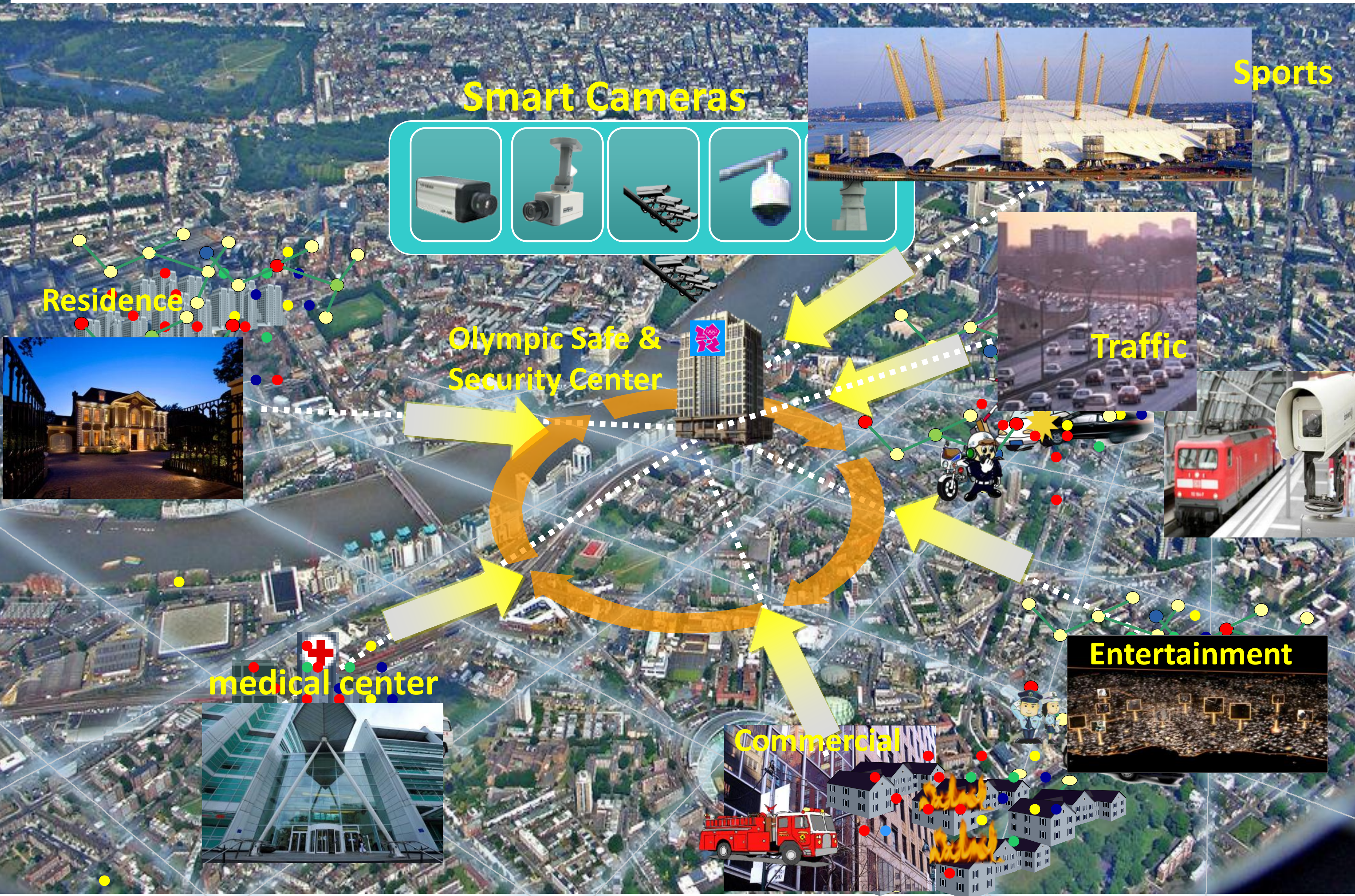
A Fast and Performance-Maintained Transcoding Method based on Background Modeling for Surveillance Video

Mingchao Geng, Xianguo Zhang, Yonghong Tian, Luhong Liang, Tiejun Huang

National Engineering Laboratory for Video Technology
Peking University, Beijing, China

From Sydney, to Beijing and then London, safety and security is always one of main concerns for each Olympics



Huge amount of surveillance video data

2012.7.10

1.5T/day for one HD Camera

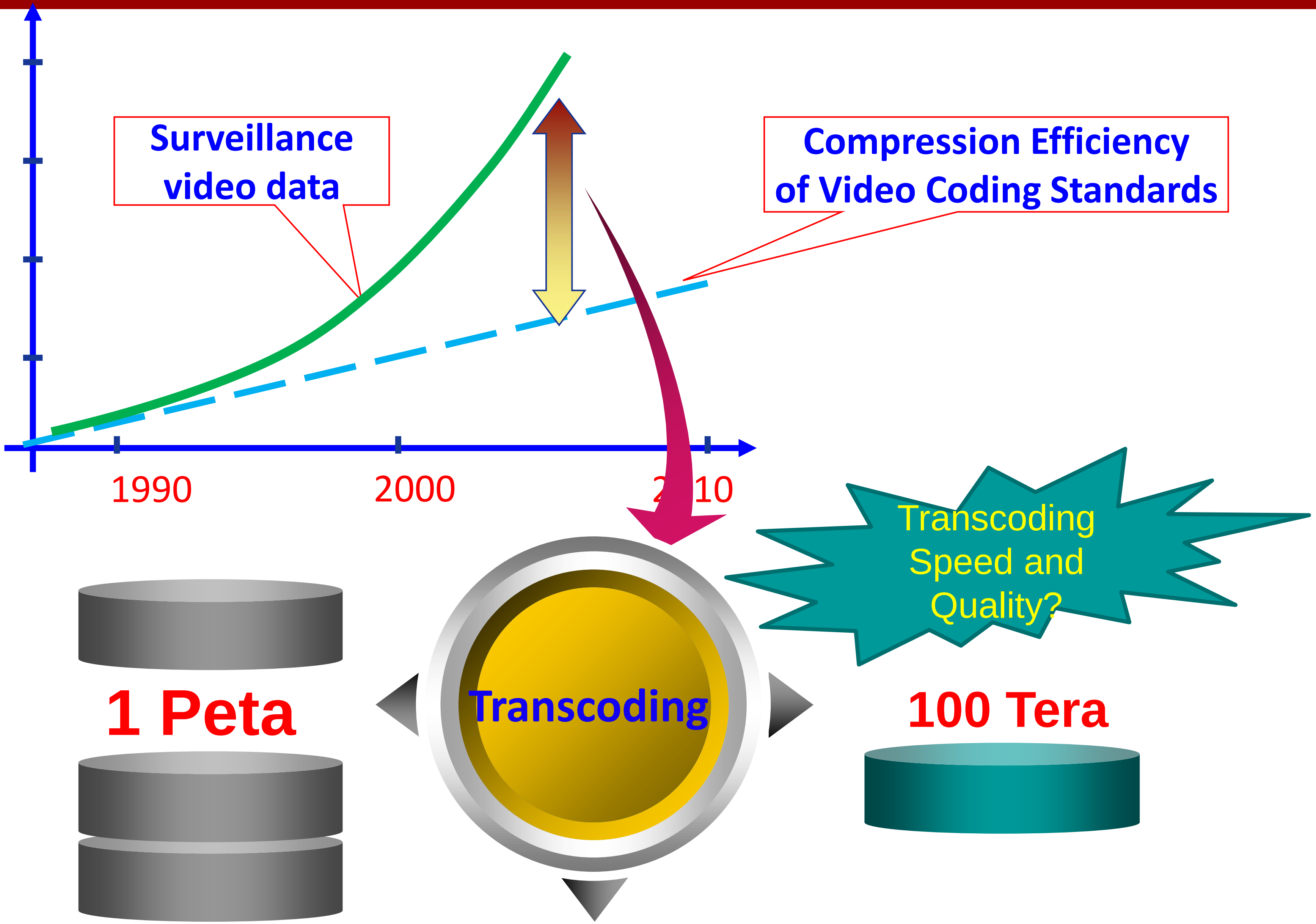
30days~6months storage time

More than 4M cameras in UK for Olympics

Huge data poses a grand challenge for transmission and long-term storage!

Surveillance video data increases extremely faster than the compression efficiency

2012.7.10



Surveillance video data

Compression Efficiency of Video Coding Standards

Transcoding Speed and Quality?

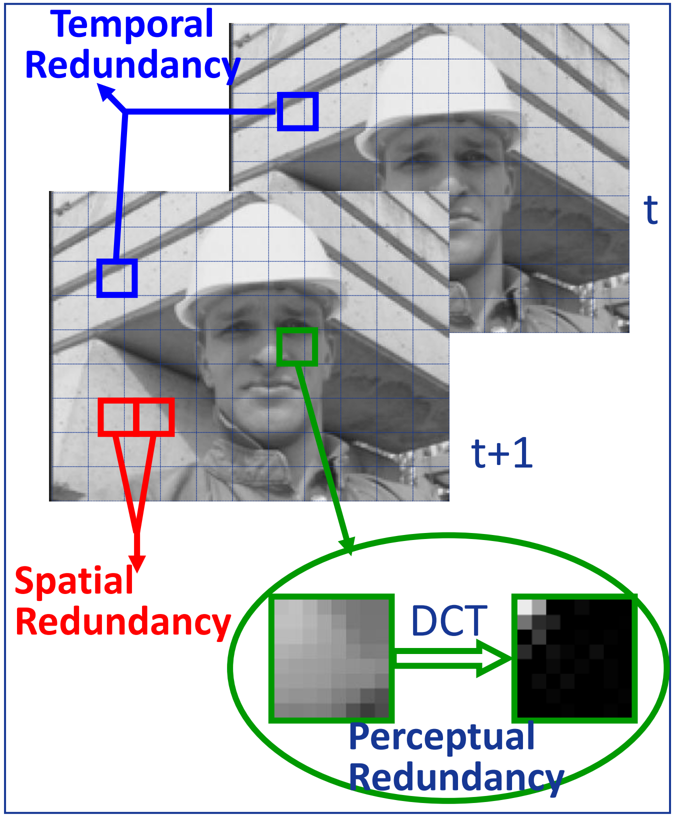
1 Peta

Transcoding

100 Tera

How to guarantee the transcoding quality and speed?

- ✓ Beyond the traditional redundancy…For surveillance video, we can remove …?
- ✓ Background is a special kind of visual redundancy for surveillance video



Temporal Redundancy

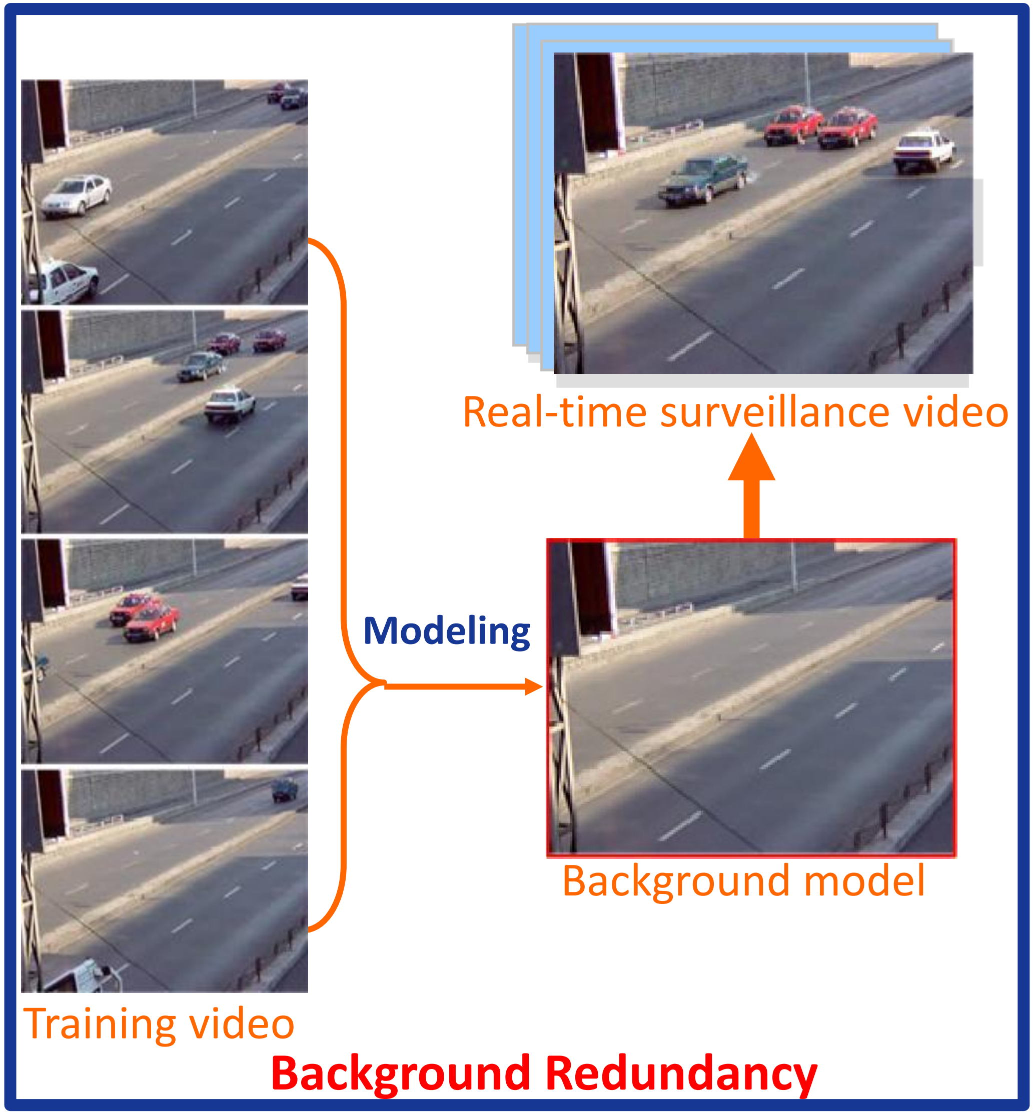
t

t+1

Spatial Redundancy

DCT

Perceptual Redundancy



Real-time surveillance video

Modeling

Background model

Training video

Background Redundancy

How to guarantee the transcoding quality and speed?

Whenever cloudy or fog, rainy or snow, in the morning or afternoon, most surveillance cameras are always deployed at a fixed position and cover a specific range of the scene for a long time even forever

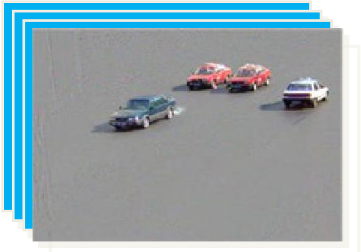
How to guarantee the transcoding quality and speed?

2012.7.10

- ✓ **Key idea:** Separately transcoding the foreground and background!

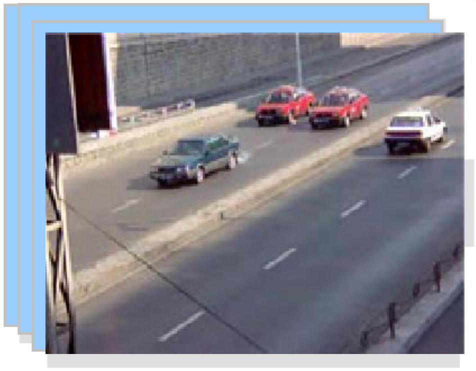
Training video

Modeling

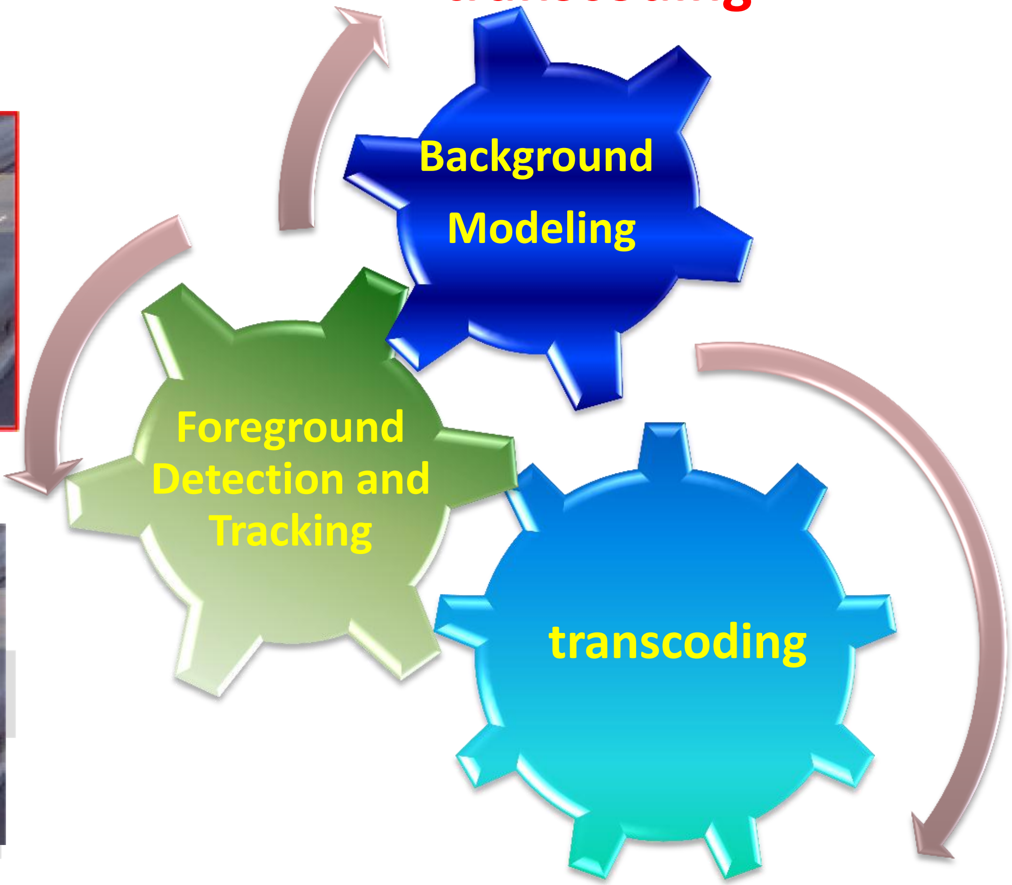


Foreground

Background model



Real-time surveillance video



Object-oriented transcoding

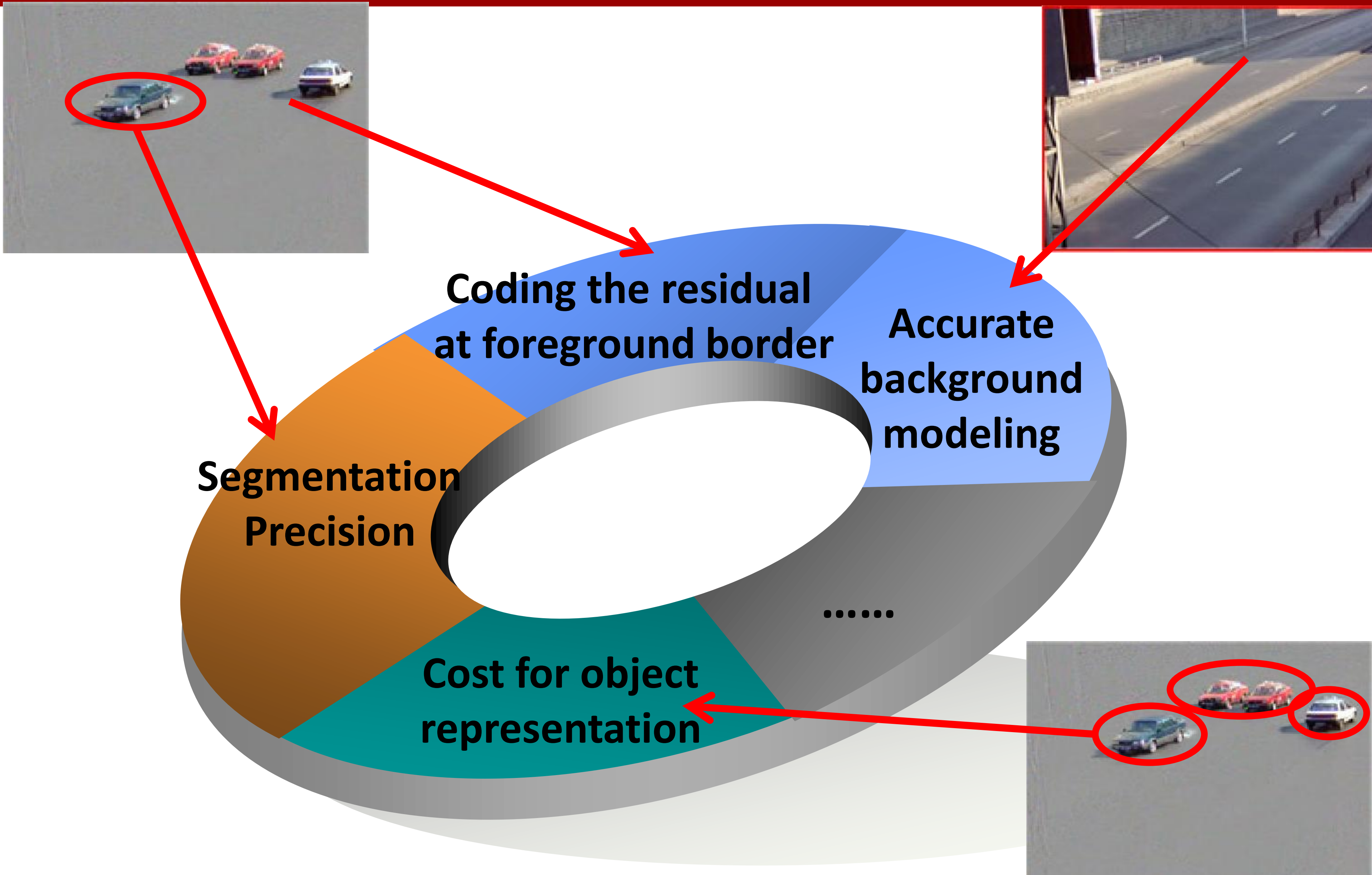
Background Modeling

Foreground Detection and Tracking

transcoding

Challenges in separately transcoding the background and foreground

2012.7.10



Coding the residual at foreground border

Accurate background modeling

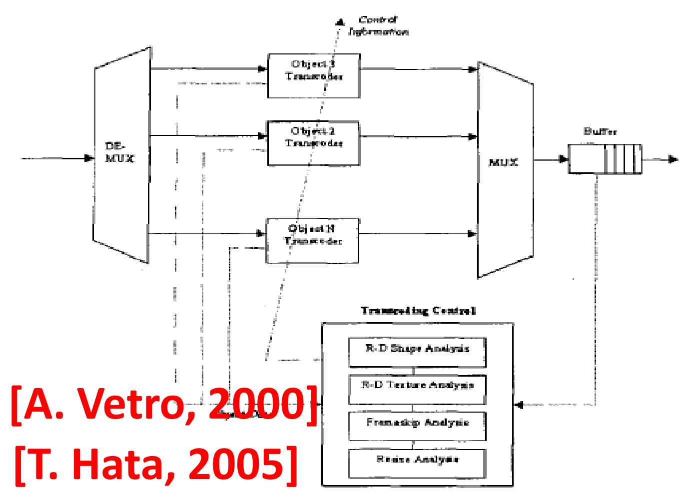
Segmentation Precision

Cost for object representation

……

How to handle these challenges in object-oriented coding?

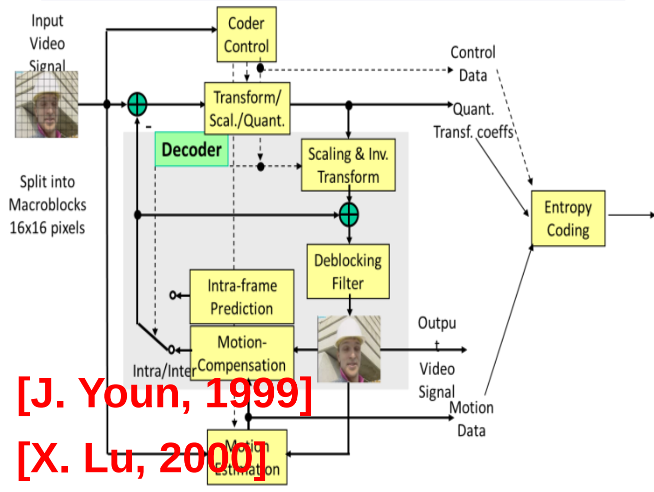
Object-oriented Transcoding



[A. Vetro, 2000]
[T. Hata, 2005]


China's
high-speed train

Hybrid Block-based Transcoding

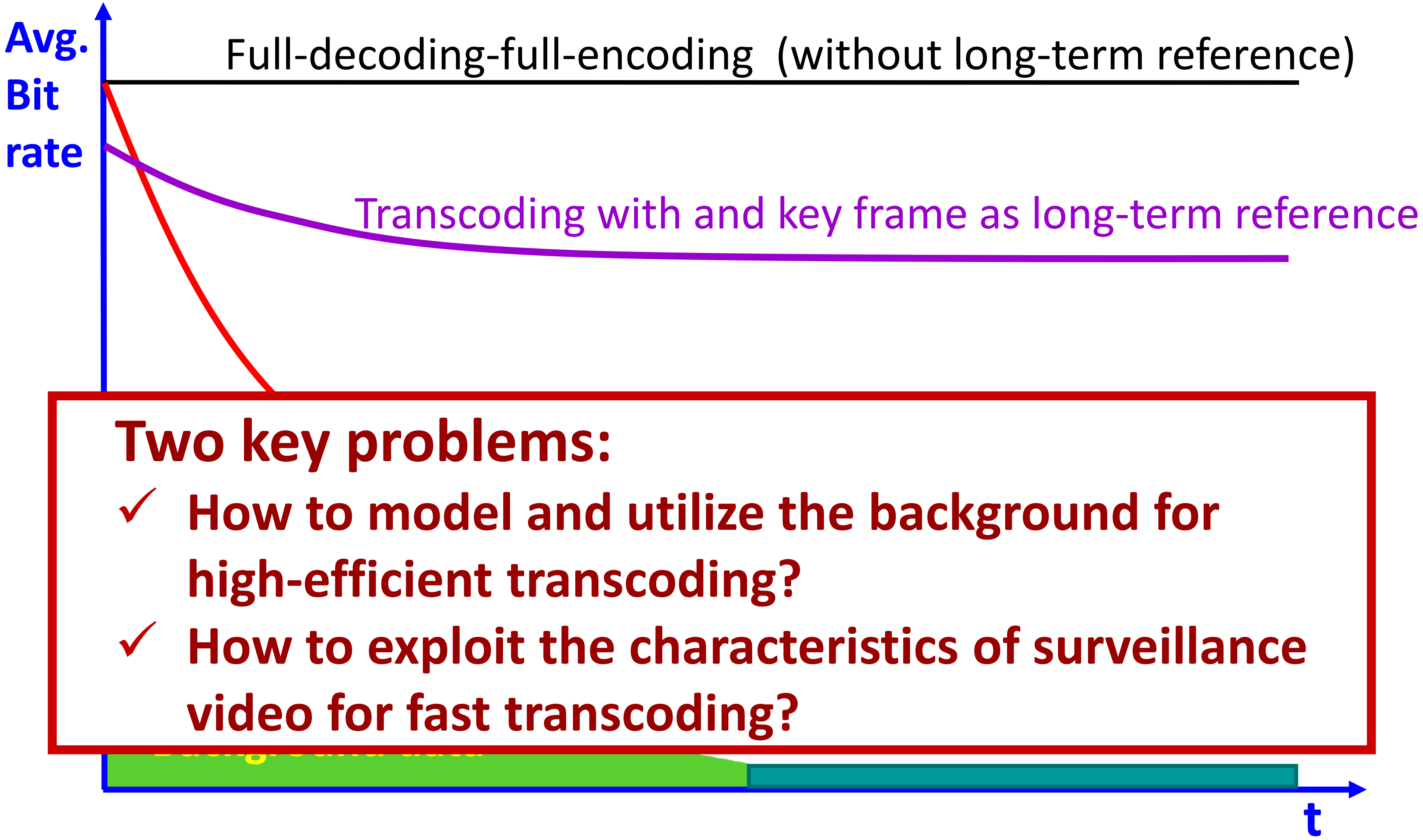


[J. Youn, 1999]
[X. Lu, 2000]

Can we integrate object-oriented transcoding with hybrid block-based methods for a best performance?

Hybrid block-based transcoding with background model

The expected bitrate curve



Two key problems:

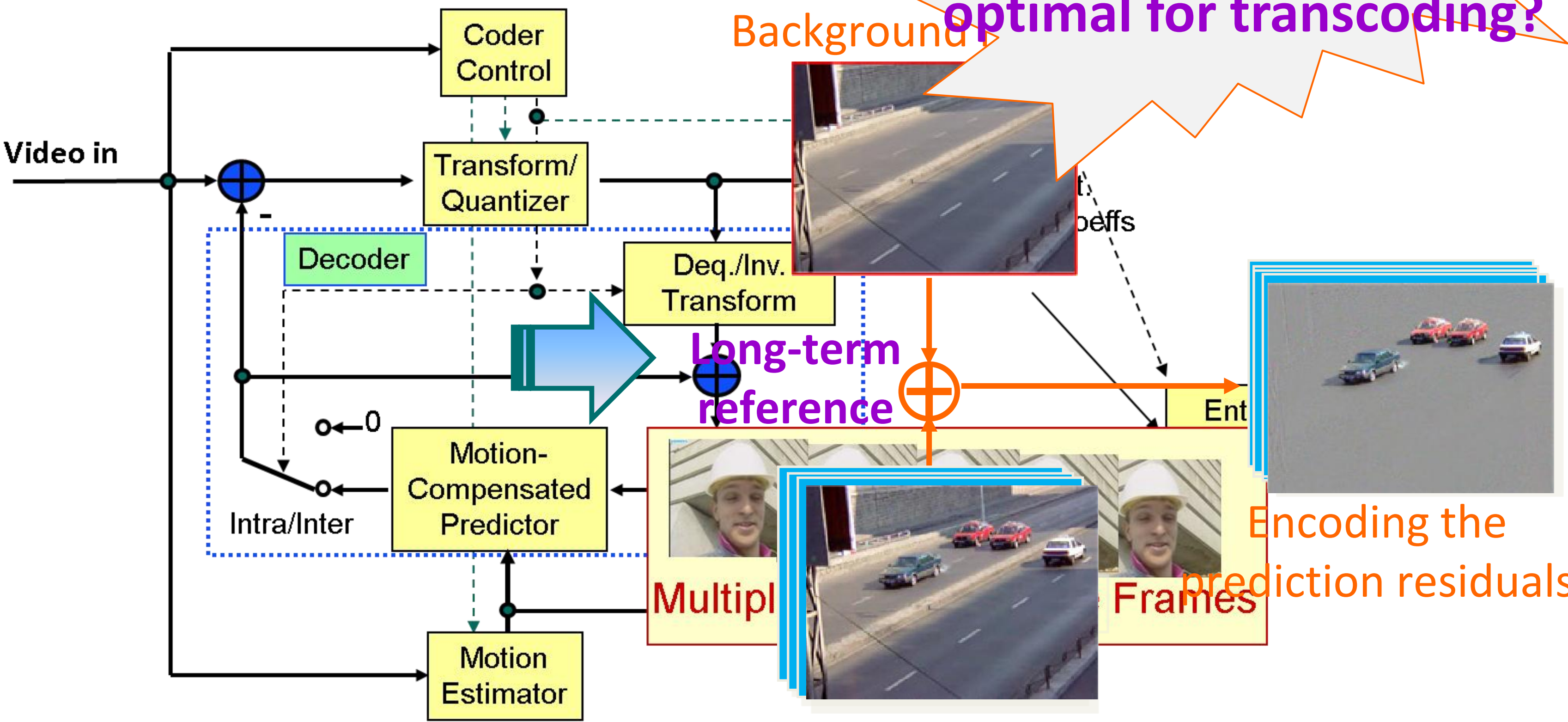
- ✓ **How to model and utilize the background for high-efficient transcoding?**
- ✓ **How to exploit the characteristics of surveillance video for fast transcoding?**

High-Efficient Transcoding:
Background Model as the Reference

2012.7.10

From: multiple past reference frames
To: background model as the reference

What kind of background model is optimal for transcoding?



Background

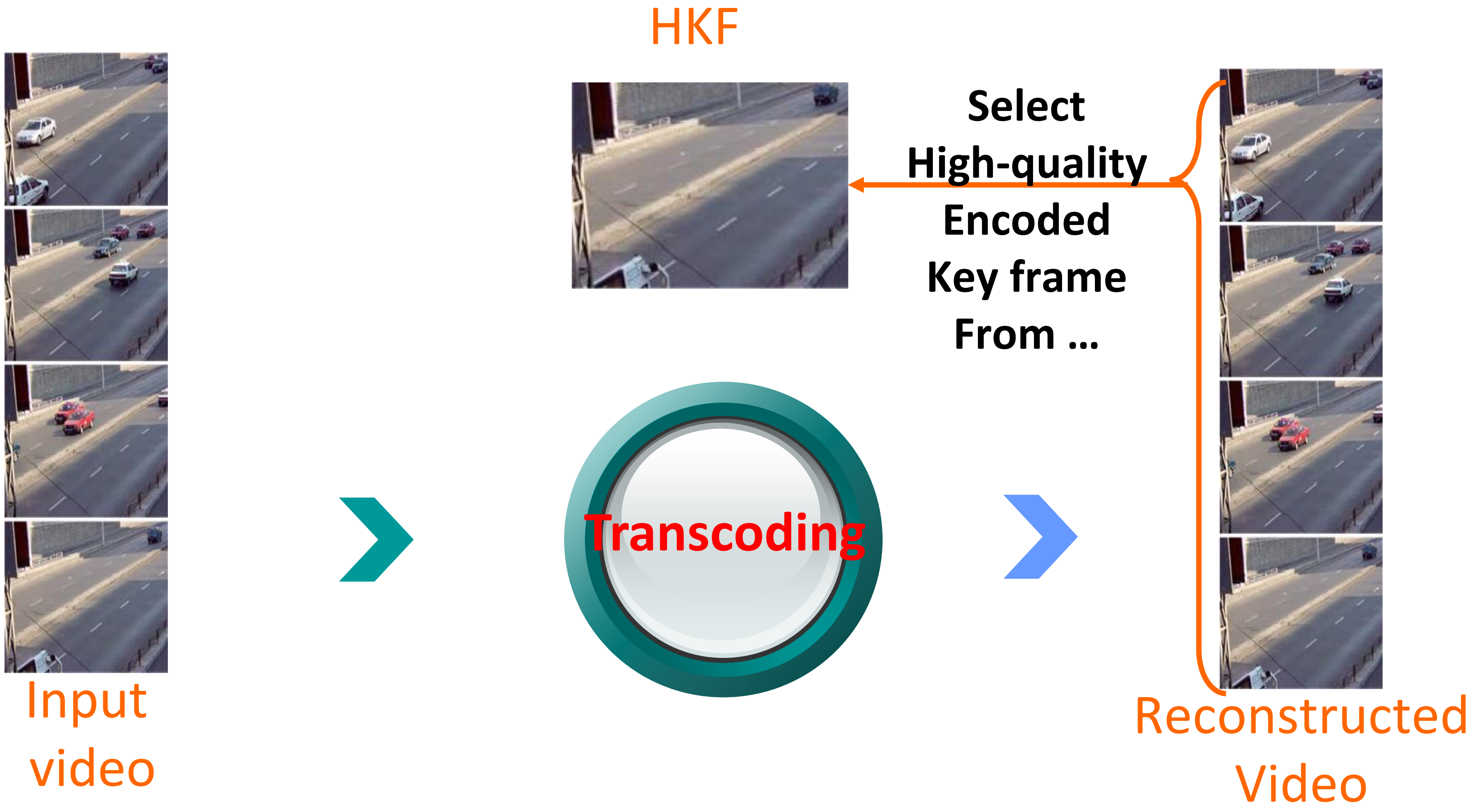
Long-term reference

Encoding the prediction residuals

Real-time surveillance video

High-Efficient Transcoding: *Optimal Background Model*

(1) Using the high-quality-transcoded key frame (HKF) as long-term reference



HKF

Select
High-quality
Encoded
Key frame
From …

Transcoding

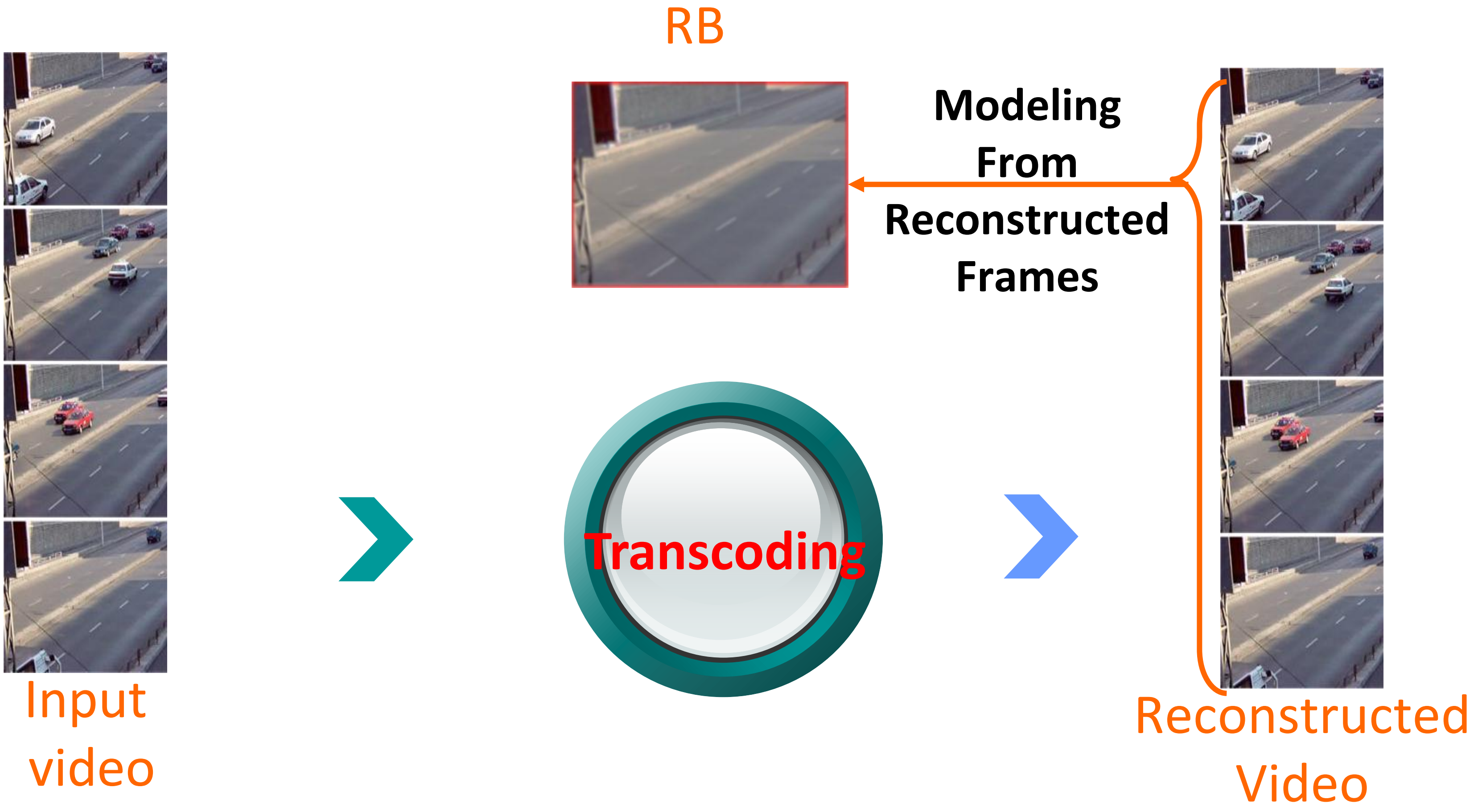
Input
video

Reconstructed
Video

High-Efficient Transcoding: *Optimal Background Model*

2012.7.10

(2) Using the background modeled from the reconstructed frames (RB) as long-term reference



RB

Modeling From Reconstructed Frames

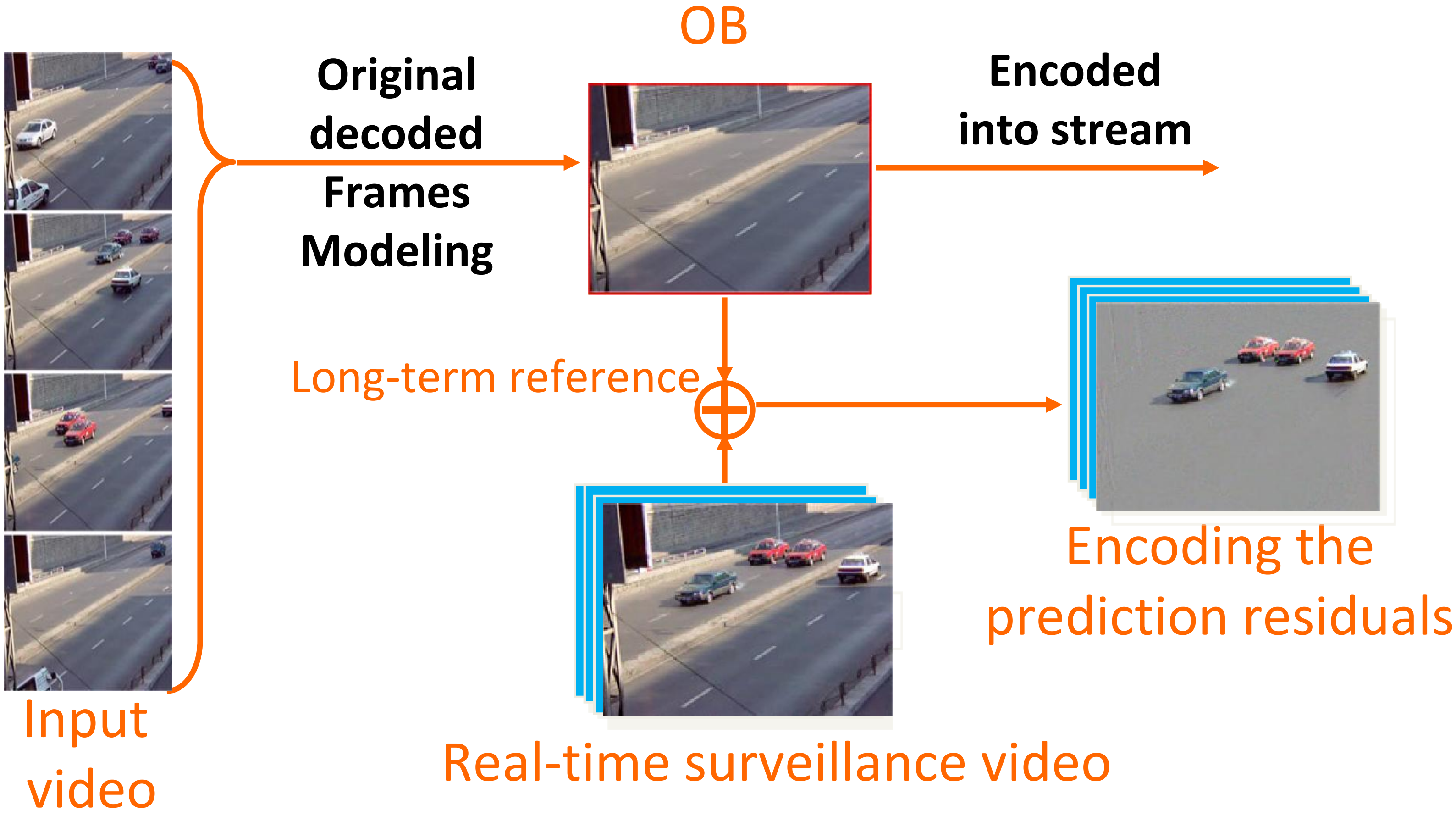
Transcoding

Input video

Reconstructed Video

High-Efficient Transcoding: *Optimal Background Model*

(3) Using the background modeled from the original-decoded frames (OB) as long-term reference



OB

Original decoded Frames Modeling

Encoded into stream

Long-term reference

Input video

Real-time surveillance video

Encoding the prediction residuals

High-Efficient Transcoding:
Theoretical Analysis: Why the OB is best?

2012.7.10

Exploiting the original-decoded-frame to train the background (OB) can achieve the best transcoding gains!

Lemma 1: *The distortion of OB is minimal*

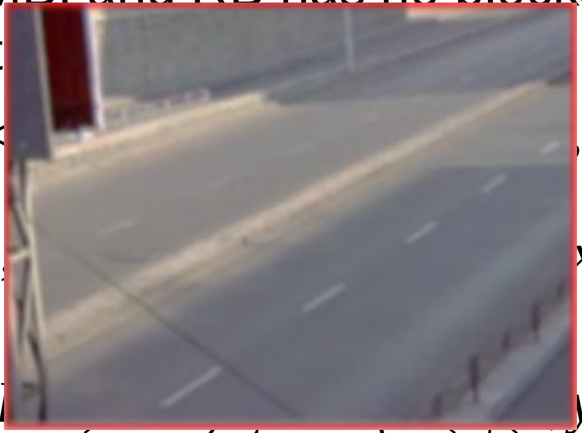
$$D(OB,\Lambda) < \min\{D(HKF,\Lambda), D(RB,\Lambda)\}$$

Proof: For any MB with the position (x, y) in an input frame I_i ϵΛ, HKF has no similar blocks with the MB, and RB has no blocks with little quantization lo… t…

HKF

High quality
But
Not clean

RB



Clean
But
Not high quality

OB

Clean
and
High quality

We get

$$D(I_i,OB) = \sum_{x,y} D(I_i(x,y),OB) < \min\{D(I_i,RB), D(I_i,HKF)\}$$

By integrating results for all frames in Λ, we can obtain the lemma.

High-Efficient Transcoding:
Theoretical Analysis: Why the OB is best?

Lemma 2: *The power spectral density of OB is minimal*

$$\Phi_{OB}(\Lambda) < \min\left\{\Phi_{RB}(\Lambda), \Phi_{HKF}(\Lambda)\right\}$$

Proof of Lemma 2 can be derived from [Leontaris & Cosman, 2007] and [Girod, 1987]

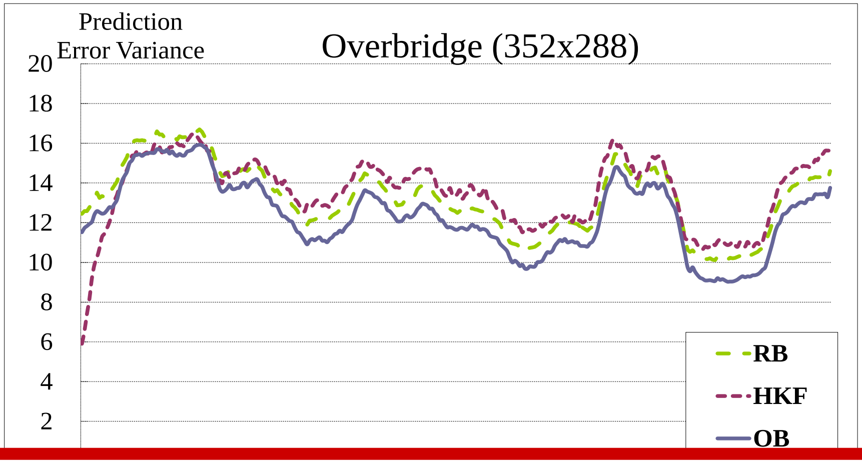
$$\Delta\Phi_{OB,HKF}(\Lambda) = \Phi_{OB}(\Lambda) - \Phi_{HKF}(\Lambda)$$

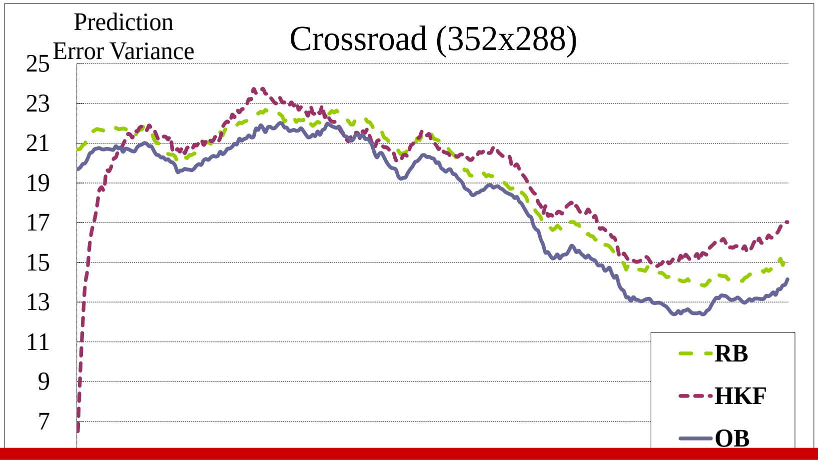
$$\approx \frac{1}{4}\times\sum_{k=1}^{n}\left(\Psi\left(\Gamma\left(I_k(x,y),OB\right)\right) - \Psi\left(\sum_{x,y}\left(\Gamma\left(I_k(x,y),HKF\right)\right)\right)\right) < 0$$

$$\boxed{\Delta\Phi_{i,j}(\Lambda) \approx \frac{\Phi_{nn\_l\_i}(\Lambda) - \Phi_{nn\_l\_j}(\Lambda)}{4}}$$

$$\Delta\Phi_{OB,RB}(\Lambda) = \Phi_{OB}(\Lambda) - \Phi_{RB}(\Lambda)$$

$$\approx \sum_{k=1}^{n}\left(\Psi\left(\Gamma\left(I_k(x,y),OB\right)\right) - \Psi\left(\sum_{x,y}\left(\Gamma\left(I_k(x,y),RB\right)\right)\right)\right) < 0$$


Overbridge (352x288)


Crossroad (352x288)

After the initial several frames, the PEV of the OB becomes less than that of the HKF and RB.

High-Efficient Transcoding:

Theoretical Analysis: Why the OB is best?

Theorem 1: *Let RD(OB, Λ)/RD(RB, Λ)/RD(HKF, Λ) denote the rate-distortion performance between an input long surveillance sequence and OB/RB/HKF; Using the same motion search method,*

$$RD(OB,\Lambda) < \min\{RD(OB,\Lambda), RD(RB,\Lambda)\}$$

Proof:

As stated in [Liu et al. 2010], when given two prediction reference frames, the one with smaller prediction distortion and smaller Φ(Λ) leads to a better compression efficiency.

Thus by integrating Lemma 1 and 2, we can derive Theorem 1.

So the remaining problem becomes

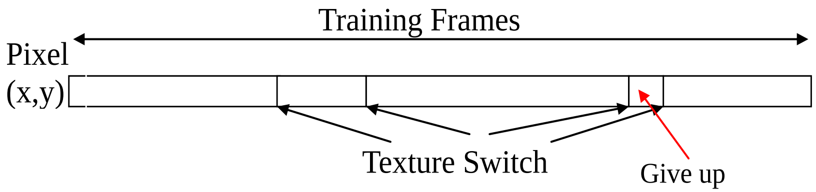
how to model OB from the decoded sequence?

High-Efficient Transcoding:
Low-Complexity Background Modeling

2012.7.10

Segment-and-Weight based Running Averaging (SWRA)

- Step 1: Divide training frames into segments



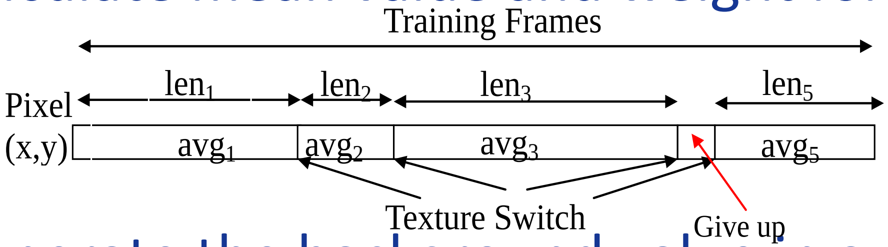
Training Frames

Pixel (x,y)

Texture Switch

Give up

- Step 2: Calculate mean value and weight for each segment



Training Frames

Pixel (x,y)

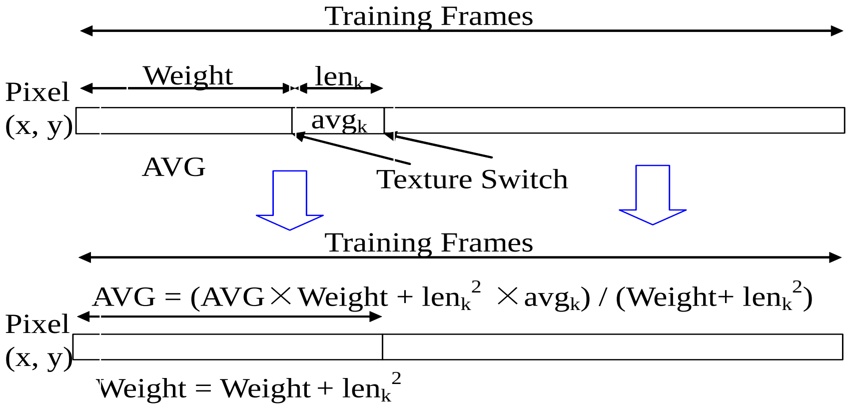
len_1 len_2 len_3 len_5

avg_1 avg_2 avg_3 avg_5

Texture Switch

Give up

- Step 3: Generate the background value in a running way



Training Frames

Pixel (x, y)

Weight len_k

avg_k

AVG

Texture Switch

Training Frames

$AVG = (AVG \times Weight + len_k^2 \times avg_k) / (Weight + len_k^2)$

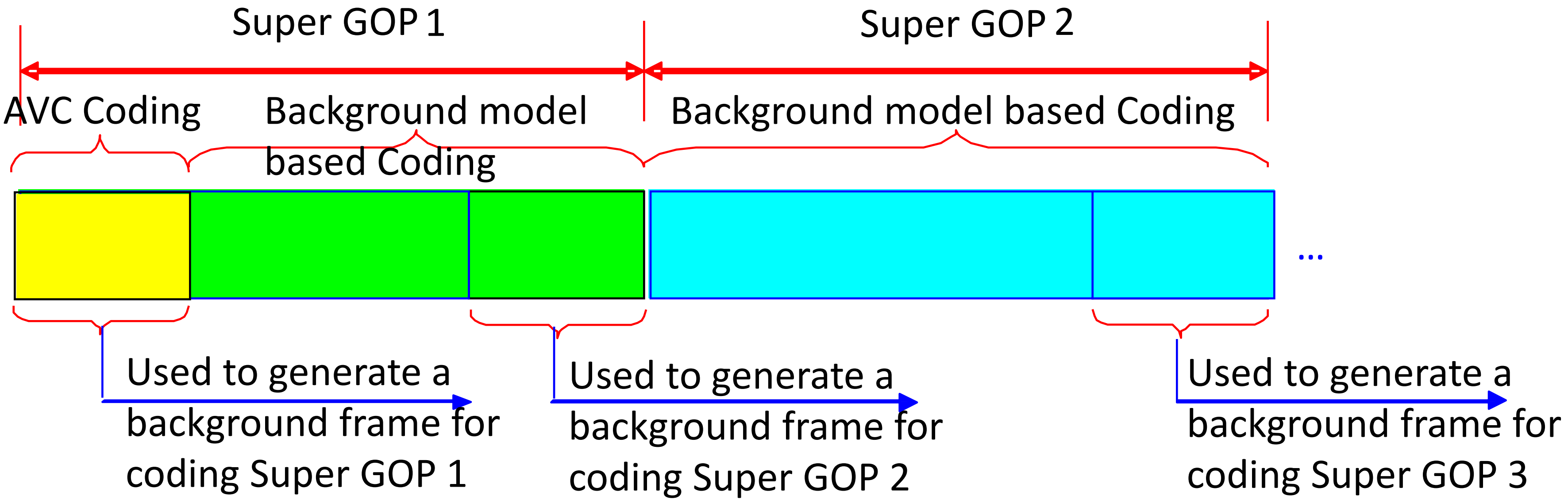
Pixel (x, y)

$Weight = Weight + len_k^2$

XG Zhang, YH Tian, et. al, LOW-COMPLEXITY AND HIGH-EFFICIENCY BACKGROUND MODELING FOR SURVEILLANCE VIDEO CODING, VCIP'2012

High-Efficient Transcoding: Background Model Update

- The no-delay sequence structure background updating



Super GOP 1

Super GOP 2

AVC Coding

Background model based Coding

Background model based Coding

…

Used to generate a background frame for coding Super GOP 1

Used to generate a background frame for coding Super GOP 2

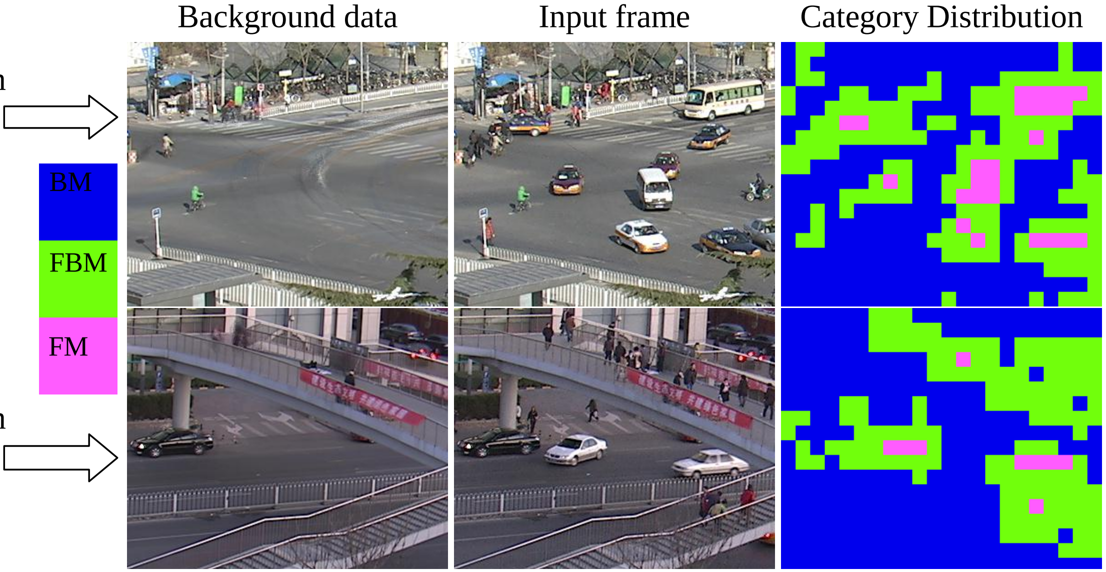
Used to generate a background frame for coding Super GOP 3

Fast Transcoding: MB Classification Based Algorithms

2012.7.10

- Key idea: *To classify MBs into three categories, and then separately design suitable fast transcoding methods for each MB category.*

$$\sum_{i=1}^{N_B}\left(\left(C_i - BG_i\right) - \beta\right) > 0$$



The 206th frame in Crossroad(CIF)

The 170th frame in Overbridge(CIF)

FM: With few background pixels

BM: With few foreground pixels

FBM: foreground border MBs

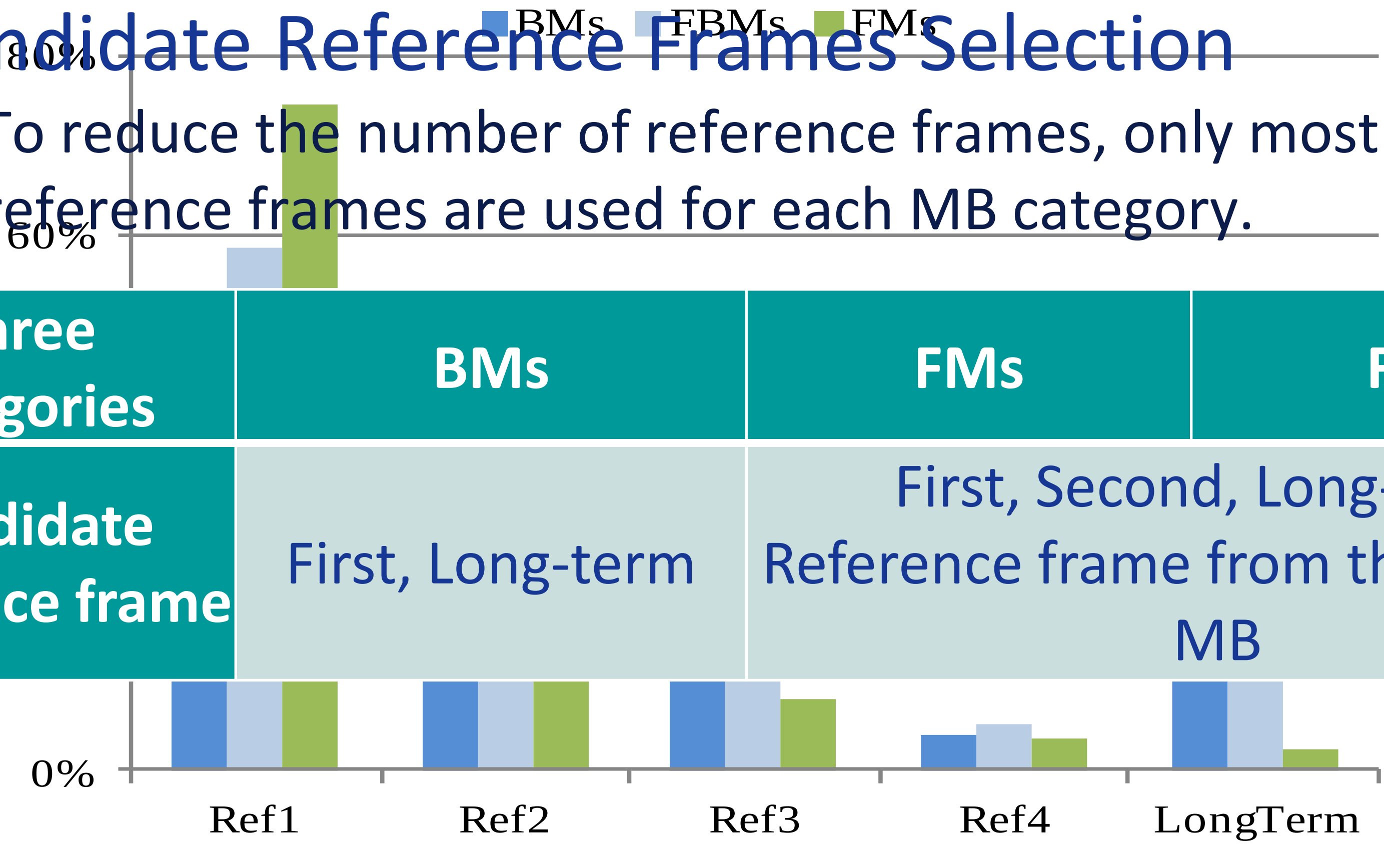
Fast Transcoding Algorithm 1:
Reference Frame Selection

MB Category based Analysis for reference frames

- The 1st reference frame is necessary for all the MBs
- The long-term reference frame also takes up a large percent for BMs and FBMs

Candidate Reference Frames Selection

- To reduce the number of reference frames, only most frequent reference frames are used for each MB category.



Three categories	BMs	FMs	FBMs
Candidate reference frame	First, Long-term	First, Second, Long-term, Reference frame from the decoded MB	

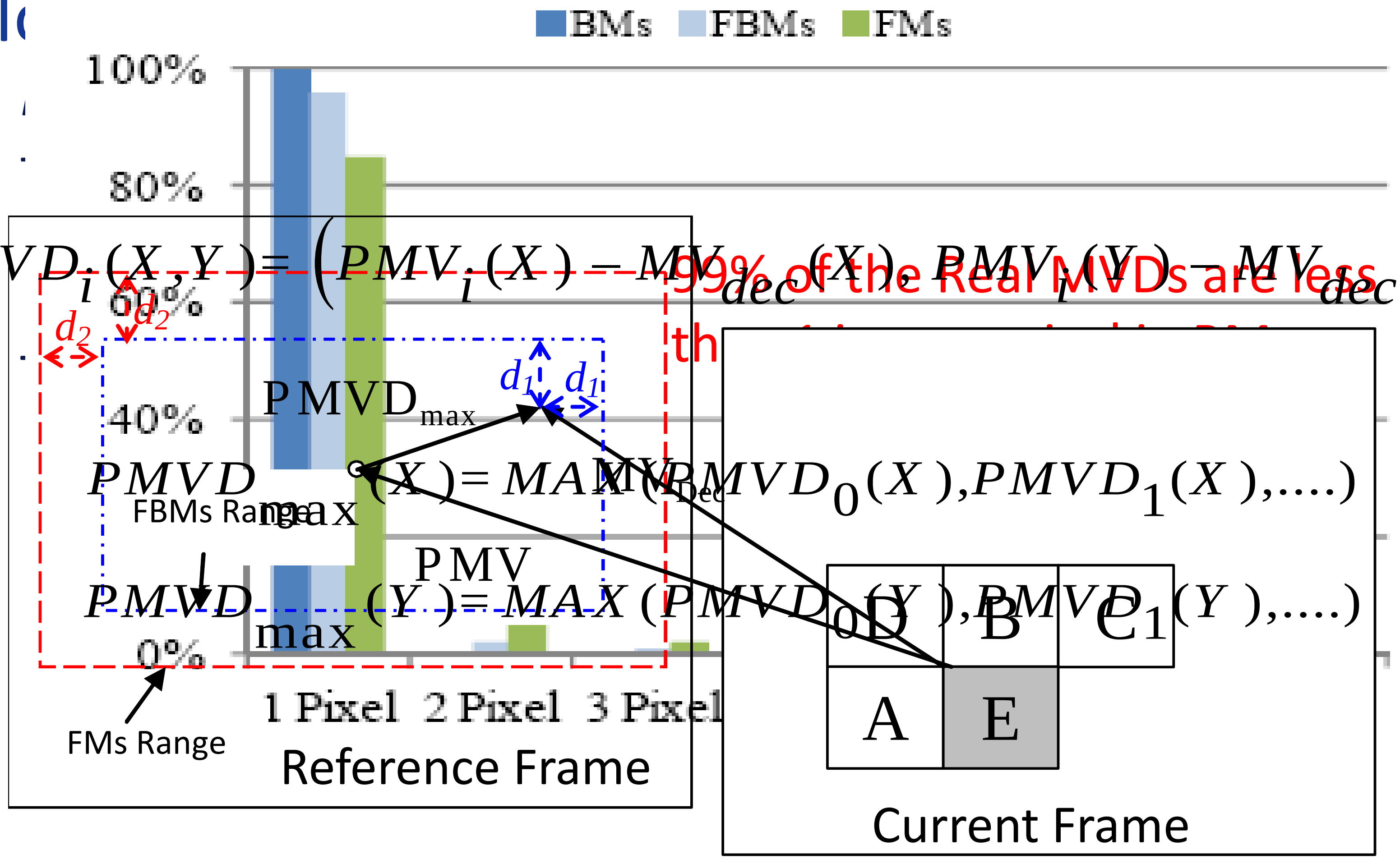
Fast Transcoding Algorithm 2:
Adaptive Motion Search Range Calculation

2012.7.10

- Analysis for the distribution of Real MVDs

  Real MVDs: difference between decoded MV and predicted MV

- Mc



$$PMVD_i(X,Y) = \left(PMV_i(X) - MV_{dec}(X),\ PMV_i(Y) - MV_{dec}(Y)\right)$$

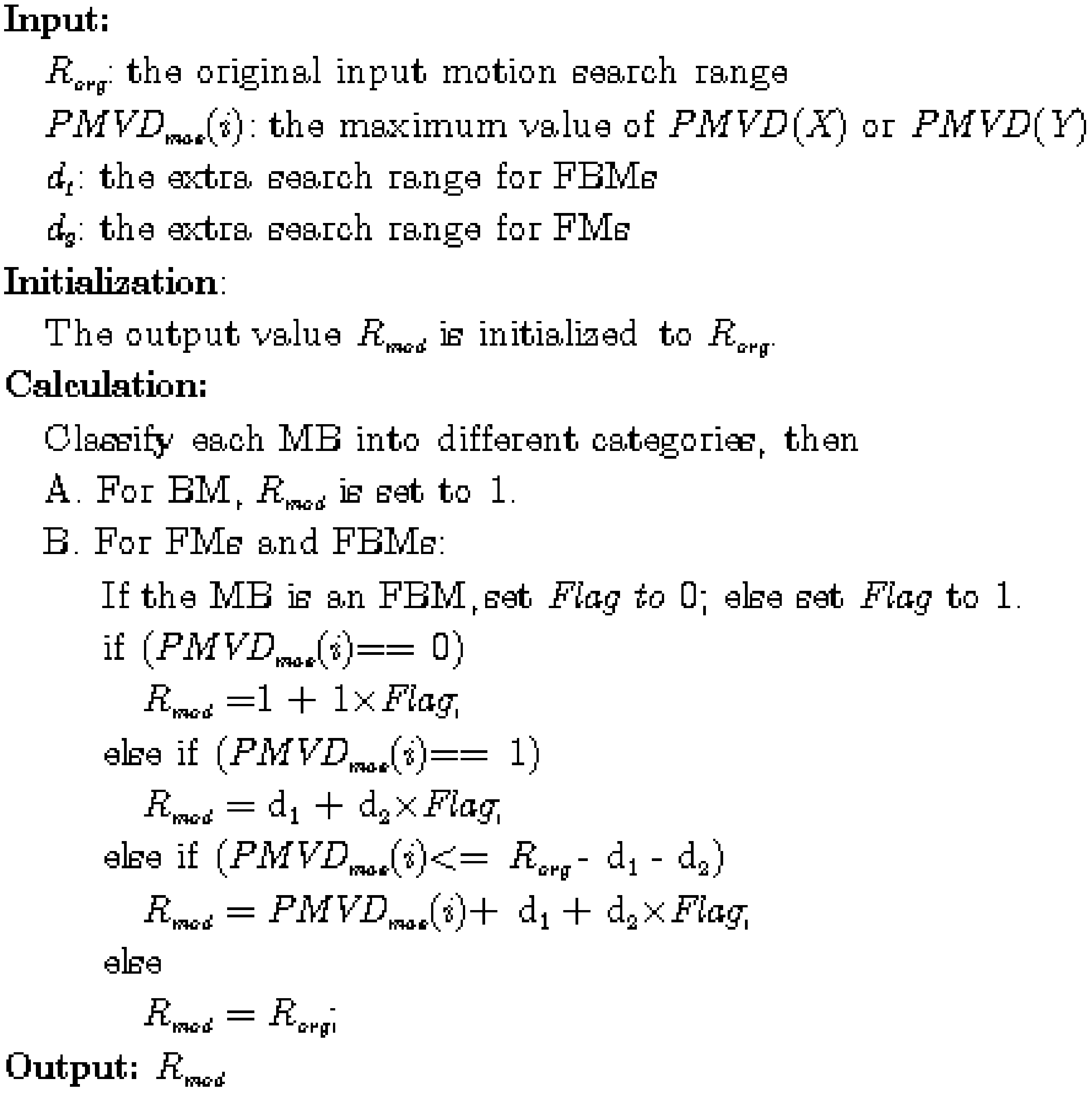
99% of the Real MVDs are less

th

$$PMVD_{max}(X) = MAX\left(PMVD_0(X), PMVD_1(X), \ldots.\right)$$

$$PMVD_{max}(Y) = MAX\left(PMVD_0(Y), PMVD_1(Y), \ldots.\right)$$

Fast Transcoding Algorithm 2: *Adaptive Motion Search Range Calculation*



```
Input:
    $R_{org}$: the original input motion search range
    $PMVD_{max}(i)$: the maximum value of $PMVD(X)$ or $PMVD(Y)$
    $d_1$: the extra search range for FBMs
    $d_2$: the extra search range for FMs
Initialization:
    The output value $R_{mod}$ is initialized to $R_{org}$.
Calculation:
    Classify each MB into different categories, then
    A. For BM, $R_{mod}$ is set to 1.
    B. For FMs and FBMs:
        If the MB is an FBM, set $Flag$ to 0; else set $Flag$ to 1.
        if ($PMVD_{max}(i) == 0$)
            $R_{mod} = 1 + 1 \times Flag$;
        else if ($PMVD_{max}(i) == 1$)
            $R_{mod} = d_1 + d_2 \times Flag$;
        else if ($PMVD_{max}(i) <= R_{org} - d_1 - d_2$)
            $R_{mod} = PMVD_{max}(i) + d_1 + d_2 \times Flag$;
        else
            $R_{mod} = R_{org}$;
Output: $R_{mod}$
```

Fast Transcoding Algorithm 3: *Mode Decision Refinement*

2012.7.10

❖ Mode refinement

Analysis for the distribution of candidate modes

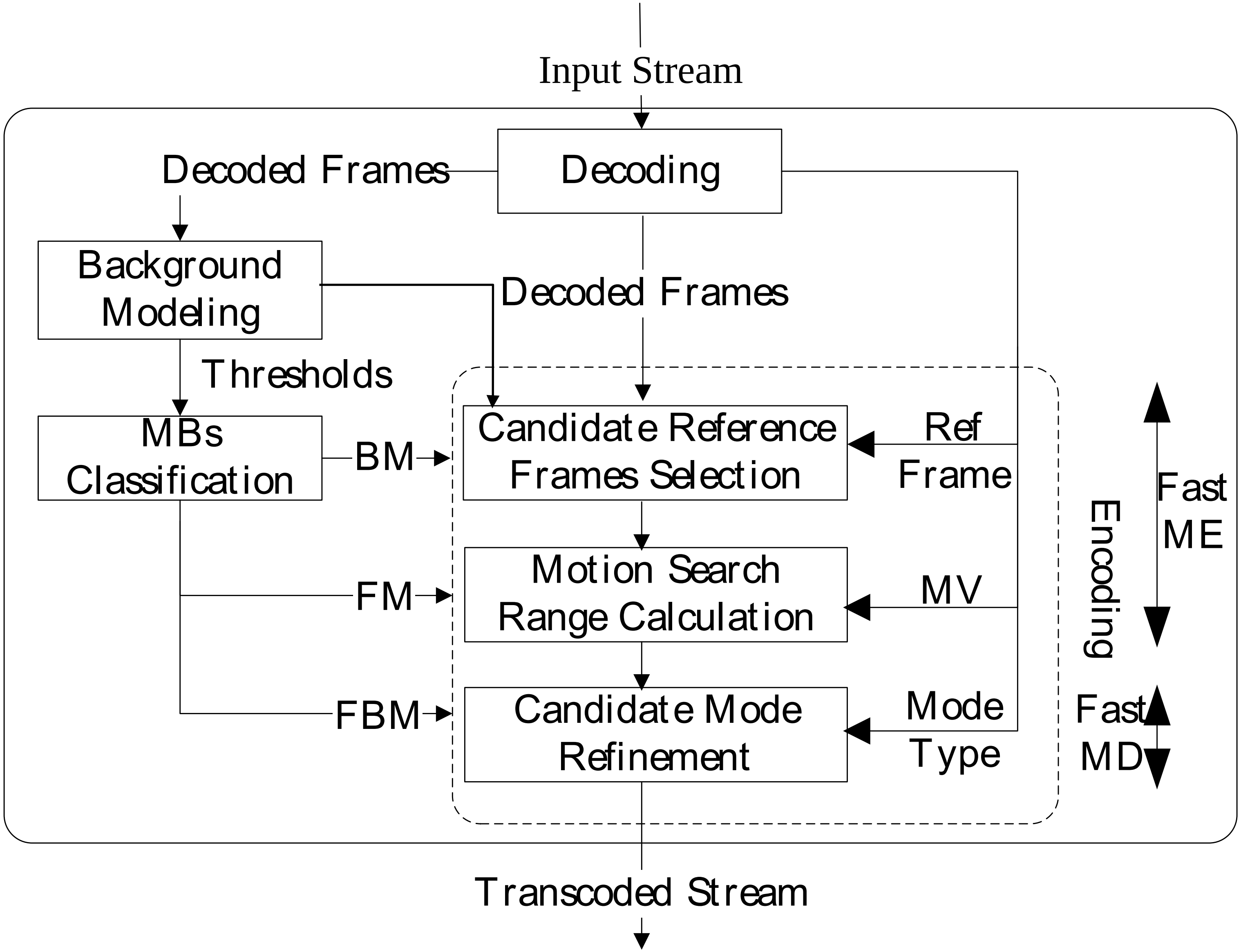
- Unnecessary mode types are forbidden in this step to reduce ME and MD complexity
- SKIP and inter 16×16 prediction modes are selected in almost 96% BMs.
- I16x16 intra prediction is seldom used in FBMs

Decode Mode	FMs	FBMs	BMs
S	level1, I16MB	level2	16×16, SKIP
8×4	level2, I16MB		
4×8			
4×4	level3, I16MB	level3	
I4MB			

***S: Decode Mode Size={ SKIP,16×16,16×8,8×16,8×8, I16MB},**
***level1={SKIP,16×16,16×8,8×16,8×8,I4MB},**
***level2={SKIP,16×16,16×8,8×16,8×8,I4MB,8×4,4×8},**
***level3={SKIP,16×16,16×8,8×16,8×8,I4MB,8×4,4×8,4×4}**

System Framework

2012.7.10



Experiments

Extensive Datasets

2 Mbps

Crossroad (CIF)

Overbridge(CIF)

Snowroad(CIF)

Snowway(CIF)

4 Mbps

Crossroad(SD)

Overbridge(SD)

Office(SD)

Bank(SD)

32 Mbps

CarRoad(HD)

Crossroad(HD)

Experimental Setup

Transcoding Anchors

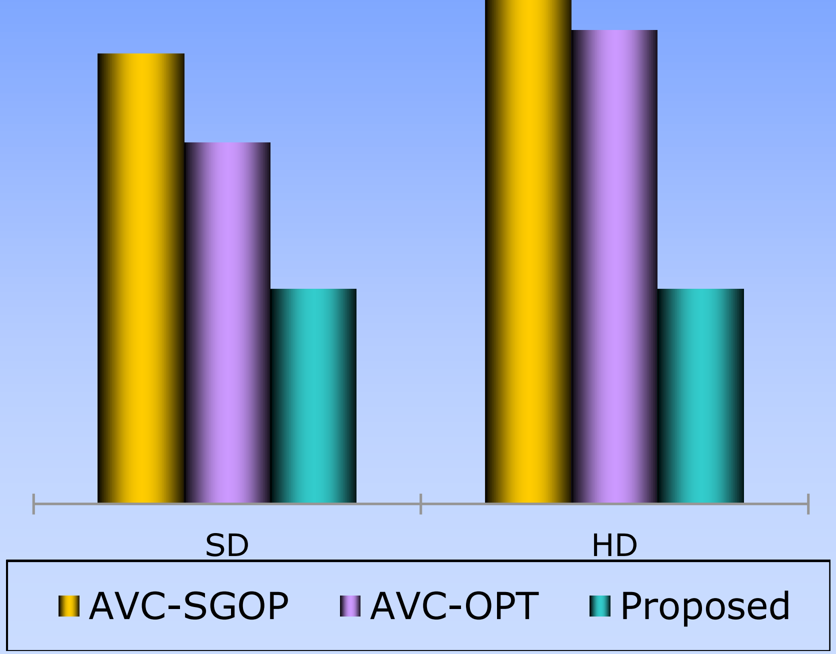
- FDFE+AVC-SGOP: H.264 with IntraPeriod=Length of S-GOP
- FDFE+AVC-OPT: H.264 using Key Frame as long-term reference

Parameter	Value	Parameter	Value
Porfile	Baseline	Used MODE	ALL
Rate Control	Disable	Framerate	25
Entropy Coding	UVLC	Frame Structure	IPPP
Search Range	32	IntraPeriod	0
RD Optimization	High	SAD Method	Hadamard
Motion Search	Fast Full	Reference Num	5

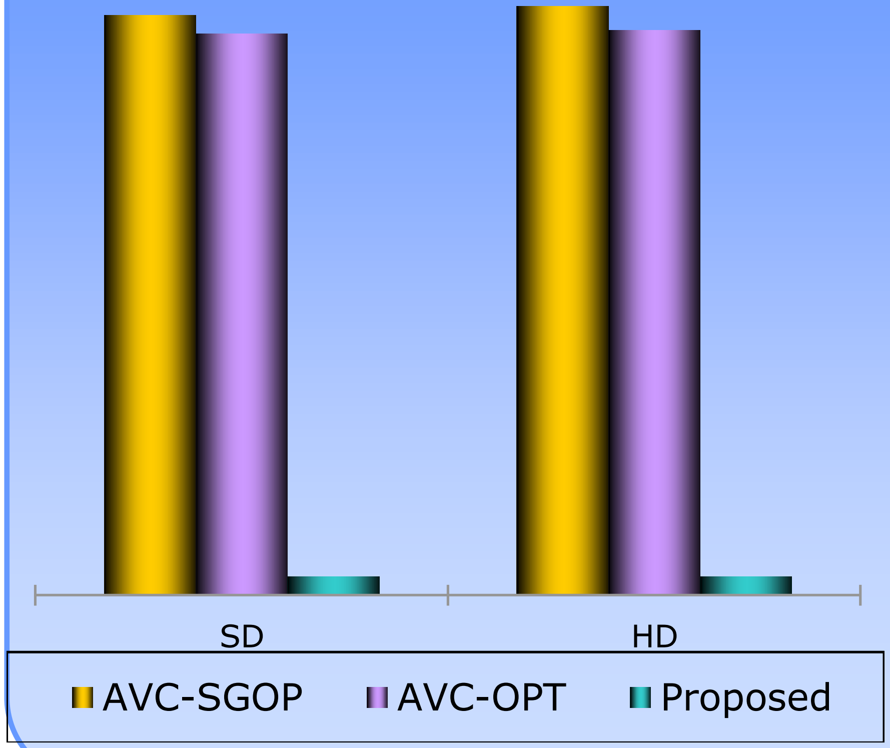
Experimental Results (1)

2012.7.10

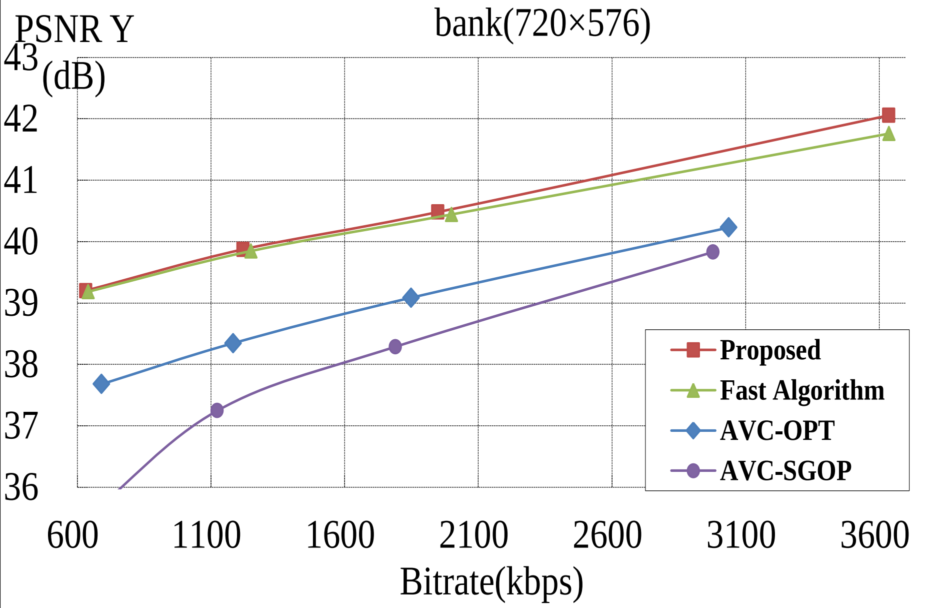
Bitrate

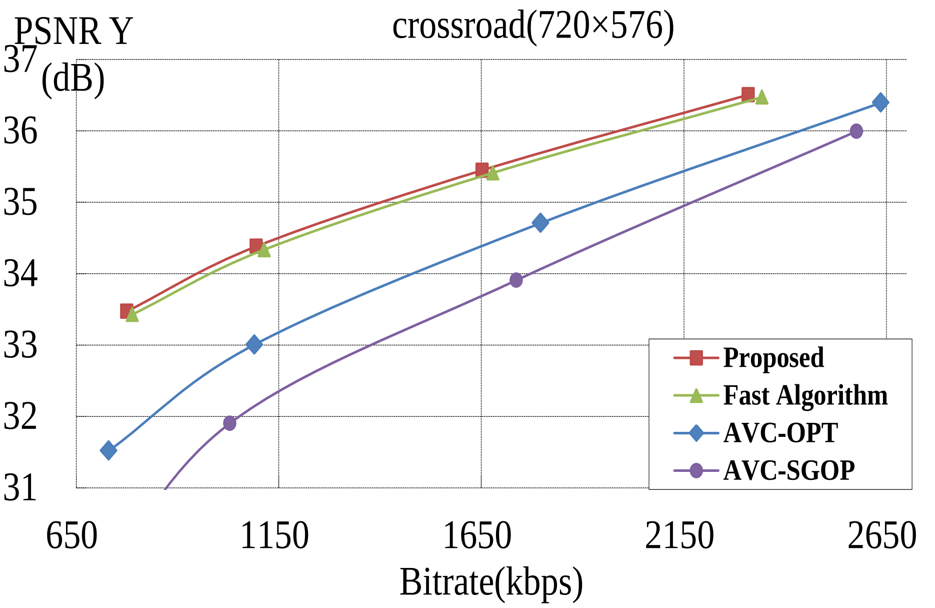


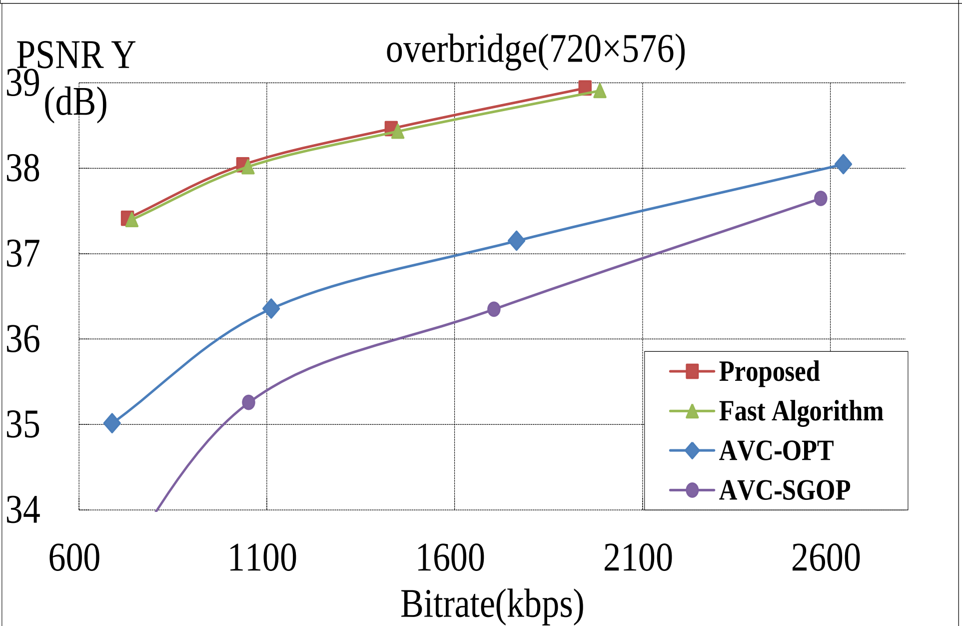
Complexity

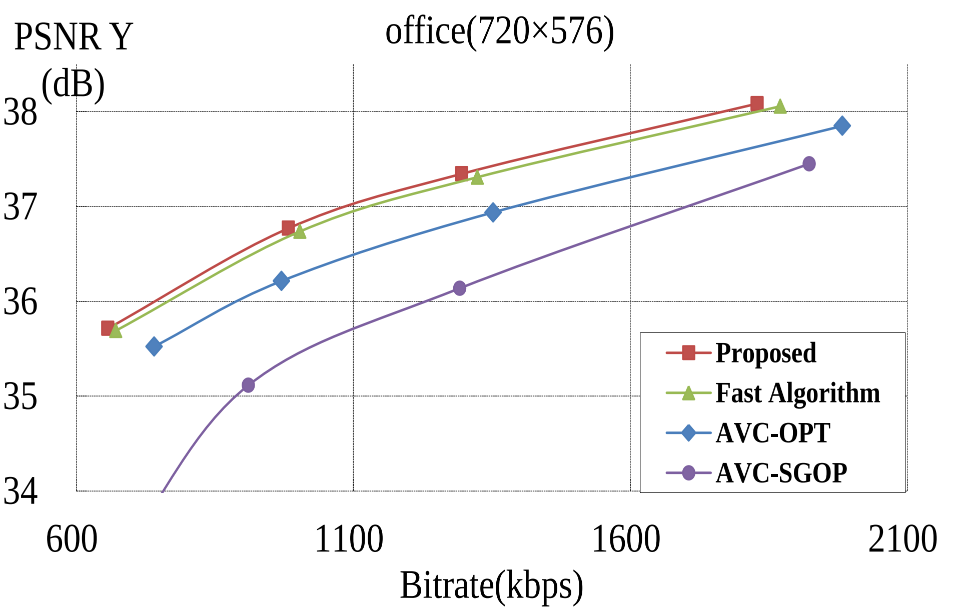


Experimental Results (2)









Experimental Results (3)

Contribution Distribution of Different Components

- Motion search range results in significant total-time saving

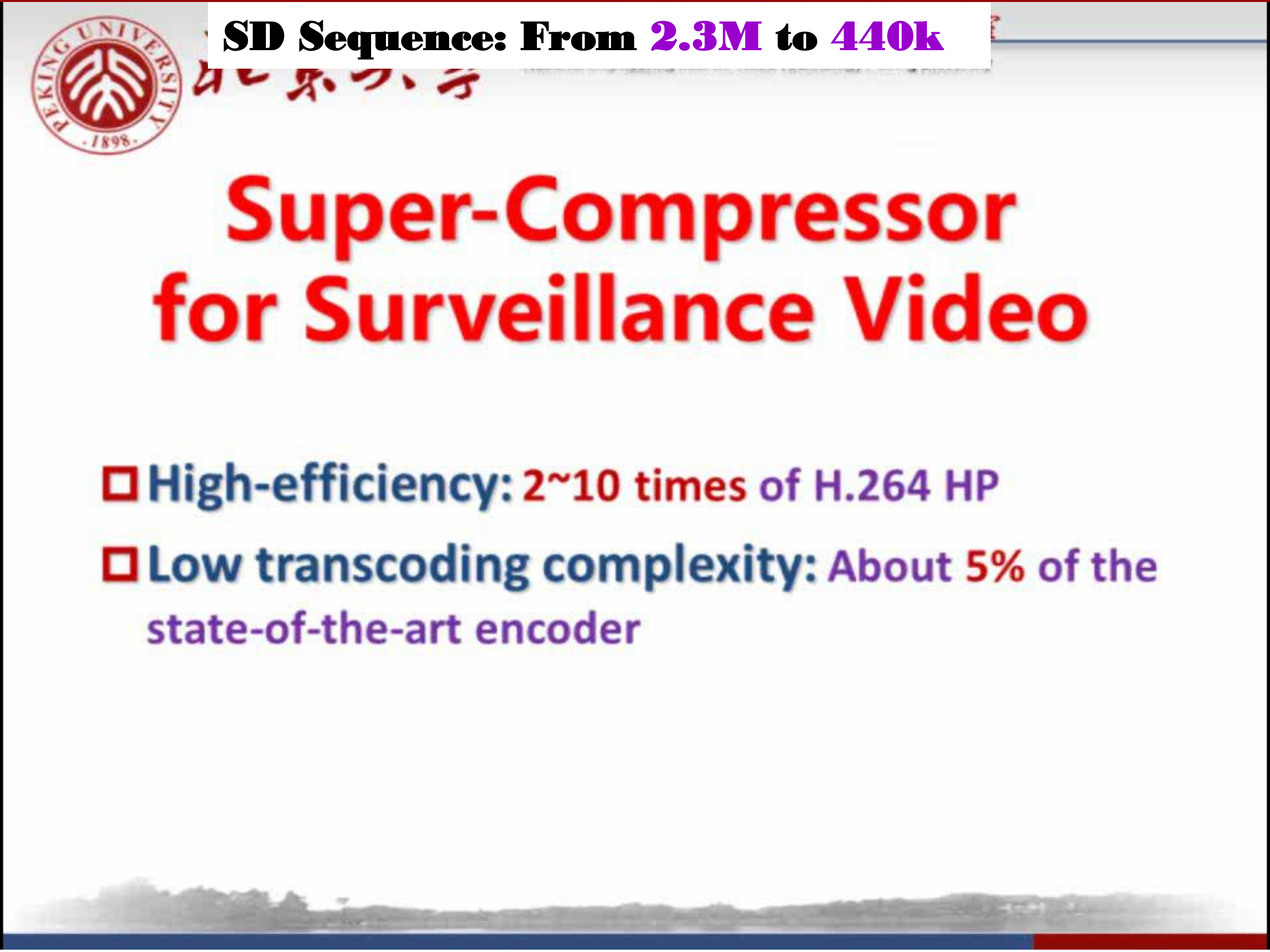
Table 5. Separate contribution for each algorithm (%)

	Crossroad	Overbridge	Snowroad	Snowate.	**average**
Search Range Calculation	93.91	93.02	94.77	94.90	**94.15**
Mode Refinement	12.29	10.69	16.73	14.00	**13.43**
Reference Frame Selection	34.05	31.17	47.89	42.08	**38.80**

Table 6. Search Points Reduction(%)

Crossroad	Overbridge	Snowroad	Snowate.	**average**
87.42	80.00	93.48	95.35	**89.06**

DEMO



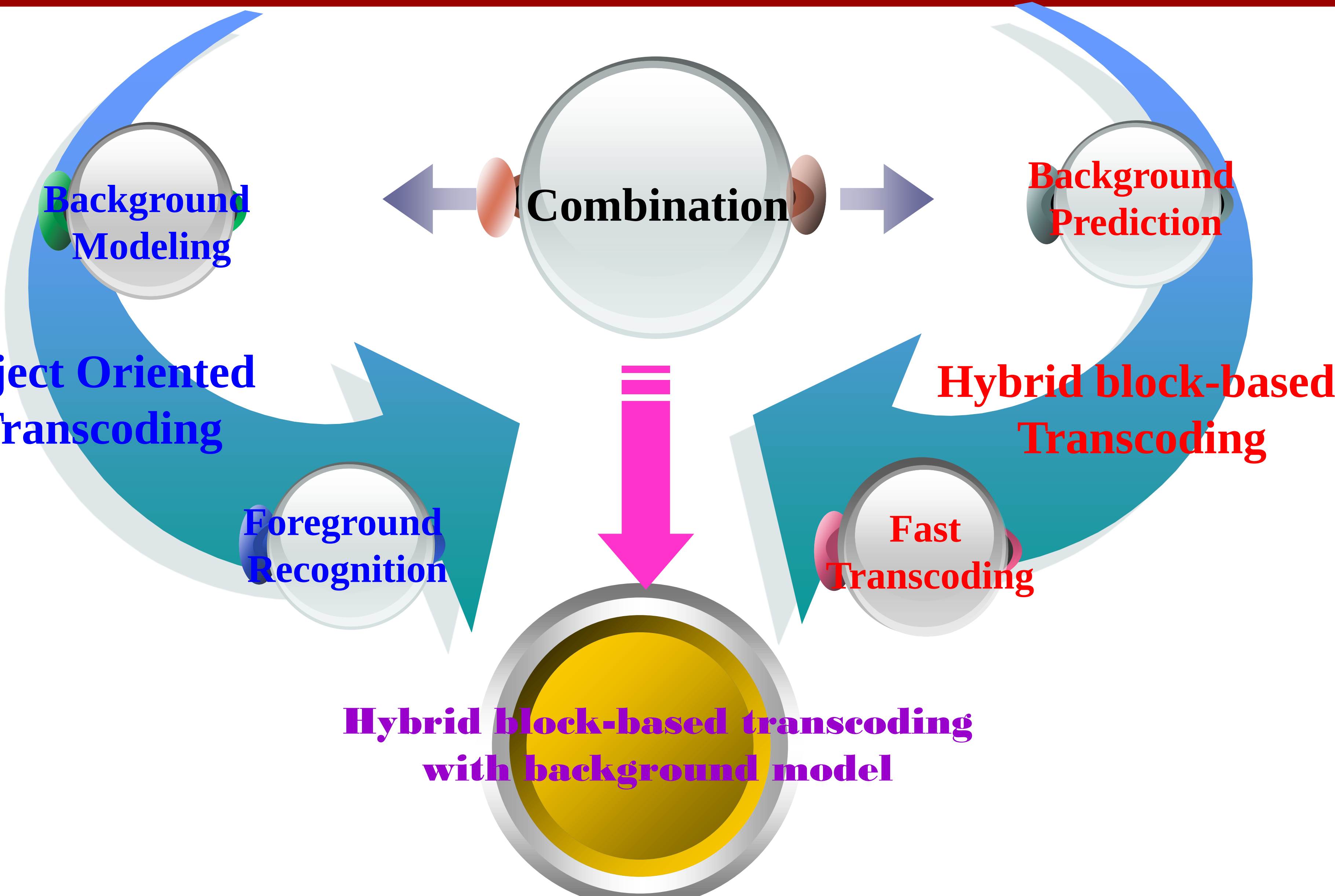
SD Sequence: From 2.3M to 440k

Super-Compressor for Surveillance Video

- ☐ **High-efficiency:** 2~10 times of H.264 HP
- ☐ **Low transcoding complexity:** About 5% of the state-of-the-art encoder

Conclusion

2012.7.10



Background Modeling

Combination

Background Prediction

Object Oriented Transcoding

Hybrid block-based Transcoding

Foreground Recognition

Fast Transcoding

Hybrid block-based transcoding with background model

Conclusion

Contribution

- Theoretically prove OB is a optimal reference, transcoding With this background model saves half the bit-rate
- MB-classification based Analysis and design three methods to speed-up the transcoding
- Extensive Experiments On CIF/SD/HD Videos
- Practically co-operating with Hisense Co. LTD. for Surveillance Video transcoding

Thank You !