



Object Tracking Podium on Region Covariance for Recognition and Classification

Md Alamgir Hossain¹, Prof (Dr) Goutam Sanyal²

¹Department of Computer Application (MCA),
Calcutta Institute of Technology,
Uluberia, West Bengal, India

²Department of Computer Science and Engineering,
National Institute of Technology,
Durgapur, West Bengal, India

¹alamgir_sim@yahoo.com, ²nitgsanyal@gmail.com

Abstract: Recent research has advocated the use of a covariance matrix of image features for tracking objects instead of the conventional histogram object representation models used in popular algorithms. In this paper we extend the covariance tracker and propose efficient algorithms with an emphasis on both improving the tracking accuracy and reducing the execution time. The algorithms are compared to a baseline covariance tracker and the popular histogram-based mean shift tracker. Quantitative evaluations on a publicly available dataset demonstrate the efficacy of the presented methods. Our algorithms obtain significant speedups factors up to 330 while reducing the tracking errors by 86 –90% relative to the baseline approach. We show that it is capable of accurately detecting the non-rigid, moving objects in non-stationary camera sequences while achieving a promising detection rate of 98.61 percent

Keywords: region-of-interest; base-point; span-space-detector; covariance-matrix;

I. Introduction

The importance of Image/Object tracking and its classification with the aim at detection as well as selection of its feature is being accelerated day by day. Features should be considered as good if it is distinguishable, robust and uncomplicated to work out. Good qualitative features must be evaluated and filtered out with the help of proficient algorithms and unique as well as logically corrected method. So recognition and tracking tasks are not easy as it is accompanied by a variety of related-tasks.

The apparatus such as color, gradient, filtering techniques used in image statistics with respect to raw pixel values are in common choice for image features selection in the field of computer vision during earlier stage of tracking system. However, it has been proved that extracted features would not be robust if the illumination changes; non-rigid images are accompanied with motions as there are few efficient matching algorithms dealing with high dimensional representation. The raw pixel values can easily be represented depending on the histogram-approach if it is applied on a specific region of an image where the region is identified with the help of estimating methods of joint distribution.

The versatility of [6] histogram for tracking of non-rigid object was used in wider way in many analyses. Moreover, for searching a fast global matching histogram has been able to prove its superiority [7].

Apart from tracking, histogram-based application is found in texture depiction [8, 9] and finding their correspondence [10] in the field of image processing.

The idea of integral image for getting the flavor of rapidity in computation process is noticed [11] in Haar-computational features extraction methods.

Many cascaded classifiers have been shown their superiority with respect to their performances on the way of image recognition, but those algorithms necessitate a huge training time to read out and classify the object. Histogram should be considered as base-point descriptor if we apply it for scaling of a span-space-detector. It is very efficient with respect to aiming at narrow neighborhoods but we can't consider it as global apparatus for span-space-detector.

Juan Pavona, Jorge Gomez-Sanz, Antonio Fernandez-Caballero, Julian J. Valencia-Jimenez have proposed [2] to give a new methodical frame of surveillance scheme managed by as a multi-agent method. They have imposed the idea of applying an agent-orientated technique, which is weighed up with tangible state of affairs.

Takeo Kanade, Robert T Collins and Alan J, Lipton have propose [3] to widen up a accommodating, multi-sensor video surveillance scheme that makes available with continuous reporting over battle field areas through

their thesis paper. Robert T. Collins, Alan J. Lipton, Takeo Kanade, Hironobu Fujiyoshi, David Duggins, Yanghai Tsin, David Tolliver, Nobuyoshi Enomoto, Osamu Hasegawa, Peter Burt and Lambert Wixson [4] have proposed a method for independent Video observation and Monitoring application. Their scientific procedures are applied in multiple and helpful video sensors for availing spontaneous reporting of people and vehicles in a mixed-up surroundings.

Kim C. Ng, Hiroshi Ishiguro, Mohan Trivedi and Takushi Sogo have developed [5] an integrated system with the aim to monitor a surveillance work. Their system was build up with manifold Omni-directional vision sensors and was made up emphasizing to deal with two unambiguous surveillance tasks. The First one is to track and silhouette human activity and the second one are to synthesis the virtual observations of people for keeping track of them for their environmental belongingness from random vantage positions.

Anthony R. Dick and Michael J. Brooks proposed [14] an automated visual surveillance with the aim as that it becomes capable to get free the operators from the burden by encapsulating the relevant software in their methodology that can analyze surveillance behaviour routinely. Their thesis paper reveals major evolution in the field of visual surveillance and takes into account to some of the prime problems in programmed video surveillance.

Liang Wang, Weiming Hu and Tieniu Tan have proposed [15] complete study on computer vision depending on person motion analysis. They have stressed on three prime issues directly related to general people motion analysis scheme: people recognition, people tracking and their activity understanding.

Norman Poh, Chi Ho Chan, Josef Kittler, Sébastien Marcel, Christopher Mc Cool, Enrique Argones Rúa, Jose Luis Alba Castro, Mauricio Villegas, Roberto Paredes, Vitomir Struc, Nikola Pavesic, Albert Ali Salah, Hui Fang, and Nicholas Costen [16] have proposed an assessment of a person's identity authentication using facial video data. Their systems dealt with a varied set of guesses, along with feature depiction and pre-processing deviations. They have been capable to weigh up the effect of unfavorable conditions, to access qualitative information, to process the query selection, and are in the position to construct template for video-to-video face verification.

II. Use of Covariance for tracking

The concise portrayal of the tracking methodology has been depicted in the following steps:

Considering frame by frame, we have tried to build a complete image. With the aim to capture a region with respect to an image, we have figured out the covariance-matrix based on the features of the images. Now from our frame, we are segregating those regions with minimal covariance distances and allocating them in their probable positions. To acclimatize the deviations, we are maintaining a set of preceding covariance matrices and hauling out a mean on the multiple samples.

A. Covariance Matrix for Distance Measurement

The distance can be measured based on the algorithm of Euclidean-space-measured method about it is proposed and proved that covariance-matrices are not depended on Euclidean technique. Generally, it has been seen that the popular machine based knowledge gathering methodology depends on Euclidean-spaces so those are not appropriate for our features-gathering methods. The adjoining neighborhood procedures which have been used in the subsequent part only necessitate a method of measuring a methodological distances between two position. The distance measuring methodology of the two covariance matrices can be performed by the equation.

$$\eta(a_1, a_2) = \sqrt{\sum_{m=1}^c x c^2 \phi_m(a_1, a_2)} \text{ --- (1)}$$

where $\{\phi_m(a_1, a_2)\}_{m=1, \dots, c}$ are the indiscriminate Eigen-values of m_1 and m_2 evaluated from the equation:

$$\phi_m a_1 p_m - a_2 p_m = 0 \quad m = 1 \dots v \quad (2)$$

And $p_m \neq 0$ are the indiscriminate Eigen-vectors. The distance computation η satisfies the maxims of matrix for optimistic distinct symmetric matrices a_1 and a_2

The Lie-group structure for optimistic distinct symmetric matrices required for the distance computation and their algebraic expression that is an equivalent form for optimistic affirmative matrices measuring tactic. For explanation of the Eigen-value computation we require a degree of three arithmetical maneuvers based on numerical methods along with one logarithm computation for distance calculation. This procedure is better compared to histogram approach.

B. Acceleration of Computation

The Integral-images carry a prime role by keeping itself in the intermediate stage between two images and accelerate quick computation based on regions of interest. We know that for every pixel of the Integral-image is the summation of all the pixels within the rectangle surrounded by the upper left corner of the image and the pixel-of-interest. For an image based on its intensity and its Integral-image is represented as,

Integral-image

$$(p', q') = \sum_{p < p', q < q'} \phi(p, q) \quad (3)$$

By the help of the above representation the summation of any rectangular region should be calculated in invariable time. The Integral-images would have been used in extended way for rapid computation if histogram approach is used based on region-of-interest. Here we have proposed an alike idea for speedy computation of region-of-interest based on covariance.

Now based on the (m, n)th component of the covariance matrix the aforesaid equation can be defined as

$$a_s(m, n) = \frac{1}{c-1} \sum_{k=1}^c (u_k(m) - \psi(m))(u_k(n) - \psi(n)). \quad (4)$$

By rearranging the expanded mean we can rewrite the above equation as:

$$a_s(m, n) = \frac{1}{n-1} \left[\sum_{k=1}^c u_k(n)u_k(n) - \frac{1}{c} \sum_{k=1}^c u_k(m) \sum_{k=1}^c u_k(n) \right] \quad (5)$$

To locate the covariance bounded by a rectangular region s, we must calculate the summation of every feature dimension, $u(m)m=1\dots c$, along with the addition after the multiplication of any two dimensional features, say, $\{u(m), u(n)\}$ $m,n=1\dots c$. Now we must have to build $v + v^2$ Integral-images for $u(m)$ and multiplication of $\{u(m), u(n)\}$.

Let there be (I) with $w \times h \times v$ tensor of the Integral- images. So we get the modified equation as:

$$I(p', q', m) = \sum_{p < p', q < q'} \Pi(p, q, m) \quad m = 1\dots v \quad (6)$$

next order Integral- images reforming it as:

$$Y(p', q', m, n) = \sum_{p < p', q < q'} \Pi(p, q, m) \Pi(p, q, n) \quad m, n = 1\dots v \quad (7)$$

We know that Integral- image is calculated in a single pass depending on our image. In our notation, I p, q is a vector with d dimension and γ x, y is a matrix with $v \times v$ dimension. So they can be represented as

$$I_{p,q} = [I(p, q, 1) \dots I(p, q, v)]^{time}$$

$$Y_{p,q} = \begin{pmatrix} Y(p, q, 1, 1) \dots Y(p, q, 1, v) \\ Y(p, q, v, 1) \dots Y(p, q, v, v) \end{pmatrix} \quad (8)$$

We see that γ p, q is a matrix with a symmetric and $v + (v^2 + v) / 2$ passes are considered for calculating both and γ . The timing complexity to build the Integral- images is $\text{img}(v^2 w h)$.

Let there be a rectangular shaped region s (p', q' ; p'', q''), where (p', q') is the west-north coordinate and (p'', q'') south-east coordinate. Now let covariance of the region bounded by (1, 1) and (p', q') so from there we can achieve the following equation:

$$a_s(1, 1; p', q') = \frac{1}{c-1} \left[Y_{p', q'} - \frac{1}{c} I_{p', q'} I_{p', q'}^{time} \right] \quad (9)$$

where $c = p' \cdot q'$. On the same way, after a small number of manipulations, the covariance based on the region-of-interest s (p', q' ; p'', q'') can be calculated as:

$$a_s(p', q'; p'', q'') = \frac{1}{c-1} \left[Y_{p', q'} + Y_{p'', q''} - Y_{p', q''} - Y_{p'', q'} - \frac{1}{c} (I_{p', q'} + I_{p'', q''} - I_{p', q''} - I_{p'', q'}) (I_{p', q'} + I_{p'', q''} - I_{p', q''} - I_{p'', q'})^{time} \right] \quad (10)$$

Where $c = (p'' - p') \cdot (q'' - q')$. However, after building Integral-images the covariance of any rectangular shaped-region can be calculated and we have the timing complexity of $O(v^2)$.

C. Matching with Best Suitable Technique

The attainment of a analogous region-of-interest of an apparatus image, we should have to emphasize on the distances among covariance matrices, their related intended images and the nominees of region-of-interest. The

covariance- matrices-space is not same as vector-space. There is no possibility of measuring the distances between two matrices to find out the region-of-interest based on the arithmetic operations only. As no features in the feature-vector-space would be accurately the same that may depict the covariance-matrices depending on their positive values so we must have to use different distance measuring technique. Hence we have used the following equation for removing the discrepancy between covariance- matrices:

$$\eta(a_m, a_n) = \sqrt{\sum_{m=1}^c x c^2 \phi_m(a_m, a_n)} \quad (11)$$

Where $\{\phi_m(a_1, a_2)\}_{m=1, \dots, c}$ are the indiscriminate Eigen-values considering as the parameter of a_m and a_n of the above equation (2).

The measurement of distance η depends on metric maxims, their positive and symmetric values that can be achieved from triangular shaped region of intensity-of-interest. We are trying to find out the smallest distance between the region-of-interest and the current image frame. The most exact matching of region-of-interest identifies the required image in the current image-frame.

D. Feature Extraction by Covariance Tracking

The features extraction through covariance tracking has been shown an improved effect. This tracking output can be achieved by scanning the whole image as that we can locate the target location. To trim down the computation time we have taken help of Integral histograms and a simulated concurrent tracker was gained. The achievement of better alike image with respect to global searching is dawdling for many concurrent functions and is no better scaled for bigger images. Furthermore, penetrating the whole image may lead to inaccuracy that may be diverted by disrupting from analogous image therein in the panorama. The tracking algorithms to anticipate the position of the tracked image from frame to frame alter slowly. So to stop this we have used some optimization techniques as that we can achieve a healthy and proficient tracking methodology depending on covariance features.

III. Image identification

During the phases of image identification, the input image is considered as an object, for the reason of pointing the image arbitrarily and pretence following a flexible alteration. Here we use the pixels locations (p, q) with respect the colour values (RGB). We have taken the first and second order derivation for acquiring the intensities considering the pixels p and q as the parameters. Then it is converted to the required target dimension for acquiring the final features in the form of vector by means of the following equation:

$$\eta(p, q) = \left[\begin{array}{c} p \ q \ R(p, q) \ G(p, q) \ B(p, q) \ \left| \frac{\partial \phi(p, q)}{\partial p} \right| \left| \frac{\partial \phi(p, q)}{\partial q} \right| \left| \frac{\partial^2 \phi(p, q)}{\partial p^2} \right| \left| \frac{\partial^2 \phi(p, q)}{\partial q^2} \right| \end{array} \right]^{time} \quad (12)$$

Where R=Red, G=Green, B=Blue color, and ϕ is the intensity- region-of -interest.

A. Covariance as Image Identifier

We know that the variance of pixel locations (p, q) is identical for the entire regions of the equal size; they are at rest vital since their correlation is measured with other features with respect to the covariance matrix. We have used an image with randomly generated covariance matrices and then the image features are calculated contained by the image region. Firstly, we have calculated the covariance of the entire region from the base-image. Then we look for the goal image for a region with an analogous covariance matrix.

We act upon a novel search, because we wanted to calculate the covariance of a random region rapidly. If the sizes of the same regions changes then their variance should changes so we have taken the normalized values of the rows and columns.

Initially we have identified identical locations along with their scales. Secondly we have reiterated the same with the help of the covariance matrices. Then we have find out variation of the source image model and the goal ny the following formula:

$$\eta(img, time) = \min_n \left[\sum_m^5 \eta(a_m^{img}, a_m^{time}) - \eta(a_n^{img}, a_n^{time}) \right] \quad (13)$$

where a_m^{img} and a_m^{time} are the source and goal covariance respectively. The region with the smallest distinction is chosen as the identical region.

IV. Image Classification

Presently, a large amount of flourishing procedures for image categorization are by pointing the centre position of it and then the feature-span-space is derived from it. The feature-span-space is constructed from the yielded result of the sifted features of the gained pixels. For searching an image the usual methodology is to take help of the k-means-clustering procedure, although it may not be the best choice. Here with the help of covariance along with time series we have tried to identify the image. The procedure is computationally costly as the images are to keep in the large database and it needs an elevated dimensional space clustering technique.

A. Covariance as a Classifier

We propose a novel methodology for image categorization predicament except consideration of the whole image. We begin with extracting numerous features from every pixel. For image categorization predicament we consider image intensities and follow 1st and 2nd order derivatives with respect to p and q. Now the dimensional feature-span-space can be derived from the equation:

$$\eta(p, q) = \left[\phi(p, q) \left\| \frac{\partial \phi(p, q)}{\partial p} \right\| \left\| \frac{\partial \phi(p, q)}{\partial q} \right\| \left\| \frac{\partial^2 \phi(p, q)}{\partial p^2} \right\| \left\| \frac{\partial^2 \phi(p, q)}{\partial q^2} \right\| \right] \quad (14)$$

The image space is chosen randomly from each image and the sizes will lie within 64×64 and 556×556. Integral images are used for computation of every region along with covariance matrix. Every source image (simg) is then stands for sm covariance matrices and we propose exercise image(eimg) from every image cluster and a sum of simg*eimg covariance matrices. The above process is repeated as earlier.

B. Classification Experiment

We have tried to achieve our experimental result based on Brodatz-image-dataset where more than 100 images are available and we have taken challenge considering the heterogeneous images within the dataset. Every image has been subdivided into sub-images as that we can make it available as exercise image set for our testing. Here we have used k-means-clustering formula and we have achieved 98.61% accuracy based on our image-dataset.

With our exercise image set and our 70 images from our image dataset identified them by arbitrarily chosen covariance for every image we have achieved the above mentioned result i.e. 98.61% It is better with respect to faster identification as well as reorganization rate.

We have taken into considered some basic features namely-intensity, gradient and Laplacian as that we can achieve better performance for classification and we reached to attain 96.41%. Thus we have gained the following tabular result:

	INT	GRAD	LM	ARBITRARY COVARIANCE
RESULT	87.31	96.41	97.91	98.61

Table 1. Performance Result set.

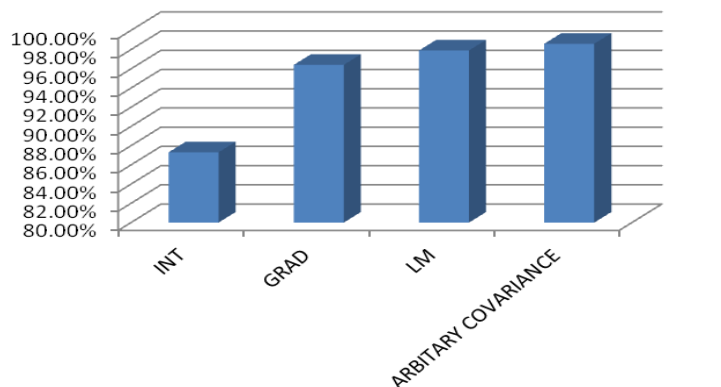


Fig: Resultant Chart

V. Conclusion

In our paper we presented the all features based on covariance and the k-means-formula is used for image identification and their categorization. We have given emphasized on some basic features: intensity, gradient and Laplacian to gain better performance for classification and identification. We have the idea to apply it in

wider way for video and as well as audio application for tracking of images in a complex environment and circumferences irrespective of distance. Our methodology has been based on actual images. If we can detect it without noise then efficacy will raise to a greater extend and it will be much more robust.

VI. References

- [1] Ismail Oner Sebe, Jinhui Hu, Suyu You, Ulrich Neumann, "3D Video Surveillance with Augmented Virtual Environments", ACM SIGMM 2003 Workshop on Video Surveillance, pp 107-112, Berkeley, California, November 2003, Berkeley, CA, pp.107-112..
- [2] Juan Pavona, Jorge Gomez-Sanz, Antonio Fernandez-Caballero, Julian J. Valencia-Jimenez "Development of intelligent multisensor surveillance systems with agents" Elsevier, Robotics and Autonomous Systems 55 (2007) 892–903
- [3] Takeo Kanade, Robert T Collins, Alan J. Lipton, "Advances in Cooperative Multi-Sensor Video Surveillance", Peter Burt, Lambert Wixson - Proceedings of DARPA Image Understanding Workshop, 1998.
- [4] Robert T. Collins, Alan J. Lipton, Takeo Kanade, Hironobu Fujiyoshi, David Duggins, Yanghai Tsin, David Tolliver, Nobuyoshi Enomoto, Osamu Hasegawa, Peter Burt and Lambert Wixson, "A System for Video Surveillance and Monitoring" CMU-RI-TR-00-12 citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.73...pdf.
- [5] Kim C. Ng, Hiroshi Ishiguro, Mohan Trivedi, Takushi Sogo "An Integrated Surveillance System—Human Tracking and View Synthesis using Multiple Omni-Directional Vision Sensors", Image and Vision Computing Journal, June 2002.
- [6] Comaniciu, D., Ramesh, V., Meer, P.: Real-time tracking of non-rigid objects using mean shift. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Hilton Head, SC. Volume 1. (2000) 142–149.
- [7] Porikli, F.: Integral histogram: A fast way to extract histograms in cartesian spaces. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition, San Diego, CA. Volume 1. (2005) 829 – 836.
- [8] N Leung, T., Malik, J.: Representing and recognizing the visual appearance of materials using three-dimensional textons. Intl. J. of Comp. Vision 43 (2001) 29–44. 2001 Kluwer Academic Publishers. Manufactured in The Netherlands
- [9] Varma, M., Zisserman, A.: Statistical approaches to material classification. In: Proc. European Conf. on Computer Vision, Copenhagen, Denmark. (2002), IEEE PAMI, VOL. X, NO. Y, APRIL 2006.
- [10] Georgescu, B., Meer, P.: Point matching under large image deformations and illumination changes. IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 26, NO. 6, JUNE 2004-Pgs: 674–688.
- [11] Viola, P., Jones, M.: "Rapid object detection using a boosted cascade of simple features". In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Kauai, HI. Volume 1. (2001) 511–518.
- [12] Kentaro Toyama, John Krumm, Barry Brumitt, Brian Meyers, "Wallflower: Principles and Practice of Background Maintenance", Computer Vision, 1999. The Proceedings of the Seventh IEEE, Vol: 1 Dig Obj Iden: 10.1109/ICCV.1999.791228, 1999, Pg: 255 – 261.
- [13] C. E. Liedtke and R. Koch, "MODELLING OF 3D SCENES FROM THE ANALYSIS OF STEREOSCOPIC IMAGE SEQUENCES" citeseerx.ist.psu.edu/viewdoc/download
- [14] Anthony R. Dick, Michael J. Brooks, "Issues in Automated Visual Surveillance", New Scientist, 12 July 2003, page 4.
- [15] Liang Wang, Weiming Hu, Tieniu Tan, "Recent Developments in Human Motion Analysis", National Laboratory of Pattern Recognition Institute of Automation, Chinese Academy of Sciences, Beijing, P. R. China, 100080.
- [16] Norman Poh, Chi Ho Chan, Josef Kittler, Sébastien Marcel, Christopher Mc Cool, Enrique Argones Rúa, Jose Luis Alba Castro, Mauricio Villegas, Roberto Paredes, Vitomir Struc, Nikola Pavesic, Albert Ali Salah, Hui Fang, Nicholas Costen, "An Evaluation of Video-to-Video Face Verification", IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, VOL. 5, NO. 4, DECEMBER 2010.
- [17] Norman Poh, Chi Ho Chan, Josef Kittler, Sébastien Marcel, Christopher Mc Cool, Enrique Argones Rúa, Jose Luis Alba Castro, Mauricio Villegas, Roberto Paredes, Vitomir Struc, Nikola Pavesic, Albert Ali Salah, Hui Fang, Nicholas Costen, "Face Video Competition", M. Tistarelli and M.S. Nixon (Eds.): ICB 2009, LNCS 5558, pp. 715–724, 2009. Springer-Verlag Berlin Heidelberg 2009.