International Journal of Information Processing, 6(4), 1-12, 2012
ISSN : 0973-8215
IK International Publishing House Pvt. Ltd., New Delhi, India

A Local Histogram Based Descriptor for Object Tracking in Wide Area Motion Imagery

Alex Mathew[a], Vijayan K. Asari[b]

[a]Department of Electrical and Computer Engineering, Kettering Laboratories Room 341, 300 College Park, Dayton, USA, OH 45469-0232, Contact: mathewa3@udayton.edu

[b]Professor, Department of Electrical and Computer Engineering and Ohio Research Scholars Chair Wide-Area Surveillance, Department of Electrical and Computer Engineering, Kettering Laboratories Room 341, 300 College Park, Dayton, USA, OH 45469-0232, Contact: vasari1@udayton.edu

In this paper we propose a novel feature extraction technique and show its application in tracking in low resolution videos. The aspects that make tracking particularly challenging are global camera motion, large target movement, poor gradient and texture information and absence of color information. Global camera motion is reduced or eliminated by registering the images from frame to frame employing SURF (Speeded Up Robust Feature). The proposed method is based on intensity histogram, but with a variant that encodes both spatial and intensity information. The method is evaluated on CLIF (Columbus Large Image Format) data. The robustness of the feature eliminates the need for background subtraction in videos. A performance comparison of our approach with other object descriptors such as HOG (Histogram of Gradients), SURF and SIFT (Scale Invariant Feature Transform) shows the effectiveness of the proposed descriptor.

Keywords : Aerial Images, CLIF Data, Low Resolution Object Detection, Tracking, Wide Area Surveillance.

1. INTRODUCTION

Finding reliable features for visual tracking in low resolution imagery is a hard problem. The proposed method focuses finding robust features for detection and tracking in wide area motion imagery. This problem is of immense interest in the area of Wide Area Surveillance. Although there are several feature extraction techniques available for high resolution images, these methods fail in low resolution scenarios. The method described in this paper is evaluated on Columbus Large Image Format(CLIF) data [1]. The data is captured using cameras mounted on flying platforms at a height of approximately 7000 ft. The camera captures images of size 2672 x 4008 at a rate of 2 frames per second. Targets of interest span only a few pixels.

In object recognition, any available cue is of interest. They include-

Physical characteristics: Color, texture, physical structure, location, orientation, depth map etc. These features are not in CLIF data.

Behavioral Patterns: This includes the behavioral patterns such as the gait of a person, the trajectory, acceleration and speed of a vehicle *etc.*. Low video frame rate in this data results in large variations between consecutive frames making trajectory prediction hard.

Environment Context: This includes the physical environment features. For example, a car is likely to be on a road or a parking spot rather than on the top of a building. Such contexts are very useful for effective search and recognition of objects. However in this data, areas such as road cannot be reliably extracted since there is very little texture information.

There are two sub-areas in object recognition-

(a) Target. (b) Background. (c) Reference.

Figure 10. Target, Background and Reference Images.

Bhattacharya coefficient is

$$\rho(p, q) = \sum_{i=1}^{m} \sqrt{p_i q_i} \tag{7}$$

Using Equation 7, the distance between the two distributions is defined as

$$d(y) = \sqrt{1 - \rho(p, q)} \tag{8}$$

Figure 9 shows tracking three targets using mean-shift. As can be seen only one target is tracked correctly. This is because mean-shift tracking cannot cope with large target movements. For the computation of the mean-shift vector, successive target locations need to be overlapping. Figure 10 shows a target image, part of the background region and the reference image. For the target to be matched correctly with the reference, the distance between the reference and the target has to be less than that between the reference and the background. Table 1 shows a comparison of the metric used in our method, using chi-square metric without dividing the image into regions and the metric used in mean-shift tracking (Equation 8). Dividing the image into grids is the key to correct matching and localization. As noted in Table 1, the target cannot be correctly matched when the histograms are compared with chi-square metric without dividing the image or when the metric used in mean-shift tracking is employed. Figure 11-13 shows the pixel errors for three targets.

5. CONCLUSIONS

We have presented a novel feature extraction technique that addresses the problem of detection and tracking in low resolution domain. A comparison with other techniques such as HOG, SURF, SIFT and mean-shift tracking shows that our descriptor is far superior to such techniques when there is very little textural and structural information. In addition, since the approach is based on histogram, it is less computationally intensive than most other methods. We have shown the effectiveness of our approach in tracking in wide area CLIF data. A scheme that uses cross bin similarity metrics instead of bin-bin similarity metrics like chi-square is currently being developed.

REFERENCES

1. AFRL CLIF Dataset 2007, https://www.sdms .afrl.af.mil/index.php?collection=clif2007.
2. Veenman C, Reinders M, Backer E. Resolving Motion Correspondence for Densely Moving Points, *IEEE Trans. on Pattern Analysis Machine Intelligence,* 23(1):54-72, 2001.
3. Serby D, Koller-Meier S, Gool L V. Probabilistic Object Tracking using Multiple Features, *In IEEE International Conference on Pattern Recognition (ICPR),* pages 184-187, 2004.
4. Comaniciu D. Bayesian Kernel Tracking, *In Annual Conference of the German Society for Pattern Recognition,* pages 438-445, 2002.
5. Ballard D, Brown C. Computer Vision, *Prentice-Hall,* 1982.
6. Yilmaz A, Li X, Shah M. Contour based Object Tracking with Occlusion Handling in Video Acquired using Mobile Cameras, *IEEE*

Table 1
Performance Comparison of Histogram Similarity Metrics.

	Our Method	Undivided Image Histogram	Metric in Mean shift Tracking
Dist(reference,target)	3.1654	0.2968	0.3252
Dist(reference,background)	4.2863	0.2902	0.3120

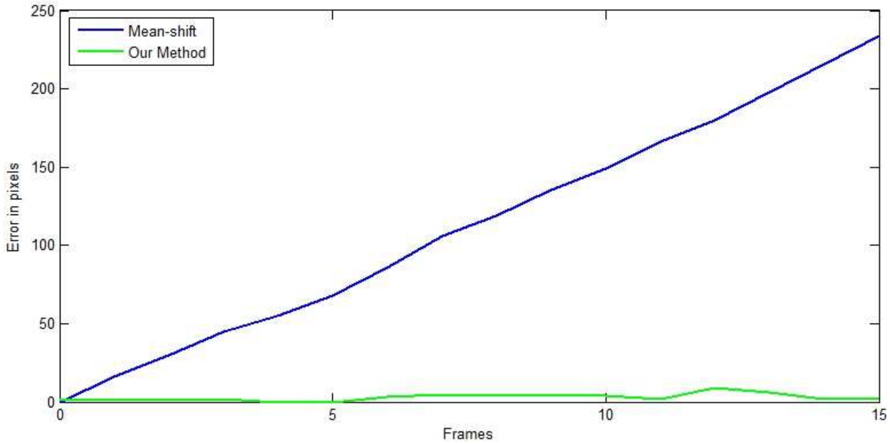

Figure 11. Error Plot for Target 1

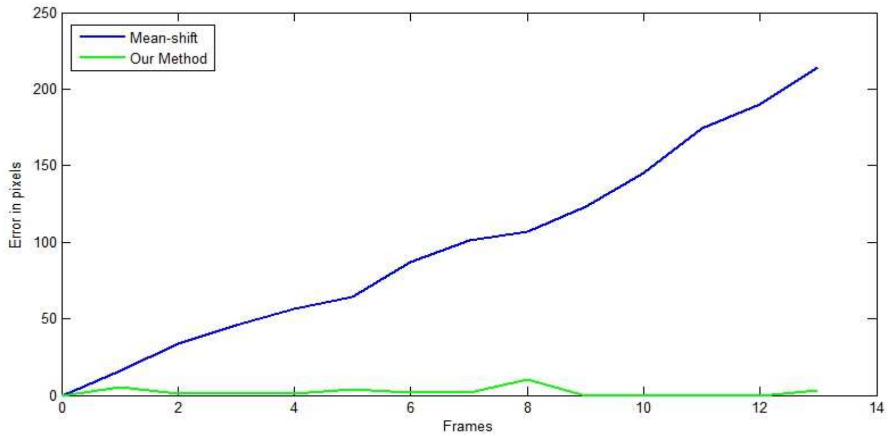

Figure 12. Error Plot for Target 2

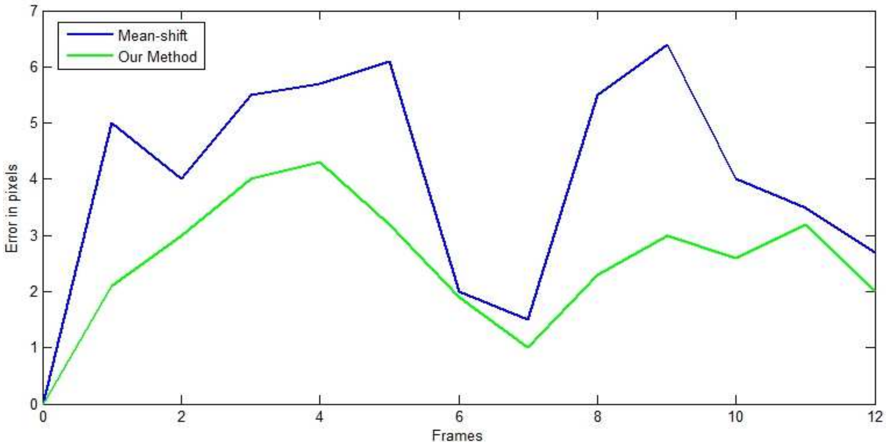

Figure 13. Error Plot for Target 3

Trans. on Pattern Analysis Machine Intelligence, 26(11):1531-1536, 2004.

7. Fieguth P, Terzopoulos D. Color-based Tracking of Heads and other Mobile Objects at Video Frame Rates, *In IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* pages 21-27, 1997.
8. Canny J. A Computational Approach to Edge Detection, *IEEE Trans. Pattern Analysis Machine Intelligence,* 8(6):679-698, 1986.
9. Haralick R, Shanmugam B, Dinstein I. Textural Features for Image Classification, *IEEE Trans. System Man Cybern.* 33(3):610-622, 1973.
10. Mallat S. A Theory for Multiresolution Signal Decomposition: The Wavelet Representation, *IEEE Trans. Pattern Analysis Machine Intelligence,* 11(7):674-693, 1989.
11. Greenspan H, Belongie S, Goodman R, Perona P, Rakshit S and Anderson C. Overcomplete Steerable Pyramid Filters and Rotation Invariance, *In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* pages 222-228, 1994.
12. Bay H, Ess A, Tuytelaars T, Gool L. SURF: Speeded Up Robust Features, *International Journal on Computer Vision and Image Understanding (CVIU),* 110(3):346-359, 2008.
13. Lowe D G. Distinctive Image Features from Scale-invariant Keypoints, *International Journal on Computer Vision,* 60(2):91-110, 2004.
14. Juan L, Gwun O. A Comparison of SIFT, PCA-SIFT and SURF, *International Journal of Image Processing (IJIP),* 3(4):143-152, 2009.
15. http://www.mathworks.com/matlabcentral/fileexchange/28300-opensurf-including-image-warp
16. Dalal N, Triggs B. Histograms of Oriented Gradients for Human Detection, *In the Proceedings of International Conference on Computer Vision and Pattern Recognition,* vol. 2, pages 886-893, 2005.
17. Perez P, Hue C, Vermaak J, Gangnet M. Color-Based Probabilistic Tracking, *Proceedings of European Conference on Computer Vision,* vol. 1, pages 661-675, 2002.
18. Ahonen T, Hadid A, Pietikäinen M. Face Recognition with Local Binary Patterns, *In the Proceedings of Eighth European Conference on Computer Vision,* pages 469-481, 2004.
19. Comaniciu D, Meer P. Mean Shift: A Robust Approach Toward Feature Space Analysis, *IEEE Trans. on Pattern Analysis and Machine Intelligence,* 24(5):603-619, 2002.
20. Ester M, Kriegel H, Sander J, Xu X. A Density Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise, *In the Proceedings of* 2^{nd} *International Conference on Knowledge Discovery and Data Mining,* pages 226-231, 1996.
21. Comaniciu D, Ramesh V, Meer P. Real-time Tracking of Non-rigid Objects Using Mean Shift, *In proceedings of IEEE Conference on Computer Vision and Pattern Recognition,* vol. 2, pages 142-149, 2000.

Alex Mathew received his MS in Electrical Engineering from University of Rochester, NY in October 2006. He received his B. Tech in Electronics and Communication Engineering from National Institute of Technology (NIT), Calicut, India in June 2-002. He is currently doing his Ph.D at the University of Dayton OH. His research interests include Pattern Recognition, Computer Vision, Image Processing and High Performance Computing.

Dr. Vijayan K. Asari is a Professor in Electrical and Computer Engineering and Ohio Research Scholars Endowed Chair in Wide Area Surveillance at the University of Dayton, Dayton, Ohio, USA. He is the director of the Computer Vision and Wide Area Surveillance Laboratory (UD Vision Lab) at UD. Dr. Asari received his Bachelor's Degree in Electronics and Communication Engineering from the University of Kerala, India in 1978, MTech and Ph.D Degrees in Electrical Engineering from the Indian Institute of Technology, Madras in 1984 and 1994 respectively. Prior to joining UD in February 2010, Dr. Asari worked as Professor in Electrical and Computer Engineering at Old Dominion University, Norfolk, Virginia, Research Fellow at National University of Singapore and Nanyang Technological University, Singapore and Assistant Professor at University of Kerala, India. Dr. Asari holds two patents and has published more than 330 research papers, including 64 peer-reviewed journal papers in the areas of Image Processing, Computer Vision, Machine Learning, Pattern Recognition and High Performance Digital Architectures. He has supervised 14 Doctoral Thesis and 28 Masters Dissertations since joining ODU in August 2000. Dr. Asari received several teaching, research and advising awards. He is a Senior Member of IEEE and SPIE.