

学校编码：10384

学号：23020071151261

廈門大學

硕士学位论文

基于显著性的运动目标检测与跟踪

Object Detection and Tracking Based on
Visual Saliency

胡妙君

指导教师：李翠华

专业名称：计算机应用技术

答辩日期：2010年6月

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下，独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果，均在文中以适当方式明确标明，并符合法律规范和《厦门大学研究生学术活动规范(试行)》。

另外，该学位论文为()课题(组)的研究成果，获得()课题(组)经费或实验室的资助，在()实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称，未有此项声明内容的，可以不作特别声明。)

声明人(签名)：

年 月 日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文(包括纸质版和电子版)，允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

()1. 经厦门大学保密委员会审查核定的保密学位论文，于
年 月 日解密，解密后适用上述授权。

()2. 不保密，适用上述授权。

(请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。)

声明人(签名)：

年 月 日

摘 要

基于视频的目标检测与跟踪是计算机视觉领域热点的研究方向之一，它在智能视频监控、军事侦察监视、交通管理和无人驾驶等领域有着广泛的应用，并发挥着举足轻重的作用。

在机器视觉中，一般的视频跟踪技术需要在第一帧手动地标记出运动目标。本文针对这一问题，研究如何让机器自动发现显著物并进行跟踪：利用视觉显著性对目标进行检测，通过词袋模型形成对运动目标的观测，结合粒子滤波跟踪算法对运动目标进行跟踪。主要的研究工作及创新点如下：

1. 提出一种基于多线索视觉显著性融合的运动目标检测算法。利用中央周边差异显著性来检测局部对比度强的显著区域，利用谱残差显著性检测图像在空间域上的显著区域，利用动态显著性来检测具有运动信息的显著区域。基于这三种显著性融合的目标检测算法，查全率高、误检率低，能够更快更准地定位到运动目标。
2. 提出一种基于词袋模型的运动目标识别算法。算法针对交通场景下的行人和车辆，利用词频直方图表示这两类前景目标，通过朴素贝叶斯分类器形成运动目标的观测器，可以有效地判断出运动目标所属的类别。
3. 设计一个基于粒子滤波的运动目标跟踪算法。这个算法是一个全自动的过程，采用先检测再识别后跟踪的模式对运动目标进行跟踪：利用视觉显著性发现目标、利用词袋模型判断目标、利用显著性结合粒子滤波跟踪目标。实验表明，算法可以很好地跟踪到显著目标，而且与目标实际位置非常逼近。

关键词：显著性特征；词袋模型；粒子滤波跟踪

Abstract

Video-based target detection and tracking is one of the research hotspots in the field of computer vision. It plays a very important role in many applications, such as smart surveillance, military reconnaissance and surveillance, traffic management and auto driving.

In machine vision, tracking always needs to label the object by human on the first frame. According to this problem, this thesis researchs how to make the machine find the salient object and track it automaticly. Detect object baesd on the salient features. Recognize object based on bag-of-words model. And track object based on the salient features using the particle filter. The main researches and innovations are as follows:

1. We propose an object detection algorithm baesd on many clues of visual saliency. Includeing the center surround different saliency which is used to detect the local contrast regions, spectral residual saliency which is used to detect the regions in spatial domain, and dynamic saliency which is used to detect the motion regions. This algorithm is effective with high recall and low false discovery rate.

2. We propose an object classification algorithm based on bag-of-words model. The algorithm focuses on the vehicles and pedestrians. Represent the front objects by the histogram of words. Recognize the objects by the naïve Bayesian classifier. This algorithm could help us to recognize the objects detected effectively.

3. We propose an object tracking algorithm baed on the particle filter. This algorithm is qutomatic by detection-recognition-tracking model. Detect object baesd on the salient features. Recognize object based on bag-of-words model. And track object based on the salient features using the particle filter. From the experiments, we know this algorithm could track the salient object effectively, and

is very close the object.

Keywords: Salient features; bag-of-words model; tracking based on particle filter

厦门大学博硕士学位论文摘要库

参考资料

- [1] 郑南宁. 计算机视觉与模式识别. 国防工业出版社[M]. 1998.
- [2] Spagnolo, P. and Orazio, T.D. and Leo, M. and Distante, A. Moving object segmentation by background subtraction and temporal analysis [J]. *Image and Vision Computing*, 2006, 24: 411 – 423.
- [3] Yan-bing, H. the Moving Object Detection and Tracking Based on Video Sequence Images [M]. M.E. Thesis, Dalian University of Technology, 2008.
- [4] Wren, C. and Azarbayejani, A. and Darrell, T. and Pentland, A. Real-Time Tracking of the Human Body [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1997, 19(7): 780-785.
- [5] Stauffer, C. and Grimson, WEL. Learning Patterns of Activity Using Real-Time Tracking [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, 22(8): 747-757.
- [6] Morellas, V. and Pavlidis, I. and Tsiamyrtzis, P. Deter: detection of events for threat evaluation and recognition [J]. *Machine Vision and Applications*, 2003, 15(1): 29-45.
- [7] 王华伟, 吴琦颖, 白建华, 李翠华. 一种新的背景模型建立及更新方法[J]. *厦门大学学报*. 2005, 44(6): 774-777.
- [8] Kim, K. and Chalidabhongse, T.H. and Harwood, D. and Davis, L. Real-time foreground – background segmentation using codebook model [J]. *Real-Time Imaging*, 2005, 11: 172 – 185.
- [9] Oral, M. and Deniz, U. Centre of mass model – A novel approach to background modeling for segmentation of moving objects [J]. *Image and Vision Computing*, 2007, 25(8): 1365-1376.
- [10] Lipton, A.J. and Fujiyoshi, H. and Patil, R.S. Moving target classification and tracking from real-time video [A]. *Fourth IEEE Workshop on Applications of Computer Vision*, 1998. WACV'98, 1998: 8-14.
- [11] Horn, B.K.P. and Schunck, B.G. Determining optical flow [J]. *Artificial intelligence*, 1981, 17: 185-203.
- [12] Youfu, W. and Jun, S. Properties of Orthogonal Gaussian-Hermite Moments and Their Applications [J]. *EURASIP Journal on Applied Signal Processing*, 2005, 4: 588-599.
- [13] Marr, D. *Vision* [M], 1982. 中译本: 视觉计算理论[M], 姚国正、刘磊、汪云九译, 科学出版社, 1988.
- [14] 侯志强, 韩崇昭. 视觉跟踪技术综述[J]. *自动化学报*, 2006, 32(4): 603-617.
- [15] Marfil, R. and Molina-Tanco, L. and Rodriguez, J.A. and Sandoval, F. Real-time object tracking using bounded irregular pyramids [J]. *Pattern Recognition Letters*, 2007, 28: 985-1001.
- [16] Babu, R.V. and Perez, P. and Bouthemy, P. Robust tracking with motion estimation and local Kernel-based color modeling [J]. *Image and Vision Computing*, 2007, 25(8): 1205-1216.
- [17] Dewan, M. and Hager, G.D. Toward Optimal Kernel-based Tracking [C]. *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2006, 1.
- [18] Matthews, I. and Ishikawa, T. and Baker, S. The template update problem [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, 26(6): 810-815.
- [19] Roh, M.C. and Kim, T.Y. and Park, J. and Lee, S.W. Accurate object contour tracking based on boundary edge selection [J]. *Pattern Recognition*, 2007, 40: 931-943.
- [20] Park, S.C. and Lim, S.H. and Sin, B.K. and Lee, S.W. Tracking non-rigid objects using probabilistic Hausdorff distance matching [J]. *Pattern Recognition*, 2005, 38(12): 2373-2384.
- [21] Gui, C.L.C.X.C. and Fox, M.D. Level set evolution without re-initialization: A new variational formulation [C]. *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005, 1: 430-436.
- [22] Sum, K.W. and Cheung, P. Boundary vector field for parametric active contours [J]. *Pattern Recognition*, 2007, 40(6): 1635-1645.
- [23] Caselles, V. and Cattedo, F. and Coll, T. and Dibos, F. A geometric model for active contours in image processing [J]. *Numerische Mathematik*, 1993, 66(1): 1-31.
- [24] Malladi, R. and Sethian, J.A. and Vemuri, B.C. Shape modeling with front propagation: a level set approach [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1995, 17(2): 158-175.
- [25] Tissainayagam, P. and Suter, D. Object tracking in image sequences using point features [J]. *Pattern*

Recognition, 2005, 38:105-113.

- [26] McKenna, S.J. and Nait-Charif, H. Tracking human motion using auxiliary particle filters and iterated likelihood weighting [J]. *Image and Vision Computing*, 2007, 25(6):852-862.
- [27] Juan, J. and Pantrigo, Angel, S. and JoseMira. On knowledge modelling of the Visual Tracking task [J]. *Expert Systems with Applications: An International Journal*, 2008, 35(1-2):69-81..
- [28] Wang, Y. and Loe, K.F. and Wu, J.K. A Dynamic Conditional Random Field Model for Foreground and Shadow Segmentation [J]. *IEEE transactions on pattern analysis and machine intelligence*, 2006, 28(2):279-290.
- [29] Munoz-Salinas, R. and Aguirre, E. and Garcia-Silvente, M. People detection and tracking using stereo vision and color [J]. *Image and Vision Computing*, 2007, 25:995-1007.
- [30] Lippiello, V. and Siciliano, B. and Villani, L. Adaptive extended Kalman filtering for visual motion estimation of 3D objects [J]. *Control Engineering Practice*, 2007, 15:123-134.
- [31] Julier, S.J. and Uhlmann, J.K. A New Extension of the Kalman Filter to Nonlinear Systems [C]. *Int. Symp. Aerospace/Defense Sensing, Simul. and Controls*, 1997, 3068:182-193.
- [32] Kamel, H. and Badawy, W. A Real-Time Multiple Target Tracking Algorithm Using Merged Probabilistic Data Association Technique and Smoothing Particle Filter [C]. *2006 IEEE Conference on Radar*, 2006:24-27.
- [33] Koch, C. and Ullman, S. Shifts in selective Visual attention: towards the underlying neural circuitry [J]. *Human Neurobiology*, 1985, 4(4):219-227.
- [34] Itti, L. and Koch, C. and Niebur, E. A model of saliency-based visual attention for rapid scene analysis [J]. *IEEE Trans. on Pattern Analysis and Machine Intelligence*. 1998, 20(11):1254-1259.
- [35] Itti, L. and Koch, C. Computational modeling of visual attention [J]. *Nature Reviews Neuroscience*. 2001, 2(3):194-230.
- [36] Milanese, R. Detecting salient regions in an image: from biological evidence to computer implementation [M], University of Geneva. 1993.
- [37] Viola, P. and Jones, M. Rapid object detection using a boosted cascade of simple features [C]. *Proc. IEEE CVPR 2001*, 2001:511 – 518.
- [38] Papageorgiou, C. and Poggio, T. A trainable object detection system: Car detection in static images [J]. *International Journal of Computer Vision*, 2000, 38(1):15 – 33,
- [39] Knudsen, E.I. Fundamental Components of Attention [M]. *Annual Review of Neuroscience*. 2007, 30:57-58.
- [40] Achanta, R. and Estrada, F. and Wils, P. and Susstrunk, S. Salient region detection and segmentation [C]. *International Conference on Computer Vision Systems (ICVS)*, 2008, 5008:66 – 75.
- [41] Bay, H. and Tuytelaars, T. and Van Gool, L. SURF: Speeded Up Robust Features [J]. *Computer Vision and Image Understanding (CVIU)*, 2008, 110(3): 346-359.
- [42] Hou, X. and Zhang, L. Saliency detection: A spectral residual approach [C]. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR07)*. 2007:1-8.
- [43] 张焱, 张志龙, 沈振康, 一种融入运动特征的显著性特征提取方法 [J], *国防科技大学学报*, 2008, 30(3):109-115.
- [44] Jia, Y. and Zhang, C. Front-view Vehicle Detection by Markov Chain Monte Carlo Method [J]. *Pattern Recognition*, 2009, 42(3):313-321.
- [45] Baumberg, A. Reliable feature matching across widely separated views [C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2000:774 – 781.
- [46] Schaffalitzky, F., and Zisserman, A. Multi-view matching for unordered image sets, or “ How do I rganize my holiday snaps? ” [C]. *Proceedings of the 7th European Conference on Computer Vision*, 2002:414 – 431.
- [47] Schmid, C. and Mohr, R. Local gray-value invariants for image retrieval [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997, 19(5):530 – 534.
- [48] Ferrari, V., Tuytelaars, T., and Van Gool, L. Simultaneous object recognition and segmentation by image exploration [C]. *Proceedings European Conference on Computer Vision*, 2001:40 – 54.
- [49] Grabner, H. and Grabner, M. and Bischof, H. Real-time Tracking via On-line Boosting [C]. *Proceedings*

British Machine Vision Conference, 2006,1:47-56.

[50] Sivic, J., and Zisserman, A. Video google: A text retrieval approach to object matching in videos [C].

Proceedings of the International Conference on Computer Vision, 2003: 1470-1477.

[51] Brown, M. and Lowe, D. Recognizing panoramas [C]. Proceedings of the International Conference on Computer Vision, 2003:1218 – 1225.

[52] Csurka, G. and Dance, C. and Fan, L. and Willamowski, J. and Bray, C. Visual categorization with bags of keypoints [C]. Workshop on Statistical Learning in Computer Vision, ECCV, 2004, 1:22.

[53] Mikolajczyk, K. and Schmid, C. Scale & affine invariant interest point detectors [J]. International Journal of Computer Vision, 2004, 60(1):63 – 86.

[54] Mikolajczyk, K. and Schmid, C. Indexing Based on Scale Invariant Interest Points [C]. Proc. Eighth Int'l Conf. Computer Vision, 2001:525-531.

[55] Lindeberg, T. and Garding, J. Shape-adapted smoothing in estimation of 3-D shape cues from affine deformations of local 2-D brightness structure [J]. Image and Vision Computing, 1997,15(6):415 – 434.

[56] Lowe, D. Distinctive Image Features from Scale-Invariant Key-points [J]. Int'l J. Computer Vision, 2004,2(60):91-110.

[57] Harris, C. and Stephens, M. A combined corner and edge detector [C]. Alvey Vision Conference, 1988:147 – 151.

[58] Mikolajczyk, K. and Tuytelaars, T. and Schmid, C. and Zisserman, A. and Matas, J. and Schaffalitzky, F. and Kadir, T. and Gool, L.V. A Comparison of Affine Region Detectors [J]. International Journal of Computer Vision, 2005,65(1): 43-72.

[59] Matas, J. Chum, O. Urban, M. and Pajdla, T. Robust wide-baseline stereo from maximally stable extremal regions [C]. Proceedings of the British Machine Vision Conference, 2002:384 – 393.

[60] Mikolajczyk, K. and Schmid, C. A performance evaluation of local descriptors [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27:1615 – 1630.

[61] Lazebnik, S. Schmid, C. and Ponce, J. A sparse texture representation using affine-invariant regions [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2003:319 – 324.

[62] Ojala T. Pietikainen M. and Mäkelä T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns [J], IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002,24(7):971 – 987.

[63] Lewis, D. Naïve (Bayes) at Forty: The Independence Assumption in Information Retrieval [C]. Machine Learning: ECML-98, 1998:4-15.

[64] Blei, D.M. and Ng, A.Y. and Jordan, M.I. Latent Dirichlet Allocation [J]. Journal of Machine Learning Research, 2003,3:993-1022.

[65] Cortes, C. and Vapnik, V. Support-vector Networks [J]. Machine Learning, 1995,20:273 – 297.

[66] Breiman L. Random forests [J]. Machine Learning, 2001, 45(1):5-32

[67] Fei-Fei, L. Fergus, R. and Perona, P. Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories. Computer Vision and Pattern Recognition Workshop of Generative Model Based Vision. 2004.

[68] Shivani, A. Aatif, A. and Dan, R. Learning to detect objects in images via a sparse, part-based representation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 26(11):1475-1490, 2004.

[69] Leibe, B. Leonardis, A. and Schiele, B. Robust object detection with interleaved categorization and segmentation [J]. International Journal on Computer Vision, 2008, 77(1-3):259-289.

[70] Isard, M. and Blake, A. Condensation—conditional density propagation for visual tracking [J]. International Journal on Computer Vision, 1998, 29(1):5-28.

[71] Gordon, N. and Salmond, D. and Ewing, C. Bayesian state estimation for tracking and guidance using the bootstrap filter [J]. Journal of Guidance, Control and Dynamics, 1995, 18(6):1434-1443.

[72] Kitagawa, G. Monte Carlo and smoother for non-Gaussian nonlinear state space models [J]. Journal of

Computational and Graphical Statistics, 1996, 5(1):1-25.

[73] Arulampalam, MS and Maskell, S. and Gordon, N. and Clapp, T. and Sci, D. and Organ, T. and Adelaide, SA A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking [J]. IEEE Transactions on Signal Processing, 2002, 50(2): 174-188.

[74] Bergman, N. Recursive Bayesian estimation: Navigation and tracking applications [M]. Ph.D. dissertation, Linköping University, 1999

厦门大学博硕士论文摘要库

Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.

厦门大学博硕士论文摘要库