

3 Estimação e Compensação de movimento na codificação de vídeo

O objetivo do modelo temporal (que engloba as fases de estimação e compensação de movimento) é diminuir a redundância entre quadros transmitidos a partir da subtração de um quadro estimado e o quadro atual. A saída desse processo é um quadro residual e, quanto menor a energia deste, maior é a precisão do processo de predição. O quadro residual é codificado e enviado para o decodificador que recria o quadro estimado, adiciona o resíduo decodificado e reconstrói o quadro atual. O quadro estimado é criado a partir de um ou mais quadros anteriores ou futuros (quadros de referência). A precisão do quadro estimado pode geralmente ser melhorada pela compensação do movimento entre o quadro de referência e o quadro atual.

Neste capítulo apresentaremos os métodos de predição a partir de quadro de vídeo anterior, estimação e compensação de movimento baseada em bloco, compensação de movimento de subpixel e os tamanhos de blocos de compensação de movimento.

3.1 Resíduo formado pela diferença entre quadros

O método mais simples de predição temporal é usar o quadro anterior como um preditor para o quadro atual. Dois quadros sucessivos de uma seqüência de vídeo são mostrados nas figuras 13 e 14. O quadro 1 é usado como um preditor para o quadro 2. O quadro residual ou resíduo, formado a partir da diferença entre o preditor (quadro 1) e o quadro atual (quadro 2), é mostrado na figura 15. Verifica-se nesta figura: a parte acinzentada representa a diferença sendo zero, a parte mais clara corresponde a diferença sendo positiva e a parte mais escura representa a diferença sendo negativa. Essa predição apresenta o problema de, no quadro residual, permanecer uma grande quantidade de energia, o que verifica-se pelas áreas luminosas e escuras. Boa parte dessa energia residual é devida a

movimentos de objetos. Portanto, depois da predição temporal, há ainda uma significativa quantidade de informação para ser comprimida. Uma redução dessa energia poderia ser realizada pela compensação do movimento entre os dois quadros.

3.2 Fluxo óptico

Mudanças entre quadros de vídeo podem ser causadas por:

- movimento do objeto, que pode ser rígido ou deformável. Para o primeiro caso, um exemplo seria um carro em movimento e, para o segundo caso, um exemplo seria um braço em movimento;
- movimento de câmera causado, por exemplo, pela visualização de panorama, inclinação, zoom, rotação;
- regiões descobertas, por exemplo, uma parte do fundo da cena descoberta pelo movimento do objeto; e
- mudanças de luz.

Com exceção das regiões descobertas e das mudanças de luz, essas diferenças correspondem a movimento de pixels entre quadros. A trajetória de cada pixel entre sucessivos quadros de vídeo é conhecida como fluxo óptico, sendo possível estimá-la. A figura 16 mostra o fluxo óptico para os quadros da figura 13 e 14. O fluxo é representado, por simplicidade, somente pelo vetor para o segundo pixel.

Uma vez que o fluxo óptico é conhecido, então é possível determinar com precisão a predição da maioria dos pixels do quadro atual pelo movimento de cada pixel a partir do quadro de referência. No entanto, uma vez que os métodos mais precisos de compensação de movimento usam um procedimento iterativo para todo pixel, um cálculo preciso do fluxo óptico seria computacionalmente intenso. Além disso, seria necessário enviar o vetor de fluxo óptico para todo pixel do decodificador a fim de recriar o quadro estimado. Isto resulta em uma grande quantidade de dados transmitidos e, portanto, perde a vantagem de gerar um pequeno resíduo.



Figura 13- Quadro 1: preditor
Fonte: Ref [1]



Figura 14- Quadro 2: quadro atual
Fonte: Ref [1]



Figura 15- Diferença entre os quadros 1 e 2: quadro residual

Fonte: Ref [1]



Figura 16- Fluxo óptico

Fonte: Ref [1]

3.3 Estimação e Compensação de movimento baseada em bloco

Estimação e compensação de movimento baseada em bloco é um método amplamente usado que considera o movimento de seções retangulares ou blocos do quadro atual. Para cada bloco de amostras $M \times N$ no quadro atual, o seguinte procedimento é realizado:

Passo 1: Pesquisar numa área no quadro de referência (quadro passado ou futuro, previamente codificado e transmitido) para encontrar um bloco $M \times N$ que forneça o melhor casamento. Isto é realizado através da comparação do bloco $M \times N$ do quadro atual com alguns ou todos as possíveis blocos $M \times N$ da área de pesquisa (Figura 17). Um critério de casamento é a minimização da energia do resíduo formado pela subtração da região candidata pelo bloco $M \times N$ atual. Este processo de encontrar o melhor casamento é conhecido como estimação de movimento.

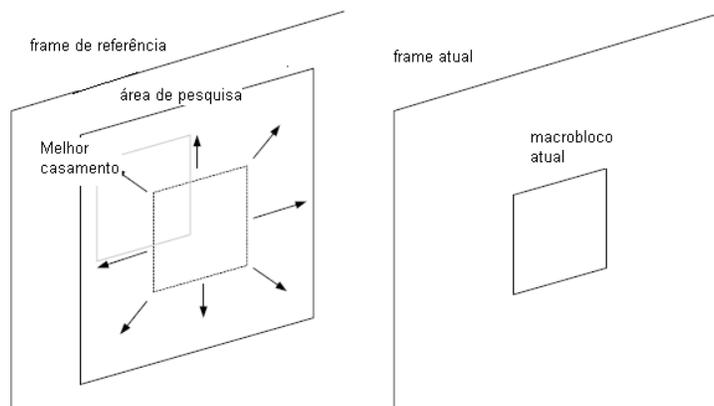


Figura 17- Estimação de movimento

Fonte: Ref [1]

Passo 2: A região candidata escolhida torna-se um estimador para o bloco $M \times N$ atual e é subtraída deste bloco para gerar um bloco residual $M \times N$. Este processo é denominado de compensação de movimento.

Passo 3: O bloco residual é codificado e transmitido e o vetor de movimento gerado pelas posições (geralmente centrais) do bloco atual e da região candidata é também transmitido.

Passo 4: O bloco residual codificado é também decodificado e adicionado à região de melhor casamento para gerar um bloco reconstruído que é armazenado como uma referência para futuras estimações usando compensação de movimento. É necessário usar um resíduo decodificado para reconstruir o bloco a fim de assegurar que tanto o codificador quanto o decodificador usem um quadro de referência idêntico para a compensação de movimento.

Passo 5: No decodificador o vetor de movimento recebido é usado para recriar a região estimada. Além disso, o bloco residual é decodificado e adicionado ao estimado, reconstruindo a versão do bloco original.

Vários são os motivos que levam a compensação de movimento baseada em blocos ser bastante usada. O primeiro é que este método é relativamente direto e computacionalmente implementável. O segundo é que ajusta-se bem com quadros de vídeo retangulares e com transformadas de imagens baseadas em blocos, por exemplo, a Transformada Cosseno Discreta, fornecendo um efetivo modelo temporal para muitas seqüências de vídeos. Há contudo, algumas desvantagens. A primeira delas é que objetos reais raramente tem extremidades perfeitas que casam com fronteiras retangulares. A segunda é que objetos geralmente movem-se por um número fracionário de posições de pixel entre os quadros. A terceira desvantagem é que muitos outros tipos de movimentos de objetos são difíceis de compensar pelo uso de métodos baseados em blocos, por exemplo, objetos deformáveis, rotação e distorção e movimentos complexos tais como o de uma nuvem de fumaça [1]. Apesar de todas estas desvantagens, a compensação de movimento baseada em blocos é usada por todos os padrões de codificação de vídeos atuais.

O processo de estimação e compensação de movimento apresenta muitas variações. O quadro de referência pode ser um quadro anterior (na ordem temporal), um quadro futuro ou uma combinação de estimações a partir de dois ou mais quadros codificados previamente. Se um quadro futuro é escolhido como a referência, é necessário codificar este quadro antes do quadro atual, isto é, quadros devem ser codificados fora de ordem. Por exemplo, numa mudança de cenário, onde existe uma mudança significativa entre os quadros de referência e atual, pode ser mais eficiente codificar um macrobloco pelo modo intra, onde não existe compensação de movimento, do que no modo inter, onde existe compensação de

movimento. Objetos em movimento numa cena de vídeo raramente seguem uma fronteira perfeita nos blocos de tamanho 16x16 pixels. Logo, pode ser mais eficiente usar um tamanho de bloco variável para estimação e compensação do movimento, o que será apresentado no item 3.4. Objetos podem mover-se por um número fracionário de pixels entre quadros (por exemplo, 4,97 pixels ao invés de 5 pixels na direção horizontal). Então uma melhor estimação pode ser realizada utilizando resolução de sub-pixel (através da interpolação de posições inteiras dos pixels) no quadro de referência antes de pesquisar estas posições para o melhor casamento, o que será apresentado no item 3.5.

3.4

Tamanho de bloco variável

Dois quadros sucessivos de uma seqüência de vídeo são mostrados na Figura 18 e Figura 19. Quadro 1 é subtraído do quadro 2 para produzir o quadro residual, sem que se utilize compensação de movimento (Figura 20). A Figura 21 representa a mesma subtração do quadro 1 pelo quadro 2, só que agora usando compensação de movimento para cada bloco de tamanho 16x16 pixel (macrobloco). A energia no bloco residual é reduzida neste caso. Usando a compensação de movimento para cada bloco 8x8 (ao invés de ser para cada macrobloco 16x16) reduz-se a energia residual ainda mais (Figura 22) e a compensação de movimento para blocos 4x4 gera uma energia residual menor ainda (Figura 23).

Esses exemplos, extraídos de [1], mostram que a compensação de movimento usando blocos de tamanhos menores, em certos casos, pode produzir resultados melhores. No entanto, deve-se observar que um bloco de tamanho menor leva a um aumento de complexidade, em função de um maior número de operações de pesquisa que devem ser realizadas, e um aumento no número de vetores de movimento que precisam ser transmitidos. Enviar cada vetor de movimento requer aumento do número de bits a ser enviado o que poderia ser menos benéfico do que os benefícios de uma energia residual reduzida.

Um compromisso em termos de efetividade consiste em adaptar o tamanho do bloco para as características da figura. Por exemplo, escolher um bloco de tamanho grande para regiões de um quadro planas e homogêneas e escolher um

tamanho de bloco pequeno para áreas de alto nível de detalhes e movimentos complexos.

H.264 usa um tamanho de bloco adaptativo na compensação de movimento. A variação de tamanho destes blocos será apresentada no item 4.4.



Figura 18- Quadro 1

Fonte: Ref [1]



Figura 19- Quadro 2

Fonte: Ref [1]



Figura 20- Frame residual (sem compensação de movimento)

Fonte: Ref [1]



Figura 21- Resíduo (com compensação de movimento e tamanho de bloco 16x16)

Fonte: Ref [1]



Figura 22- Resíduo (com compensação de movimento e tamanho de bloco 8x8)

Fonte: Ref [1]

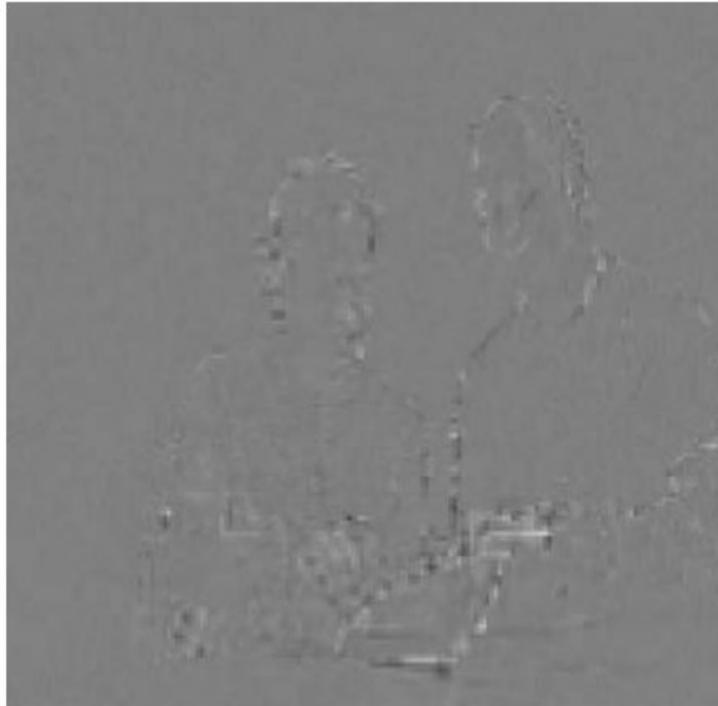


Figura 23- Resíduo (com compensação de movimento e tamanho de bloco 4x4)

Fonte: Ref [1]

3.5 Compensação de movimento com precisão de subpixel

Algumas vezes, uma melhor estimação por compensação de movimento pode ser realizada utilizando posições de amostras interpoladas no quadro de referência. Estimação e compensação de movimento com precisão de sub-pixel envolve pesquisar, além de posições inteiras de pixel, posições interpoladas de sub-pixel, escolhendo a posição que fornece o melhor casamento, ou seja, que minimiza a energia residual.

Figura 24 mostra o conceito da estimação de movimento de um quarto de pixel. No primeiro estágio, estimação de movimento encontra o melhor casamento para o grid de amostras inteiras (círculos). O codificador pesquisa as posições de meio-pixel em seguida para verificar o melhor casamento (quadrados). Para ver se o casamento pode ser melhorado, pesquisa posições de um quarto de pixel (triângulos). O bloco no quadro de referência cuja posição foi descoberta é subtraído do bloco atual ou macrobloco.

O resíduo na Figura 25 é produzido usando bloco de tamanho 4x4 com interpolação de meio-pixel e tem energia residual mais baixa que na Figura 23. Esse método pode ser estendido para uma interpolação com precisão de um quarto de pixel para gerar uma energia residual menor, como mostra a Figura 26.

Em geral, uma interpolação “apurada” fornece melhor desempenho em termos de compensação de movimento, ou seja, fornecendo um resíduo com energia menor, aliada a um aumento da complexidade. O ganho de desempenho tende a diminuir quando aumentam-se as pesquisas com interpolação.

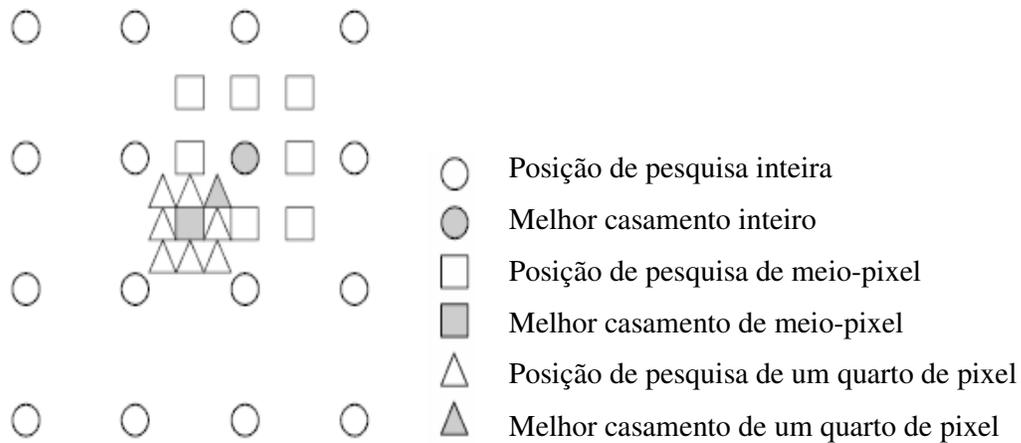


Figura 24- Estimação de movimento de pixel inteiro, metade e um quarto.

Fonte: Ref [1]

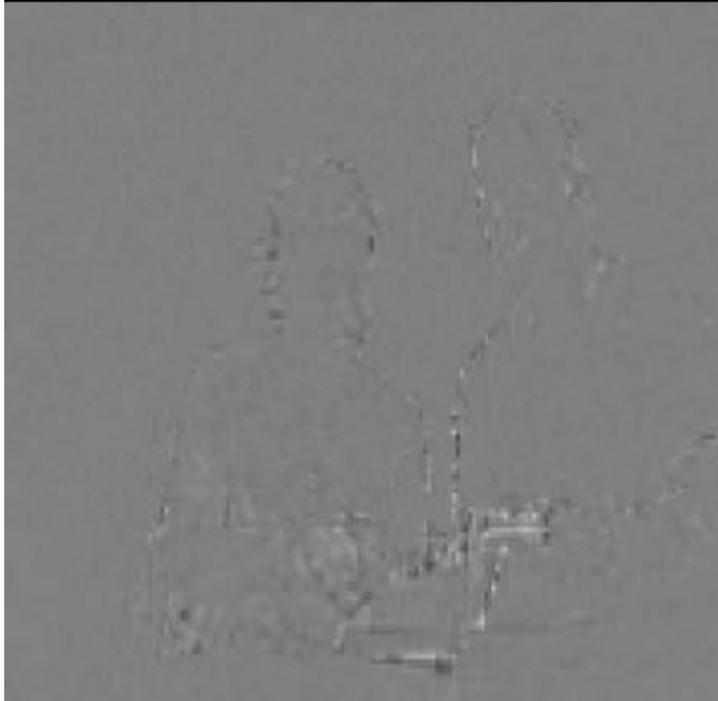


Figura 25- Bloco residual (bloco 4x4, compensação com precisão de meio-pixel)

Fonte: Ref [1]

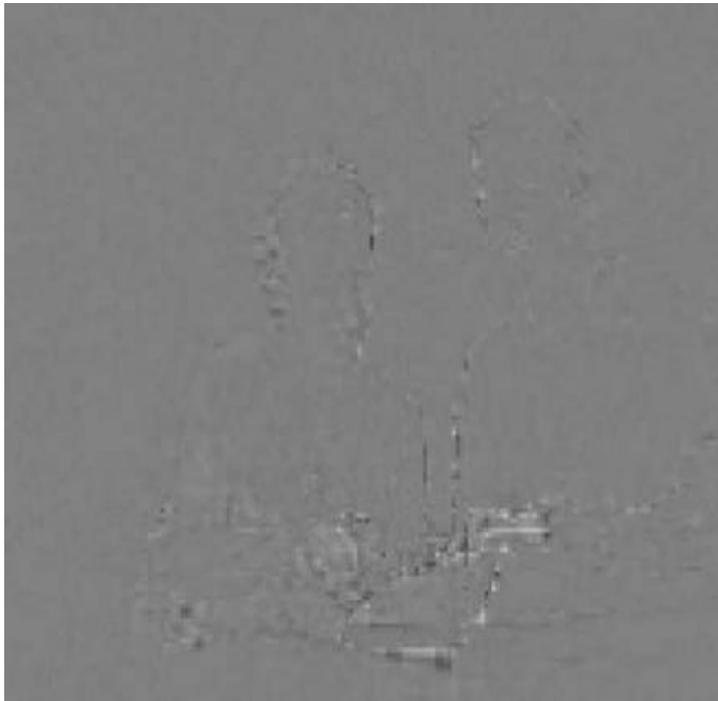


Figura 26- Bloco residual (bloco 4x4, compensação com precisão de um quarto de pixel)

Fonte: Ref [1]

Pesquisar o casamento de blocos 4x4 com interpolação de um quarto de pixel é consideravelmente mais complexo que pesquisar blocos 16x16 sem interpolação. Adicionar esta complexidade, gera uma penalidade de codificação desde que o vetor para todo bloco deve ser codificado e transmitido para o receptor a fim de reconstruir a imagem corretamente. Como o tamanho do bloco é reduzido, o número de vetores a serem transmitidos aumenta. Mais bits são necessários para representar vetores de metade ou um quarto de pixel porque a parte fracionária do vetor (isto é, 0,25 ou 0,5) deve ser codificada também. Há conseqüentemente uma relação proporcional entre eficiência de compressão e complexidade da compensação de movimento.

3.6

Resumo e conclusão do capítulo

Neste capítulo foram apresentados os métodos de predição a partir de quadro de vídeo anterior, estimação e compensação de movimento baseada em bloco, compensação de movimento de subpixel e os tamanhos de blocos de compensação de movimento.

A estimação e compensação de movimento tem por objetivo diminuir a redundância entre quadros transmitidos a partir da subtração de um quadro estimado e o quadro atual. A saída desse processo é um quadro residual e, quanto menor a energia deste, maior é a precisão do processo de predição. O quadro residual é codificado e enviado para o decodificador que recria o quadro estimado, adiciona o resíduo decodificado e reconstrói o quadro atual. O quadro estimado é criado a partir de um ou mais quadros anteriores ou futuros (quadros de referência). A precisão do quadro estimado pode geralmente ser melhorada pela compensação do movimento entre o quadro de referência e o quadro atual.

Vários são os motivos que levam a compensação de movimento baseada em blocos ser bastante usada. O primeiro é que este método é relativamente direto e computacionalmente implementável. O segundo é que ajusta-se bem com quadros de vídeo retangulares e com transformadas de imagens baseadas em blocos, por exemplo, a Transformada Cosseno Discreta. Há contudo, algumas desvantagens. A primeira delas é que objetos reais raramente tem extremidades perfeitas que casam com fronteiras retangulares. A segunda é que objetos

geralmente movem-se por um número fracionário de posições de pixel entre os quadros. A terceira desvantagem é que muitos outros tipos de movimentos de objetos são difíceis de compensar pelo uso de métodos baseados em blocos, por exemplo, objetos deformáveis, rotação e distorção e movimentos complexos tais como o de uma nuvem de fumaça. Apesar de todas estas desvantagens, a compensação de movimento baseada em blocos é usada por todos os padrões de codificação de vídeos atuais.

Estimação e compensação de movimento com precisão de sub-pixel envolve pesquisar, além de posições inteiras de pixel, posições interpoladas de sub-pixel, escolhendo a posição que fornece o melhor casamento, ou seja, que minimiza a energia residual.