

## فشرده‌سازی اطلاعات

۰۱-۰۲-۰۰۱۰-۱۰-۱۴۰

بخش هشتم

قسمت اول

مقدمه‌ای بر  
MPEG2



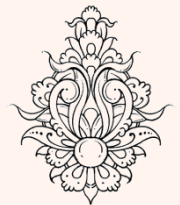
دانشگاه شهید بهشتی  
پژوهشکده‌ی فضای مجازی

بهار ۱۴۰۰

احمد محمودی ازناوه

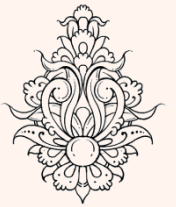
# فهرست مطالب

- پیش‌گفتار
- ادراک حرکت در سیستم بینایی بشر
- فشردده‌سازی ویدئو
- انواع نمونه‌برداری
  - پیش‌رونده
  - درهم تنیده
- افزودنی زمانی و مکانی
- انواع فریم
- تخمین بردار حرکت



# پیش‌گفتار

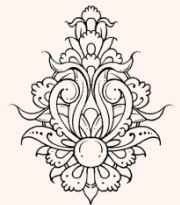
- محیط زندگی ما، پر از دوربین‌های مداربسته است.
- افزون بر این، روزانه تعداد قابل توجهی ویدئو که با دوربین‌های موبایل و تبلت گرفته می‌شوند، در سایت‌های مختلف بارگذاری می‌شوند.
- این دنباله‌های ویدئویی که با اهداف مختلف اخذ شده‌اند، حاوی اطلاعات مفیدی است که باید ذخیره، ارسال و پردازش شوند.
- از این رو، «پردازش سیگنال‌های ویدئویی» اهمیت روزافزونی یافته است.



# برخی کاربردها



- ویدئوکنفرانس
- ارسال ویدئو از طریق شبکه‌های ارتباطی
- - **فشرده‌سازی**، یکی از مهم‌ترین نیازمندی‌های این حوزه است.
- افزایش کیفیت ویدئو
- شناسایی هویت با استفاده از حرکات
- تجزیه و تحلیل حرکات
- تشخیص اعمال انسان (Human Action Recognition)
- نهان‌نگاری و واترمارکینگ ویدئو
- تشخیص جعل در دنباله‌های ویدئو



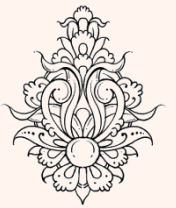
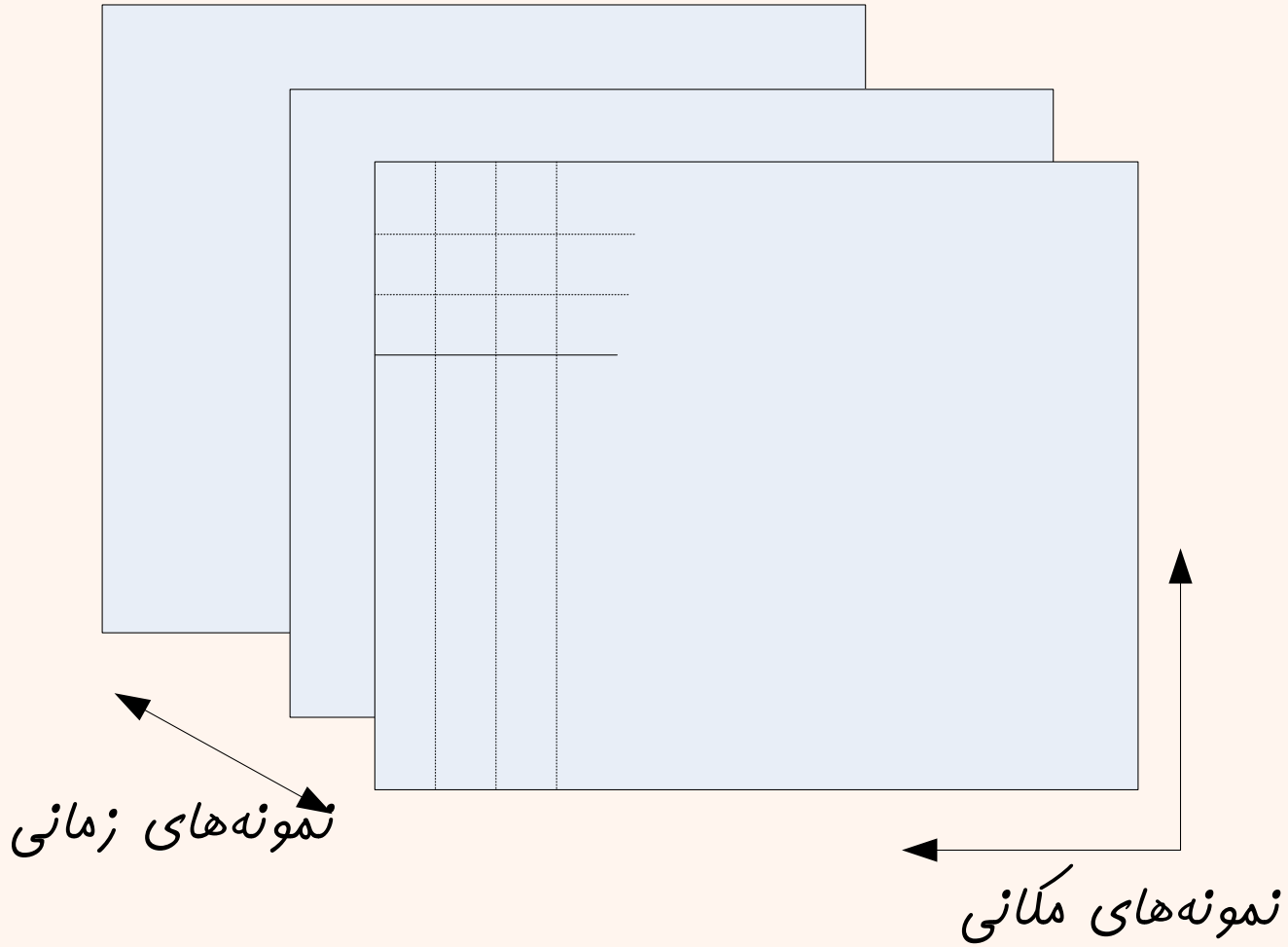
# معرفی

- یک صحنه‌ی طبیعی ویدئویی متشکل از تعداد بی‌شماری شیء همراه با خصوصیات ظاهری مانند عمق، طرح و قالب، بافت و روشنایی می‌باشد.
- خصوصیات یک صحنه ویدئویی:
  - شامل خواص زمانی
  - خصیصه‌های وابسته به مکان



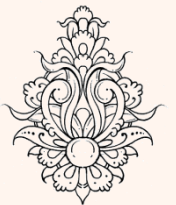
# شمای کلی

نمونه‌های مکانی و زمانی یک جریان ویدئویی



# نمونه برداری زمانی

- نمونه برداری زمانی همان ذخیره نمودن تصاویر لحظه‌ای در فواصل زمانی مشخص می‌باشد.
  - نمایش این تصاویر به صورت پی‌درپی حرکت اشیا را نشان می‌دهد.
  - نرخ نمونه برداری بسیار اهمیت دارد. هر چه نرخ مذکور بالاتر باشد، حرکت نرم‌تر و به حالت طبیعی نزدیک‌تر است.
  - برای سیستم تلویزیونی ۲۵ تا ۳۰ فریم در ثانیه می‌باشد.



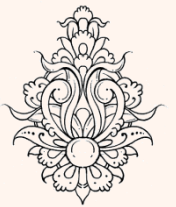
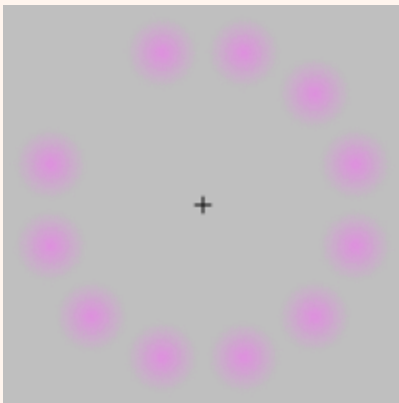
# Motion perception

- هر تصویر حدود 25ms بر روی شبکه ثابت می ماند.

Persistence of vision



- دو شی که دارای تخییر سریع هستند، در سیستم بینایی بشر به صورت یک حرکت پیوسته ادراک می شود.

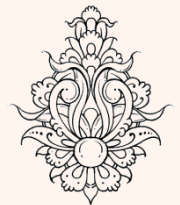
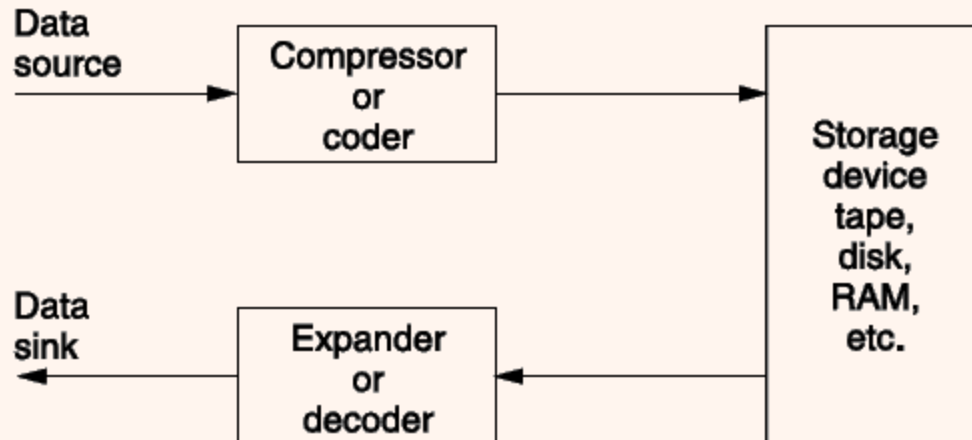




# اطلاعات حجیم

- برخلاف تصویر حجیم اطلاعات در رسانه‌ی ویدئو بسیار بالاست.
- از این رو معمولاً کدینگ بدون اتلاف در این مقوله کاربرد ندارد.

**Too Much Information !!**



# انواع نمونه برداری

## Progressive Sampling

### • پیش رونده

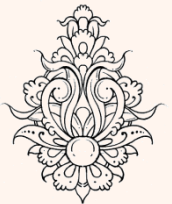
- هنگامی که یک سیگنال ویدئویی به صورت سری کاملی از فریم‌های کامل نمونه برداری گردد، نمونه برداری را پیش رونده می نامند.

## Interlaced Sampling

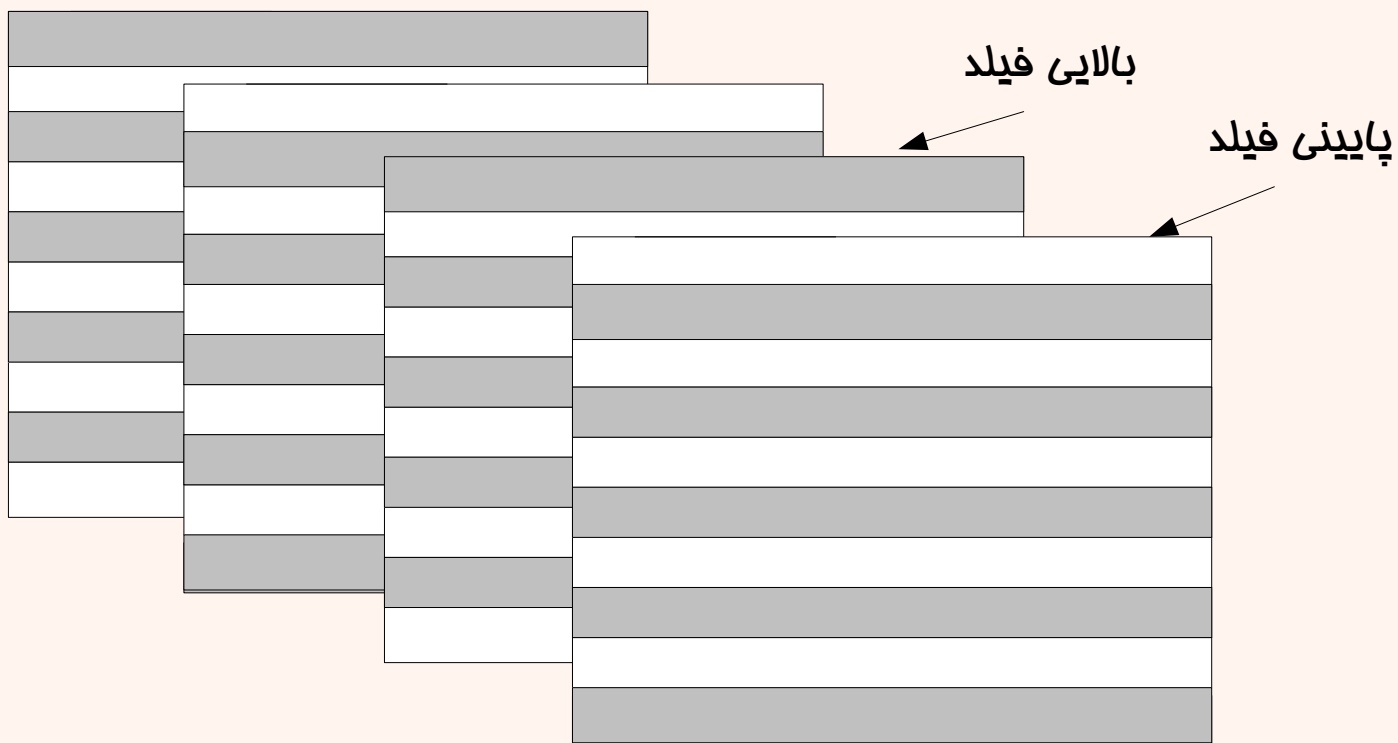
### • نمونه برداری درهم تنیده

- در هر فریم نیمی از داده‌ها (یک نامیه) نمونه برداری صورت می پذیرد. یک نامیه متشکل از تعداد زوج و یا فرد از خطوط داده‌های هر فریم است. در روش نمونه برداری درهم تنیده، در هر بازه زمانی، نیمی از داده‌ها (نامیه بالایی و یا نامیه پایینی) جهت نمونه برداری در نظر گرفته می شوند.

• مزیت این روش به گونه‌ای است که در این شیوه با نرخ بیت یکسان نسبت به روش پیش رونده، قابلیت ارسال داده دوبرابر از هر نامیه وجود دارد.



# جریان داده‌ای ویدئویی همراه با درهم‌تنیدگی



# فضا رنگ

$$Y = K_r R + K_g G + K_b B$$

$$C_b = B - Y$$

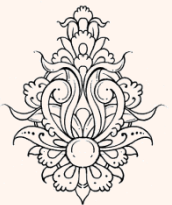
$$C_r = R - Y$$

$$C_g = G - Y$$

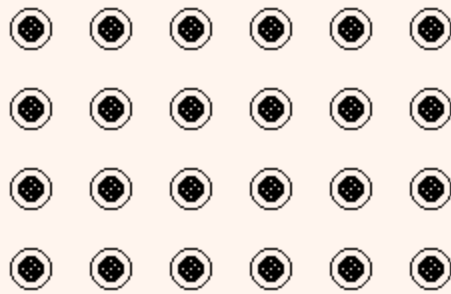
$$Y' C_b C_r \bullet$$

## معادلات تفاضلی رنگ یا مولفه‌های تفاوت کرومینانس

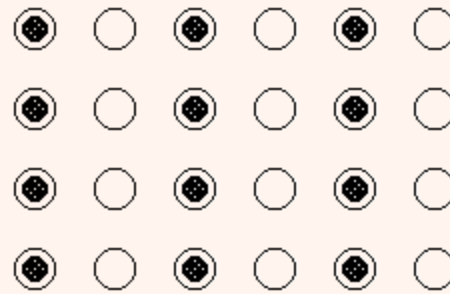
- از آنجا که  $C_b + C_r + C_g$  مقداری متناسب با  $Y$  دارد، ارسال دو مقدار از سه مؤلفه کروما کفایت می‌نماید و مؤلفه سوم با داشتن دومؤلفه دیگر محاسبه می‌گردد.
- در سیستم  $YC_b C_r$  مولفه روشنایی و دو مؤلفه رنگی ارسال می‌گردد. مزیت عمده سیستم مذکور قابلیت ارسال دو مؤلفه کروما با دقت که‌تر از مؤلفه روشنایی است زیرا سیستم بینایی انسان نسبت به روشنایی حساسیت بیشتری دارد.



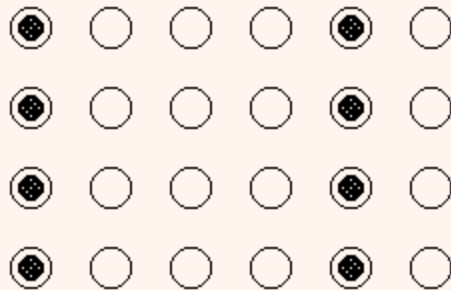
# ساختار سیستم نمونه برداری



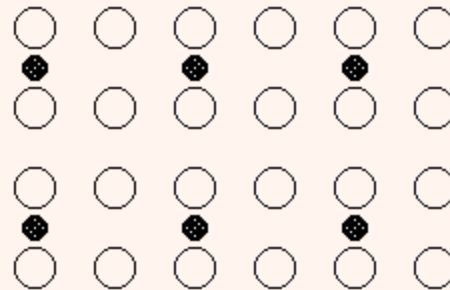
4:4:4



4:2:2

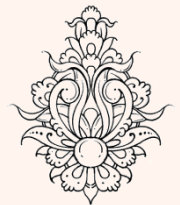


4:1:1

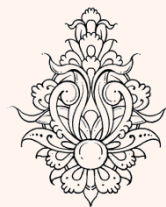
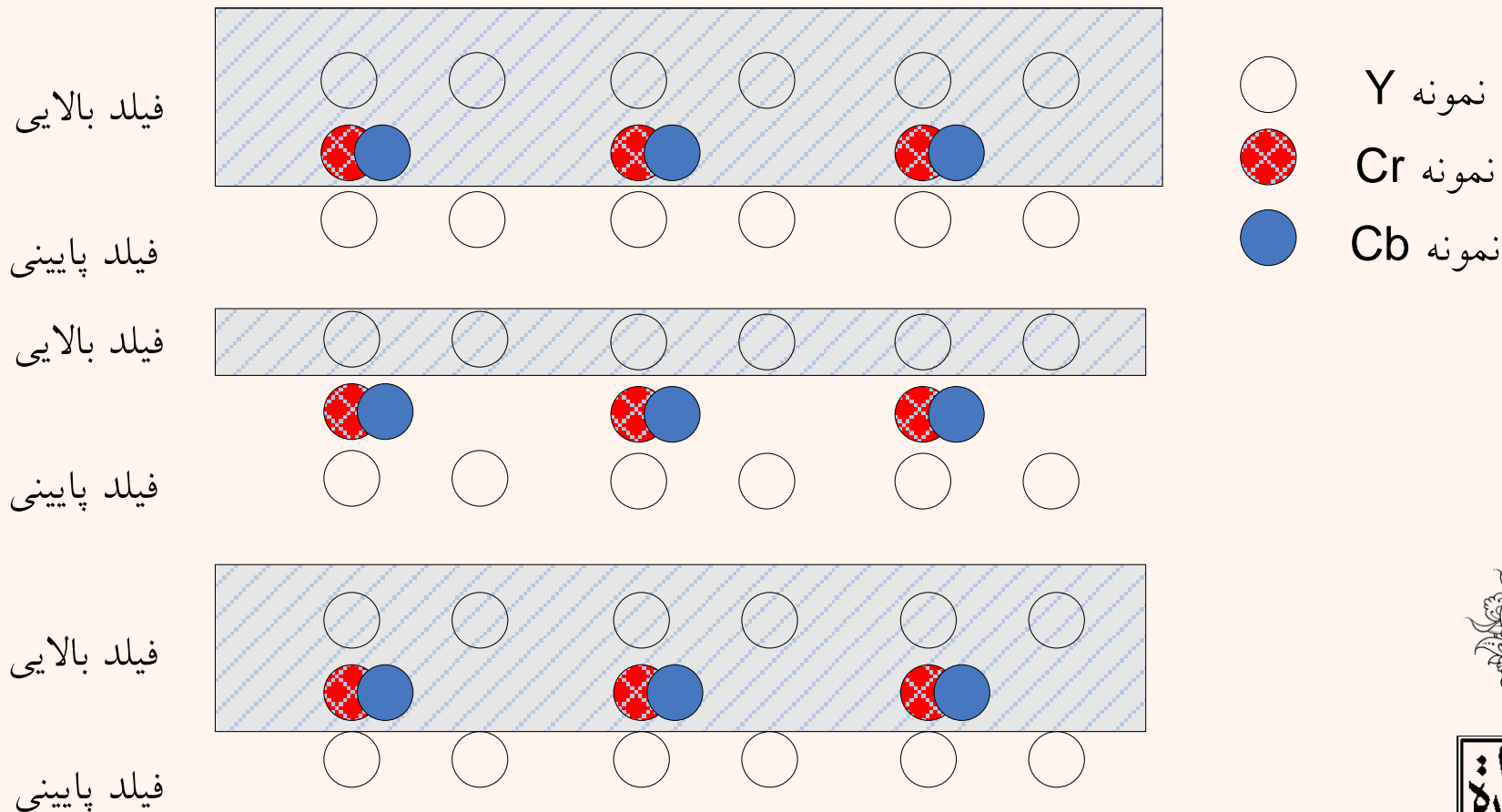


4:2:0

- -- Pixel with only Y value
- -- Pixel with only Cr and Cb values
- ⊙ -- Pixel with Y, Cr and Cb values

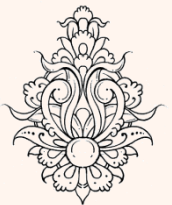


# نمونه برداری (۴:۲:۰) در ویدئوهای در هم تنیده



# ساختارهای ویدئویی

- ساختار CIF پایه‌ای برای بسیاری از قالب‌های فایل‌های ویدئویی می‌باشد. بسته به برنامه کاربردی مورد نظر و همچنین ظرفیت کانال مورد استفاده جهت ارسال فایل ویدئویی از ساختارهای متفاوت استفاده می‌شود.
- برای مثال 4CIF برای سیستم استاندارد تلویزیونی و DVD کاربرد دارد و در همین راستا برای کاربردهایی با درجه‌ی دقت پایین‌تر چون ویدئوکنفرانس استاندارد CIF و یا QCIF به کار می‌رود.





4CIF

**704×576**

CIF

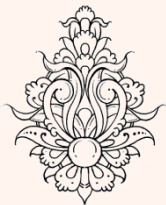
**352×288**

QCIF

**176×144**

SOCIF

**128×96**



تراشه‌نگاره  
به‌پشتی



**(Source Input Format)**

*is practically identical to CIF, but taken from MPEG-1 rather than ITU standards*

Format	Video Resolution
SQCIF	128 × 96
QCIF	176 × 144
SCIF	256 × 192
SIF(525)	352 × 240
CIF/SIF(625)	352 × 288
4SIF(525)	704 × 480
4CIF/4SIF(625)	704 × 576
16CIF	1408 × 1152
DCIF	528 × 384

**Sub Quarter CIF, sometimes subQCIF**

**Quarter CIF**

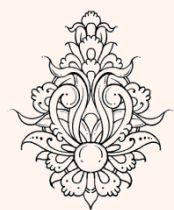
**("NTSC") based systems**

**("PAL") based systems**

**video conferencing systems**

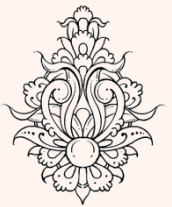
**Double CIF**

**common CCTV equipment**

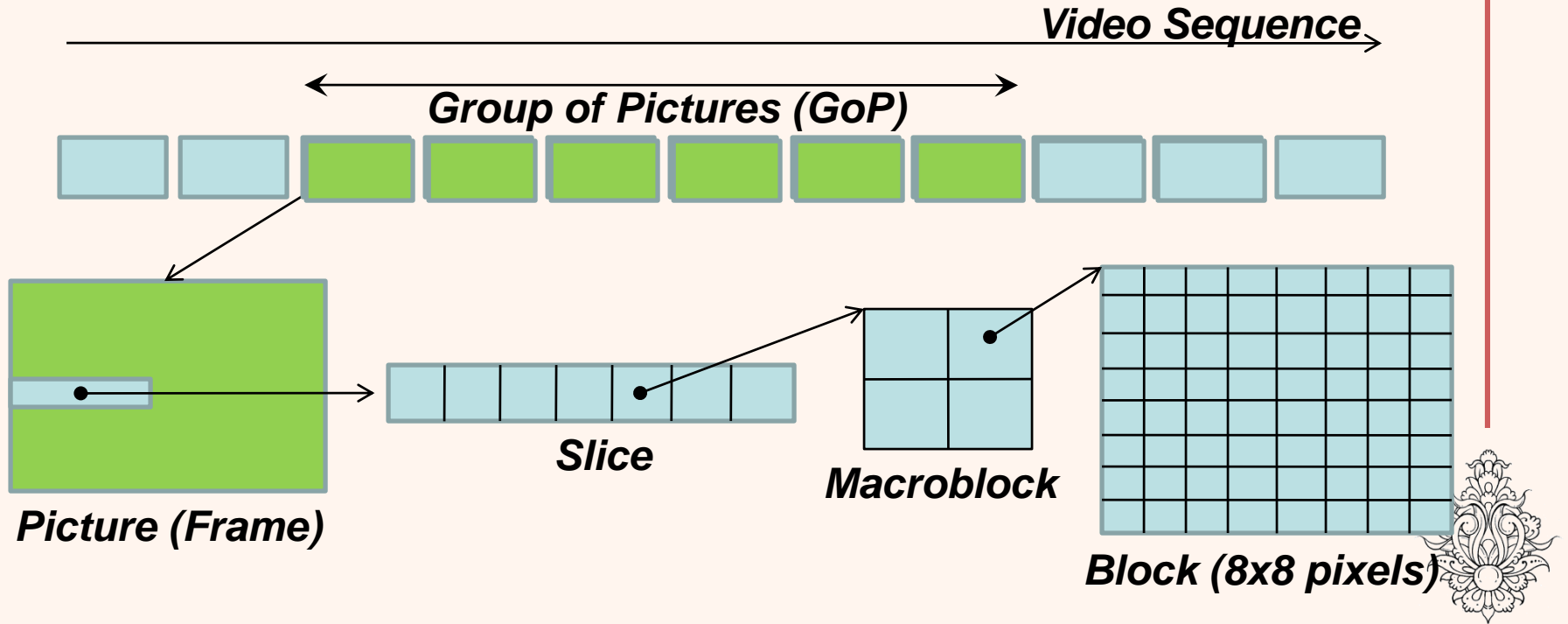


# مفهوم کدینگ در ویدئو

- فشردگی‌سازی فرایندی را مطرح می‌نماید که در آن داده به تعداد بیت‌های کم‌تر فشرده می‌گردد.
- بسیاری از روش‌های کدینگ ویدئو از افزودگی زمانی و مکانی داده‌ها در یک ساختار ویدئویی سود می‌جویند. در حوزه زمان در بیشتر مواقع نرخ این شباهت میان فریم‌های متوالی بسیار بالاست.
- فریم‌های متوالی که به لحاظ زمانی‌های متوالی که به لحاظ زمانی مرتب شده‌اند، به ویژه در مواردی که نرخ نمونه‌برداری بالا باشد، بسیار شبیه به هم می‌باشند.
- در حوزه مکان، هنگامی که یک فریم در نظر گرفته می‌شود، این شباهت در پیکسل‌های مجاور هم به چشم می‌خورد.



# ساختار اصلی



# مفهوم فشردگی در ویدئو

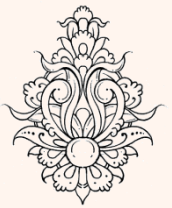


فریم ۱



فریم ۲

*Spatial model*



# مفهوم فشردگی در ویدئو

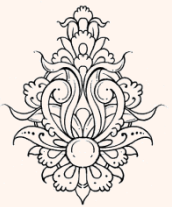


فریم ۱



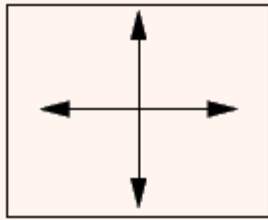
فریم ۲

*Temporal model*



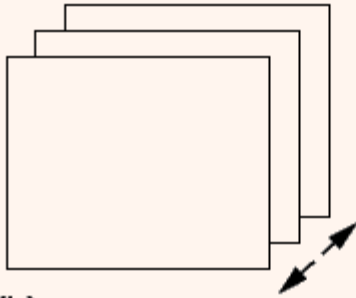
# کدگذار ویدئو

- مدل زمانی
- مدل مکانی
- کدگذار آنتروپی



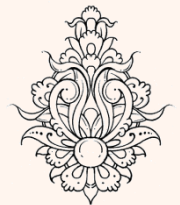
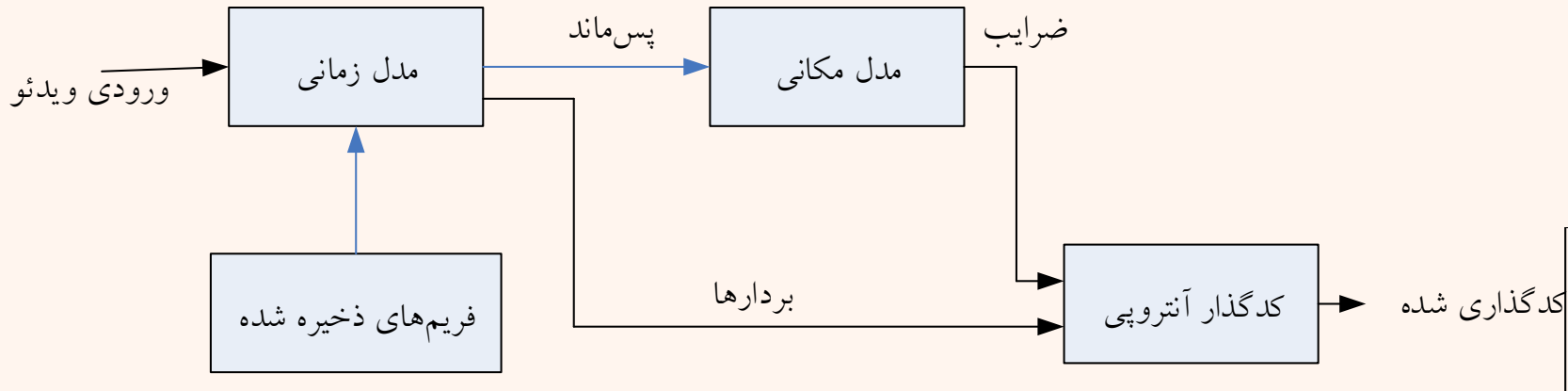
(a)

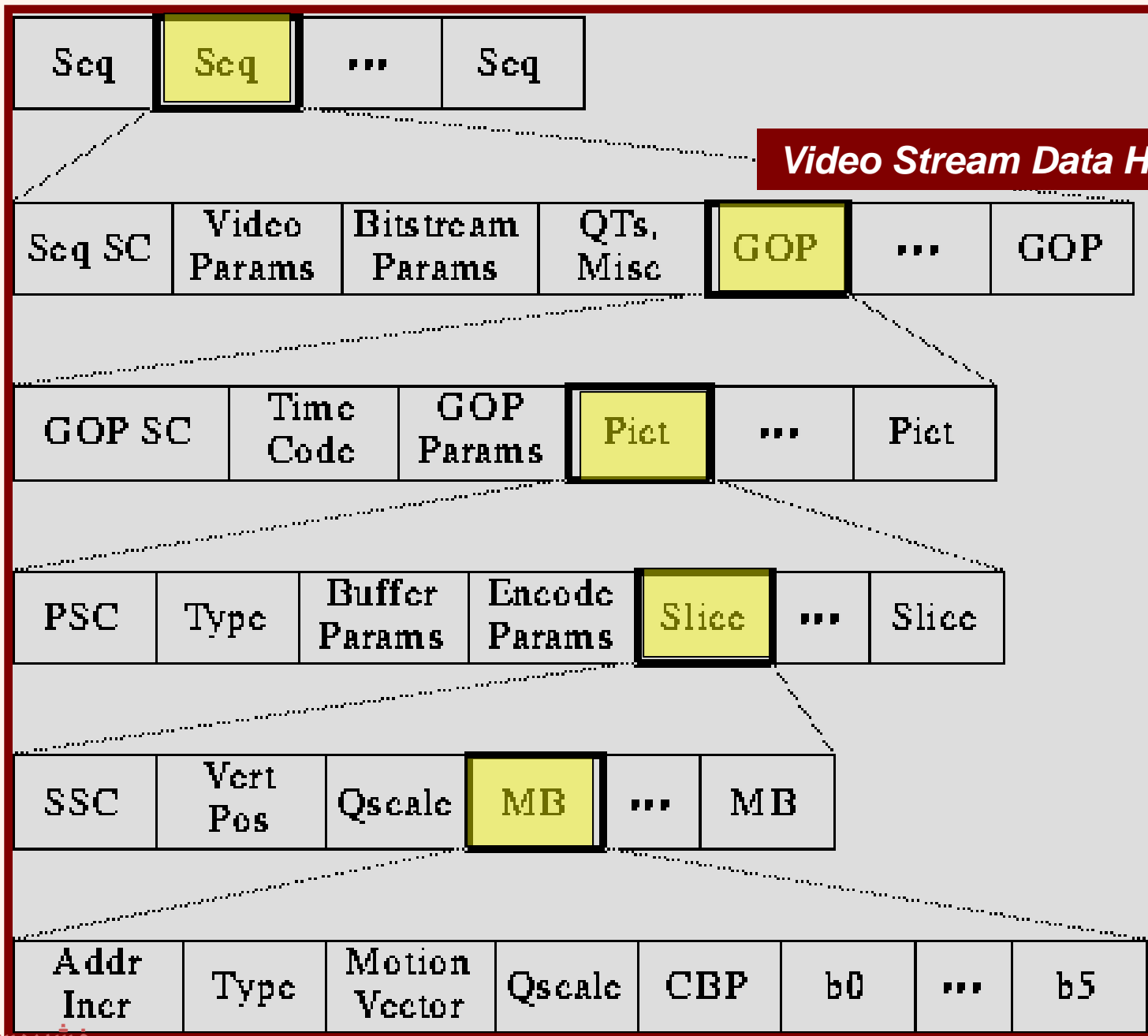
*Spatial or intra-coding explores redundancy within a picture*



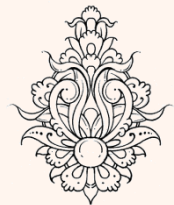
(b)

*Temporal or inter-coding explores redundancy between pictures*





**Video Stream Data Hierarchy**



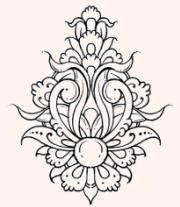
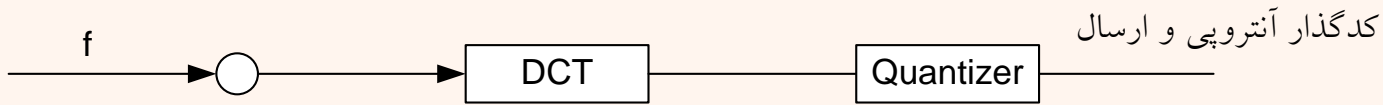
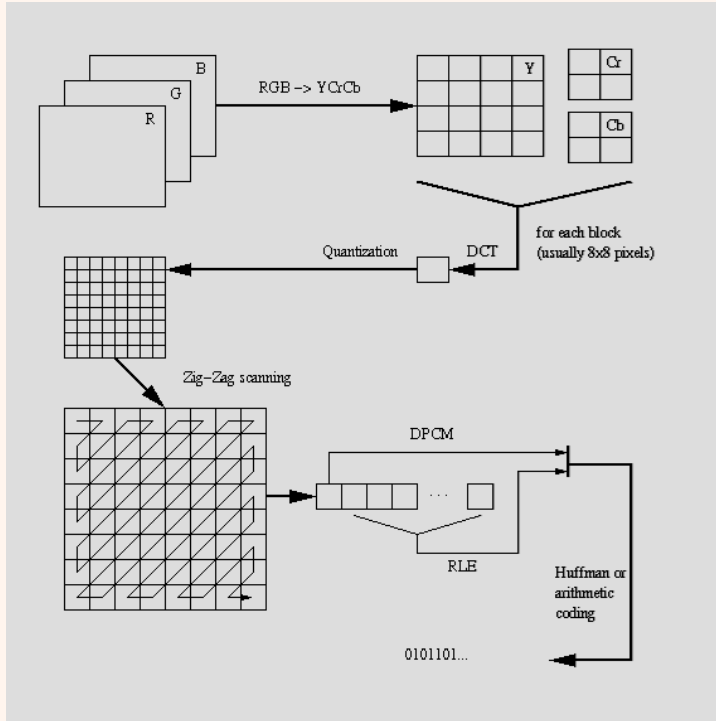
تراشگاه  
سپهر  
بهشتی

# ارسال اطلاعات یک فریم

Intra Frame

• یک فریم

$$f(n_1, n_2)$$



• چند فریم متوالی

Inter Frame

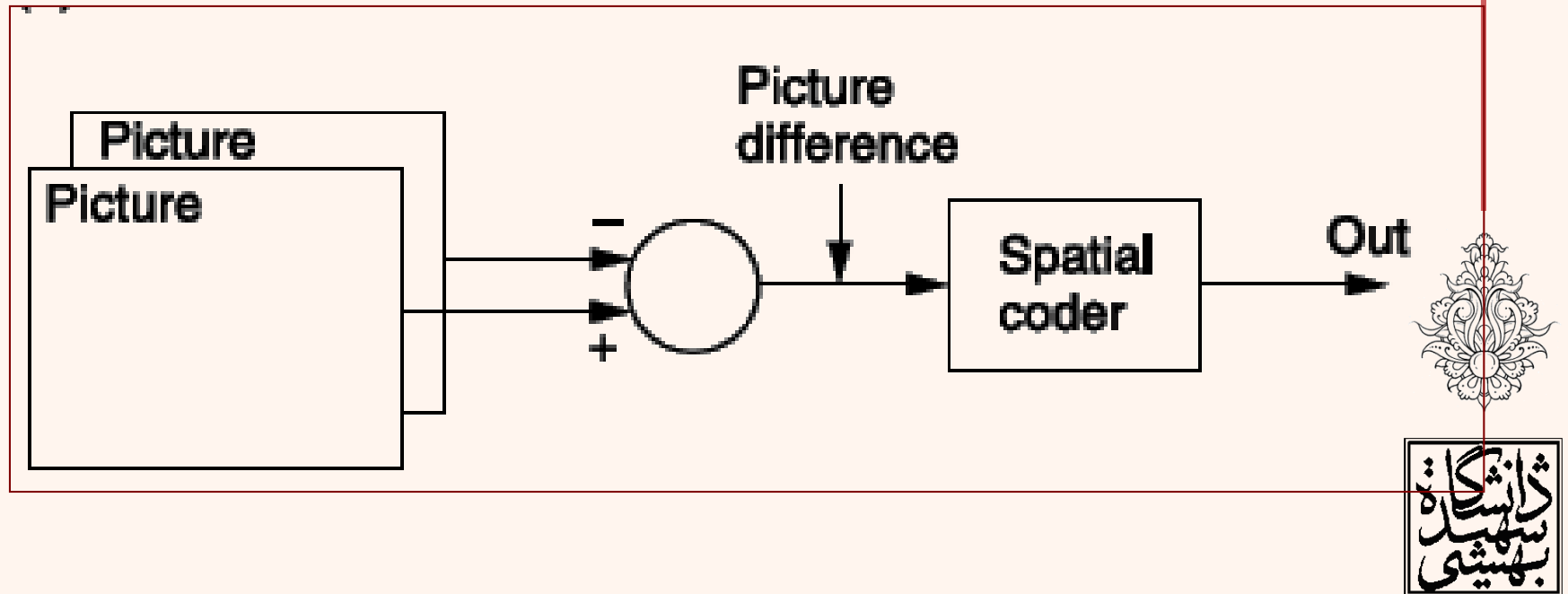
$$f(n_1, n_2, n_3)$$





# ارسال اطلاعات دو فریم متوالی

- ساده ترین روش این است که دو فریم متوالی را از یکدیگر کم کرده اختلاف را برای کد کردن ارسال کنیم.

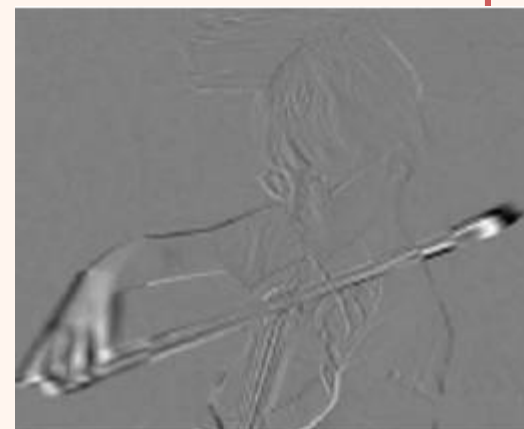




**Frame 1**



**Frame 2**



**Difference**

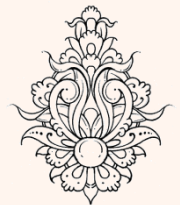
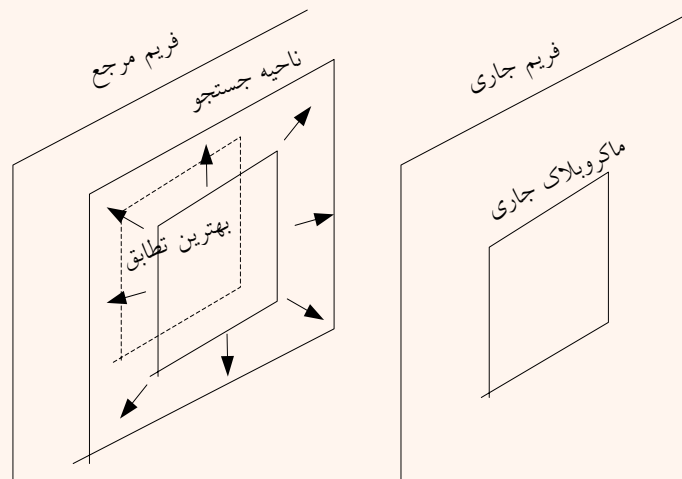
پنجاه و هشت

# تخمین حرکت (برپایه‌ی بلوک)

## Motion Estimation

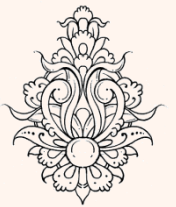
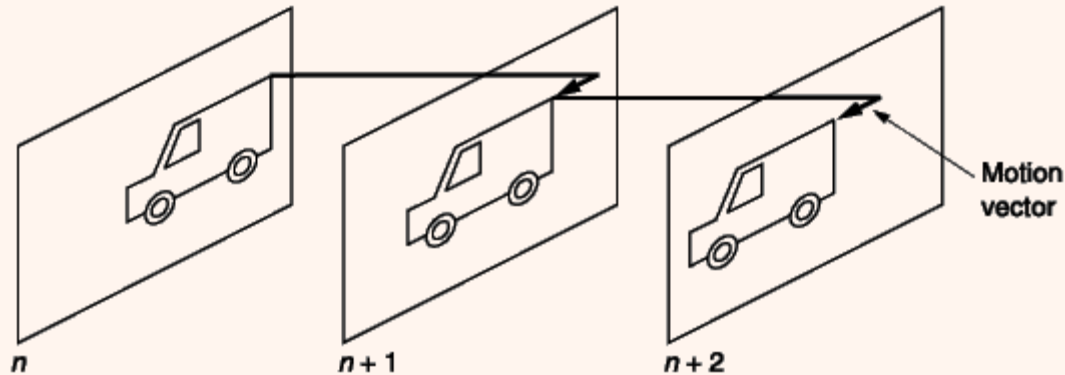
- «تخمین حرکت» یک ماکروبلای یافتن نامیه‌ای  $16 \times 16$  در فریم مرجع است که است که بیشترین میزان شباهت را با فریم جاری دارا باشد.
- فریم مرجع، فریم قبلی کدگذاری شده است که می‌تواند به لحاظ ترتیب نمایش یکی از فریم‌های پسین یا پیشین را شامل گردد. به بهترین گزینه انتخابی «بهترین انطباق» گفته می‌شود.

## Best Match

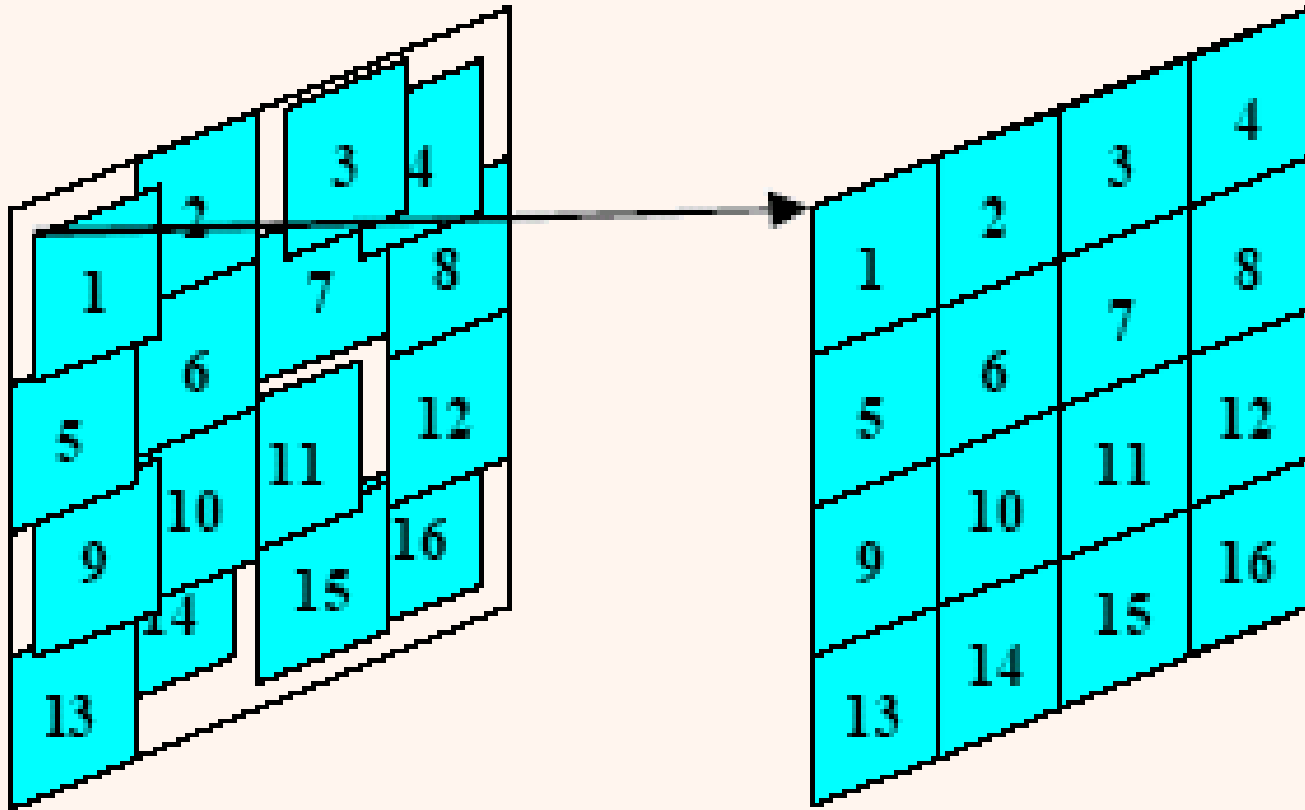


# جبران حرکت (ادامه...)

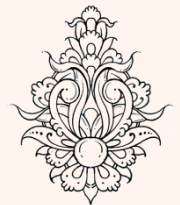
- روشی است که به واسطه‌ی آن میزان افزونگی داده در فریم‌های متوالی را به کمینه‌ی آن برسانیم.
- به واسطه‌ی این فرآیند داده‌هایی از نوع بردار حرکت و خطای ناشی از تخمین خواهیم داشت که می‌باید ارسال گردد.



# جبران حرکت (ادامه...)



<http://www.cs.ucf.edu/courses/cap5015/lectures.html>



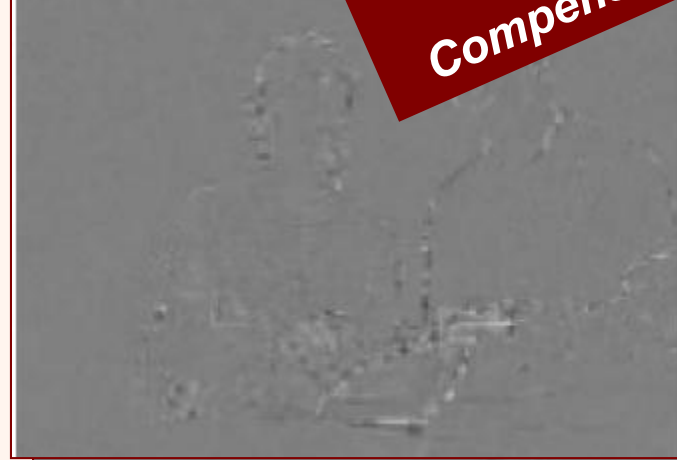
# اندازه‌ی بلوک‌ها



Frame 1



Frame 2



4x4 Difference Motion Compensation



8x8 Difference Motion Compensation

Difference No Motion Compensation

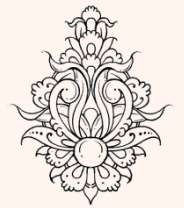
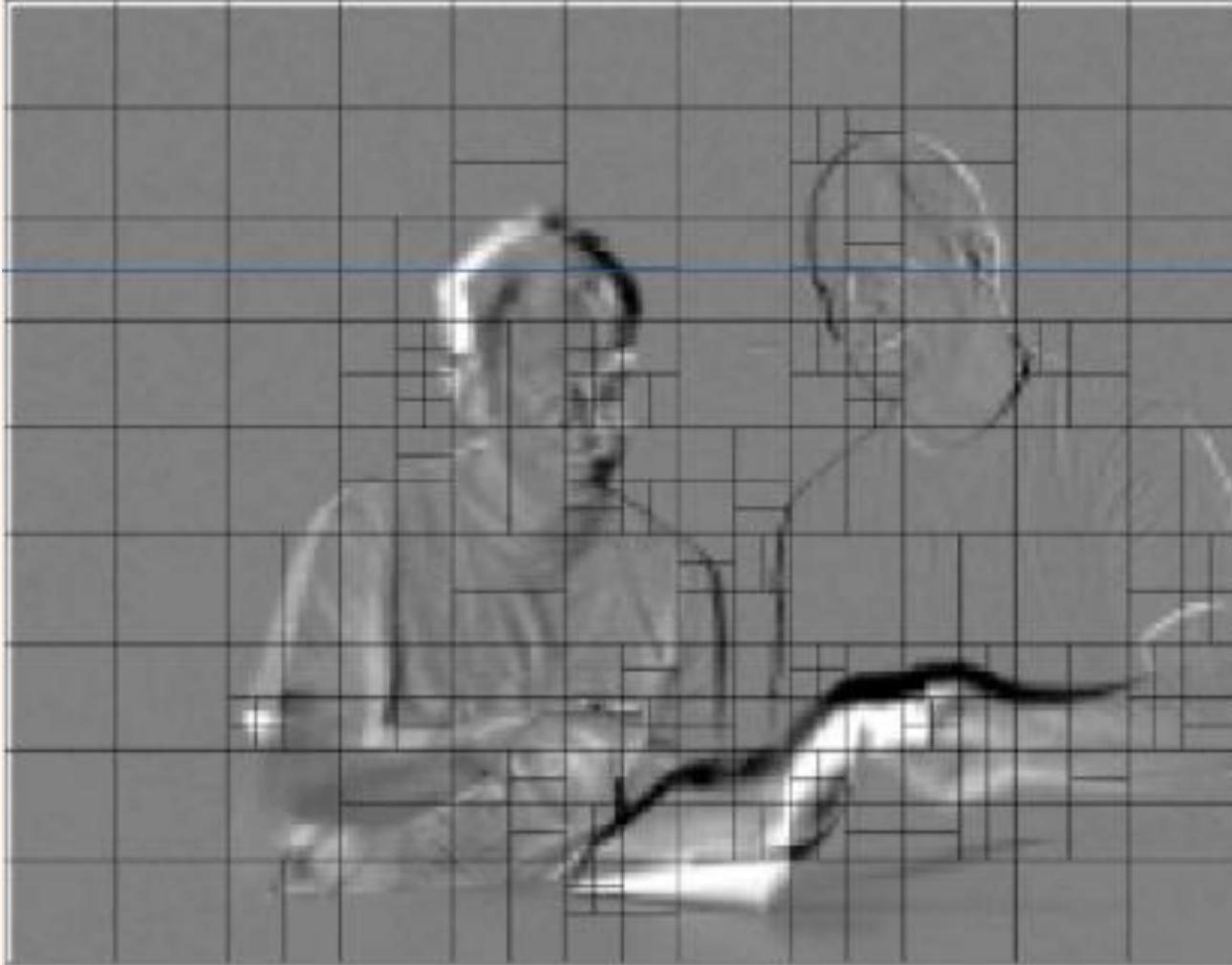


16x16 Difference Motion Compensation



سپهر

# اندازه‌ی بلوک‌ها



# Picture Types

- Intra

کمترین میزان فشرده‌سازی

همان فریم نوع I است که به صورت مستقل کد می‌گردد.

- Predicted

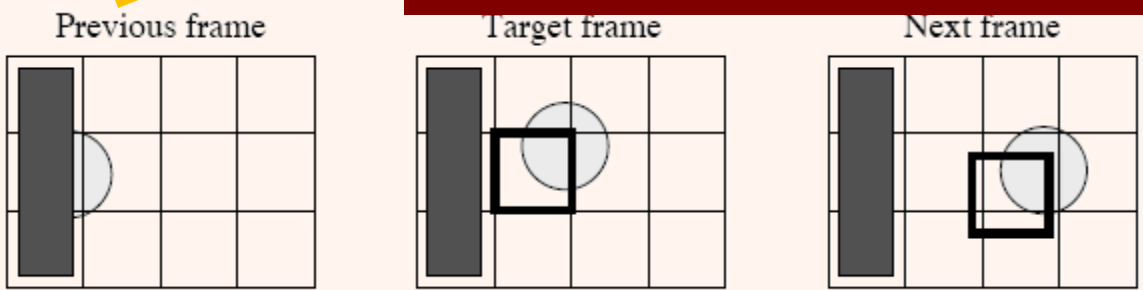
فشرده‌سازی میانه

همان فریم نوع P است که با توجه به نزدیک‌ترین فریم نوع I یا P کد می‌شود. فریم‌های نوع P می‌توانند فقط در کل جریان داده در یک GOP انتقال دهند.

- Bidirectional

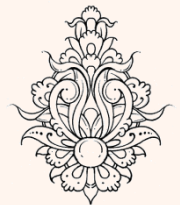
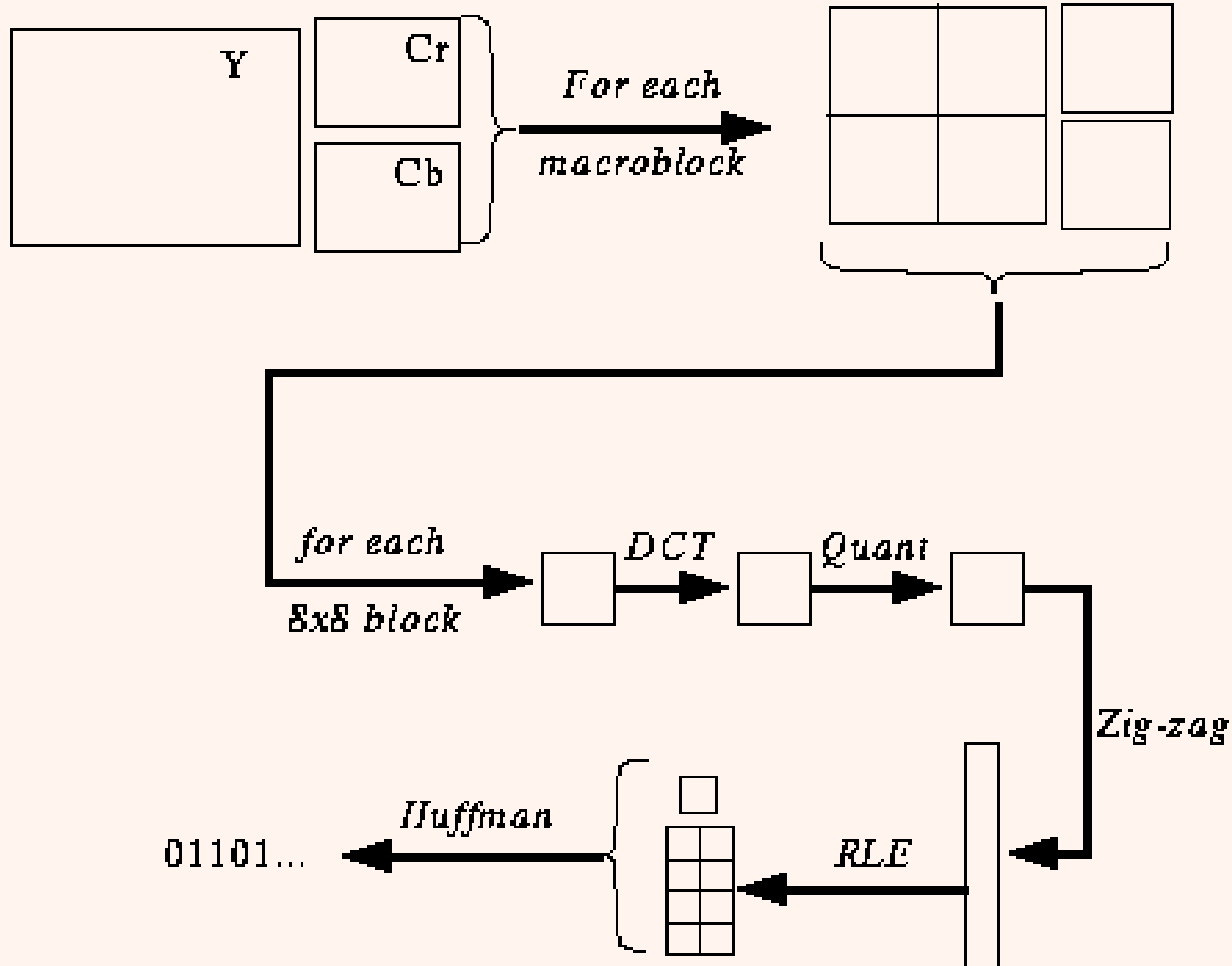
بیشترین میزان فشرده‌سازی

همان فریم نوع B است که با توجه به دو فریم دیگر کد می‌شود. (قبلی و بعدی) چون هیچ گاه مبدا واقع نمی‌شود فقط را منتقل نمی‌کند. به دلیل استفاده از میانگین دو فریم عموماً اثر نویز را هم کاهش می‌دهد.

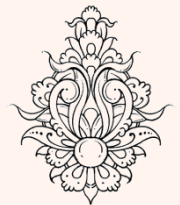
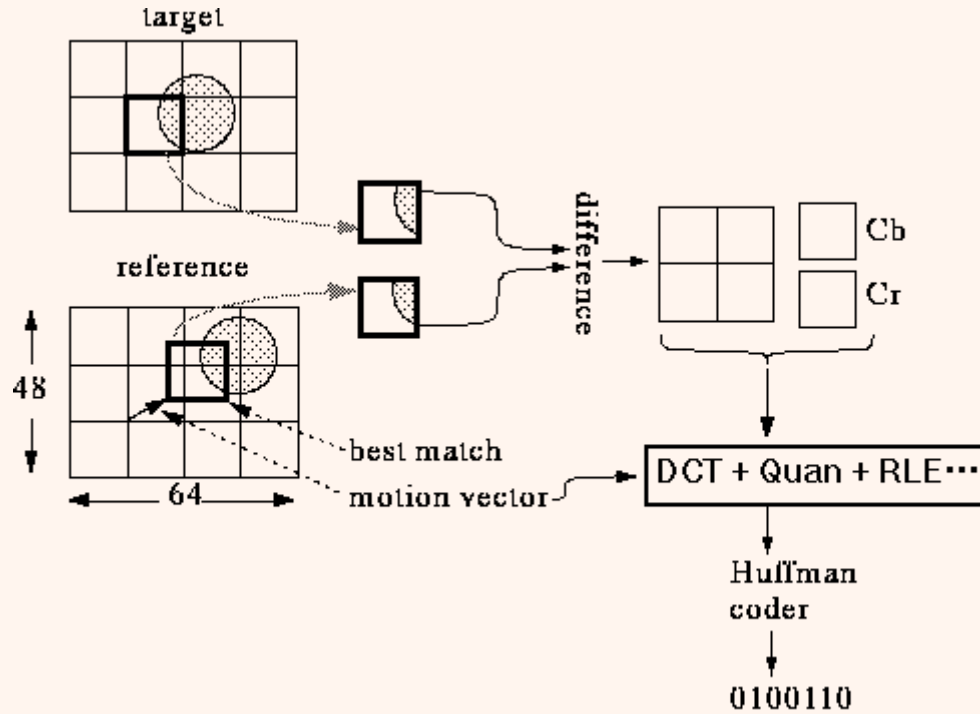




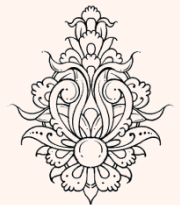
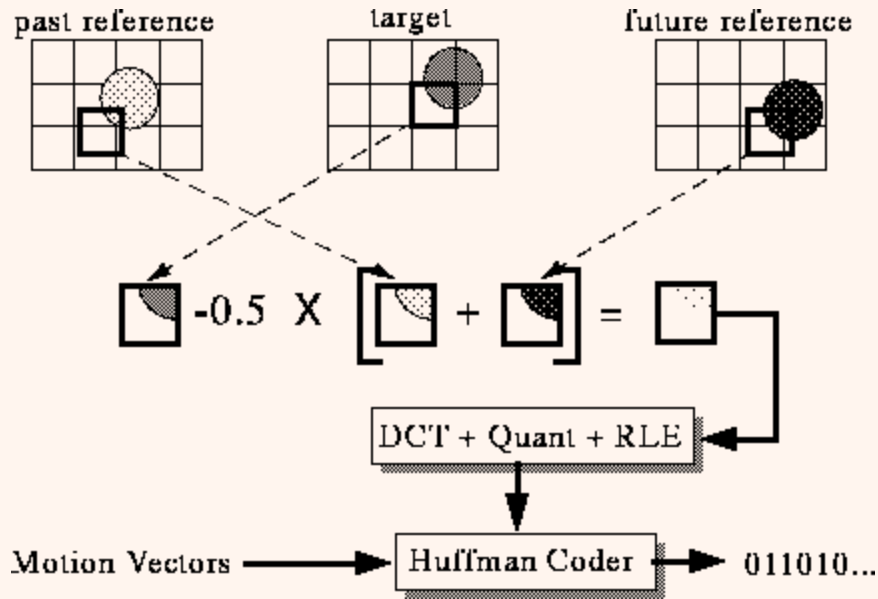
# ساختار کلی ارسال فریم ۱



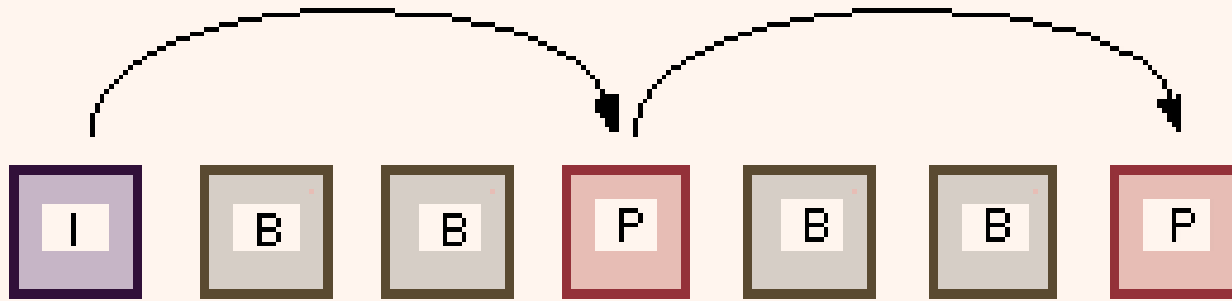
# ساختار کلی ارسال فریم P



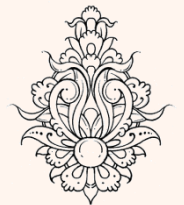
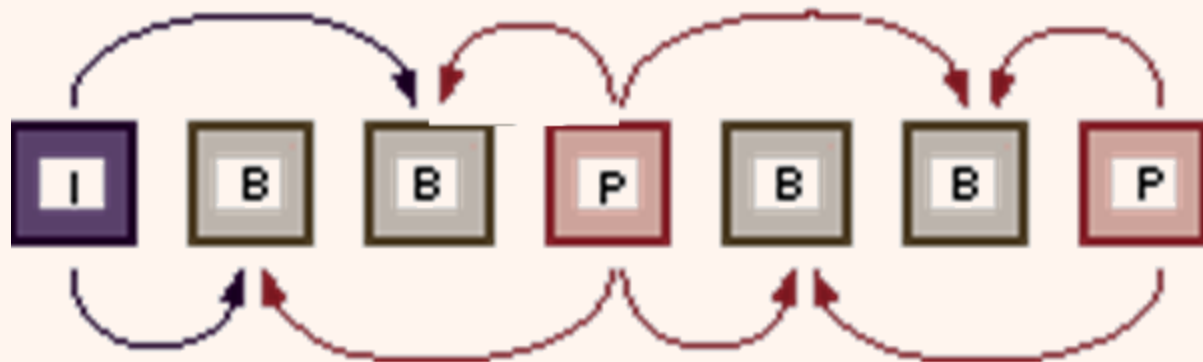
# ساختار کلی ارسال B



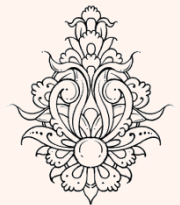
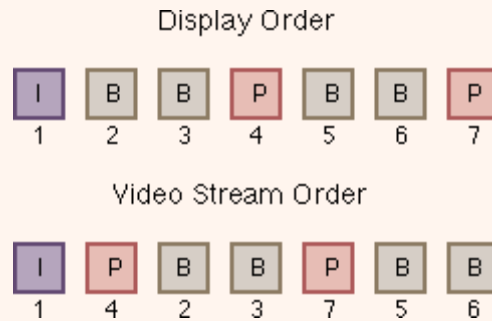
## Forward Prediction



## Bidirectional Prediction

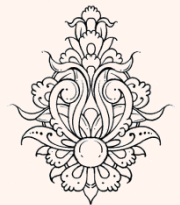
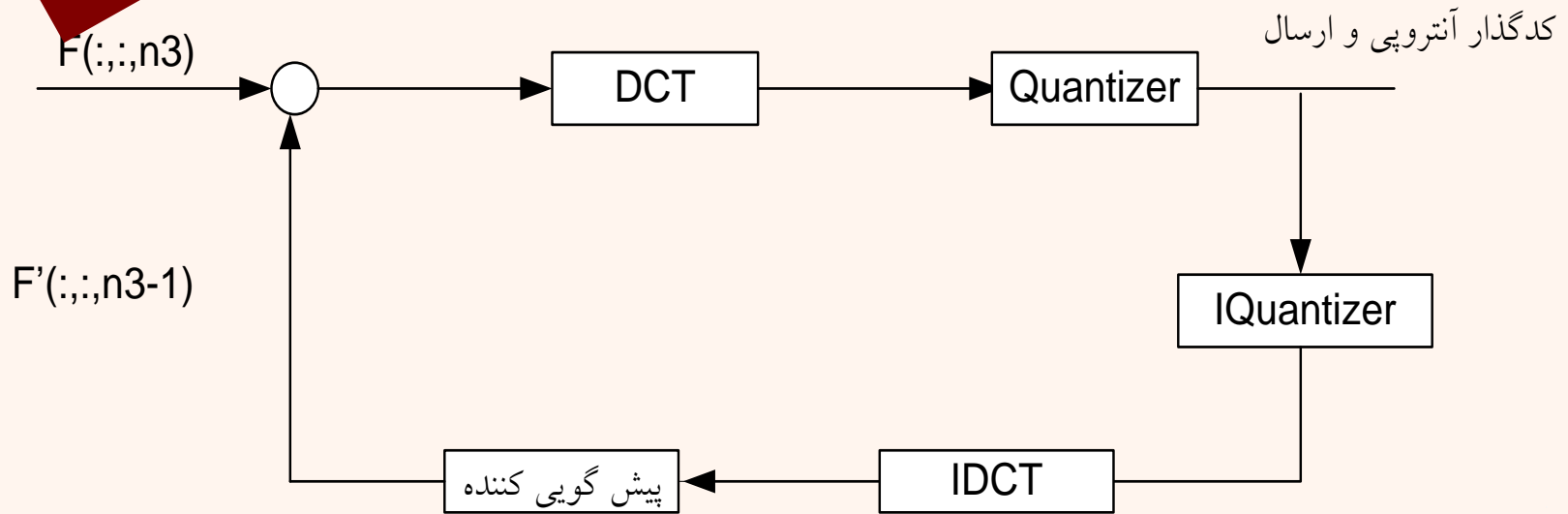


# نمودی ذخیره‌کردن فریم‌ها در کدگذار



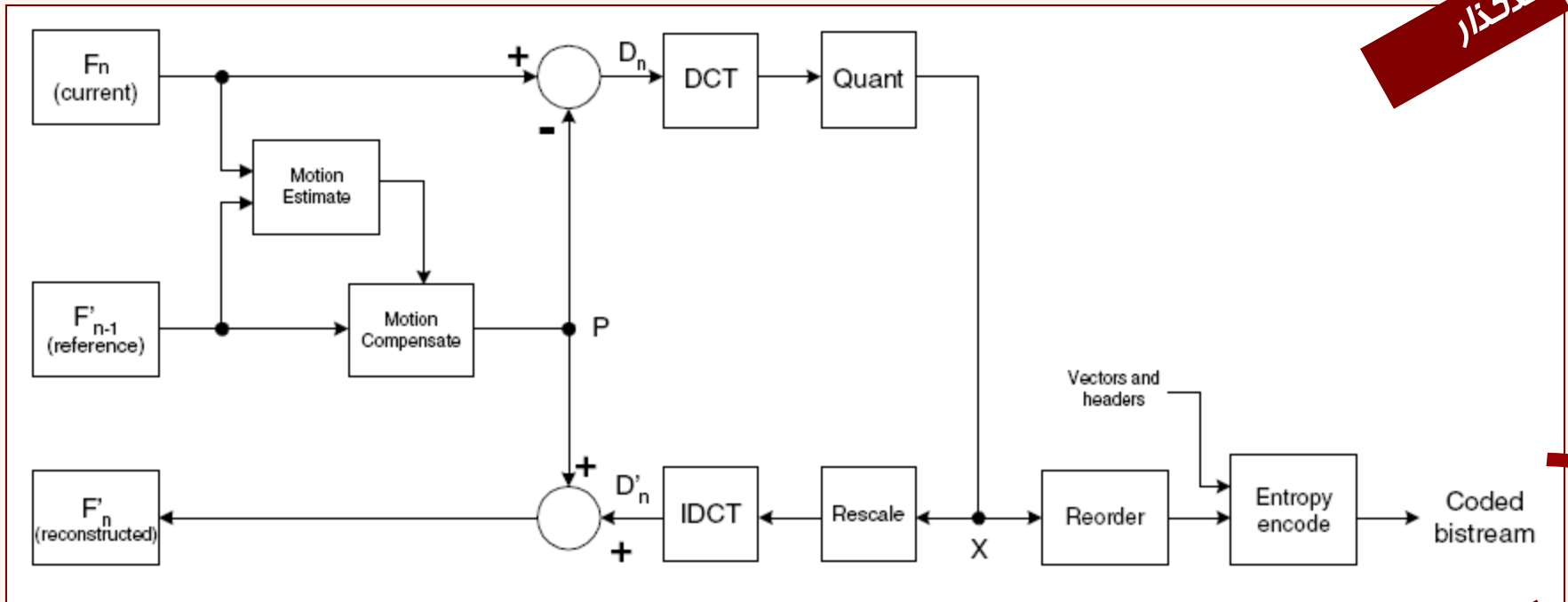
# شمای کلی کدگذار

کدگذار

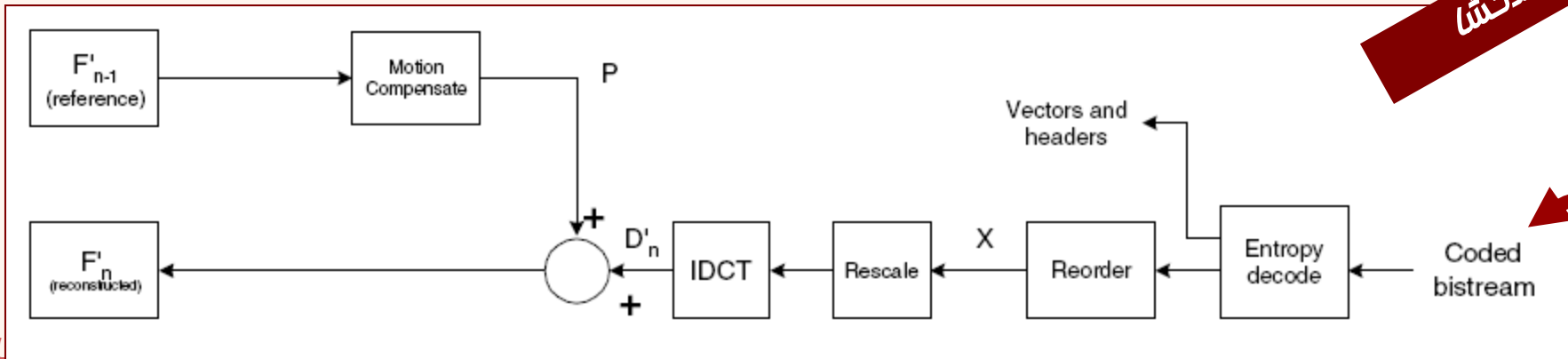


**compatible with H.261, H.263, MPEG-1, MPEG-2, MPEG-4 Visual and H.264**

کدگذار



کدگشا



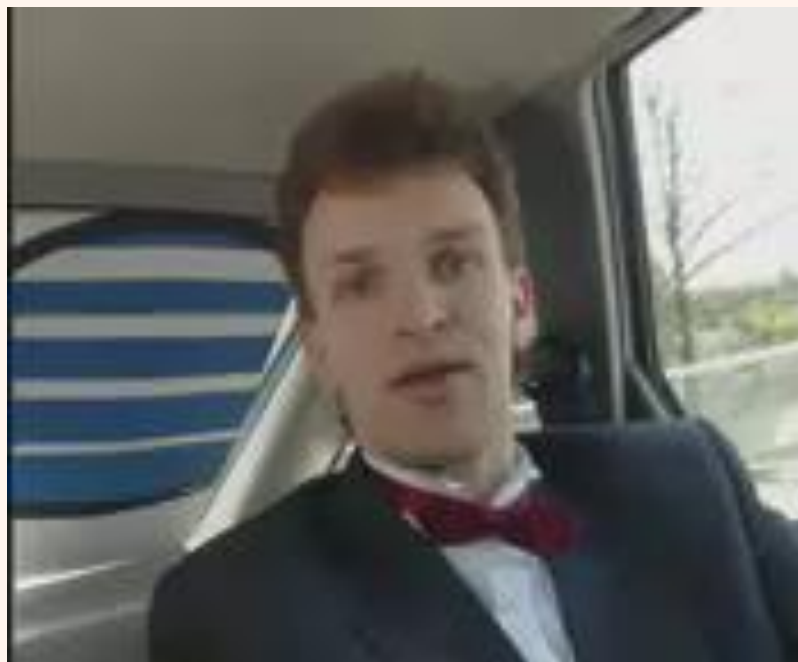
لازی

# کیفیت زمانی و مکانی



کیفیت مکانی

کیفیت زمانی





# کیفیت در رسانه‌ی ویدئو

- بررسی کیفیت به صورت شکستن ویدئو به سری داده‌های دوبعدی

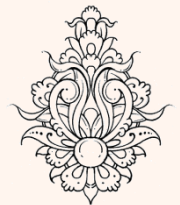
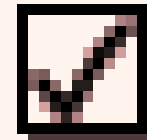
– میانگین‌گیری از فریم‌ها

• PSNR

• SSIM

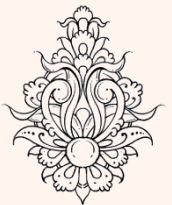
- بررسی مستقل رسانه‌ی ویدئو

• VQM



# کیفیت در رسانه‌ی ویدئو (ادامه...)

- در صورت بررسی هر فریم ویدئو به صورت یک سیگنال دو بعدی و سپس اعمال میانگین‌گیری – خصوصیتی که مختص سیگنال سه بعدی است ممکن است از دست برود.
- اثر فرآیندهایی همچون سوسو زدن در این چنین پارامتر کیفیتی لحاظ نخواهد شد.
- بهتر است برای محاسبه‌ی کیفیت در یک سیگنال ویدئویی سیگنال به صورت سه بعدی مورد ارزیابی قرار گیرد.



## (Exhaustive Block Matching Algorithm)

- ممکن است اندازه‌ی بردارهای ضریب صحیحی از پیکسل‌ها نباشد!

- اندازه‌ی گام را معمولاً ۰.۵ در نظر می‌گیرند تا جستجوی بهتری صورت گیرد.

- **فریم مرجع تنها پیکسل‌های صحیح دارد.**

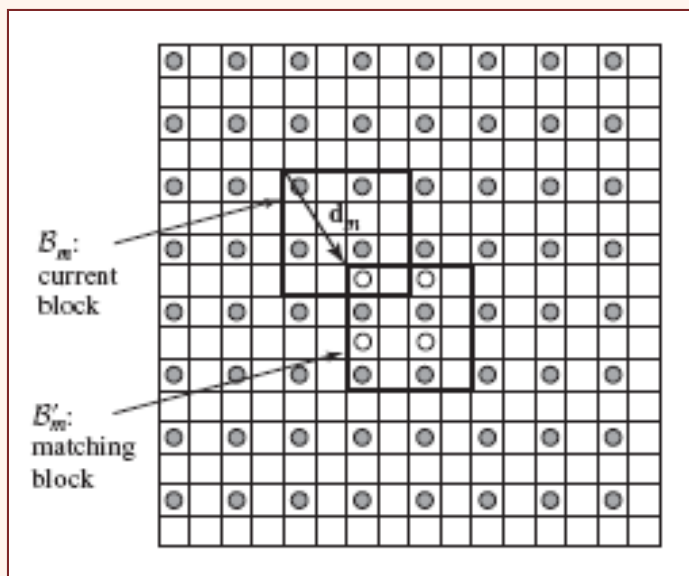
- **راه‌حل:**

- قبل از جستجو میانگین‌گیری صورت گیرد.

- معمولاً Bilinear interpolation صورت می‌گیرد.

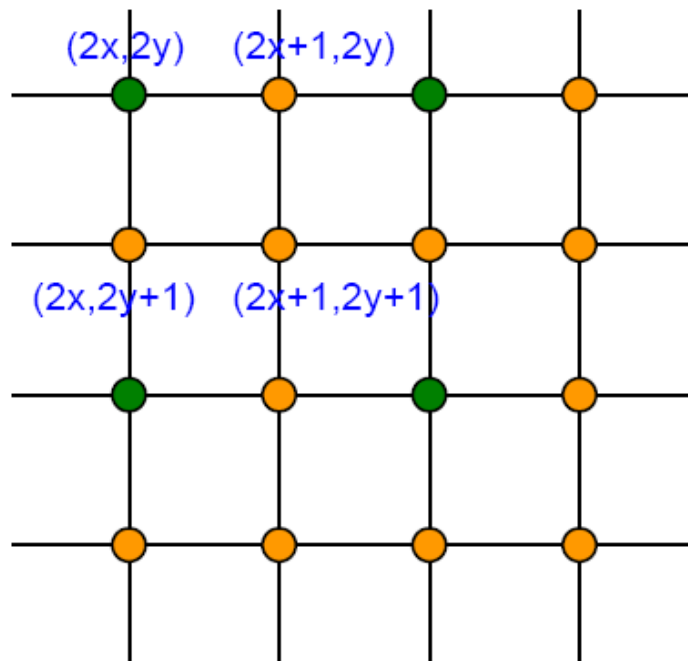
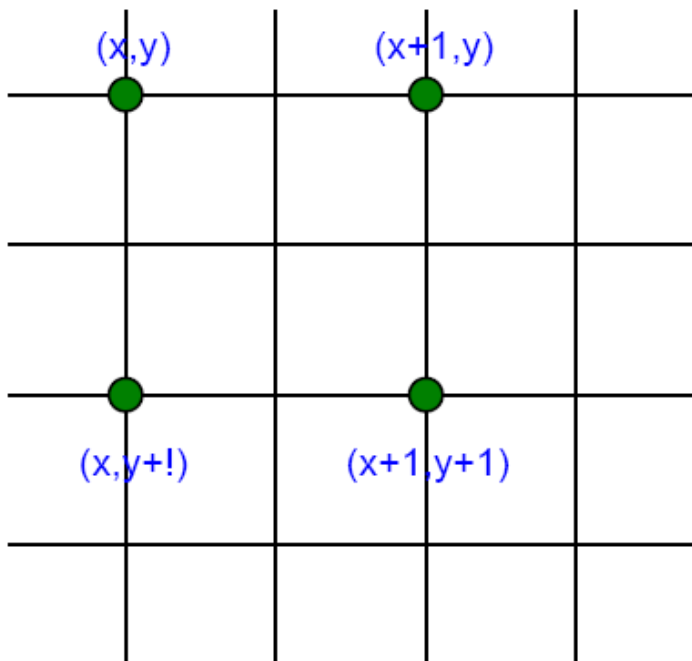


- پیچیدگی: چهار برابر نسخه‌ی صحیح به علاوه‌ی سایر فرآیندهای مورد نیاز
- الگوریتم سریع: ابتدا جستجو را برای گام‌های صحیح و سپس برای محدوده‌ی کوچک‌تری گام نیم‌پیکسلی جستجو صورت می‌گیرد.



# Bilinear Interpolation

# درون‌یابی دوخطی

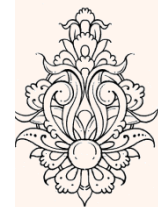


$$O[2x,2y]=I[x,y]$$

$$O[2x+1,2y]=(I[x,y]+I[x+1,y])/2$$

$$O[2x,2y+1]=(I[x,y]+I[x+1,y])/2$$

$$O[2x+1,2y+1]=(I[x,y]+I[x+1,y]+I[x,y+1]+I[x+1,y+1])/4$$

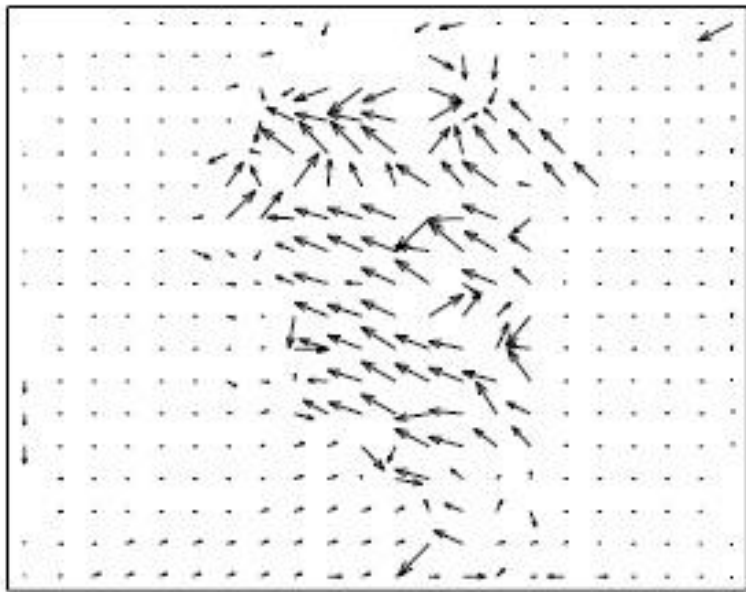




(a)



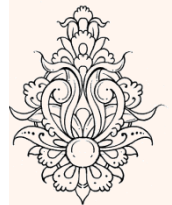
(b)



(c)

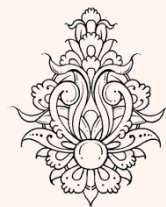


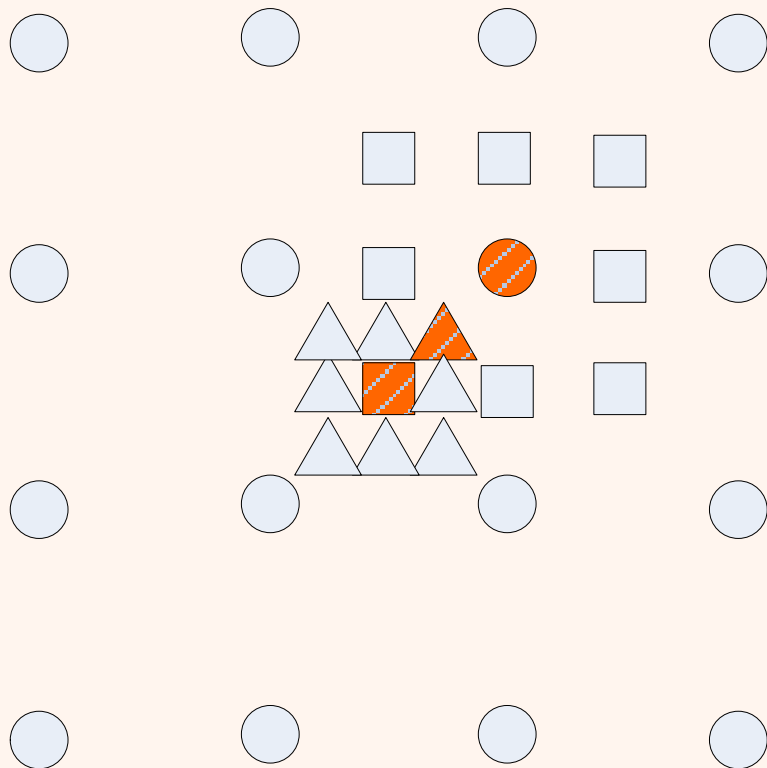
(d)





**(a) the tracked frame; (b) the anchor frame; (c-d) motion field and predicted image the anchor frame (PSNR=29.86 dB) obtained by half-pel accuracy EB**


- مفهوم جستجوی یک چهارم نمونه: در مرحله اول تخمین حرکت با استفاده از نمونه‌های صحیح که با علامت دایره نشان داده شده است صورت می‌پذیرد.
  - کدگذار همزمان در محدوده بهترین انتخاب، مکان‌های نیم‌نمونه را جستجو می‌کند (اشکال مربع)
  - پس از انتخاب بهترین نیم نمونه جستجو به جهت یافتن بهترین ربع‌نمونه صورت می‌پذیرد (اشکال مثلث).
  - در پایان بهترین انتخاب از میان انتخاب‌های صورت گرفته برگزیده خواهد شد.







صحيح جستجوی های مکان 

صحيح تطابق بهترین 

صحيح نیم جستجوی های مکان 

صحيح نیم تطابق بهترین 

صحيح ربع جستجوی های مکان 

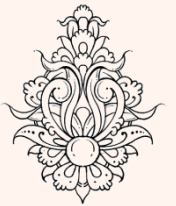
صحيح ربع تطابق بهترین 



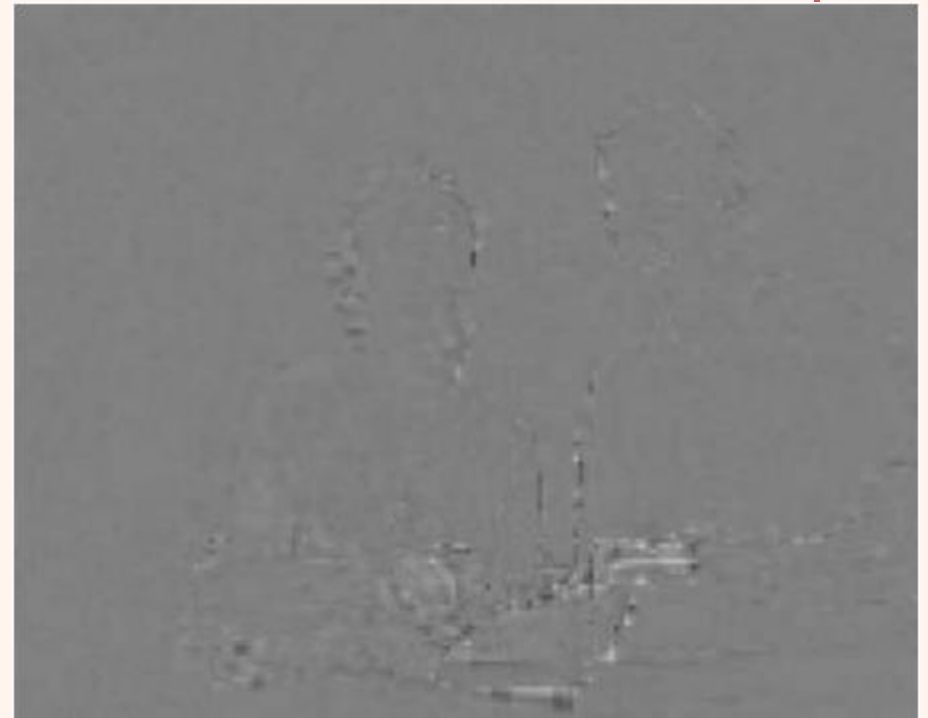
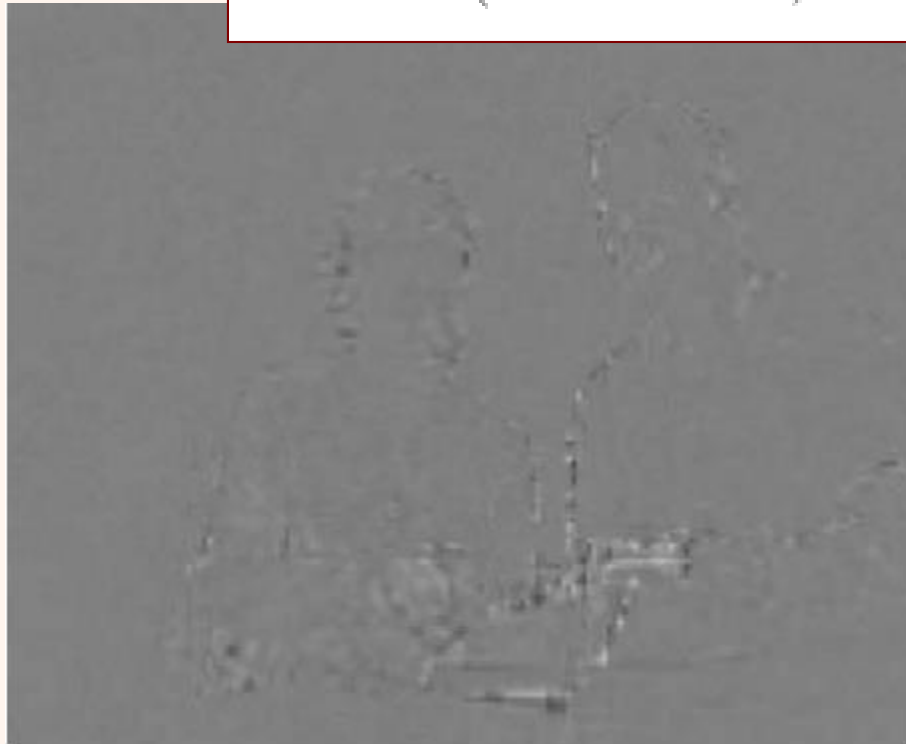


# نمونه برداری غیرصمیمی

- با توجه به آزمایش‌های صورت گرفته فشرده‌سازی مدل نیم‌نمونه‌ای نسبت به مدل صمیمی، بازده بسیار بیشتری را نسبت به گزینش مدل ربع‌نمونه‌ای نسبت به نیم‌نمونه‌ای دارا می‌باشد.  
– استفاده از گام کوچک‌تر اگر چه کیفیت بالاتری را ارائه می‌نماید لیکن پیچیدگی بسیار بالایی را طلب می‌کند.



Residual (4 × 4 blocks, half-pixel compensation)



Residual (4 × 4 blocks, quarter-pixel compensation)