Target Tracking in FLIR Images Using Mean Shift

Alper YILMAZ
Computer Vision Lab.
University of Central Florida

Characteristics of Targets

- Low contrast with the background.
- Similar gray level distribution to the overall frame gray level distribution.
- Slightly brighter than the background (not always).
- Targets are most of the time 5 to 10 pixels.
- Fast global and ego motion.
- No specific shape information

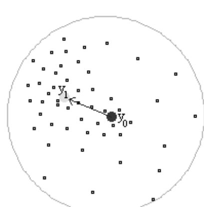
Mean Shift Vector

Given:
Data points and approximate location of the mean of this data.

Task:
Estimate the exact location of the mean of the data by determining the shift vector from initial mean.

UCF Computer Vision Lab. 3

Mean Shift Vector Example



$$M_h(\mathbf{y}) = \left[\frac{1}{n_x} \sum_{i=1}^{n_x} (\mathbf{x}_i - \mathbf{y}_0) \right] - \mathbf{y}_0$$

Mean shift vector always points towards the direction of the maximum increase in the density.

UCF Computer Vision Lab. 4

Modified Mean Shift

$$M_h(\mathbf{y}_0) = \left[\frac{\sum_{i=1}^{n_x} w_i(\mathbf{y}_0)\mathbf{x}_i}{\sum_{i=1}^{n_x} w_i(\mathbf{y}_0)} \right] - \mathbf{y}_0$$

n_x : number of points in the kernel
$\mathbf{y}_0$: initial mean location
$\mathbf{x}_i$: data points
h : kernel radius

Weights are determined using kernels (masks):
Uniform, Gaussian or Epanechnikov

UCF Computer Vision Lab. 5

Properties of Mean Shift

- Mean shift vector has the direction of the gradient of the density estimate.

- It is computed iteratively for obtaining the maximum density in the local neighborhood.

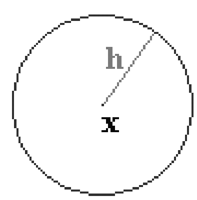
UCF Computer Vision Lab. 6

Outline

1. Introduction
    1. Kernel density estimate
    2. Possible kernels
    3. Epanechnikov profile
    4. Estimate of density gradient
    5. Mean shift & Epanechnikov Kernel
2. Feature space
    1. Target gray level distribution
    2. Distribution and tracking
        1. Similarity of target & candidate distributions
        2. Distance minimization
        3. Bhattacharya maximization using mean shift
        4. Algorithm
    3. Target standard deviation
3. Target localization using 2 features
4. FLIR results
5. Future work

Kernel Density Estimate

$$\hat{f}(\mathbf{x}) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{\mathbf{x} - \mathbf{x}_i}{h}\right)$$

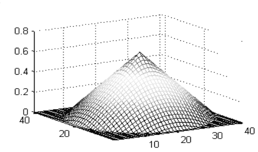


n : number of points in the kernel
h : window radius
x : mean vector
d : number of dimensions
K : Kernel density function

Possible Kernels

- Uniform kernel
- Epanechnikov kernel (convex, monotonic decreasing)

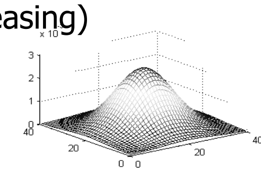
$$K_E(\mathbf{x}) = \begin{cases} \frac{1}{2} c_d^{-1} (d+2)(1 - \|\mathbf{x}\|^2) & if\ \|\mathbf{x}\| < 1 \\ 0 & otherwise \end{cases}$$



c_d : volume of unit d-dim sphere
d : number of dimensions

- Normal kernel (convex, monotonic decreasing)

$$K_N = (2\pi)^{-d/2} e^{-\|\mathbf{x}\|^2/2}$$



d : number of dimensions

UCF Computer Vision Lab.

Kernel & Profile

- **Kernel function :** defined in terms of vector
- **Profile function :** defined in terms of variable

$$k_E\left(\|\mathbf{x}\|^2\right) = K_E(\mathbf{x})$$

UCF Computer Vision Lab. 10

Epanechnikov Profile (2D)

Epanechnikov profile yields minimum mean integrated square error

$$k_E(x) = \frac{1}{2} c_d^{-1} (d+2)(1-x)$$

$$k_{E,d=2}(x) = \frac{2}{\pi}(1-x) \qquad \frac{\partial k_{E,d=2}}{\partial x} = \frac{-2}{\pi}$$

Estimate of Density Gradient

density estimate:

$$\hat{f}(\mathbf{x}) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{\mathbf{x}-\mathbf{x}_i}{h}\right)$$

gradient of density estimate:

$$\hat{\nabla} f(\mathbf{x}) \equiv \nabla \hat{f}(\mathbf{x}) = \frac{1}{nh^d} \sum_{i=1}^{n} \nabla K\left(\frac{\mathbf{x}-\mathbf{x}_i}{h}\right)$$

Mean Shift Vector in Terms of Epanechnikov Kernel

$$\hat{\nabla} f(\mathbf{x}) \equiv \nabla \hat{f}(\mathbf{x}) = \frac{1}{nh^d} \sum_{i=1}^{n} \nabla K\left(\frac{\mathbf{x}-\mathbf{x}_i}{h}\right)$$

Using Epanechnikov kernel: $K_E(\mathbf{x}) = \frac{1}{2} c_d^{-1} (d+2)(1-\|\mathbf{x}\|^2)$

$$\hat{\nabla} f(\mathbf{x}) = \frac{d+2}{nh^{d+2} c_d} n \left(\frac{1}{n} \sum_{x_i \in S_h(x)}^{n} [\mathbf{x}_i - \mathbf{x}]\right) = \frac{d+2}{h^{d+2} c_d} M_h(x)$$

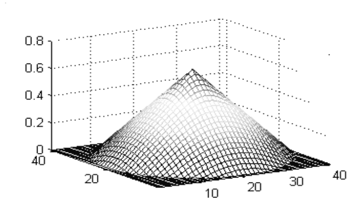
mean shift vector

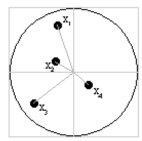
n : number of points in unit d-dimensional sphere

Target Model for Tracking

- Features used for tracking include:
  - Gray level
  - Standard deviation
- Feature probability distribution are calculated by using weighted histograms.
- The weights are derived from Epanechnikov profile.

Target Model for Tracking



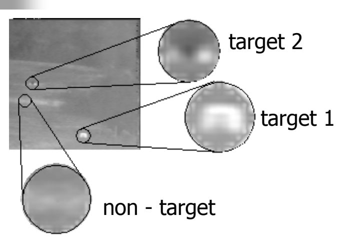


x_1, x_2, x_3, x_4 has the same feature, such as gray level.

$$p(u|K) = C \sum_{\mathbf{x}_i \in S} k\left(\|\mathbf{x}_i\|^2\right) \delta[S(\mathbf{x}_i) - u]$$

UCF Computer Vision Lab. 15

Target Gray Level Feature



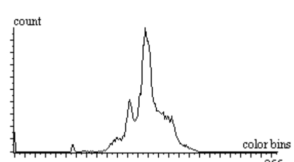
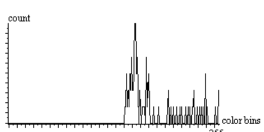
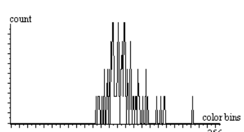


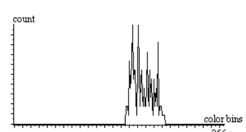
image histogram



target 1 distribution



target 2 distribution



non target distribution

UCF Computer Vision Lab. 16

Similarity of Target and Candidate Distributions

Target : q_u.
Candidate : $\hat{p}_u$.

$$d(\mathbf{y}) = \sqrt{1-\rho(\mathbf{y})}$$

$$\rho(\mathbf{y}) = \rho[\hat{p}(\mathbf{y}), q] = \sum_{u=1}^{m} \sqrt{\hat{p}_u(\mathbf{y}) q_u}$$

$\rho(\mathbf{y})$:Bhattacharya coefficient.

Distance Minimization

Minimizing the distance corresponds to *maximizing* Bhattacharya coefficient.

$$\rho[\hat{p}(\mathbf{y}), q] = \sum_{u=1}^{m} \sqrt{\hat{p}_u(\mathbf{y}) q_u}$$

Taylor expansion around $\hat{p}(\mathbf{y}_0)$

$$\rho[\hat{\mathbf{p}}(\mathbf{y}), \mathbf{q}] \cong \rho[\hat{\mathbf{p}}(\mathbf{y}_0), \mathbf{q}] + \frac{1}{2} \sum_{i=1}^{m} \hat{p}_u(\mathbf{y}) \sqrt{\frac{q_u}{\hat{p}_u(\mathbf{y}_0)}}$$

Maximizing Bhattacharya coefficient can be obtained by ***maximizing the blue term***.

Likelihood Maximization

$$\rho[\hat{\mathbf{p}}(\mathbf{y}),\mathbf{q}] \cong \rho[\hat{\mathbf{p}}(\mathbf{y}_0),\mathbf{q}] + \frac{1}{2}\sum_{i=1}^{m} \hat{p}_u(\mathbf{y})\sqrt{\frac{q_u}{\hat{p}_u(\mathbf{y}_0)}}$$

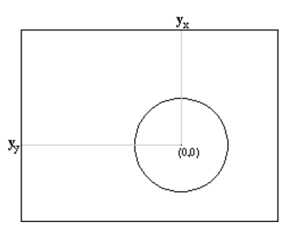
$$\frac{C_h}{2}\sum_{i=1}^{n_x}\left[\sum_{u=1}^{m}\delta[S(\mathbf{x}_i)-u]\sqrt{\frac{q_u}{\hat{p}_u(\mathbf{y}_0)}}\right] k\left(\left\|\frac{\mathbf{y}-\mathbf{x}_i}{h}\right\|\right)$$

likelihood maximization depends on maximizing w_i.

h : radius of sphere
C_h : normalization constant
$S(x_i)$: gray level at x
y : kernel center
m : number of bins

UCF Computer Vision Lab. 19

Likelihood Maximization Using Mean Shift Vector



- **y** : Spatial coordinates of target
- $\mathbf{y}_0 = [0,0]^T$.

$$M_h(\mathbf{y}_o) = \left[\frac{1}{n_x}\sum_{i=1}^{n_x}(\mathbf{x}_i - \mathbf{y}_o)\right] - \mathbf{y}_o \Rightarrow M_h(\mathbf{y}_0) = \frac{\sum_{i=1}^{n_x}\mathbf{x}_i}{n_x} = \frac{\sum_{i=1}^{n_x}\mathbf{x}_i}{\sum_{i=1}^{n_x}1}$$

new mean location

UCF Computer Vision Lab. 20

Likelihood Maximization Using Mean Shift Vector

Maximization of the likelihood of target and candidate depends on the weights:

$$w_i(\mathbf{y}_o) = \sum_{u=1}^{m} \delta[S(\mathbf{x}_i) - u] \sqrt{\frac{q_u}{\hat{p}_u(\mathbf{y}_o)}} \quad where \quad 0 \leq w_i \leq 1$$

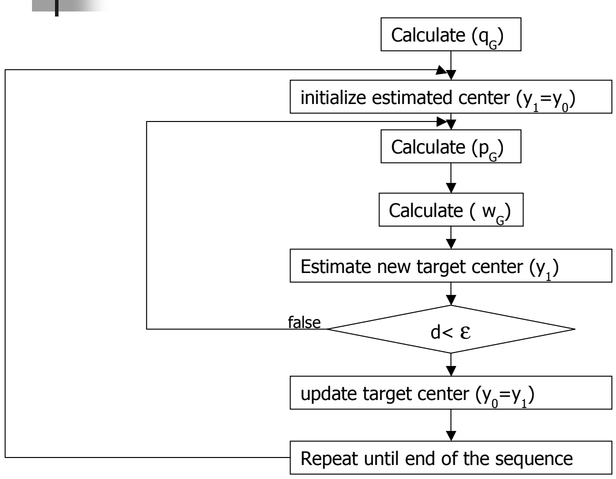
Since $\sum_{i=1}^{n_x} w_i(\mathbf{y}_0)$ is strictly positive, mean shift vector can be written as

$$M_h(\mathbf{y}_0) = \frac{\sum_{i=1}^{n_x} w_i(\mathbf{y}_0)\mathbf{x}_i}{\sum_{i=1}^{n_x} w_i(\mathbf{y}_0)} - \mathbf{y}_0$$

Thus, new target center is $\hat{\mathbf{y}} = \mathbf{y}_0 + M_h(\mathbf{y}_0)$

UCF Computer Vision Lab. 21

Algorithm



UCF Computer Vision Lab. 22

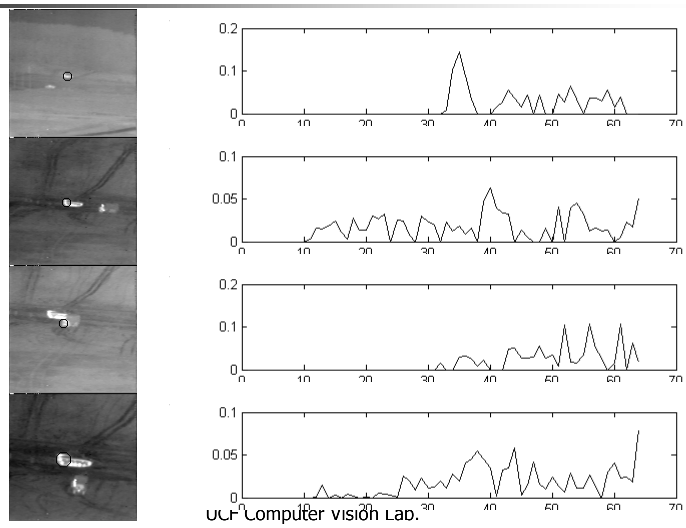
Target Gray Level Distribution (1)

target (frame 0)

frame 220

frame 246

frame 288



UCF Computer Vision Lab. 23

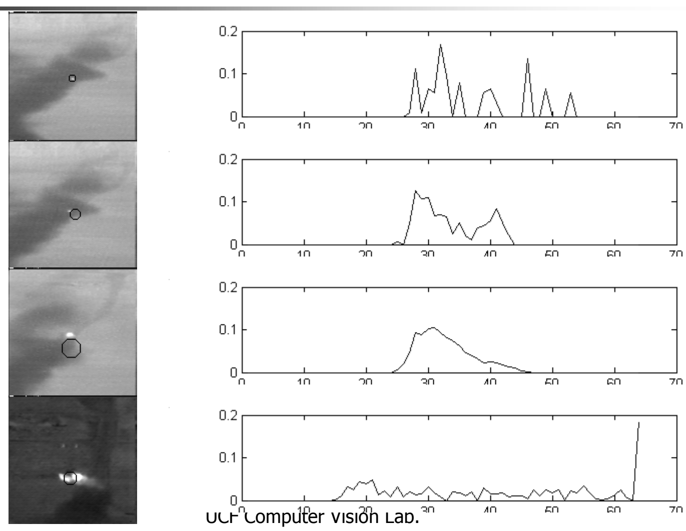
Target Gray Level Distribution (2)

target (frame 173)

frame 221

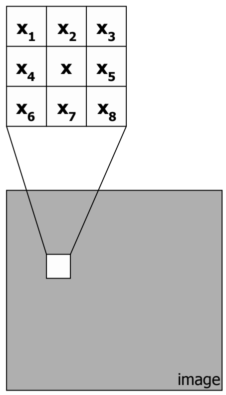
frame 382

frame 695



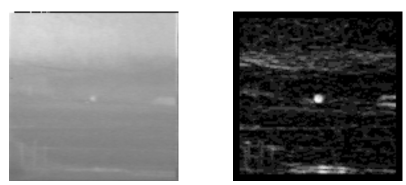
UCF Computer Vision Lab. 24

Target Std. Deviation Feature



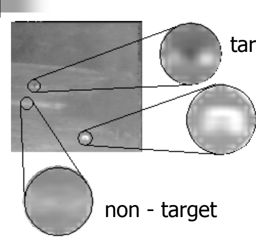
$$\sigma^2 = \frac{1}{8}\sum_{i=1}^{8}(x_i - x)^2$$

Modify **x** by $\sqrt{\sigma^2}$



UCF Computer Vision Lab. 25

Gray Level (Reminder)



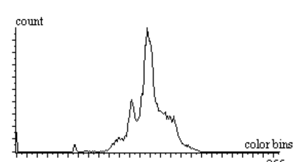
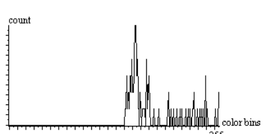
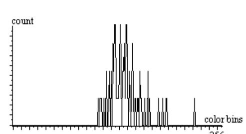


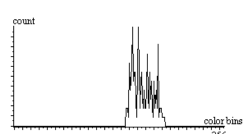
image histogram



target 1 distribution



target 2 distribution



non target distribution

UCF Computer Vision Lab. 26

Target Std. Deviation Feature

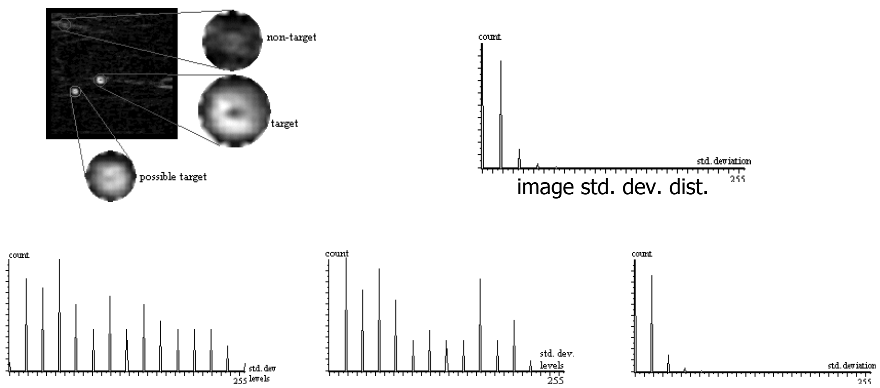


image std. dev. dist.

target std. dev. dist. | possible target std. dev. dist. | non - target std. dev. dist.

UCF Computer Vision Lab. 27

Combining Distributions of Gray Level and Std. Deviation

q_G : target gray level
q_S : target standard deviation
p_G : candidate gray level
p_S : candidate standard deviation

$$w_{G_i}(\mathbf{y}_o) = \sum_{u=1}^{m} \delta[S(\mathbf{x}_i) - u] \sqrt{\frac{q_{G_u}}{\hat{p}_{G_u}(\mathbf{y}_o)}}$$

$$w_{S_i}(\mathbf{y}_o) = \sum_{u=1}^{m} \delta[S(\mathbf{x}_i) - u] \sqrt{\frac{q_{S_u}}{\hat{p}_{S_u}(\mathbf{y}_o)}}$$

UCF Computer Vision Lab. 28

Combining Distributions of Gray Level and Std. Deviation

New target center is determined using

$$\hat{\mathbf{y}}_{k+1} = \mathbf{y}_k + \frac{\sum_{i=1}^{n_x} w_{S_i}(\mathbf{y}_k) w_{G_i}(\mathbf{y}_k)\mathbf{x}_i}{\sum_{i=1}^{n_x} w_{S_i}(\mathbf{y}_k) w_{G_i}(\mathbf{y}_k)}$$

Iteratively calculate the new target center until the distance is minimized.

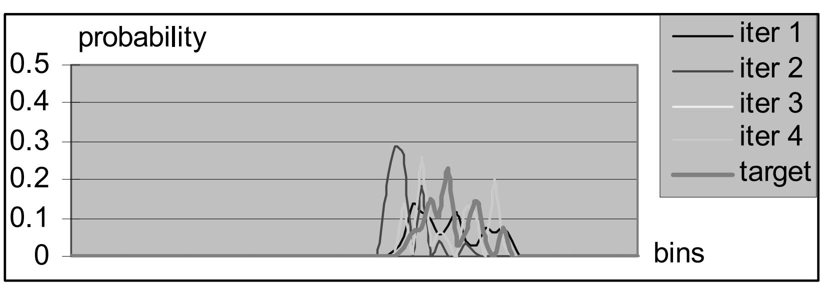
$d(\mathbf{y}) = \sqrt{1-\rho(\mathbf{y})}$ where

$$\rho(\mathbf{y}) = \tau\sqrt{\mathbf{p}_S(\mathbf{y})^T \mathbf{q}_S} + (1-\tau)\sqrt{\mathbf{p}_G(\mathbf{y})^T \mathbf{q}_G}$$

UCF Computer Vision Lab. 29

Gray Level Distribution at Each Iteration

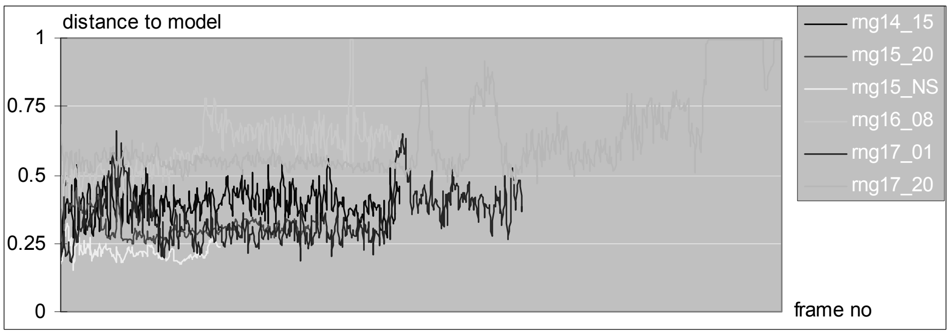
Sequence rng14_15, frames 87 and 88



UCF Computer Vision Lab. 30

Distance Between Consecutive Frames

For Different Sequences



UCF Computer Vision Lab. 31

Experiments

Features used:

- Intensity
- Standard deviation
- Gradient magnitude
- Intensity & standard deviation
- Intensity & gradient magnitude

Mutual probabilities are combined using

- Geometric mean
- Weighting (described in slide 20)

We filtered the frames using 2D-Regularization filter.

UCF Computer Vision Lab. 32

Test Set

Our test set is composed of 21 FLIR sequences:
`rng14_15, rng15_20, rng15_NS, rng16_04, rng16_07, rng16_08, rng16_18, rng17_01, rng17_02, rng17_20, rng18_03, rng18_05, rng18_07, rng18_12, rng18_13, rng18_16, rng18_18, rng19_01, rng19_06, rng19_07, rng19_11`

We manually initialize one target in the first frame and track the target in the sequence.

We have visually confirmed the results (not from the ground truth)

Comparison of Results

			Variance and Color together					
			Arithmetic mean of probabilities					Geom. mean
	Variance	Color	sqrt(var/max)	(>255)=255	Var/max	$(var/max)^2$	sqrt(var)	
14_15	Fine	Ok	Fine	Fine	Fine	Fine	Fine	Fine
15_20	Fine -	Bad	Fine	Bad	Fine	Fine	Fine -	Fine
15_NS	Fine	Bad	Ok	Ok	Ok	Ok	Fine	Fine
16_04	Bad	Fine	Ok	Ok	Ok	Ok	Fine	Fine
16_07	Bad	Bad	Fine	Fine	Fine	Bad	Ok (last part)	Ok (last part)
16_08	Ok	Fine	Fine	Fine	Fine	Ok	Fine	Fine
16_18	Fine	Bad	Ok	Bad	Fine	Ok	Fine	Fine
17_01	Bad	Fine	Fine	Fine	Bad	Bad	Fine	Fine
17_02	Bad	Bad	Bad	Bad	Bad	Bad	Bad	Bad
17_20	Bad	Bad	Ok	Fine	Fine	Bad	Fine	Fine
18_03	Ok	Bad	Bad	Fine	Ok	Bad	Ok	Ok
18_05	Ok	Bad	Bad	Fine	Fine	Bad	Fine	Fine
18_07	Fine	Ok	Ok-fine	Fine	Ok-fine	Ok-fine	Ok-fine	Fine
18_12	Bad	Bad	Bad-ok	Bad-ok	Bad-ok	Bad	Bad-ok	Bad-ok
18-13	Fine	Bad	Fine	Fine	Fine	Fine	Fine	Fine
18-16								Ok
18-18								Ok
19-01								Fine
19-06								Fine
19-07								Fine
19-11								Fine

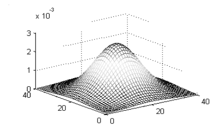
Fine:
tracks very well

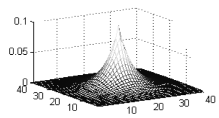
Okay:
tracks well except some sections

Bad:
does not track

Comparison of Results

Using "Geometric Mean" and
- Regularization pre-filtering
- Gaussian pre-filtering
- Median filtering
- Without filtering

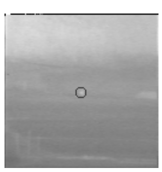

Gaussian filter

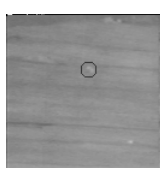

Regularization filter

	14_15	15_20	15_NS	16_04	16_07	16_08	16_18	17_01	17_02	17_20	18_03	18_05	18_07	18_12	18-13	18-16	18-18
With Regularization filtering	Fine	Fine	Fine	Fine	Ok	Fine	Fine	Fine	Bad	Fine	Ok	Fine -	Fine	Bad-ok	Fine	Ok	Ok -
With Gaussian filtering	Fine -	Ok	Ok	Ok	Bad	Ok	Ok	Ok	Bad	Ok	Ok	Ok	Ok	Bad-ok	Ok	Ok	Ok -
With median filter-art mean	Fine	Fine	Fine	Fine	Fine -	Fine	Fine	Fine	Bad	Ok	Ok	Fine -	Fine	Bad	Fine	Fine -	Ok -
Without any filtering	Fine	Fine	Fine	Fine	Bad	Fine	Fine	Fine	Bad	Ok	Ok	Fine -	Fine	Bad-ok	Fine	Ok	Ok -

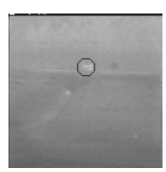
UCF Computer Vision Lab. 35

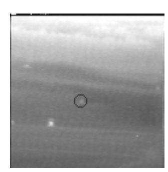
FLIR Tracking Results

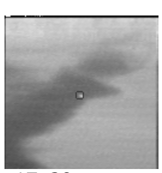

14_15


15_20


15_NS


17_01


17_02


17_20


18_03


18_05


19_07


19_11

UCF Computer Vision Lab. 36

Future Work

- Resolve the drawback of the system for large global motion
- Enhance the sequence to have more distinctive features of the target
- Obtain target model using ellipsoidal region instead of circular region.
- Update initial model periodically.

UCF Computer Vision Lab. 37