International Journal of Electrical, Electronic and Communication Sciences
ISSN: 2517-9438
Vol:3, No:12, 2009

Object Tracking using MACH filter and Optical Flow in Cluttered Scenes and Variable Lighting Conditions

Waqar Shahid Qureshi, Abu-Baqar Nisar Alvi

***Abstract*—**Vision based tracking problem is solved through a combination of optical flow, MACH filter and log r-θ mapping. Optical flow is used for detecting regions of movement in video frames acquired under variable lighting conditions. The region of movement is segmented and then searched for the target. A template is used for target recognition on the segmented regions for detecting the region of interest. The template is trained offline on a sequence of target images that are created using the MACH filter and log r-θ mapping. The template is applied on areas of movement in successive frames and strong correlation is seen for in-class targets. Correlation peaks above a certain threshold indicate the presence of target and the target is tracked over successive frames.

***Keywords*—**Correlation filters, optical flow, log r-θ mapping.

I. Introduction

FOR vision based tracking, a practical system should be able to identify and track object in cluttered scenes and variable lighting conditions. To track target objects in video sequences, a system must first be designed to detect and correctly classify objects. For a video sequence, an object of interest in one frame can then be associated with the same object in the coming frames. There are number of approaches to address this problem. Most of the algorithms have been applied on indoor scenes in ideal lighting and background conditions. All the approaches are based upon some assumptions like fixed camera distance from the object, defined range in the variations of object size and orientation, known lighting conditions etc. In case of fixed camera position, the process of searching an object in individual video frames can be greatly optimized by only searching for those portions of the image that are moving. The moving parts of the image can be extracted by keeping a running average image of the previous n frames and subtracting it from the current frame. However, such a scheme cannot be used in a scenario where either the camera is moving or both object and camera are moving.

The problem of recognizing objects of interest or targets in video sequences despite distortions in position, orientation and scale has been addressed using various techniques. One of these techniques is based on the development of Synthetic Discrimination Function (SDF) [1]. SDF is a filter design method to create invariance to such distortions. Based on the SDF, attempts have been made to introduce a class of filters that use correlation for object recognition. Some of these filters are Maximum Average Correlation Height (MACH) filter, Minimum Average Correlation Energy (MACE), Minimum Variance Synthetic Discriminant Function (MVSDF) [2] and Distance Classifier Correlation Filter (DCCF) [3]. All these filters can be tuned to give maximum performance and allow object recognition that is immune to background clutter. Recently two techniques have been combined to solve the invariance problem by Bone, Young and Chatwin [4]. The technique combines MACH filter to achieve out-of-plane rotation and scale and in-plane rotation invariance is created with the use of a log r-θ mapping (log-map) of a localized region of the image space. The results show reasonable success when applied in ideal lighting and uniform background. However, the technique works only for a fixed camera video sequence as it uses background subtraction for isolating moving objects in video frames. We have used the same model as proposed by Bone, Young and Chatwin [4] but instead of using background subtraction, the technique uses optical flow for object detection. This has allowed the system to be used in moving camera scenarios. The MACH filter parameters have also been altered to allow it to function in variable lighting conditions.

II. Maximum Average Correlation Height (MACH) filter

The object recognition MACH filter is based on the SDF in which linear combination of the various reference or training images are combined to make a composite image. The MACH filter maximizes the relative height of the average correlation peak with respect to the expected distortions [2]. Unlike the SDF, the MACH filter can be tuned to maximize the correlation peak height, peak sharpness and noise suppression while also being tolerant to distortions in the target object that fall between the distortions given in the training set [4]. MACH filter has been studied, improved and used for various Automatic Target Recognition (ATR) applications and problems [5], [6]. The details of MACH filter and its usage have been explained in [7] and [8]. The basic construction of the MACH filter is given in (1).

Engr. W.S.Qureshi is Lecturer at Air University, PAF- Complex, E-9 Islamabad (Phone: +923028177897, E-mail: waqar.shahid@mail.au.edu.pk).
Engr. A.N.Alvi is PhD Student at Center for Advance Studies in Engineering, Sir-syed Memoiral building, G-5 Islamabad (Phone: +923335261723, E-mail: ab_alvi@yahoo.com).

$$H(f) = \frac{N_x^*}{\alpha C + \beta D_x + \gamma S_x} \quad (1)$$

where α, β, and γ are non-negative OT parameters, Nx is the average of the training image vectors X1, X2, ….XN (in the frequency domain), and C is the diagonal power spectral density matrix of additive input noise. Dx is the diagonal average power spectral density of the training images given in (2).

$$D_x = \frac{1}{N}\sum_{i=1}^{N} X_i^* X_i \quad (2)$$

and Sx is given in (3).

$$S_x = \frac{1}{N}\sum_{i=1}^{N} (X_i - M_x)^* (X_i - M_x) \quad (3)$$

where Mx is the average of Xi.

III. Use of Log r-Θ Mapping

Log r-θ mapping has been used for invariance to in-plane rotation and scale variations. The algorithm has been explained in [9] for detection of straight lines, circles and line segments. The technique exploits the phenomenon of foveal sampling seen in mammalian eyes. In the mammalian retina, photoreceptive cells are small and densely packed in the fovea and increase in size exponentially to create a blurred periphery. This technique maps each sensor pixel on the circular region of the x-y Cartesian space into a rectangular region in Polar r-θ space. The sensor's geometry maps concentric circles in the Cartesian space into vertical lines in the polar space and radial lines in the Cartesian space into horizontal lines in the Polar space. The mapping follows the mathematical relationship shown in (4), (5), (6) and (7).

$$Z = e^{-w} \quad (4)$$

$$w = \log Z \quad (5)$$

where

$$Z = x + iy \quad (6)$$

and

$$w = u + iv \quad (7)$$

Fig.1 shows a 128 x 128 grey scale image and its log-map respectively.

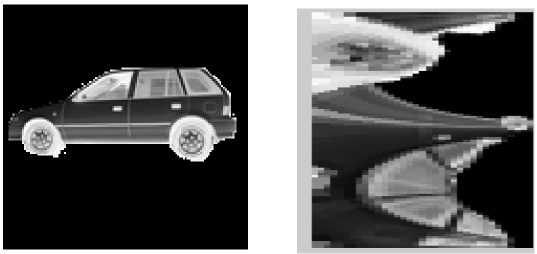

Fig. 1 Image of car with its log r-θ Map

IV. Optical Flow

Object recognition or target tracking problems in a video or sequence of frames becomes very complex. Recently methods have been developed for such object recognition and problems and one of such method is Optical Flow [10]. This method is very useful in videos that contain moving objects. Though it is computationally expensive, but still a robust method to detect moving objects in a scene. Optical flow is a relation of the motion field. It is the 2D projection of the physical movement of points relative to the observer, to the 2D displacement of pixel patches on the image plane. The main assumption and limitation behind this method is that the Image Intensity (I) in a frame remains the same for a small change in position for the next frame. The relationship between two consecutive frames is given in (8),

$$I(x+d_x, y+d_y, t+d_t) = I(x,y,t) \quad (8)$$

Using the Taylor's Series and eliminating higher order terms we get,

$$I(x+d_x, y+d_y, t+d_t) = I(x,y,t) + I_x d_x + I_y d_y + I\ldots \quad (9)$$

thus (8) and (9) becomes

$$I_t = -(I_x\,[d_x / d_t] + I_y\,[d_x / d_t]) \quad (10)$$

where I_x is the differential w.r.t x in successive frame, I_y is the differential w.r.t y in successive frame and I_t is the temporal differential (difference of two consecutive frames). Thus [dx/dt] and [dy/dt] are the required velocity vectors magnitude. The equation to get the velocity vector is non trival. Lucas & Kanade [10] proposed a technique to solve the problem. The velocity vector for each frame has been obtained using this method, thus, we find all the moving objects in the frame that have motion relative to the frame.

V. Implementation

The implementation of the project was done similar to the work presented in [4]. The major difference was the creation of template images and video sequences as they were done in outdoor scenes. The template images were created offline by taking pictures from a high-resolution digital camera. The images were then cleaned in terms of background noise using offline techniques. White background was selected for the actual images. The images were then converted to gray scale and complemented. Some of the training images used in simulations are shown in Fig.2. All images were then down-scaled to 128 by 128 sized for speed optimization of the algorithm.

Video sequences were taken in outdoor environment using a digital camera at a resolution that was less than the still image resolution. The system was expected to correctly detect and track the White Suzuki Cultus car in the video sequences despite variations in its out-of-plane orientation, in-plane rotation, and scale, position with noisy or cluttered backgrounds.

International Journal of Electrical, Electronic and Communication Sciences
ISSN: 2517-9438
Vol:3, No:12, 2009

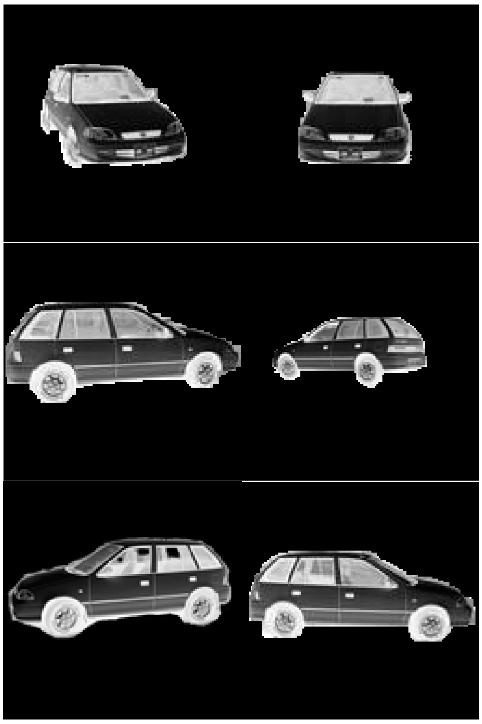

Fig. 2 Training Images used in MACH filter design

To create invariance to changes in out-of-plane orientation of the car, the MACH filter was created using a training image set consisting of the expected range of rotation. The log-map of each training image was created and then all these log-maps were combined using the MACH filter. The resulting template was then used for target detection in all frames of the video sequence.

To detect the presence of moving objects in video frames, optical flow was used. Optical flow highlighted all possible regions where the moving object of interest or target could be found. A 128 by 128 sub-image, centered over each region of interest (ROI), was log-mapped and correlated with the MACH filter. The correlation peak height of the ROI was compared to a threshold - above which the ROI was classified as containing an in-class target. The threshold was calculated by correlating each of the reference images with the MACH filter and taking an average of their resultant peak intensities. This value then gave a value close to what would be expected after correlation between the filter and an in-class target.

VI. Results

A video sequence in the outdoors was captured using a digital camcorder. Initially, the system was tested using a single moving target. The Cultus car was driven in a way that its orientation was fixed from the point of view of the camera. The camera was kept stationary throughout the capturing of the video sequence. Only those template images were used for training of the MACH filter that covered that specific orientation of the car. The size of the images and the video sequence was reduced to 128 by 128 to optimize the algorithm speed and decrease the computational burden. The system performed well and classified the moving car as a target in every frame. This was despite the fact that the video image was of slightly lower quality than the training images, which had been taken using a high resolution digital still camera, due to video compression. The outdoor lighting conditions during the filming of the video sequences also made it different to that of the training images, but the filter was able to cope well with this. Fig.3 shows the result of optical flow and the correlation with MACH filter in one of the frames of the video. The arrows indicate presence of a moving object and the mesh shows the correlation plane.

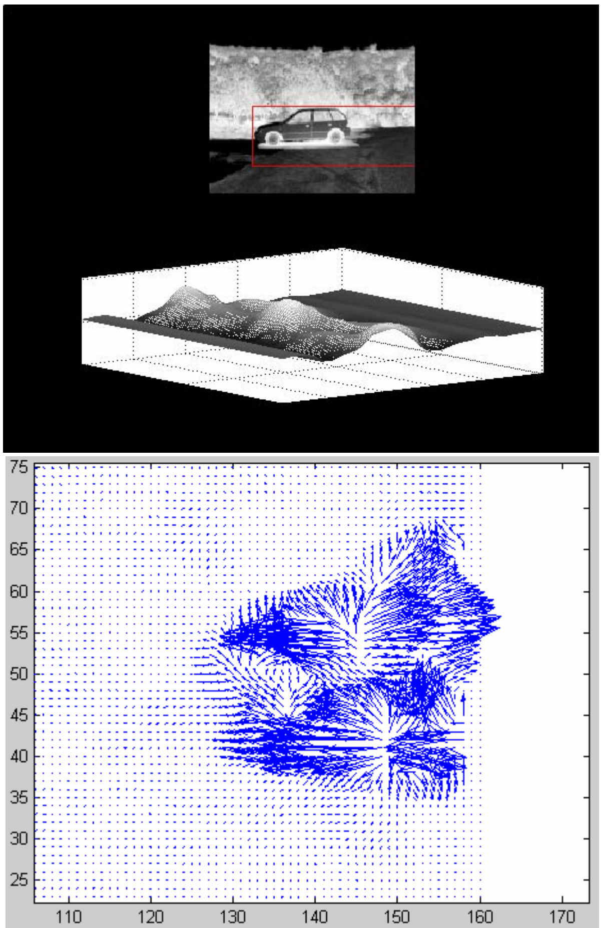

Fig. 3 Optical Flow and MACH filter results on the target for a video frame filmed with a stationary camera

More video sequences were obtained in the same outdoor environment but this time there were multiple objects in the scene. The camera was also moved during the filming of the video sequence. Optical flow was still able to detect the motion in the video sequence and MACH filter was able to classify the in-class target. The results are shown for various frames in Fig.4 where the car is moving across another car in the scene and the MACH filter only recognizes the moving car as the target in the scene with the camera also moving during the filming of the video sequences.

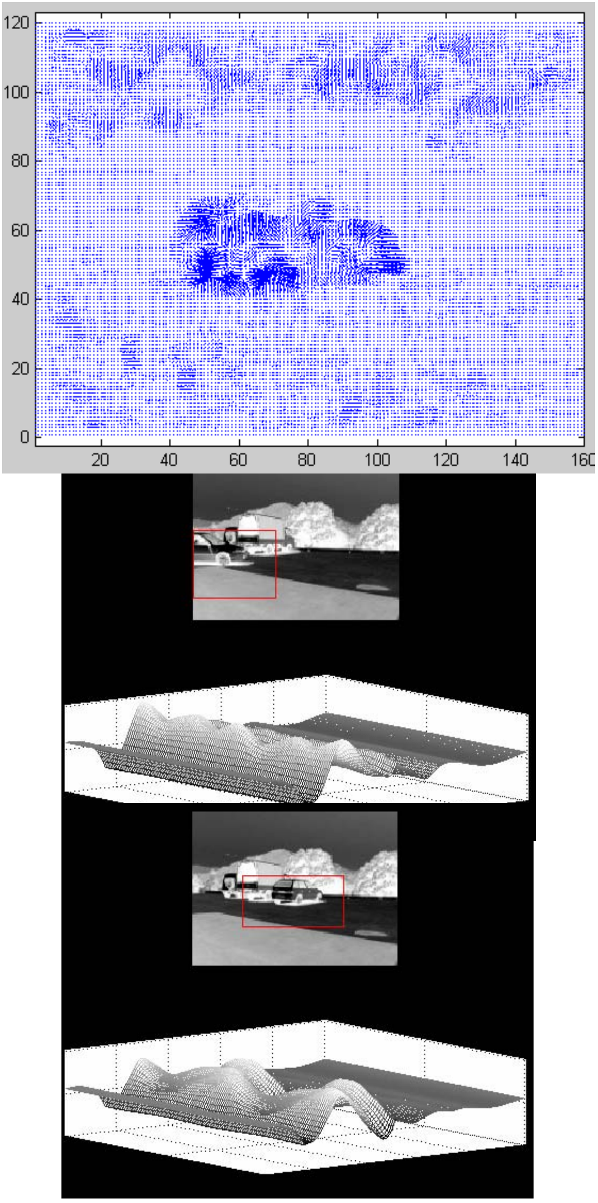

Fig. 4 Optical Flow and MACH filter results on the target for video frames with multiple objects in the scene and the camera also moving

VI. Conclusion

The combination of MACH filter, log-mapping and optical flow was able to create a system that could be used to detect targets in cluttered environment for both stationary and moving camera scenarios. Both MACH filter and log-map provide invariance to position, orientation and scale changes in cluttered backgrounds and variable lighting and optical flow provides information about presence of motion in video sequences. The main objective to use the optical flow as opposed to the method used by Bone, Young and Chatwin [4] was to find the moving objects even when the camera is moving with a constant velocity. The method used in [4] fails to detect and track when camera is moving. A method of classification of correlation peaks as in-class or out-of-class objects was employed by computing the average expected peak intensity. This made object detection simple since the height of the peak for a potential target could directly be compared to a predefined threshold that defines whether the ROI contains a target or not. All simulations were run using MATLAB and all results have been obtained within the same simulation environment. The most time critical portion of the routine is the two-dimensional Fourier transform used in MACH filter correlation. For practical and realizable systems, MATLAB code has to be converted to some higher level language like C so that it can be run on an embedded platform. The work has the potential to be used in real-time systems where the entire algorithm can be optimized for use on an embedded system.

Acknowledgment

W.S Qureshi and A.N. Alvi acknowledge the efforts of Dr. Imtiaz. A. Taj who taught Digital Image Processing and Computer Vision and thanks him for his motivation for this research. Engr. W.S Qureshi also wants to thank Engr. Bahman R. Alyaei for proof reading and giving guidelines for the preparation of the paper.

References

[1] C. F. Hester and D. Casasent, “Multivariant technique for multiclass pattern recognition”, Applied Optics, Vol. 19, 1758-1761 (1980).

[2] A. Mahalanobis, B.V.K. Vijaya Kumar, S. Song, S.R.F. Sims, J.F. Epperson, “Unconstrained correlation filters”, Applied Optics, Vol. 33, pp. 3751-3759, 1994.

[3] A. Mahalanobis, B.V.K. Vijaya Kumar, S.R.F. Sims, “Distance-classifier correlation filters for multiclass object recognition”, Applied Optics, Vol. 35, pp. 3127-3133, 1996.

[4] P. Bone, R. Young and C. R. Chatwin, “Position, rotation, scale and orientation invariant multiple object recognition from cluttered scenes”, Optical Engineering, Vol. 45, pp. 077203-1to -8, 2006.

[5] *A.* Mahalanobis, A. Forman, M. Bower, N. Day, and R. F. Cherry, “A quadratic distance classifier for multi-class SAR ATR using correlation filters,” in Ultrahigh Resolution Radar, R. S. Vickers, ed. Proc. SPIE 1875, 84–95, 1993.

[6] A. Mahalanobis, D. W. Carlson, and B. V. K. Vijaya Kumar,“Evaluation of MACH and DCCF correlation filters for SAR ATR using MSTAR public data base,” in *Algorithms for Synthetic Aperture Radar Imagery V*, E. G. Zelnio, ed. Proc. SPIE 3370, 460–469, 1998.

[7] I. Kypraios, R. C. D. Young, P. Birch, C. Chatwin, "Object recognition within cluttered scenes employing a Hybrid Optical Neural Network (HONN) filter", *Optical Engineering*, Special Issue on Trends in Pattern Recognition, Vol. 43, No. 8, 1839-1850, 2004.

[8] Ph. Refregier, “Optimal trade-off filters for noise robustness, sharpness of the correlation peak and Horner efficiency”, *Optics Letters*, Vol. 16, No. 11, 829-831, 1991.

[9] D. Young, “Straight lines and circles in the log polar image”, *Proc. Of British Machine Vision conference,* 2000.

[10] B.D. Lucas, T. Kanade, “An iterative image registration technique with an application to stereo vision”, *Proceedings of Imaging understanding workshop*, pp 121—130, 1981.