



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2017-0140655
(43) 공개일자 2017년12월21일

(51) 국제특허분류(Int. Cl.)
G06K 9/00 (2006.01) G06K 9/38 (2006.01)
(52) CPC특허분류
G06K 9/00597 (2013.01)
G06K 9/0061 (2013.01)
(21) 출원번호 10-2016-0073242
(22) 출원일자 2016년06월13일
심사청구일자 2016년06월13일

(71) 출원인
인하대학교 산학협력단
인천광역시 남구 인하로 100, 인하대학교 (용현동)
(72) 발명자
이필규
인천광역시 남구 매소홀로475번길 18, 17동 601호(학익동, 신동아아파트3차)
(74) 대리인
특허법인충정

전체 청구항 수 : 총 6 항

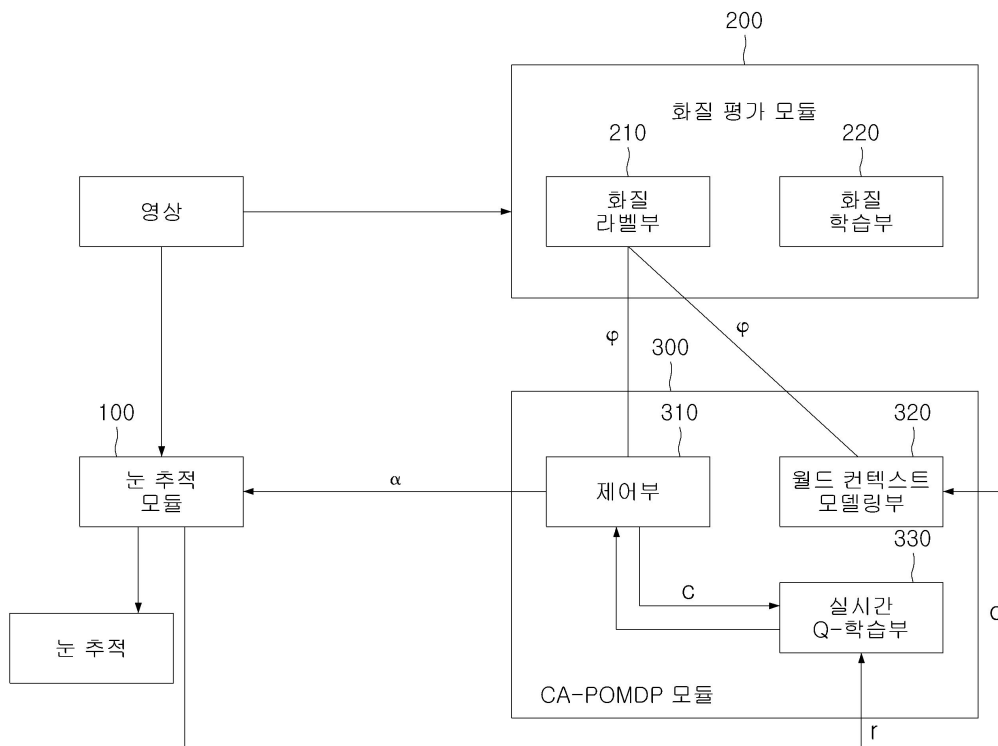
(54) 발명의 명칭 강건한 휴먼 컴퓨터 상호작용을 위한 CA-POMDP를 이용한 눈 추적 장치

(57) 요약

본 발명은 강건한 휴먼 컴퓨터 상호작용을 위한 CA-POMDP를 이용한 눈 추적 장치에 관한 것이다.

또한, 본 발명에 따르면, 얼굴 영상의 입력시에, 눈 영역을 추출한 후에, 이진화 문턱값으로 눈 영상을 이진화하며, 이진화된 눈 영상에서 눈동자 위치를 추적하여 눈 추적을 수행하고, 시스템 제어 파라미터의 조정 행위에 대(뒷면에 계속)

대표도



한 보상값을 출력하는 눈 추적 모듈; 얼굴 영상을 입력받아 화질을 평가하여 화질 라벨을 제공하는 화질 평가 모듈; 및 상기 시스템 제어 파라미터를 상태로 정의하고, 화질 라벨별로 다양한 객체들의 조합으로 월드 컨텍스트 모델을 구성하고, 상기 화질 평가 모듈에서 제공되는 화질 라벨에 따라 다양한 객체들의 속성과 속성 값들로 표현된 월드 컨텍스트 모델상의 정보를 기초로 현재 상태를 판단하고, 판단된 현재 상태를 근거로 실행 가능한 시스템 제어 파라미터 조정 행위를 수행하고, 상기 시스템 제어 파라미터 조정 행위의 보상값을 기초로 행위 함수값을 갱신하고 갱신된 행위 함수값에 따라 최적의 시스템 제어 파라미터 조정 행위를 선택하는 CA-POMDP 모듈을 포함하는 눈 추적 장치를 제공한다.

(52) CPC특허분류

G06K 9/38 (2013.01)

G06K 9/66 (2013.01)

G06T 2207/20081 (2013.01)

명세서

청구범위

청구항 1

얼굴 영상의 입력시에, 눈 영역을 추출한 후에, 이진화 문턱값으로 눈 영상을 이진화하며, 이진화된 눈 영상에서 눈동자 위치를 추적하여 눈 추적을 수행하고, 시스템 제어 파라미터의 조정 행위에 대한 보상값을 출력하는 눈 추적 모듈;

얼굴 영상을 입력받아 화질을 평가하여 화질 라벨을 제공하는 화질 평가 모듈; 및

상기 시스템 제어 파라미터를 상태로 정의하고, 화질 라벨별로 다양한 객체들의 조합으로 월드 컨텍스트 모델을 구성하고, 상기 화질 평가 모듈에서 제공되는 화질 라벨에 따라 다양한 객체들의 속성과 속성 값들로 표현된 월드 컨텍스트 모델상의 정보를 기초로 현재 상태를 판단하고, 판단된 현재 상태를 근거로 실행 가능한 시스템 제어 파라미터 조정 행위를 수행하고, 상기 시스템 제어 파라미터 조정 행위의 보상값을 기초로 행위 함수값을 갱신하고 갱신된 행위 함수값에 따라 최적의 시스템 제어 파라미터 조정 행위를 선택하는 CA-POMDP 모듈을 포함하는 눈 추적 장치.

청구항 2

청구항 1항에 있어서,

상기 시스템 제어 파라미터는 이진화 문턱값, 부분 허프 변환의 각도 파라미터 및 칼만 필터의 노이즈 파라미터를 포함하는 눈 추적 장치.

청구항 3

청구항 1항에 있어서,

상기 화질 평가 모듈은

수집된 훈련 영상을 사용하여 화질 인덱스 프로토타입으로부터 화질 인덱스를 학습하는 화질 학습부; 및

영상을 입력받아 관심 영역을 추출하고, 사각 패치로 분할한 후에, 상기 화질 학습부의 프로토타입을 로드하고, 스캐닝 윈도우의 픽셀을 추출하여 화질 라벨을 산출하는 화질 라벨부를 포함하는 눈 추적 장치.

청구항 4

청구항 3항에 있어서,

상기 화질 라벨부는 화질 라벨을 산출하는데 있어서 조명 변화를 고려하여 간략화된 화질 라벨을 산출하는 것을 특징으로 하는 눈 추적 장치.

청구항 5

청구항 1항에 있어서,

상기 CA-POMDP 모듈은

상기 시스템 제어 파라미터를 상태로 정의하고, 다양한 객체들의 속성과 속성 값들로 표현된 월드 컨텍스트 모델상의 정보를 기초로 각 상태 변수들의 현재 상태를 판단하며, 갱신된 행위 함수값에 근거한 실행 가능한 행위를 수행하는 제어부;

화질 라벨별로 다양한 객체들의 조합으로 월드 컨텍스트 모델을 구성하는 월드 컨텍스트 모델링부; 및

현재 상태에 대한 행위의 보상값을 기초로 행위 함수값을 갱신하고 갱신된 행위 함수값을 저장하는 실시간 Q-학습부를 포함하는 눈 추적 장치.

청구항 6

청구항 1항에 있어서,

상기 실시간 Q-학습부는 행위를 획득하고, 내부 보상을 계산하며, 즉각적인 보상을 근거로 새로운 내부 상태를 관찰하여 월드 컨텍스트 모델의 행위 함수값을 업데이트하는 것을 특징으로 하는 눈 추적 장치.

발명의 설명

기술 분야

[0001] 본 발명은 강건한 휴먼 컴퓨터 상호작용을 위한 CA-POMDP를 이용한 눈 추적 장치에 관한 것이다.

배경 기술

[0002] 최근, 많은 시선 추적(eye-tracking) 처리 방법들이 다양한 목표들과 어플리케이션들에 대해 제안되었다.

[0003] 자동 시선 추적은, 인간의 지각, 주의 및 인식의 상태와 강하게 연결되어 있기 때문에, 많은 어플리케이션 영역에서 이용되어 왔다.

[0004] 시선 추적 기술은, 상호작용을 위한 효과적인 툴로서, 방해받지 않고 손을 쓰지 않고도 이용할 수 있는 휴먼 컴퓨터 상호작용(human computer interaction, HCI) 및 컴퓨터와 사람간의 통신에 이용될 수 있다.

[0005] 그럼에도 불구하고, 기술 개발에 많은 노력과 조명, 시야각, 스케일 등 많은 문제가 존재하고, 개개인의 눈의 모양 및 지터링의 다양한 종류에 따른 동작 환경의 변경으로 문제가 존재한다.

[0006] 상업적인 영역에서 가장 성공적인 시선 추적 기술은 고 비용의 이미지 캡처 장치(예를 들어, 고 비용의 카메라 및 렌즈)를 요구하거나 또는 강력하게 제어된 상황 내에서 매우 제한된 동작이 요구된다.

[0007] 최근 시선 추적 기술에서 최대의 관심사는, 웹 유용성, 광고, 스마트 TV 및 모바일 어플리케이션을 위한 HCI 분야에서 생성되고 있다.

[0008] 그러나, 종래의 시선 추적 기술은 저 비용 산업 어플리케이션에서 이용하기에 높은 비용과 제한성이라는 문제점을 갖고 있다.

[0009] 조명의 제어와 상황의 제어 없이, 비용을 줄일 뿐만 아니라 복잡한 초기 사용자 설정을 간소화하여 더 이용 가능하고 더 일반적인 시선 추적 기술이 성공적인 HCI 어플리케이션을 위해 필요하다.

선행기술문헌

특허문헌

[0010] (특허문헌 0001) 공개특허번호 2015-002741

(특허문헌 0002) 등록번호 10-138775

비특허문헌

[0011] (비특허문헌 0001) 1. L. J. G. Vazquez, et. al.: Low Cost Human Computer Interface Voluntary Eye Movement as Communication System for Disabled People with Limited Movements. PAHCE, pp. 165-170. (2011)

(비특허문헌 0002) 2. J. S. Agustin, et. al.: Low-Cost Gaze Interaction: Ready to Deliver the Promises. CHI 2009 - Spotlight on Works in Progress - Session 2, Boston, MA, USA (2009)

(비특허문헌 0003) 3. A. Sein et. al.: Eyeing a Real-Time Human-Computer Interface to Assist Those with Motor Disabilities. IEEE Potentials, pp. 19-25 (2008)

(비특허문헌 0004) 4. R. S. Sutton, A. G. Barto: Reinforcement Learning: An Introduction. Cambridge, MA: MIT Press, (1998)

(비특허문헌 0005) 5. S. Ross et. al.: A Bayesian Approach for Learning and Planning in Partially

Observable Markov Decision Processes. Journal of Machine Learning Research 12, pp.1729-1770. (2011)

(비특허문헌 0006) 6. G. Shani: A survey of point-based POMDP solvers. Auton Agent Multi-Agent Syst, Published online, Springer (2012)

(비특허문헌 0007) 7. A. Atrash et. al.: Development and validation of a robust speech interface for improved human-robot interaction. International Journal of Social Robotics, 1, pp.345-356. (2009)

(비특허문헌 0008) 8. G. Shani, et. al.: An MDP-based recommender system. Journal of Machine Learning Research, 6, pp.1265-1295. (2005)

(비특허문헌 0009) 9. M. Hauskrecht, H. S. F. Fraser: Planning treatment of ischemic heart disease with partially observable Markov decision processes," Artificial Intelligence in Medicine, 18(3), pp. 221-244. (2000)

(비특허문헌 0010) 10. J. Hoey, et. al.: Automated hand washing assistance for persons with dementia using video and a partially observable Markov decision process. Computer Vision and Image Understanding, 114(5), pp. 503-519. (2010)

(비특허문헌 0011) 11. J. D. Williams, S. Young: Partially observable Markov decision processes for spoken dialog systems. Computer Speech & Language, 21(2), pp. 393-422. (2007)

(비특허문헌 0012) 12. L. P. Kaelbling et. al.: Planning and acting in partially observable stochastic domains. Artificial Intelligence, 101, pp. 99-134. (1998)

(비특허문헌 0013) 13. J. Pineau, et. al., "Point-based value iteration: An anytime algorithm for POMDPs," In International joint conference on artificial intelligence, pp. 1025-1032. (2003)

(비특허문헌 0014) 14. T. Smith and R. G. Simmons, "Point-based POMDP algorithms: Improved analysis and implementation," In Conference on uncertainty in artificial intelligence (UAI), pp. 542-547. (2005)

(비특허문헌 0015) 15. S. Paquet: Distributed Decision-Making and Task Coordination in Dynamic, Uncertain and Real-Time Multiagent Environments. Ph.D. thesis, Laval University (2006)

(비특허문헌 0016) 16. S. Ross, et. al.: Online Planning Algorithms for POMDPs. Journal of Artificial Intelligence Research, 32, pp. 663-704. (2008)

(비특허문헌 0017) 17. M. L. Littman et. al.: Learning policies for partially observable environments: scaling up. International Conference in Machine Learning (1995)

(비특허문헌 0018) 18. G. Shani, et. al.: Forward search value iteration for POMDPs. In International joint conference on artificial intelligence, (IJCAI), (2007)

(비특허문헌 0019) 19. T. Smith, and R. Simmons, "Heuristic search value iteration for POMDPs," In Conference on uncertainty in artificial intelligence (UAI), 2003.

(비특허문헌 0020) 20. H. Kurniawati, et. al. "SARSOP: Efficient Point-Based POMDP Planning by Approximating Optimally Reachable Belief Spaces," Proc. Robotics: Science & Systems, 2008.

(비특허문헌 0021) 21. A. Barto, et. al.: Learning to act using realtime dynamic programming. AIJ, Vol. 72, pp. 81-138. (1995)

(비특허문헌 0022) 22. S. Paquet, et. al., "An online POMDP algorithm for complex multiagent environments, AAMAS (2005)

(비특허문헌 0023) 23. J. Wang, et. al., "A Framework for Moving Target Detection, Recognition and Tracking in UAV Videos," Affective Computing and Intelligent Interaction, ed., Jia Luo, volume 137 of Advances in Intelligent and Soft Computing, Springer Berlin / Heidelberg, 2012, pp. 69-76.

(비특허문헌 0024) 24. B. Bonet and H. Geffner, "Solving POMDPs: RTDP-bel vs. point based algorithms," Proceedings of the 21st international joint conference on Artificial intelligence, IJCAI, 2009, p.

1641-1646.

(비특허문헌 0025) 25. S. Ross and B. Chaib-draa et. al., "Aems: An anytime online search algorithm for approximate policy refinement in large POMDPs," Proceedings of the 20th International Joint Conference on Artificial Intelligence, 2007, pp. 2592-2598.

(비특허문헌 0026) 26. C. A. P. Chaneil, "POMDP-based online target detection and recognition for autonomous UAVs," ECAI, 2012, pp. 955-960.

(비특허문헌 0027) 27. S. Paquet, et. al., "Hybrid POMDP algorithms," Proceedings of The Workshop on Multi-Agent Sequential Decision Making in Uncertain Domains, 2006.

(비특허문헌 0028) 28. S. Ross, et. al., "Theoretical analysis of heuristic search methods for online POMDPs," Advances in Neural Information Processing Systems, 2008.

(비특허문헌 0029) 29. H. Geffner and B. Bonet, "Solving large POMDPs using real time dynamic Programming," in Working Notes Fall AAAI Symposium on POMDPs, 1998, pp. 61-68.

(비특허문헌 0030) 30. R. Washington, "BI-POMDP: Bounded, Incremental Partially-Observable Markov-Model Planning," Proceedings of the 4th European Conference on Planning, volume 1348 of Lecture Notes in Computer Science, Toulouse, France, Springer, 1997, pp. 440-451.

(비특허문헌 0031) 31. T. Wang, et. al., "A Bayesian sparse sampling for on-line reward optimization," ICML'05: Proceedings of the 22nd international conference on Machine Learning, New York, NY, USA: ACM, 2005, pp. 956-963.

(비특허문헌 0032) 32. J. Asmuth, et. al., "A Bayesian sampling approach to exploration in reinforcement learning," Conference on Uncertainty in Artificial Intelligence (2009)

(비특허문헌 0033) 33. F. Doshi-Velez, et. al., "Reinforcement learning with limited reinforcement: using Bayes risk for active learning in POMDPs," Artificial Intelligence, 2012, pp. 187-188.

(비특허문헌 0034) 34. M. Strens, "A Bayesian framework for reinforcement learning," International Conference in Machine Learning, 2000.

(비특허문헌 0035) 35. R. Jaulmes, "Active learning in partially observable Markov Decision Processes," ECML, 2005.

(비특허문헌 0036) 36. S. Ross, et. al., "Bayes-adaptive POMDPs. In Neural Information Processing Systems," 2008.

(비특허문헌 0037) 37. F. Doshi-Velez, "The infinite partially observable Markov decision process," Advances in Neural Information Processing Systems, 22, pp. 477-485 (2009)

(비특허문헌 0038) 38. P. Poupart and N. Vlassis, "Model-based Bayesian reinforcement learning in partially observable domains," ISAIM, 2008.

(비특허문헌 0039) 39. A. Krause, et. al., "Trading off prediction accuracy and power consumption for context-aware wearable computing," Proc. 9th IEEE Int. Symp. Wearable Computers, Oct. 2005, pp. 20-26.

(비특허문헌 0040) 40. L. Au, et. al, "Episodic sampling: Towards energy-efficient patient monitoring with wearable sensors," Proc. IEEE Annu. Int. Conf. Engineering in Medicine and Biology Society, Sep. 2009, pp. 6901-6905.

(비특허문헌 0041) 41. B. Bhanu and J. Peng, "Adaptive integrated image segmentation and object recognition," IEEE Trans. on Systems, Man, and Cybernetics-PART C: Applications and Reviews, Vol. 30, No. 4, Nov. 2000, pp. 427-441.

(비특허문헌 0042) 42. R. D. Smallwood and E. J. Sondik. "The optimal control of partially observable Markov processes over a finite horizon," Operations Research, 21(5): Sep/Oct, 1973, pp. 1071-1088.

(비특허문헌 0043) 43. K. Murphy, "Dynamic Bayesian Networks: Representation, Inference and Learning,"

PhD thesis, UC Berkeley, 2002.

(비특허문헌 0044) 44. M. Abdel-Mottaleb and M. H. Mahoor, "Application notes - Algorithms for assessing the quality of facial images", IEEE Computational Intelligence Magazine, Volume 2, Issue2, May 2007, pp. 10 - 17.

(비특허문헌 0045) 45. H. Sellaheewa and S. A. Jassim, "Image-quality-based adaptive face recognition," IEEE Trans. on Instrumentation and Measurement, Vol. 59, No. 4, April, 2010, pp. 805-813.

(비특허문헌 0046) 46. Q. Li and Z. Wang, "Reduced-reference image quality assessment using divisive normalization-based image representation," IEEE Journal of Selected Topic in Signal Processing, Vol. 3, No. 2, Apr. 2009, pp. 202-211.

(비특허문헌 0047) 47. C. J. C. H. Watkins, Learning from delayed rewards, Ph.D. Thesis, Cambridge University, Cambridge, England, 1989.

(비특허문헌 0048) 48. Z., Wang, A.C., Bovik: A universal image quality index. IEEE Signal Process. Lett., Vol. 9, No. 3, pp. 81-84. (2002)

(비특허문헌 0049) 49. P.K., Rhee, et. al.: Context-aware evolvable system framework for environment identifying systems. Part II, pp. 270-283. KES (2005)

(비특허문헌 0050) 50. E.J., Koh, M.Y., Nam, P.K., Rhee: A context-driven Bayesian classification method for eye location. ICANNGA, 2, pp.517-524 (2007)

(비특허문헌 0051) 51. R. Dearden, N. Friedman, D. Andre: Model based Bayesian exploration. Proceedings UAI (1999)

(비특허문헌 0052) 52. P. K. Rhee, et. al.: Pupil location and movement measurement for efficient emotional sensibility Analysis. ISSPIT (2011)

(비특허문헌 0053) 53. Z. Kadal, et. al.: Tracking-Learning-Detection. IEEE Trans on PAMI, Vol. 34, No. 7, pp. 1409-1422. (2012)

(비특허문헌 0054) 54. Y. Shen, et al.: Evolutionary adaptive eye tracking for low-cost human computer interaction applications. Journal of Electronic Imaging, 22.1, 013031-013031 (2013)

발명의 내용

해결하려는 과제

[0012] 본 발명은 상기와 같은 문제점을 해결하기 위하여, 고비용 이미지 캡처 기기나 매우 제한적인 상황을 이용하는 대신, CA-POMDP를 사용하여 시스템 제어 파라미터를 최적화하여 성능을 보장할 수 있는 강건한 휴먼 컴퓨터 상호작용을 위한 CA-POMDP를 이용한 눈 추적 장치를 제공하는 데 있다.

과제의 해결 수단

[0013] 본 발명의 일 측면은 얼굴 영상의 입력시에, 눈 영역을 추출한 후에, 이진화 문턱값으로 눈 영상을 이진화하며, 이진화된 눈 영상에서 눈동자 위치를 추적하여 눈 추적을 수행하고, 시스템 제어 파라미터의 조정 행위에 대한 보상값을 출력하는 눈 추적 모듈; 얼굴 영상을 입력받아 화질을 평가하여 화질 라벨을 제공하는 화질 평가 모듈; 및 상기 시스템 제어 파라미터를 상태로 정의하고, 화질 라벨별로 다양한 객체들의 조합으로 월드 컨텍스트 모델을 구성하고, 상기 화질 평가 모듈에서 제공되는 화질 라벨에 따라 다양한 객체들의 속성과 속성 값들로 표현된 월드 컨텍스트 모델상의 정보를 기초로 현재 상태를 판단하고, 판단된 현재 상태를 근거로 실행 가능한 시스템 제어 파라미터 조정 행위를 수행하고, 상기 시스템 제어 파라미터 조정 행위의 보상값을 기초로 행위 함수값을 갱신하고 갱신된 행위 함수값에 따라 최적의 시스템 제어 파라미터 조정 행위를 선택하는 CA-POMDP 모듈을 포함한다.

[0014] 또한, 본 발명의 일 측면의 상기 시스템 제어 파라미터는 이진화 문턱값, 부분 히프 변환의 각도 파라미터 및

칼만 필터의 노이즈 파라미터를 포함한다.

- [0015] 또한, 본 발명의 일 측면의 상기 화질 평가 모듈은 수집된 훈련 영상을 사용하여 화질 인덱스 프로토타입으로부터 화질 인덱스를 학습하는 화질 학습부; 및 영상을 입력받아 관심 영역을 추출하고, 사각 패치로 분할한 후에, 상기 화질 학습부의 프로토타입을 로드하고, 스캐닝 윈도우의 픽셀을 추출하여 화질 라벨을 산출하는 화질 라벨부를 포함한다.
- [0016] 또한, 본 발명의 일 측면의 상기 화질 라벨부는 화질 라벨을 산출하는데 있어서 조명 변화를 고려하여 간략화된 화질 라벨을 산출하는 것을 특징으로 한다.
- [0017] 또한, 본 발명의 일 측면의 상기 CA-POMDP 모듈은 상기 시스템 제어 파라미터를 상태로 정의하고, 다양한 객체들의 속성과 속성 값들로 표현된 월드 컨텍스트 모델상의 정보를 기초로 각 상태 변수들의 현재 상태를 판단하며, 갱신된 행위 함수값에 근거한 실행 가능한 행위를 수행하는 제어부; 화질 라벨별로 다양한 객체들의 조합으로 월드 컨텍스트 모델을 구성하는 월드 컨텍스트 모델링부; 및 현재 상태에 대한 행위의 보상값을 기초로 행위 함수값을 갱신하고 갱신된 행위 함수값을 저장하는 실시간 Q-학습부를 포함한다.
- [0018] 또한, 본 발명의 일 측면의 상기 실시간 Q-학습부는 행위를 획득하고, 내부 보상을 계산하며, 즉각적인 보상을 근거로 새로운 내부 상태를 관찰하여 월드 컨텍스트 모델의 행위 함수값을 업데이트하는 것을 특징으로 한다.

발명의 효과

- [0019] 본 발명은 고비용 이미지 캡처 기기나 매우 제한적인 상황을 이용하는 대신, CA-POMDP를 사용하여 시스템 제어 파라미터를 최적화하여 성능을 보장할 수 있다.

도면의 간단한 설명

- [0020] 도 1은 본 발명의 일실시예에 따른 강건한 휴먼 컴퓨터 상호작용을 위한 POMDP를 이용한 눈 추적 장치의 구성도이다.
- 도 2는 눈이 떠지는 정도에 따라 홍채의 외부 경계를 나타내는 부분 원들을 표시한 도면이다.
- 도 3은 화질 라벨부에서 화질을 라벨링하기 위한 과정을 설명하기 위한 도면이다.
- 도 4를 참조하면, 일반적인 POMDP와 본 발명의 CA-POMDP를 비교해서 보여준다.
- 도 5는 도 1의 눈 추적 장치의 동작을 설명하기 위한 흐름도이다.

발명을 실시하기 위한 구체적인 내용

- [0021] 본 발명과 본 발명의 동작상의 이점 및 본 발명의 실시예에 의하여 달성되는 목적을 설명하기 위하여 이하에서는 본 발명의 바람직한 실시예를 예시하고 이를 참조하여 살펴본다.
- [0022] 먼저, 본 출원에서 사용한 용어는 단지 특정한 실시예를 설명하기 위해 사용된 것으로서, 본 발명을 한정하려는 의도가 아니며, 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함할 수 있다. 또한 본 출원에서, "포함하다" 또는 "가지다" 등의 용어는 명세서 상에 기재된 특징, 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.
- [0023] 본 발명을 설명함에 있어서, 관련된 공지 구성 또는 기능에 대한 구체적인 설명이 본 발명의 요지를 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명은 생략한다.
- [0024] 도 1은 본 발명의 일실시예에 따른 강건한 휴먼 컴퓨터 상호작용을 위한 POMDP를 이용한 눈 추적 장치의 구성도이다.
- [0025] 도 1을 참조하면, 본 발명의 일실시예에 따른 강건한 휴먼 컴퓨터 상호작용을 위한 POMDP를 이용한 눈 추적 장치는 눈추적 모듈(100), 화질 평가 모듈(200) 및 CA-POMDP 모듈(300)을 포함한다.
- [0026] 상기 눈추적 모듈(100)은 얼굴 영상의 입력시에, 눈 영역을 추출한 후에, 이진화 문턱값으로 눈 영상을 이진화하며, 이진화된 눈 영상에서 눈동자 위치를 추적하여 눈 추적을 수행한다.

- [0027] 이와 같은 눈추적 모듈(100)에서 수행되는 눈 추적 방법의 성능은 각 단계들에서 알고리즘의 선택 그리고 선택된 알고리즘과 관련된 문턱값들과 파라미터들(이를 통칭하여 시스템 제어 파라미터라고 한다)에 매우 의존적이다.
- [0028] 상기 눈 추적 방법은, 눈 추적 알고리즘 구조 및, 알고리즘 구조와 관련된 문턱값들과 파라미터들을 결정하는 것의 중대한 동적 제어 문제에서 가장 잘 체계화될 수 있다.
- [0029] 이와 같은 눈 추적 방법이 성능이 저하되는 주된 이유가, 조명, 포즈 및 이미지의 모양변화로 인한 이미지 질의 변화임을 고려하면 상기 문턱값들과 파라미터들은 환경의 변화에 많은 영향을 받는다. 본 발명에서는 눈추적 모듈(100)의 시스템 제어 파라미터를 화질 라벨(image-quality labels), 관측값들(observation) 및 보상값들(rewards)에 의해 적응적으로 제어한다.
- [0030] 눈 영역은, 이미 잘 알려진 얼굴 위치 및 눈 영역 위치 방법을 이용하여 정확한 위치를 찾아낼 수 있고, 눈 추적 역시 이미 잘 알려진 눈 추적 방법에 기반하여 처리될 수 있다.
- [0031] 초기 눈의 중심이 전처리, 특성 추출 및 부분 허프 변환(Hough transformation)에 의해 추정될 수 있다.
- [0032] 전처리는 적응적으로 선택된 알고리즘 하부 구조 및 선택된 알고리즘 하부 구조와 관련된 문턱값을 이용하여 수행될 수 있고, 특성 추출은 윤곽이나 엣지 이미지를 생성하기 위해 수행될 수 있다.
- [0033] 윤곽 또는 엣지 이미지는 최적의 트래킹 포인트를 결정하기 위해, 홍채 경계의 조절된 호각 파라미터를 이용한 부분 허프 변환에 의해 처리될 수 있다.
- [0034] 최종적으로, 칼만 필터가 적용될 수 있고, 다음 눈 중심이 예측될 수 있다. 상술한 단계들은 눈 추적이 종료될 때까지 반복될 수 있다.
- [0035] 눈 추적 제어 공간은, 예를 들어 가능한 알고리즘 구조와 문턱값들 또는 파라미터들의 다양성은, 추적의 정확도와 실행 시간 제약간의 트레이드 오프를 고려한 이전 지식 또는 경험에 기반하여 결정될 수 있다.
- [0036] 이와 같은 눈 추적 방법에 있어서 전처리 단계는 히스토그램 평활화, 하나의 문턱값을 이용한 레티넥스 및 두 개의 파라미터들을 이용한 엔드인 컨트라스트 스트레칭(end-in contrast stretching)으로 구성될 수 있다.
- [0037] 예를 들어, 전처리 단계의 가능한 알고리즘 구조들은 히스토그램 평활화(HE:Histogram equalization), 레티넥스, 직렬 조합에서 히스토그램 평활화를 이용한 레티넥스 또는 엔드 인 컨트라스트 스트레칭일 수 있다.
- [0038] 다음으로, 특징 추출 단계는 (1) 이진화(BN:binarization), (2) 캐니 엣지 탐색 알고리즘, (3) AND 연산을 이용한 병렬 조합에서 이진화와 캐니 또는 (4) OR 연산을 이용한 병렬 조합에서 이진화와 캐니가 될 수 있다. 여기서, 직렬 조합에서 이진화와 윤곽은 단순성을 위해 이진화에 의해 나타나질 수 있다.
- [0039] 특징 추출은 상술한 (1) 내지 (4)의 알고리즘 구조 중 하나를 이용하여 선택적으로 수행될 수 있다. 이진화는 하나의 문턱값(X_{BN})을 가질 수 있고, 캐니 알고리즘은 엣지 탐지를 위한 두 개의 문턱값을 가질 수 있다. 여기서, 캐니
- [0040] 알고리즘을 위한 두 개의 문턱값 중 하나는 엣지 링킹과 관련될 수 있고, 다른 하나는 스트롱 엣지의 초기 세그먼트를 찾는 것과 관련될 수 있다.
- [0041] 한편, 눈 추적 방법에서 홍채 경계는, 낮은 해상도 이미지를 이용하여 실시간 성능을 제공하는 것을 목표로 하기 때문에 원에 근사할 수 있다.
- [0042] 알고리즘은 홍채 이미지의 외부 경계를 계산함으로써, 눈(홍채)의 중심을 탐지할 수 있다. 이때, 노이즈와 눈꺼풀 가림의 영향을 덜 미치는 특성을 추출하기 위해, 수정된 허프 변환 방법(부분 허프 변환이라 불리는데)이 이용될 수 있다.
- [0043] 완전한 원 대신에 원의 두 부분(부분 원들)이 연결되어 위와 아래의 눈꺼풀들이 교합되는 효과를 최소화할 수 있다.
- [0044] 눈의 중심 트래킹을 위한 4차원의 제어 공간은 부분 관심사(예를 들어, 부분 원들)가 되는 홍채 경계를 나타내는 각도값을 포함할 수 있다.
- [0045] 도 2는 눈이 떠지는 정도에 따라 홍채의 외부 경계를 나타내는 부분 원들을 표시한 도면이다. 도 2는 눈꺼풀 교합 효과를 피하기 위해, 모든 홍채 경계 대신에, Φ_1 과 Φ_2 의 각도들과, Φ_3 와 Φ_4 의 각도들간의 홍채 외부 경계

부분이 고려되는 모습을 나타내고 있다.

[0046] 서클피팅을 위한 부분 허프 변환 알고리즘은 다음과 같이 묘사될 수 있다. 우선, 홍채 외부 원은 다음 수학식 1과 같이 표현될 수 있다.

[0047] (수학식 1)

$$[0048] \quad (x-a_0)^2 + (y-b_0)^2 = r^2$$

[0049] 여기서, (a_0, b_0) 는 눈 중심(홍채 외부 경계의 중심)의 좌표를 의미할 수 있고, r 은 원의 반지름을 의미할 수 있다. 이때, 다음과 같은 두 개의 스테이지 알고리즘이 이용될 수 있다. 첫 번째 스테이지는 눈 원 중심을 찾고, 두 번째 스테이지는 눈 중심을 교차하는 부분 원들에서 홍채 외부 경계점의 정상적인 방향을 예측할 수 있다. (x, y) 가 홍채 외부 경계에서의 경사 방향점(gradient direction point)이라고 가정하자. (x, y, X_{PTH}) 삼원수로부터 센터 파라미터 (a, b) 까지의 맵핑은 직선이 된다. 여기서, X_{PTH} 는 경사 방향의 각도를 의미할 수 있다. 이러한 직선들의 많은 교차는 눈 중앙의 좌표를 식별할 수 있다. (x, y, X_{PTH}) 과 (a, b) 간의 관계는 다음 수학식 2와 같이 주어질 수 있다.

[0050] (수학식 2)

$$[0051] \quad X_{PTH} = \arctan((y-b)/(x-a))$$

[0052] 부분 허프 변환 알고리즘에서, 홍채 외부 경계의 범위는 눈꺼풀 교합의 효과를 피하기 위해 다음 수학식 3의 일레와 같이 네 개의 각도 파라미터 X_{PTH1} , X_{PTH2} , X_{PTH3} , 및 X_{PTH4} 에 의해 제한될 수 있다.

[0053] (수학식 3)

$$[0054] \quad X_{PTH} = [X_{PTH1}, X_{PTH2}] \cup [X_{PTH3}, X_{PTH4}]$$

[0055] 한편, 칼만 필터(KF:Kalman filter)는 제곱 오차를 최소화하는 상태 프로세스의 추정에 의한 불연속성의 선형 필터 문제를 위한 반복적인 접근법일 수 있다.

[0056] 칼만 필터는 베이저안 트래킹 문제로서 체계화되는 눈 중앙의 트래킹의 근사치를 계산하기 위해 이용될 수 있다.

[0057] 눈 중앙의 움직임이 다음 수학식 4 및 수학식 5로서 결정될 수 있는 일정 속도 F_t , 상태 천이 모델의 공분산 행렬 및 프로세스 노이즈 모델을 갖는다고 가정하자.

[0058] (수학식 4)

$$[0059] \quad Q_{t-1} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & q^2 & 0 \\ 0 & 0 & 0 & q^2 \end{bmatrix}$$

[0060] (수학식 5)

$$[0061] \quad F_t = \begin{bmatrix} 1 & 0 & \Delta T & 0 \\ 0 & 1 & 0 & \Delta T \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

[0062] 여기서, ΔT 는 인접한 프레임들간의 시간 주기(보통 매우 짧은)를, q 는 프로세스 노이즈와 연관된 파라미터를 각각 의미할 수 있다. H_t 가 상수라 가정하면, 측정 모델의 매트릭스들과 측정 노이즈 공분산은 각각 다음 수학식

6 및 7과 같이 결정될 수 있다.

(수학식 6)

$$H_t = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

(수학식 7)

$$R_t = \begin{bmatrix} r & 2 \\ 0 & r \end{bmatrix}$$

여기서, r 은 측정 노이즈와 관련된 파라미터를 의미할 수 있다. 칼만 필터 접근법은 종종 노이즈 공분산 매트릭스 Q_{t-1} 과 R_t 의 추정에 어려움이 있다. 많은 어플리케이션들에서, 노이즈 공분산의 Q_{t-1} 과 R_t 는 빠르게 안정화될 수 있고, 불변할 수 있다. 파라미터들 r 과 q 는 오프라인에서 필터를 동작시키거나 또는 상태 값을 결정함으로써 미리 계산될 수 있다. 그러나, 노이즈는 동적으로 변화하는 환경의 불확실성 때문에 눈 추적에서 불변하지 않는다. 파라미터들 r 과 q 는 본 발명의 실시예들에 따른 눈 추적 스킴에서 최적의 성능을 달성하기 위해 조정될 수 있다.

여기에서, 시스템 제어 파라미터는 표기상의 통일을 위하여 r 은 X_{KF1} 으로 표현하고, q 는 X_{KF2} 로 표기한다.

결론적으로, 시스템 제어 파라미터 공간 X^Ψ 는 파라미터들을 내부 집합으로 갖는 아래 수학식 8로 표현된다.

(수학식 8)

$$X^\Psi = (X_{BN}, X_{PHT1}, X_{PHT2}, X_{PHT3}, X_{PHT4}, X_{KF1}, X_{KF2})$$

여기에서, 각각의 파라미터는 해당하는 이산 영역에 포함된다.

상기 눈 추적 모듈(100)은 CA-POMDP 모듈(300)에 의해 시스템 제어 파라미터가 상태 s 로 정의되고 행위 a 에 의해 이러한 시스템 제어 파라미터가 변화될 때 실시간 Q-학습부(330)의 강화학습을 위하여 각 행위의 실행 결과를 평가하여 보상값을 결정하여 실시간 Q-학습부(330)에 제공한다. 이때, 보상값을 월드 컨텍스트 모델링부(320)에도 제공하여 월드 컨텍스트 모델을 업데이트 하도록 한다.

한편, 외부 환경이 주어지면, 관련된 문턱값과 파라미터들을 갖는 최적의 알고리즘 구조가 인지된 이미지 컨텍스트에 따라 결정될 수 있다.

일반적으로, 컨텍스트는 관심사의 시스템 성능에 영향을 미치는 어떤 정보가 될 수 있다. 이미지 컨텍스트는 조명의 방향, 밝기, 대비 및 분광 조성에 의해 영향을 받을 수 있다. 화질(image quality)의 분석 기술을 이용한 이미지 컨텍스트의 개념은 시스템 성능을 향상시킬 수 있다.

최근, 화질의 분석 방법은 이미지 저장, 압축, 통신, 디스플레이, 분할 및 인식 등과 같은 다양한 어플리케이션에 성공적으로 적용되고 있으며, 적응적 가

중 파라미터를 이용하여 적응 빛 정규화를 위한 전처리를 결정하는데 이용될 수 있다.

이를 위하여 화질 평가 모듈(200)은 화질을 분석하고 분석된 결과를 화질 라벨(Φ)로 CA-POMDP 모듈(300)로 알려준다.

이를 위하여 화질 평가 모듈(200)은 화질 라벨부(210)와 화질 학습부(220)로 구성되어 있다.

여기에서, 화질 라벨부(210)는 이미지 프레임 I 가 입력되면 도 3에 도시되어 있는 바와 같이 관심 영역(ROI: region of interest)(210-1)을 추출한다.

[0081] 그리드(grid) Θ 는 n 개의 그리드 셀로 이루어지며 이에 따라 $\Theta = \{\omega_1, \dots, \omega_n\}$ 으로 정의되며, ROI는 $I(\Theta)$ 로 정의된다.

[0082] 그리고, 화질 라벨부(210)는 관심 영역을 사각 패치(rectangular patch)(210-2)로 분할하고, 사각 패치(210-2)의 인덱스를 만든다.

[0083] 사각 패치를 $I(\omega)$ 라 하고, 각각의 ω 가 중심 그리드 셀을 의미할 때에 $I(\Theta) = I(\omega_1) \cup \dots \cup I(\omega_n)$ 로 표현된다.

[0084] n 개의 그리드 셀의 라벨 집합을 $\Lambda_i (i=1, \dots, n)$ 이라고 하고, $\Lambda^+_i = \{0\} \cup \Lambda_i$ 라고 하며(여기에서 $\{0\}$ 은 라벨이 없는 널 라벨을 의미함), Φ^j_i 를 Λ^+_i 의 j 번째 라벨이라 하면, $\Lambda_i = \{\Phi^1_i, \dots, \Phi^l_i\}$ 이 된다. 이에 따라, 관심 영역의 화질 라벨 공간은 $\Phi(I(\Theta)) \in \mathcal{T}$ 로 정의될때 $\mathcal{T} = \Lambda^+_1 \times \dots \times \Lambda^+_n$ 에 속한다.

[0085] 다음으로, 화질 라벨부(210)는 프로토타입의 픽셀을 로드하고, 사각 패치의 W 스캐닝 윈도우의 픽셀을 추출한 후에, 스캐닝 윈도우의 픽셀을 프로토타입의 픽셀로 비교하면서 화질을 계산하여 화질 인덱스를 생성하여 출력한다.

[0086] 여기에서, W 스캐닝 윈도우는 사각 패치의 좌에서 우로 그리고 위에서 아래로 수행된다.

[0087] 이와 관련하여 상세히 살펴보면, 스캐닝 윈도우를 k 라 할때, 각각 Φ^j_i 를 갖는 클러스터 프로토타입(cluster prototype)의 픽셀은 $z^k(\Phi^j) = \{z^k_1, \dots, z^k_m\}$ 으로 표현된다. 그리고, 스캐닝 윈도우 k 의 테스트 픽셀의 화질값은 $x^k = \{x^k_1, \dots, x^k_m\}$ 이라 하며, W 스캐닝 윈도우에서 예측값 $\Delta(\Phi^j)$ 는 아래 수학식9와 10으로 정해진다.

[0088] (수학식 9)

$$\Delta^k(\Phi^j) = \frac{4\sigma_{xz}^k \bar{x}^k \bar{z}^k}{[(\sigma_x^k)^2 + (\sigma_z^k)^2][(\bar{x}^k)^2 + (\bar{z}^k)^2]}$$

[0089]

$$\begin{aligned} \bar{x}^k &= \frac{1}{m} \sum_{i=1}^m x_i^k, & \bar{z}^k &= \frac{1}{m} \sum_{i=1}^m z_i^k, & (\sigma_x^k)^2 &= \frac{1}{m-1} \sum_{i=1}^m (x_i^k - \bar{x}^k)^2, & (\sigma_z^k)^2 &= \frac{1}{m-1} \sum_{i=1}^m (z_i^k - \bar{z}^k)^2 \end{aligned}$$

[0090] 여기에서, $\sigma_{xz}^k = \frac{1}{m-1} \sum_{i=1}^m (x_i^k - \bar{x}^k)(z_i^k - \bar{z}^k)$ 이다.

[0091] (수학식 10)

$$\Delta(\Phi^j) = \frac{1}{W} \sum_{i=1}^W \Delta^k(\Phi^j)$$

[0092]

[0093] 여기에서, $\Delta(\Phi^j) \in (-1, 1)$ 이며, $\Delta(\Phi^j)$ 의 최대값은 모든 $i=1, \dots, m$ 일때 $z^k(\Phi^j) = x^k$ 즉, $z^k_i = x^k_i$ 이면 1.0이다.

[0094] 이미지 패치 $I(\omega)$ 에서 화질 라벨은 다음 수학식 11로 결정된다.

[0095] (수학식 11)

$$\phi_i^* = \operatorname{argmax}_j \Delta(\phi_i^j)$$

[0096]

[0097] 관심 영역 $I(\Theta)$ 에서 화질 라벨은 다음 수학식 12로 얻어진다.

[0098] (수학식 12)

[0099] $\Phi(I(\Theta)) = \Phi_1^* \Phi_2^* \dots \Phi_n^*$

[0100] 각각 Φ_i^j 의 화질값은 x^k 는 아래 수학적 식 13과 같이 상관 관계, 조명 왜곡 및 컨트라스트 왜곡의 3개부분으로 나뉘어진다.

[0101] (수학적 식 13)

[0102]
$$\Delta(\phi^j) = \frac{1}{W} \sum_{k=1}^N \frac{\sigma_{xz}^k}{\sigma_x^k \sigma_z^k} \cdot \frac{2\bar{x}^k \bar{z}^k}{(\bar{x}^k)^2 + (\bar{z}^k)^2} \cdot \frac{2\sigma_x^k \sigma_z^k}{(\sigma_x^k)^2 + (\sigma_z^k)^2}$$

[0103] 여기에서, 눈 추적의 성능의 주요한 요소와 실시간의 제약을 고려할 때에 조명 영향이 가장 중요하면 다음 수학적 식 14로 화질값은 간략화된다.

[0104] (수학적 식 14)

[0105]
$$\tilde{\Delta}(\phi^j) = \frac{1}{W} \frac{2\bar{x}^k \bar{z}^k}{(\bar{x}^k)^2 + (\bar{z}^k)^2}$$

[0106] 이에 따라, 관심 영역의 화질은 다음 수학적 식 15로 표현된다.

[0107] (수학적 식 15)

[0108]
$$\tilde{\Phi}(I(\Theta)) = \tilde{\Phi}_1^* \tilde{\Phi}_2^* \dots \tilde{\Phi}_n^*$$

[0109]
$$\tilde{\Phi}_i^* = \operatorname{argmax}_j \tilde{\Delta}(\tilde{\Phi}_i^j)$$
이다.

[0110] 결론적으로, 화질 라벨부(210)는 조명변화를 반영하여 화질을 계산한다.

[0111] 다음으로, 화질 학습부(220)는 수집된 훈련 영상을 사용하여 화질 인덱스 프로토타입으로부터 화질 인덱스를 학습한다.

[0112] 한편, CA-POMDP 모듈(300)는 제어부(310), 월드 컨텍스트 모델링부(320) 및 실시간 Q-학습부(330)로 이루어져 있다.

[0113] 상기 CA-POMDP 모듈(300)은 시스템 제어 파라미터를 상태 s로 정의한다(제어부가 이 역할을 수행한다).

[0114] 그리고, CA-POMDP 모듈(300)은 화질 라벨별로 다양한 객체들의 조합으로 월드 컨텍스트 모델을 구성하고(월드 컨텍스트 모델링부가 수행함), 다양한 객체들의 속성과 속성 값들로 표현된 월드 컨텍스트 모델상의 정보를 기초로 각 상태 변수들의 불리언 값을 결정함으로써 현재 상태를 판단한다(제어부가 수행함).

[0115] 이와 같은 CA-POMDP 모듈(300)에서 판단된 현재 상태는 행위와 Q-학습이 중요한 기초가 된다.

[0116] 상기 CA-POMDP 모듈(300)은 현재 상태에 대한 행위의 보상값을 기초로 그 행위의 가치 함수인 Q 값을 갱신하고 갱신된 Q 값을 저장하며(실시간 Q-학습부가 수행함), 갱신된 Q 값에 근거한 실행 가능한 행위를 수행하고(제어부가 수행함) 눈 추적 모듈(100)에서 보상값을 제공받아 Q 값을 갱신하면서 가치 함수인 Q 값이 가장 우수한 행위를 선택한다(제어부와 실시간 Q-학습부가 수행함).

[0117] 일반적으로, 부분 관찰 마르코프 의사 결정 모델(partially observable Markov decision process, POMDP)은 부분적으로 관찰 가능한 문제에서의 일반적인 의사 결정 프레임워크이다.

[0118] 이는 불확실성이 존재하는 실제 문제들에 적합한 모델이고, 실제 세계에 대한 정확한 모델을 가정하며, 주어진 모델에서의 최적 행위 정책을 찾는다.

[0119] POMDP는 부분적으로 관측할 수 있는 불확실한 상황을 고려하여 기본적으로 다음{S,A,O,T,Ω,R,γ}의 구성요소를

가진다.

[0120] 여기에서 S는 상태(state)s의 집합이며, A는 제어부가 취할 수 있는 행위(action) a의 집합이고, T는 상태 s의 변이 확률 분포이며, R은 제어부의 행위에 따른 보상r의 확률 분포이며, O는 관측치 o의 집합이며, Ω는 실제 상태에 대한 관측치의 확률 분포이다. γ는 할인율로 0과 1사이의 실수 값을 가진다.

[0121] 즉, 현재 상태를 s, 다음 상태를 s'라 하면, $T(s, a, s')=P(s' | s, a)$ 가 되며, $\Omega(s, a, o')=P(o' | s, a)$ 가 된다.

[0122] MDP에서 특정 상태를 기반으로 판단하게 되지만 POMDP에서는 이것이 불가능해지기 때문에 부분적으로 관측 정보를 이용해서 확률적으로 상태를 결정하도록 하는 신뢰 공간(belief space)에 대한 정의가 필요하다. 이러한 신뢰공간으로 인해 POMDP 해결은 MDP 해결로 가능하게 된다.

[0123] 시점 t에서의 상태 s의 분포 확률을 $b_t(s)$ 라 하면 에이전트 히스토리에 의해 $b_t(s)=P(s_t | h)$ 로 정해진다. 여기에 서, $h=((a_0, o_1, r_1), (a_1, o_2, r_2), \dots, (a_{t-1}, o_t, r_t))$ 이다.

[0124] POMDP에서 최적의 정책을 구하는 문제는 신뢰 상태 MDP에서 4가지 튜플 $[B, A, \tau, R_b]$ 를 구하는 것과 같다.

[0125] 신뢰 상태 b는 상태 평가자 τ를 이용하여 다음 신뢰 상태 b'으로 업데이트되며, $b'=\tau(b, a, o)$ 로 정해진다. 상태의 업데이트는 아래 베이지안(Bayesian) 규칙[5]에 의해 정해진다.

[0126] (수학식 16)

$$b'(s') = \Pr(s' | b, a, o) \propto \frac{\Omega(s', a, o) \sum_{s \in S} T(s, a, s') b(s)}{\sum_{s \in S} \Omega(s', a, o) \sum_{s \in S} T(s, a, s') b(s)}$$

[0127]

[0128] 그리고, 보상 함수는 다음 수학식 17로 정의된다.

[0129] (수학식 17)

$$R(b, a) = \sum_{s \in S} b(s) R(s, a)$$

[0130]

[0131] 여기서, V^* 를 정책 π의 가치함수(value function)라 한다. 만약 V^* 가 '정의할 수 있는 다른 모든 타 정책의 가치함수'보다 크거나 같다면, 이때의 정책을 최적정책(optimal policy)이라고 하고, π^* 로 표기한다.

[0132] 결과적으로, MDP에서 해를 구하는 과정은 최적정책(π^*)을 구하는 과정이라 할 수 있다.

[0133] 최적 정책은 아래 수학식 18의 벨만(Bellman) 방정식에 의해 정의된다.

[0134] (수학식 18)

$$V^*(b) = \max_{a \in A} Q^*(b, a)$$

[0135]

[0136] 여기에서, $Q(b, a)$ 는 행위 값 함수로 최적의 행위 값 함수는 $Q^*(b, a)$ 이다.

[0137] 최적의 행위 값 함수는 다음 수학식 19로 얻을 수 있다.

[0138] (수학식 19)

$$Q^*(b, a) = R(b, a) + \gamma \sum_{o \in O} \Omega(b, a, o) V^*(\tau(b, a, o))$$

[0139]

$$\Omega(b, a, o) = \sum_{s \in S} \Pr(o | s, a) b(s)$$

[0140] 한편, 관찰 분포 함수 이다.

- [0141] 상기 CA-POMDP에서 상태 공간은 컨텍스트 공간에 의존하는 시스템 제어 공간에 의해 한정된다. 여기에서, 컨텍스트는 눈 추적 모듈(100)의 시스템 성능에 영향을 미치는 어떤 정보가 될 수 있다. 이미지 컨텍스트는 조명의 방향, 밝기, 대비 및 분광 조성등이 될 수 있다.
- [0142] 행위의 집합 A는 상태 공간에서 시스템 제어 파라미터의 적응을 나타낸다. 그리고, 관찰의 집합 O는 객체가 출현하는 추적 성능과 화질의 상태는 나타낸다.
- [0143] 행위 주체는 현재의 환경 상태를 직접적으로 알 수 없고, 대신 관찰만을 얻을 수 있다. 따라서 행위 주체는 현재까지의 모든 행위와 그에 대응되는 관측을 통해 현재의 환경 상태에 대한 확률 분포를 유지한다. 이와 같은 것을 신뢰 상태(belief state)라고 하며, 이러한 공간을 신뢰 공간 B라고 한다. 따라서, 행위 주체는 신뢰 공간에서 근거해서 불확실성을 탐지한다.
- [0144] 하지만, 일반적인 POMDP에서는 상태 번호 $|S|$ 에서 신뢰 공간 B의 급증때문에 실시간 해결을 제공하고 있지 않다.
- [0145] 상태 천이들, 관측들 그리고 보상들의 확률 분포는 그러한 처리상의 어려움을 피하기 위하여 한정된 월드 컨텍스트 모델들과 결합되어야 한다.
- [0146] 상기 월드 컨텍스트 모델들은 눈 추적의 불확실성에 대한 영역에 대한 이해를 나타내며, 영역에 대한 이해는 시간 범위 모색에 사용된다.
- [0147] 일반적인 POMDP와 달리 단일 모델에서 상태는 관찰되며 전이 동작은 현재의 월드 컨텍스트 모델에 의존하는 국부 상태에 집중될 수 있다.
- [0148] 제안된 CA-POMDP는 추가적인 간략화로 인해 정확도가 떨어지나 실시간의 한계를 만족하는 계산상의 이득이 있다는 것은 분명하다.
- [0149] 상기 CA-POMDP는 $\langle S, A, O, T, \Omega, R, \gamma, \Phi \rangle$ 의 8개 요소로 정의된다.
- [0150] 여기에서, Φ 는 컨텍스트 공간을 나타내며, 컨텍스트 공간은 눈 영상에 영향을 주는 조명의 방향, 밝기, 시야각 및 스케일등이 될 수 있다.
- [0151] 상기 컨텍스트 공간은 영상을 획득한 환경을 고려하여 화질을 측정하는 것을 나타낸다. 일례로, 컨텍스트 공간은 여러가지 변수를 포함하는데, Φ_1 은 조명의 방향, Φ_2 는 밝기, Φ_3 는 시야각 그리고 Φ_4 는 물체의 스케일을 나타낸다.
- [0152] 이와 같은 요소는 이산 값으로 측정되며, 컨텍스트 상태 ϕ 는 각 변수의 정렬된 값으로, 즉 $\phi = \{\Phi_1 = \phi_1, \dots, \Phi_M = \phi_M\}$ 으로 표현된다.
- [0153] 상태 공간 S는 눈 추적 공간에서 시스템 제어 파라미터로 정의되며, 환경의 변화에 영향을 받는다. 정확한 컨텍스트가 불가능하기 때문에, 상태 공간 S는 부분적으로 관찰 가능하게 감안하여 처리한다. 화질 상태가 실질적으로 측정될 수 없기 때문에 화질 라벨 Φ' 에 의해 추정된다.
- [0154] 상태 공간 S는 컨텍스트 공간 Φ 에 의존하는 시스템 제어 공간 X의 집합이다. 즉, $S = (X | \Phi)$ 이며, 여기에서 $X = \{X_1, \dots, X_L\}$ 이고, $\Phi = \{\Phi_1, \dots, \Phi_M\}$ 으로, 컨텍스트 공간에 의존하는 시스템 제어 공간의 랜덤 변수 집합을 포함한다.
- [0155] 상기 랜덤 변수 X_i 는 눈 추적에서 문턱값 또는 파라미터를 의미한다. 개별적인 랜덤 변수의 시스템 제어 파라미터의 정렬 $x = \{X_1 = x_1, \dots, X_L = x_L\}$ 을 나타낸다.
- [0156] 상태 공간 S에서의 상태 s는 다음 수학적 식 20로 표현된다.
- [0157] (수학적 식 20)
- $$s = \{ \{ \{ x_i \}_{i=1}^L \mid x_i \in D_i, \phi \in \Phi \}$$
- [0158]
- [0159] 여기에서, $s \in S$ 이며, s는 제어부 양상의 나타내는 스칼라이며, D_i 는 시스템 제어 파라미터 X_i 의 이산화된 영역이

다.

[0160] 제어부(310)는 눈 추적 장치에 있어서 시스템 제어 파라미터를 행위에 의해 변경할 수 있으나, 컨텍스트 공간에서 화질을 변경할 수는 없다.

[0161] 화질은 실제로 완전하게 측정될 수 없으며 다만 부분적으로 그리고 제한적으로 이용가능하다. 또한, 컨텍스트 공간에서 불확실성으로 인해 상태 s 는 부분적으로 관찰가능하다.

[0162] 행위 집합 a 은 행위 공간 A 에서 정의된다. 행위 a_i 는 컨텍스트 공간 Φ 에 의존하는 시스템 제어 공간 X 의 랜덤 변수 X_i 에 대하여 각각 정의되며 이러한 관계는 수학식 21으로 표현된다.

[0163] (수학식 21)

$$a = \{(\{a_i\}_{i=1}^L | a_i \in [-d_i, 0, +d_i], \phi \in \Phi)\}$$

[0164]

[0165] 여기에서, 행위 a 는 행위 공간 A 에 속하며, 각각의 X_i 수평에서 d_i 단위의 하락을 나타내거나 포화, 상승을 나타내는 제어부의 $-d_i$ 행위, 0 , $+d_i$ 행위를 나타내는 스칼라이다.

[0166] 눈 추적 알고리즘에 있어서 다음 상태는 불확실성을 모델링한 전이 확률 함수 T 에 의해 나타내어진다.

[0167] 함수 T 는 현상태 s , 컨텍스트 상태 $\phi \in \Phi$, 행위 a 가 주어지면 현상태로부터 다음 상태 s' 으로의 전이 확률이다. 상태 전이의 확률 함수 $T: (S \times A \times S | \Phi) \mapsto [0, 1]$ 는 다음 상태 s' 의 확률 분포로 아래 수학식 22와 같이 정의된다.

[0168] (수학식 22)

$$T(s, a, s' | \phi) = P(s_{i+1} = s' | s_i = s, a_i = a, \phi_i \in \Phi) = P(x_{i+1} = x' | x_i = x, a_i = a, \phi_i \in \Phi)$$

[0169]

$$\sum_{s' \in S, \phi \in \Phi} T(s, a, s' | \phi) = 1$$

[0170]

[0171] 관측 공간 O 는 눈 추적 성능과 화질에 걸쳐 있는 분포 공간으로, $O = \{\tau, \phi\}$ 로 나타낸다. 여기에서, τ 는 추적 정확도이고, ϕ 는 화질 측정치이다. 관찰 o 는 관찰 확률 함수 $\Omega: S \times A \times O \mapsto [0, 1]$ 를 사용해서 관찰 공간에 걸쳐 아래 수학식 23로 정의된다.

[0172] (수학식 23)

$$\Omega(s', a, o') = P(o'_\tau, o'_\phi | o_\tau, o_\phi, a_i = a) = P(o'_\tau | a) P(o'_\phi | a)$$

[0173]

[0174] 여기에서, o_τ 와 o_ϕ 는 추적 정확도와 화질 측정치를 나타낸다.

[0175] 더불어, 보상은 행위에 의한 즉각적으로 주어지는 것으로, 함수 $R: S \times A \rightarrow \mathbb{R}$ 로 표현되며, 보상 r 은 $r = R(s, a)$ 로 표현된다. 보상 함수는 상태 s 에서 행위 a 를 취했을 때 얻는 보상을 나타내며, 보상 함수는 추적 성능에 근거해서 측정되며, 문턱값보다 추적 신뢰가 크면 높은 점수가 보상되고, 이와 달라지면 적은 점수가 보상된다. γ 는 할인율로 0과 1사이의 실수 값을 가진다.

[0176] 상기 CA-POMDP에서 막대한 계산상의 요구를 극복하기 위하여, 유연한 월드 컨텍스트 모델들과 각각의 월드 컨텍스트 모델들에 대한 온라인 학습을 통해 간략화한다.

[0177] 상기 월드 컨텍스트 모델들은 온라인 학습에서 전이, 관찰 및 보상 분포의 결합 공간에 걸쳐 월드 컨텍스트 모

텔링부(320)에서 미리 컴파일한 것이다.

- [0178] 온라인 학습은 불충분한 정보와 실시간의 한계로 인하여 눈 추적 알고리즘의 시스템 제어 파라미터의 동적 적응으로 공식화되며 실시간 Q-학습 접근[16]으로 해결된다.
- [0179] 정책 학습은 온라인 정책 학습과 오프라인 정책 학습으로 구분된다[10,11,12]. 오프라인 학습은 행위 선택 정책을 결정하기 위하여 막대한 계산을 요구한다. 이것은 실시간 임무에서 때때로 유리하나, 실행시간에 매우 빈번하게 발생하는 새로운 환경에서 적용하기는 어렵다.
- [0180] 온라인 접근은 실시간으로 약간의 계산을 요구하면서 적응성이 뛰어나다. 월드 컨텍스트 모델링부(320)에 의한 월드 컨텍스트 모델에 대하여 오프라인 학습을 그리고 실시간 Q-학습부(330)에 의한 실시간 국부 정책 결정에 대하여 온라인 학습을 결합하며 계산상의 이점이 있다.
- [0181] 상기 CA-POMDP는 월드 컨텍스트 모델링부(320)를 통하여 실행 동안에는 확률 분포 T, 및 R에 대하여 가장 좋은 상태를 선택하며 오프라인 월드 컨텍스트 모델을 구성한다.
- [0182] 상기 월드 컨텍스트 모델이 실시간 성능 저하 없이 미리 컴파일되기 때문에, 온라인 학습은 제한적인 신뢰 상태를 탐색하여 좋은 정책을 발견하도록 한다.
- [0183] 현재의 월드 컨텍스트 모델에서 최대값이 계산됨에 따라 상기 온라인 학습이 수행되며 정책은 국부적으로 결정된다.
- [0184] 월드 컨텍스트 모델은 개개의 행위보다 긴 시간 동안 충분한 행위가 발생됨에 따라 여러 스텝 동안 유지된다. 동일한 월드 컨텍스트 모델이 여러 시간 간격동안 유지되는 것은 탐색 전략을 일치시킬 수 있는 이점이 있다. 온라인 정책 산출은 시스템 정확도와 실시간 한계 사이에 균형을 고려할 필요가 있다.
- [0185] 한편, CA-POMDP는 월드 컨텍스트 모델에서 오프라인 학습을 수행하고(월드 컨텍스트 모델링부(320)가 수행함), 현재 월드 컨텍스트 모델에 의존하는 상태에서 온라인 학습을 수행한다(실시간 Q-학습부(330)가 수행함).
- [0186] S,A,O 및 R 집합의 카디널리티(cardinality)가 제한되어 있기 때문에, CA-POMDP는 전이, 관찰 및 보상에서 충분한 한계를 허용한다.
- [0187] 월드 컨텍스트 모델링부(320)는 오프라인 학습을 위하여 결합 공간 $CM=Tx\Omega \times R$ 에서 월드 컨텍스트 모델을 미리 컴파일한다. 유사한 접근이 참고문헌 [17]에서 발견되는데, 화질 라벨을 반영하는 컨텍스트 공간을 고려하지 않는다.
- [0188] 도 4를 참조하면, 일반적인 POMDP와 본 발명의 CA-POMDP를 비교해서 보여준다.
- [0189] 일반적인 POMDP는 단일 월드 컨텍스트 모델로 행위가 제어되는데 반해, CA-POMDP에서 전이 동작은 현재 상태와 월드 컨텍스트 모델에 영향을 받는다.
- [0190] CA-POMDP는 다중의 월드 컨텍스트 모델로 구성되며, MDP 아래에서 국부 신뢰 공간을 가지고 있다. CA-POMDP는 환경과 상호작용에 있어서 부분적인 관찰에 대한 마코비안 성질에 근거하고 있다.
- [0191] 랜덤 변수 A,O 및 R은 확률 분포 T, Ω 및 R과 관련되어 있다. $O_1, O_2, \dots, O_n$ 는 관찰 확률 분포 Ω 의 랜덤 변수이다. $A_0, A_1, \dots, A_{N-1}$ 은 전이 확률 분포 T에 대한 랜덤 상수이며, $R_1, R_2, \dots, R_N$ 은 보상 확률 분포 R에 랜덤 상수이다.
- [0192] 행위 $A_0=a_0, A_1=a_1, \dots, A_{N-1}=a_{N-1}$ 에 대하여 각각 관찰 $O_1=o_1, O_2=o_2, \dots, O_N=o_N$ 에 보상 $R_1=r_1, R_2=r_2, \dots, R_N=r_N$ 을 구할 수 있다.
- [0193] CM 공간에 대한 월드 컨텍스트 모델 c에 대한 수집된 훈련 데이터는 $h=\{(a_0, o_1, r_1), (a_1, o_2, r_2), \dots, (a_{n-1},$

$o_n, r_n)\}$ 로 표현된다. 모든 월드 컨텍스트 모델에 대한 전체 훈련 데이터는 $H = \bigcup_{c \in CM} \{h\}$ 로 나타내어진다. 월드 컨텍스트 모델과 상태는 부분적으로 관찰가능하며, 결합 신뢰는 아래 수학적 24로 표현된다.

[0194] (수학식 24)

$$b(s, c | h, \phi) = b_c(s) b_{CM}(c | \phi)$$

[0195]

[0196] 여기에서, $b_c(s)=b(s | c, h)$ 이며, 신뢰 업데이트에 의해 결정된다. 행위는 국부 수행 시간에 걸친 기대 할인 보상을 최대화하는 월드 컨텍스트 모델에 대한 $b_{CM}(c | \phi)$ 에 의해 선택된다.

[0197] $b_{CM}(c)$ 는 월드 컨텍스트 모델 c 에 걸쳐 있는 신뢰를 나타내며, 아래 수학식 25로 간략화된다[17].

[0198] (수학식 25)

$$b_{CM}(c | \phi) \approx w(c | \phi) = \sum w_i \delta(c_i, c, \phi)$$

[0199]

[0200] 여기에서, $\delta(c_i, \cdot)$ 는 화질 라벨 ϕ 에서 월드 컨텍스트 모델의 델타 함수를 나타내며, w_i 는 c_i 와 관련된 가중치를 나타낸다. 컨텍스트 상태 ϕ 는 화질 라벨 ψ 의 간략화이며, 닫힌 형식이 아니다. 상기 월드 컨텍스트 모델링 부(320)에 의해 구성된 월드 컨텍스트 모델의 신뢰는 영역 행위, 관찰 및 보상의 히스토리에 의해 결정된다. 전이, 관찰 및 보상 모델들이 이산 다중으로, 분포[19]를 학습하기 위해 디클레(Dirichlet) 정리와 베이시언(bayesian) 방법을 사용한다.

[0201] 새로운 랜덤변수를 (a, o, r) 이라 하면, 확률은 현재 월드 컨텍스트 모델에 의해 계산되며, 다음 컨텍스트 모델은 MAP 군집화 규칙에 의해 결정된다. 확률은 월드 컨텍스트 모델 c 에 의해 다음 수학식 26로 결정된다.

[0202] (수학식 26)

$$P(c | (a, o, r), H, \phi) = \frac{p((a, o, r) | c, h, \phi) P(c | \phi)}{\sum_{c \in CM} p((a, o, r) | c, h, \phi) P(c | \phi)}$$

[0203]

[0204] 여기에서, $p((a, o, r) | c, h)$ 는 현재 월드 컨텍스트 모델의 조건 밀도를 나타내며, 정리 확률 $P(c)$ 는 현재 월드 컨텍스트 모델의 신뢰 $b_{CM}(c)$ 이다. 그리고, h 와 H 는 개별적인 월드 컨텍스트 모델과 전체 월드 컨텍스트 모델의 각각의 훈련 집합을 의미한다. 새로운 랜덤 변수에 대한 다음 월드 컨텍스트 모델은 MAP 규칙을 사용하여 다음 수학식 27으로 결정된다.

[0205] (수학식 27)

$$\hat{c}_{MAP} = \arg \max_{c \in CM} p((a, o, r) | c, h, \phi) P(c | \phi)$$

[0206]

[0207] 약간의 시간 동안 동일한 월드 컨텍스트 모델을 유지하는 것은 탐색 전략의 일치성을 향상시킨다. 월드 컨텍스트 모델은 개별적인 시간 단계보다 약간의 시간 단계의 기간 동안에 걸쳐 행위, 관찰 및 보상의 충분한 동기화에 기초하여 변경된다.

[0208] 한편, 월드 컨텍스트 모델링부(320)는 오프 라인 학습을 통하여 월드 컨텍스트 모델들을 모델링한다.

[0209] 그리고, 실시간 Q-학습부(330)는 월드 컨텍스트 모델에 근거하여 최적 정책을 산출한다.

[0210] 이처럼 CA-POMDP는 현재의 월드 컨텍스트 모델에 근거하여 최적의 정책을 결정한다. 이는 제어부(310)가 월드 컨텍스트 모델의 국부 시간 범위를 고려하고 온라인 학습에 근거한 신뢰 상태의 작은 집합에 집중하도록 한다.

[0211] 비록 월드 컨텍스트 모델을 오프라인을 통하여 구축하는 것이 CA-POMDP에서 많은 계산 시간의 오버 헤드를 감소시키지만 여전히 시간 복잡도는 지수 함수적으로 증가하기 때문에, 즉 $O(|A||O|)^k$ (여기에서 k 는 재귀 계산의

깊이이고, "0"는 큰 0 표기임)이기 때문에 다루기가 불가능하다. CA-POMDP의 탐색 공간의 감소는 탐색 깊이가 만족하지 못하기 때문에 새로운 온라인 접근을 사용해도 해결할 수 없다. 눈추적에 있어 실시간의 제약은 간단하고 빠른 알고리즘을 요구하며, 실시간 Q-학습은 직접인 신뢰 상태에서 해결하기보다는 불완전한 정보를 가지고 비동기적으로 동적 프로그램을 가진다.

[0212] CA-POMDP의 근사 최적 가치 함수 V^* 는 다음 시간 단계에 모든 불확실성이 해결되는, 즉 다음 행위 이후에 현재 신뢰 나머의 모든 불확실성이 해결되는 것으로 간략화된다.

[0213] QMDP의 경험적인 접근은 아주 큰 장시간을 가지는 모든 상태에서부터 보상을 가지는 행위가 각 시간 단계에서 신뢰 상태가 선택되어 가중되며 아래 수학적 28로 표시된다.

[0214] (수학적 28)

$$V^*(b_c) \cong \tilde{V}(b_c) = \sum_{s \in S_c} b_c(s) V_{MDP}^*(s)$$

[0215]

[0216] 여기에서, MDP 아래에서 $V_{MDP}^*(S)$ 는 상태 s 의 최적 가치 함수를 나타내며, 신뢰 상태 b_c 는 수학적 16을 사용하여 산출된다.

[0217] 간략화된 최적값은 MDP 아래에서 선형이며, 효과적으로 해결가능하다. 월드 컨텍스트 모델 c 에 대한 MDP 아래에서 수행 시간은 아래 수학적 29로 정의된다.

[0218] (수학적 29)

$$S_c = \{(\{X_i = x_i\}_{i=1}^L) \mid x_i \in D_{c,i}, \phi \in \Phi\}$$

[0219]

$$S = \bigcup_{c \in CM} S_c, \quad D_j = \bigcup_{c \in CM} D_{c,i} \text{ 이다.}$$

[0220]

[0221] 각각의 월드 컨텍스트 모델에 대한 MDP 아래에서 각각의 국부 탐색 공간은 MDP 해결을 작은 시간 범위로 감소시킨다. MDP 아래에서 여전히 불완전한 정보를 가지고 있다.

[0222] 눈 추적 모듈을 제거하는 것과 최적 성능을 달성하면서 행위를 추적하는 것의 충돌 문제는 민감한 이슈이다. 이러한 종류의 모순은 많은 탐색과 일반적이지 않은 매커니즘을 필요로 한다.

[0223] MDP는 베이시안 방법 또는 베이시안 방법이 아닌 방법을 사용하여 이러한 문제를 해결한다. 실시간 Q-학습은 직접적인 베이시안 방법은 아니다.

[0224] 상기 실시간 Q-학습부(330)에 의한 각각의 월드 컨텍스트 모델에 대하여 Q-학습은 Q-테이블을 유지하며, 평가 함수 대신에 각각의 볼 수 있는 상태 행위 쌍에 대하여 최적 행위 함수 값인 Q-값의 평가에 의해 백업된다.

[0225] 월드 컨텍스트 모델 c 의 Q-테이블은 수집된 훈련 데이터에 의해 초기화된다. $Q_k^c(s_i, a_i)$ 는 스테이지 k 에서

$$Q_k^c(s_i, a_i) \text{의 평가를 나타내며, 즉 } \hat{Q}_k^c(s_i) = \max_{a_i \in A(c)} Q_k^c(s_i, a_i) \text{ 이다.}$$

[0226] $k=0, 1, \dots, n$ 의 각각의 스테이지에서 실현가능한 상태-행위 쌍이 $SA_k^c \subseteq \{(s_i, a_i) \mid s_i \in S_c, a_i \in A_i\}$ 로 표현되도록

하면, Q-값은 스테이지 k 에서 업데이트된다. $(s_i, a_i) \in SA_k^c$ 에 대하여, Q-값은 다음 수학적 30에 따라 스테이지 $k+1$ 에서 백업된다.

[0227] (수학식 30)

$$Q_{t+1}(s_i, a_i) = (1 - \alpha(s_i, a_i))Q_t(s_i, a_i) + \alpha(s_i, a_i)[r_i(a_i) + \gamma Q_t(s_i, ')]$$

[0228]

[0229] 여기에서, s_i' 는 다음 상태이며, 즉 $s_i' = T(s_i, a_i)$ 이고, $\alpha(s_i, a_i)$ 는 현재 상태-행위 쌍에서 학습률 파라미터이며, r_i 는 보상 함수이다. MDP 아래에서 명시적인 모델의 구성하지 않고도 최적의 정책을 결정할 수 있다.

[0230] 실시간 적용에 있어서, 제어부(310)는 각각의 시간 단계 t 에서 상태 s_t 를 관찰하며, 초기화 상태 이후에 모든 진행하는 시간 단계에서 최적의 Q-값을 산출한다.

[0231] 제어부(310)는 상태 $a_t \in A_c(s_t)$ 를 선택하며, 그것을 수행하고, 눈추적 모듈(200)의 상태가 다음 상태로 전이될 때에 즉각적인 보상 $r_{s_t}(a_t)$ 을 받는다.

[0232] 이때, 월드 컨텍스트 모델 c 의 Q-값은 시간 $t+1$ 에서 아래 수학식 31에 따라 백업된다.

[0233] (수학식 31)

$$Q_{t+1}(s_i, a_i) = (1 - \alpha(s_i, a_i))Q_t(s_i, a_i) + \alpha(s_i, a_i)[r(a_i) + \gamma \max_{a \in A(s_{t+1})} Q_t(s_i, a)]$$

[0234]

[0235] 여기에서, $\alpha(s_t, a_t)$ 는 시간 단계 t 에서 현재 상태-행위 쌍에서 학습률 파라미터이며, r_t 는 보상 함수이다.

[0236] 실시간 Q-학습부(330)는 오프라인 Q-학습의 제안된 경우로, 상태-행위 쌍의 집합에서 Q-값이 각 단계 t 에서 백업되며, $S_t^{\mathcal{Q}^c}$ 는 (s_t, u_t) 가 된다.

[0237] 실시간 Q-학습부(330)는 오프라인 학습이 발산되는 것을 요구하는 상태의 Q^* 에서는 발산한다.

[0238] 최적의 정책에서 Q-학습이 각각의 실현가능한 행위가 무한한 수의 시간 단계가 반복되는 각 상태에서 백업되며 점진적인 방법으로 시간 단계 t 가 학습률 α 로 줄어들때 발산하는 것은 Watkins[47]에 의해 제공된다.

[0239] 실시간 Q-학습부(330)는 비동기 DP 백업보다 적은 백업을 실시간 응용에서 필요하기 때문에 이점이 있다.

[0240] 만약, n 상태가 있고 어떤 상태에서 허용되는 동작의 가장 큰 수가 m 이면, Q-학습의 백업은 $O(m)$ 이며 DP 백업의 $O(mn)$ 보다 작다.

[0241] 본 발명에서 국부적인 정책의 수립이 최적화되지 않을 때에, 즉 성공적인 눈 추적 기준에 부합되지 않을 겨웅에 추가적인 계산이 요구된다. 그 경우에, 다음 실현 가능한 월드 컨텍스트 모델이 시도되며, 시간 슬라이스는 종료된다.

[0242] 눈 추적의 정확도와 실시간의 요구는 시간 슬라이스의 길이를 조정하여 균형을 이룰 수 있다.

[0243] 이와 같은 강건한 휴먼 컴퓨터 상호작용을 위한 CA-POMDP를 이용한 눈 추적 장치의 동작을 도 5의 흐름도를 참조하여 설명하면 다음과 같다.

[0244] CA-POMDP 모듈(300)의 제어부(310)는 각각의 월드 컨텍스트 모델에 대한 상태 s 를 초기화한다(S100).

[0245] 그리고, 이때 화질 평가 모듈(200)의 화질 라벨부(210)는 첫번째 영상을 획득하고 화질 라벨을 산출하여 CA-POMDP 모듈(300)의 제어부(310)와 월드 컨텍스트 모델링부(320)에 제공한다(S200).

[0246] 한편, CA-POMDP 모듈(300)의 제어부(310)는 실시간 Q-학습부(330)에 월드 컨텍스트 모델 c 를 질의한다(S300).

[0247] 이후에, 제어부(310)는 시간 슬라이스를 초기화한다(S400).

[0248] 이에 따라 실시간 Q-학습부(330)는 그리디 폴리시를 이용하여 현재 상태 s 로부터 행위 a 를 선택한다(S500).

[0249] 그리디 폴리시는 현재 낮은 값 행위를 학습하는 것을 허용하지 않기 때문에, 시험해 보지 않은 동작 시퀀스를

탐색하지 않는다. 그러나, 미래에 높은 값을 유도할 수 있다. 이와 관련하여, ϵ -그리스 폴리스가 탐색과 개발 간의 행위의 균형을 맞추기 위해 이용될 수 있다.

[0250] 실시간 Q-학습부(330)는 행위를 얻고, 내부 보상을 계산함으로써 즉각적인 보상 r 을 관찰하고, 새로운 내부 상태 s' 을 관찰한다(S600).

[0251] 다음으로, 실시간 Q-학습부(330)는 월드 컨텍스트 모델 c 의 Q-테이블을 업데이트하며, 이때 수학식 31이 이용된다(S700).

[0252] 그리고, 실시간 Q-학습부(330)는 상태를 다음 상태로 변경하며(S800), 시간 슬라이스가 종료되거나 눈 추적이 성공 조건을 만족하는 경우에(S900), 단계 S200으로 진행하고, 종료되지 않거나 만족하지 않는 경우에 단계 S500으로 진행한다.

[0253] 상기와 같은 본 발명에 따르면 고비용의 이미지 캡처 기기나 매우 제한적인 상황을 이용하는 대신, CA-POMDP를 사용하여 시스템 제어 파라미터를 최적화하여 성능을 보장할 수 있다.

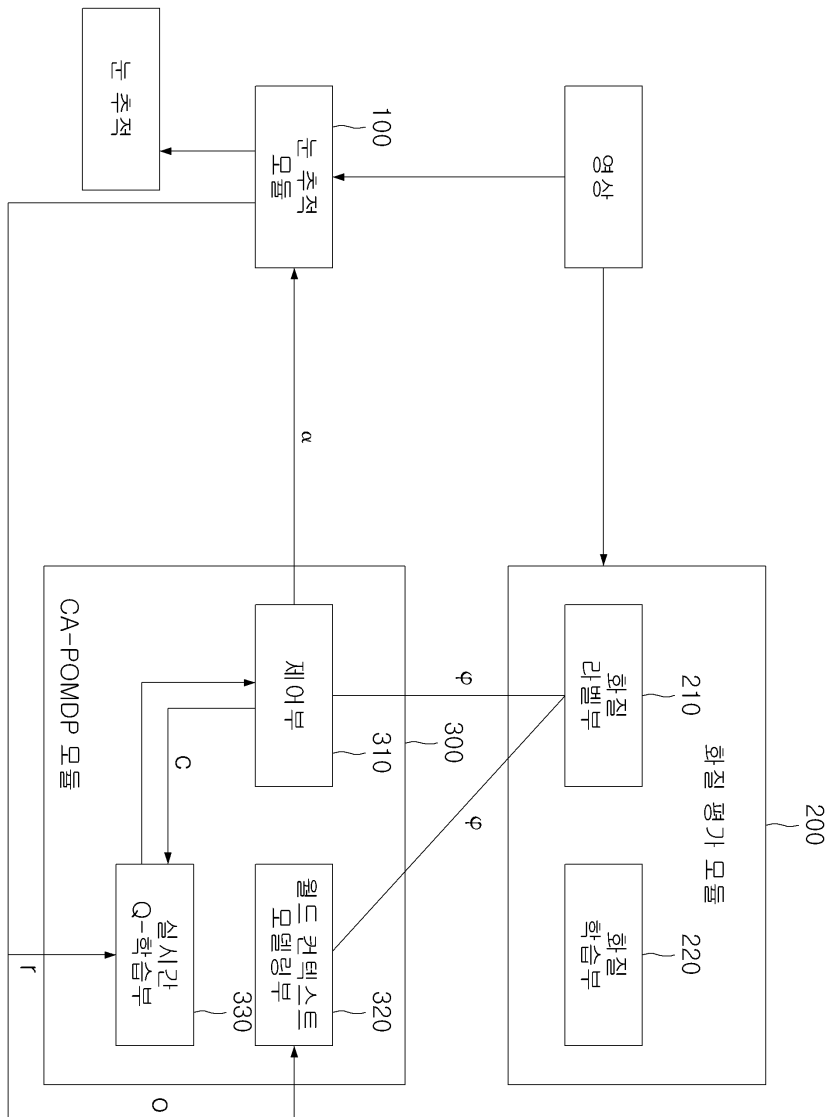
[0254] 이상의 설명은 본 발명의 기술 사상을 예시적으로 설명한 것에 불과한 것으로서, 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자라면 본 발명의 본질적인 특성에서 벗어나지 않는 범위에서 다양한 수정 및 변형이 가능할 것이다. 따라서 본 발명에 기재된 실시예들은 본 발명의 기술 사상을 한정하기 위한 것이 아니라 설명하기 위한 것이고, 이러한 실시예에 의하여 본 발명의 기술 사상이 한정되는 것은 아니다. 본 발명의 보호 범위는 아래의 청구범위에 의해서 해석되어야 하며, 그와 동등한 범위 내에 있는 모든 기술 사상은 본 발명의 권리범위에 포함되는 것으로 해석되어야 할 것이다.

부호의 설명

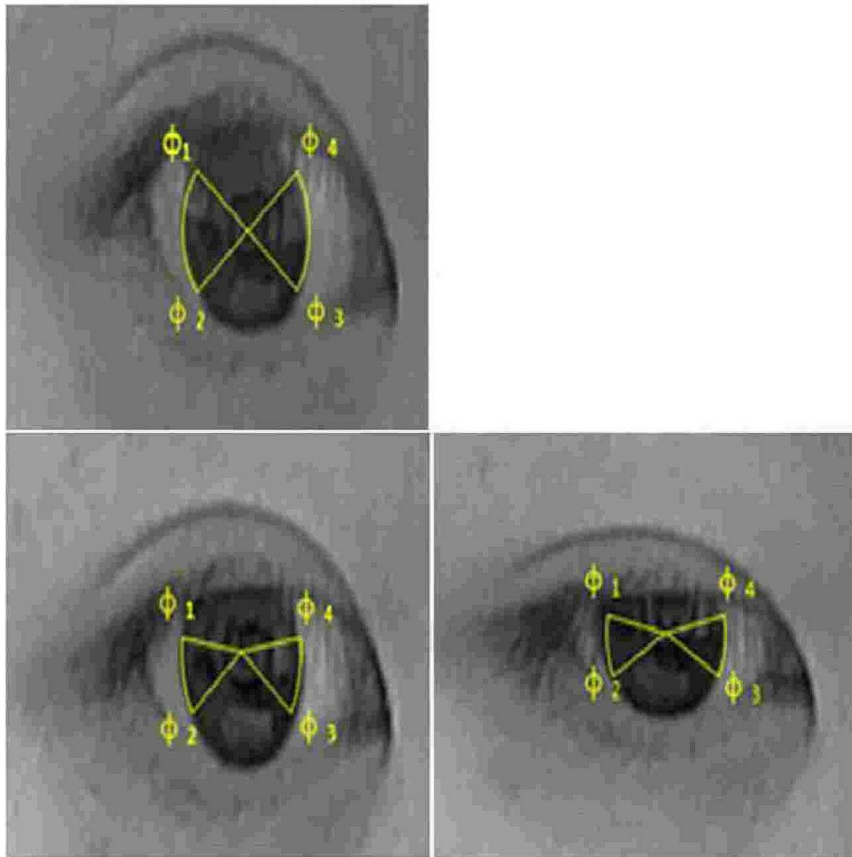
[0255]	100 : 눈 추적 모듈	200 : 화질 평가 모듈
	210 : 화질 라벨부	220 : 화질 학습부
	300 : CA-POMDP 모듈	310 : 제어부
	320 : 월드 컨텍스트 모델링부	330 : 실시간 Q-학습부

도면

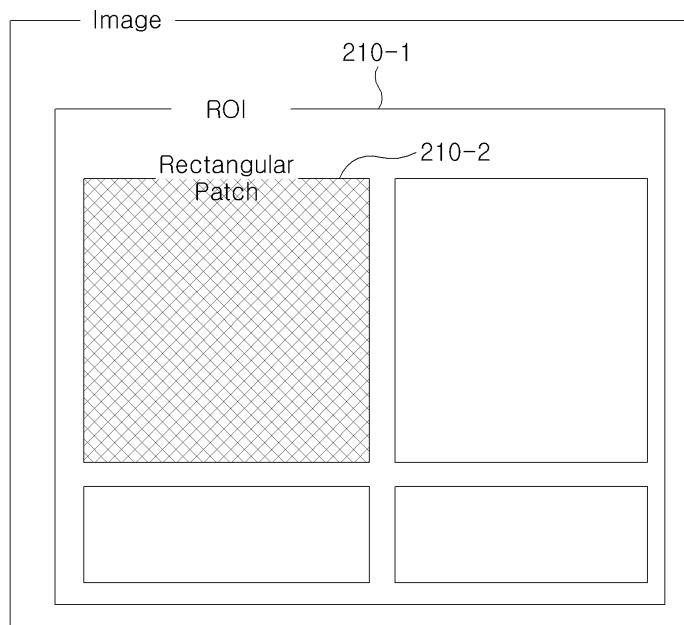
도면1



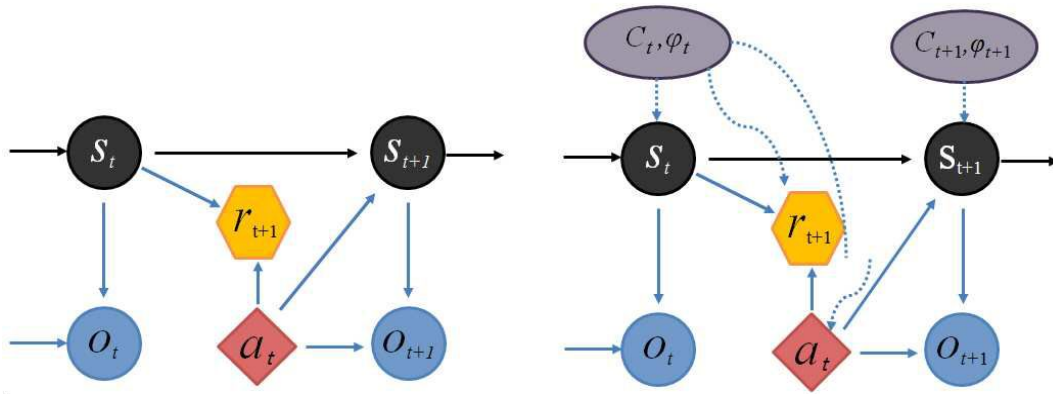
도면2



도면3



도면4



도면5

