SumShift 실시간 물체추적기의 성능 개선

최성록O, 이재영, 정재찬, 김지완, 조재일
한국전자통신연구원(ETRI) 지능형인지기술연구부
sunglok@etri.re.kr / http://sites.google.com/site/sunglok

요 약

영상기반 물체추적기술은 지능형 영상기반 감시시스템(IVS)이나 자율주행 자동차 등에 널리 사용되는 요소기술이다. 본 논문은 실시간 영상기반 물체추적기술인 SumShift를 개선하는 세 가지 방법과 각각의 방법을 적용하였을 때의 성능 개선 결과를 소개한다. 개선된 SumShift 물체추적기는 VOT Challenge 2015에 참가하였는데, 대회 결과와 함께 향후 성능 향상을 위한 시사점을 제시한다.

1. 서론

영상기반 물체추적기술은 지능형 영상기반 감시시스템(intelligent visual surveillance; IVS)의 사람이나 목표물 추적, 자율주행 자동차의 인접 자동차와 보행자 추적, 드론의 목표지점 및 착륙지점 추적 등 많은 컴퓨터비전의 응용분야에 폭 넓게 사용되는 요소기술이다. 이러한 필요성과 중요성으로 지금까지 많은 영상기반 물체추적기술들이 연구되었고, 성능평가를 위한 데이터셋과 성능지표에 대한 연구[1,2]도 많이 진행되었다.

본 논문에서 컬러 히스토그램을 이용한 실시간 영상기반 물체추적기술인 SumShift[3]를 개선하는 세 가지 방법과 이들을 적용하였을 때의 실험 결과를 보인다. 성능 평가를 위한 실험에는 추적 물체의 크기 변화, 가려짐 등 다양한 특성들이 반영된 60개의 영상으로 구성된 VOT Challenge 2015 데이터셋을 이용하였다. 또 성능지표는 VOT Challenge에 사용되는 정확도(accuracy), 실패 횟수(robustness)를 이용하였다. 또한 별로 실시간성을 살펴보기 위해 알고리즘이 물체추적에 걸리는 시간을 측정하였다. 개선된 SumShift 추적기는 VOT Challenge 2015에 참가하였고, 대회 결과[4]와 함께 향후 추적기의 성능 향상을 위한 시사점을 제시한다.

2. SumShift 추적기의 세 가지 성능 개선

2.1 SumShift 추적기

SumShift 추적기[3]는 RGB 히스토그램을 물체모델로 사용하는 실시간 영상기반 물체추적기이다. RGB 히스토그램과 히스토그램 역투영(backprojection)을 통해 유사도를 평가한다는 점에서 기존의 MeanShift 알고리즘과 유사하다. 그러나 SumShift와 MeanShift는 크게 두 가지 큰 차이점이 있다. 우선 그림 1과 같이 기존의 MeanShift는 유사도(밀도)의 최고값을 추적하는데, SumShift는 물체의 범위(region)을 고려하여 유사도(밀도)의 합이 최대가 되는 방향으로 추적한다. 이러한 방법을 통해 유사한 배경에 의한 물체의 놓침을 크게 줄일 수 있다. 또한 SumShift는 물체를 하나의 덩어리가 아닌 여러 개의 파티션으로 표현한다. 예를 들어, 물체를 2x1 파티션으로 표현하는 경우, 윗부분의 물체/배경 히스토그램, 아랫부분의 물체/배경 히스토그램, 총 4개의 히스토그램을 이용해 물체를 표현한다. 물체를 하나의 덩어리가 아닌 파티션으로 표현한다는 점은 단순히 RGB 색상의 비율뿐만 아니라 그 분포도 고려했다는 점에서 물체의 형태(appearance)를 어느 정도 반영하는 것으로 볼 수 있다. 실제 사람의 경우 상의와 하의가 크게 다른 경우가 많은데, 파티션을 이용한 물체의 표현은 사람의 위치를 보다 정확하게 찾을 수 있도록 한다.

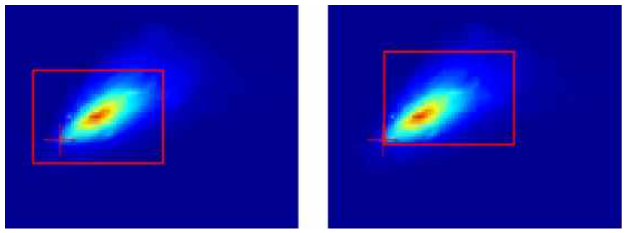

그림 1. MeanShift(좌)와 SumShift(우)의 비교

2.2 세 가지 성능 개선

개선 #1: 물체/배경 히스토그램 업데이트 방법

물체는 물체 영역의 RGB 히스토그램 H_m와 물체 주변의 배경 히스토그램 H_b로 표현되고, 이 두 가지 히스토그램을 물체추적이 진행되면서 아래와 같은 방법으로 업데이트할 수 있다.

- 방법1: H_m, H_b를 모두 업데이트 한지 않음
- 방법2: H_m, H_b를 같은 가중치로 업데이트
- 방법3: H_m, H_b를 다른 가중치로 업데이트

실험적으로 방법3 > 방법2 > 방법1의 순서로 좋은 결과를 나타냈다. 업데이트 가중치가 너무 큰 경우, 물체가 가려짐이 생길 때 쉽게 실패하였고, 물체가 점차 밀려가는 현상(drift)이 일어났다. 그러나 어느 정도 작은 값에서는 안정적으로 물체를 추적하며 물체와 배경의 변화에 따라 히스토그램을 적응적으로 변화시켰다. 또한 H_m의 업데이트 가중치를 H_b의 업데이트 가중치보다 작게 하였을 때(방법3), 보다 좋은 결과가 나타났는데, H_m은 물체의 고유한(invariant)한 특성을 나타내는 부분이고, H_b는 물체가 움직임에 따라 바뀌는 부분이기 때문이라고 실험 결과를 해석하였다.

개선 #2: 반-전역탐색 (Semi-Global Search)

SumShift에서 추적된 물체의 위치는 물체 범위의 히스토그램 유사도의 합이 최대가 되는 위치를 탐색한다. 기존의 SumShift는 이러한 위치를 반복적인 지역탐색을 통해 찾는다. 즉, 물체 주변의 상하좌우 ±1 픽셀 씩 이동하여 유사도의 합이 증가하는 경우에만 탐색을 계속하는 방식이다. 그러나 물체의 움직임이 큰 경우 지역탐색으로는 물체를 놓치기 쉽다. 또 영상 전체로 전역탐색을 하는 경우 물체의 직전 위치를 고려하지 않아 물체와 유사한 배경을 물체로 오인하기 쉽다. 따라서 지역탐색과 전역탐색의 장점을 합쳐서 물체 주변을 탐색하되, 인접 픽셀 이동 후에 유사도의 합이 증가하지 않더라도 지정된 크기만큼까지는 탐색을 계속하는 반-전역탐색을 도입하였다.

개선 #3: 물체의 크기와 종횡비 변화 고려

SumShift의 경우 물체의 크기를 처음에 설정된 값을 고정값으로 사용한다. 그러나 많은 추적 문제에서 물체의 크기와 종횡비가 달라지는 경우가 잦다. 따라서 SumShift의 반-전역탐색에 물체의 크기와 종횡비의 변화도 고려하도록 하였다. 따라서 기존의 x, y 방향뿐만 아니라 너비 w와 길이 h를 고려하도록 탐색공간을 확장하였다. 따라서 기존의 단순한 유사도 합 대신 물체의 넓이와 둘레를 고려한 유사도의 합을 이용해 물체를 탐색하였다.

3. 실험 결과 및 분석

SumShift와 각각의 개선이 반영된 SumShift를 기존의 물체추적기술과 VOT 2015 데이터셋을 이용하여 비교하였다. 모든 알고리즘은 C/C++로 구현되었고 Intel Core i7-5500U 2.4GHz (RAM 8GB)에서 테스트되었다. 실험 결과 개선 #1을 반영하였을 때

표 1. 개선된 SumShift 추적기의 실험 결과

Name	Accuracy	# of Fails	Robustness	Time msec]
NCC	0.477428	362	0.817822	2.932
DSST(dlib)	0.491650	194	0.897827	37.602
MIL(OpenCV)	0.428124	129	0.930841	88.477
SS(SumShift)	0.496298	133	0.928775	13.043
SS+#1	0.505445	92	0.950173	14.419
SS+#1+#2	0.505548	87	0.952816	23.668
SS+#1+#2+#3	0.498027	84	0.954405	24.246

Tracker	A	R	$\hat{\Phi}$	Speed	Impl.
MDNet*	**0.60**	**0.69**	**0.38**	0.87	M C G
DeepSRDCF*	0.56	1.05	*0.32*	0.38	M C
EBT	0.47	*1.02*	0.31	1.76	M C
SRDCF*	0.56	1.24	0.29	1.99	M C
LDP*	0.51	1.84	0.28	4.36	M C
sPST*	0.55	1.48	0.28	1.01	M C
SC-EBT	0.55	1.86	0.25	0.80	M C
NSAMF*	0.53	1.29	0.25	5.47	M
Struck*	0.47	1.61	0.25	2.44	C
RAJSSC	0.57	1.63	0.24	2.12	M
S3Tracker	0.52	1.77	0.24	14.27	C
SumShift	0.52	1.68	0.23	16.78	C
SODLT	0.56	1.78	0.23	0.83	M C G
DAT	0.49	2.26	0.22	9.61	M
MEEM*	0.50	1.85	0.22	2.70	M
RobStruck	0.48	1.47	0.22	1.89	C
OACF	*0.58*	1.81	0.22	2.00	M C
MCT	0.47	1.76	0.22	2.77	C
HMMTxD*	0.53	2.48	0.22	1.57	C

그림 2. VOT Challenge 2015 결과 [4]
(총 62개 추적기 중 상위 19개의 결과)

추적 실패 횟수가 크게 줄었고, 나머지 개선 #2와 #3를 조금씩 실패 횟수를 줄일 수 있었다. 수행시간은 약 0.8배 늘어났다. VOT 2015에 개선 #1과 #2를 반영한 SumShift와 모든 개선이 반영된 S3Tracker(Scalable SumShift Tracker)를 제출하였고, 각각 전체 62개의 추적기 중 12위와 11위를 하였다. 상위 알고리즘 중 실시간 동작이 가능한 것은 S3Tracker와 SumShift뿐이기 때문에 실시간성이 필요한 응용에서 SumShift는 매우 효과적인 방법으로 생각된다.

감사의 글

본 연구는 국토교통부와 국토교통과학기술진흥원의 연구과제(철도역사 서비스 표준화 및 안전관리 자동화 기술 개발, 14RTRP-B091404-01)에 의해서 수행되었음.

참고문헌

[1] Visual Object Challenge, http://votchallenge.net/
[2] Visual Tracker Benchmark, http://visual-tracking.net/
[3] Lee and Yu, Visual Tracking by Partition-based Histogram Backprojection and Maximum Support Criteria, in ROBIO, 2011.
[4] Kristan et al., The Visual Object Tracking VOT2015 Challenge Results, in ICCV-Workshop, 2015.

스틱셀 기반 배경 변환을 이용한 다시점 동영상 합성

Multi-view video stitching using stixel-based background warping

이규열, 심재영 ○

울산과학기술원(UNIST) 전기전자컴퓨터공학부

ever1135@unist.ac.kr, jysim@unist.ac.kr

요 약

영상 합성 기법은 다수의 영상을 넓은 화각을 갖는 단일 영상으로 만든다. 기존의 기법은 촬영된 영상의 위치가 가깝고, 전반적인 장면을 평면으로 근사하는 가정이 있었다. 본 논문에서는 영상에서 근경과 원경을 구분하는 경계선을 찾고, off-plane 픽셀 변환을 적용하여, 촬영된 영상의 위치가 멀고 전반적인 장면이 평면이 아니더라도 투영 왜곡을 적게 만드는 영상 합성 방법을 제안한다.

1. 서론

영상 합성은 서로 시점이 다른 다수의 영상을 합성하여 넓은 화각을 가지는 단일 영상으로 만드는 기술이다. 다수의 카메라를 이용하는 스포츠, 영상기반 감시, 차량용 서라운드 뷰와 같이 다양한 분야에서 널리 활용되고 있다.

전통적인 영상 합성 기법은 특징점 추출과 매칭(matching), 호모그래피 변환(homography transform) 추정, 영상 워핑(warping)과 블렌딩(blending) 과정으로 이루어진다. 이러한 기법은 다수의 영상이 매우 가까운 위치에서 촬영되고, 전반적인 장면이 평면으로 근사된다는 가정을 하기 때문에 한정된 영상에서만 성공적인 결과를 얻을 수 있다는 한계가 있다.

최근에 소개되는 영상 합성 기법은 보다 먼 위치에서 촬영된 영상과, 다양한 장면을 합성하기 위해 적응적 호모그래피 변환을 사용한다[1]. 호모그래피가 만드는 투영 왜곡을 줄이기 위해 호모그래피와 닮음 변환(similarity transform), 아핀 변환(affine transform)을 혼합해 사용하는 방법도 개발되었다[2]. 하지만 영상이 서로 매우 먼 위치에서 촬영되며, 전반적으로 평면이 아닌 장면에서는 여전히 정확한 합성에 한계가 있다. 한편, 영상의 기하학적인 구조를 분석하여 전경 객체와 같이 호모그래피 변환과 현저하게 다른 변환이 필요한 영역도 합성할 수 있는 off-plane 픽셀 변환 기법이 소개되었다[3].

본 논문에서는 시점이 매우 상이한 두 영상을 합성할 때, 배경 영역을 적응적으로 변환하여 왜곡을 줄이는 동영상 합성 알고리즘을 제안한다. 배경 영역은 근경 영역과 원경 영역으로 이루어졌다는 가정을 한다. 입력 영상을 스틱셀(stixel) 구조로 나타내고 두 영역을 구분하는 경계선을 추정한다. 경계선을 기준으로 off-plane 픽셀 변환을 적응적으로 적용함으로써 합성 영상을 얻는다.

2. 스틱셀 기반 영상 합성

2.1 Off-Plane 픽셀 변환

호모그래피 변환은 3 차원 평면 π의 점 X_1이 투영된 영상 좌표 p_1, q_1의 관계를 호모그래피 변환 매트릭스 H_π를 이용하여 아래 식으로 표현한다.

$$q_1 = H_\pi p_1 \quad (1)$$

평면 π위에 있지 않은 off-plane 점 X_2이 투영된 영상 좌표 p_2, q_2의 관계는 투영 깊이 ρ와 에피폴(epipole) e를 이용하여 아래 식으로 표현할 수 있다.

$$q_2 = H_\pi p_2 + \rho e \quad (2)$$

임의의 3 차원 점 X_i가 투영된 영상 좌표 p_i, q_i의 관계는 fundamental matrix F를 이용하여 아래 조건을 만족한다.

$$q_i^T F p_i = 0 \quad (3)$$

3 차원 평면 π에서 투영된 픽셀을 on-plane 픽셀, π가 아닌 위치에서 투영된 픽셀을 off-plane 픽셀이라고 부른다. Off-plane 픽셀 쌍 p_2, q_2 사이의 변환 기법이 제안되었다[3]. 먼저 픽셀 p_2에 해당하는 on-plane 픽셀 $\bar{p}_2$를 아래 식으로 표현한다.

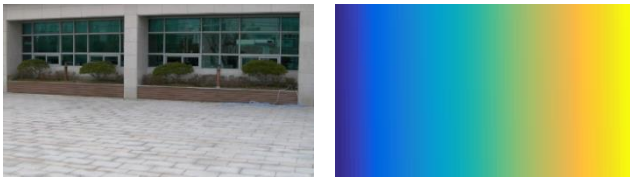

그림 1. 입력 영상과 스틱셀(stixel).

$$\bar{p}_2 = p_2 + h\theta_1 \quad (4)$$

여기서, h 는 $\bar{p}_2$, p_2 사이의 거리를 나타내고, θ_1는 입력 영상의 소실점(vanishing point)을 나타내며 임의의 영상에서 구할 수 있도록 무한상의 점으로 근사한다.

픽셀 q_2에 해당하는 on-plane 픽셀 $\bar{q}_2$은 호모그래피 변환으로 계산한다.

$$\bar{q}_2 = H\bar{p}_2 \quad (5)$$

픽셀 q_2의 위치는 소실점과 on-plane 픽셀을 잇는 선과 에피폴라 선(epipolar line) Fp_2의 교점으로 다음과 같이 표현한다.

$$q_2 = (Fp_2) \times (\theta_2 \times \bar{q}_2) \quad (6)$$

Off-plane 픽셀 변환은 $p_2, \bar{p}_2$ 사이의 거리 h에 의해 정해지며, 영상의 구조와 기하 관계를 통해 공간 특징점 매칭이 없이도 추정할 수 있는 장점이 있다. 예를 들어 전경 객체의 off-plane 픽셀 변환은 전경 객체 영역의 모양으로 결정할 수 있다[3].

2.2 스틱셀 기반 배경 합성

Off-plane 픽셀 변환은 영상의 구조와 기하 관계를 통해 대응점을 추정할 수 있다. 본 논문에서는 근경과 원경 영역의 경계선을 추정하고, 원경 픽셀에 대한 h를 구한다. 원경 영역은 off-plane 픽셀 변환을 적용하고, 근경 영역에는 호모그래피 변환을 적용하여 전체적인 영상 변환을 수행한다.

효과적인 경계선 추정을 위해, 입력 영상을 그림 1 과 같이 막대기 모양의 스틱셀로 분할한다. 서로 다른 스틱셀은 서로 다른 색상으로 표현하였다. 각 스틱셀 $s_1, s_2, \cdots, s_n$에 대한 분할점 $b_1, b_2, \cdots, b_n$을 추정하며, 분할점 b_i 는 스틱셀의 m 개 픽셀 $\{s_{i1}, s_{i2}, \cdots, s_{im}\}$ 중 하나로 표현한다. 분할점의 집합을 분할선으로 정의한다.

분할선의 특징은 대체로 경계부에 있으며, 이를 기준으로 근경과 원경의 히스토그램이 변하고, 최적의 분할선은 기준 영상과 목표 영상의 색상 유사도를 높이도록 결정한다. 이를 반영하도록 다음과 같은 에너지 함수를 정의한다.

$$E = E_{edge} + \alpha E_{histogram} + \beta E_{warping} + \gamma E_{smooth} \quad (7)$$

여기서 각 항에 대한 중요도 상수를 실험적으로

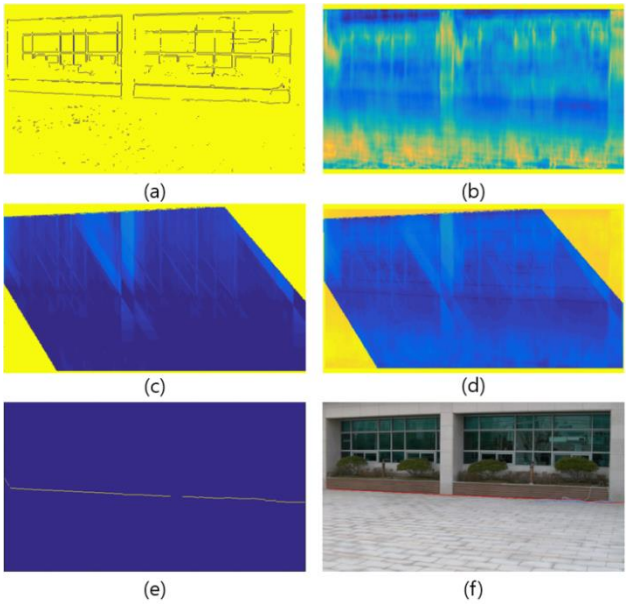

그림 2. (a) 경계 맵, (b) 히스토그램 맵, (c) 워핑 맵, (d) 전체 에너지 맵, (e) 예측한 경계선, (f) 예측한 경계선을 원본 영상에 표시.

$\alpha = 4.5, \beta = 0.05, \gamma = 1$로 설정하고 그래프 컷을 이용하여 최적의 해를 구한다[6].

첫번째 항은 다음과 같이 정의한다.

$$E_{edge} = \sum_{i=1}^{n} 1 - edge(b_i) \quad (8)$$

여기서 $edge(b_i)$는 분할점 b_i의 Sobel 경계 검출기의 결과 값이다.

히스토그램 항을 표현하기 위해, 분할점 b_i를 기준으로 근경의 히스토그램을 $gndHist(b_i)$로, 원경의 히스토그램을 $distHist(b_i)$로 정의한다. 근경과 원경의 히스토그램이 가장 달라지는 분할점 b_i를 선택하도록 다음과 같이 정의한다.

$$E_{histogram} = \sum_i \exp(-\|gndHist(b_i) - distHist(b_i)\|/\tau) \quad (9)$$

τ는 지수의 변화량을 정하는 상수로써 0.1 로 실험적으로 정한다.

색상 유사도 조건을 만족하기 위한 항은 다음과 같이 정의한다.

$$E_{warping} = \sum_{i=1}^{n} \frac{1}{m} \sum_{j=1}^{m} \|I(s_{ij}) - J(t_{ij})\| \quad (10)$$

$I(s_{ij})$는 스틱셀 i의 j번째 픽셀의 색상 값, $J(t_{ij})$는 s_{ij}를 워핑하여 얻은 픽셀 t_{ij}의 색상 값이다. 만약 s_{ij}가 분할점 b_i보다 위에 있다면 원경 영역으로 가정하여 $h_{ij} = \|s_{ij} - b_i\|$를 계산하고 off-plane 픽셀 변환을 한다. 그렇지 않다면 s_{ij}를 근경 영역으로 가정하여 호모그래피 변환을 한다.

이웃한 스틱셀의 분할점이 유사하도록 다음과 같은 smoothness 항을 정의한다.

(a)

(b)

(c)

그림 3. “Office” 영상 합성 결과. (a) 입력 영상, (b) 호모그래피 기반 영상 합성 결과. (c) 제안 알고리즘 결과.

$$\mathrm{E}_{\mathrm{smooth}} = \sum_{i=1}^{n-1}(b_i - b_{i+1})^2 \qquad (11)$$

그림 2 는 경계 맵, 히스토그램 맵, 워핑 맵과 함께 예측한 경계선을 원본 영상에 덧씌운 그림을 보여준다. 각 맵은 s_{ij}의 에너지 값을 시각화 했고, 히스토그램 맵의 상단과 하단의 10 픽셀은, 충분한 개수의 픽셀이 없기 때문에 계산하지 않았다. 워핑 맵은 호모그래피 변환을 기준으로 두 영상에서 촬영된 영역만 계산한다. 정의되지 않은 영역의 에너지는 가질 수 있는 최대 에너지인 $\sqrt{3 * 255^2}$로 설정하였다.

3. 실험 결과 및 분석

그림 3, 4 에서는 360×640 크기의 두 영상을 합성한 두 가지 실험 결과를 도시한다. 그림 3(a)와 4(a)의 입력 영상에 대해, 그림 3(b)와 4(b)는 두 영상의 지면 간 호모그래피 변환을 이용하여 모든 픽셀을 합성한 결과이며, 그림 3(c)와 4(c)는 제안하는 영상 합성 기법의 결과를 보여준다. 제안 기법이 영상에서 투영 왜곡을 현저하게 줄임을 확인할 수 있다.

4. 결론

본 논문에서는 스틱셀 기반으로 off-plane 픽셀 변환을 적응적으로 적용하는 영상 합성 기법을 소개하였다. 입력 영상에서 스틱셀 별로 할당된 에너지를 최적화함으로써 근경과 원경을 구분하는 경계선을 찾고, 이를 기준으로 적응적인 영상 변환을 수행하였다.

(a)

(b)

(c)

그림 4. “Badminton” 영상 합성 결과. (a) 입력 영상, (b) 호모그래피 기반 영상 합성 결과. (c) 제안 알고리즘 결과.

감사의 글

이 논문은 2013 년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(2013R1A1A2011920).

참고문헌

[1] J. Zaragoza, T.-J. Chin, Q.-H. Tran, M.S. Brown, and D. Suter, “As-projective-as-possible image stitching with moving dlt,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 36, no. 7, pp. 1285–1298, July 2014.

[2] C.-H. Chang, Y. Sato, and Y.-Y. Chuang, “Shape-preserving half-projective warps for image stitching,” in Proc. IEEE CVPR, June 2014, pp. 3254–3261.

[3] K.–Y. Lee and J.-Y. Sim, “Robust video stitching using adaptive pixel transfer,” in Proc. IEEE Intl. Conf. Img. Proc, pp. 831-817, 2015

[4] S.-Y. Lee, J.-Y. Sim, C.-S. Kim, and S.-U. Lee, “Correspondence matching of multi-view video sequences using mutual information based similarity measure,” IEEE Trans. Multimedia, vol. 15, no. 8, pp. 1719–1731, Dec 2013.

[5] R. Hartley, and A. Zisserman. Multiple view geometry in computer vision. Cambridge university press, 2003.

[6] Y. Boykov, O. Veksler, and R. Zabih. "Fast approximate energy minimization via graph cuts IEEE Trans. Pattern Anal. Mach. Intell., 2001.

다중 영상 인식을 이용한 사용자 인증 방법

김계경 [O,1], 강상승 [1], 지수영 [1], 김진호 [2]

[1] 한국전자통신연구원 지능형인지기술연구부

[2] 경일대학교 전자공학과

kyekyung@etri.re.kr, kss@etri.re.kr, chisy@etri.re.kr, ho@kiu.ac.kr

요 약

본 논문에서는 다중 카메라로 획득한 얼굴 영상과 사용자 정보가 기록된 문서 영상 인식을 이용하여 사용자를 인증하는 방법에 대하여 제안하였다. 출입 통제, 금융, 보안 등 각종 분야에서 얼굴 인식 기술을 적용하여 사용자를 인증하여 왔으나, 얼굴 인식은 조명, 포즈, 표정, 시간 변화에 민감하여 얼굴 영상을 획득하는 환경 변화에 의존적인 얼굴 인식 성능을 나타냄으로써 안정된 얼굴 인식 성능을 제공할 수 없는 문제점이 발생하였다. 따라서, 주변 환경 변화에 덜 민감한 사용자 인증 방법으로 얼굴 영상뿐만 아니라 사용자 정보가 기재된 문서 영상을 결합한 다중 영상 인식 방법으로 사용자 정보를 인식함으로써 안정된 사용자 인증 성능을 제공할 수 있도록 하였다. 실제 환경에서 다양한 종류의 카메라로 획득한 얼굴 영상 및 문서 영상을 대상으로 얼굴 인식 및 문서 인식 실험을 수행하여 제안된 다중 영상 인식 방법의 타당성을 평가하였다.

1. 서론

얼굴 인식 기술은 영상을 획득하고 처리하는 과정에서 경제적이고 설치가 용이하며 사용자 편의성을 제공할 수 있다는 장점으로 인하여 출입 통제, 금융 보안, 인터넷 거래 및 사용자 맞춤형 서비스 제공 등의 목적으로 다양한 분야에서 활용되어 왔다[1-3]. 최근 인터넷 거래 및 사용자 맞춤형 서비스 제공을 위하여 원격 접속된 사용자 인증에 대한 요구가 증대됨에 따라 얼굴 인식 기술의 상용 기술 적용 방안에 대한 연구가 활발히 진행 중이다. 이러한 환경에서의 얼굴 인식은 사용자의 휴대폰 카메라나 웹 카메라를 이용하여 무제한 환경에서 사용자의 얼굴 영상을 획득한 다음 사용자 인증 절차를 수행하게 됨으로, 복잡한 배경, 조명, 포즈, 표정 변화에 강인하게 안정된 얼굴 인식 성능을 보장할 수 있는 얼굴 인식 기술이 요구된다. 최근, 딥러닝 기술이 개발되어 대용량 얼굴 영상을 이용하여 얼굴 인식 성능 저하 문제를 현저히 개선함으로써 얼굴 인식의 상용 기술로의 활용도가 높아졌다.

본 논문에서는 얼굴 영상뿐만 아니라 사용자 정보가 기재된 문서 영상을 이용한 다중 영상 기반 사용자 인증 방법을 제안하였다. 다양한 얼굴 포즈에 강인한 얼굴 인식 성능을 보장하기 위하여 다중 얼굴 모델을 생성하고 가버 특징을 추출하여 안정된 얼굴 인식 성능을 나타낼 수 있도록 하였다. 환경 변화에 따른 얼굴 인식 성능 저하 문제점을 보완하기 위하여 사용자 정보가 기재된 문서 영상을 카메라로 획득한 다음 사용자 정보에 대한 문자 인식을 수행하였다. 실제 환경에서 다양한 시각 센서로 획득한 얼굴 영상과 다양한 글자 폰트 및 크기, 복잡한 배경 문서 또는 한글과 영어가 혼용된 문서 영상에 대한 인식 실험을 통하여 사용자 정보 인증 방법을 평가하였다.

2. 다중 영상 기반 사용자 인증

실생활 환경에서 획득한 사용자의 얼굴 영상 및 사용자 정보가 기재된 문서 영상을 획득하여 영역 추출 및 인식 과정을 수행하였다.

2.1 다중 영상 기반 얼굴 인식

환경 변화가 다양한 실제 환경에서 안정된 얼굴 인식 성능을 보장하기 위하여 조명 및 포즈 변화에 강인한 가버 특징을 추출하고, 포즈 변화에 강인한 얼굴 특징 추출을 위하여 다양한 포즈 변화에 따른 얼굴 모델을 생성하여 그림 1 과 같이 인식하였다.

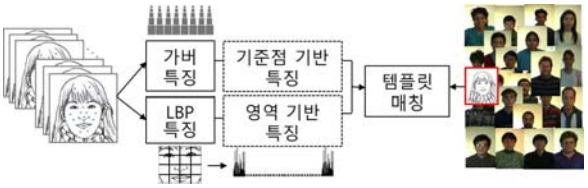

그림 1. 다중 영상 얼굴 인식 방법

2.2 문서 영상 인식

사용자 정보 인식을 위하여 명함이나 사용자 카드 영상 정보를 이용하였다. 휴대형 카메라로 획득한 문서 영상을 대상으로 DOG 필터 및 적응적 이진화 [4] 기법을 적용하여 이진 영상을 추출한 다음, 블럽 분석을 통하여 문자와 그림 영역을 추출하였다. 카메라로 획득한 문서 영상은 기울어지거나 다양한 문자 크기 및 폰트로 구성되어 있으며, 한글과 영어가 혼재되어 있거나 이웃하는 문자끼리 서로 접촉된 경우가 흔히 나타난다. 본 논문에서는 블럽의 유형을 분석하여 한글과 영어 또는 접촉된 문자를 분할하도록 하였으며 분할된 문자 영상의 인식 결과를 통해 문자 분할 여부를 결정하였다. 접촉된 문자의 경우, 유형별 접촉 문자 분할 방법[5]을 적용하여 문자를 분할한 다음 한영 혼용 문자를 인식하도록 하였다. 다양한 폰트 종류 및 크기를 가지는 문자를 인식하기 위하여 실제 환경에서 카메라로 획득한 문자를 검출하고 문자의 기하학적인 특징 및 통계적인 특징 정보를 추출하여 특징 벡터를 구성한 다음, 신경회로망을 이용하여 학습한 다음 학습시키지 않은 문자 영상을 대상으로 인식 성능을 테스트 하였다. 그림 2 는 카메라로 획득한 문서 영상을 대상으로 문자 영역과 그림 영역을 분할한 다음 문자 영역 추출, 특징 추출 및 인식 단계를 거쳐 문자 인식하는 흐름도를 나타낸 것이다.

그림 2. 문자 인식 흐름도

3. 실험 결과 및 분석

본 논문에서 제안한 다중 영상 기반 사용자 인식 방법의 타당성을 평가하기 위하여 실제 환경에서 획득한 얼굴 영상 데이터베이스 및 표준 얼굴 영상 데이터베이스를 이용하여 얼굴 인식 실험을 수행하였다. 실제 환경에서 웹 카메라로 획득한 사용자의 얼굴 및 사용자 정보가 기재된 문서에서 사용자 얼굴 인식을 그림 3 과 같이 수행하였다.

그림 3. 사용자 얼굴 인식

또한, 사용자 정보가 기재된 문서 영상을 분석한 다음 문자 영역 추출 과정을 거쳐 한영 혼용 문자를 대상으로 인식하였고, 그 결과를 그림 4 에 나타내었다.

그림 4. 문서 영상 분석 및 인식 결과

4. 결론

본 논문에서는 다중 영상을 이용하여 사용자를 인식하는 방법에 대하여 제안하였다. 사용자 인증을 위해 실제 환경에서 조명 및 포즈를 변화시켜 가면서 카메라로 획득한 얼굴 영상 및 사용자 정보가 기재된 문서 영상에서 얼굴 및 문자를 인식하였다. 안정된 얼굴 인식 성능을 보장하기 위하여 다중 얼굴 모델을 생성한 다음 특징 벡터를 조합하여 얼굴 특징 벡터를 생성하였다. 문서 영상은 문자 영역 분할, 접촉된 문자 분할 및 한영 혼용 문자 인식 결과로부터 사용자 정보를 인식하였다. 본 논문에서는 사용자 인증을 위해 얼굴 및 문자 인식 결과를 결합하여 사용함으로써 실제 환경에서 사용자 인증에 대한 활용도를 높일 수 있도록 하였다.

Acknowledgement

본 연구는 미래창조과학부 및 한국산업기술평가관리원의 산업원천기술개발사업의 일환으로 수행하였음. [10041627, 다축 모션 플랫폼을 기반으로 한 범용 오감 융합형 스포츠 시뮬레이터 개발]

참고문헌

[1] R. Ramadan and R. A. Kader, “Face Recognition Using Particle Swarm Optimization-Based Selected Features,” Int'l Journal of Signal Processing, Image Processing and Pattern Recognition (June), vol. 2, pp. 51-65, 2009.

[2] M. Yang, D. Kriegman, and N. Ahuja, “Detecting Faces in Images: A survey,” IEEE Trans. PAMI, vol. 24, no. 1, pp. 34-58, 2002.

[3] M. Turk, and A. Pentland. 1991, Eigenfaces for Recognition, Journal of Cognitive Neuroscience (March), 3, 1, 71-86.

[4] 김계경, 김재홍, 이재연, “조명 및 회전에 강인한 물체 인식,” 한국콘텐츠학회논문지, pp. 1-8, 2012.

[5] 김계경, “ 인쇄체 한영 혼용 문서 인식을 위한 기록블록기반의 문자 분할,” 경북대학교 전자공학과 박사학위 논문, 1997.

Multi-thread based depth estimation system from 4D light field data

한미선, 최은정, 김정태 O

이화여자대학교 전자공학과

miseon.han@ewhain.net, ejeong_choi@naver.com, jtkim@ewha.ac.kr

요 약

라이트 필드 데이터는 공간 정보뿐만 아니라 방향 정보를 포함하는 대용량 데이터이므로 빠른 시간 내에 깊이 추정을 수행하는 것이 어렵다는 문제점을 가지고 있다. 본 논문에서는 라이트 필드 데이터의 공간 정보와 방향 정보를 활용하여 epipolar plane image (EPI) 를 생성하고, EPI 에서 밝기 값에 대한 유사도를 측정함으로써 깊이 정보를 추정하는 알고리즘을 설계한다. 또한 알고리즘을 병렬처리가 가능하도록 설계하여 멀티 쓰레드 기반의 프로그래밍을 적용함으로써 빠르고 정확한 깊이 추정 알고리즘 기법을 연구한다.

1. 서론

라이트 필드 카메라는 공간 정보만을 기록하는 일반 카메라와는 달리 방향 정보를 포함하여, 단 한 번의 노출 만으로 여러 방향 정보를 가지는 영상들을 동시에 획득할 수 있다는 장점을 가진다. 이러한 장점으로 인하여 최근 라이트 필드 데이터를 이용한 깊이 추정 알고리즘에 대한 관심은 더욱 증가하고 있다. 하지만 라이트 필드 데이터는 공간 정보와 방향 정보를 모두 기록하므로 데이터 량이 방대하여 라이트 필드 데이터를 활용한 알고리즘을 수행하기 위해서는 오랜 연산 시간을 요구하는 문제가 발생한다. 그러므로 라이트 필드 데이터를 이용하여 정확하면서 빠르게 깊이 추정이 가능한 알고리즘에 대한 연구가 필요하다.

본 논문에서는 라이트 필드 데이터의 공간 정보와 방향 정보를 활용하여 EPI 를 생성하고 EPI 에서 밝기 값에 대한 유사도를 측정함으로써 정확한 깊이 추정이 가능한 알고리즘을 설계한다. 또한 이 알고리즘을 병렬적으로 처리 가능하도록 시스템을 설계하고 멀티 쓰레드 기반의 프로그래밍을 적용함으로써 정확하면서 빠른 연산이 가능한 깊이 추정 알고리즘을 제안한다.

2. Multi-thread based depth estimation system

제안하는 깊이 추정 알고리즘은 크게 2 부분으로 구분할 수 있다. 깊이 추정 알고리즘을 설계하는 부분과 설계한 알고리즘을 병렬화하는 부분이다. 깊이 추정 알고리즘은 라이트 필드 데이터를 이용하여 EPI 를 생성하고 EPI 에서 깊이 정보를 의미하는 라인의 기울기를 추정하기 위해, 각도에 따른 밝기 값에 대한 유사도를 측정하여 깊이 정보를 획득한다 [1]. 더욱 정확한 깊이 정보를 추정하기 위해 주변 공간 정보를 고려한 정규화 항을 추가하여 목적함수를 설계한다. 설계한 목적함수를 병렬 시스템으로 구성함으로써 멀티 쓰레드 기반의 고속 깊이 추정 알고리즘 시스템을 설계한다.

2.1 정규화 항을 고려하여 정확한 깊이 추정 알고리즘 설계

깊이 정보를 추정하기 위해 라이트 필드 데이터를 이용하여 EPI 를 생성한다. EPI 는 공간 정보와 방향 정보를 포함하며 라이트 필드 데이터인 $L(x,y,s,t)$ 에 대하여 $x-s$ 그리고 $y-t$ 축을 가지는 epipolar plane 영상이다. 그림 1 과 같이 EPI 에서 표현되는 라인은 깊이 정보를 의미하므로 영상 위치에서 각도에 따른 밝기 값에 대한 유사도를 위한 유사도 지수를 연산하고 분석한다 [1].

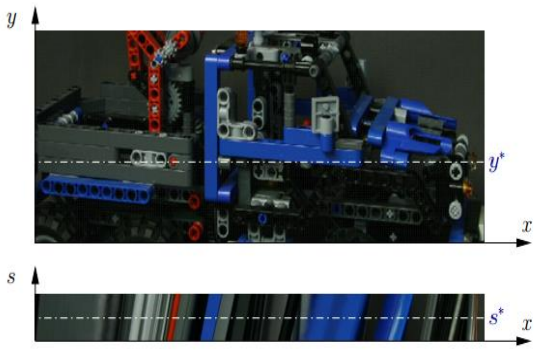

그림 1. $L(x,y,s,t)$ 중 (s^*,t^*) 를 특정하여 얻은 영상과 $x-s$ 축을 가진 EPI 영상 [4]

획득한 유사도 지수를 이용하여 식 (1) 과 같이 정규화 항을 고려한 목적함수를 설계하여 깊이 정

보를 추정한다.

$$\hat{\alpha} = \underset{\alpha}{\operatorname{argmin}} \sum_{x,y} V(\alpha(x,y)) + \lambda R(\alpha) \tag{1}$$

EPI 에서 라인의 기울기인 $\alpha = [\alpha(x_i, y_j)]$ 이고 $i = 1, \cdots, N$, $j = 1, \cdots, M$ 이다. 여기서 N 과 M 은 영상의 가로와 세로의 크기이다. $V(\alpha(x,y))$ 는 α 값에 따라 저장된 픽셀간의 밝기 유사도 지수를 의미하며, λ 는 분산 정보와 정규화 항 간의 조절계수를 의미하고 $R(\alpha)$ 는 주변 공간 정보를 고려한 정규화 항으로 주변 깊이 정보와의 차이를 최소화하여 잡음을 감소시키는 역할을 한다. 그러므로 목적함수는 라이트 필드 데이터를 통해 획득한 유사도 지수가 최소가 되면서 주변 깊이 정보와의 차이가 최소가 되는 깊이 정보인 α 를 찾아 정확한 깊이 정보를 추정할 수 있다.

2.2 병렬처리가 가능한 깊이 추정 알고리즘 설계

식 (1) 과 같이 구성된 목적함수에서 정규화 항인 $R(\alpha)$ 은 주변 픽셀 정보를 이용하므로 모든 픽셀 정보들이 중첩되어 있다 [2]. 그러므로 정규화 항을 병렬처리가 가능한 형태로 재구성하여 목적함수를 최소화하는 기울기 α 를 멀티 쓰레드 기반 프로그래밍을 통해 수행하여 빠르면서 정확한 깊이 추정 영상을 획득한다.

Open muti-processing (OpenMP) 프로그래밍을 사용하면 CPU 개수만큼 쓰레드를 생성하여 연산을 수행할 수 있으므로 단일 쓰레드만을 사용하여 연산을 수행하는 것보다 빠르게 연산을 수행할 수 있다. 또한 멀티 쓰레드 기반의 프로그래밍 중 하나인 compute unified device architecture (CUDA) 는 OpenMP 와는 달리 GPU 에서 생성하는 쓰레드를 사용하므로 OpenMP 보다 더 많은 쓰레드를 이용하여 연산을 수행할 수 있다. 본 논문에서는 CUDA 와 OpenMP 프로그래밍을 통하여 멀티 쓰레드 기반의 고속 깊이 추정 알고리즘을 수행하였다.

3. 실험 결과 및 분석

HCI Light Field Research 에서 제공하는 dataset 중 Mona, StillLife, Buddha dataset 을 이용하여 실험을 수행하였고, dataset 은 9 x 9 의 방향 정보와 768 x 768 x 3 의 공간 정보를 포함한다 [3]. 그 중 StillLife dataset 에 대한 깊이 추정 영상은 그림 2 와 같다.

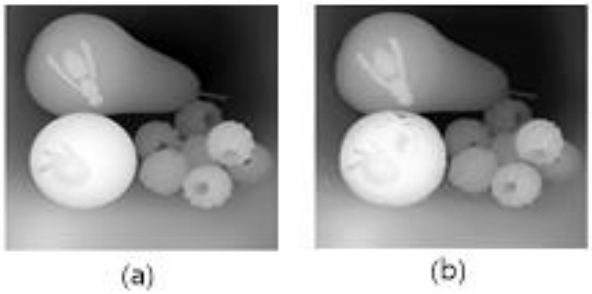

그림 2. StillLife dataset 의 깊이 추정 영상
(a) StillLife dataset 의 ground truth [2]
(b) 제안한 알고리즘을 적용 결과

사용한 그래픽 카드는 NVdia 의 Quadro K600 이며, 1GB 의 GPU memory 와 192 개의 CUDA 코어를 지닌다. 사용한 그래픽 카드에 CUDA 프로그래밍을 적용하였을 때, 병렬 프로그래밍을 적용하기 전보다 3.5 배 정도 향상하였음을 확인하였으나 이 값은 GPU 의 사양에 따라 의존적인 결과를 보인다. 또한 깊이 추정 결과와 ground truth 와의 영상을 비교하기 위하여 correlation coefficient 를 연산하였다. 이 값은 1 에 가까울수록 ground truth 와 실험 결과 영상이 유사하다는 것을 의미한다. Correlation coefficient 값 또한 영상에 따라 의존적이나 실험한 dataset 에 대하여 모두 0.9 이상의 값을 가지므로 ground truth 와 거의 비슷한 결과를 보임을 확인할 수 있었다.

4. 결론

본 논문에서는 라이트 필드 데이터를 이용한 깊이 추정 기법을 제안하였고 그 성능은 실험을 통해 검증하였다. 향후 알고리즘의 속도 개선을 위하여 추가적으로 연구를 진행할 계획이다.

참고문헌

[1] Bolles, R. C., Baker, H. H., & Marimont, D. H. (1987). Epipolar-plane image analysis: An approach to determining structure from motion. International Journal of Computer Vision, 1(1), 7-55.

[2] Fessler, J. A. (2008). Image reconstruction: Algorithms and analysis. Under preparation.

[3] Wanner, S., Meister, S., & Goldluecke, B. (2013). Datasets and benchmarks for densely sampled 4D light fields. In Annual Workshop on Vision, Modeling and Visualization: VMV (pp. 225-226).

[4] Wanner, S., & Goldluecke, B. (2012, June). Globally consistent depth labeling of 4D light fields. In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on (pp. 41-48). IEEE.