Real-Time Moving Object Segmentation and Tracking for H.264/AVC Surveillance Videos

Pei Dong (dongpei@it.usyd.edu.au) *Supervisors*: *David D. Feng & Yong Xia*
School of Information Technologies
FACULTY OF ENGINEERING & INFORMATION TECHNOLOGIES

MOTIVATION AND BACKGROUND

- Video surveillance is ubiquitous and it produces a large amount of visual data
- H.264/AVC is currently the most advanced and the most popular video coding standard
- Automatic and efficient moving object segmentation and tracking can facilitate the content analysis and management of surveillance videos

RELATED WORK

Pixel domain methods

- Carry out a *full* decoding
- Rely on *raw pixels* for further analysis

Compressed domain methods

- Carry out a *partial* decoding
- Analyze the information produced by video encoder, e.g. motion vectors, transform coefficients, and prediction modes

Shortcomings of existing methods

- Encode with small intra frame period and thus compromise the coding efficiency
- Rarely take advantage of the prediction modes

PROPOSED METHOD

Advantages

- Requires very small amount of video decoding (i.e. entropy decoding-level)
- Adaptively fuses compressed domain features on different stages
- Imposes no restrictions on intra period

Framework

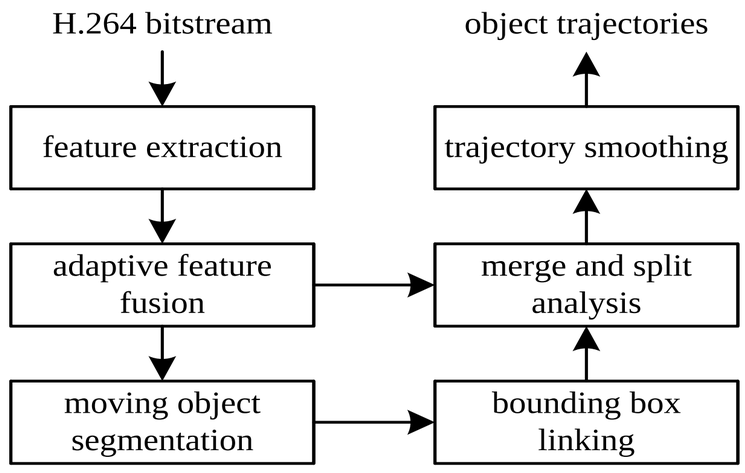

Fig. 1. The framework of the proposed algorithm

Moving object segmentation

- Step 1: Combine
  $Mask_{DCT}$ (transform coefficients),
  $Mask_{PM}$ (prediction modes), and
  $Mask_{MV}$ (motion vectors)
- Step 2: Obtain $Mask_{MO}$, i.e. the mask of moving regions (Connected regions in $Mask_{MO}$ are candidate moving objects.)

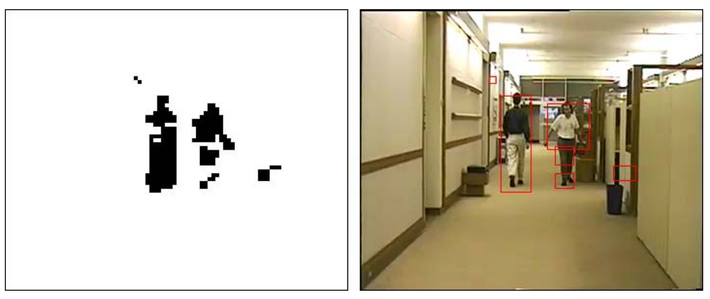

Fig. 2. Detected moving regions (left) and their bounding boxes (right) (Test video: Hall_monitor)

Tracking and trajectory construction

- Match analysis: Robustly search for bounding box links (BBL) among nearby frames
- Refinement: Split under-segmented bounding boxes, and merge over-segmented bounding boxes

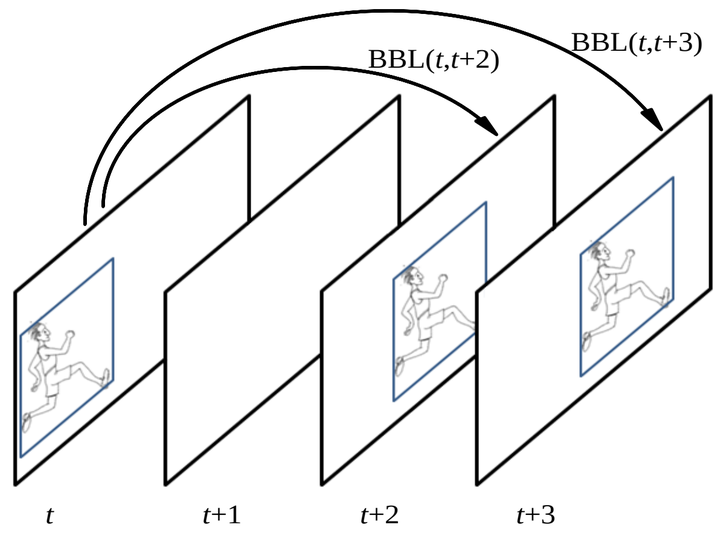

Fig. 3. Temporal linking of bounding boxes

- Trajectory smoothing: Polynomial fitting on the left, right, upper and lower bounds of the bounding boxes of the same object

EXPERIMENTS

Test videos

- Hall_monitor (352 x 288)
- Traffic (320 x 240)

Methods for comparison

- M-method (based on Ref.[1])
- The proposed method but without using prediction modes

Evaluation metrics for detection

- Precision (P)
- Recall (R)

Moving object detection results

	Hall_monitor		Traffic	
	P	R	P	R
M-method	84.56%	96.26%	54.95%	58.83%
Proposed method (without *SPM**)	**97.57%**	96.92%	85.94%	87.10%
Proposed method	96.11%	**97.80%**	**88.13%**	**90.01%**

**SPM*: small prediction modes

Speed test

- Hall_monitor: 117 fps; Traffic: 169 fps
- *Real-time processing* is achieved.

Tracking results

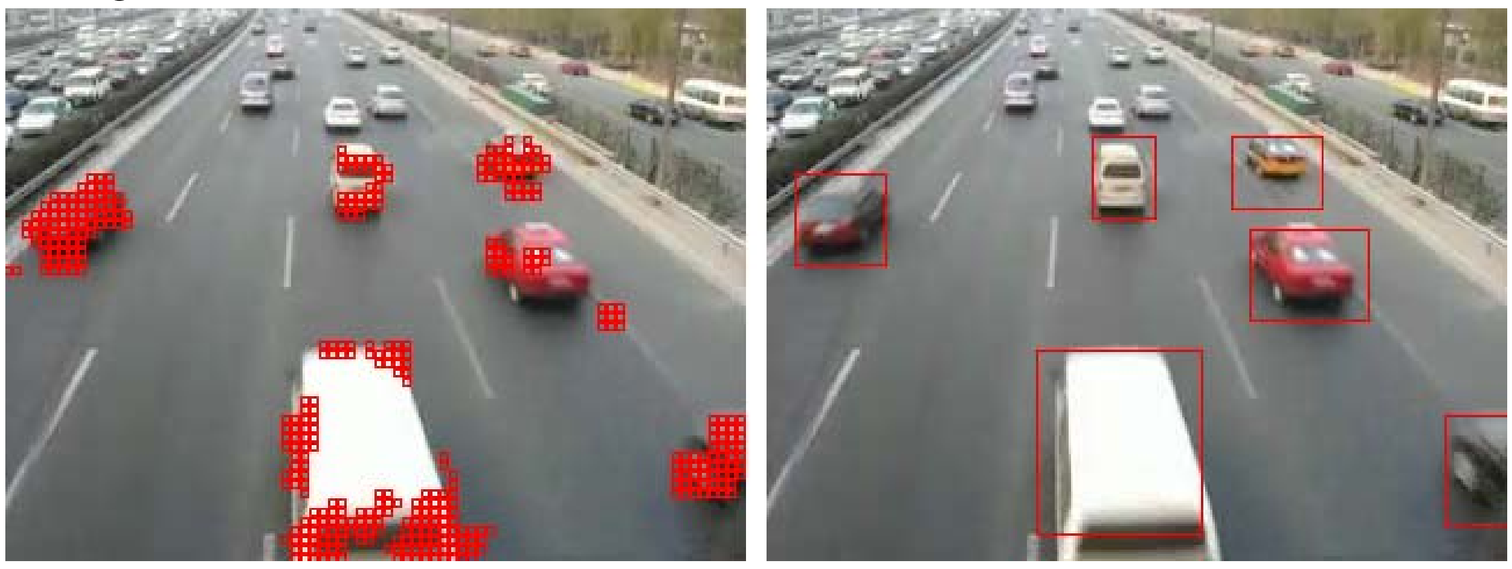

Fig. 4. Moving object mask (left) and bounding boxes after trajectory smoothing (right) (Test video: Traffic)

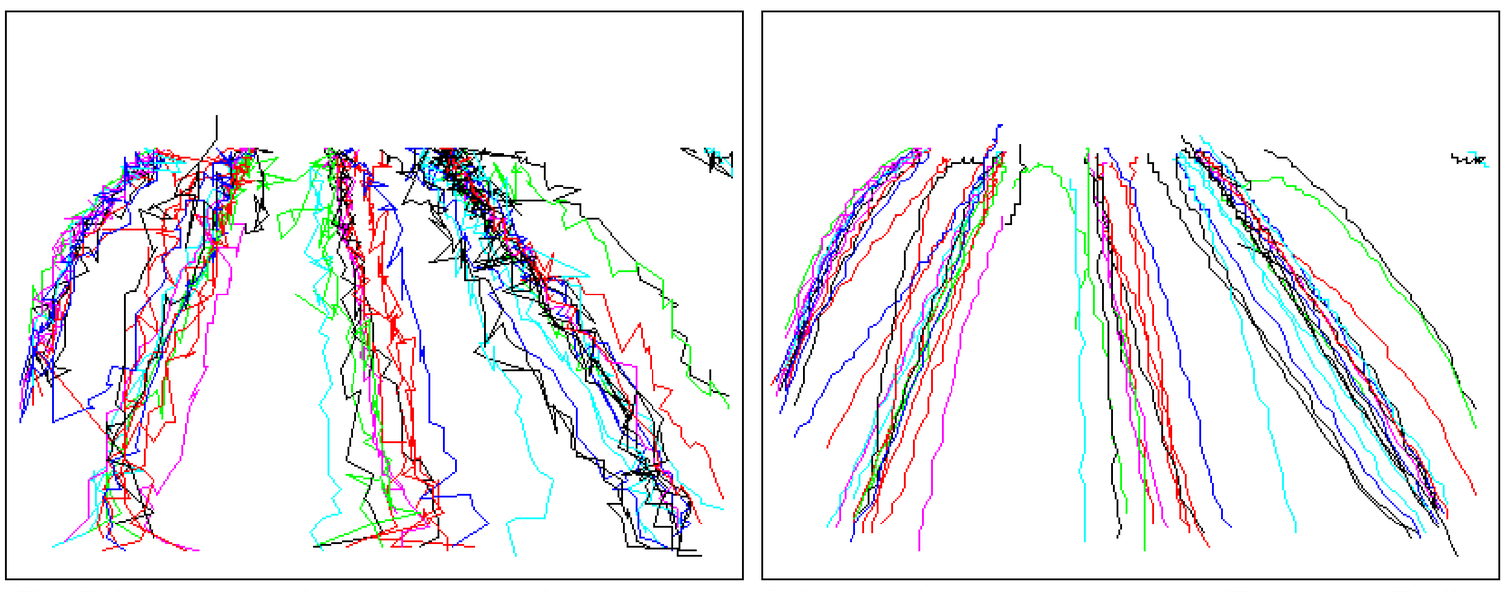

Fig. 5. Moving vehicle trajectories before smoothing (left) and after smoothing (right) (Test video: Traffic)

CONCLUSIONS AND OUTLOOK

Conclusions

An H.264/AVC compressed domain method has been proposed. It

- combines multiple types of compressed domain information and thus reduces the negative impact of their coarseness
- brings improvement to the overall performance by incorporating prediction modes
- has achieved real-time moving object segmentation and tracking

Future work

- Uses automatic thresholding based on statistics from the video
- Applies the proposed method to anomaly detection

REFERENCES

[1] Z. Liu, Y. Lu, and Z. Zhang, "Real-time spatiotemporal segmentation of video objects in the H.264 compressed domain," *Journal of Visual Communication and Image Representation*, vol. 18, no. 3, pp. 275-290, 2007.

ACKNOWLEDGEMENTS

This work was supported in part by the ARC grants and National Natural Science Foundation of China.

THIS RESEARCH IS SPONSORED BY

This work has been published in the International Conference on Image Processing, ICIP 2011, Brussels, Belgium.