

Object Tracking using SIFT Features and Mean Shift

Computer Vision and Image Understanding
vol. 113, No. 3, 2009

Huiyu Zhou, Yuan Yuan, and Chunmei Shi

Presented by Bong-Seok Choi

School of Electrical Engineering and Computer Science
Kyungpook National Univ.



COLOR & IMAGING LAB.
KYUNGPOOK NATIONAL UNIVERSITY

Abstract

□ Proposed method

- Object tracking using mean shift/SIFT strategy
 - Using SIFT feature
 - Corresponding region of interests across frame
 - Using mean shift
 - Conducting similarity search by color histogram
 - Using expectation-maximization algorithm
 - Optimizing probability function for better similarity search

Introduction

- ❑ Goal of object tracking
 - Determining position of object in image
- ❑ Previous algorithms for object tracking
 - Considering gaussian and linear problem
 - Using kalman filter based method
 - Considering non-gaussian and non-linear syetem
 - Using particle filter based method
 - Mean shift algorithm
 - Efficient algorithm to Handle occlusion and significant clutters
 - Drawback of mean shift algorithm
 - Less efficiency in presence of dramatic intensity or color change
 - » Cannot effective work in variable scaled, rotated, and translated image

-
- Scale invariant feature transform(SIFT)
 - Generating feature point
 - Invariant to any scaling, rotation or translation of images

□ Proposed method

- Integration of mean shift and SIFT feature tracking
- Using expectation-maximization algorithm
 - Estimating maximum likelihood
 - using measurements from Mean shift and SIFT correspondence

Literature review

- ❑ Previous method for object tracking
 - Feature based approaches
 - Multiple hypothesis tracking(MHT) algorithm
 - Considering multiple tracking candidate
 - Finding best fit to real image descriptors
 - Computationally expensive both in time and memory
 - » Reluctantly support real application
 - Hidden markov models(HMM) algorithm
 - Use to transformation between two images or moving 3D structure

-
- Particle filter to Kalman filter
 - Robust performance in case of non-gaussian and non-linear system
 - Solving computational problem by large particle numbers
 - Mean shift tracking algorithm
 - Measuring similarity between template region and current target region
 - » Using bhattacharyya coefficient
 - » Finding local minimum of distance measure function
 - Model based approaches
 - Requirement of grouping, reasoning, and rendering
 - Requiring Prior knowledge about investigated model

-
- Optical flow based approach
 - Optical flow
 - Vector field of images changes with time
 - Normally use for generating dense flow field
 - Computing flow vector of each pixel under brightness constancy constraints
 - Example of optical flow based approach
 - Shi-Tomasi-Kanade(STK) tracking
 - » Computing iteratively translation of region centered on interest point
 - » Requiring feature work
 - » Reducing incorrect point correspondence

□ Similarity search

– Similarity measure by mean shift

- Searching similarity across two neighborhood image frames
- Measuring similarity based on color information

– Sample point in current frame

$$\gg \mathbf{I}_x = (\mathbf{x}_i, \mathbf{u}_i)_{i=1}^N$$

– Sample point in target image

$$\gg \mathbf{I}_y = (\mathbf{y}_j, \mathbf{v}_j)_{j=1}^M$$

– Estimating PDF of object in current image using kernel density estimation

$$\tilde{p}_x(\mathbf{x}, \mathbf{u}) = \frac{1}{N} \sum_{i=1}^N W \left(\left| \frac{\mathbf{x} - \mathbf{x}_i}{\sigma} \right|^2 \right) k \left(\left| \frac{\mathbf{u} - \mathbf{u}_i}{h} \right| \right) \quad (1)$$

where W is weight function,

k is kernel function,

σ and h are the bandwidths in the spatial and feature spaces,

- Measuring affinity between two distributions

$$\int p_y(u) \log \frac{p_y}{p_x} du \quad (2)$$

Latter is represented as

$$\begin{cases} B(I_x, I_y) = \sqrt{1 - \rho(p_x, p_y)} \\ \rho(p_x, p_y) = \int \sqrt{\tilde{p}_x(u) \tilde{p}_y(u)} du \end{cases} \quad (3)$$

- Finding mode of $\tilde{p}_x(\mathbf{x}, \mathbf{u})$

$$\frac{\partial \tilde{p}_x(\mathbf{x}, \mathbf{u})}{\partial \mathbf{u}} = \frac{2}{N} \sum_{i=1}^N W \left(\left| \frac{\mathbf{x} - \mathbf{x}_i}{\sigma} \right|^2 \right) k' \left(\left| \frac{\mathbf{u} - \mathbf{u}_i}{h} \right|^2 \right) \sum_i^{-1} (\mathbf{u} - \mathbf{u}_i) = 0 \quad (4)$$

where $k' = dk/dt$, and \sum_i is covariance matrix

- Hessian of $\tilde{p}_x(\mathbf{x}, \mathbf{u})$

$$\nabla^2 \tilde{p}_x(\mathbf{x}, \mathbf{u}) = -2c \times \sum_{i=1}^N w \left(\left| \frac{\mathbf{x} - \mathbf{x}_i}{\sigma} \right|^2 \right) \left(k \left(\frac{\mathbf{u} - \mathbf{u}_i}{h} \right)^2 I + 2k' \left(\frac{\mathbf{u} - \mathbf{u}_i}{h} \right) \right) \quad (5)$$

where c is constant and I is identity matrix

$$\nabla^2 \tilde{p}_x(\mathbf{x}, \mathbf{u}) = -2c \sum_{i=1}^N w \left(\left| \frac{\mathbf{x} - \mathbf{x}_i}{\sigma} \right|^2 \right) k \left(\frac{\mathbf{u} - \mathbf{u}_i}{h} \right)^2 I \quad (6)$$

- Solving \mathbf{u}

$$f(\mathbf{u}) = \sum_{i=1}^N \frac{k' \left(\frac{\mathbf{u} - \mathbf{u}_i}{h} \right)^2}{\sum_{i'=1}^N k' \left(\frac{\mathbf{u} - \mathbf{u}_{i'}}{h} \right)^2} \mathbf{u}_i \quad (7)$$

where vector $f(\mathbf{u}) - \mathbf{u}$ is mean shift

- SIFT feature corresponding
 - Component of formulation of final k'
- SIFT theory
 - Extracting scale-invariant features by using staged filtering approach
 - Scale space of image $L(x, y, \sigma_L)$, resulting from convolution of variable-scale gaussian $G(x, y, \sigma_L)$ for image $I(x, y)$

$$L(x, y, \sigma_L) = G(x, y, \sigma_L) * I(x, y) \quad (8)$$

and

$$G(x, y, \sigma_L) = \frac{1}{2\pi\sigma_L^2} \exp^{-(x^2+y^2)/2\pi\sigma_L^2} \quad (9)$$

- Different of gaussian function

$$G(x, y, s\sigma_L) - G(x, y, \sigma_L) \approx (s-1)\sigma_L^2 \nabla^2 G \quad (10)$$

$$\frac{\partial G}{\partial \sigma_L} = \sigma_L^2 \nabla^2 G \quad (11)$$

– SIFT and mean shift-based similarity measure

- PDF of object in current image

$$\tilde{p}_x(\mathbf{x}, \mathbf{u}) = \frac{1}{N} \sum_{i=1}^N \left(W_1 \left(\left| \frac{\mathbf{x} - \mathbf{x}_i}{\sigma} \right|^2 \right) k' \left(\left| \frac{\mathbf{u} - \mathbf{u}_i}{h} \right|^2 \right) + W_2 \left(\left| \frac{\mathbf{x} - \mathbf{x}_i}{\sigma} \right|^2 \right) f_s(\mathbf{x}, \mathbf{u}) \right) \quad (13)$$

where f_s is gaussian distribution based on SIFT feature correspondence

w_1, w_2 are two weight functions

Updated on pair-wise frames

$$\sum_{i=1}^N \left(W_1 \left(\left| \frac{\mathbf{x} - \mathbf{x}_i}{\sigma} \right|^2 \right) + W_2 \left(\left| \frac{\mathbf{x} - \mathbf{x}_i}{\sigma} \right|^2 \right) \right) = 1 \quad (14)$$

and

$$f_s(\mathbf{x}, \mathbf{u}) = \frac{1}{2\pi\sigma_s^2} \exp \left(-\frac{((v_{x_i} - v_{x_0})^2 - (v_{y_i} - v_{y_0})^2)}{2\sigma_s^2} \right) \quad (15)$$

- Estimating mean shift algorithm
 - Using established expectation-maximization(EM) algorithm
 - Expectation step
 - » Evaluating posterior probabilities for each mixture component

$$\begin{cases} f(\mathbf{u}) = \sum_{r=1}^N q(r|\mathbf{u}) \mathbf{u}_r \\ q(r|\mathbf{u}) = \frac{p(r|\mathbf{u}) \sigma_r^{-2}}{\sum_{r'=1}^N p(r'|\mathbf{u}) \sigma_{r'}^{-2}} \end{cases} \quad (16)$$

where $q(r|\mathbf{u})$ is posterior probability or responsibility $p(r|\mathbf{u})$ re-weight by inverse variance and re-normalized

- » Log-likelihood of image data

$$\sum_{i=1}^N \mathcal{L} = \sum_{i=1}^N \log q(\mathbf{u}, \mathbf{z} | \eta) \quad (17)$$

» Expectation with respect to posterior distribution

$$Q\left(\mathbf{z}^\tau \mid \mathbf{z}^{\tau-1}\right) = \sum_{i=1}^N \sum_{j=1}^M q\left(\mathbf{z} \mid \mathbf{u}, \boldsymbol{\eta}^\tau\right) \log q\left(\mathbf{u} \mid \mathbf{z}, \boldsymbol{\eta}\right) + C \quad (18)$$

where C term is independent of \mathbf{z}

- Maximization step

» New estimates are deductive if a maximization is reached

$$\mathbf{z}^{\tau+1} = \arg \max Q(\mathbf{z} | \mathbf{z}^{\tau-1}) \quad (19)$$

$$\frac{\partial Q}{\partial \eta} = \sum_{i=1}^N \sum_{j=1}^M \frac{q(\mathbf{z} | \mathbf{u}, \eta^\tau)}{q(\mathbf{u} | \mathbf{z}, \eta)} \frac{\partial q(\mathbf{u} | \mathbf{z}, \eta)}{\partial \eta} = 0 \quad (20)$$

Let \mathbf{u}_z be a mean value, then

$$\frac{\partial q(\mathbf{u} | \mathbf{z}, \eta)}{\partial \eta} = q(\mathbf{u} | \mathbf{z}, \eta) \sum_z (\mathbf{u} - \mathbf{u}_z - \eta) \quad (22)$$

Finally, solution for η is

$$\eta^{\tau-1} = \left(\sum_{i=1}^N \sum_{j=1}^M q(\mathbf{z} | \mathbf{u}, \eta^\tau) \sum_z^{-1} \right)^{-1} \sum_{i=1}^N \sum_{j=1}^M q(\mathbf{z} | \mathbf{u}, \eta^\tau) \sum_z^{-1} (\mathbf{u}_z - \mathbf{u}) \quad (23)$$

– Proposed algorithm

- Procedure of proposed method
 - Defining rectangle on region of interest in first frame
 - Computing color histogram of this region
 - » Extracting SIFT features
 - Similarity measure using eq.2,3 and 13
 - » Applying SSD method
 - Launch proposed EM algorithm
 - Iterate above steps till difference between two mean shift

Experimental walk

- Evaluation of proposed method
 - Test sequences



Fig. 1. Test sequences used in current evaluation

– Configuration of each image sequence

Table 1. Details of four image sequences used in the evaluation (fps, frames per second)

Sequences	Size	Frame-number	fps	Object-number
'Single person in darkness'	720 × 576	680	25	1
'Four person'	720 × 576	1006	25	4
'Traffic condition'	720 × 576	3748	25	>8
'Fast movement'	720 × 576	448	25	1

– Comparison of previous method in sequence 1

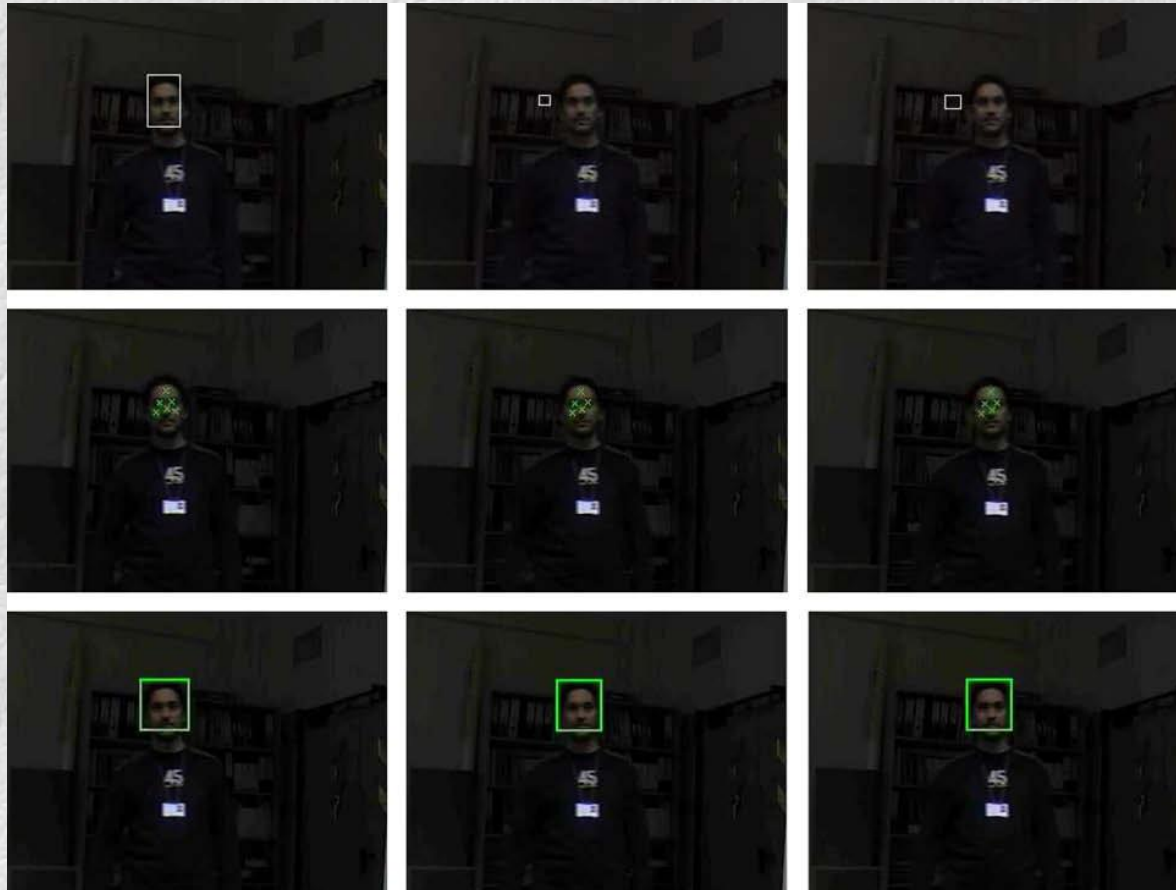


Fig. 2. Sequence 1: tracking comparison of the classical mean shift (first row), SIFT feature correspondence (2nd row, SIFT features marked as “x”) and proposed tracker (3rd row).

– Comparison of previous method in sequence 2

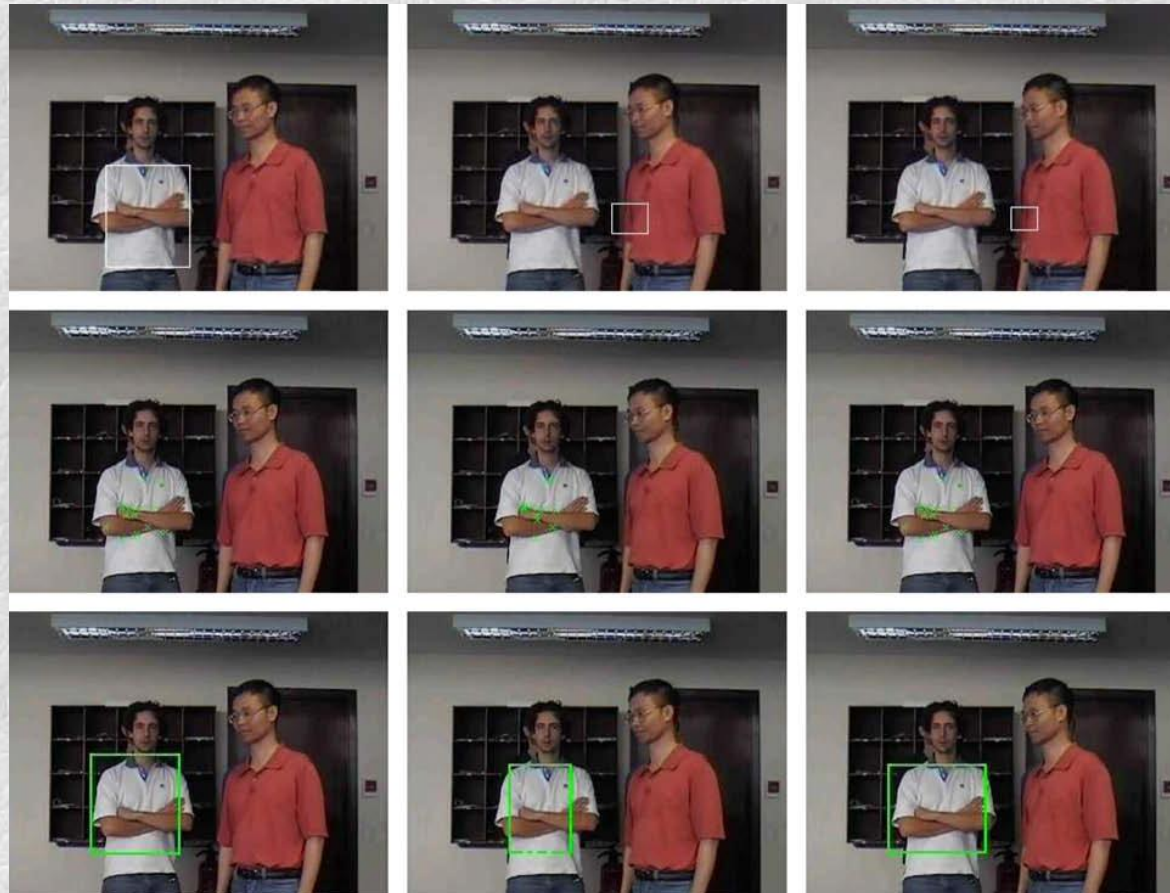


Fig. 3. Sequence 2: tracking comparison of the classical mean shift (first row), SIFT feature correspondence (2nd row, SIFT features marked as “”) and proposed tracker (3rd row).

- Comparison of previous method in sequence 2
 - Object occlusion



Fig. 4. Performance comparison of classical mean shift (first row), SIFT feature correspondence (2nd row, SIFT features marked as “x”) and proposed tracker (3rd row) in case the SIFT approach fails in object occlusions.

– Statistics of tracking errors

Table 2. Statistics of tracking errors in different scenarios by individual approaches (units:pixels)

Sequences	Mean shift	SIFT-SSD	SIFT-mean shift
'Single person in darkness'	8.2	6.9	4.6
'Four person'	6.8	6.1	5.2
'Traffic condition'	5.1	4.4	3.6
'Fast movement'	2.3	2.0	1.7

- Illustration of tracking accuracy in sequence
 - Single person in darkness

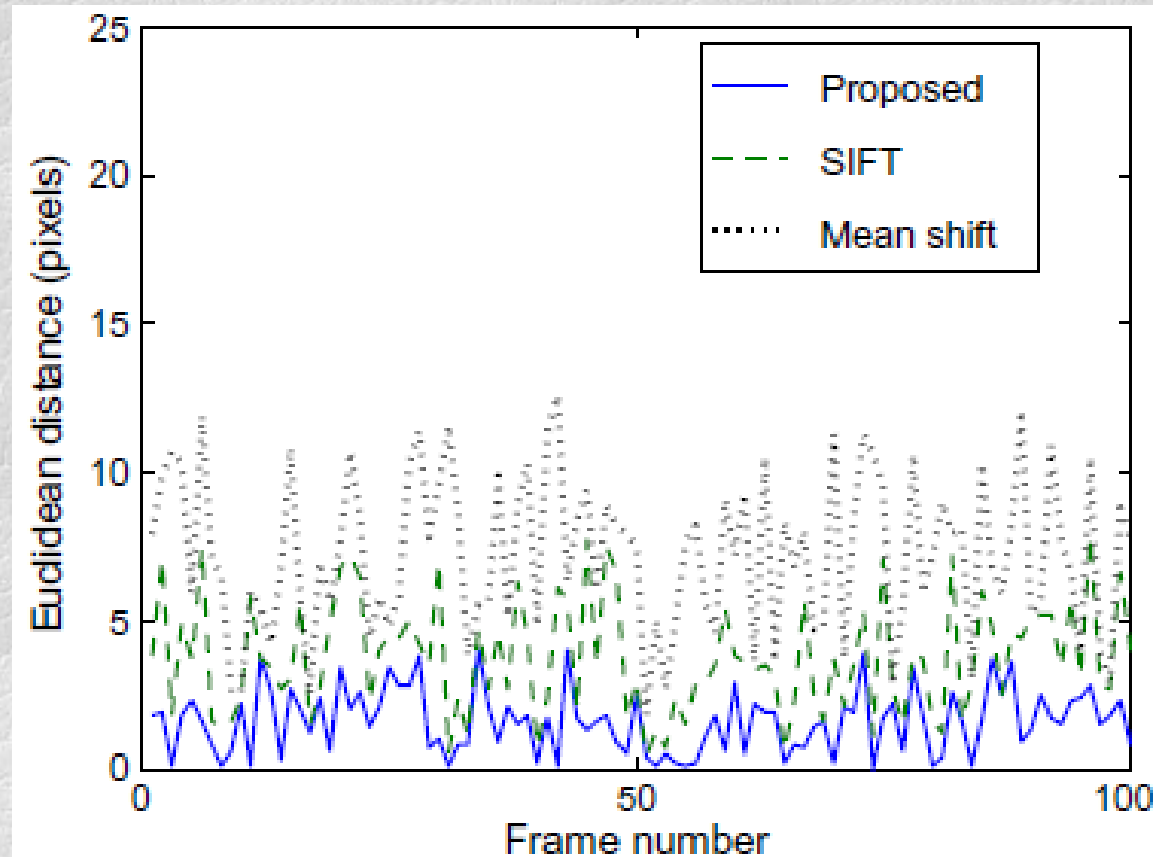


Fig. 5. Illustration of tracking accuracy in sequence “single person in darkness”: the Euclidean distance between the estimated objection position and the ground truth is plotted against frame numbers.



- Illustration of tracking accuracy in sequence
 - Traffic condition

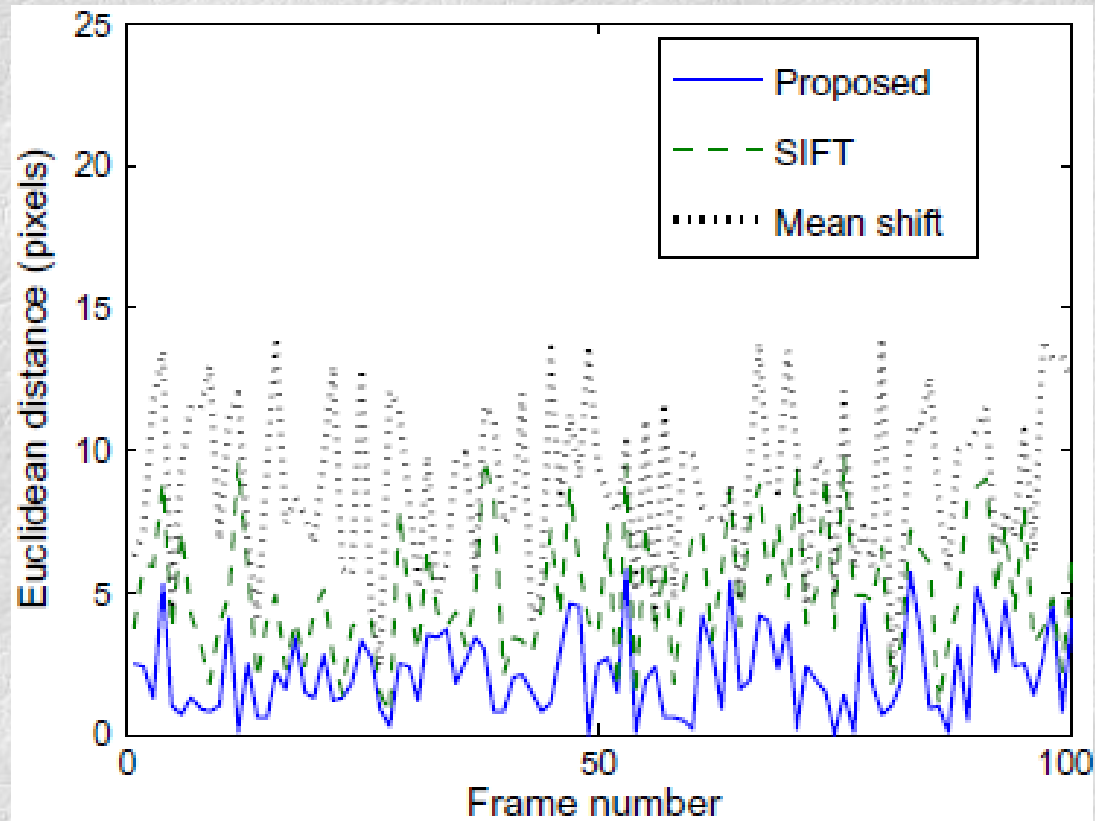


Fig. 5. Illustration of tracking accuracy in sequence “traffic condition”: the Euclidean distance between the estimated objection position and the ground truth is plotted against frame numbers.

Conclusion and future work

- Proposed method
 - Enhancing classical mean shift object tracking
 - Integrating SIFT feature correspondence and mean shift tracking
 - Using expectation-maximization algorithm
 - Optimizing probability function for better similarity search