Convolutional Regression Gradient Descent method for Robust Visual Tracking

N.Bhuvaneswari [1], M.Nuthal Srinivasan[2],

ECE Department,
E.G.S Pillay engineering college,
Nagapattinam.
bhuvanaraj145@gmail.com
mnuthalsrinivasan@egspec.org

***Abstract*— Discriminatively algorithm has attracted much attention in visual object tracking community. To predict the location of an object and to train the ridge regression model large number of samples are utilized in the discriminatively learned correlation filter (DCF). To solve this problem, the samples are generated from a searching patch by circular shifting. However, these circular synthetic samples also induce some negative effects that weaken the robustness of DCF-based video trackers. In this project, we propose a new approach to learn the visual tracking of regression model using single convolutional layer with gradient descent (GD). Contrary to DCF, the kernel size of the convolution layer with GD is set to the size of the object hence all "real" samples clipped from the whole image. In addition, the security system is added to verify the biometrics of the image using Arnold transformation. The result shows that the proposed algorithm achieves the outstanding performance than the existing DCF-based algorithms**.

Keywords—Visual tracking, linear regression, gradient descent, convolutional neural network.

I .INTRODUCTION

Video tracking predicts the state of object in the video to obtain an understanding of the scene that it describes. . It is an essential component of a number of technologies, including video surveillance, robotics, and multimedia. From a basic science perspective, methods in video analysis are motivated by the need to develop machine algorithms that can mimic the capabilities of human (and other animal) visual systems. It is an area of research that has seen huge growth in the recent past. Researchers in video analysis have varied backgrounds, including signal/image processing, computer science, system theory statistics, and applied mathematics.

In general discriminative algorithm can be divided into two parts .The first is to represent the object with handcrafted features, such as HOG, color names and original RGB colors [1].The second is deeply learned convolutional features from networks like ResNet [2] and VGGnets [3] The existing DCF algorithms utilizes thousands of samples for both training and detecting .since it is important for visual tracking algorithm it induces some negative effects. (a)The training and detecting samples are all synthetic hence they decreases the effectiveness of the regression model.(b) Discriminative power of the regression model weaken due to the too much background information included in the training and predicting samples.(c) The search space for predicting the object location is limited to the size of the base sample. These three negative effects significantly limits the performance of DCF based trackers.

In this paper, we try to address this issues in a different way by combining the DCF algorithm with Gradient Descent (GD) method. Instead of looking for an analytic solution to the regression problem, we try to obtain an approximate solution via gradient descent (GD). Contrary to DCF, the kernel size of the convolution layer with GD is set to the size of the object hence all "real" samples clipped from the whole image**.**

II. RELATED WORK

Discriminative correlation filter has two types of trackers DCF based trackers and CNN based trackers. The proposed method is fully CNN based trackers. The convolutional neural network has truncated loss function to eliminate easy negatives and a weight function to enhance positives.

A.DCF based trackers

The efficient and robust trackers are built using Discriminative correlation filters. The large number samples are utilized in the DCF based trackers for detecting and training

.A number of trackers [4], [9] developed from the DCF framework have been proposed to improve the performance. In the spatial Regularization discriminative correlation filter (SRDCF) framework, the searching patch can be much larger than in DCF, so that more negative samples can be utilized. As a result, the performance of SRDCF is significantly improved compared with the other DCF based trackers. In [5], the performance is further improved by adopting deeply learned convolutional features.

B.CNN Based Trackers

CNN has achieved a great success in computer vision task like object detection and image classification. In general it is impossible to train a deep CNN because of limited training data. Instead, we can transfer a deep CNN like VGGnets [2] trained for image classification to extract convolutional features for tracking. In [6], both shallow and deep convolutional features extracted from a pre-trained CNN are utilized in the DCF frame work. In a two-stream fully convolutional network to capture both general object information and specific discriminative information for visual tracking.

III. VIDEO TO FRAME CONVERSION AND REGRESSION VIA SINGLE CONVOLUTION LAYER

The regression of samples extracted by placing a window over an image patch can be computed via a single convolution layer. Then the coefficients can be optimized using gradient descent together with the back-propagation technique [12]. Compared to the conventional discriminative correlation filters, the convolutional regression is trained on “real” samples with no background context included, and unlimited negative samples can be incorporated.

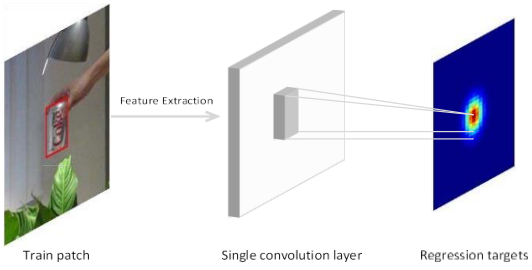

Fig 1 Regression via single convolution layer.

Instead of looking for an analytic solution to the regression problem, we try to obtain an approximate solution via gradient descent (GD). In our framework, the regression model is built over a one-channel-output convolution layer, as used in typical convolutional neural networks except that the kernel size is set to the object size.

III. TRACKING VIA CONVOLUTIONAL METHOD AND FOREGROUND AND BACKGROUND EXTRACTION

The feature extraction block is employed to calculate the feature points from the face for additional process. Feature extraction starts from an initial set of measured knowledge and builds derived values. The frame rate is that the range of frames or pictures that area unit projected or displayed per second.

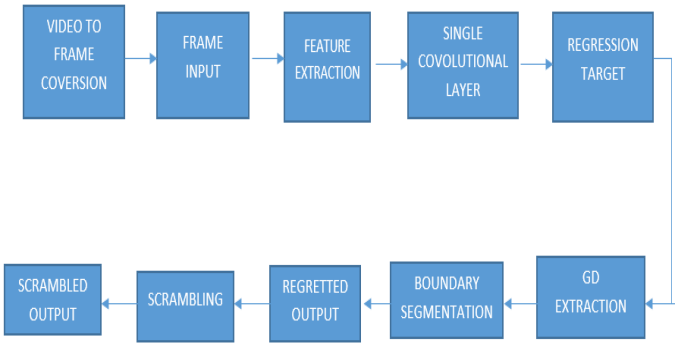

Fig 2. Block diagram of proposed method

Regression analysis may be a kind of prognostic modelling technique that investigates the link between a dependent (target) and independent variables (predictor)[14]. This system is employed for statement, statistic modelling and finding the causal result relationship between the variables. Motion maps area unit in a different way to represent the motion of associate object (other representations are graphical and mathematical models). Motion maps are a visible thanks to represent associate object's motion at numerous time. Boundary-based strategies area unit typically accustomed explore for express or implicit boundaries between regions admire completely different tissue sorts.

Edge detection includes a range of mathematical strategies that aim at distinguishing points in a digital image at that the image brightness changes sharply or, additional formally, has discontinuities. The points at that image brightness changes sharply area unit usually organized into a collection of arched line segments termed edges [15]. An equivalent downside of finding discontinuities in one-dimensional signals is thought as step detection and the matter of finding signal discontinuities over time is thought as change detection [16]. Edge detection may be a basic tool in image process, machine vision and computer vision, notably within the areas of feature detection and feature extraction

Ridge detection is that the try, via computer code, to find ridges in a picture. In arithmetic and laptop vision, the ridges of a sleek perform of 2 variables area unit a collection

of curves whose points area unit, in one or additional ways in which to be created precise below, native maxima of the perform in a minimum of one dimension[20]. Gradient descent is a first-order iterative optimization algorithm for finding the minimum of a perform. To search out a local minimum of a perform victimization gradient descent, one takes steps proportional to the negative of the gradient (or approximate gradient) of the perform at the present purpose. If instead one takes steps proportional to the positive of the gradient, one approaches a local maximum of that function; the procedure is then acknowledged as gradient ascent [21]. The gradient descent will take several iterations to cipher an area minimum with a required accuracy, if the curvature in directions is extremely different for the given perform. For such functions, preconditioning, that changes the pure mathematics of the area to form the perform level sets like concentric circles, cures the slow convergence.

IV.PRIVACY USING ARNOLD TRANSFORMATION

Privacy has become a problem of nice anxiety within the transmission and distribution of police work videos. For instance, personal image mustn't be browsed while not authorization. Face image scrambling has emerged as an easy answer for privacy-related applications. Facial biometric verification has to be administered within the disorganized domain, therefore transferral a brand new challenge to face classification [17]. However, this kind of scrambling will simply lose the facial information, and hence, face recognition becomes unsuccessful in this case [18]. In addition, for security reasons, it is obviously not a good choice to really erase human faces from surveillance video.

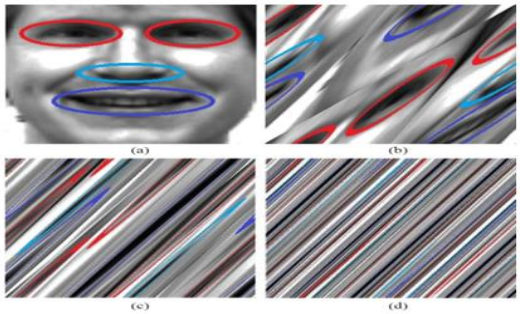

Fig .3 Scrambling using Arnold Transformation

Arnold remodel primarily based scrambling as our specific take a look at platform. Automatic police work systems are put in with on-line facial biometric verification [19]. Whereas it's going to not be allowable to unscramble detected faces while not authorization thanks to privacy protection policies, the flexibility to hold out facial biometric verification within the disorganised domain becomes fascinating for several rising police work systems.

V.RESULTS AND DISCUSSION

A.Frame Extraction:

Frame extraction can be defined as an action of retrieving a frames from video source, usually a hardware based source. After uploading video we get 100 frames for 3 second.

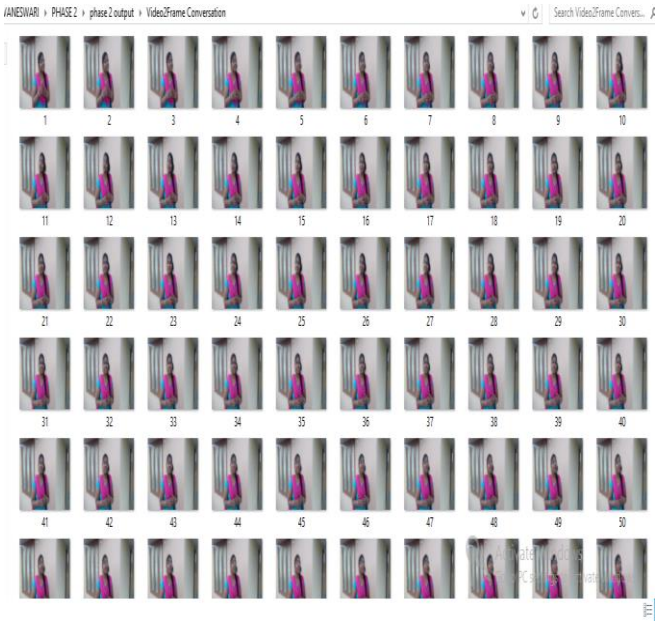

Fig .4 Extracted Frame (1-50)

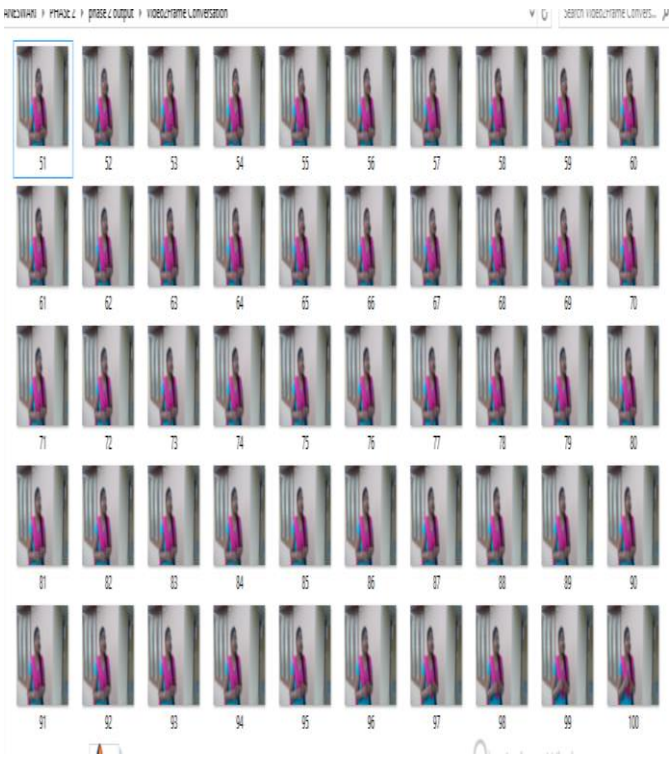

Fig .4 Extracted Frame (51-100)

B.Foreground and Background Extraction:

Foreground and Background extraction detects the moving objects by finding the difference between the current frame and a reference frame.

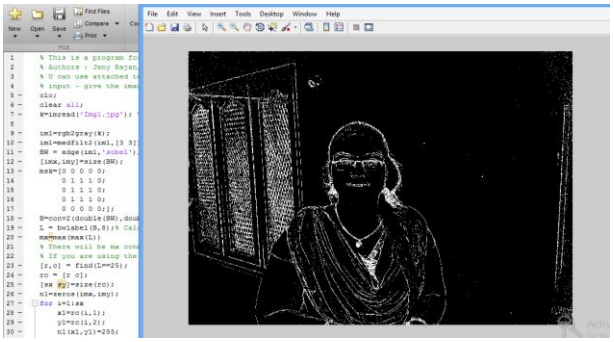

Fig .5 Foreground Extracted image

Edge detection is a fundamental tool in image processing, machine and computer vision, particularly in the areas of feature detection and feature extraction

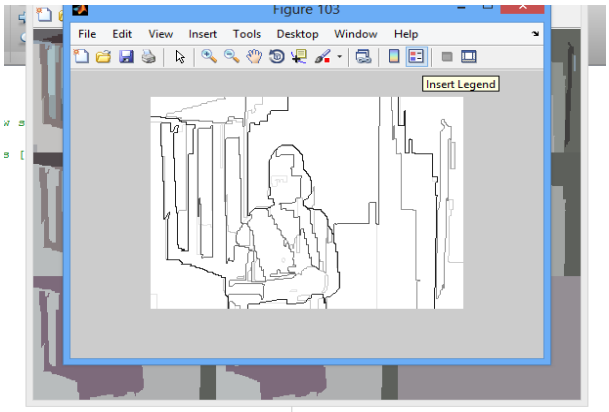

Fig .6 Edge detected image

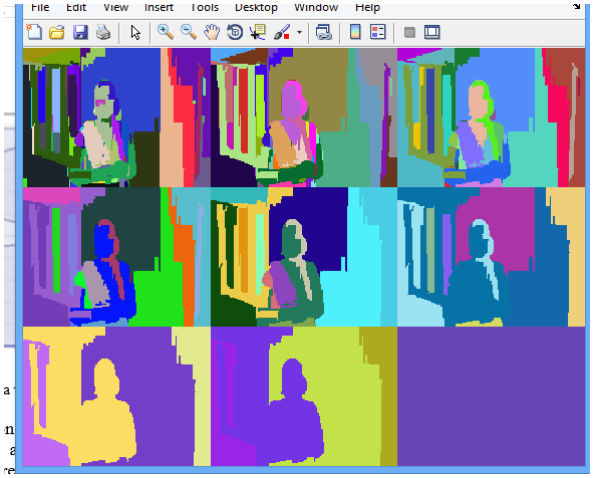

Fig .7 Background Extracted image

C.Scrambling process:

The security system is added to verify the biometrics of the regretted image using Arnold transformation.

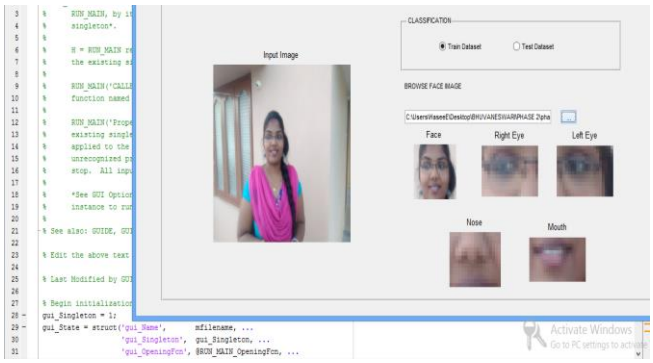

Fig .8 Biometric verification for scrambling.

The scrambled datasets are generated, the scrambling pattern is different from one images to another. Dataset is generated after the Arnold transformation

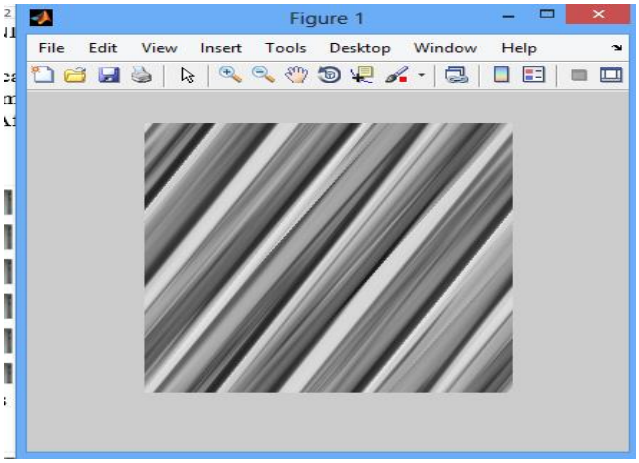

Fig .9 Scrambled image

The test image is compared to the trained dataset, if the images is already trained in the dataset means it execute the command as “Face Matched” if not means “Face Doesn’t Matched”.

Fig .10 Biometric testing for descrambling

VI. CONCLUSION

We propose a new approach to learn the visual tracking of regression model using single convolutional layer with gradient descent (GD).An improved objective function to eliminate easy negatives and enhance positives. In addition, the security system is added to verify the biometrics of the image using Arnold transformation. The result shows that the proposed algorithm achieves the outstanding performance than the existing DCF-based algorithms.

References

[1] M. Danelljan, F. S. Khan, M. Felsberg, and J. van de Weijer, "Adaptive color attributes for real-time visual tracking," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2014, pp. 1090–1097.

[2] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in Proc. Int. Conf. Learn. Represent., 2015, pp. 1–14.

[3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," CoRR, vol. abs/1512.03385, pp. 1–12, Dec. 2015. [Online]. Available: http://arxiv.org/abs/1512.03385

[4] C. Ma, X. Yang, C. Zhang, and M.-H. Yang, "Long-term correlation tracking," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2015, pp. 5388–5396.

[5] M. Danelljan, G. Häger, F. S. Khan, and M. Felsberg, "Convolutional features for correlation filter based visual tracking," in Proc. IEEE Int. Conf. Comput. Vis. Workshop, Dec. 2015, pp. 621–629.

[6] C. Ma, J.-B. Huang, X. Yang, and M.-H. Yang, "Hierarchical convoltional features for visual tracking," in Proc. IEEE Int. Conf. Comput. Vis., Dec. 2015, pp. 3074–3082.

[7] L. Wang, W. Ouyang, X. Wang, and H. Lu, "Visual tracking with fully convolutional networks," in Proc. IEEE Int. Conf. Comput. Vis., Dec. 2015, pp. 3119–3127.

[8] H. Nam and B. Han, "Learning multi-domain convolutional neural networks for visual tracking," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2016, pp. 1–10.

[9] Y. Li, J. Zhu, and S. C. H. Hoi, "Reliable patch trackers: Robust visual tracking by exploiting reliable patches," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2015, pp. 353–361.

[10] X. Jia, H. Lu, and M.-H. Yang, "Visual tracking via adaptive structural local sparse appearance model," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2012, pp. 1822–1829.Transactions on Antennas and Propagation, vol. 62, no. 8, pp. 4015–4020, 2014.

[11]A.Melle and J.-L. Dugelay, "Scrambling faces for privacy protection usingbackground self-similarities," in Proc. IEEE Int. Conf. Image Process., 2014, pp. 6046–6050.2002, p. 215.

[12]A. Erdlyi, T. Bart, P. Valet, T. Winkler, and B. Rinner, "Adaptive cartooning for privacy protection in camera networks,"inProc. Int. Conf. Adv.Video Signal Based Surveillance, 2014, pp. 44–49.

[13]A. Ghazanfar and D. Takahashi, "Facial expressions and the evolution of the speech rhythm," J. Cognitive Neurosci., vol. 26, no. 6, pp. 1196–1207, Jun. 2014.

[14]D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," CoRR, vol. abs/1412.6980, pp. 1–15, Dec. 2014. [Online]. Available: http://arxiv.org/abs/1412.6980

[15] M. Danelljan, G. Häger, F. S. Khan, and M. Felsberg, "Adaptive decon-tamination of the training set: A unified formulation for discriminative visual tracking," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2016, pp. 1430–1438.

[16] Z. Jianming, M. Shugao, and S. Stan, "MEEM: Robust tracking via multiple experts using entropy minimization," in Proc. Eur. Conf. Comput. Vis., 2014, pp. 188–203.

[17] J. Gao, H. Ling, W. Hu, and J. Xing, "Transfer learning based visual tracking with gaussian processes regression," in Proc. Eur. Conf. Com- put. Vis., 2014, pp. 188–203.

[18] A.Melle and J.-L. Dugelay, "Scrambling faces for privacy protection usingbackground self-similarities," in Proc. IEEE Int. Conf. Image Process., 2014, pp. 6046–6050.2002, p. 215.

[19] A. Erdlyi, T. Bart, P. Valet, T. Winkler, and B. Rinner, "Adaptive cartooning for privacy protection in camera networks,"inProc. Int. Conf. Adv.Video Signal Based Surveillance, 2014, pp. 44–49.

[20] A. Ghazanfar and D. Takahashi, "Facial expressions and the evolution of the speech rhythm," J. Cognitive Neurosci., vol. 26, no. 6, pp. 1196–1207, Jun. 2014.

[21] B. Draper, K. Baek, M. Bartlett, and J. Beveridge, "Recognizing faces with PCA and ICA," Comput. Vision Image Understanding, vol. 91, nos. 1/2, pp. 115–137, 2003.Comput. Vision Pattern Recog., 2014, pp. 1701–1708.

[22] D. J. Kim and Z. Bien, "Design of „personalized" classifier using soft computing techniques for „personalized" facial expression recognition," IEEE Trans. Fuzzy Syst., vol. 16, no. 4, pp. 874–885, Aug. 2008.

[23] F.Dufaux and T. Ebrahimi, "Scrambling for video surveillance with privacy," in Proc. Conf. Comput. Vision Pattern Recog. Workshop, Washington, DC, USA, 2006, pp. 106–110.

[24] F. Dufaux, "Video scrambling for privacy protection in video surveillance: Recent results and validation framework," Proc. SPIE, vol. 8063, p. 806302, 2011.Gap to human-level performance in face verification," in Proc. IEEE Conf.