

공간적 특징을 이용한 신경 회로망 기반 객체 분할 및 움직임 예측

김형진, 이동규, 이두수

한양대학교 전자공학과

전화 : (02) 2290-0358 / 팩스 : (02) 2298-1796

Neural network based Object segmentation and optical flow estimation using spatial feature

Hyung-Jin Kim, Dong-Gyu Lee, Doo-Soo Lee

Dept. of Electronic engineering, Hanyang University

E-mail : f5tiger@ses.hanyang.ac.kr

요약문

동영상에서 움직이는 객체 분할 및 모션 예측을 동시에 수행할 수 있는 연구는 다양한 방법으로 시도 되어 왔다. 실제 이미지를 서로 다른 움직임이나 서로 다른 공간적인 특정 영역으로 분리 될 수 있다고 가정 한다면 복수의 객체 또는 객체의 움직임으로 표현 할 수 있다.

객체 분할 측면에서 볼 때 효율적인 분할을 위해서는 특징 입력 벡터의 선택이 중요한 변수로 작용한다. 본 연구에서는 정밀한 객체 분할을 위해 밝기, 질감(Texture) 정보와 같은 정지영상의 특징 입력 벡터와 움직임 벡터 같은 동영상의 특징 입력 벡터를 동시에 사용한다. 분리된 객체는 각각의 클래스를 구성하게 되고 이를 위한 클래스 분류기로서 Median Radial Basis 신경 회로망을 사용한다.

객체 분할과 움직임 예측을 위해서 확률적 방법을 통한 에너지 함수를 구하고 비용함수를 도입한다. 신경 회로망의 각 Basis 함수는 영상의 특정한 영역에서 활성화되며 객체의 분류를 위해 신경 회로망 출력으로 가중치의 합으로서 나타나게 된다.

I. 서론

동영상에서 움직임(Motion) 표현은 동영상을 해석하고 이해하는데 있어서 중요한 단계로 인식 되어진다.

움직임을 찾기 위한 방법으로 광학적 흐름(Optical flow)을 이용하며 이를 이용해 모션을 분할하는 방법은 A posteriori criterion[2]을 이용하는 방법이 제안되었다. 방법론적으로는 Markov random fields 와 Gibbs distribution[3]을 이용한 방법이 사용되며 모션 필드를 평탄화 하기 위한 방법으로 ICM(Iterated conditional mode) 이 사용되어 왔다. 본 연구에서는 광학적 흐름과 객체분할 예측을 동시에 수행하기 위한 방법으로 전역적 기준(global criterion)[1]을 최소화시키는 방법을 사용했으며 이를 위해 비용함수의 개념을 도입한다. 이미지를 일정한 크기의 블록으로 나누고 각각의 블록을 공간적 매개변수인 블록 평균 밝기, 질감(texture)과 시간적 매개변수인 움직임 벡터의 4 차원 특징 벡터를 구성한다. 블록의 4차원 특징 벡터를 이용해서 Unsupervised classification을 수행하게 되며 이 분류 작업은 매개변수 공간의 거리와 Bayesian 이론[2]의 판단 기준으로부터 결정된다.

광학적 흐름과 객체 분할의 모델링을 위해 MRBF (Median Radial Basis Function)[5] 이 사용 된다. MRBF는 2 단계의 신경 회로망으로 구성되며 각 커널은 가우시안 함수로서 첫 단계 뉴런으로 구성하게 되고 두 번째 단계에서는 광학적 흐름과 객체분할을 위한 각 커널의 가중치의 합으로 구성된다.

MRBF 신경회로망은 2 단계의 학습 단계를 거쳐 매개변수를 구하게 된다. 첫 번째 단계에서 LVQ(learning vector quantization)[7] 과 비슷한 클러스터링 기법으로

로 각 커널의 중심(Mean) 과 공분산(Covariance)을 구하고 두 번째 단계에서는 역전파 학습 방법에 의한 신경회로망의 출력 가중치를 구하게 된다. 학습이 끝난 신경회로망에 블록별 입력벡터를 입력 하면 최종 출력 뉴런이 활성화 되는 객체로 입력 블록을 할당 함으로써 움직임 예측과 객체 분류를 하게 된다.

II. 이미지 분할 기준

객체 분할을 위한 첫 단계로 각 블록별 4 차원 특징 벡터 u_{IJ} 를 구하고 분할을 위한 대표 예측 벡터인 μ_k 와의 에너지 함수 비교를 통해서 우세 클래스를 구하게 된다.

1. 블록 특징 벡터

각 블록을 B_{IJ} , $I = 0, \dots, n_x - 1$, $J = 0, \dots, n_y - 1$ 으로 나타내고 B_{IJ} 에 따른 특징 벡터를 u_{IJ} 라고 표시한다. S_{IJ} 는 공간적 특징 벡터이고 M_{IJ} 는 시간적 특징 벡터이다.

$$u_{IJ} = [S_{IJ}, M_{IJ}] \quad (1)$$

$$S_{IJ} = [G_{IJ}, T_{IJ}] \quad (2)$$

$$G_{IJ} = \frac{1}{D_x D_y} \sum_{h=-\frac{D_x}{2}}^{\frac{D_x}{2}} \sum_{v=-\frac{D_y}{2}}^{\frac{D_y}{2}} f(I+h, J+v) \quad (3)$$

G_{IJ} 는 평균 흑백(Grey) 값이며 $f(I+h, J+v)$ 는 이미지의 흑백 값을 의미하고 D_x, D_y 는 블록 크기를 말한다. T_{IJ} 는 B_{IJ} 의 질감 정보(Texture)를 말한다.

질감 정보를 얻기 위해 2차원 가보(Gabor) 필터[8] 를 사용하며 다음과 같이 정의한다.

$$h(x, y) = e^{-\frac{1}{2}(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2})} \times \cos(2\pi\mu_0(x\cos\theta + y\sin\theta)) \quad (4)$$

σ_x, σ_y 는 2차원 가우스 함수의 공간 상수를 말하며 μ_0 는 변조 되는 주파수 성분이다.

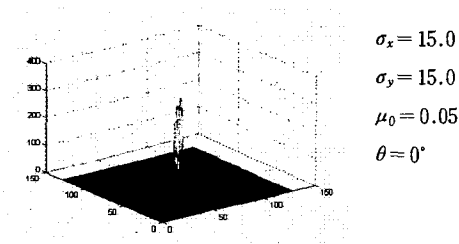


그림 1. 2차원 가보 필터의 주파수 영역
그림 1 에서 가보 필터는 주파수 상에서 두 개의 가우

스 분포 함수 형태로 나타나며 가우시안 정점의 위치는 $\sigma_x, \sigma_y, \mu_0, \theta$ 에 따라 바뀐다.

식(4)의 질감 정보 $Gabor_{IJ}$ 는 통계적으로 대역통과(Bandpass filtered) 된 형태로서 Rician 분포를 가진다

$$Gabor_{IJ} = h_j(x, y) * B_{IJ} \quad (4)$$

$$T_{IJ} = ||Gabor_{IJ}|| \quad (5)$$

시간적 특징 벡터 M_{IJ} 는 움직임 벡터로서 각 움직임 벡터의 유사성을 효율적으로 검출하기 위해 길이 R 과 각도 θ 로 변환한다.

$$M_{IJ} = [R, \theta] \quad (6)$$

$$R = \sqrt{m_{x,IJ}^2 + m_{y,IJ}^2}, \quad \theta = \tan^{-1}\left(\frac{m_{y,IJ}}{m_{x,IJ}}\right) \quad (7)$$

2. 객체 분류 (Object classification)

동영상의 각 프레임은 L 개의 다른 영역으로 분할 될 수 있다. X_1, X_2, \dots, X_L 각 영역은 각각의 클래스를 구성하게 된다. X_k 는 대표 특징 벡터 μ_k 로 표현되며 광학 흐름과 분할 정보를 담고 있다.

$$\mu_k = [S_k, M_k] \quad (8)$$

\hat{S}_k, \hat{M}_k 를 예측 매개변수로 나타내면 a posteriori 확률 $P(\hat{M}_k, \hat{S}_k | f_{t-1}, f_t)$ 가 가장 큰 값일 때 B_{IJ} 는 X_k 에 속하게 된다. 즉 $B_{IJ} \in X_k$

$$P(\hat{M}_k, \hat{S}_k | f_{t-1}, f_t) > P(\hat{M}_j, \hat{S}_j | f_{t-1}, f_t) \quad (9)$$

$$j = 1, \dots, L, \quad j \neq k$$

a posterior 확률을 기준으로 Bayes' rule를 적용하면 a posterior 분포는 다음과 같다.

$$P(\hat{M}_k, \hat{S}_k | f_{t-1}, f_t) = \frac{P(f_t | f_{t-1}, \hat{M}_k, \hat{S}_k) P(\hat{M}_k, \hat{S}_k | f_{t-1}) P(\hat{S}_k | f_{t-1})}{P(f_t | f_{t-1})} \quad (10)$$

(10)의 각 조건부 확률을 에너지 함수 $E(X)$ 로 나타내면 다음과 같다.

$$P(X) = \frac{1}{Z} \exp\left[-\frac{E(X)}{\beta}\right] \quad (11)$$

Z는 정규화 상수이며 β 는 각 $E(X)$ 를 조절하기 위한 상수로서 사용된다. 따라서 식(9)의 확률 예측의 문제는 식 (10) 과 (11) 에 의해 식(12)의 에너지 함수를 최소화시키는 문제로 전환시킬 수 있다.

$$E_j = E(f_t | f_{t-1}, \hat{M}_j, \hat{S}_j) + E(\hat{M}_j | \hat{S}_j, f_{t-1}) + E(\hat{S}_j | f_{t-1}) \quad (12)$$

첫째 항은 f_{t-1} 프레임 기준에서 f_t 프레임으로의 광학 흐름과 객체분할의 예측 에너지 함수이고 두 번째 항은 광학적 흐름에 관한 예측 함수이며 세 번째

함은 분할에 관계된 에너지 함수이다. E_j 를 최소화하기 위해서 세가지의 에너지 함수가 동시에 최소 값을 가져야 한다. 따라서 E_j 가 최소값을 가질 때 광학적 흐름과 분할의 기능을 동시에 만족하게 되고 이 에너지 함수를 비용함수로 사용하게 된다.

전체 에너지의 성능 평가는 특징벡터 공간에서 모든 에러의 제곱승으로 나타내며 에너지 함수는 clustering 미터법으로 표현된다.

$$E_f(u_{ij}) = \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} (u_{ij} - \hat{\mu}_i)^T (u_{ij} - \hat{\mu}_i) + EDFD(\hat{M}_i) \quad (13)$$

$E(\hat{M}_i, \hat{S}_i, f_{i-1}), E(\hat{S}_i | f_{i-1})$ 은 (13) 식의 첫 번째 항으로 예측 중심 벡터 $\hat{\mu}_i$ 와 블록 특징 벡터 u_{ij} 의 거리로 표현할 수 있으며 $E(f_i | f_{i-1}, \hat{M}_i, \hat{S}_i)$ 는 EDFD (Estimate Displaced Frame Difference) 로 나타낼 수 있다. 식 (14)에서 S_x, S_y 는 search 영역을 말한다.

$$EDFD(\hat{M}) = \left[\frac{d_{ij}(\hat{M})^2}{\sum_{k=-\frac{S_x}{2}}^{\frac{S_x}{2}} \sum_{m=-\frac{S_y}{2}}^{\frac{S_y}{2}} d_{i+k, j+m}(\hat{M})} \right]^2 \quad (14)$$

III. MRBF 신경회로망

단락II 에서 구한 블록 특징벡터와 분할 기준 비용함수를 이용해 신경회로망의 매개변수를 구하고 이를 이용해 특징 입력 벡터가 신경망 회로망에 입력 될 때 해당 객체의 특정 출력 뉴런이 활성화 되도록 전체적인 시스템을 구축 한다.

1. 은닉층(Hidden unit) 매개변수

비용함수 (13)을 이용해 이미지의 영역분할을 하게 되는데 Covariance 매트릭스 C^{-1} 를 고려한 에너지 함수를 적용해서 Radial basis 함수를 구하게 된다.

$$\phi_j(u_{ij}) = \exp \left[- (u_{ij} - \hat{\mu}_j)^T C^{-1} (u_{ij} - \hat{\mu}_j) - EDFD(\hat{M}_i) \right] \quad (15)$$

그림3 의 출력 뉴런은 (0,1) 사이의 단극 Sigmoid 함수이며 다음과 같이 표현 할 수 있다.

$$Y_k(u_{ij}) = \frac{1}{1 + \exp \left[- \sum_{j=1}^L \lambda_{kj} \phi_j(u_{ij}) \right]} \quad (16)$$

여기서 λ_{kj} 는 은닉 뉴런 k에서 출력 j로의 가중치이다. 은닉층 매개변수는 첫 번째 학습 단계에서 LVQ 알고리즘을 사용하며 식(17) 에 따라 우세 클래스(Winner class) μ_j 의 중심벡터와 표준편차를 갱신한다.

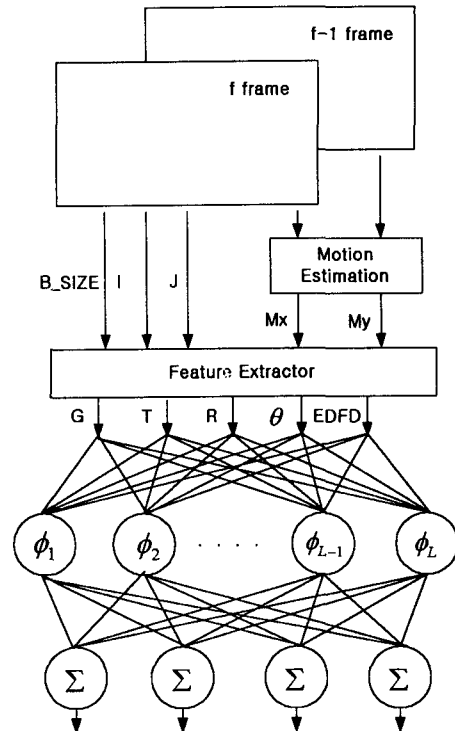


그림 3. 객체 분할을 위한 MRBF 신경회로망 구조

$$\min_{k=1}^L E_k(u_{ij}) = E_j(u_{ij}) \quad (17)$$

노이즈의 강인한 학습을 위해 MRBF 학습방법을 사용하며 클래스 해당 벡터를 순서대로 나열하고 그 중간 (Median) 값을 취해 중심값과 표준편차를 구한다.

$$\hat{\mu}_k = \text{median}(u_0, u_1, u_2, \dots, u_{p-1}) \quad (18)$$

$$\hat{C}_k = \frac{\text{median}(|u_0 - \hat{\mu}_k|, \dots, |u_{p-1} - \hat{\mu}_k|)}{0.6745} \quad (19)$$

\hat{C}_k 는 공분산 매트릭스의 대각 성분이고 0.6745 는 일정 분포를 위한 스케일 상수이다.

2. 출력층(Output unit) 매개변수

출력 뉴런은 각 이미지 블록을 객체 분할된 클래스로 할당해 주는 역할을 하게 된다.

$$\text{IF } \phi_k(u_{ij}) = \max \phi_j(u_{ij}) \text{ then } B_{ij} \in X_k \quad (20)$$

식(20) 에 의해 블록의 클래스가 결정되고

$$\text{IF } B_{ij} \in X_k \text{ then } F_k(u_{ij}) = 1, F_l(u_{ij}) = 0 \text{ for } l = 1, \dots, N, \quad l \neq k \quad (21)$$

출력 가중치를 구하기 위해 역전파 학습 방법을 통해 λ_{kj} 를 구하게 된다.

$$\lambda_{kj} = \sum_{i=0}^K \sum_{j=0}^K [F_k(\mathbf{u}_{ij}) - Y_k(\mathbf{u}_{ij})] Y_k(\mathbf{u}_{ij}) \times [1 - Y_k(\mathbf{u}_{ij})] \phi_k(\mathbf{u}_{ij}) \quad (22)$$

$F_k(\mathbf{u}_{ij})$ 는 출력 뉴런의 예상 값이고 $Y_k(\mathbf{u}_{ij})$ 는 식 (16)의 출력 Sigmoid 함수 결과 값이고 $\phi_k(\mathbf{u}_{ij})$ 는 식 (15)의 Radial basis 함수이다.

두 단계의 학습이 끝난 후 입력 특징벡터를 신경회로망에 넣으면 출력이 활성화되는 뉴런으로 객체의 클래스가 정해진다.

IV. 실험 및 결과

실험에 사용된 영상은 Table tennis 이며 이미지 크기는 352x240 이다. 전체 이미지를 4*4 블록으로 나누고 6번과 7번 프레임을 이용해서 전 영역 탐색(Full search) 알고리즘으로 움직임 벡터를 구하고 7번 프레임의 공간상의 특징벡터(평균 밝기, 질감정보)를 추출한다. 단락 III 의 과정을 통해 대표 중심 벡터의 값이 일정한 값으로 수렴 할때까지 학습을 시킨후 신경 회로망의 매개 변수로 사용하였다.

각 블록의 특징 벡터를 학습이 종료된 신경 회로망 입력에 입력 시킴으로써 객체분류된 결과를 얻었다.



그림4 table tennis 원그림에 모션 벡터를 표시한 그림



그림 5 신경망 출력의 블록별 모션 예측 그림(좌)
신경망 출력의 블록별 객체 분할된 그림(우)

테스트 이미지의 경우 6개의 객체 (1) 배경, (2) 테이블, (3) 팔, (4) 공, (5) 라켓, (6) 손 순으로 나누어지며 그림5 (우)에서 보듯이 6개의 영역으로 자동 분류됨을

알 수 있었다. 제안한 분류기의 성능 평가를 위해 원 그림을 수동 분류를 통해 객체 분리하고 제안한 분류기를 통과한 객체 분류 결과를 Classification 에러의 백분율로 표시하였다.

| 배경 | 테이블 | 팔 | 공 | 라켓 | 손 | Total |
|------|-------|-------|------|-------|-------|-------|
| 1.78 | 14.44 | 18.80 | 9.52 | 48.52 | 13.95 | 5.32 |

표 1. 객체별 분류 에러표 (단위: %)

표 1의 결과에서 라켓과 같은 움직임 벡터가 일정하지 않고 클래스의 갯수가 적은 곳 일수록 분류 에러가 많음을 알 수 있다.

결론적으로 그림 4와 그림 5의 움직임 벡터를 비교 했을때 움직임 벡터가 평탄해졌으며 질감정보를 이용함으로써 평균 배경색과 움직임 벡터가 비슷한 이미지에 서도 객체 분류가 가능함을 알 수 있었다.

참고문헌

- [1] D.W. Murray and B.F. Buxton, "Scene segmentation from visual motion using global optimization," IEEE trans. Pattern anal. machine intell. volume PAMI-9, pages 220-228, 1987.
- [2] J.Konrad and E. Dubis, "Bayesian estimation of motion vector fields", IEEE trans. Pattern anal. machine intell. volume 14, pages 910-927, sep, 1992.
- [3] F.Heitz and P. Bouthemy, "Multimodal estimation of discontinuous optical flow using Markov random fields," IEEE trans. Pattern anal. machine intell. volume 15, pages 1217-1232, dec, 1993.
- [4] C.Kotropoulos, N.Nikolaidisucel "Robust and adaptive techniques in self-organizing neural networks", Intern. J Computer Math., vol. 67, pages 183-200, 1998.
- [5] A. G. Bors and I. Pitas, "Median radial basis function neural network," IEEE trans. Neural Networks, volume 7, pages 1351- 1364, July, 1996.
- [6] A. G. Bors I. Pitas, "Optical Flow Estimation and Moving Object Segmentation Based On Median Radial Basis Function network", IEEE trans. Image Process, vol. 7, No 5, pp 693-702, May 1998.
- [7] Pitas et al, "order statics learning vector quantizer," IEEE trans. Image process. vol 5, pp 1048-1053, June, 1996.
- [8] Thomas Weldon, William E.Higgins, "Designing multiple Gabor filters for multi texture image segmentation", Opt. Eng, vol 38, pages 1478-1489, Sept. 1998.