



República Federativa do Brasil
Ministério da Economia
Instituto Nacional da Propriedade Industrial

(21) BR 102017024085-1 A2



(22) Data do Depósito: 09/11/2017

(43) Data da Publicação Nacional: 04/06/2019

(54) **Título:** SISTEMA DE RASTREAMENTO VISUAL BASEADO EM TRANSFORMADA TRIDIMENSIONAL DE BAIXA COMPLEXIDADE

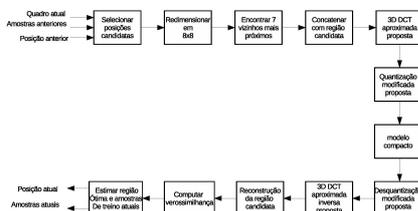
(51) **Int. Cl.:** G06T 7/262; G06K 9/62.

(52) **CPC:** G06T 7/262; G06K 9/62.

(71) **Depositante(es):** UNIVERSIDADE FEDERAL DE PERNAMBUCO; UNIVERSIDADE FEDERAL DE SANTA MARIA.

(72) **Inventor(es):** VÍTOR DE ANDRADE COUTINHO; RENATO JOSÉ DE SOBRAL CINTRA; FÁBIO MARIANO BAYER.

(57) **Resumo:** Esta invenção se trata de um dispositivo de rastreamento de objeto em movimento em um vídeo digital de reduzido consumo energético. O método pode ser empregado em diferentes aplicações, tais como vigilância digital, navegação autônoma, interação homem-robô e realidade aumentada. Este esquema é fundamentado em uma nova transformação tridimensional baseada em uma matriz de transformação associada a uma aproximação para a DCT de baixo custo computacional. O bloco da transformada usualmente implementada pela transformada discreta do cosseno tridimensional (3D DCT) é substituído pela transformada tridimensional de baixa complexidade apresentada, que possui alta performance de concentração de energia de vídeos e imagens tridimensionais. Além da substituição do bloco da 3D DCT por uma outra transformada, o bloco de quantização é modificado de modo a tornar a complexidade da transformada apresentada baixa. O presente invento apresenta grande eficiência energética por fazer uso de uma matriz com coeficientes apenas em 0 e ± 1 . Isso implica em eliminação da complexidade multiplicativa da etapa de processamento do sinal tridimensional e torna desnecessária a utilização de aritmética de ponto flutuante. Este comportamento da presente dispositivo configura um dispositivo ambientalmente correto - livre tradução de eco-friendly ou green device.



SISTEMA DE RASTREAMENTO VISUAL BASEADO EM TRANSFORMADA TRIDIMENSIONAL DE BAIXA COMPLEXIDADE

→ CAMPO DE APLICAÇÃO

[01] A invenção proposta trata de um sistema de rastreamento visual de objetos em vídeos digitais, com alto grau de correlação espaço-temporal, como vídeos de câmeras de segurança e de sistemas de navegação autônoma. A transformada discreta do cosseno (DCT) é largamente empregada na área de compressão de imagem e vídeo, representando uma alternativa prática para a transformada de Karhunen-Loève (KLT), sendo esta o método ótimo de descorrelação, mas de difícil implementação prática. A transformada discreta tridimensional do cosseno (3D DCT) pode ser utilizada para compressão de imagens concatenadas, como uma sequência de quadros em um vídeo ou uma coleção de regiões próximas de uma imagem estática.

[02] Diversos sistemas de rastreamento visual aplicam análise de componente principal (PCA) como modelo de representação do objeto a ser rastreado. Matematicamente, a PCA é equivalente à KLT. Sendo assim, a DCT pode ser considerada para tais sistemas de rastreamento. Em [28], um sistema de rastreamento visual, baseado na 3D DCT, foi proposto como uma alternativa prática aos métodos de rastreamento visual, baseados em PCA. No entanto, tal método ainda apresenta as desvantagens de possuir alto custo computacional, sendo necessárias operações de multiplicação e de aritmética de ponto flutuante para o cálculo da 3D DCT, que demandam alto consumo de energia e recursos de *hardware*. A presente invenção se baseia na substituição do bloco da 3D DCT por um novo esquema de baixa complexidade, para computar a transformada tridimensional de bloco com alta capacidade de

compactação de energia, resultando num sistema de menor custo computacional.

→ ANTECEDENTE DA INVENÇÃO

[03] Vídeos digitais são matematicamente representados por *arrays* multidimensionais de *pixels* que, em geral, apresentam alto grau de correlação e podem ser modelados como processos markovianos de primeira ordem, com elevado coeficiente de correlação [29]. Nesse caso, a transformada discreta do cosseno (DCT) representa uma aproximação para a transformada de Karhunen-Loève (KLT) [30], que é a transformada ótima para descorrelação de dados [31]. O emprego da DCT é vantajoso pelo fato desta apresentar núcleo de transformação fixo, enquanto a KLT apresenta núcleo que depende de propriedades estatísticas do conjunto de dados de entrada. Em consequência, a DCT emerge como a principal ferramenta matemática em diferentes aplicações de compressão de dados [32].

[04] A DCT transforma um sinal discreto $\mathbf{x} = [x(0) \ x(1) \ \dots \ x(N - 1)]^T$ no sinal de saída $\mathbf{X} = [X(0) \ X(1) \ \dots \ X(N - 1)]^T$, de acordo com a seguinte expressão [33]:

$$X[k] \triangleq \alpha_N[k] \cdot \sum_{n=0}^{N-1} x[n] \cdot \cos\left(\frac{\pi(2n+1)k}{2N}\right),$$

$$k = 0, 1, \dots, N - 1, \quad (1)$$

em que

$$\alpha_N[k] \triangleq \sqrt{\frac{1}{N}} \cdot \begin{cases} 1, & \text{if } k = 0, \\ \sqrt{2}, & \text{otherwise.} \end{cases} \quad (2)$$

Um vídeo é geralmente representado por uma sequência de matrizes – *arrays* de três dimensões – em que cada elemento da matriz carrega

um valor inteiro, que representa a intensidade de cor para cada uma das cores, vermelha, verde e azul (RGB). Considerando a representação usual de 8 bits, tais valores de intensidade variam entre 0 a 255. Em escala de cinza, é considerado apenas um valor para cada elemento da matriz. Cada matriz representa um instante do vídeo e é denominada de quadro. Em [28], um método de rastreamento visual de um objeto previamente definido em um vídeo é proposto. Tal sistema considera a 3D DCT como modelo de representação do objeto rastreado. A 3D DCT de um sinal T com entradas $t[n_1, n_2, n_3]$, $n_i = 0, 1, \dots, N_i - 1$, para $i = 1, 2, 3$, é dado pelo sinal Y , cujas entradas são dadas por [28, 34]:

$$y[k_1, k_2, k_3] \triangleq \alpha_{N_1}[k_1] \cdot \alpha_{N_2}[k_2] \cdot \alpha_{N_3}[k_3] \cdot \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} \sum_{n_3=0}^{N_3-1} t[n_1, n_2, n_3] \cdot \cos\left(\frac{\pi(2n_1+1)k_1}{2N_1}\right) \cdot \cos\left(\frac{\pi(2n_2+1)k_2}{2N_2}\right) \cdot \cos\left(\frac{\pi(2n_3+1)k_3}{2N_3}\right),$$

$$k_i = 1, 2, \dots, N_i - 1, \quad i = 1, 2, 3.$$

(3)

[05] Vetores e matrizes podem ser modelados como tensores de primeira e segunda ordem, respectivamente [35]. De forma análoga, um sinal 3D pode ser entendido como um tensor de terceira ordem [36, 37, 28]. Um tensor de ordem R -ésima é, simplesmente, um *array* que requer R índices. Seja $A \in \mathbb{C}^{N_1 \times N_2 \times \dots \times N_R}$ um tensor de R -ésima ordem, cujas entradas são dadas por $a[n_1, n_2, \dots, n_R]$, em que \mathbb{C} é o conjunto dos números complexos, e $n_i = 0, 1, \dots, N_i - 1$, para $i = 1, 2, \dots, R$. O produto, em i -ésimo modo, de tensor A pela matriz $M \in \mathbb{C}^{H \times N_i}$ [38, página xxxv], denotado por $A \times_i M$, é definido como o tensor $B \in \mathbb{C}^{N_1 \times N_2 \times \dots \times N_{i-1} \times H \times N_{i+1} \times \dots \times N_R}$, cujas entradas são dadas por:

$$b[n_1, \dots, n_{i-1}, h, n_{i+1}, \dots, n_R] \triangleq \sum_{n_i=0}^{N_i-1} a[n_1, \dots, n_i, \dots, n_R] \cdot m[h, n_i],$$

(4)

em que $m[h, n_i]$ são as entradas de M e $h = 0, 1, \dots, H - 1$. O produto em i -ésimo modo generaliza o produto matricial, viabilizando operar

arrays de alta ordem com matrizes. A DCT de múltiplas dimensões pode ser expressa em termos de produtos em i -ésimo modo pela matriz de transformação da DCT. A DCT unidimensional em (1) pode ser expressa por $\mathbf{X} = \mathbf{x} \times_1 \mathbf{C}_N$, em que \mathbf{C}_N é a matriz da DCT, cujas entradas são dadas por:

$$c_N[k, n] = \alpha_N[k] \cdot \cos\left(\frac{\pi(2n+1)k}{2N}\right),$$

$$k, n = 0, 1, \dots, N-1. \quad (5)$$

e a 3D DCT em (3) pode ser escrita na forma [28]:

$$\mathcal{Y} = \mathcal{T} \times_1 \mathbf{C}_{N_1} \times_2 \mathbf{C}_{N_2} \times_3 \mathbf{C}_{N_3}. \quad (6)$$

Uma vez que a matriz da DCT é ortogonal, ou seja, $\mathbf{C}_N^{-1} = \mathbf{C}_N^T$, as transformadas inversas para a 1D DCT e 3D DCT são dadas, respectivamente, por:

$$\mathbf{x} = \mathbf{X} \times_1 \mathbf{C}_N^T \quad (7)$$

$$\mathcal{T} = \mathcal{Y} \times_1 \mathbf{C}_{N_1}^T \times_2 \mathbf{C}_{N_2}^T \times_3 \mathbf{C}_{N_3}^T. \quad (8)$$

[06] Após o estágio da transformação 3D, no esquema da Figura 3, os coeficientes obtidos são quantizados por um volume de quantização, e reorganizados de acordo com um esquema *zig-zag* 3D [39]. Seja $\mathbf{Q} \in \mathbb{C}^{N_1 \times N_2 \times N_3}$ o volume de quantização, tensor de terceira ordem, cujas entradas são dadas por $q[k_1, k_2, k_3]$, para $k_1, k_2, k_3 = 0, 1, \dots, N_i - 1, i = 1, 2, 3$. A etapa de quantização realiza a seguinte operação [40, 41]:

$$\hat{y}[k_1, k_2, k_3] = \text{round}\left(\frac{y[k_1, k_2, k_3]}{q[k_1, k_2, k_3]}\right),$$

$$k_1, k_2, k_3 = 0, 1, \dots, N_i - 1, \quad i = 1, 2, 3. \quad (9)$$

onde $y[k_1, k_2, k_3]$ são os coeficientes no domínio da frequência, de acordo com (3). Os coeficientes do volume de quantização devem ser projetados para preservar ou descartar informação de vídeo no domínio

da frequência, de acordo com algum critério de *trade-off* entre qualidade visual e taxa de compressão. O processo de desquantização é definido por $\hat{y}[k_1, k_2, k_3] = \tilde{y}[k_1, k_2, k_3] \cdot q[k_1, k_2, k_3]$, onde $\hat{y}[k_1, k_2, k_3]$ é uma estimativa de $y[k_1, k_2, k_3]$ [42].

[07] Algoritmos multidimensionais para a DCT frequentemente aplicam a propriedade de separação de núcleo, que permite o uso de algoritmos para computar transformadas de dimensões maiores [43, 44]. Esse método é, frequentemente, denominado *decomposição linha-coluna* [45]. Tendo em vista a grande demanda de aplicações, vários circuitos esquemáticos rápidos para a 1D DCT foram propostos [46, 47, 48, 49, 50, 51]. De fato, a complexidade multiplicativa mínima teórica, prevista por Heideman [52], foi atingida pelo algoritmo de Loeffler para o comprimento $N = 8$ [50]. Uma vez que se trata de um campo de pesquisa já bem desenvolvido, é improvável que novos algoritmos apresentem melhorias significativas em termos de complexidade computacional. Neste contexto, diferentes métodos aproximados foram visados para diminuir o custo do cômputo da DCT, tal como a DCT inteira [31, 53, 54, 55], DCTs aproximadas [56, 57, 58, 59, 60], algoritmos podados para a DCT [61, 62, 63], assim como técnicas combinadas [64, 65, 66, 67]. Em particular, DCTs aproximadas são transformadas que apresentam custo computacional muito baixo (em geral, o custo multiplicativo é nulo), enquanto preservam características importantes da transformação original, como compactação de energia e capacidade de descorrelação. Diferentemente da DCT original – também chamada de DCT exata –, as aproximações não estão sujeitas aos limites multiplicativos teóricos.

[08] Diversas aproximações para a DCT vêm sendo propostas nos últimos anos. Neste contexto, a 2D DCT aproximada vem sendo usada com sucesso em aplicações de compressão de imagem, obtendo

performance competitiva a um custo computacional consideravelmente menor [65, 66, 68]. Apesar de vasto espectro de aplicações, o projeto de aproximações para a DCT de múltiplas dimensões, e suas decorrentes aplicações, representam um campo industrial inexplorado.

→ SUMÁRIO DA INVENÇÃO

[09] A presente invenção propõe um dispositivo para o rastreamento de objetos em vídeos, baseado em uma transformada tridimensional proposta, que possui baixo custo computacional e forte poder de decorrelação espaço-temporal. O esquema se baseia em aplicar produto em i -ésimo de tensores de alta ordem por uma matriz aproximada para DCT, proposta em [59], derivando uma 3D DCT de tamanho $8 \times 8 \times 8$. Tal transformada possui a menor complexidade aritmética da literatura, com apenas 14 adições em 1D, e não necessita de operações com ponto flutuante. A aproximação para a 3D DCT proposta requer apenas 2688 adições, enquanto a 3D DCT exata necessita de 2112 multiplicações e 5568 adições, para o mesmo tamanho de transformada, derivada a partir do algoritmo de Loeffler [50].

[10] Os blocos das transformadas direta e inversa, no esquema de compactação e descompactação de imagem e vídeo, usualmente implementado pela 3D DCT e 3D DCT inversa (3D IDCT), respectivamente, são substituídos pelos métodos propostos nesta invenção. O bloco de quantização 3D e desquantização 3D também são modificados, de modo a explorar a baixa complexidade aritmética da transformação tridimensional. A invenção propõe um método industrializável relevante, diferindo dos métodos já existentes, se tratando de um dispositivo de rastreamento visual de um objeto, cujos blocos de transformada 3D, direta e inversa, da Figura 3, são baseados

em aproximações para a 3D DCT e 3D IDCT, que possuem menor complexidade computacional que os métodos já existentes.

→ BREVE DESCRIÇÃO DAS FIGURAS

[11] A Figura 1 mostra o algoritmo rápido referente para o cálculo da DCT unidimensional, que é empregado para computar a 3D DCT, de acordo com (14) e (15).

[12] A Figura 2 mostra o algoritmo rápido referente para o cálculo da DCT inversa unidimensional, que é empregado para computar a 3D DCT inversa, de acordo com (13) e (23).

[13] A Figura 3 apresenta esquema do dispositivo proposto para o rastreamento visual de um objeto em um vídeo, cujo modelo de representação do objeto é baseado na aproximação para 3D DCT proposta.

→ DESCRIÇÃO DETALHADA

[14] O novo esquema proposto para rastreamento visual é apresentado a seguir. O esquema se baseia em uma nova transformada tridimensional, aqui proposta. A presente invenção difere significativamente de sistemas já propostos em trabalhos prévios.

[15] Em US 20030012408 A1 [69], é proposto um método de rastreamento facial em uma sequência de imagens. O sistema propõe o uso de PCA como modelagem do objeto rastreado. Dessa forma, tal sistema difere da presente invenção, que se baseia em uma aproximação para a 3D DCT proposta.

[16] A invenção US 20150304634 A1 [70] consiste em um sistema de rastreamento espaço-temporal de objetos, baseado em PCA, para modelagem de objeto. A presente invenção aplica uma aproximação para a 3D DCT proposta, que apresenta baixo consumo energético.

[17] Em US 20030161500 A1 [71], é proposto um sistema de rastreamento de objetos, baseado em exemplares probabilísticos, diferindo da presente invenção, que se baseia em uma aproximação proposta para a 3D DCT.

[18] O sistema US 20050185826 A1 [72] de rastreamento de objetos considera PCA para modelagem de componentes do objeto. A presente invenção, por sua vez, considera uma nova aproximação para a 3D DCT, como modelo de treinamento.

[19] Em US 20080187174 A1 [73], é proposto um sistema de rastreamento de objetos em imagens, que considera PCA para modelagem de características de treinamento. O invento proposto na presente patente propõe uma aproximação para a 3D DCT de baixa complexidade, para modelagem 3D do objeto rastreado.

[20] Um sistema de rastreamento é proposto em US 5570094 A [74]. Tal sistema é baseado em radar de efeito Doppler. A presente invenção se baseia em uma aproximação para a 3D DCT proposta.

[21] Em US 7574020 B2 [75], é proposto um sistema de rastreamento de objetos em imagens, que se baseia no cômputo da PCA, diferindo da presente invenção, que aplica uma aproximação para a 3D DCT proposta, como base para modelagem do objeto rastreado.

[22] Em US 20150055821 A1 [76], é proposto um sistema de rastreamento de objetos que emprega a PCA como extração de características do objeto. A presente invenção se baseia em uma nova aproximação para a 3D DCT, de baixo custo computacional.

[23] O esquema proposto se baseia em aplicar álgebra tensorial, combinado com a matriz aproximada DCT, proposta em [59], e obter uma aproximação tridimensional para a DCT de tamanho $8 \times 8 \times 8$. Tal transformada tridimensional apresenta custo multiplicativo nulo, substitui a 3D DCT exata no esquema de rastreamento visual de

objetos em vídeo, proposto em [28], substituindo o bloco da 3D DCT. Um novo design para o bloco de quantização e desquantização multidimensionais é proposto, de modo a explorar as propriedades de baixa complexidade das transformadas propostas.

[24] Seja a matriz de transformação de baixa complexidade

$$\mathbf{T} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & -1 & -1 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 0 & -1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 \end{bmatrix}, \quad (10)$$

que corresponde a uma aproximação para a DCT dada por $\hat{\mathbf{C}} = \mathbf{S} \cdot \mathbf{T}$, em que $\mathbf{S} = \text{diag}(1/\sqrt{8}, 1/\sqrt{2}, 1/2, 1/\sqrt{2}, 1/\sqrt{8}, 1/\sqrt{2}, 1/2, 1/\sqrt{2})$.

Definimos uma aproximação 3D de um sinal $T \in \mathbb{C}^{N_1 \times N_2 \times N_3}$, representado por tensor de terceira ordem, o sinal tridimensional de saída dado por

$$\begin{aligned} \mathcal{Y} &= \mathcal{T} \times_1 \hat{\mathbf{C}} \times_2 \hat{\mathbf{C}} \times_3 \hat{\mathbf{C}} \\ &= \mathcal{T} \times_1 (\mathbf{S} \cdot \mathbf{T}) \times_2 (\mathbf{S} \cdot \mathbf{T}) \times_3 (\mathbf{S} \cdot \mathbf{T}). \end{aligned} \quad (11)$$

Considerando propriedades algébricas do produto em i -ésimo modo [36, 35], verifica-se que a propriedade da separabilidade é preservada. Além disso, pode-se rearranjar (11) da seguinte forma:

$$\mathcal{Y} = \mathcal{T} \times_1 \mathbf{T} \times_2 \mathbf{T} \times_3 \mathbf{T} \times_1 \mathbf{S} \times_2 \mathbf{S} \times_3 \mathbf{S}. \quad (12)$$

Sendo $\hat{\mathbf{C}}$ uma matriz de transformação ortogonal, e $\mathbf{S} = \mathbf{S}^T$ uma matriz diagonal, tem-se que $\hat{\mathbf{C}}^{-1} = \hat{\mathbf{C}}^T = \mathbf{T}^T \cdot \mathbf{S}$. Consequentemente, a transformação inversa para (12) é dada por:

$$\begin{aligned} \mathcal{T} &= \mathcal{Y} \times_1 \hat{\mathbf{C}}_N^{-1} \times_2 \hat{\mathbf{C}}_N^{-1} \times_3 \hat{\mathbf{C}}_N^{-1} \\ &= \mathcal{Y} \times_1 (\mathbf{S} \cdot \mathbf{T})^T \times_2 (\mathbf{S} \cdot \mathbf{T})^T \times_3 (\mathbf{S} \cdot \mathbf{T})^T \\ &= \mathcal{Y} \times_1 \mathbf{S} \times_2 \mathbf{S} \times_3 \mathbf{S} \times_1 \mathbf{T}^T \times_2 \mathbf{T}^T \times_3 \mathbf{T}^T. \end{aligned} \quad (13)$$

[25] A 3D DCT aproximada proposta pode ser computada em duas etapas separadas: a primeira, envolvendo os produtos em i -ésimo modo com as matrizes de baixa complexidade \mathbf{T} , e, na etapa subsequente,

pelas matrizes S , que apresentam entradas dadas por números irracionais. Sugerimos um método de embutir tais valores irracionais na etapa subsequente de quantização, de modo que a matriz S não introduza complexidade adicional. Definimos o sinal seguinte, tensor de terceira ordem, intermediário:

$$\mathcal{A} = \mathcal{T} \times_1 \mathbf{T} \times_2 \mathbf{T} \times_3 \mathbf{T}, \quad (14)$$

Para cada produto em i -ésimo modo, e mantendo os demais índices das outras dimensões fixos, temos o equivalente a um produto matricial unidimensional. Sejam $t[n_1, n_2, n_3]$ e $a[k_1, n_2, n_3]$ as entradas dos sinais T e A , respectivamente, para $n_1, k_1 = 0, 1, \dots, 7$, fixando n_2 e n_3 . Então, $A = T \times_1 T$ pode ser representado pelo produto

$$\begin{bmatrix} X[0] \\ X[1] \\ \vdots \\ X[7] \end{bmatrix} = \mathbf{T} \cdot \begin{bmatrix} x[0] \\ x[1] \\ \vdots \\ x[7] \end{bmatrix},$$

$$X[k_1] \triangleq a[k_1, n_2, n_3] \quad x[n_1] \triangleq t[n_1, n_2, n_3]. \quad (15)$$

O produto (15) pode ser calculado pelo circuito esquemático rápido mostrado na Figura 1. Ao variar os valores de n_2 e n_3 , é necessário $N^2 = 8^2$ cálculos da expressão (15), e, portanto, tal quantidade de aplicações do circuito esquemático rápido da Figura 1. Para os demais produtos \times_2 e \times_3 , o desenvolvimento é análogo, totalizando $3 \cdot N^2 = 3 \cdot 8^2 = 192$ aplicações do circuito esquemático rápido. Sendo assim, a expressão (14) aproximada proposta pode ser computada através do circuito esquemático rápido unidimensional, apresentado na Figura 1, que demanda 14 adições individualmente, totalizando $192 \cdot 14 = 2688$ adições para o cálculo da 3D DCT aproximada completa.

[26] Para obter a 3D DCT aproximada proposta (12), é necessário realizar os produtos, em i -ésimo modo, do sinal A pela matriz diagonal S , de acordo com:

$$\mathcal{Y} = \mathbf{A} \times_1 \mathbf{S} \times_2 \mathbf{S} \times_3 \mathbf{S}. \quad (16)$$

Como S é uma matriz diagonal, é possível demonstrar que:

$$y[k_1, k_2, k_3] = a[k_1, k_2, k_3] \cdot d_{k_1} \cdot d_{k_2} \cdot d_{k_3}, \quad (17)$$

em que $a[k_1, k_2, k_3]$ são as entradas do sinal A dadas em (14), e d_k é o k -ésimo elemento diagonal de S . O cálculo da expressão (17) pode ser embutido na etapa subsequente de quantização. De (17) em (9), com $N_1 = N_2 = N_3 = 8$, obtemos:

$$\tilde{y}[k_1, k_2, k_3] = \text{round} \left(\frac{a[k_1, k_2, k_3] \cdot d_{k_1} \cdot d_{k_2} \cdot d_{k_3}}{q[k_1, k_2, k_3]} \right), \\ k_1, k_2, k_3 = 0, 1, \dots, 7. \quad (18)$$

Propomos um novo volume de quantização $Q^* \in \mathbb{C}^{8 \times 8 \times 8}$, cujas entradas são dadas por:

$$q^*[k_1, k_2, k_3] \triangleq \frac{q[k_1, k_2, k_3]}{d_{k_1} \cdot d_{k_2} \cdot d_{k_3}}. \quad (19)$$

Substituindo (19) em (18), obtemos o volume de quantização modificado, como se segue:

$$\tilde{y}[k_1, k_2, k_3] = \text{round} \left(\frac{a[k_1, k_2, k_3]}{q^*[k_1, k_2, k_3]} \right), \\ k_1, k_2, k_3 = 0, 1, \dots, 7. \quad (20)$$

Apenas o sinal A é necessário ser computado. Tal sinal é aplicado como entrada ao volume de quantização modificado (20). Uma vez que o cômputo de A requer apenas a transformação 3D de baixa complexidade (14), o excedente computacional imposto pela matriz diagonal S é descartado. O procedimento de quantização inversa modificado é

obtido de forma análoga, resultando num tensor \hat{A} , cujas entradas são dadas por:

$$\hat{a}[k_1, k_2, k_3] = \tilde{y}[k_1, k_2, k_3] \cdot q^\dagger[k_1, k_2, k_3], \quad (21)$$

em que

$$q^\dagger[k_1, k_2, k_3] \triangleq q[k_1, k_2, k_3] \cdot d_{k_1} \cdot d_{k_2} \cdot d_{k_3} \quad (22)$$

são as entradas do volume de quantização inverso modificado Q^\dagger . De maneira análoga ao caso direto, na presente invenção, apenas as matrizes de baixa complexidade são empregadas na etapa inversa, de acordo com

$$\hat{T} = \hat{A} \times_1 \mathbf{T}^\top \times_2 \mathbf{T}^\top \times_3 \mathbf{T}^\top, \quad (23)$$

em que \hat{T} é um bloco recuperado do vídeo comprimido. A transformação inversa também pode ser aplicada com a decomposição linha-coluna, fazendo uso do algoritmo rápido 1D, mostrado na Figura 2, analogamente ao caso direto.

[27] O diagrama esquemático do dispositivo é mostrado na Figura 3. No índice de tempo t , o sistema recebe, como entradas, o novo quadro do vídeo $T(t)$, a posição anterior do objeto rastreado $Z(t - 1)$, e amostras de treinamento prévias, positivas e negativas, dadas por amostras de dimensão 8×8 , de regiões próximas $Z^+(t - 1)$ e de regiões mais distantes $Z^-(t - 1)$, respectivamente. São amostradas e escalonadas para dimensão 8×8 V regiões candidatas $Z^*(t, v)$, para $v = 0, 1, \dots, V - 1$. Para cada região candidata $A(v) = Z^*(t, v)$, os sete vizinhos mais próximos nos conjuntos $Z^+(t - 1)$ e $Z^-(t - 1)$ são obtidos e concatenados com a região candidata, obtendo blocos de $8 \times 8 \times 8$, dados por $T^+(t, v)$ e $T^-(t, v)$. Para cada bloco, é calculado a 3D DCT aproximada proposta, e os coeficientes são quantizados com o volume de quantização modificado proposto, de acordo com (14) e (19),

respectivamente, resultando nos blocos quantizados $\tilde{Y}^+(v)$ e $\tilde{Y}^-(v)$. A seguir, o procedimento inverso é realizado. Após desquantização e 3D DCT aproximada inversa, dados por (22) e (23), respectivamente, são reconstruídas representações de $T^+(t, v)$ e $T^-(t, v)$, dadas por $\tilde{T}^+(t, v)$ e $\tilde{T}^-(t, v)$, cujas entradas são, respectivamente, dadas por $\tilde{t}_v^+(n_1, n_2, n_3)$ e $\tilde{t}_v^-(n_1, n_2, n_3)$, para $n_1, n_2, n_3 = 0, 1, 2, 3, 4, 5, 6, 7$. As matrizes $\tilde{A}^+(v) = \tilde{t}_v^+(n_1, n_2, 7)$ e $\tilde{A}^-(v) = \tilde{t}_v^-(n_1, n_2, 7)$ são representações aproximadas para a região candidata $A(v)$. A partir das diferenças quadráticas $\|A(v) - \tilde{A}^+(v)\|^2$ e $\|A(v) - \tilde{A}^-(v)\|^2$, são calculadas as verossimilhanças $L^+(v)$ e $L^-(v)$. Por fim, tais métricas são combinadas em $L(v) = \rho(L^+(v) - \gamma L^-(v))$, em que γ é um fator de ponderação, e $\rho(x) = 1/[1+\exp(-x)]$. A região candidata $T(t)$ ótima é, então, obtida maximizando $L(v)$. Novas amostras de treino $Z^+(t)$ e $Z^-(t)$ são selecionadas. Como dados de saída, o sistema fornece $T(t)$, $Z^+(t)$ e $Z^-(t)$, representando o estado do objeto no quadro de índice t .

→ REFERÊNCIAS

[28] Xi Li, A. Dick, C. Shen, A. van den Hengel, e H. Wang. **Incremental learning of 3D-DCT compact representations for robust visual tracking.** *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35:863-881, 2013.

[29] K. R. Rao e P. Yip. **The Transform and Data Compression Handbook.** CRC Press LLC, 2001.

[30] R. J. Clarke. **Relation between the Karhunen-Loève and cosine transforms.** *IEE Proceedings F Communications, Radar and Signal Processing*, 128:359-360, 1981.

[31] V. Britanak, P. Yip, e K. R. Rao. **Discrete Cosine and Sine Transforms.** Academic Press, 2007.

[32] K. R. Rao e P. Yip. **Discrete Cosine Transform: Algorithms, Advantages, Applications**. Academic Press, San Diego, CA, 1990.

[33] N. Ahmed, T. Natarajan, e K. R. Rao. **Discrete cosine transform**. *IEEE Transactions on Computers*, C-23:90-93, 1974.

[34] S. Boussakta. **Fast algorithm for the 3-D DCT-II**. *IEEE Transactions on Signal Processing*, 52:992-1001, 2004.

[35] L. De Lathauwer, B. De Moor, e J. Vandewalle. **On the best rank-1 and rank-(r_1, r_2, \dots, r_n) approximation of higher-order tensors**. *SIAM Journal on Matrix Analysis and Applications*, 21(4):1324-1342, 2000.

[36] L. De Lathauwer e B. De Moor. **From matrix to tensor: Multilinear algebra and signal processing**. In *Institute of Mathematics and Its Applications Conference Series*, volume 67, páginas 1-16. Citeseer, 1998.

[37] D. G. Northcott. **Multilinear algebra**. Cambridge University Press, 2008.

[38] D. S. Bernstein. **Matrix Mathematics: Theory, Facts, and Formulas**. Princeton University Press, 2009.

[39] Y.-L. Chan e W.-C. Siu. **Variable temporal-length 3-D discrete cosine transform coding**. *IEEE Transactions on Image Processing*, 6(5):758-763, 1997.

[40] M. C. Lee, Raymond K. W. Chan, e D. A. Adjeroh. **Quantization of 3D-DCT coefficients and scan order for video compression**. *Journal of Visual Communication and Image Representation*, 8(4):405-422, 1997.

[41] R. K. W. Chan e M. C. Lee. **3D-DCT quantization as a compression technique for video sequences**. *Proceedings of the International Conference on Virtual Systems and MultiMedia*, páginas 188-196, 1997.

- [42] V. Bhaskaran e K. Konstantinides. **Image and Video Compression Standards**. Kluwer Academic Publishers, 1997.
- [43] R. E. Blahut. **Fast Algorithms for Signal Processing**. Cambridge University Press, 2010.
- [44] Y. Zeng, G. Bi, e A. Rahim Leyman. **New polynomial transform algorithm for multidimensional DCT**. *IEEE Transactions on Signal Processing*, 48(10):2814-2821, 2000.
- [45] V. K. Madisetti. **The Digital Signal Processing Handbook**. CRC Press, 2009.
- [46] W. H. Chen, C. Smith, e S. Fralick. **A fast computational algorithm for the discrete cosine transform**. *IEEE Transactions on Communications*, 25:1004-1009, 1977.
- [47] B. G. Lee. **A new algorithm for computing the discrete cosine transform**. *IEEE Transactions on Acoustics, Speech and Signal Processing*, ASSP-32:1243-1245, 1984.
- [48] Z. Wang. **Fast algorithms for the discrete W transform and for the discrete Fourier transform**. *IEEE Transactions on Acoustics, Speech and Signal Processing*, ASSP-32:803-816, 1984.
- [49] H. S. Hou. **A fast recursive algorithm for computing the discrete cosine transform**. *IEEE Transactions on Acoustic, Signal, and Speech Processing*, 6:1455-1461, 1987.
- [50] C. Loeffler, A. Ligtenberg, e G. S. Moschytz. **Practical fast 1-D DCT algorithms with 11 multiplications**. *ICASSP International Conference on Acoustics, Speech, and Signal Processing*, 2:988-991, 1989.
- [51] E. Feig e S. Winograd. **Fast algorithms for the discrete cosine transform**. *IEEE Transactions on Signal Processing*, 40:2174-2193, 1992.

[52] M. T. Heideman. **Multiplicative Complexity, Convolution, and the DFT.** *Signal Processing and Digital Filtering.* Springer-Verlag, 1988.

[53] Y.-J. Chen, S. Oraintara, e T. Nguyen. **Video compression using integer DCT.** In *Proceedings of International Conference on Image Processing*, volume 2, páginas 844-845. IEEE, 2000.

[54] Y. Yokotani, R. Geiger, G. D. T. Schuller, S. Oraintara, e K. R. Rao. **Lossless audio coding using the intMDCT and rounding error shaping.** *IEEE Transactions on Audio, Speech, and Language Processing*, 14(6):2201-2211, 2006.

[55] L. O. Hnativ. **Integer cosine transforms: Methods to construct new order 8, 16 fast transforms and their application.** *Cybernetics and Systems Analysis*, 50(6):913-929, 2014.

[56] T. I. Haweel. **A new square wave transform based on the DCT.** *Signal Processing*, 82:2309-2319, 2001.

[57] K. Lengwehasatit e A. Ortega. **Scalable variable complexity approximate forward DCT.** *IEEE Transactions on Circuits and Systems for Video Technology*, 14:1236-1248, 2004.

[58] R. J. Cintra e F. M. Bayer. **A DCT approximation for image compression.** *IEEE Signal Processing Letters*, 18:579-582, 2011.

[59] F. M. Bayer e R. J. Cintra. **DCT-like transform for image compression requires 14 additions only.** *Electronics Letters*, 48:919-921, 2012.

[60] S. Bouguezel, M. O. Ahmad, e M. N. S. Swamy. **Low-complexity 8×8 transform for image compression.** *Electronics Letters*, 44:1249-1250, 2008.

[61] Z. Wang. **Pruning the fast discrete cosine transform.** *IEEE Transactions on Communications*, 39:640-643, 1991.

[62] L. Makkaoui, V. Lecuire, e J. Moureaux. **Fast zonal DCT-based image compression for wireless camera sensor networks.** In

2nd International Conference on Image Processing Theory Tools and Applications (IPTA), páginas 126-129. IEEE, 2010.

[63] V. Lecuire, L. Makkaoui, e J.-M. Moureaux. **Fast zonal DCT for energy conservation in wireless image sensor networks.** *Electronics Letters*, 48:125-127, 2012.

[64] N. Kouadria, N. Doghmane, D. Messadeg, e S. Harize. **Low complexity DCT for image compression in wireless visual sensor networks.** *Electronics Letters*, 49:1531-1532, 2013.

[65] V. A. Coutinho, R. J. Cintra, F. M. Bayer, S. Kulasekera, e A. Madanayake. **A multiplierless pruned DCT-like transformation for image and video compression that requires ten additions only.** *Journal of Real-Time Image Processing*, 12(2):247-255, 2016.

[66] R. J. Cintra, F. M. Bayer, V. A. Coutinho, S. Kulasekera, A. Madanayake, e A. Leite. **Energy-efficient 8-point DCT approximations: Theory and hardware architectures.** *Circuits, Systems, and Signal Processing*, 35(11):4009-4029, 2016.

[67] K. Mechouek, N. Kouadria, N. Doghmane, e N. Kaddeche. **Low complexity DCT approximation for image compression in wireless image sensor networks.** *Journal of Circuits, Systems and Computers*, 2016.

[68] U. S. Potluri, A. Madanayake, R. J. Cintra, F. M. Bayer, S. Kulasekera, e A. Edirisuriya. **Improved 8-point approximate DCT for image and video compression requiring only 14 additions.** *IEEE Transactions on Circuits and Systems I: Regular Papers*, PP:1-14, 2014.

[69] J.-Y. Bouguet, R. Grzeszczuk, e S. Gokturk. **Method and system using a data-driven model for monocular face tracking**, 2001. US 20030012408 A1.

[70] J. G. Karvounis. **Mapping and tracking system**, 2011. US 20150304634 A1.

[71] A. Blake e K. Toyama. System and method for probabilistic exemplar-based pattern tracking, 2002. US 20030161500 A1.

[72] B. Georgescu, X. Zhou, D. Comaniciu, e R. Rao. Method and system for multi-modal component-based tracking of an object using robust information fusion, 2004. US 20050185826 A1.

[73] D. Metaxas e A. Kanaujia. System and method for detecting and tracking features in images, 2006. US 20080187174 A1.

[74] B. S. R. Armstrong. Three dimensional tracking by array doppler radar, 1995. US 5570094 A.

[75] A. Shamaie. Detecting and tracking objects in images, 2005. US 7574020 B2.

[76] D. A. Fotland. Multi-tracker object tracking, 2013. US 20150055821 A1.

REIVINDICAÇÕES

1. “Sistema de Rastreamento Visual Baseado em Transformada de Baixa Complexidade” caracterizado por computar transformada tridimensional sem operação de ponto flutuante, em que a matriz \mathbf{T} em (10) aplicada a expressão (12) e (14) no bloco de Transformação 3D no esquema da Figura 3.
2. “Sistema de Rastreamento Visual Baseado em Transformada de Baixa Complexidade” caracterizado por computar transformada tridimensional sem operação de ponto flutuante, em que o circuito esquemático rápido da Figura 1 é aplicado sucessivamente em decomposição linha-coluna no bloco de Transformação 3D no esquema da Figura 3.
3. “Sistema de Rastreamento Visual Baseado em Transformada de Baixa Complexidade” de acordo com as reivindicações 1 e 2, caracterizado por computar transformada tridimensional inversa sem operação de ponto flutuante, em que a matriz \mathbf{T}^T em (10) aplicada a expressão (13) e (23) no bloco de Transformação 3D inversa no esquema da Figura 3.
4. “Sistema de Rastreamento Visual Baseado em Transformada de Baixa Complexidade” de acordo com as reivindicações 1 e 2, caracterizado por computar transformada tridimensional inversa sem operação de ponto flutuante, em que o circuito esquemático rápido da Figura 2 é aplicado sucessivamente em decomposição linha-coluna no bloco de Transformação 3D inversa no esquema da Figura 3.
5. “Sistema de Rastreamento Visual Baseado em Transformada de Baixa Complexidade” de acordo com as reivindicações 1, 2, 3 e 4, caracterizado por substituir o bloco de quantização usual pela quantização modificada proposta, cujas entradas são dadas em (19).
6. “Sistema de Rastreamento Visual Baseado em Transformada de Baixa Complexidade” de acordo com as reivindicações 1, 2, 3 e 4, caracterizado por substituir o bloco de desquantização usual pela desquantização modificada proposta, cujas entradas são dadas em (22).

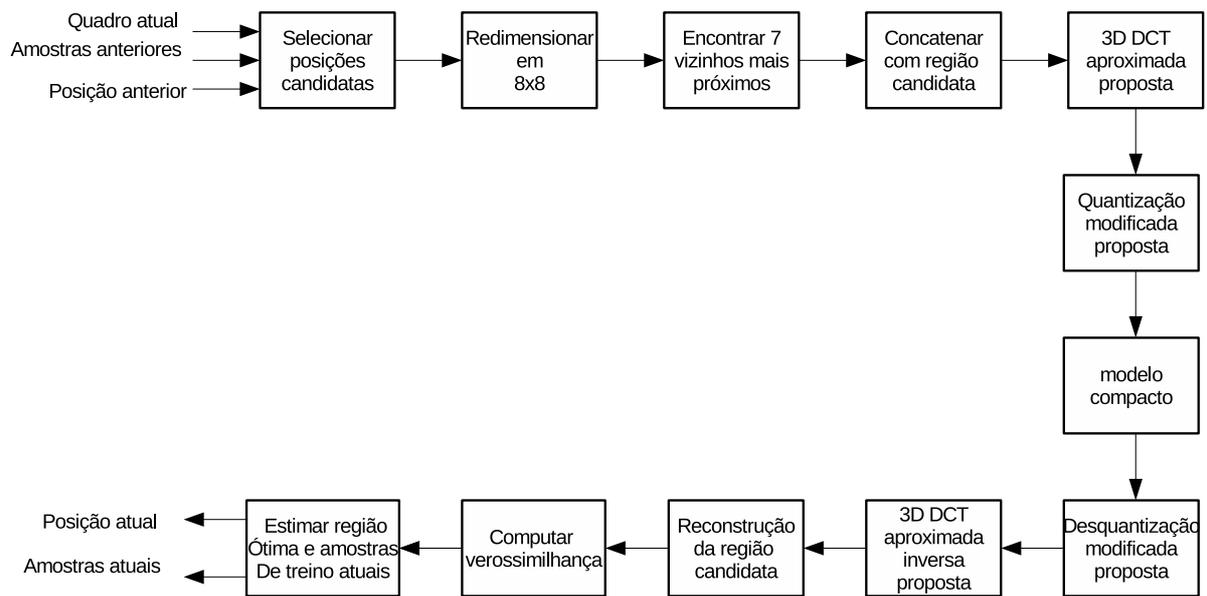


Figura 3

RESUMO

SISTEMA DE RASTREAMENTO VISUAL BASEADO EM TRANSFORMADA TRIDIMENSIONAL DE BAIXA COMPLEXIDADE

Esta invenção se trata de um dispositivo de rastreamento de objeto em movimento em um vídeo digital de reduzido consumo energético. O método pode ser empregado em diferentes aplicações, tais como vigilância digital, navegação autônoma, interação homem-robô e realidade aumentada. Este esquema é fundamentado em uma nova transformação tridimensional baseada em uma matriz de transformação associada a uma aproximação para a DCT de baixo custo computacional. O bloco da transformada usualmente implementada pela transformada discreta do cosseno tridimensional (3D DCT) é substituído pela transformada tridimensional de baixa complexidade apresentada, que possui alta performance de concentração de energia de vídeos e imagens tridimensionais. Além da substituição do bloco da 3D DCT por uma outra transformada, o bloco de quantização é modificado de modo a tornar a complexidade da transformada apresentada baixa. O presente invento apresenta grande eficiência energética por fazer uso de uma matriz com coeficientes apenas em 0 e ± 1 . Isso implica em eliminação da complexidade multiplicativa da etapa de processamento do sinal tridimensional e torna desnecessária a utilização de aritmética de ponto flutuante. Este comportamento da presente dispositivo configura um dispositivo ambientalmente correto - livre tradução de eco-friendly ou green device.