

관심영역 기반 전역 움직임 보상을 이용한 ST-MRF 기반 추적기 고속화 방법

*박동민 **이동규 ***김상민 ****오승준

광운대학교 전자공학과

*freelyfly@media.kw.ac.kr

Fast ST-MRF based tracking using ROI-based GMC

*Park, Dong-Min **Lee, Dong-Kyu ***Kim, Sang-Min ****Oh, Seung-Jun

Department of Electronic Engineering at Kwangwoon University

요약

동영상에서의 객체 추적 알고리즘에 대한 활발한 연구가 진행되고 있음에도 불구하고 실시간 객체추적을 위해서는 여전히 정확도, 복잡도 등에서의 성능향상이 필요하다. 압축영역 기반 방식에서는 전역 움직임 보상(GMC : Global Motion Compensation)과정을 거쳐 추적하려는 객체와 배경을 구분한다. 전역 움직임 보상방법은 프레임 전 영역을 대상으로 하는 연산으로 전체 추적 시스템에서 차지하는 복잡도가 높다. 본 논문은 관심영역(ROI : Region Of Interest) 기반 전역 움직임 보상방법을 이용한 ST-MRF(Spatio-Temporal Markov Random Field)기반 추적기 고속화 방법을 제안한다. 관심영역을 기반으로 전역 움직임 보상을 적용함으로써 객체와 배경을 분리할 뿐만 아니라 알고리즘의 복잡도를 효과적으로 줄일 수 있다. 제안하는 방법의 추적성능은 평균 precision 87.29%, recall 82.58%, F-measure 83.78%로 기존방법과 비교하여 약 1%의 차이를 유지하였으며 전체 시스템의 수행시간은 평균 29.95ms로 기존방법과 비교하여 1.74배의 속도향상을 보였다.

1. 서론

동영상에서의 객체 추적은 비디오 감시, 색인 및 검색, 통신, 압축 등 여러 분야에서 중요한 문제 중 하나이다. 동영상에서의 객체 추적(VOT : Video Object Tracking) 방법은 크게 두 가지가 있다. 첫 번째로, 화소영역 기반 방식은 일반적으로 높은 정확도를 가지지만 그만큼 높은 복잡도를 요구한다. 오늘날 대부분의 비디오 콘텐츠는 압축된 비트스트림(bitstream) 형태로 존재하기 때문에 화소 정보를 추출하기 위해서는 복호화 과정이 필수적이다. 두 번째 방식은 압축영역 기반 방식이다. 압축영역 기반 방식은 압축된 비트스트림 상에 존재하는 움직임 벡터(MV : Motion Vector)나 부호화 모드, 잔여신호, 변환계수 등의 정보를 사용한다. 화소 정보를 사용하지 않기 때문에 화소영역 기반 방식보다는 정확도가 떨어지지만 화소 정보를 얻기 위한 복호화 과정을 생략할 수 있어 낮은 복잡도의 장점을 가진다. 따라서, 압축영역 기반 방식이 실시간 응용프로그램에 더 적합하다.

압축영역 기반 방식으로 전역 움직임 추정(GME : Global Motion Estimation)과 매크로블록 제외(macrobloc rejection) 방법을 이용하여 H.264/AVC 비트스트림 상에서 매크로블록 단위로 VOT를 수행하는 방법이 있다.[1] 하지만 이 방식은 배경 움직임에 비해 물체의 움직

임이 충분히 크지 않으면 추적하지 못하는 한계가 있다. Kas와 Nicolas는 움직이는 물체의 궤적을 H.254/AVC SVC의 움직임 벡터를 이용하여 추정하였다.[2] 우선 배경분리 기술을 통해 전경에 존재하는 물체를 분리하고, 시간적 필터링(temporal filtering)을 통해 잡음을 제거한다. Timed motion history images 방법을 통해 VOT가 수행된다. H.264/AVC 비트스트림의 움직임 벡터와 분할 크기(partition size) 정보를 mean shift clustering을 이용하여 객체를 분리하는 방법이 제안되었다.[3] 두드러진 움직임 벡터를 공간-시간적 중앙값 필터(median filter)와 전역 움직임 보상(GMC : Global Motion Compensation)을 적용하여 추출한 후, 공간적 mean shift를 통해 움직임이 균일한 지역을 찾아내고, 시간적 mean shift를 통해 그 지역을 부드럽게 해준다.

압축영역 기반 방식에서 MRF(Markov Random Field) 모델을 이용하여 VOT를 수행하는 연구들이 있다. Treetasanatavorn은 움직임 벡터를 깁스-MRF 이론(Gibbs-MRF theory)과 베이저안 추정(Bayesian estimation)을 이용하여 VOT를 수행하였다.[4] 첫 번째 프레임(frame)에 대한 객체 분리 작업은 확률적 움직임 일관성 모델(stochastic motion coherence model)[5]을 통해 이뤄지고, 연속적인 프레임들에 대해서는 어파인 모델(affine model)을 이용하여 이전 프레임에 존재하는 분할 정보를 현재 프레임에 사영(projection) 시켜 현

제 프레임에 대한 분할 정보를 예측한다. 그 후, 최종적인 분할 라벨(label)을 부여한다. Zeng은 MRF 분류(classification)를 적용하여 VOT를 수행하였다.[6] 움직임 벡터들은 배경 에지(edge), 전경, 잠음 중에서 한 형태로 분류된다. 그 후, 최대사후확률(MAP : maximum a posteriori probability)을 최대화 하는 방법으로 객체를 추출한다. 하지만 낮은 복잡도를 갖는 압축영역 기반 방식 역시 실시간 객체 추적을 위해서는 추가적인 고속화 연구가 필요하다.

본 논문에서는 압축영역 기반 방식에서 VOT를 위해 공간-시간적 MRF(ST-MRF : Spatial-Temporal MRF)모델을 이용한 VOT 알고리즘[7]에서 높은 복잡도를 차지하는 전역 움직임 보상 모듈을 관심영역 기반으로 수행함으로써 VOT 시스템을 고속화하는 방법을 제안한다. ST-MRF 모델은 물체의 움직임의 공간적 및 시간적 양상을 반영 해주며 이 알고리즘은 비트스트림에서 추출할 수 있는 움직임 벡터와 블록의 부호화 모드 정보만을 이용해 객체를 추적한다.

2장에서는 본 논문에서 제안하는 알고리즘을 자세하게 기술하고, 3장에서는 제안하는 알고리즘에 대한 실험 결과를 보이며 4장에서는 결론을 짓는다.

2. 제안하는 알고리즘

본 논문은 ST-MRF 모델을 사용하여 객체를 추적하는 시스템을 기반으로 한다.[7] 이 알고리즘은 비트스트림에 존재하는 움직임 벡터와 블록의 부호화 모드 정보를 이용하여 VOT를 수행한다. 첫 번째 프레임에서 사용자가 추적하고자 하는 물체를 직접 선택하면 이 알고리즘을 통해 연속된 프레임들에 대해서 자동으로 VOT가 이루어진다. 그림 1은 본 논문에서 제안하는 ST-MRF 모델 기반 VOT 시스템의 흐름도이다.

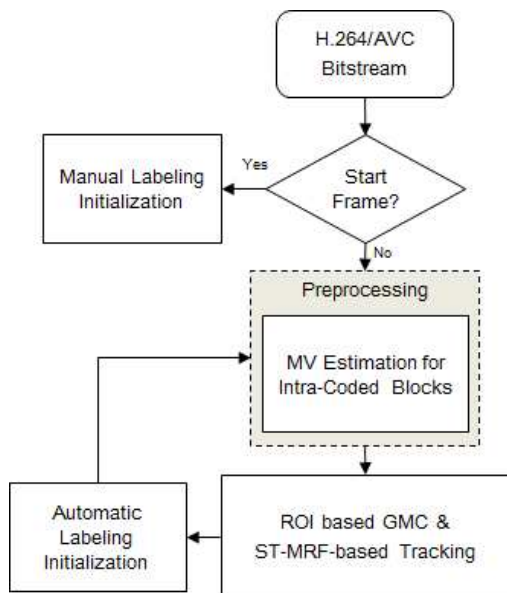


그림 1. ROI-based GMC가 접목된 ST-MRF 모델을 활용한 객체 추적 시스템의 흐름도

추적을 수행하기 전에 화면 내 예측 모드(intra mode)로 부호화된 블록들은 움직임 벡터를 포함하지 않으므로 주변 블록들의 움직임 벡터들을 이용하여 움직임 벡터를 추정한다.[7] 이와 같은 전처리과정을

거치고 나면 관심영역 기반 전역 움직임 보상 알고리즘을 통해 관심영역 내 배경의 움직임 벡터를 제거시킨 후 ST-MRF 추적 알고리즘을 객체를 추적한다.

2.1 ROI-GMC 기반 ST-MRF를 이용한 객체 추적 알고리즘

추적 알고리즘에서 통상의 전역 움직임 보상방법은 배경과 물체의 움직임을 분리하기 위한 알고리즘으로 프레임 내의 전 영역을 기반으로 전역 움직임 예측 및 보상을 수행한다. 그러나 ST-MRF 모델 기반 추적기에서는 객체가 존재하는 관심영역 내에서만 ST-MRF 모델을 적용하여 객체의 추적이 이루어지기 때문에 관심영역 내에서만 전역 움직임 보상을 수행하여도 객체 추적성능에 큰 영향을 미치지 않으면서 알고리즘의 고속화가 가능하다. 이 때, 관심영역은 이전 추적 결과로부터 예측된 현재 객체의 경계로부터 프레임 크기의 20% 만큼 더 넓게 설정하여 객체 주변의 배경을 충분히 포함시킴으로써 관심영역 내에서도 전역 움직임 보상이 잘 이루어지도록 하였으며, 20%는 실험을 통해 고속화 대비 추적성능의 안정성을 기반으로 설정하였다.

전역 움직임 보상을 위한 전역 움직임 추정 방법은 6개의 파라미터를 사용하는 어파인 모델을 사용한다.[8] 주어진 전역 움직임 파라미터(GMP : Global Motion Parameter) $\mathbf{m} = [m_1, \dots, m_6]$ 에 의해 (x, y) 에 위치하던 블록은 (x', y') 위치로 이동하게 되고, 움직임의 정도는 움직임 벡터 $\mathbf{v}(x, y)$ 로 표시된다.

$$x' = m_1 + m_2 + m_3y, y' = m_4 + m_5 + m_6y, \mathbf{v}(x, y) = (x' - x, y' - y). \quad (1)$$

\mathbf{m} 을 추정하기 위해 M-Estimator를 사용한다.[9] 이 알고리즘은 \mathbf{m} 의 예측오차를 기반으로 각각의 움직임 벡터에 대한 가중치를 갱신한다. 이 과정에서 outlier는 가중치가 작아져 제외된다. Arvanitidou에 따르면 8x8 이상의 큰 블록들은 배경에 속할 확률이 높지만 작은 블록들은 움직이는 물체에 속할 확률이 높다.[10] 따라서, 본 연구에서는 8x8~16x16 크기를 가지는 블록들을 전역 움직임 추정에 사용하고, 나머지 작은 블록들은 제외한다. 또한, 이전 프레임에서 움직이는 물체로 판명된 블록들은 전역 움직임과 상이한 움직임을 보일 확률이 높기 때문에 전역 움직임 추정과정에서 제외시킨다.

MRF 모델을 기반으로 하는 방법들은 대부분 공간적 MRF 모델(2-D MRF)을 적용하여 공간적인 특성만을 반영한다. 하지만 본 연구에서 사용하는 ST-MRF 모델은 공간적 특성뿐만 아니라 시간적 특성까지 고려하여 한 차원 높은 3-D MRF 모델을 사용한다.

$$\omega' = \arg \min_{\omega \in \Omega} \left\{ \frac{1}{\lambda_T} \cdot \sum_{\mathbf{n}; \psi(\mathbf{n})=1} \epsilon_T(\mathbf{n}; \omega^{t-1}, \kappa^t) + \frac{1}{\lambda_A} \cdot \sum_{\mathbf{n}; \psi(\mathbf{n})=1} \epsilon_A(\mathbf{n}; \kappa^t) \right. \\ \left. + \frac{1}{\lambda_\Phi} \cdot \sum_{\mathbf{n}; \psi(\mathbf{n})=1} \epsilon_\Phi(\mathbf{n}) \right\}. \quad (2)$$

수식 (2)에서 ω^{t-1} 은 $t-1$ 번째 프레임의 블록 라벨(0, 1)을 나타내며 $\kappa^t = \{\mathbf{v}^t, o^t\}$ 로 \mathbf{v}^t 는 움직임 벡터, o^t 는 블록의 부호화 모드를 나타낸다. 또한 세 에너지 함수 $\epsilon_T, \epsilon_\Phi, \epsilon_A$ 는 각각 시간적 불연속성(temporal discontinuity), 움직임 비일관성(motion incoherence), 물체의 산개성(incompactness)을 나타낸다. $\lambda_T, \lambda_A, \lambda_\Phi$ 는 계수 인자(scaling factor)이다. 수식 (2)의 해를 구하기 앞서 이

전 프레임에서 객체로 판정된 블록을 가리키는 움직임벡터를 갖는 각각의 블록들로 현재 물체의 위치를 초기화하고, ST-MRF 모델에서 정의한 MAP criterion을 기반으로 Iterated Conditional Modes (ICM) 알고리즘을 통해 움직이는 물체의 정확한 위치가 예측된다.[11]

3. 실험결과

제안하는 알고리즘의 성능평가를 위해 3.4GHz 클럭을 갖는 Intel i7-4770 프로세서에서 실험을 진행하였다. 추적기 알고리즘은 Windows8 운영체제에서 MATLAB을 이용하여 구현되었다. 실험에는 H.264/AVC에서 사용되는 YUV 4:2:0 포맷의 CIF(352x288)와 SIF(352x240) 시퀀스들이 사용되었으며 H.264/AVC JM v10.1 부호화기를 통해 부호화되었다. QP는 28로 고정하였으며 IPPP형태의 GOP 구조를 지녔고 참조픽처의 수는 1장으로 제한하였다. 또한 움직임 벡터는 4x4 블록 단위로 제한하였다.

표 1. 기존 방법과 제안하는 방법의 수행시간 비교

		기존 방법 [7](ms)	제안하는 방법(ms)	Speed up
City (CIF)	Pre	0.12	0.12	1
	GMC	51.34	37.5	1.37
	MRF	6.62	6.37	1.04
	Total	58.07	43.98	1.32
Garden (SIF)	Pre	0.16	0.16	1
	GMC	40.59	23	1.76
	MRF	5.16	5.1	1.01
	Total	45.91	28.26	1.62
TableTennis (SIF)	Pre	0.11	0.12	0.92
	GMC	39.13	29.76	1.31
	MRF	5.33	5.36	0.99
	Total	44.56	35.24	1.26
Stefan (CIF)	Pre	0.41	0.4	1.03
	GMC	50.83	22.12	2.3
	MRF	3.13	3.03	1.03
	Total	54.37	25.55	2.13
HallMonitor (CIF)	Pre	0.48	0.47	1.02
	GMC	46.71	21.5	2.17
	MRF	3.22	3.11	1.04
	Total	50.41	25.08	2.01
Mobile (CIF)	Pre	0.15	0.16	0.94
	GMC	52.5	11.83	4.44
	MRF	2.69	2.54	1.06
	Total	55.34	14.53	3.81
Coastguard (CIF)	Pre	0.22	0.22	1
	GMC	49.52	16.67	2.97
	MRF	2.61	2.56	1.02
	Total	52.35	19.45	2.69
Foreman (CIF)	Pre	0.6	0.56	1.07
	GMC	47.4	38.9	1.22
	MRF	8.21	8.01	1.02
	Total	56.21	47.47	1.18
Total 평균		52.15	29.95	1.74

표 1 에 기존 방법[7]과 제안하는 방법의 pre-processing (Pre), Global Motion Compensation(GMC), ST-MRF tracking(MRF)의 수행시간을 나타내었다. 이를 통해 기존 방법[7]에 비해 제안하는 방법이 평균 1.74배 속도가 향상되었음을 확인할 수 있다. Mobile의 경우 최대 3.81배 속도가 향상되었지만 Foreman의 경우 1.18배로 상대적으로 미비한 성능향상을 보였다. 이는 Foreman 시퀀스의 객체가 다른 시퀀스의 객체보다 상대적으로 커 GMC에 적용되는 관심영역이 크기 때문이다.

표 2. 기존 방법[7]과 제안하는 방법의 추적성능 비교

	기존 방법 [7](%)			제안하는 방법(%)		
	P	R	F	P	R	F
City	96.09	94.51	95.27	96.36	94.29	95.30
Garden	86.39	93.18	89.51	87.95	92.28	89.97
TableTennis	98.37	84.28	90.62	98.46	84.22	90.63
Stefan	89.52	62.72	72.42	90.75	62.20	72.52
HallMonitor	79.47	80.21	79.74	80.58	79.69	80.00
Mobile	82.16	82.78	81.99	85.85	72.61	77.89
Coastguard	63.78	89.30	73.88	64.00	88.13	73.36
Foreman	94.28	86.66	90.17	94.38	87.19	90.53
평균	86.26	84.21	84.20	87.29	82.58	83.78

표 2 에 기존방법[7]과 제안하는 방법의 precision(P), recall(R), F-measure(F) 수치를 나타내었다. 기존방법은 각각 평균 86.26%, 84.21%, 84.20%의 수치를 보였고 제안하는 방법은 각각 평균 87.29%, 82.58%, 83.78%의 수치를 보여 약 1%의 성능차가 남을 알 수 있다. 이를 통해 기존 방법[7]의 추적성능과 제안하는 방법의 추적성능이 유사하다는 점을 확인할 수 있다.

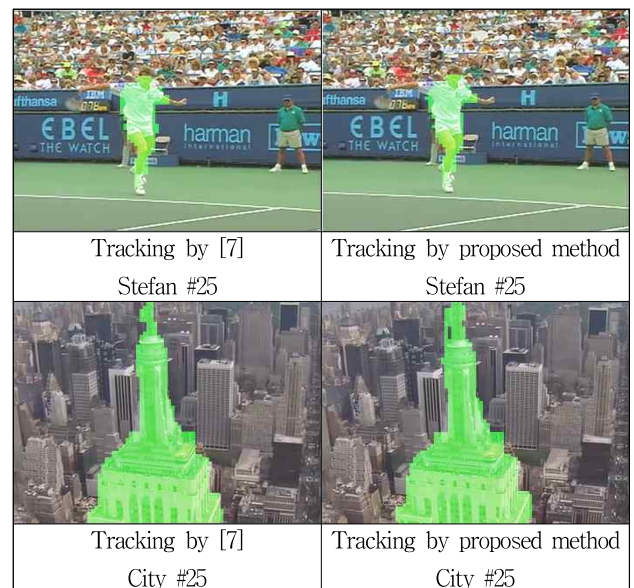


그림 2. 기존 방법[7]과 제안하는 방법의 주관적 추적성능 비교

그림 2 를 통해 주관적인 추적성능 역시 기존 방법[7]과 제안하는 방법이 큰 차이를 보이지 않고 있음을 확인할 수 있다.

4. 결론

본 논문에서는 관심영역 기반 전역 움직임 보상을 이용한 ST-MRF 기반 추적기 고속화 방법을 제안하였다. 공간-시간적 MRF(ST-MRF : Spatial-Temporal MRF)모델을 이용한 VOT 알고리즘은 움직임 벡터와 블록 부호화 모드를 이용하여 MAP 해법을 통해 객체의 추적이 이루어진다. 이때 카메라 움직임으로 발생하는 전역 움직임을 보상해주기 위해 관심영역 기반 전역 움직임 보상방법을 이용하여 물체의 순수한 움직임만을 반영하여 추적을 수행한다. 복잡도 비중이 높은 전역 움직임 보상과정을 관심영역을 기반으로 수행함으로써 전체 시스템의 복잡도를 효과적으로 줄일 수 있었다. 그 결과 H.264/AVC의 표준 시퀀스들 상에서의 추적성능은 평균 precision 87.29%, recall 82.58%, F-measure 83.78%로 기존방법과 비교하여 약 1%의 차이를 유지하였으며 전체 시스템 수행시간은 평균 29.95ms로 기존방법과 비교하여 1.74배 속도향상을 보였다.

감 사 의 글

본 연구는 미래창조과학부 및 정보통신기술진흥센터의 정보통신·방송 연구개발 사업의 일환으로 수행하였음.

[14-000-11-002, 방송용 영상 인식 기반 객체 중심 지식융합 미디어 서비스 플랫폼 개발]

참 고 문 헌

- [1] V. Mezaris, I. Kompatsiaris, N. V. Boulgouris, and M. G. Strintzis, "Real-time compressed-domain spatiotemporal segmentation and ontologies for video indexing and retrieval," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 5, pp. 606 - 621, May 2004.
- [2] C. Käs and H. Nicolas, "An approach to trajectory estimation of moving objects in the H.264 compressed domain," in *Proc. 3rd Pacific Rim Symp. Adv. Image Video Technol.*, pp. 318 - 329, 2009.
- [3] W. Fei and S. Zhu, "Mean shift clustering-based moving object segmentation in the H.264 compressed domain," *IET Image Process.*, vol. 4, no. 1, pp. 11 - 18, Feb. 2010.
- [4] S. Treetasanatavorn, U. Rauschenbach, J. Heuer, and A. Kaup, "Bayesian method for motion segmentation and tracking in compressed videos," in *Proc. 27th DAGM Conf. Pattern Recognit.*, pp. 277 - 284, 2005.
- [5] S. Treetasanatavorn, U. Rauschenbach, J. Heuer, and A. Kaup, "Stochastic motion coherency analysis for motion vector field segmentation on compressed video sequences," in *Proc. IEEE Workshop Image Anal. Multimedia Interact. Services*, pp. 1 - 4, Apr. 2005.
- [6] W. Zeng, J. Du, W. Gao, and Q. Huang, "Robust moving object

segmentation on H.264/AVC compressed video using the block-based MRF model," *Real-Time Imaging*, vol. 11, no. 4, pp. 290 - 299, Aug. 2005.

[7] Sayed Hossein Khatoonabadi and Ivan V. Bajić, "Video object tracking in the compressed domain using spatio-temporal Markov Random Fields," *IEEE Transactions on Image Processing*, vol. 22, no. 1, pp.300-313, Jan. 2013.

[8] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2004, pp. 39 - 44.

[9] M. G. Arvanitidou, A. Glantz, A. Krutz, T. Sikora, M. Mrak, and A. Kondoz, "Global motion estimation using variable block sizes and its application to object segmentation," in *Proc. IEEE Workshop Image Anal. Multimedia Interact. Services, London, U.K.*, May 2009, pp. 173 - 176.

[10] A. Smolic, M. Hoeynck, and J.-R. Ohm, "Low-complexity global motion estimation from P-frame motion vectors for MPEG-7 applications," in *Proc. IEEE Int. Conf. Image Process.*, vol. 2. Vancouver, BC, Canada, Sep. 2000, pp. 271 - 274.

[11] J. Besag, "On the statistical analysis of dirty pictures," *J. Royal Stat. Soc. B*, vol. 48, no. 3, pp. 259 - 302, 1986.