

К УСКОРЕНИЮ АНАЛИЗА ДВИЖЕНИЯ ПРИ КОДИРОВАНИИ ДИНАМИЧЕСКИХ ИЗОБРАЖЕНИЙ

Дворкович В.П., Дворкович А.В., Соколов А.Ю.

Научно-Исследовательский Институт Радио
Ул. Казакова, 16, Москва 103064

Аннотация: Приведены результаты по ускорению процедуры анализа движения, являющейся одной из наиболее затратных при кодировании динамических изображений по стандарту MPEG-1/2. В предлагаемом способе достигается ускорение поиска векторов движения в 16 раз при потере в коэффициенте сжатия не более 2%.

Кодирование динамических изображений согласно стандартам сжатия MPEG-1/2 [1,2], MPEG-4 [3], H.261 [4] и H.263 [5,6] в качестве основного элемента включает межкадровую компенсацию движения. Использование межкадровой компенсации движения позволяет достигнуть большего коэффициента сжатия (в 2-3 раза), чем без ее использования. Стандартным алгоритмом компенсации движения является блочный метод, в котором компенсация движения выполняется по макроблокам. Для этого текущий кадр разбивается на прямоугольные элементы фиксированного размера (макроблоки), обычно размер макроблока составляет 16x16 пикселей для яркостной компоненты и 8x8 пикселей для цветоразностных компонент, учитывая тот факт, что пространственное разрешение цветоразностного сигнала вдвое меньше разрешения сигнала яркости. Каждому макроблоку ставится в соответствие вектор движения $(\vec{V} = (V_x, V_y))$, характеризующий положение данного макроблока в предыдущем кадре:

$$H = \sum_{x,y=1,16} |C(x, y, t)| \equiv \sum_{x,y=1,16} |F(x, y, t) - F(x - V_x, y - V_y, t - \Delta t)| \quad (1)$$

Здесь F - значение яркости, (x, y) - пространственные координаты каждого пикселя макроблока, отсчитываемые от его левого верхнего угла, t - временной индекс кадра, суммирование производится по всем пикселям макроблока. Значения C в (1) представляют собой разность между текущими значениями яркости макроблока и значениями, соответствующими положению макроблока в предыдущем кадре, сдвинутыми на вектор движения. Если изменения от кадра к кадру сводятся, главным образом, к пространственному смещению участков изображения (что практически имеет место), величина C мала. В этом случае для достижения большего сжатия выгоднее передавать разностный сигнал C вместо исходного сигнала Y . Величина H в (1) имеет смысл нормы (мощности) сигнала после компенсации движения. Компенсация движения тем эффективнее, чем меньше значение H . Аналогичная схема используется и для цветоразностных сигналов. Следует также отметить, что в MPEG-1/2 стандарте существует возможность кодирования макроблока с использованием двух векторов движения, в этом случае передается разность между исходным сигналом и интерполяционным значением, вычисленным с использованием двух векторов и двух опорных кадров. Кроме того, в MPEG-1/2 конечная точность векторов движения может составлять 0.5 пикселя. При этом значение сигнала в опорном кадре $F(x - V_x, y - V_y, t - \Delta t)$, соответствующее нецелому значению вектора, вычисляется с использованием интерполяции между соседними пикселями.

Существуют различные алгоритмы поиска векторов движения. Наиболее простым и наиболее затратным является алгоритм перебора всех возможных значений векторов в некотором окне поиска $\min < V_x, V_y < \max$. Значение \vec{V} , для которого норма H в (1) имеет наименьшее значение, принимается за искомый вектор. Предполагая, что размер окна поиска равен $\pm N$ пикселей по координатам x и y , получим для числа операций, необходимых для определения вектора движения одного макроблока размером 16x16 пикселей, величину порядка $3 \cdot 256 (2N + 1)^2$. На один пиксел макроблока количество операций составляет $3 \cdot (2N + 1)^2$, что уже при типичном значении $N = 15$ (вектор движения в пределах ± 15 точек) составляет значительную величину более 10^3 операций / пиксел. Данный способ дает наилучший результат, но является сложным с точки зрения объема вычислений. Поэтому он обычно используется в качестве эталонного для оценки качества других способов поиска векторов движения.



(a)



(б)

Рис. 1. Текущий (а) и опорный (б) кадры видеопоследовательности «Сад цветов».

Одним из наиболее эффективных способов поиска векторов является способ, разработанный в [7]. В этом способе суммирование в (1) заменяется на суммирование лишь по небольшому числу пикселей макроблока. Идея метода заключается в выборе небольшого числа пикселей макроблока, которые наилучшим образом характеризуют изменение рельефа сигнала яркости макроблока. На Рис. 1 приведены два кадра из последовательности «Сад цветов», а на Рис.2 – один макроблок и соответствующий ему участок изображения в опорном кадре. На Рис.3 показан рельеф сигнала яркости в выбранном макроблоке и таблица значений яркости в каждом пикселе.

Согласно методу характерных пикселей макроблок размером 16x16 пикселей делится на 16 участков (Рис.4), в каждом из которых выбирается только один пиксел, имеющий максимальное, либо минимальное значение. Максимумы и минимумы чередуются в шахматном порядке, как показано на Рис.4. В дальнейшем только характерные пиксели используются для расчета контрольной суммы макроблока:

$$H1 = \sum_{(x,y) \text{--координаты выбранных точек в соответствующих участках макроблока}} |F(x, y, t) - F(x - V_x, y - V_y, t - \Delta t)| \quad (2)$$

С использованием 16 характерных пикселей расчет контрольной суммы (2), по сравнению с суммой, в которую входят все 256 пикселя, ускоряется в 16 раз.

Следует отметить, что, несмотря на свою простоту, изложенный алгоритм обладает высокой эффективностью. При значительном ускорении поиска векторов движения (в 16 раз), потери сжатия за счет ошибок в поиске векторов не превышают нескольких процентов (см. Табл. 1).

Таблица 1. Размер кода MPEG-2 (байт) для последовательности "Сад цветов", 97 кадров с разрешением 640x480 пикселей.

Конечная точность векторов движения	Предлагаемый алгоритм с использованием 16 характеристических пикселей (2)	Алгоритм с использованием всех 256 пикселей макроблока (1)
0.5 пикселя	3295558	3242122
1 пиксел	3699264	3545362

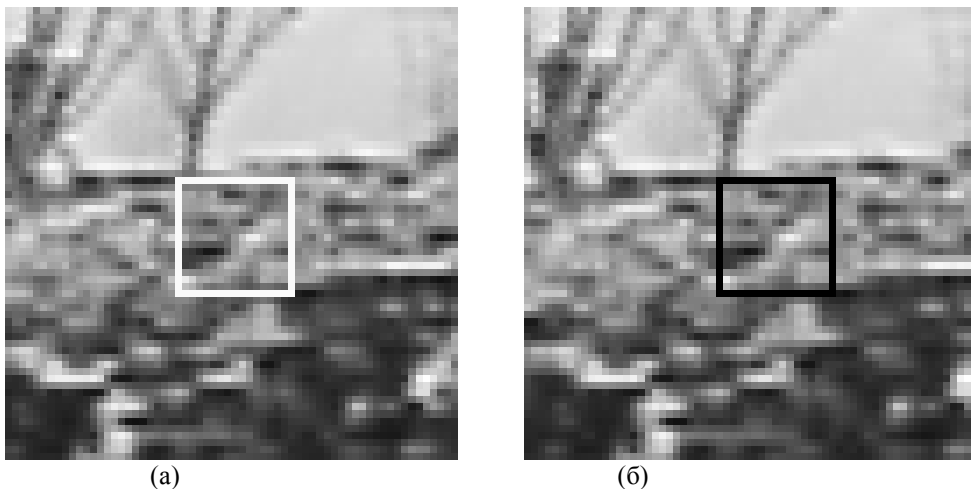
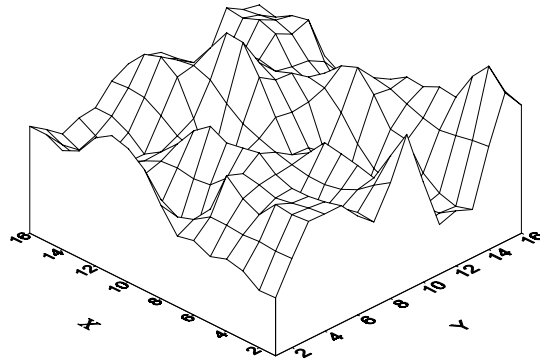


Рис. 2. Увеличенное изображение одного макроблока из текущего кадра (а - помечен белой рамкой) и соответствующий ему участок изображения в опорном кадре (б - помечен черной рамкой).

Литература:

[1] ISO/IEC 11172-2. Information Technology - Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 Mbit/s. Part 2: Video./Ed.1 JTS I/SC 29, 1993.
 [2] ISO/IEC 13818-2. Information Technology - Generic Coding of Moving Pictures and Associated Audio Information. Part 2: Video./ Ed.1 JTS I/SC 29, 1994.
 [3] ISO/IEC 14496. Information Technology - MPEG-4 Overview (San-Jose Version), Feb. 1998.
 [4] "Line transmission of non-telephone signals/ video codec for audiovisual services at px64 Kbits", ITU-T Recommendation H.261, March 1993.
 [5] "Line transmission of non-telephone signals/ video coding for low bit rate communication", ITU-T Recommendation H.263, March 1996.
 [6] "Line transmission of non-telephone signals/ video coding for low bit rate communication", ITU-T Recommendation H.263, Jan. 1998.
 [7] А.В. Дворкович, В.П. Дворкович, Ю.Б. Зубарев, А.Ю. Соколов, «Способ анализа векторов движения деталей в динамических изображениях», Патент РФ № 2137194 от 15.07.98.



(а)

y \ x	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	82	81	97	102	94	105	100	132	171	189	196	181	151	140	151	151
2	126	125	127	125	127	127	112	104	117	137	150	144	125	117	125	135
3	167	165	168	142	153	174	115	62	64	87	118	121	86	84	106	117
4	166	163	158	154	161	165	147	126	123	127	125	114	101	103	115	128
5	166	160	156	156	164	183	178	174	186	168	138	115	100	119	131	131
6	140	140	135	128	125	132	145	153	147	128	117	126	143	157	159	157
7	120	122	116	104	75	80	117	136	123	80	82	139	186	203	195	170
8	164	149	131	120	115	115	120	126	134	143	160	182	191	187	178	170
9	221	179	136	142	147	151	125	111	152	197	235	237	204	172	166	163
10	142	120	100	90	87	85	88	109	148	182	195	188	170	150	132	125
11	70	61	55	39	23	24	41	91	151	180	171	141	139	126	95	78
12	78	63	61	64	69	80	103	136	159	163	153	148	153	151	139	131
13	68	61	68	85	120	124	161	188	179	157	134	143	178	186	177	180
14	146	143	138	134	140	148	158	162	151	133	124	140	166	174	170	166
15	210	236	210	169	172	161	163	148	124	110	97	127	174	169	155	161
16	181	189	176	145	120	113	121	124	112	94	86	97	112	110	99	94

(б)

Рис. 3. Рельеф макроблока (а) и исходные цифровые значения сигнала яркости (б).

y \ x	1 2 3 4	5 6 7 8	9 10 11 12	13 14 15 16
1				
2	Max	Min	Max	Min
3				
4				
5				
6	Min	Max	Min	Max
7				
8				
9				
10	Max	Min	Max	Min
11				
12				
13				
14	Min	Max	Min	Max
15				
16				

Рис. 4. Разбиение макроблока на участки, в каждом из которых выбирается один характерный пиксел, имеющий максимальное, либо минимальное значение в данном участке.



Radio Research & Development Institute (NIIR)
16, Kazakova str., Moscow 103064

Abstract: The results on the acceleration of motion analysis procedure being one of the most computationally consuming elements in dynamic image coding according to MPEG-1/2 standard are reported. In the proposed method the acceleration of motion vector search algorithm in 16 times in comparison with the standard algorithm results in the loss in compression ratio less than 2 %.

Coding of dynamic images according to MPEG-1/2 [1,2], MPEG-4 [3], H.261 [4] and H.263 [5,6] standards includes motion compensation as the main element. The use of the motion compensation allows one to obtain higher compression ratios (in 2-3 times) than without it. The standard method of motion compensation is the blocked based one, in which motion compensation is applied to each block of the frame. In the method current image frame is divided into the rectangular blocks of fixed size (called later as "macroblocks"). The typical size of the macroblock is 16 x 16 pixels for the luminance component and 8x8 pixels for each of the chrominance components, taking into account that resolution of the chrominance is usually twice less than that for the luminance. Each macroblock is assigned a motion vector ($\vec{V} = (V_x, V_y)$) characterizing the position of the macroblock in the previous frame:

$$H = \sum_{x,y=1,16} |C(x,y,t)| \equiv \sum_{x,y=1,16} |F(x,y,t) - F(x-V_x, y-V_y, t-\Delta t)| \quad (1)$$

Here F is the pixel luminance value, (x,y) are spatial coordinates of the pixels of the macroblock being counted from its left-up corner, t is temporal reference to the current frame, $t - \Delta t$ is the temporal reference to the previous frame. The summation in (1) is carried out for all pixels in the macroblock. The values of C represent the difference between the current luminance values in the macroblock and the values in the previous frame, corresponding to the the position of the macroblock with the shift on the motion vector. If all changes from the current and referenced frames can be almost reduced to the spatial shift of the image elements (which is the case in many situations), the value of C is small. In this case higher compression can be achieved by encoding the value of C instead of the original signal Y . The value of H in (1) plays the role of the norm (power) of the motion compensated signal C . The smaller the value of H , the more effective motion compensation is. The similar coding scheme is applied to the chrominance signals. It is necessary to mention that there is a possibility in MPEG-1/2 standard to code the macroblock with the use of two motion vectors. In this case the difference between the original signal and the interpolation reference value, calculated with the use of two motion vectors and two referenced frames, is encoded. In addition, the motion vectors in MPEG 1/2 standards can be represented with half pixel accuracy. In this case the value of the luminance in the referenced frame $F(x-V_x, y-V_y, t-\Delta t)$ corresponding to sub pixel position is calculated with the use of the linear interpolation between neighboring pixels.

There are various algorithms of motion vector search. The most simple and the most computationally consuming is the algorithm based on the full search among all possible values of the motion vector in some search window $\min < V_x, V_y < \max$. The value of \vec{V} for which H in (1) has the smallest value is taken as the sought vector. Considering the rectangular search window of the size of $\pm N$ pixels along coordinates x and y , the estimation of the necessary value of computational operations to find the motion vector per macroblock gives the value of $3 \cdot 256(2N+1)^2$. This is equal to $3 \cdot (2N+1)^2$ operations per pixel, and for the typical value of the search window size $N=15$ becomes the significant number of the order of 10^3 operations/pixel. This method provides the best result, but it is complicated one from the view point of the amount of computations. For this reason it is usually used only as the standard referenced method for the estimation of the quality of other motion vector search techniques.

One of the most effective methods of motion vector search is the method described in [7]. In this method the summation in (1) is replaced by the summation only with the small number of macroblock pixels. The idea of the method is to select the small number of macroblock pixels which characterize in the best manner the shape of the luminance signal in the macroblock. Fig.1 shows two frames from the test sequence "Flower Garden". Fig.2 shows one macroblock and the region in the reference frame that corresponds to this macroblock. Fig.3 shows the relief of the luminance signal in the macroblock and the table of values at each pixel.

According to the method of characteristic pixels each macroblock of the size of 16x16 pixels is divided in 16 areas (Fig.4), and only one pixel having maximum or minimum value is selected in each area. Maximums and minimums have the chess-order as it is shown in Fig.4. Only these characteristic pixels are used for the calculation of the control sum of the macroblock:

$$H1 = \sum_{(x,y)\text{-координаты выбранных точек в соответствующих участках макроблока}} |F(x, y, t) - F(x - V_x, y - V_y, t - \Delta t)| \quad (2)$$

The calculation of the control sum (2) with the use of 16 characteristic pixels is 16 times faster compared to the calculation of the control sum (1).

In spite of its simplicity the proposed method is characterized by high efficiency. Under the considerable acceleration of motion vector search procedure (x16 times), the loss in compression ratio due to the errors in motion vector search does not exceed 2% (see Table 1).



(a)



(b)

Fig. 1. Current (a) and referenced (b) frames of the sequence "Flower Garden"

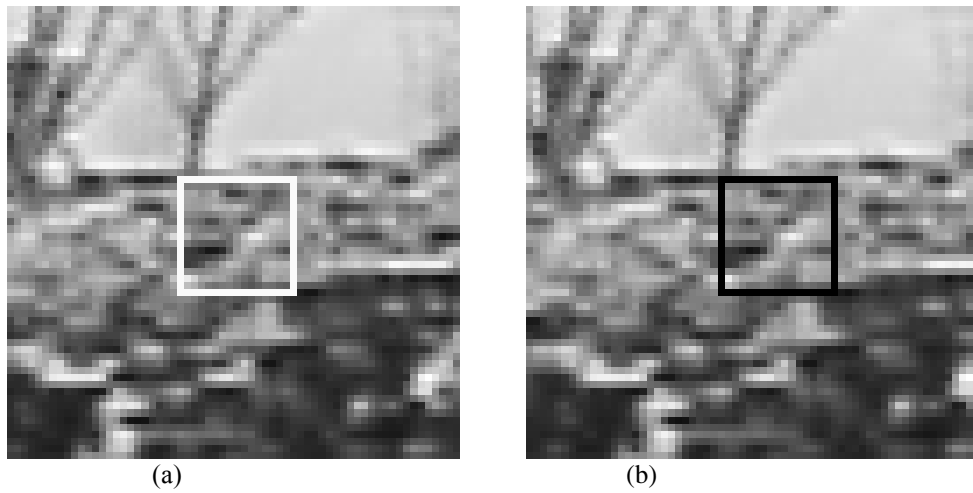
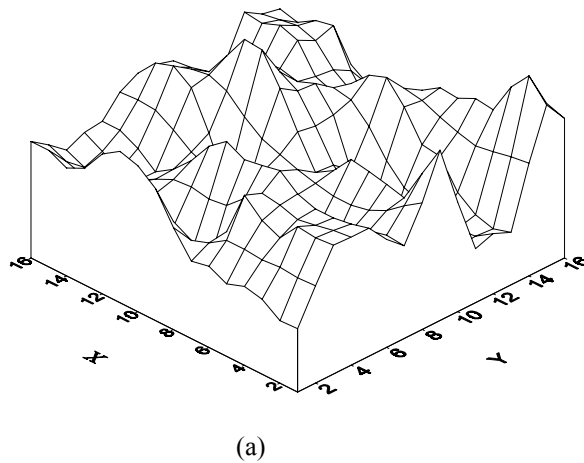


Fig. 2. Enhanced view of one macroblock from the current frame (a: marked by white color) the corresponding to it region in the referenced frame (b: marked with black color).



y \ x	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	82	81	97	102	94	105	100	132	171	189	196	181	151	140	151	151
2	126	125	127	125	127	127	112	104	117	137	150	144	125	117	125	135
3	167	165	168	142	153	174	115	62	64	87	118	121	86	84	106	117
4	166	163	158	154	161	165	147	126	123	127	125	114	101	103	115	128
5	166	160	156	156	164	183	178	174	186	168	138	115	100	119	131	131
6	140	140	135	128	125	132	145	153	147	128	117	126	143	157	159	157
7	120	122	116	104	75	80	117	136	123	80	82	139	186	203	195	170
8	164	149	131	120	115	115	120	126	134	143	160	182	191	187	178	170
9	221	179	136	142	147	151	125	111	152	197	235	237	204	172	166	163
10	142	120	100	90	87	85	88	109	148	182	195	188	170	150	132	125
11	70	61	55	39	23	24	41	91	151	180	171	141	139	126	95	78
12	78	63	61	64	69	80	103	136	159	163	153	148	153	151	139	131
13	68	61	68	85	120	124	161	188	179	157	134	143	178	186	177	180
14	146	143	138	134	140	148	158	162	151	133	124	140	166	174	170	166
15	210	236	210	169	172	161	163	148	124	110	97	127	174	169	155	161
16	181	189	176	145	120	113	121	124	112	94	86	97	112	110	99	94

Fig. 3. Macroblock's relief (a) and the values of luminance at each pixel (b).

Table 1. The size of the MPEG-2 encoded "Flower Garden" sequence (bytes). 97 frames of the resolution [640x480] have been encoded.

Motion vector accuracy	Proposed algorithm with 16 characteristic pixels (2)	Standard algorithm with all 256 pixels (1)
0.5 pixel	3295558	3242122
1 pixel	3699264	3545362

y \ x	1 2 3 4	5 6 7 8	9 10 11 12	13 14 15 16
1	Max	Min	Max	Min
2				
3				
4				
5	Min	Max	Min	Max
6				
7				
8				
9	Max	Min	Max	Min
10				
11				
12				
13	Min	Max	Min	Max
14				
15				
16				

Fig. 4. Division of macroblock into regions, only one pixel having maximum or minimum value in the particular region is selected as the characteristic pixel.

References:

- [1] ISO/IEC 11172-2. Information Technology - Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 Mbit/s. Part 2: Video./Ed.1 JTS I/SC 29, 1993.
- [2] ISO/IEC 13818-2. Information Technology - Generic Coding of Moving Pictures and Associated Audio Information. Part 2: Video./ Ed.1 JTS I/SC 29, 1994.
- [3] ISO/IEC 14496. Information Technology - MPEG-4 Overview (San-Jose Version), Feb. 1998.
- [4] "Line transmission of non-telephone signals/ video codec for audiovisual services at px64 Kbits", ITU-T Recommendation H.261, March 1993.
- [5] "Line transmission of non-telephone signals/ video coding for low bit rate communication", ITU-T Recommendation H.263, March 1996.
- [6] "Line transmission of non-telephone signals/ video coding for low bit rate communication", ITU-T Recommendation H.263, Jan. 1998.
- [7] A.V. Dvorkovich, V.P. Dvorkovich, Yu.B. Zoubarev, A.Yu. Sokolov "A Method of Motion Vector Analysis for the Details of Dynamic Images", Russian Patent No. 2137194 of July 15th, 1998.