Research Article Impact Factor: 4.226 ISSN: 2319-507X
Aaruni Bhugul, IJPRET, 2016; Volume 5 (2): 189-203 IJPRET

INTERNATIONAL JOURNAL OF PURE AND APPLIED RESEARCH IN ENGINEERING AND TECHNOLOGY

A PATH FOR HORIZING YOUR INNOVATIVE WORK

SPECIAL ISSUE FOR
INTERNATIONAL CONFERENCE ON
"INNOVATIONS IN SCIENCE & TECHNOLOGY: OPPORTUNITIES & CHALLENGES"

OBJECT RECOGNITION USING TEMPLATE MATCHING, AN APPLICATION OF ALGORITHM

AARUNI BHUGUL[1], SPARSH PATHAK[2]

1. Accenture Pvt. Ltd., Magarpatta city, Pune.
2. University of Texas in Dallas, United states of America.

Accepted Date: 07/09/2016; Published Date: 24/09/2016

Abstract: *A computer vision system has been developed for real-time Motion detection and human motion tracking of 3 D objects including those of variable internal parameters. A fast algorithm based on various algorithms of Template matching like correlation matrix, absolute difference matrix, and their normalized parts have been implemented along with a Template Updating technique using sliding window object localization approach to track the motion of a detected body in the surveillance video. A fast algorithm based on color based differentiation technique is also implemented which tracks the moving object on the basis of its dominant color. Furthermore, a data structure implementation algorithm has been proposed to reject the non-useful areas of a binary image formed after various filtering techniques. The algorithms implemented provide accurate results for the human surveillance. The method allows for larger frame to frame motion and can robustly track models with degrees of freedom while running on relatively inexpensive hardware. These provide a reasonable compromise between the simplicity of parameterization and the expressive power for subsequent scene understanding. The proposed applications of algorithms implemented in this report could be human motion analysis in visual surveillance, where path of the person is required.*

Keywords: *Template, Algorithm*

PAPER-QR CODE

Corresponding Author: MISS. AARUNI BHUGUL

Co Author: MR. SPARSH PATHAK

Access Online On:

www.ijpret.com

How to Cite This Article:

Aaruni Bhugul, IJPRET, 2016; Volume 5 (2): 189-203

INTRODUCTION

Object detection is an important computer vision building block. Object tracking in object detection is an important computer vision building block. Object tracking in videos involves verifying the presence of an object in image sequences and possibly locating it precisely for recognition. Object tracking is to monitor the objects for 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. Object detection, path tracking & Action Recognition are the most active fields of research in the field of Computer Vision & Image Processing. Traditional surveillance systems require human beings to continuously monitor several incoming videos. Surveillance cameras are already prevalent in commercial establishments, while camera outputs are usually recorded in tapes or stored in video archives. Such systems are prone to human errors. That's why there is need of an automated intelligent system to detect classify and track human motion. Major concern is to detect the required object or required human in a video, which is essentially required in most of real life applications like robotics, defence etc.

The areas where the object detection and human motion analysis systems can be used are: For surveillance and monitoring of the people to ensure that they are within the norms, For Military and Police surveillance, In the field of Robotics where path tracing and motion analysis is required and in Educational & Manufacturing industries.

Object Detection, Classification and Tracking is an important task within the field of computer vision. Object detection in video streams has been a popular topic in the field of computer vision. Tracking is a particularly important issue in human motion analysis since it serves as a means to prepare data for pose estimation and action recognition. In contrast to human detection, human tracking belongs to a higher-level computer vision problem. However, the tracking algorithms within human motion analysis usually have considerable intersection with motion segmentation during processing. As one of the most active research areas in computer vision, visual analysis of human motion attempts to detect, track and identify people, and more generally, to interpret human behavior, from image sequences involving humans. Human motion analysis has attracted great interests from computer vision researchers due to its promising applications in many areas such as visual surveillance, perceptual user interface, content-based image storage and retrieval, video conferencing, athletic performance analysis, virtual reality, etc. A general framework [1-8] for Object detection analysis involves stages such as motion detection with the help of background subtraction and foreground segmentation, object classification, and motion tracking. Wang [9] classifies object motion analysis into three parts, namely object detection, object tracking & object behavior understanding. The importance and popularity of object motion analysis has led to several previous surveys. Each such survey is discussed in the following in order to put the current review in context. The focuses were on three major areas related to interpreting human motion: (a) motion analysis involving human body parts, (b) tracking moving human from a single view or multiple camera perspectives, and (c) recognizing human activities from image sequences. Collins et al. [10] classified moving object blobs into four classes such as single human, vehicles, human groups and clutter, using two factors, namely area and shape factor. Bo Wu and Ram Nevatia [11] proposed an approach to automatically track multiple, possibly partially occluded humans in a walking or standing pose from a single camera, which may be stationary or moving. A human body is represented as an assembly of body parts. Part detectors are learned by boosting a number of weak classifiers which are based on edge-let features. Responses of part detectors are combined to form a joint likelihood model that includes an analysis of possible occlusions. The combined detection responses and the part detection responses provide the observations used for tracking. Liang Xiao [12] talks about two types of Image sequences formed by the moving target one is the static background, the other is the varying background. It states that former case usually occurs in the camera which is in a relatively static state, produces moving image sequences with static background while the latter occurs in the target movement, when camera is also in the relative movement state. It also talks about optical flow methods but criticizes them for their need of specialized hardware. Recent years have seen consistent improvements in the task of automated tracking of pedestrians in visual data. The problem of tracking of multiple targets can be viewed as a combination of two intertwined tasks: inference of presence and locations of targets; and data association to infer the most likely tracks. Research in the analysis of objects in general, and humans in particular, has often attempted to leverage the parts that the objects are composed of. Indeed, the state-of-the-art in human detection has greatly benefited from explicit and implicit detection of body parts [13]. A model of spatial relationships between detected parts is learned in an online fashion so as to split pedestrian track lets at points of low confidence.

The main objective of present work is to develop an automated Object detection system for analyzing motion of target object in a video stream from video surveillance

2. Proposed Technique

In this work Object detection is to be done by using template matching method. Template matching is a technique for finding areas of an image that match (are similar) to a template image (patch).It is a technique in digital image processing for finding small parts of an image which match a template image. It can be used in manufacturing as a part of quality control, a way to navigate a mobile robot, or as a way to detect edges in images. The Algorithm is implemented in OPENCV and the approach used for object tracking is as follows:

1. First a template image is to be loaded. A Template image **(T)** in the patch image which will be compared to the source image.

2. After that video in which detection is to be done is loaded.

3. After loading a video, matching method is to be applied on the first frame

4. Then an object is detected in the first frame by making rectangular box around the object in the first frame.

5. Gaussian Filters are applied on each consecutive frames of the video.

6. The next objective is to find the object in the image sequence. Foreground detection is done by using sliding window approach followed by template matching which is described later.

2.1. Sliding Window Object Localizations

Many different definitions of object localization exist in the literature. Typically, they differ in the form that the location of an object in the image is represented, e.g. by its centre point, its contour, a bounding box, or by a pixel-wise segmentation.

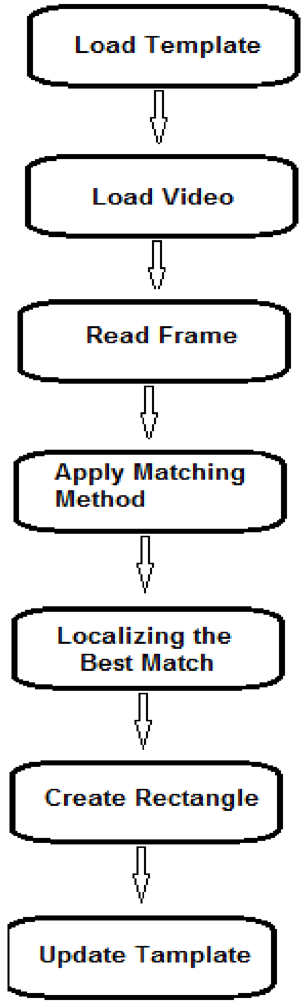


Fig.1 Algorithm flow diagram:

In the following we will only study localization where the target is to find a bounding box around the object. This is a reasonable compromise between the simplicity of the parameterization and its expressive power for subsequent scene understanding. An additional advantage is that it is much easier to provide ground truth annotation for bounding boxes than e.g. for pixel- wise segmentations.

In sliding-window-based approaches for object detection, sub-images of an input image are tested whether they contain the object of interest. Potentially, every possible sub-window in an input image might contain the object of interest. However, in a VGA image there are already 23;507;020;800 possible sub-windows and the number of possible sub windows grows as n for images of size n _n .We restrict the search space to a subspace R by employing the following constraints. First, we assume that the object of interest retains its aspect ratio. Furthermore, we introduce margins dx and dy between two adjacent sub windows and set dx and dy to be 1/10 of the values of the original bounding box. In order to employ the search on multiple scales, we use a scaling factor s = 1.2a, a ∈ {-10……10} g for the original bounding box of the object of interest. We also consider sub windows with a minimum area of 25 pixels only.

$$|R|=\Sigma_{S=-10}^{10}[n- s(w + dx)][m- (h+ dx)]$$

'w' and 'h' denote the size of the initial bounding box and n and m the width and height of the image respectively. For sliding window we need two primary components:

a. Source image (I): The image in which we expect to find a match to the template image.

b. Template image (T): The patch image which will be compared to the source image.
Our goal is to detect the highest matching area.

Figure 2.1.(a) Sliding window object localization

To identify the matching area, we have to compare the template image against the source image by sliding it.

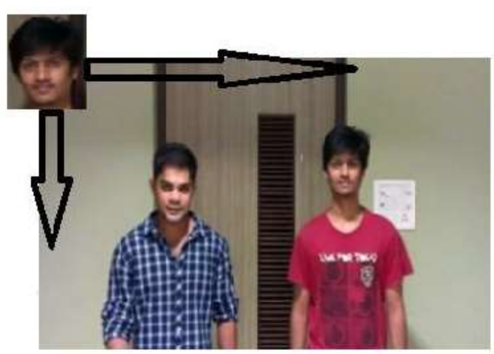


Figure 2.1.(b) Siding template image over source image

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). For each location of **T** over **I**, you store the metric in the result matrix **(R)**. Each location in **R** contains the match metric.

Figure 3. Resultant showing maximum match

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 minMaxLocto locate the highest value (or lower, depending of the type of matching method) in the *R* matrix

2.2 Template Matching Methods

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:

a) **Source Histogram (I):** The histogram of image in which we expect to find a match to the template image histogram.

b) **Template Histogram (T):** The histogram of patch image which will be compared to the template image histogram.

The goal is to detect the highest matching area. To identify the matching area, the template image histogram is compared against the source image histogram by sliding it using sliding window approach explained in previous topic. For each location of **T** over **I**, you store the metric in the result matrix**(R)**. We use following methods [9] for matching:-

a. Absolute Sequence Difference method:

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

b. Normalized Sequence Difference method:

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

c. Absolute Correlation Method:

$$R(x,y) = \sum_{x',y'}^{\infty} [T(x',y') \sum I(x+x',y+y')]^2$$

d. Normalized Correlation Method:

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

e. Absolute Coefficient Method:

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

f. Normalized Coefficient Method:

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

$$T(x',y') = T(x',y') - \frac{1}{w.h} \sum_{x'',y''} T(x''y'')$$

$$I'(x+x',y+y') = I(x+x',y+y') - \frac{1}{w.h} \sum_{x'',y''} I(x+x'',y+y'')$$

Then the location with higher matching probability is localized and a rectangle is drawn around the area corresponding to the highest match and objected is detected.

2.3 Template Matching by Cross Correlation

Correlation is an important tool in image processing, pattern recognition, and other fields. The correlation between two signals (cross correlation) is a standard approach to feature detection [3, 4] as well as a building block for more sophisticated recognition techniques. Textbook presentations of correlation commonly mention the convolution theorem and the attendant possibility of efficiently computing correlation in the frequency domain via the fast Fourier transform. Unfortunately the normalized form of correlation (correlation coefficient) preferred in many applications does not have a correspondingly simple and efficient frequency domain expression, and spatial domain implementation is recommended instead.

Template matching techniques [3] attempt to answer some variation of the following question: Does the image contain a specified view of some feature, and if so, where? The use of cross correlation for template matching is motivated by the distance measure. The resulting correlation term c(u,v) is a measure of the similarity between the image and the feature.

2.4 Normalized Cross Correlation

If the image energy Σf2(x, y) is not constant however, feature matching by cross correlation can fail. For example, the correlation between the template and an exactly matching region in the image may be less than the correlation between the template and a bright spot. Another drawback of cross correlation is that the range of c(u, v) is dependent on both the size of the template and the template and image amplitudes.

Variation in the image energy under the template can be reduced by high-pass filtering the image before cross correlation. In a transform domain implementation the filtering can be conveniently added to the frequency domain processing, but selection of the cut-off frequency is problematic – a low cut-off may leave significant image energy variations, whereas a high cut-off may remove information useful to the match. Normalized cross correlation overcomes these difficulties by normalizing the image and template vectors to unit length, yielding a cosine-like correlation coefficient.

The main aim of present work is to detect the object so that the required object can be tracked. The location of an object in the image is represented by its center point or its contour, or a bounding box, or by a pixel-wise segmentation. Here the target is only to find a bounding box around the object. This is a reasonable compromise between the simplicity of the parameterization and its expressive power for subsequent scene understanding. An additional advantage is that it is much easier to provide ground truth annotation for bounding boxes than for pixel-wise segmentations. In sliding-window-based approaches for object detection, sub images of an input image are tested whether they contain the object of interest. Potentially, every possible sub window in an input image might contain the object of interest. The template used in the previous iteration is no more useful to us because with the motion of the moving body, the template might not match any area after a few frames have passed in further iterations. Moreover a moving body might change its angle of orientation towards the camera when the next few frames are read.

To overcome these shortcomings the template update approach comes in quite handy. Whenever the template is matched with a certain area in a frame, the detected area is bounded by a rectangle whose size as same as the size of the template. This rectangle is then cropped from the frame and the cropped image becomes our new template in the next iteration. This approach where at every frame our template is updated gives accurate results until and unless the frames are missed or the motion is so rapid that matching a template fails in the very next frame. These conditions are rarely observed in our day to day life so template matching and update technique tracks the path of a human very accurately. In case of multiple human motions tracking this approach is quite useful as it distinguishes between two blobs directly on the basis of template matching and updating. Various features like orientation, area, color, contrast etc come into play when template matching is used as the area most alike would obviously give the minimum difference. This difference is plot in terms of grey scale and is shown in the results. The following color based approach can be said to be a sub-part of this approach but the time reduction in tracking the motion that we achieve with color based approach is quite good

3. Results and Discussions

The tracked region based on template matching and updating gives accurate results. Only error is when the template is lost in any frame due to rapid motions. The rectangles formed across the faces of the detected humans in the results are the exact match to their faces being supplied as templates in the beginning and being updated in every frame. Even if the faces are moved by some angle and the orientation towards the camera is changed the results are not affected as the templates are updated. The six approaches for template matching which have been described provide different results in different scenarios. Some are more accurate in one while others are more accurate in the other. So there is no trade off. Here 2 sample results are shown with original frame image and initial templates. First image is the frame input from the video and template based and updating algorithm searches for the templates of the faces provided in the beginning and being updated at each frame.

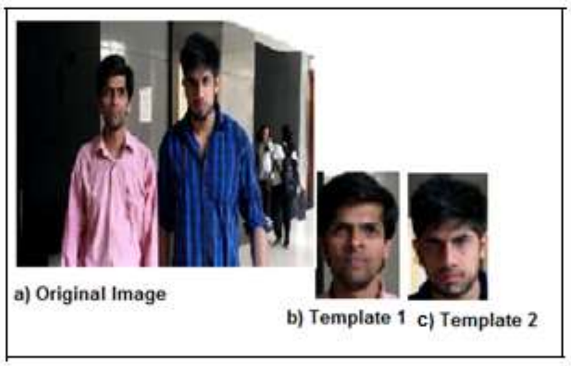
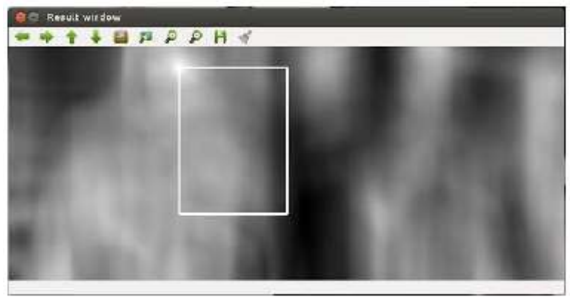


Fig.4 Template based detection sample

d) Image after applying algorithm

e) updated Template 1

f) Updated Tamplate 2



g) Resulting Window 1

h) Resulting Window 2

a) Original Image

b) Tamplate 1

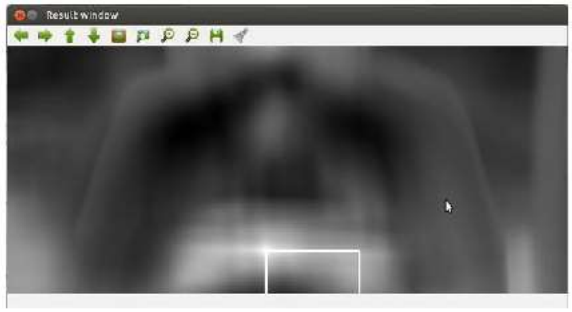
c) Template 2

d) Image after applying algorithm

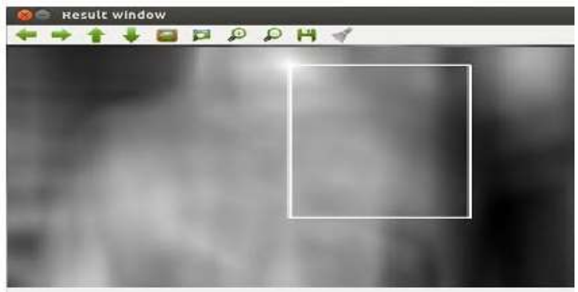
Figure 5. Template base detection sample

e) Updated Tamplate 1

f) Updated Tamplate 2



g) Resultant Window 1



h) Resultant Window 2

4. **REFERENCES**

1. Ff. R.T. Collins, A.J. Lipton, T. Kanade, Introduction to the special section on video surveillance, IEEE Trans. Pattern Anal. Mach. Intel. 22 (8) (2000) 745–746

2. A. R. Francois and G. G. Medioni. Adaptive colour background modeling for real-time segmentation of video streams. In Proceedings of the International Conference on Imaging Science, Systems, and Technology, pages 227{232, 1999. [3] R.O. Duda and P.E.Hart, Pattern Classification and Scene Analysis, New York: Wiley, 1973.

3. R. C. Gonzalez and R. E. Woods, Digital Image Processing (third edition), Reading, Massachusetts: Addison-Wesley, 1992

4. G. R. Bradski and J. Davis, Motion Segmentation and Pose Recognition with Motion History Gradients, Machine Vision and Applications, 2002

5. D. Meyer, J. Denzler, H. Niemann, Model based extraction of articulated objects in image sequences, Proceedings of the Fourth International Conference on Image Processing, 1997

6. R. Brunelli. Template Matching Techniques in Computer Vision: Theory and practice. Wiley Publishing, 2009

7. W. C. Abraham and A. Robins. Memory retention–the synaptic stability versus plasticity dilemma. Trends in neurosciences, 28(2):73–78, Feb. 2005.

8. OpenCV, Learning. "Computer vision with the OpenCV library." *GaryBradski, Adrian Kaehler*(2008).

9. Wang, Liang, Weiming Hu, and Tieniu Tan. "Recent developments in human”

10. J.K. Aggarwal, Q. Collin, Human motion analysis: a review, Proceedings of the IEEE Workshop on Motion of Non-Rigid and Articulated Objects, 1997, pp. 90–102.

11. Bo Wu and Ram Nevatia, Detection and Tracking of Multiple, Partially Occluded Humans by Bayesian Combination of Edgelet based Part Detectors, Tenth IEEE International Conference, Computer Vision, 2005. ICCV 2005.

12. Liang Xiao and Tong-qiang Li, Moving Object Detection and Tracking, 2010

13. Mikolajczyk, K., Schmid, C., Zisserman, A.: Human detection based on a probabilistic assembly of robust part detectors. In: ECCV. (2004).