

# Packet Loss Visibility for View+Depth Compressed Stereo 3D Video

Qing Song, Pamela Cosman

University of California, San Diego

December 12, 2013



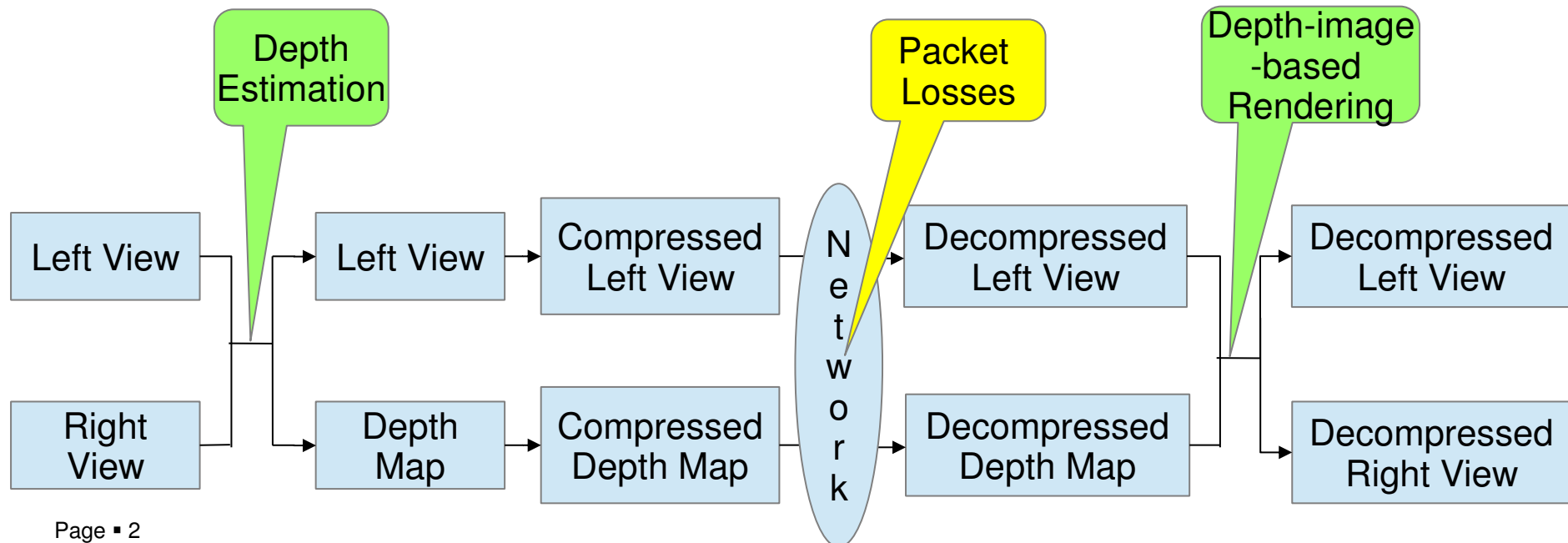
- ❖ 3D video: more data for storage and transmission
  - View+depth compression.
  - Depth includes no texture, can be compressed efficiently
- ❖ Lossy channel: packets can be lost.
  - Visibility of the packet loss, unequal error protection (UEP)
  - Human observer experiment



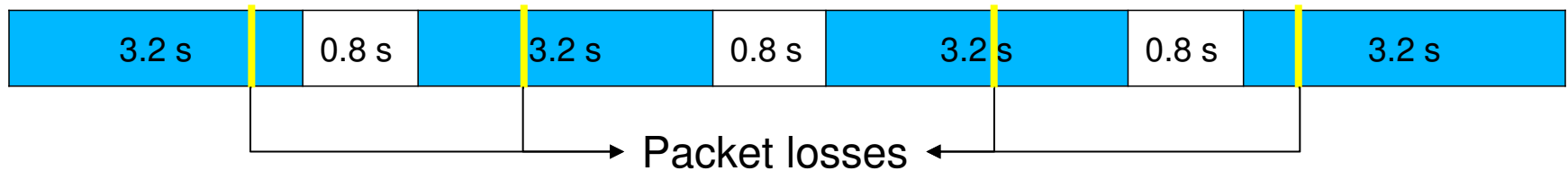
Color



Depth



- Insert packet loss every 4 sec, losses occur in the first 3.2 sec randomly, last 0.8 sec will allow any error propagation to stop



- Show lossy videos to people, viewers press space bar when see a glitch
- Response within 1 second considered as hit



- Video: 21'20"
- 5 versions of lossy videos
- Each version includes 300 losses
- Each version of the lossy video was shown to 12 people (ages 18 – 28)
- All participants have good stereo vision (tested by stereo fly test)
- A 3-min pilot training video was shown before the experiment

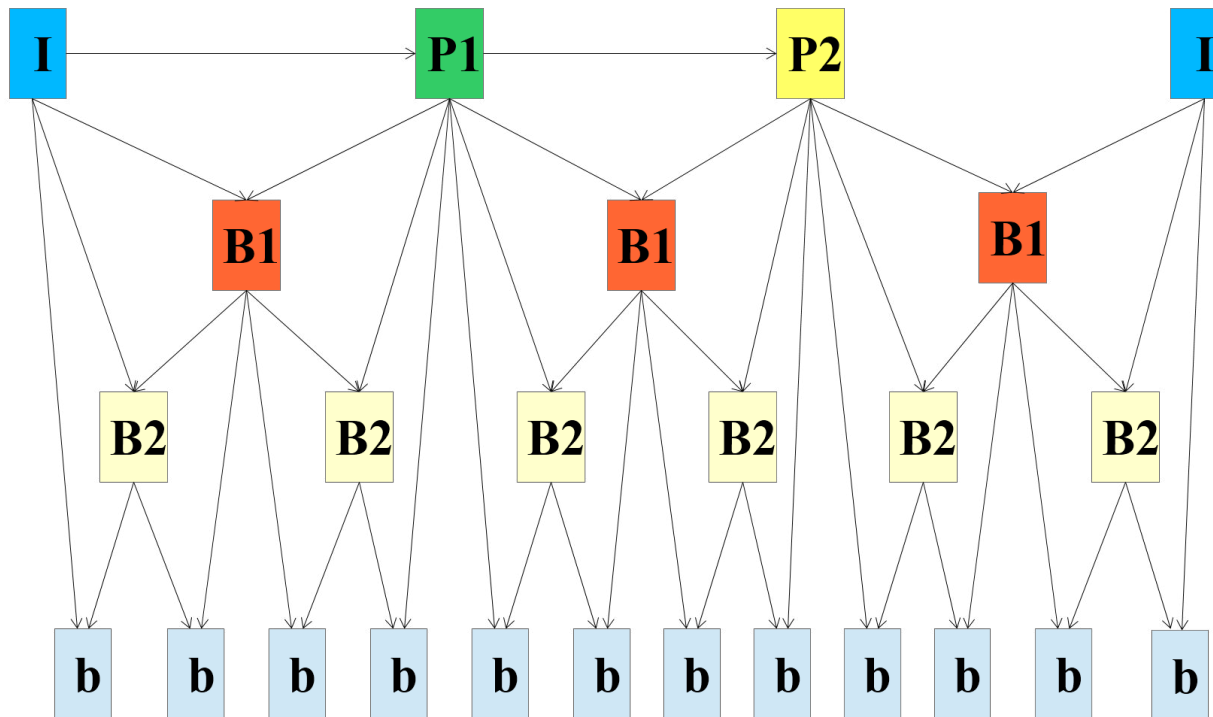
- Depth maps downsampled by 2 in horizontal and vertical directions [5]
- Color and depth encoded separately

Color resolution	1920 x 1080
Frame rate	30 fps
Profile	H. 264 High Profile
GOP structure	Hierarchical
I-frame insertion	Every 24 frames
QP	26 [5]
Codec	JM 18.1/16.2
Packet size	1316 bytes [6]

[5] D. Rusanovskyy, H. Schwarz, “Common Test Conditions for 3DV experimentation”, ISO/IEC JTC1/SC29/WG11 MPEG, N12745 2012.

[6] DSL Forum Technical Report TR-126: “Triple-play Services Quality of Experience (QoE) Requirements”. Dec 2006

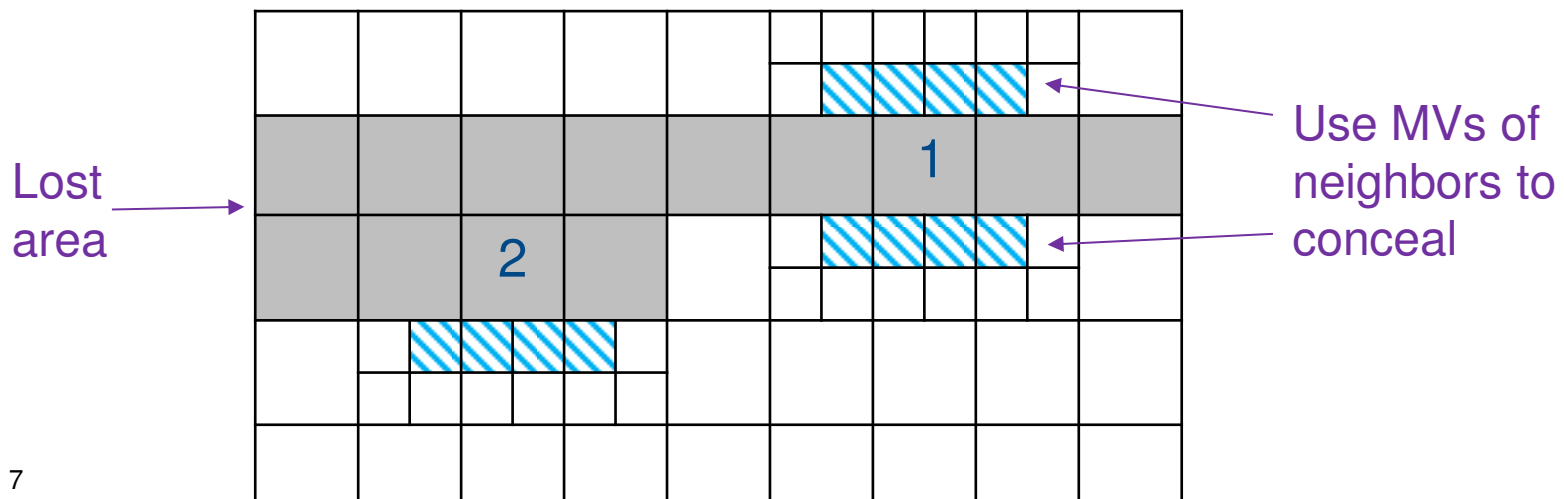
- 4 layers
- Same structure for color and depth



Type	Frames Affected
I	31
P1	23
P2	15
B1	7
B2	3
b	1

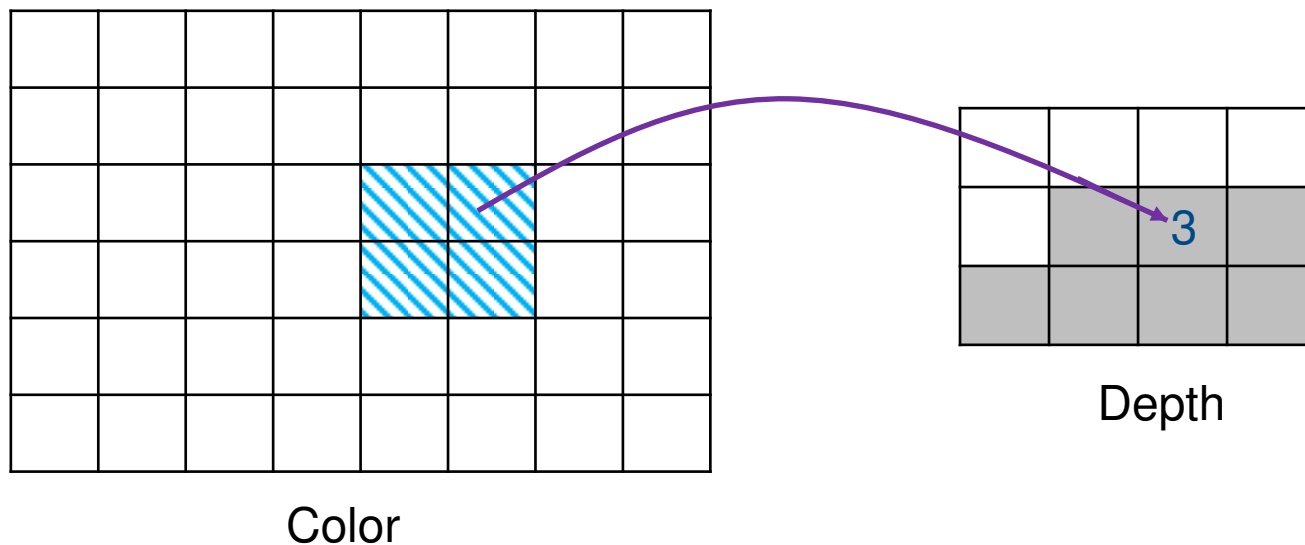
## ❖ Color packet loss

- I slice: spatial interpolation
- P/B slice: Motion-compensated error concealment (MCEC)
  - Extract motion vector(MV) of neighboring available macroblocks(MBs)
  - Find the one that minimizes boundary matching error
  - MV of correctly decoded MBs used if available
  - If not, use MV of concealed MBs
  - All neighboring MBs in intra-mode / whole frame loss:  $MV = 0$



## ❖ Depth packet loss

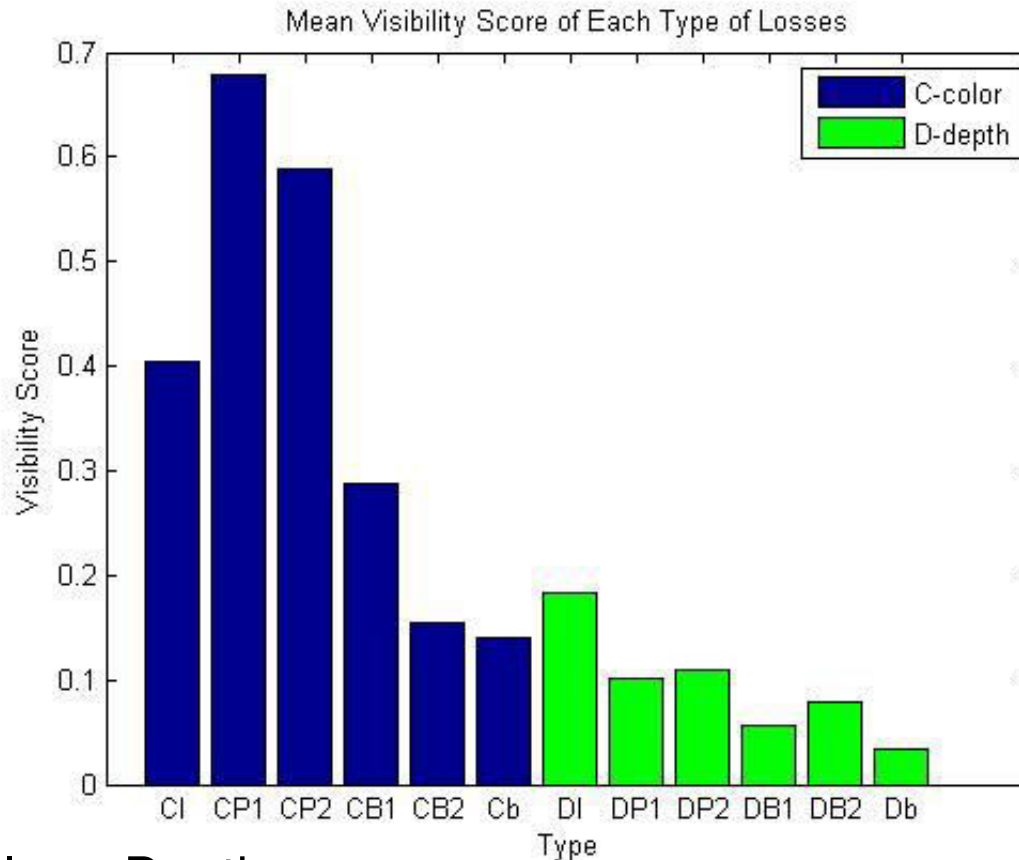
- I slice: spatial interpolation
- P/B slice: concealed using motion vectors of co-located color MBs [7]
  - If co-located color MBs are intra coded:  $MV = 0$



[7] Y. Liu; J. Wang; H. Zhang, "Depth Image-Based Temporal Error Concealment for 3-D Video Transmission," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol.20, no.4, pp.600,604, April 2010



Visibility score = (# viewers who saw the loss) / (# viewers who assessed video)



In general, Color > Depth:

- Color losses affect both views, depth losses affect only right view
- people are probably more sensitive to the blocky artifacts

# Experimental Results

- Color packet loss



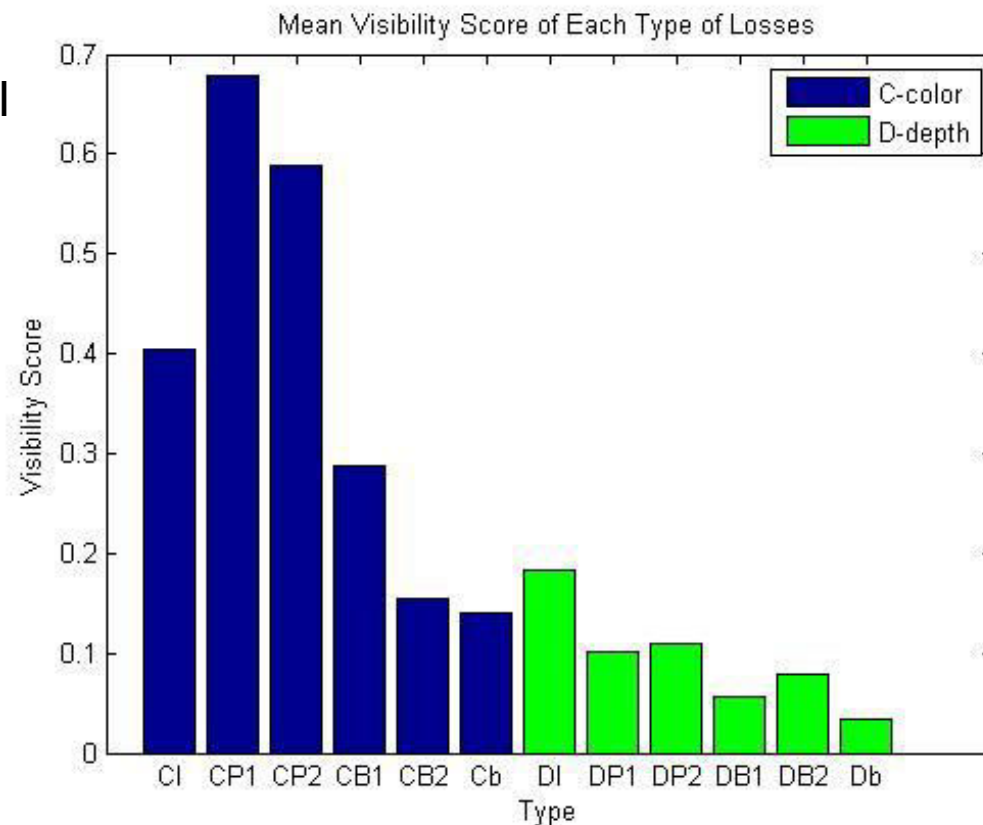
- Depth packet loss



## ❖ CP > CI, CP > CB:

- Time duration: I > P > B
- Spatial area affected: B > P > I

Video	Color	Depth
Avg # Pkt in I frame	50.8	2.3
Avg # Pkt in P frame	39.3	2.0
Avg # Pkt in B frame	20.4	1.5

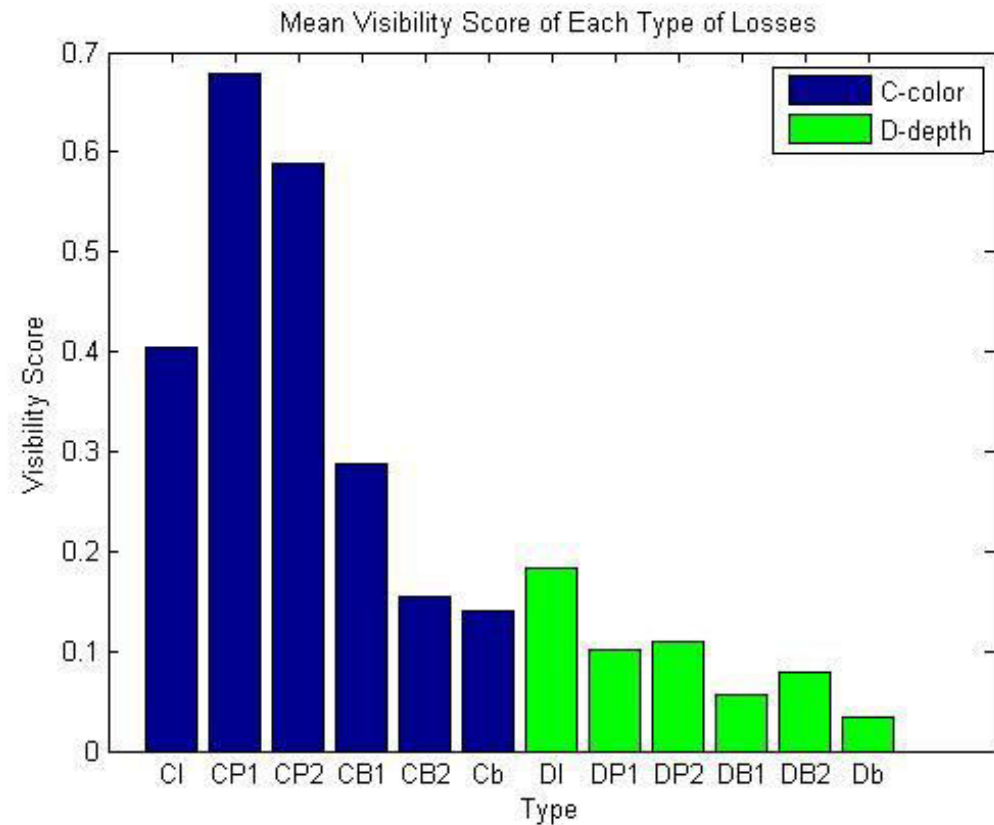


[8] Y.-L. Chang, T.-L. Lin, and P.C. Cosman, “Network-based IP packet loss importance model for H.264 SD videos”, IEEE PacketVideo 2010.

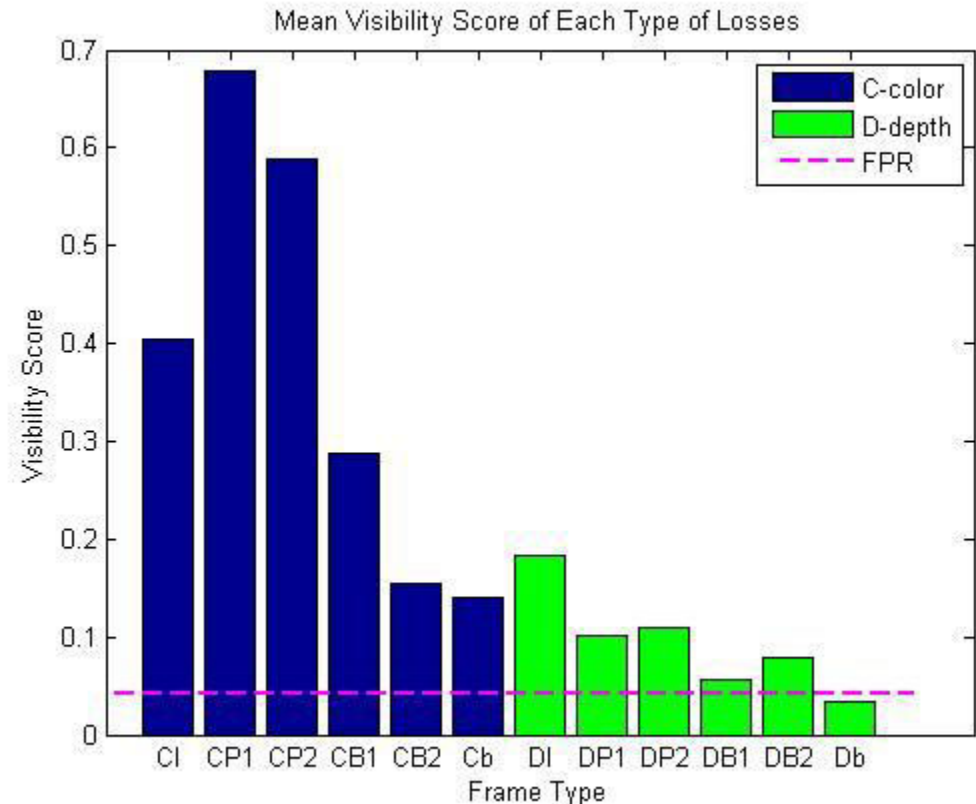
## ❖ $DI > DP$ , $DI > DB$ :

- Inter-coded MBs concealed better than intra-coded MBs
- Time duration of I frame is the longest

Video	Color	Depth
Avg # Pkt in I frame	50.8	2.3
Avg # Pkt in P frame	39.3	2.0
Avg # Pkt in B frame	20.4	1.5



- ❖ False positive rate = 4.17%
- May be due to compression artifacts, view synthesis artifacts, or just inattention
- Well below the mean visibility score of losses in all the color frame types and in depth I and P frames
- But close to the mean visibility score of depth B frames



## ❖ Content independent features:

- IsColor: whether the loss is in a color video (not depth)
- TMDR: time duration of the loss, determined by frame type
- DevFromBorder =  $\text{floor}(N/2) - \text{abs}(\text{Height} - \text{floor}(N/2))$ : deviation from border  
N = number of rows of MBs in a frame. HD: N = 68
- IsCIframe, IsP1frame, IsB1frame, IsD1frame, IsDPframe, IsDBframe: frame type

Type	Time duration
I	31
P1	23
P2	15
B1	7
B2	3
b	1

## ❖ Content dependent features

- NumMB: number of MBs in the affected area
- PktSize: packet size in bytes

Packets < 1316 when

- (1) whole frame is included in that packet: low motion complexity
- (2) the last packet in the frame: spatial location

- ❖ Extract the following features from both color and depth MBs
  - NumIntra, NumInter, Num(Skip/Direct): number of MBs in the affected area coded in intra, inter, skip/direct mode
  - IntraRatio, InterRatio, (Skip/Direct)Ratio =  
$$\frac{\text{Num MB in intra, inter, skip or direct mode}}{\text{Num MB in the whole affected area}}$$
  - MaxInterparts: maximal sub-partitions in MBs
  - Motion vectors: x, y components, magnitude, angle
  - RSENGY: residual energy per pixel after motion compensation



- ❖ IMSE, ISSIM: initial MSE and SSIM per MB within the frame where loss occurs.
  - Compute between compressed error-free video ( $f$ ) and degraded video ( $f'$ )
  - MSE per MB:  $\sum_{i=1}^{16} \sum_{j=1}^{16} (f_{ij} - f'_{ij})^2$ , larger value means worse quality
  - SSIM per MB:  $\frac{(2\mu_f \mu_{f'} + C_1)(2\sigma_{ff'} + C_2)}{(\mu_f^2 + \mu_{f'}^2 + C_1)(\sigma_f^2 + \sigma_{f'}^2 + C_2)}$ , smaller value means worse quality
  - Compute MaxMSE, MeanMSE, VarMSE,  
MinSSIM, MeanSSIM, VarSSIM

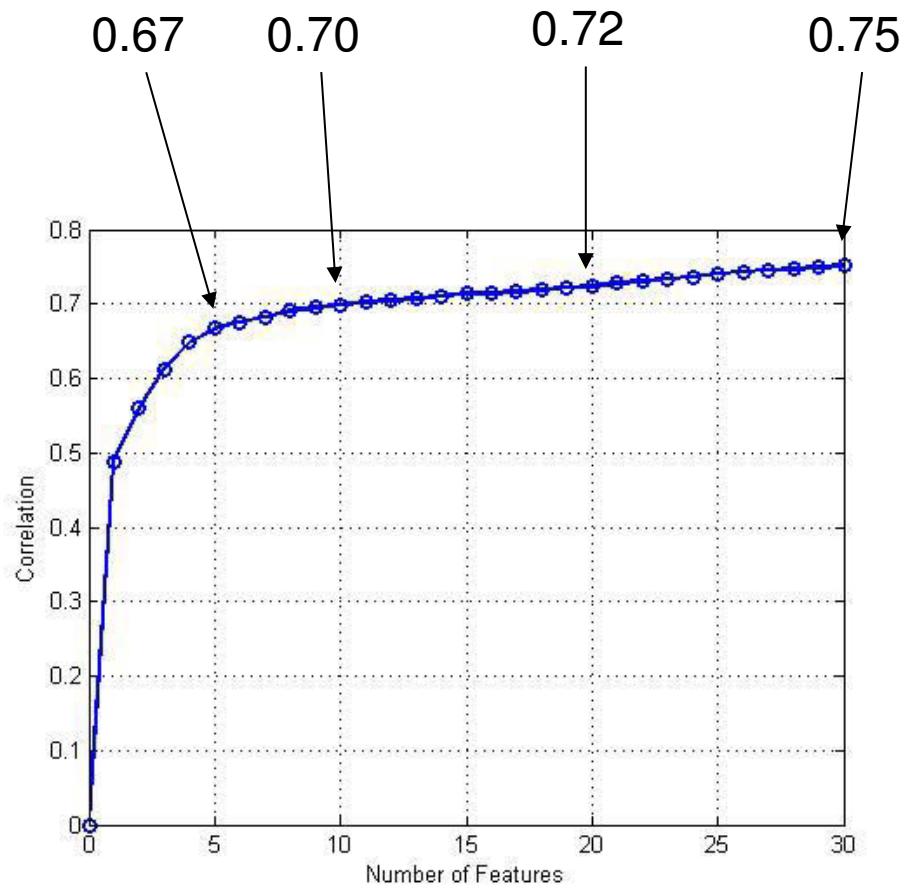
- ❖ Foreground MB: attracts people's attention
  - Extract foreground pixels by setting a threshold to depth:  
depth < threshold
  - A MB classified as foreground MB if over half of pixels are foreground
  
- Features related to foreground MB
  - FGNum: number of foreground MBs in the packet.
  - FGRatio: FGNum divided by the total number of MBs in the packet.

- Generalized linear model with logit as link function

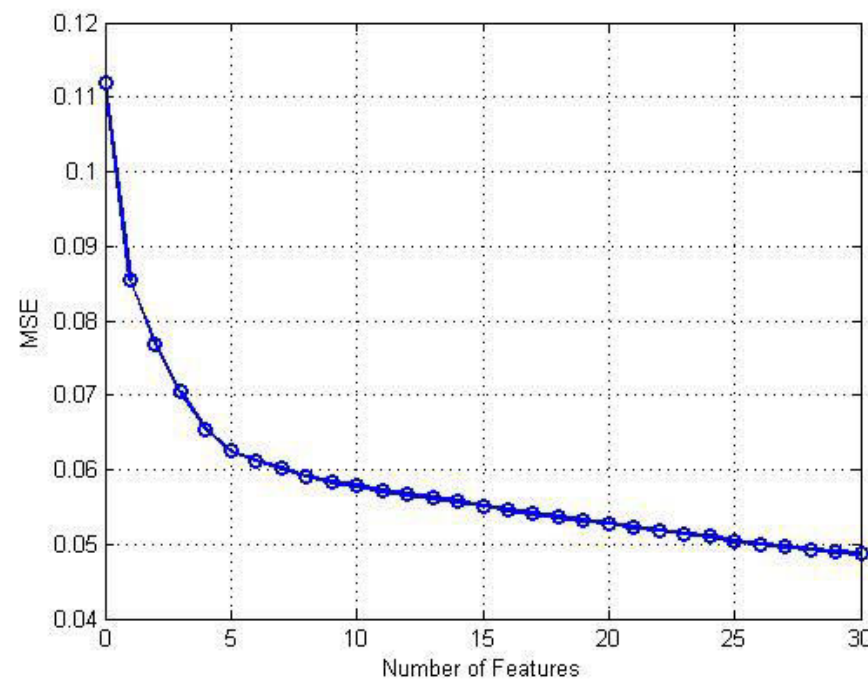
$$\log\left(\frac{p}{1-p}\right) = \gamma + \sum_{j=1}^K x_j \beta_j$$

$p$  – visibility score,  $x_j$  – feature,  $\beta_j$  – coefficient,  $\gamma$  – constant term

- 5-fold cross validation to select efficient features and prevent overfitting
- MSE and correlation coefficient between prediction and ground truth to measure performance



Correlation Coefficient



Mean Squared Error

Ten most important features: (x means interaction)

Feature #	Feature
gamma	1
1	IsColor x CIntraRatio
2	TMDR x MaxMSE
3	IsColor x DevFromBorder
4	IsCBframe x CMaxMotA
5	IsColor x CMaxMotM
6	IsCPframe x CNumIntra
7	D(Skip/Direct)Ratio x MinSSIM
8	PktSize
9	DInterRatio x DVarMotA
10	IsDBframe x DMaxMotX

- ❖ IsColor: the most important feature, has positive coefficients
  - CIntraRatio: more MBs in intra mode, harder to conceal
  - DevFromBorder: larger value, closer to the center of the screen
  - CMaxMotM: larger MotM, higher motion

Feature #	Feature	Coefficient Sign
1	IsColor x CIntraRatio	+
3	IsColor x DevFromBorder	+
5	IsColor x CMaxMotM	+

## ❖ Frame type:

- IsCPframe x CNumIntra: more MBs in intra mode implies the motion is complicated or there is a scene cut. Loss won't be concealed well.
- IsCBframe, IsDBframe, negative coefficients: B frames less visible

Feature #	Feature	Coefficient Sign
4	IsCBframe x CMaxInterparts	-
6	IsCPframe x CNumIntra	+
10	IsDBframe x DMaxMotX	-

## ❖ Objective metrics:

- TMDR x MaxMSE: big distortion lasting longer, more visible
- D(Skip/Direct)Ratio x MinSSIM: smaller SSIM means larger distortion

Feature #	Feature	Coefficient Sign
2	TMDR x MaxMSE	+
7	D(Skip/Direct)Ratio x MinSSIM	-



- Color packet losses are generally more visible than depth packet losses
- Color:  $P > I > B$ 
  - Competing factors of time duration and spatial area
- Depth:  $I > P > B$ 
  - Same 2 competing factors but also different error concealment effects
- Prediction model of visibility score
  - Correlation coefficient = 0.67 with 5 features, 0.7 with 10 features, and 0.75 with 30 features

## ❖ Unequal error protection (UEP)

- UEP for video+depth vs. UEP for MVC
- Test for Rayleigh fading channels

## ❖ Comparing with MVC

- Losses in secondary view have similar visibility of depth map?
- False positive rate caused by view synthesis artifacts?