A MULTI-FRACTAL FORMALISM FOR STABILIZATION, OBJECT DETECTION AND TRACKING IN FLIR SEQUENCES

H. Shekarforoush *R. Chellappa*

Center For Automation Research
University of Maryland
College Park, MD 20742, U.S.A.
email: {hshekar,rama}@cfar.umd.edu

ABSTRACT

In this paper, we investigate the problem of stabilization, and detection and tracking of moving or stationary objects in a forward-looking infrared (FLIR) sequence. A multi-fractal formalism is proposed for stabilization and activity detection and the method has been compared to three other classical techniques in image processing.

1. INTRODUCTION

In surveillance applications, stabilizing FLIR data is crucial for detection and tracking of objects [1, 2, 3, 4, 5, 6], and also for fusion of data across possibly different sensors (e.g. FLIR and LADAR) [7, 8, 9, 10]. Existing algorithms for FLIR data [1, 2, 3, 4, 5, 6] usually solve the problems of stabilization, object detection and tracking separately. For FLIR data however these three problems may be combined into a single process, because in general inter-frame motions are restricted to lateral translations or tilts and scale changes. The method proposed in this paper exploits these sensor motion constraints for performing simultaneous activity detection and stabilization.

We have compared our method with three other classical methods of multi-scale optical flow, phase correlation and Kalman filtering.

2. MULTIFRACTAL FORMALISM

In this section, we first recall some fundamental results on the applications of multifractal techniques in image processing. We then formulate the problem of object detection as that of segmenting an image function using a measure of its local singularity. The measure that we propose is the Hölder exponent of some hybrid capacity at a given scale. Using this hybrid measure, we define Lipschitz signatures, which reflect the singularity of the image function along each spatial axis. These signatures can be used for detection and tracking of objects, and hence stabilization of the image sequences.

In order to define the Lipschitz signatures, we shall use the concept of a Choquet capacity [11], which is defined in terms of some basic concepts in fractal theory.

Let E be a set, and $\mathcal{P}(E)$ a sequence of partitions of E. The pair (E, E_p) is refered to as a paved space, where E_p is a collection of all subsets of E. A Choquet capacity is defined as a set function c from $\mathcal{P}(E)$ to $\mathbb{R}^+$ such that

- c is non-decreasing.
- For any increasing sequence of points $\{e_n\}$ in (E, E_p)

$$c\left(\cup_n e_n\right) = \sup_n c(e_n)$$

- For any decreasing sequence of points $\{e_n\}$ in (E, E_p)

$$c\left(\cap_n e_n\right) = \inf_n c(e_n)$$

For image processing applications [12] the set E can be considered as the unit square $[0,1)\times[0,1)$, in which case the collection of partitions can be found by using an increasing sequence of positive integers ν_n

$$E_p = E_{ijn} = \left\{ \left[\frac{i}{\nu_n}, \frac{i+1}{\nu_n}\right) \times \left[\frac{j}{\nu_n}, \frac{j+1}{\nu_n}\right) \right\}$$

The local singularity for each point in the unit square is then given by the Hölder coefficients

$$\alpha(x,y) = \lim_{\delta \to 0} \frac{\log c(\mathcal{B}_\delta(x,y))}{\log \mu(\mathcal{B}_\delta(x,y))} \tag{1}$$

where $\mathcal{B}_\delta(x,y)$ is an open ball of diameter δ centered at (x,y), and μ is a measure used as a reference. We shall hereafter assume that for our applications $\log \mu(\mathcal{B}_\delta(x,y))$ tends to a positive finite constant.

The Hölder exponents given in Eq. (1) reflect the local behavior of the measure c in the neighborhood of (x,y) with respect to μ. Image analysis using multifractal theory is based on computing the fractal spectrum. The spectrum is given by the pair $(\alpha, \dim E(\alpha))$, where this graph is determined by either the Hausdorff spectrum, the Legendre spectrum or the large deviation spectrum. The key in using multifractal tools in image analysis is to design a sequence of Choquet capacities (measures) c that can extract the appropriate information on local and global behavior of the image function. Herein, we have used α for characterizing the local behavior, while a probabilistic method is used for characterizing the global behavior. The latter may be viewed as providing similar information as the large deviation spectrum. In order to apply Eq. (1), we interpret the points (x,y) in the paved space as the pixels in the image function, the open balls as windows centered at (x,y), and

Prepared through collaborative participation in the Advanced Sensors Consortium (ASC) sponsored by the U.S. Army Research Laboratory under the Federated Laboratory Program, Cooperative Agreement DAAL01-96-2-0001. The authors would also like to thank Dr. Richard Sims of AMCOM for providing the video sequences and for interesting discussions.

0-7803-6297-7/00/$10.00 © 2000 IEEE

the measure c as a function of gray levels.

Several Choquet capacities have been introduced in the literature [12, 13] and their properties have been investigated in the context of image processing. Choquet capacities have the general form of $c(x,y) = \mathfrak{O}_{(i,j)\in\mathcal{B}_\delta(x,y)} f(i,j)$, where $\mathfrak{O}$ is an operator acting either on the gray level (i.e. $f(i,j) = g(i,j)$) or on some function of it (e.g. $\Delta(i,j) = g(x,y) - g(i,j)$). Some useful capacities and their properties are summarized below:

- Sum: $\sum_{(i,j)\in\mathcal{B}_\delta(x,y)} g(i,j)$. Not a very informative measure.
- max: $\max_{(i,j)\in\mathcal{B}_\delta(x,y)}\{g(i,j)\}$. Sensitive to the amplitude of the singularity.
- min: $\min_{(i,j)\in\mathcal{B}_\delta(x,y)}\{g(i,j)\}$. Sensitive to the amplitude of the singularity.
- Iso: Binary image of iso-singularities. Sensitive to the spatial distribution of singularities.
- Self-similar: $\exp\left(-\frac{(\Delta(i,j))^2}{\sigma^2}\right)$. Sensitive to the spatial distribution of singularities.

The drawback of all these measures is their limitations in terms of their sensitivity to either the amplitude or the spatial distribution of the singularities. In this paper, we introduce new hybrid measures which are sensitive to both amplitude and spatial distribution of the singularities. The proposed measures are the following two self-similar capacities along each spatial axis:

$$c_y(x) = \max_y |G_{x,\sigma} g(x,y)| \quad (2)$$

$$c_x(y) = \max_x |G_{y,\sigma} g(x,y)| \quad (3)$$

where the operators $G_{x,\sigma}$ and $G_{y,\sigma}$ are the derivatives of the Gaussian applied along x and y axes, respectively. The parameter σ clearly defines the extent of the operator. Considering the decay of this operator, in practice we only need to apply it in a window size of 6σ. Using these two measures and Eq. (1) we can evaluate the following singularity coefficients along each spatial axes:

$$\alpha_y(x) = \lim_{\delta\to 0} \frac{\log \max_{y,\delta} |G_{x,\sigma} g(x,y)|}{\log \mu(\mathcal{B}_\delta(x,y))} \quad (4)$$

$$\alpha_x(y) = \lim_{\delta\to 0} \frac{\log \max_{x,\delta} |G_{y,\sigma} g(x,y)|}{\log \mu(\mathcal{B}_\delta(x,y))} \quad (5)$$

where $\max_{y,\delta}$ is the maximum taken over all (i,y) with i in some neighborhood of x, and $\max_{x,\delta}$ is the maximum taken over all (x,j) with j in some neighborhood of y. For image processing applications $\delta \to 0$ can be interpreted as the window size reducing to one-pixel wide, in which case $\max_{y,\delta}$ would be for instance the maximum value of $|G_{x,\sigma} g(x,y)|$ along each column. In this sense, $\alpha_y(x)$ and $\alpha_x(y)$ may be viewed as projections of $|G_{x,\sigma} g(x,y)|$ and $|G_{y,\sigma} g(x,y)|$ along the x and y axes. We shall refer to these one-dimensional signals as Lipschitz signatures, since for differentiable functions $\exp(\alpha_y(x))$ and $\exp(\alpha_x(y))$ can provide some measure of the Lipschitz bounds along each axes.

In the next section, we will discuss how these Lipschitz signature can be used to provide crucial information about local image characteristics, and hence use them for object detection, tracking and stabilization.

3. IMPLEMENTATION

The main assumption herein is that "active regions" (i.e. regions of interest), where stationary or moving objects are present, exhibit some higher level of singularity in the Lipschitz signatures. Therefore these signatures can provide information about the spatio-temporal characteristics of events in a sequence, e.g. FLIR data. In other words, singularities can be detected and tracked both in time and space. In particular the spatial information (e.g. the location of a particular singularity) can be used for alignment of the image frames and hence yield a stabilized sequence. Other aspects which can also be exploited are the size and the aspect ratio of the detected object regions.

Figure 1 shows an example of a FLIR image frame from an AMCOM sequence. The Lipschitz signatures clearly localize the objects and their spatial characteristics. By tracking these signatures in time one can also characterize the associated temporal characteristics.

The detection process consists of a simple thresholding. We describe the choice of the threshold for $\alpha_y(x)$. But the approach equally applies to $\alpha_x(y)$. In order to choose a threshold value, we estimate the probability of $\alpha_y(x)$ exceeding some value $T \in [0, 255]$, which is given by

$$p(\alpha_y(x), T|g(x,y)) = \sum_{\alpha_y(x)>T} h(\alpha_y(x)|g(x,y)) \quad (6)$$

where $h(\alpha_y(x)|g(x,y))$ is the conditional pdf of $\alpha_y(x)$ over the rescaled interval $[0, 255]$. The threshold corresponds to the point where the gradient of $p(\alpha_y(x), T|g(x,y))$ attains its minimum value with a probability $p(\alpha_y(x), T|g(x,y))$ exceeding 0.5. The condition that $p(\alpha_y(x), T|g(x,y))$ should exceed 0.5 corresponds to a likelihood ratio test for binary hypotheses when no a priori information is available.

This minimum gradient point can be readily computed for each frame from which a threshold τ_y for $\alpha_y(x)$ can be found according to

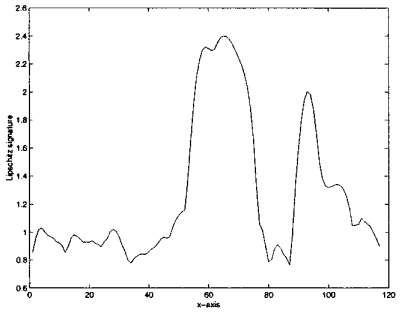
$$\tau_y = \frac{T_\tau}{255}(\alpha_{y,max} - \alpha_{y,min}) + \alpha_{y,min} \quad (7)$$

where T_τ is the point in the graph of $p(\alpha_y(x), T|g(x,y))$ where the minimum gradient is detected, and $[\alpha_{y,min}, \alpha_{y,max}]$ defines the range of values of $\alpha_y(x)$.

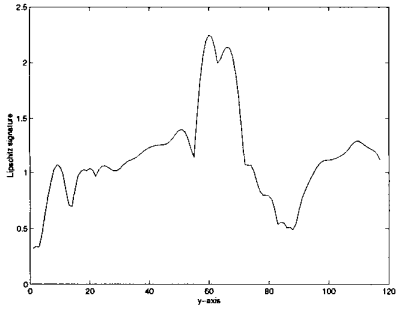
Figure 2 shows $p(\alpha_y(x), T|g(x,y))$ and the minimum gradient point for $p(\alpha_y(x), T|g(x,y)) > 0.5$, corresponding to $\alpha_y(x)$ shown in Figure 1. The detection of the objects is simply achieved by projecting the Lipschitz signatures back onto the image plane. The back-projection process involves a simple book-keeping problem. Figures 3 and 4 show two examples from AMCOM FLIR sequences where the moving or stationary object contours have been detected using Lipschitz signatures.

Once the objects have been detected, one can stabilize the sequence with respect to any object by compensating for lateral translations or tilts, which can be computed as the amount by which the centroid of the object is moving from one frame to another.

(a)



(b)



(c)

Figure 1: (a) A FLIR image, (b) Lipschitz signature along x-axis, (c) Lipschitz signature along y-axis.

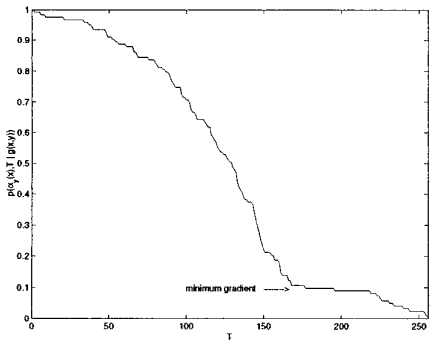


Figure 2: Probability of $\alpha_y(x) > T$, $T \in [0, 255]$.

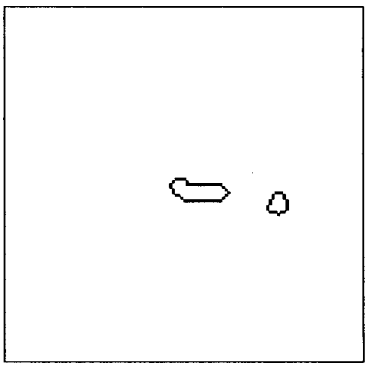
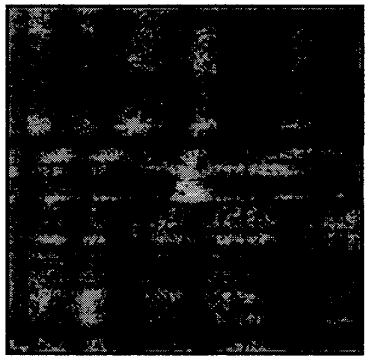


Figure 3: Detected stationary/moving objects.



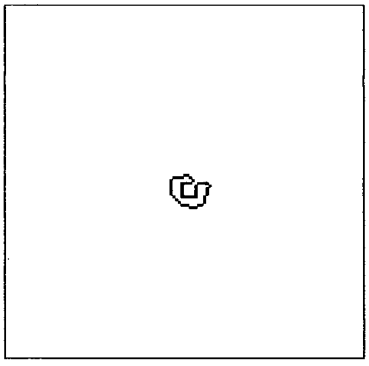


Figure 4: A detected stationary object.

4. EVALUATION AGAINST OTHER METHODS

We also implemented three other classical methods and evaluated our method against them. These methods are multi-scale optical flow, phase correlation and Kalman filtering. Each of these methods have their pros and cons summarized as below:

- Multi-scale optical flow: This method works well when inter-frame motions are relatively small compared to the image size. However, when the object is considerably unstable (i.e. its motions are too large compared to the size of the image), results are not satisfactory. Our explanation is as follows. A well known issue in multi-scale optical flow method is the tradeoff between reducing the scale for aperture adjustment and retaining finer scale features for accurate registration. When the object is very unstable, achieving an optimal tradeoff is very difficult and sometimes impossible. Also note that this method is only used for stabilization and hence additional processing is needed for detecting activities.

- Phase correlation: The application of this method to FLIR data is based on the assumption that the scale changes between two successive frames are negligible. The method fails when the sensor gets fairly close to the scene, resulting in violation of this assumption. Again the method is only a stabilizing algorithm and provides no information on the activities in a sequence.

- Kalman filter: The method requires a detector (e.g. CFAR detector) for initialization and tracking. For tracking, we used the coordinates of the centroid of the object of interest as the state variable of the Kalman filter, and used linear prediction for the innovation step. The object is assumed to be initially centered in the image or pointed at approximately by an operator. Once initialized the object is tracked automatically and the changes in the state variable are used for motion compensation and stabilization. Results are good when a single object is tracked. However, for multiple objects the algorithm can fail to track moving objects when they cross each other (occlusion). Also initialization can be problematic when the object can not be detected.

5. CONCLUSION

We have proposed a method for stabilization and activity detection in FLIR sequences. The method is robust to image scale variations and can handle multiple moving or stationary objects. The proposed method involves projections along the spatial axes, and hence can be implemented for real-time processing. An important feature of the method is its capability of providing spatio-temporal information on the objects in the scene.

References

[1] L. Baoxin, Q. Zheng, S.Z. Der, R. Chellappa, N.M. Nasrabadi, L.A. Chan, and W. LinCheng. Experimental evaluation of neural, statistical and model- based approaches to FLIR ATR. In *Proc. SPIE*, volume 3371, pages 388–397, 1998.

[2] V.K. Gonzalez and P.K. Williams. Summary of progress in FLIR/LADAR fusion for target identification at Rockwell. In *Proc. IUW*, volume 1, pages 495–499, 1994.

[3] D. McReynolds, S. Yunlong, L. Gagnon, and L. Sevigny. Stabilization of infrared image sequence with rotation, scaling and view angle changes. In *Proc. SPIE*, volume 3491, pages 992–997, 1998.

[4] R. Murenzi, D. Johnson, L. Kaplan, and K. Namuduri. Detection of targets in low resolution FLIR images using two- dimensional directional wavelets. In *Proc. SPIE*, volume 3371, pages 510–18, 1998.

[5] H.S. Parry, A.D. Marshall, and K.C. Markham. Tracking targets in FLIR images by region template correlation. In *Proc. SPIE*, volume 3086, pages 221–232, 1997.

[6] R.L. Sinclair and S. Tritchew. Stabilized FLIR for long range airborne surveillance. In *Proc. SPIE*, volume 2744, pages 366–372, 1996.

[7] J.R. Beveridge, A. Hanson, and D. Panda. Model-based fusion of FLIR, color and LADAR. In *Proc. SPIE*, volume 2589, pages 2–11, 1995.

[8] J.G. Romanski and K.R. Wegner. Method for registering low resolution ir sensor images with second generation FLIR sensors. In *Proc. SPIE*, volume 2075, pages 47–63, 1993.

[9] B.W. Smith. Multi-frame enchancement of FLIR and infrared seeker images. In *Proc. SPIE*, volume 3377, pages 231–239, 1998.

[10] B.H. Chen and S.C.A. Thomopoulos. Multiresolution fusion of FLIR and ladar data for target detection with JDEF. In *Proc. SPIE*, volume 2059, pages 305–314, 1993.

[11] J. Lévy Véhel and R. Vojak. Multifractal analysis of choquet capacities. *Advances in applied mathematics*, 20(1):pp. 1, 1998.

[12] J.-P. Berroir and J. Lévy Véhel. Multifractal tools for image processing. In *Proc. Scandinavian Conference on Image Analysis*, volume 1, pages 209–216, 1993.

[13] J. Lévy Véhel and P. Mignot. Multifractal segmentation of images. *Fractals*, 2(3):371–377, 1994.