www.ijsrnsc.org

Available online at www.ijsrnsc.org

IJSRNSC

Volume-5, Issue-3, June 2017
Review Paper

Int. J. Sc. Res. in Network Security and Communication

ISSN: 2321-3256

Object Detection and Tracking using Congnitive Approach

Mahajan J.R[1*], C.S. Rawat[2]

[1*]Department of ETE, Pacific University, Udaipur, India
[2]Department of ETE, Vivekanand Institute of Technology, Chembur, India

Corresponding Author: mahjayant@gmail.com

Received 10th Mar 2017, Revised 25th Apr 2017, Accepted 18th May 2017, Online 30th Jun 2017

Abstract - In the world of computer vision the object tracking is biggest challenge.The performance is suscaptible to various parameters such as occlusion,background clutter,change in illumination and scale variation.The development of high powered computers,the availability of high quality and inexpensive video cameras and increase in automated ,video analysis aid the purpose of object detection.Three key steps in video analysis are detection of moving objects,tracking of such objects from frame to frame and analysis of object to recognize their behavior.Different approaches have been proposed for object tracking.This paper combines tracking methods from braod categories that provide a unique solution to the widely applied object and tracking problem.

Key Word - Object detection, Object tracking, PCA, LSR Camera

I. INTRODUCTION

A video is sequences of images.These images are called a frame.The frames displayed fast enough so that human eyes can percieve and the link its content. Hence all image processing methods can be applied to individual frames. The contents of two consecutive frames are close with each other.

Visual content can be shaped by a series of concept. At the first level are the raw pixels with color or brightness information. Further processing takes features such as edges, corners, lines, curves, and color regions. A higher abstraction layer may combine and interpret these features as objects and their dimension. At the highest level are the human level concepts involving one or more objects and relationships among them.

Object detection in videos involves verifying the presence of an object in image sequences and possibly locating it precisely for recognition. Object tracking is to monitor an object spatial and temporal changes during a video sequence, including its presence, position, size, shape, etc. This is done by solving the temporal correspondence problem, the problem of matching the target region in successive frames of a sequence of images taken at closely-spaced time intervals. These two processes are closely related because tracking usually starts with detecting objects, while detecting an object repeatedly in subsequent image sequence is often necessary to help and verify tracking.

There are many approaches in object tracking, they are mainly varied from each other depending upon (i) object representation for suitable for tracking , (ii) image features (iii) The motion, appearance, and (iv) shape of the object be modeled.
The aptitude to constantly detect and track human motion is a valuable means for high-level application that depend on image input. Interact with human and accepting their activities is one of core problems in intelligent systems, such as human-computer interaction and robotics. An algorithm for human motion detection digests high-bandwidth video into a compact description of the human presence in that scene. This high-level description can then be put to use in other applications.

Some examples of applications that could be realized with reliable human motion detection and tracking are:

- Automated surveillance for security-conscious venues
- Human interaction for mobile robotics
- Safety devices for pedestrian detection on motor vehicles
- Automatic motion capture for film and television

With the availability of economical and high motion cameras, and the impressive increase in the information processing capabilities of computers, more and more visual sensors are being addicted up to computers. Now a day Robotics, visual surveillance, manufacturing and medicine are few domains where such addicted can be seen most commonly Robotics is a major field with endless possibilities for computer vision. Vision based tracking systems in robotics have been used for various tasks ranging from robot soccer to docking of space craft's. Robots must interact with the environment to perform various tasks. Even more so for a mobile robot as it needs to negotiate obstacles and perform its tasks. Therefore sensing

mechanisms are required for these robots to track obstacles and other moving objects.

II. LITERATURE REVIEW

D. Du, et al [1] have proposed new and high-octane method for online change in object tracking. Apart from most of all present methods, this method works on higher order structural dependences of different parts of the tracking target in multiple consecutive frames. It is necessary to build a structure aware hyper-graph to getting such higher order dependences, and solve the tracking problem by searching dense sub graphs on it. A new evaluating data set for online deformable object tracking, which includes 50 challenging sequences with full annotations that represent realistic tracking challenges, such as deformations and severe occlusions.

X. Zhang, et al [2] propose a tracking algorithm that is robust and which fuses information from both visible images and infrared (IR) images. This tracking algorithm not only incorporates convolutional feature maps from the visible channel, but also employs a scale pyramid representation from IR channel. It estimates the target location by fusing multilayer convolutional feature maps. It also predicts the target scale from a scale pyramid.

M. Stommel, et al [3] planned a spatiotemporal segmentation of key-points provided by a skeletonization of depth contours. A vector-shaped pose descriptor allows for the retrieval of similar poses and is easier to use with many machine learning libraries. A visualization method based on the Hilbert curve provides valuable insight in the detected poses.

H. Liu, et al [4] proposes a new robust object tracking method based on the Principal Component Analysis (PCA) and Local Sparse Representation (LSR). First, candidates are reconstructed through the PCA subspace model in global manner. To handle occlusion, a patch-based similarity estimation strategy is proposed for the PCA subspace model. In the patch-based strategy, the PCA representation error map is divided into patches to estimate the similarity between target and candidate considering the occlusion. Second, the LSR is introduced to detect the occluded patches of the object and estimate the similarity through the residual error in the sparse coding. Finally, the two similarities of each candidate from the PCA subspace model and LSR model are fused to predict the tracking result.

W. Hu, et al [5] proposed a tracking algorithm based on a multi-feature joint sparse representation. The templates for the sparse representation can include pixel values, textures, and edges. In the multi-feature joint optimization, noise or occlusion is dealt with using a set of trivial templates. A sparse weight constraint is introduced to dynamically select the relevant templates from the full set of templates. A variance ratio measure is adopted to adaptively adjust the weights of different features. The multi-feature template set is updated adaptively.

Y. Yuan, et al [6], presented a novel visual object tracking algorithm based on the Observation Dependent Hidden Markov Model (OD-HMM) framework. The observation dependency is computed by Structure Complexity Coefficients (SCC) which is defined to predict the target appearance change. Different from conventional methods addressing the appearance change problem by investigating different online appearance models, this problem is handled by addressing the fundamental reason of motion-related appearance change during visual tracking. Based on the analysis of motion-related appearance change, this work investigates the relationship between the structure of the object surface and the appearance stability.

X. Sun, et al [7] presented a novel approach to non-rigid objects contour tracking based on a supervised level set model (SLSM). In contrast to most existing trackers that use bounding box to specify the tracked target, the proposed method extracts the accurate contours of the target as tracking output, which achieves better description of the non-rigid objects while reduces background pollution to the target model. The SLSM can ensure a more accurate convergence to the exact targets in tracking applications.

T. Suwannatat, et al [8] proposed a visual tracking approach that detects unexpected moving human that appears in the scene of a monocular camera. This method is implemented using a novel combination of multi-features tracking algorithm based on the particle filter, which allows robust and accurate visual tracking under the circumstance of real-time visual tracking. This work introduce multi-features probability framework which combines the advantages of color feature-based tracking and contour feature-based tracking. By taking advantage of the unique feature and properties of the color and contour feature of humans, it hope to overcome the inherent disadvantages of each, resulting in a combined feature which is more effective than either feature is individually. The accuracy and robustness were evaluated and compared on a challenging synthetic tracking problem using real visual tracking experiments. As a result, it is be possible to track robustly human's motion in the complex environment with moving camera.

X.Zhang et al [9] proposed an efficient approach to human pose estimation in static images and human pose tracking in video sequences. In this work, the human body is modeled as a three-level tree structure and the estimation or tracking process is formulated as a Bayesian inference problem. The tree structure state space is carefully parsed into a lexicographic order using an appropriate grammar, and searched by the data driven Markov Chain Monte Carlo (DDMCMC) technique.

O.Zoidi et al [10] proposed a visual object tracking framework, which employs an appearance-based representation of the target object, based on local steering

kernel descriptors and color histogram information . This framework takes as input the region of the target object in the previous video frame and a stored instance of the target object, and tries to localize the object in the current frame by finding the frame region that best resembles the input. As the object view changes over time, the object model is updated, hence incorporating these changes. Color histogram similarity between the detected object and the surrounding background is employed for background subtraction.

J.Martinez del Rincon et al [11] proposed a novel framework for visual tracking of human body parts is introduced. The approach presented demonstrates the feasibility of recovering human poses with data from a single uncalibrated camera by using a limb-tracking system based on a 2-D articulated model and a double-tracking strategy. Its key contribution is that the 2-D model is only constrained by biomechanical knowledge about human bipedal motion, instead of relying on constraints that are linked to a specific activity or camera view. These characteristics make this approach suitable for real visual surveillance application.

X.Zhao et al [12] proposed a novel online sparse Gaussian Process (GP) regression model to recover 3-D human motion in monocular videos. Particularly, this work investigates the fact that for a given test input. Its output is mainly determined by the training samples potentially residing in its local neighborhood and defined in the unified input-output space. This leads to a local mixture GP experts system composed of different local GP experts, each of which dominates a mapping behavior with the specific covariance function adapting to a local region. To handle the multimodality, the methods combine both temporal and spatial information therefore to obtain two categories of local experts. The temporal and spatial experts are integrated into a seamless hybrid system, which is automatically self-initialized and robust for visual tracking of nonlinear human motion. Learning and inference are extremely efficient as all the local experts are defined online within very small neighborhoods.

Mun Wai Lee et al [13] proposed a three-stage approach with multilevel state representation that enables a hierarchical estimation of 3D body poses. This method addresses various issues including automatic initialization, data association, self and inter-occlusion. At the first stage, humans are tracked as foreground blobs, and their positions and sizes are coarsely estimated. In the second stage, parts such as face, shoulders, and limbs are detected using various cues, and the results are combined by a grid-based belief propagation algorithm to infer 2D joint positions. The derived belief maps are used as proposal functions in the third stage to infer the 3D pose using data-driven Markov chain Monte Carlo.

R.Horaud et al [14] proposed a method that iteratively computes two things: maximum likelihood estimates for both the kinematic and free-motion parameters of a kinematic human-body representation, as well as probabilities that the data are assigned either to a body part or to an outlier class. This approach introduce a new metric between observed points and normal on one side and a parameterized surface on the other side, the latter being defined as a blending over a set of ellipsoids. This metric is well suited when one deals with either visual-hull or visual-shape observations. This work illustrates the method by tracking human motions using sparse visual-shape data (3D surface points and normals) gathered from imperfect silhouettes.

Y.Lu et al [15] proposed a cooperative hybrid visual tracking system for event detection and human tracking. This system is composed of a stationary camera and a Pan-tilt-zoom (PTZ) camera. The stationary camera has a wide field of view and is attentive to the scene for event detection. The PTZ camera is activated if an event is detected by the stationary camera. It then pans and tilts to center the target in its view and zooms in to obtain identifying details of the target that may not be clear to the stationary camera. Such a system could be used for attentive surveillance in various locations.

Q.Cai et al [16] presented a comprehensive framework for tracking coarse human models from sequences of synchronized monocular grayscale images in multiple camera coordinates. It demonstrates the feasibility of an end-to-end person tracking system using a unique combination of motion analysis on 3D geometry in different camera coordinates and other existing techniques in motion detection, segmentation, and pattern recognition. The system starts with tracking from a single camera view. When the system predicts that the active camera will no longer have a good view of the subject of interest, tracking will be switched *to* another camera which provides a better view and requires the least switching to continue tracking. The non-rigidity of the human body is addressed by matching paints of the middle line of the human image. Spatially and temporally, using Bayesian classification schemes. Multivariate normal distributions are employed to model class-conditional densities of the features for tracking, such as location, intensity, and geometric features. Limited degrees of Occlusion are tolerated within the system.

Feng et al [17], proposed a novel robust Probability Hypothesis Density (PHD) filter for multiple human tracking, where it employs a Student's-t distribution for the state and the observation models. This distribution has an advantage over the traditional Gaussian distribution in the sense that its tail is heavier and can potentially cover more widely-spread particles. In an enclosed environment, accurate measurement of the humans can be difficult to obtain due to illumination and posture changes. Therefore, this algorithm employ a one class support vector machine (OCSVM) to calculate the weights for the particle based PHD filter, which utilizes the color and oriented gradient

histogram features due to their accuracy in describing the human targets. The OCSVM is shown to be robust against background noise in the measurement due to the difference between human target and noises features.

X.T.Truong et al [18] proposed an effective socially aware navigation framework for mobile service robots in social environments. This proposed framework consists of three stages. In the first stage, RGB-D and laser data fusion-based human detection and tracking system is utilized to detect humans in the vicinity of the robot. In the second stage of framework, the extended personal space is modeled by using the human states including position and motion, and the relative motion between the human and the robot. In the third stage, the extended personal space is incorporated into the motion planning system and then a kinodynamic RRT motion planner is made use of to generate a legible trajectory of the mobile robot. The experimental results indicate that the proposed framework is able to ensure the safety of human, providing socially acceptable behaviors of the mobile service robot.

L.Wang et al [19] presented a tracklet based data association approach for multiple human tracking in surveillance scenarios, with the assumptions that the camera is static, people walk on a ground plane and camera parameters can be obtained. Unlike most of the previous data association works that only consider how to ensure correct linking, authors also attempt to improve the detections when reliable temporal information can be obtained. To this end, it first generate tracklets by conservative linking of detections, and extract the appearance, size and position information of those reliable detections that show high temporal and spatial consistency. Then the extracted information is propagated to detections within the tracklets by refining the detections' shape models. After that, local conservative tracklet association based on the Hungarian algorithm is performed so that reliable temporal information can be further propagated. The iteration stops when there are no new detection updates or new tracklet association. Finally, the Hungarian algorithm is applied globally to resolve ambiguous situations.

M.Gupta [20] aims at developing a robust vision-based algorithm using point-based features, like Speeded up robust features (SURF), which can track a human under challenging conditions including variation in illumination, pose change, full or partial occlusion and abrupt camera motion. Since the point based methods use tracking-by-detection framework, the major problem lies in finding sufficient number of descriptors in subsequent frames as the target undergoes the above mentioned variations. This looks into the problem of constructing an object model which can evolve over time to deal with short-term changes while maintaining stability on a longer term. The object model is updated by propagating useful descriptors from past templates onto the current template using affine transformation. An Support Vector Machine (SVM) classifier along with a Kalman Filter predictor is used to differentiate between a case of pose change and a case of occlusion (partial/full). An attempt is also made to detect pose change due to *out-of-plane rotations* which is a difficult problem and lead to frequent tracking failures.

H.Wang et al [21] considered the challenging problems of tracking, and focus on constructing robust and salient human action trajectories. To this goal, a novel motion tracking method is proposed, in which the Scale Invariant Feature Transform (SIFT) key-points are separately tracked at each spatial scale by the technique of dense optical flow, and camera motion elimination is employed to construct robust trajectories. The HOG, HOF and MBH descriptors are used to form the feature descriptor of trajectories. In order to avoid complexity involved in high dimensional space and further reduce noise features, Principal Component Analysis (PCA) is utilized to trim out the redundancy of feature vectors. Then, the Fisher vector model is applied to aggregate HOG, HOF and MBH descriptors, and the linear Support Vector Machine (SVM) is employed to classify human actions.

K.H.Lee et al [22] proposed a robust ground-moving platform- based human tracking system, which effectively integrates visual simultaneous localization and mapping (V-SLAM), human detection, ground plane estimation, and kernel-based tracking techniques. The proposed system systematically detects humans from recorded video frames of a moving camera and tracks the humans in the V-SLAM-inferred 3-D space via a tracking-by-detection scheme. By taking advantage of the appearance model and 3-D information, the proposed system not only achieves high effectiveness but also well handles occlusion in the tracking.

P.Feng et al [23] proposed a robust particle probability hypothesis density (PHD) filter where the variational Bayesian method is applied in joint recursive prediction of the state and the time varying measurement noise parameters. The proposed particle PHD filter is based on forming variational approximation to the joint distribution of states and noise parameters at each frame separately; the state is estimated with a particle PHD filter and the measurement noise variances used in the update step are estimated with a fixed point iteration approach. A deep belief network (DBN) is used in the update step to mitigate the effect of measurement noise on the calculation of particle weights in each frame.

P.Feng et al [24] proposed a Gaussian based social force model is employed to build a posterior distribution within the prediction stage of the particle PHD filter. In order to identify the new born targets and form the measurement set, a background subtraction step is employed to detect the targets in each frame.

G.Yang [25] proposed an indoor human tracking system based on wireless network, WiLocus. They use the

stationary smart appliances indoor to sensor users' moving behaviors. According to a series of detected moving behaviors, this system then identifies the actual moving trajectory of a user. Because of multipath, human's moving behaviors effect the wi-fi signals and it can use *Channel State Information* (CSI) to detect them. In this work, there are three moving behaviors crucial to construct a moving trajectory, *Pass-Device*, *Pass-Room*, *Enter/Exit-room*. These moving behaviors are able to be identified using machine learning technology.

III. CONCLUSION

The object tracking problem in video is defined as finding the location of a given change surface in the current frame.Usually the change surface is specified in the previous frame or in an intial image frame.The object is tracked in the video by keeping track of the locations of the change surface over the video which consist of image frame.

REFERENCES

[1]. D. Du, H. Qi, W. Li, L. Wen, Q. Huang and S. Lyu, "*Online Deformable Object Tracking Based on Structure-Aware Hyper-Graph*", in IEEE Transactions on Image Processing, vol. 25, no. 8, pp. 3572-3584, 2016.

[2]. X. Zhang, Y. Yuan and X. Lu, "*Deep object tracking with multi-modal data*", 2016 International Conference on Computer, Information and Telecommunication Systems (CITS), Kunming, pp. 1-5, 2016.

[3]. M. Stommel, M. Beetz and W. Xu, "*Model-Free Detection, Encoding, Retrieval, and Visualization of Human Poses From Kinect Data*", in IEEE/ASME Transactions on Mechatronics, vol. 20, no. 2, pp. 865-875, 2015.

[4]. H. Liu, S. Li and L. Fang, "*Robust Object Tracking Based on Principal Component Analysis and Local Sparse Representation*", in IEEE Transactions on Instrumentation and Measurement, vol. 64, no. 11, pp. 2863-2875, 2015

[5]. W. Hu, W. Li, X. Zhang, and S. Maybank, "*Single and multiple object tracking using a multi-feature joint sparse representation*", IEEE Trans. Pattern Anal. Mach. Intell., Vol.37, Issue.4, pp.816-833, 2015.

[6]. Y. Yuan, H. Yang, Y. Fang, W. Lin, "*Visual Object Tracking by Structure Complexity Coefficients*", in IEEE Transactions on Multimedia, vol. 17, no. 8, pp. 1125-1136, 2015.

[7]. X. Sun, H. Yao, S. Zhang and D. Li, "*Non-Rigid Object Contour Tracking via a Novel Supervised Level Set Model*", in IEEE Transactions on Image Processing, vol. 24, no. 11, pp. 3386-3399, 2015.

[8]. T. Suwannatat, N. Indra-Payoong, K. Chinnasarn, "*Robust human tracking based on multi-features particle filter*", Computer Science and Software Engineering (JCSSE), 2015 12th International Joint Conference on, Songkhla, pp.12-17, 2015,

[9]. X. Zhang, C. Li, W. Hu, X. Tong, S. Maybank and Y. Zhang, "*Human Pose Estimation and Tracking via Parsing a Tree Structure Based Human Model*", in IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 44, no. 5, pp. 580-592, 2014.

[10]. O. Zoidi, A. Tefas and I. Pitas, "*Visual Object Tracking Based on Local Steering Kernels and Color Histograms*", in IEEE Transactions on Circuits and Systems for Video Technology, vol. 23, no. 5, pp. 870-882, 2013.

[11]. J. Martinez del Rincon, D. Makris, C. Orrite Urunuela and J. C. Nebel, "*Tracking Human Position and Lower Body Parts Using Kalman and Particle Filters Constrained by Human Biomechanics*", in IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 41, no. 1, pp. 26-37, 2011.

[12]. X. Zhao, Y. Fu and Y. Liu, "*Human Motion Tracking by Temporal-Spatial Local Gaussian Process Experts*", in IEEE Transactions on Image Processing, vol. 20, no. 4, pp. 1141-1151, 2011.

[13]. Mun Wai Lee, Ramakant Nevatia, "*Human Pose Tracking in Monocular Sequence Using Multilevel Structured Models*", IEEE Transactions on Pattern Analysis & Machine Intelligence, vol. 31, no. , pp. 27-38, 2009

[14]. R. Horaud, M. Niskanen, G. Dewaele and E. Boyer, "*Human Motion Tracking by Registering an Articulated Surface to 3D Points and Normals*", in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 1, pp. 158-163, 2009.

[15]. Y. Lu and S. Payandeh, "*Cooperative hybrid multi-camera tracking for people surveillance*", in Canadian Journal of Electrical and Computer Engineering, vol. 33, no. 3/4, pp. 145-152, 2008.

[16]. Q. Cai and J.K. Aggarwal, "*Tracking Human Motion in Structured Environments Using a Distributed-Camera System*", IEEE Trans. Pattern Recognition and Machine Intelligence, vol. 21, no. 11, pp. 1241-1247, 1999.

[17]. Feng, Pengming, "*A Robust students-t distribution PHD filter with OCSVM updating for multiple human tracking*", 23rd European IEEE Signal Processing Conference (EUSIPCO), France, pp. pp. 2396-2400, 2015.

[18]. X. T. Truong, V. N. Yoong, T. D. Ngo, "*RGB-D and laser data fusion-based human detection and tracking for socially aware robot navigation framework*", 2015 IEEE International Conference on Robotics and Biomimetics (ROBIO), Zhuhai, pp. 608-613, 2015.

[19]. L. Wang, Q. Deng and M. Jia, "*Robust multi-human tracking by detection update using reliable temporal information*", Computer Vision Theory and Applications (VISAPP), 2014 International Conference on, Portugal, pp. 387-396, 2014.

[20]. M. Gupta, S. Kumar, N. Kejriwal, L. Behera and K. S. Venkatesh, "*SURF-based human tracking algorithm for a human-following mobile robot*", Image Processing Theory, Tools and Applications (IPTA), 2015 International Conference on, Orleans, pp. 111-116, 2015.

[21]. H. Wang, Y. Yi, "*Tracking Salient Keypoints for Human Action Recognition*", Systems, Man, and Cybernetics (SMC), 2015 IEEE International Conference on, Kowloon, pp.3048-3053, 2015.

[22]. K. H. Lee; J. N. Hwang; G. Okopal; J. Pitton, "*Ground-Moving-Platform-Based Human Tracking Using Visual SLAM and Constrained Multiple Kernels*", in IEEE Transactions on Intelligent Transportation Systems , vol.17, no.12, pp. 3602-12, 2016.

[23]. P. Feng, W. Wang, S. M. Naqvi, J. Chambers, "*Variational Bayesian PHD Filter with Deep Learning Network Updating for Multiple Human Tracking*", Sensor Signal Processing for Defence (SSPD), 2015, Edinburgh, 2015, pp. 1-5.

[24]. P. Feng, W. Wang, S. M. Naqvi, S. Dlay, J. A. Chambers, "*Social force model aided robust particle PHD filter for multiple human tracking*", 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, pp. 4398-4402, 2016.

[25]. G. Yang, "*WiLocus: CSI Based Human Tracking System in Indoor Environment*", 2016 Eighth International Conference on Measuring Technology and Mechatronics Automation (ICMTMA), Macau, pp. 915-918, 2016.

