International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868
Foundation of Computer Science FCS, New York, USA
International Conference & workshop on Advanced Computing 2013 (ICWAC 2013) – www.ijais.org

Adaptive Automatic Tracking, Learning and Detection of Real-time Objects in the Video Stream

Bhushan Nemade
AP, TCET
Kandivali (E), Mumbai, India

R. R. Sedamkar
Dean Academics,
Head of Department. TCET,
Kandivali (E), Mumbai, India

Abstract
Proposed system presents an automatic long term tracking and learning and detection of real time objects in the live video stream. In this system, Object to be tracked also called as cropped image is defined by its location and the extent in the single frame by selecting the object of interest in the live video. Many existing systems for tracking objects fails due to loss of information caused by complex shapes, rapid motion, illumination changes, scaling and projection of 3D world on 2D image. Proposed modified PN learning algorithm which uses background subtraction technique to increase speed of the frame processing for object detection. Proposed Modified PN learning algorithm considers the object to be tracked as P-Type Object and background is divided into the numbers of N-Type objects. Initially input image is matched with the N-Type of objects for rejection and then with P-type for acceptance. Proposed system uses the Template Matching algorithm to match cropped image with region of interest in the current frame to mark the Object Location. If match is found then Principle Component Analysis algorithm is used for detection of the fast moving object which is the advantage over the existing systems. If match does not found then Proposed Modified PN learning processing is applied to detect the image in rapid motion video. Proposed system uses background subtraction to increase the performance for detection of any moving object as the background remains still and we get approximate location of the moving object. Proposed System is expected to minimize delay for frame processing and reduce average localization errors to improve in matching percentage irrespective of scaling of the input image. Thus proposed system is expected to overcome the drawbacks of existing system for efficient tracking of any real time object.

General Terms
Region of interest, Principle component analysis,

Keywords
Template matching Tracking, Adaptive Learning, Object Detection

1 Existing Work on Object Detection and Tracking
Camshift algorithm uses color feature for real time object tracking. Camshift fails when video is under rapid motion illumination changes and background distraction.[1]. Adaptive Local Search and Kalman Filter are proposed to predict position of Moving object. Kalal has given work TLD framework for tracking and detection and PN learning algorithm to learn about the characteristics of the moving object in video stream [2]. Improved Camshift reduces the effect of illumination interference and judges whether the target is lost. SURF algorithm is used with improved camshift because it is invariant to scale, rotation and translation of Image. Thus lost target can be found out using improved camshift and SURF Algorithm [3]. It gives five parameter set 2-D central location, width, height and orientation of rectangular box as tunable variable in object tracking. Principle component analysis (PCA) also called Eigen object detection is used for matching two objects. [5]

2 Introductions
Existing system has the problems when object changes its appearances or object is moving out of camera view and again coming in front of camera. Tracking fails due to scaling, rotation, illumination changes and does not perform in case of full out of plane rotation.

The proposed system overcomes the above mentioned drawback and system can Track and Learn the Real-time object automatically using Principle component analysis, P-N learning, template matching and Eigen object detection.

Video surveillance plays a crucial role due to security issues involved in various areas like crowded public places, departmental stores, traffic monitoring, banks and boarders between two countries. The System is expected to track and learn the real time objects. Video stream will be processed at frame rate and process should run indefinitely long. The task is called as long term tracking.

In the simplest form, Tracking is defined as is process of continuously finding an object of interest in the video. In other words, a tracker assigns consistent labels to the tracked objects in different frames of a video. Additionally, depending on the tracking domain, a tracker can also provide object-centric information, such as orientation, area, or shape of an object.

Tracker estimates the object motion under the assumption that the object is visible and its motion is limited. A tracker can provide weakly labeled training data for a detector and thus improve it during runtime.[1]

Detector performs full scanning of the image to localize all the appearances that have been observer in the past. A detector can reinitialize a tracker and thus minimize the tracking failures. Detection based algorithms estimates the object location in every frame independently. [1][3]

Detectors do not drift and do not fail if the object disappears from the camera. However, they require an offline training stage. The starting point of my work says that neither tracking nor detection can solve long term tracking task independently. But if they operate simultaneously, there is potential to benefit one from another. [1]

International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868
Foundation of Computer Science FCS, New York, USA
International Conference & workshop on Advanced Computing 2013 (ICWAC 2013) – www.ijais.org

2.1 Tracking Algorithm provides:

1. All tracking algorithms assume that the object motion is smooth with no abrupt changes.

2. The object motion is assumed to be of constant velocity.

3. Prior knowledge about the number and the size of objects, or the object appearance and shape.

Learning: It observes performance of both, the tracker and detector, identifies errors of the detector and generates training examples to avoid these errors in the future.

2.1.1 Two subtasks

1. Build model to track Object.

2. Use of previous frame(s) to predict current frame and restrict the search

Repeat the two subtasks, possibly updating the mode

2.2 P-N Learning Algorithm:

In every frame, the P-N learning performs the following steps:
1) evaluation of the detector on the current frame.

2) estimation of the detector errors using the P-N experts, and
3) update of the detector by labeled examples output by the experts.

2.3 PCA Algorithm:

Three steps are used to reach the goal:

i) Interested clip extraction, the purpose to reduce the computational time of PCA;

ii) PCA, the process to produce PC (Principal Component) images that contain both static and dynamic information on the image sequence.

iii) Track extraction, the step to extract the tracks of moving objects from each PC.

The purpose of PCA is to produce PC (Principal) images that contain both static and dynamic information on the interested clip. Firstly, since the video images normally include three bands (R, G, and B), every band in the interested clip will be assembled to form a single-band image set. Then, PCA is performed on every image set and produce a series of Principal Components for every individual band. Finally, all Principal Components are collected together to become a PC image sequence.[15]

2.3.1 Track Extraction from PC Image

Since PCA is able to put the moving track together in certain PCs, it is the main aim of this step to extract the PCs that contain the tracking pattern. A region growing scheme is adopted to produce segmentation image for each PC (Fig. 1b). Normally three main regions can be observed in a segmented PC image: noise region, background region, and track region. Noise regions generally have small areas, which commonly are too small to be the track of the interested moving object. On the other hand, background regions generally come with relatively large areas and turn out to be the static information. As a result, the rest of the regions will be grouped together to become the track regions (Fig. 1c). In the end of track extraction, we compile the total area of track regions in pixel as the index of each PC image. The PC image with highest index would be the interested PC, which is able to represent the track of moving object.

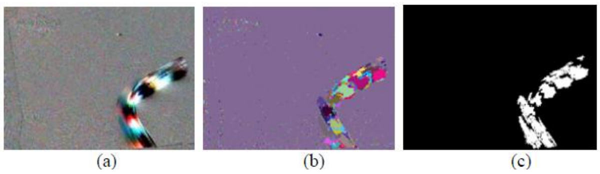

Fig. 1

2.4 Template Matching

Template matching is a technique for finding areas of an image that match (are similar) to a template image (patch).

We need two primary components:

Source image (I): The image in which we expect to find a match to the template image Template image (T): The patch image which will be compared to the template image our goal is to detect the highest matching area:[16]

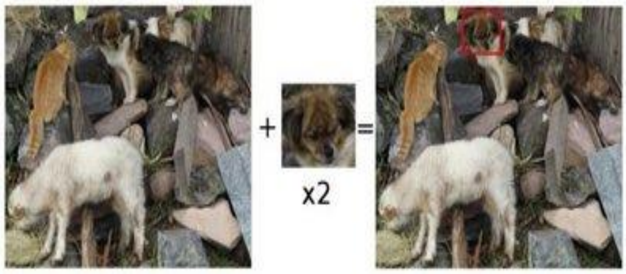

Fig 2 [16]

To identify the matching area, we have to c*ompare* the image against the source image by sliding it:[16]

Fig 3 [16]

By sliding, we mean moving the patch one pixel at a time (left to right, up to down). At each location, a metric is calculated so it represents how “good” or “bad” the match at that location is (or how similar the patch is to that particular area of the source image).[16]

For each location of T over I, you store the metric in the result matrix (R). Each location (x,y) in R contains the match metric:[16]

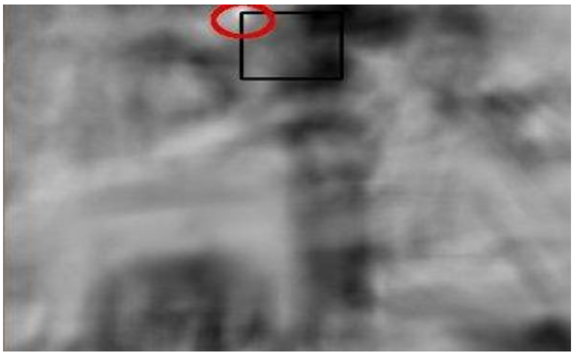

Fig 4

the image above is the result R of sliding the patch with a metric TM_CCORR_NORMED. The brightest locations indicate the highest matches. As you can see, the location marked by the red circle is probably the one with the highest value, so that location (the rectangle formed by that point as a corner and width and height equal to the patch image) is considered the match.

In practice, we use the function minMaxLoc to locate the highest value (or lower, depending of the type of matching method) in the R matrix.

Different Template matching techniques are mentioned below.[16]

1. Method=CV_TM_SQDIFF

$$R(x,y) = \sum_{x',y'} (T(x',y') - I(x+x',y+y'))^2$$

2. Method=CV_TM_SQDIFF_NORMED

$$R(x,y) = \frac{\sum_{x',y'} (T(x',y') - I(x+x',y+y'))^2}{\sqrt{\sum_{x',y'} T(x',y')^2 \cdot \sum_{x',y'} I(x+x',y+y')^2}}$$

2.Method=CV_TM_CCORR

$$R(x,y) = \sum_{x',y'} (T(x',y') \cdot I(x+x',y+y'))$$

4. method=CV_TM_CCORR_NORMED

$$R(x,y) = \frac{\sum_{x',y'} (T(x',y') \cdot I'(x+x',y+y'))}{\sqrt{\sum_{x',y'} T(x',y')^2 \cdot \sum_{x',y'} I(x+x',y+y')^2}}$$

5. method=CV_TM_CCOEFF

$$R(x,y) = \sum_{x',y'} (T'(x',y') \cdot I(x+x',y+y'))$$

Where

$$T'(x',y') = T(x',y') - 1/(w \cdot h) \cdot \sum_{x'',y''} T(x'',y'')$$
$$I'(x+x',y+y') = I(x+x',y+y') - 1/(w \cdot h) \cdot \sum_{x'',y''} I(x+x'',y+y'')$$

6. method=CV_TM_CCOEFF_NORMED

$$R(x,y) = \frac{\sum_{x',y'} (T'(x',y') \cdot I'(x+x',y+y'))}{\sqrt{\sum_{x',y'} T'(x',y')^2 \cdot \sum_{x',y'} I'(x+x',y+y')^2}}$$

3 Evaluation of existing work on Object Detection and Tracking

Object Tracking and learning has following pit falls of existing system

1. When object changes its appearance or object is moving out of camera view and coming in front of camera

2. Long term tracking fails due to rotation, illumination changes, background clutters and operate in real time.

3. TLD does not perform well in case of full out-of-plane rotation

4. Scaling (Zooming)

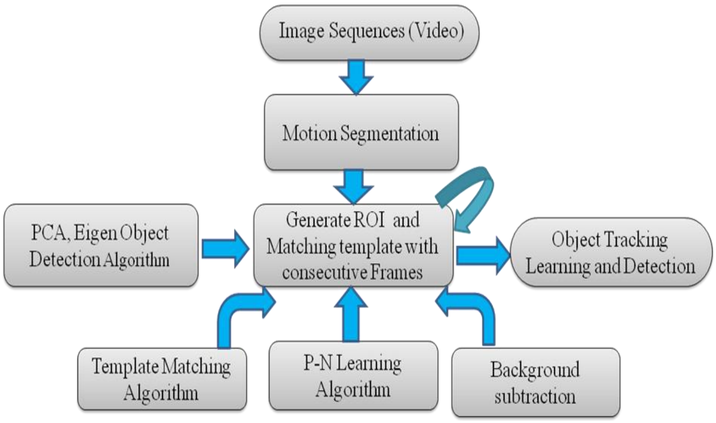

Fig 5 Proposed System Architecture

4 Proposed Algorithm

Step 1: Camera interfacing

Step 2: Take the video streams into the system

Step 3: Object selection (Cropped image) by the user.

Step 4: Creation of image array of size 50 for learning.

Step 5: Store the Cropped object at array index 0.

Step 6: Create ROI at 20 pixel distance. ROI can be increased with the increment of 20 pixels if object is not found in the current ROI.

Step 7: Fetch the next Frame from video stream.

Step 8: Apply the template matching for cropped image for cropped image at the index 0 in the ROI frame to get the Object location. Template matching algorithm to get highest intensity location and mark the object Location.[12] If ROI fails then background subtraction technique can be used. [4][14]

Step 9: If match is found then apply the PCA for tracking the moving object on the video image then Eigen Object Detector to detect the object. Check for percentage matching. If matching percentage is less then store the new object in the learned array at the index 0 and shift other images in the array.[3][4][7][8]

Step 10: If match is not found then Identification of detector errors and learning from it can be done from by P-N learning. P-N learning estimates the errors by pair of experts. P-experts detects missed detection whereas N-experts detects alarms [1]

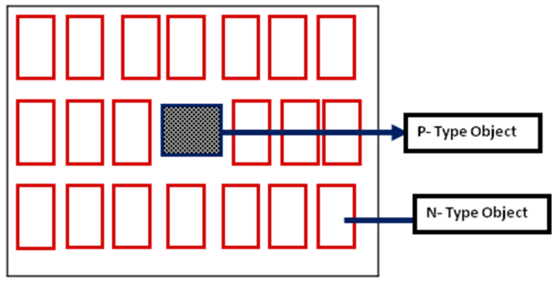

Fig 6 P N Algorithm

Modification to P-N Learning Algorithm:

When object/template from the image is selected by the user, object to be tracked is called as P-type of object and background is divided into the numbers of N-Type Objects as shown in figure.

This P-type Object becomes reference template. Input image is matched with the template. Percentage of P-Type and N-Type matching is calculated.

If input image matches with N-Type of object then matched image is neglected because it is falsely detected. At the same time, P-type of object is matched with all the images in the stored array of Image.

As Object moves, position of P-Type and N-Type of Objects may change accordingly. Percentage of P-Type of Object is calculated and behavior of object is learned. Maximum percentage of P-type of object gives the object of interest.

Step 11: Go back to step 7.

Performance Measures

1. The performance is evaluated using precision P, recall R, and f-measure F.

2. P is the number of true positives divided by the number of all responses. R is the number of true positives divided by the number of object occurrences that should have been detected.

3. F combines these two measures as $F=2PR / (P+R)$.

4. Detection was considered to be correct if its overlaps with ground truth bounding box was larger than 50 percent.

5 Major Expected Result

1. We expect the PCA matching percentage up to 0.98

2. Average localization error can be reduced up to 65 %.

3. Template matching intensity is expected to be 255 in the range of 0-255 where 0 is dark and 255 intense.

4. Improvement in object learning is expected because the modified of PN learning Algorithm

5. Proposed method can track the moving object successfully; moreover, the trace of object can be also extracted and the tracking result is unconcerned with the size of moving object. The proposed method still can track the interested target and filter out the noise by checking the size of moving regions

6. The speed of system for processing a frame in the ROI is about 300 ms approximately and may increase if ROI increases.

7. If ROI fails for fast motion video then Background subtraction technique is used.

6 Benchmark

The Benchmark is decided on the following parameters :

1. Average localization Error —average distance between center of predicted and ground truth bounding box.

2. recall —number of true positives divided by the length of the sequence (true positive is considered if the overlap with ground truth is > 50%)

3. System estimates the scale of an object. However, the algorithms compared in this experiment perform tracking in single scale only. In order to make a fair comparison, the scale estimation was not used.

7 Applications

1. Motion-based recognition: human identification based on gait, automatic object detection, etc.

2. Automated surveillance: monitoring a scene to detect activities or unlikely events

3. Video indexing: automatic annotation and retrieval of in multimedia databases

4. Human-computer interaction: gesture recognition, eye gaze tracking for data input to computers, etc.

5. Traffic monitoring: real-time gathering of traffic statistics to direct traffic flow

6. Vehicle navigation: video-based path planning and obstacle avoidance capabilities [1]

Figure- 7 Application [1]

8 Conclusions

At the end of this review, we conclude that Automatic Real-time Tracking and Learning of object will continue to grow in every direction; new audiences, new purposes, new styles of use, new modes of interaction etc.

Hence, we overviewed description of various algorithm and technique which will be useful to enhance the performances of tracking by reducing complexity caused due to loss of information caused by projection of 3D world on 2D image, complex object shapes / motion, illumination changes, scaling, rotation and partial and full object occlusions compared to existing systems.

9 References

[1] Zdenek Kalal, jiri matas, "Tracking- learning –Detection" IEEE Transactions on pattern analysis and machine intelligence", vol. 34, no. 7, july 2012 1409, 0162-8828/12/$31.00 2012 IEEE.

[2] Rangachar Kasturi, Fellow, IEEE, Dmitry Goldgof, Fellow, IEEE, Padmanabhan Soundararajan, Member, IEEE, Vasant Manohar, Student Member, IEEE, "Framework for Performance Evaluation of Face, Text, and Vehicle Detection and Tracking in Video: Data, Metrics, and Protocol", IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 31, No. 2, February 2009

[3] Swantje Johnsen and Ashley Tews, "Real-Time Object Tracking and Classification Using a Static Camera", Proceedings of the IEEE ICRA 2009

[4] Marwa abdel el Azeem Marzouk, "Modified background subtraction algorithm for motion detection in surveillance systems",vol 1,Number 2,(2010), pp -112-123.

[5] JainHong, Zhang, Zhenuan, Wei Guo, "Object tracking using improved camshaft with SURF method" , OSSC-2011 May- 2012, 987-61284-495-4/11 @2011IEEE

[6] Prof. Y. Vijaya Lata1, Chandra Kiran Bharadwaj Tungathurthi, "Facial Recognition using Eigen faces by PCA", Transaction Paper International Journal of Recent Trends in Engineering, Vol. 1, No. 1, May 2009

[7] Prof. Y. Vijaya Lata1, Chandra Kiran Bharadwaj Tungathurthi2, H. Ram Mohan Rao3, Dr. A. Govardhan4, Dr. L. P. Reddy, "Facial Recognition using Eigenfaces by PCA", International Journal of Recent Trends in Engineering, Vol. 1, No. 1, May 2009

[8] Chi-Farn Chen and Yun-Te Su "The Use of PCA for Moving Objects Tracking on the Image Sequence", Center for Space and Remote Sensing Research publication. 2012, 987-1-4577-0345

[9] Tim K. Lee and Mark S. Drew, "3D Object Recognition by Eigen-Scale-Space of Contours", Cancer Control Research Program, BC Cancer Reserach Centre, 675

[10] Swantje Johnsen and Ashley Tews, "Real-Time Object Tracking and Classification Using a Static Camera", Workshop on People Detection and Tracking Kobe, Japan, May 2009

[11]Rangachar Kasturi, Padmanabhan Soundararajan, Framework for Performance Evaluation of Face, Text, and Vehicle Detection and Tracking in Video: Data, Metrics, and Protocol", 0162-8828/09/$25.00 2009 IEEE

[12] The OpenCV Tutorials, Release 2.4.2 July, page 1-355

[13] Rita Cucchiara, Costantino Grana, Massimo Piccardi, Andrea Prati, "Detecting Moving Objects, Ghosts and Shadows in Video Streams".

[14] Priti Kuralkar*, Prof. V.T.Gaikwad "Background Subtraction and Shadow Detection Techniques A Review Paper", International Journal of Computer, Electronics & Electrical Engineering (ISSN: 2249 –9997) Volume 2–Issue1

[15] Chi-Farn Chen and Yun-Te Su "The Use of PCA for Moving Objects Tracking on the", Center for Space and Remote Sensing Research National Central University Jhongli, TAIWAN Image Sequence

[16] Open CV documentation