

PENKODEAN VIDEO DENGAN METODE SPATIAL SCALABILITY

Aan Darmawan¹ dan Riko Arlando Saragih²

Jurusan Teknik Elektro Universitas Kristen Maranatha

Jl. Prof. Drg. Suria Sumantri 65, Bandung, Indonesia

Phone: +62222012186 Fax +62222015154

E-mail: ⁽¹⁾aan.darmawan@eng.maranatha.edu, ⁽²⁾riko.as@eng.maranatha.edu

Abstrak

Dalam penggunaan video digital, diperlukan media penyimpanan dengan kapasitas yang besar untuk dapat menyimpan data video digital dan bandwidth yang besar untuk dapat mengirim data video digital. Pada tulisan ini akan dipaparkan metode Spatial Scalability MPEG-2 untuk mengkompresi data video, sehingga dalam proses penyimpanan dan transmisinya tidak memerlukan kapasitas yang besar. Masukan berupa video digital dengan resolusi 160 x 120 dan diproses dengan menggunakan encoder dan decoder spatial scalability dengan mengubah-ubah nilai faktor kompresi.

Kuantitas dan kualitas gambar hasil kompresi diperoleh dengan menghitung Mean Absolute Error (MAE). Dari penelitian diperoleh bahwa semakin besar nilai MAE, maka semakin turun kuantitas gambar tersebut. Contoh : untuk faktor kompresi $K = 3$, diperoleh nilai $MAE = 2,72$, sedangkan untuk faktor kompresi $K = 15$, diperoleh nilai $MAE = 4,94$.

Kata kunci: MPEG-2, spatial scalability, MAE.

1. Pendahuluan

Teknik kompresi video menggunakan algoritma yang kompleks dan memerlukan sirkuit elektronik yang kompleks sehingga membutuhkan biaya yang mahal, karena itu diperlukan suatu standard yang memungkinkan teknik kompresi yang efektif.

Salah satu standard yang ada adalah MPEG-2. Saat ini MPEG-2 digunakan untuk berbagai aplikasi seperti *digital terrestrial broadcasting*, *digital satellite TV*, dan DVD. MPEG-2 menggunakan *bi-directionally predictive pictures* sehingga memenuhi *delay coding* yang dapat diterima. Selain itu, dari single bit-stream MPEG-2 dapat diekstrak dua atau lebih video image dengan macam-macam resolusi, baik spatial, temporal, atau SNR.

2. Standar MPEG-2

Standar MPEG-2 adalah standar video codec generik dari ITU-T dan ISO/IEC(SG15 ITU-T) serta terdiri dari tiga elemen, yaitu:

- Bagian pertama: system standard (ISO/IEC 13818-1).
- Bagian kedua: video standard (ISO/IEC 13818-2).
- Bagian ketiga: audio standard (ISO/IEC 13818-3).

Bagian pertama menjelaskan sistem termasuk sinkronisasi audio-video, multiplexing dan informasi lain yang berhubungan dengan sistem. Bagian kedua berisi *codec representation* dari data video dan proses dekoding. Bagian ketiga berisi representasi dari data audio dan proses dekoding.

Sintaks video MPEG-2 mempunyai struktur hierarkis 6-layer, seperti yang terlihat pada Tabel 1 berikut.

Tabel 1. MPEG-2 Layer

<i>Syntax Layer</i>	<i>Functionality</i>
Sequence Layer	Context unit
Group-of-picture Layer	Random acces unit : video coding
Picture Layer	Primary coding unit
Slice Layer	Resynchronisation unit
Macroblock Layer	Motion compesation unit
Block Layer	DCT unit

Pada Tabel 1, *Sequence Layer* berisikan parameter-parameter dasar seperti ukuran video, *bit-rate*, *picture-rate*. *Group-of-picture Layer* adalah *layer* yang memungkinkan MPEG-2 mendukung *random access*, *fast search*, dan pengeditan. *Picture Layer* berisi tipe gambar yang ada, misalnya I, P, atau B. *Slice Layer* digunakan untuk sinkronisasi ulang pada saat transmisi bila ada bit yang error. Pada *macroblock Layer* dilakukan motion compensation dan *Blok Layer* berisi data koefisien DCT yang telah dikuantisasi.

Pada MPEG-2 kompresi video dan audio dilaksanakan dengan MPEG-2 video dan audio standar. Untuk aplikasi praktis *bit stream* video dan audio yang sudah dikompres harus digabungkan dalam satu single bit stream sehingga dapat disimpan dalam *Digital Storage Media* (DSM) atau ditransmisikan melalui kanal komunikasi (MPEG-2 system layer).

MPEG-2 mendefinisikan dua tipe stream, yaitu:

1. Program Stream:
 - Kompatibel dengan stream MPEG-1 *system*, tetapi dengan modifikasi sintaks dan fungsi baru, yaitu *scalability*.
 - Ukuran paket biasanya 1-2 kbyte (sampai 64 kbyte).
 - Menyediakan fitur yang tidak didukung MPEG-1 (scrambling data, prioritas paket, indikasi copyright, dan lain-lain).
2. Transport Stream:
 - Tahan terhadap kanal noisy.
 - Ukuran paket tetap 188 byte dengan header sintaks baru.
 - Cocok untuk hardware processing dan skema *error correction* (diperlukan pada broadcasting TV, satelit/cable tv, jaringan ATM).

3. Level & Profile MPEG-2

Profile adalah subset dari keseluruhan sintaks bit-stream yang didefinisikan oleh spesifikasi MPEG-2. Misal: untuk keperluan *broadcast* B-picture, maka *scalability* tidak diperlukan, tetapi untuk keperluan DSM perlu semua *picture* tetapi tidak diperlukan *scalability*. Sedangkan level adalah satu set batasan untuk parameter-parameter dalam bit-stream (didefinisikan dalam tiap profile). Sebagai contoh, untuk format SIF, kecepatan kompresinya lebih kecil daripada 4Mbit/s dan termasuk dalam low level. Sedangkan untuk resolusi 1920 x 1250, kecepatan kompresinya di atas 80 Mbit/s.

Jika hanya ada satu layer bit-stream, maka dinamakan non-scalable bit-stream. Sedangkan jika ada dua atau lebih layer bit-stream dinamakan scalable, dengan layer pertama disebut base layer (dikodekan secara independen) dan layer lainnya disebut enhancement layer (hanya bisa dikodekan dengan lower layer).

4. Prediksi Frame untuk Frame Picture

Prediksi frame untuk frame picture pada MPEG-2 identik dengan prediksi yang digunakan pada MPEG-1, yaitu:

- a. Tiap frame dapat membuat prediksi dari frame sebelumnya.
- b. Ada satu *motion vector* untuk tiap *motion compensated macroblock*.
- c. B-frame dapat menggunakan frame sebelumnya atau yang akan datang.
- d. Akan ada sampai dua *motion vector* (*forward* dan *backward*) untuk tiap B-frame *motion compensated macroblock*.

5. Field Prediction untuk Field Pictures

Selain prediksi frame bisa juga digunakan prediksi field dengan setiap frame dibagi secara interlaced menjadi field-field yang kemudian dikodekan secara terpisah sebagai berikut:

- a. Prediksi mode ini sama dengan prediksi frame, kecuali pixel-pixel dari macroblock target (MB yang akan dikodekan) dari field yang sama.
- b. Prediction MB harus datang dari satu field (*top* atau *bottom field*).
- c. Contoh prediksi MB target pada top field pada P-frame, dapat datang dari *top field* atau *bottom field* dari frame referensi.
- d. Untuk P-pictures prediction MB datang dari dua field.
- e. Prediksi MB target pada *bottom field* diperoleh dari dua field terbaru, yaitu *top field* dari frame yang sama atau bottom field dari frame referensi.
- f. Untuk prediksi B-pictures MB diambil dari dua anchor picture terbaru (I/P atau P/P). Tiap target MB dapat membuat *forward* atau *backward prediction* dari satu field.

6. Scalability

Scalable video coding disebut juga *layered coding*. Metode scalability pada awalnya diusulkan untuk menangani *cell loss* pada jaringan ATM. Ide dari metode ini adalah codec membangkitkan dua bit-stream menjadi:

- a. Base layer: memuat informasi video yang vital.
- b. Enhancement layer: memuat informasi residual untuk meningkatkan kualitas image base layer.

Ada beberapa macam metode *scalability*, yaitu:

- a. Data Partitioning
Single-coded video bit-stream yang dipartisi menjadi dua layer atau lebih untuk keperluan transmisi atau penyimpanan. Pada dekoder kedua layer dapat dikombinasikan untuk menghasilkan gambar dengan kualitas yang sama dengan *bit-stream* yang tidak dipartisi. *Base layer* berisi bagian kritis dari *bit-stream* dan ditransmisikan pada kanal dengan error yang lebih kecil. Sementara itu *enhancement layer* untuk data yang kurang kritis dan ditransmisikan pada kanal dengan error lebih besar.
- b. SNR Scalability
SNR scalability merupakan *scalability* yang menggunakan lebih dari satu layer. Tiap layer mempunyai kualitas yang berbeda dalam resolusi spasial yang sama. *SNR scalability* membutuhkan *bit* lebih sedikit daripada *data partitioning*. Secara khusus pengkombinasian pada dekoder menghasilkan kualitas yang lebih rendah dibandingkan dengan *data partitioning*. *Base layer* dalam

SNR scalability menyediakan kualitas dapat digunakan untuk menyediakan layanan multi kualitas dan pada aplikasi *error-resilient*.

- c. Spatial Scalability
Spatial scalability adalah metode pengkodean yang menyediakan dua atau lebih layer. Metode pengkodean ini lebih kompleks daripada SNR scalability tetapi juga menawarkan fleksibilitas resolusi yang berbeda pada tiap layer. Base layer dapat mencapai kualitas yang cukup baik. Kualitas dari kombinasi layer dalam beberapa kasus lebih rendah dari hasil non-scalable coding. Base layer menyediakan resolusi spasial dasar dan enhancement layer memberikan resolusi spasial penuh dari input video.
- d. Temporal Scalability
Temporal scalability merupakan scalability yang menggunakan lebih dari satu layer. Tiap layer mempunyai resolusi temporal yang sama atau berbeda, tetapi masih dalam resolusi spasial yang sama. Temporal scalability memerlukan kompleksitas pengkodean yang sama dengan yang dihasilkan oleh non-scalable coding. Pada base layer dan enhancement layer mungkin memiliki sebagian resolusi temporal dari input video. Kualitas spasial dari base layer biasanya lebih tinggi.
- e. Hybrid Scalability
Hybrid scalability merupakan gabungan kombinasi dari ketiga scalability, sehingga diperoleh kualitas yang lebih baik.

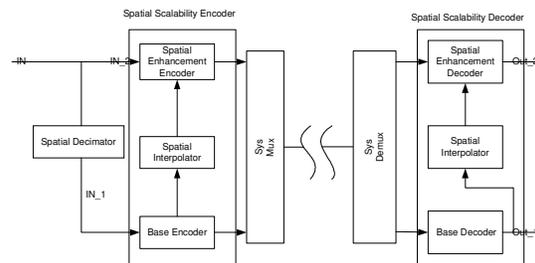
Dalam penelitian ini metode yang digunakan adalah spatial scalability. Spatial scalability mengacu pada pengkodean layer di daerah spasial. Hal ini berarti bahwa pada tipe pengkodean ini, lower layer menggunakan resolusi spasial yang sama atau lebih rendah daripada resolusi spasial dari layer yang lebih tinggi. Layer yang lebih tinggi menggunakan prediksi spasial antar layer untuk mengkodekan gambar dari layer yang lebih rendah.

Metode spatial scalability menawarkan fleksibilitas dalam pemilihan resolusi dan format video untuk digunakan pada tiap layer. Hal ini memungkinkan interoperability antara format-format video yang berbeda. Lebih jauh lagi, lower layer juga dapat dikodekan dengan menggunakan standard yang berbeda. Misalnya, spatial scalability menawarkan kecocokan dengan standar-standar yang lain seperti H.261 dan MPEG-1.

Spatial scalability mendukung interoperability antara aplikasi-aplikasi dengan menggunakan format-format video yang berbeda, interworking dengan standar yang berbeda, interworking antara broadcast dan aplikasi komputer, interworking antara berbagai macam aplikasi telekomunikasi, sistem dua layer yang terdiri atas HDTV dan TV normal, dan lain-lain.

7. Spatial Scalability Codec

Codec pada Gambar 1 hanya sebagai representasi fungsional dari operasi high-level. Walaupun menunjukkan encoder dan decoder yang terpisah untuk setiap layer, encoder/decoder untuk high layer tidak selalu sama dengan encoder/decoder untuk layer yang lebih rendah. Hal ini berarti bahwa beberapa operasional dapat dikombinasikan dengan satu encoder/decoder. Dalam spatial scalability, layer yang lebih rendah dapat dinyatakan sebagai base layer dan layer yang lebih tinggi sebagai enhancement layer.



Gambar 1. Spatial scalability codec

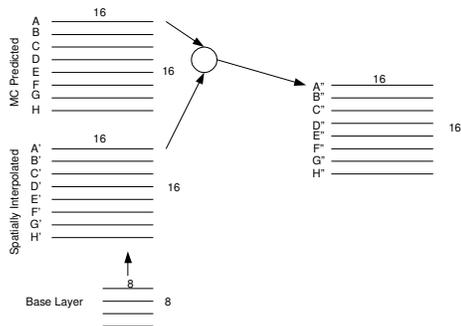
Masukan video pada *in* di *downsample* oleh *spatial decimator* dan diterapkan pada *in_1* ke *Base Encoder*, yang menghasilkan *bitstream* yang telah dikodekan yang kemudian masuk ke *multiplexer(Sys Mux)*. Sinyal video yang telah dikodekan oleh *base layer* juga masuk ke *Spatial upsampler* yang dinamakan *Spatially interpolated* dan menghasilkan sinyal yang terinterpolasi secara spasial ke *Spatial Enhancement Encoder*, yang kemudian digabungkan dengan sinyal prediksi sementara yang dibangkitkan di *Spatial Enhancement Encoder* ketika mengkodekan sinyal dari *in_2*. *Bitstream* yang telah dikodekan kemudian dimasukkan ke *Sys Mux*. *Bitstream* dari *base layer* dan *bitstream* dari *enhancement layer* dipaketkan dan *stream* dari paket-paket tersebut siap untuk di kirimkan ke *network* atau tempat penyimpanan. Pada *decoder*, *Sys Demux* memisahkan paket-paket dari *base layer* dan *enhancement layer* serta membangkitkan *base layer bitstream* dan *enhancement layer bitstream*. *Base Decoder* mendekodekan *base layer bitstream* dan menghasilkan video yang telah didekodekan pada *out_1*, yang juga dimasukkan ke *Spatial Interpolator* untuk menghasilkan sinyal terinterpolasi yang digunakan sebagai prediksi oleh *Spatial Enhancement Decoder* berdasarkan pengkodean pada saat mengkodekan *enhancement bitstream*. Video dari *enhancement layer* yang telah didekodekan keluar dari *out_2*.

8. Prediksi Spatiotemporal

Dalam *Spatiotemporal Weighted Prediction* (Gambar 2), pada basis makroblok, sinyal *lower layer* yang telah dikodekan secara *spatially*

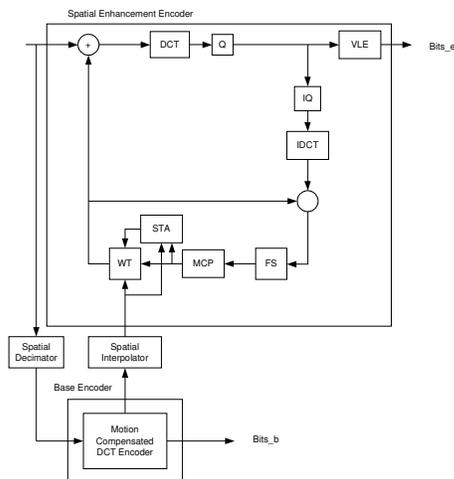
interpolated digabungkan dengan prediksi *motion-compensated*. Untuk setiap makroblok, *Spatiotemporal Weighted Prediction* dibangkitkan dan digunakan untuk perhitungan prediksi *error*.

Pada Gambar 2, blok 8 x 8 pada *base layer* dilakukan *upsampling* menjadi blok 16 x 16 yang kemudian digabungkan dengan blok prediksi temporal 16 x 16 (didapatkan dengan *motion compensation*).



Gambar 2. Spatio-temporal Prediction

Dalam Gambar 3 *spatial enhancement encoder* terdiri dari *motion-compensated DCT encoder* untuk menghitung *weighted spatiotemporal prediction* yang akan dipilih dan memilih yang terbaik pada basis makroblok. Pada *motion-compensated DCT encoding* yang normal, frame-frame video yang disalurkan pada *in* dipartisi menjadi *nonoverlapping* makroblok (16 x 16 blok *luminance* dan dua blok 8 x 8 *chrominance*) yang kemudian diprediksi dengan menggunakan frame-frame sebelumnya. Hasil dari sinyal-sinyal tersebut di DCT, dan kemudian dikuantisasi di *quantizer(Q)*.



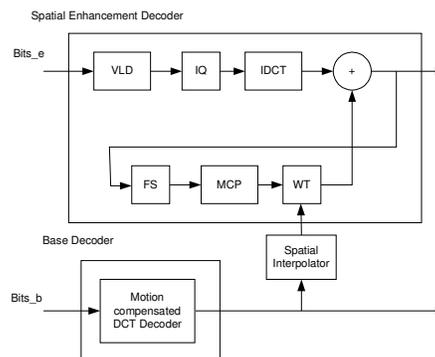
Gambar 3. Spatial scalability encoder

Indeks kuantisasi, *motion vector*, dan data-data lainnya di *variable length encoded* pada VLE untuk membangkitkan *bitstream*. Indeks kuantisasi juga diumpun-balikkan ke kalang yang terdiri atas

Inverse Quantizer(IQ), *Inverse DCT (IDCT)*, dan *Adder, Frame Store (FS)*, *Motion*

Estimation Predictor (MCP). Selain blok-blok yang ditemukan di *motion-compensated DCT coding* yang normal, *encoder spatial scalability* mempunyai blok-blok yang berbeda seperti *Spatio Temporal Analyzer (STA)* dan *WT(Weighter)*. Prediksi *temporal* dari MCP juga diteruskan ke STA dan WT. STA memilih *weight* yang memperkecil *error* prediksi dan WT membangkitkan prediksi *spatiotemporal*. Hasil *bitstream* dari *enhancement layer* adalah *bits_e* dan dikirim ke *Sys Mux* yang juga menerima *bitstream* dari *base layer(bits_b)*.

Operasi dari *Spatial Enhancement Decoder* mirip dengan *Motion Compensated DCT Decoder* yang normal. Operasi dari *decoder* ini juga indentik dengan operasi dari *decoding loop* lokal dalam *Spatial Enhancement Encoder*.



Gambar 4. Spatial scalability Decoder

Pada Gambar 4 *bitstream* dari *Enhancement layer(bits_e)* dari *Sys Demux* masuk ke *spatial Enhancement Decode*. *Bitstream* pertama kali di proses oleh *Variable Length Decoder(VLD)*, kemudian diteruskan ke *Inverse Quantizer(IQ)*, *Inverse DCT(IDCT)*, dan kemudian ke *Adder* yang mempunyai masukan lainnya berupa sinyal prediksi.

Sinyal prediksi ini dibangkitkan oleh *spatiotemporal weighted prediction*, yang mempunyai prosedur yang serupa dengan yang ada pada *Encoder*. Yang membedakannya hanya *weighting* yang digunakan oleh WT didekodekan dari *bitstream* oleh VLD bersama dengan data-data yang lain. WT menggunakan *output* dari prediksi *spatial* yang telah di-*upsample* oleh *Spatial Interpolator*, gambar *Enhancement layer* yang telah didekodekan dikeluarkan melalui *Adder*.

9. Discrete Cosine Transform (DCT)

Discrete cosine transform (DCT) mentransformasi data dari domain ruang ke domain frekuensi. Masukan proses DCT berupa matriks

data NxN. Persamaan DCT untuk matriks NxN adalah sebagai berikut:

$$F(u, v) = \frac{1}{\sqrt{2N}} C(u).C(v) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x, y) \cos\left[\frac{(2x+1)mu}{2N}\right] \cos\left[\frac{(2y+1)mv}{2N}\right] \quad (1)$$

Keterangan:

- $C(i) = \frac{1}{\sqrt{2}}$ untuk $i = 0$,
- $C(i) = 1$ untuk $i \neq 0$
- $f(x, y) =$ data pada domain ruang
- $F(u, v) =$ data pada domain frekuensi
- x dan $y =$ baris ke x dan kolom ke y pada domain ruang
- u dan $v =$ baris ke u dan kolom ke v pada domain frekuensi

Untuk mentransformasikan kembali data dari domain frekuensi ke domain ruang digunakan inverse dari *discrete cosine transform* atau IDCT. Persamaan IDCT untuk blok matriks NxN sebagai berikut:

$$f(x, y) = \frac{1}{\sqrt{2N}} \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} C(u).C(v) F(u, v) \cos\left[\frac{(2x+1)mu}{2N}\right] \cos\left[\frac{(2y+1)mv}{2N}\right] \quad (2)$$

10. Kuantisasi (Q)

Proses kuantisasi merupakan proses untuk mengurangi jumlah bit yang diperlukan untuk menyimpan suatu nilai dengan memperkecil jumlah bit. Hasil keluaran dari DCT memerlukan kuantisasi agar lebih sedikit bit yang dibutuhkan untuk dikirimkan. Kuantisasi dilakukan dengan membagi keluaran proses DCT dengan suatu nilai yang ditetapkan dalam matriks kuantisasi. Proses kuantisasi dilakukan sebagai berikut:

$$\text{Hasil kuantisasi } (u, v) = F(u, v) / \text{Quantum } (x, y) \quad (3)$$

Keterangan: Quantum = matriks kuantisasi.

Matriks kuantisasi dapat dipilih uniform atau nonuniform. Pada matriks kuantisasi uniform semua koefisien mempunyai besar yang sama. Sedangkan nilai tiap koefisien pada matriks kuantisasi nonuniform tersebut ditentukan dengan suatu persamaan, misalnya:

$$\text{Quantum } (u, v) = 1 + Q * (1 + u + v) \quad (4)$$

Keterangan:

- $u, v = 0, 1, 2, \dots, N$
- $Q =$ faktor kualitas (1,2,3,...)

Jika nilai-nilai koefisien matriks ini semakin besar, keluaran yang dihasilkan akan semakin kecil, sehingga jumlah bit yang dibutuhkan juga makin kecil. Hal ini mengakibatkan faktor kompresi semakin tinggi. Tetapi walaupun demikian, kualitas hasil rekonstruksi semakin rendah, sehingga dalam melakukan proses kuantisasi faktor kualitas Q perlu dipilih dengan pertimbangan faktor kompresi dan kualitas gambar.

Pada dekoder, proses kebalikan dari kuantisasi adalah invers kuantisasi, hasil proses invers kuantisasi cenderung mengalami distorsi

dibanding nilai aslinya. Hal ini dikarenakan pada proses kuantisasi terjadi error yang disebabkan oleh proses pembulatan, sehingga berpengaruh pada kualitas gambar hasil rekonstruksi.

11. Variable Length Coding (VLC)

Variable Length Coding digunakan untuk mengkodekan simbol dengan kode-kode tertentu yang mempunyai panjang berlainan. Pada pengkodean ini simbol yang sering muncul dikodekan dengan kode yang pendek dan simbol yang jarang muncul dikodekan dengan kode yang panjang. Dengan demikian, secara keseluruhan jumlah bit yang dibutuhkan menjadi lebih sedikit dan jumlah bit yang disimpan atau ditransmisikan menjadi lebih kecil.

12. Mean Absolute Error

Mean Absolute Error(MAE) digunakan untuk menghitung kecocokan antara dua blok dengan mencari rata-rata nilai absolut dari selisih tiap pixel yang bersesuaian posisinya. Semakin rendah MAE semakin baik hasil yang diperoleh. MAE dapat diperoleh dengan rumus:

$$MAE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} |B_1(i, j) - B_2(i, j)| \dots \dots \dots \quad (5)$$

Keterangan:

- $m =$ jumlah pixel pada baris
- $n =$ jumlah pixel pada kolom
- $B_1(i, j) =$ letak pixel citra asli pada kolom i dan baris j .
- $B_2(i, j) =$ letak pixel citra hasil kompresi pada kolom i dan baris j .

13. Rasio Kompresi

Rasio Kompresi adalah perhitungan dengan membandingkan besar data citra asli dengan besar data citra hasil kompresi, semakin besar rasio kompresi berarti semakin kecil besar data citra hasil kompresi. Perhitungan rasio kompresi dilakukan dengan persamaan:

$$\text{Rasio}_\text{Kompresi} = \frac{\text{Besar}_\text{Data}_\text{Citra}_\text{Asli}}{\text{Besar}_\text{Data}_\text{Citra}_\text{Hasil}_\text{Kompresi}} \dots \dots \dots (6)$$

14. Hasil Penelitian

Pada bagian ini akan disampaikan data hasil pengujian secara kualitatif dan kuantitatif. Untuk pengamatan secara kuantitatif, dihitung rasio kompresi dan *Mean Absolute Error*(MAE) antara video input dengan video output berdasarkan factor kompresinya (K). Sedangkan untuk pengujian secara kualitatif dilakukan *perhitungan Mean Opinion Score* (MOS). Pengujian dilakukan pada dua frame file *capture.avi* (Gambar 5) yang mempunyai ukuran frame sebesar 160x120 sebagai masukan yang akan dikompresi. Kompresi akan

Hal ini dapat ditunjukkan dari nilai MAE yang diperoleh.

15. Kesimpulan

Dari hasil pengujian yang dilakukan untuk file video yang digunakan, hasil kompresi optimal yang diperoleh adalah $K = 6$.

Daftar Pustaka

- [1] Haskel, Barry G., Atul Puri, and Arun N Netravali, *Digital Video: An Introduction To MPEG-2*, Chapman and Hall, 1997.
- [2] Solari, Stephen J., *Digital Video and Audio Compression*, Mc Graw Hill, 1997.
- [3] Tekalp, A.Murat, *Digital Video Processing*, Prentice Hall.