, 기존에 제안된 점진 학습 기반의 IVT 방법과 MCT+IVT 방법의 장점을 유지하면서 각각의 단점을 보완하여, 다양한 변이에도 잘 동작할 수 있는 얼굴 추적 방법을 제안하였다.

이를 위해 첫 째, 영상 전처리에 LBP를 사용함으로써 IVT 방법의 단점인 조명 변이에 대응하였고, 특히 extend factor(ef)를 사용한 부분 영역에 대해 전처리를 수행하여 연산 비용을 감소시켰다. 둘째, SMSLBP를 이용하여 패치들의 특징을 추출하고, cascade 구조를 통해 연산 시간을 줄일 수 있었다.

실험 결과, 실제 환경에 존재하는 다양한 변이를 포함하고 있는 Dudek과 Trellis70 비디오 실험에서, 제안한 S-IVT 방법이 기존의 IVT 방법과 MCT+IVT 방법에 비해 정확하게 얼굴을 추적하였다.

REFERENCES

- [1] F. Cootes, J. Edwards, and J. Taylor, "Active appearance models." *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 23, no. 6, pp. 681-685, 2001.
- [2] O. Balan and J. Black, "An adaptive appearance model approach for model-based articulated object tracking." *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Vol. 1, pp. 758-765, 2006.
- [3] A. Ross, et al., "Incremental learning for robust

- visual tracking.” *International Journal of Computer Vision*, Vol. 77, no. 1-3, pp. 125-141, 2008.
- [4] J. Black and D. Jepson, “Eigentracking: Robust matching and tracking of articulated objects using a view-based representation.” *International Journal of Computer Vision*, Vol. 26, no. 1, pp. 63-84, 1998.
- [5] M. La Cascia, S. Sclaroff, and V. Athitsos, “Fast, reliable head tracking under varying illumination: An approach based on registration of texture-mapped 3D models.” *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 22, no. 4, pp. 322-336, 2000.
- [6] Y. Lee and S. Choi, “Face tracking based on incremental learning under illumination variation.” *2013 Korea Signal Processing Conference*, pp. 53-56, Daejeon, Korea, March 2013.
- [7] T. Ojala, M. Pietikäinen and D. Harwood, “A comparative study of texture measures with classification based on featured distributions.” *Pattern Recognition*, Vol. 29, no. 1, pp. 51-59, 1996.
- [8] D. Kim, M. Sohn and S. Lee, “A Study on Face Recognition Method based on Binary Pattern Image under Varying Lighting Condition”, *Journal of the Institute of Electronics and Information Engineers*, Vol. 49, no. 2, pp. 61-74, 2012.
- [9] Y. Cao, et al., “Staggered multi-scale lbp for pedestrian detection.” *2012 19th IEEE International Conference on Image Processing (ICIP)*, pp. 449-452, Florida, USA, September 2012.
- [10] A. Levey and M. Lindenbaum, “Sequential Karhunen-Loeve basis extraction and its application to images.” *IEEE Trans. Image Processing*, Vol. 9, no. 8, pp. 1371-1374, 2000.
- [11] B. Froba and A. Ernst, “Face detection with the modified census transform.” *Proc. Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 91-96, 2004.
- [12] G. Cho, et al., “Real-Time Face Recognition System Based on Illumination-insensitive MCT and Frame Consistency.”, *Journal of the Institute of Electronics and Information Engineers*, Vol. 45, no. 3, pp. 123-134, 2008.
- [13] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection.” *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 886-893, San Diego, USA, June 2005.
- [14] C. Cortes and V. Vapnik, “Support-vector networks.” *Machine Learning*, Vol. 20, no. 3, pp. 273-297, 1995.
- [15] National Institute of Standards and Technology, the FERET database.
<http://www.itl.nist.gov/iad/humanid/feret/>
- [16] A.M. Martinez and R. Benavente, The AR Face Database. *CVC Technical Report #24*, June 1998.
- [17] T. Sim, S. Baker and M. Bsat, “The CMU pose, illumination, and expression database.” *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 25, no. 12, pp. 1615-1618, 2003.
- [18] O. Jesorsky, J. Kirchberg and W. Frischholz, “Robust face detection using the hausdorff distance.” *Audio-and video-based biometric person authentication*. Springer, pp. 90-95, 2001.
- [19] C. H. Kim, et al., “A new biased discriminant analysis using composite vectors for eye detection.” *IEEE Trans. Systems, Man, and Cybernetics, Part B: Cybernetics*, Vol. 42, no. 4, pp. 1095-1106, 2012.
- [20] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features.” *Proc. of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1-511, 2001.
- [21] S. H. Park, D. Cho and K. R. Park, “A Study on Eye Detection by Using Adaboost for Iris Recognition in Mobile Environments.”, *Journal of the Institute of Electronics and Information Engineers*, Vol. 45, no. 4, pp. 1-11, 2008.

— 저 자 소 개 —



이 용 겐(학생회원)
2012년 단국대학교 응용컴퓨터
공학과 공학사
2014년 단국대학교 컴퓨터학과
공학석사
2015년~현재 단국대학교 컴퓨터
학과 박사 과정

<주관심분야 : 패턴인식, 컴퓨터비전>



최 상 일(정회원)
2005년 서강대학교 전자공학과
공학사
2010년 서울대학교 전기·컴퓨터
공학부 공학박사
2010년~2011년 Post Doctoral
Fellow, BK21 Information
Technology, Seoul National
University, Seoul, Korea
2011년 Post Doctoral Fellow, Computer Science
Department, University of Southern
California, CA, USA
2011년~현재 단국대학교 대학원 컴퓨터학과
조교수

<주관심분야 : 패턴인식, 컴퓨터비전>