FIGURE 1 A typical composited scene in the multi-point video conferencing application. (Scene 1)

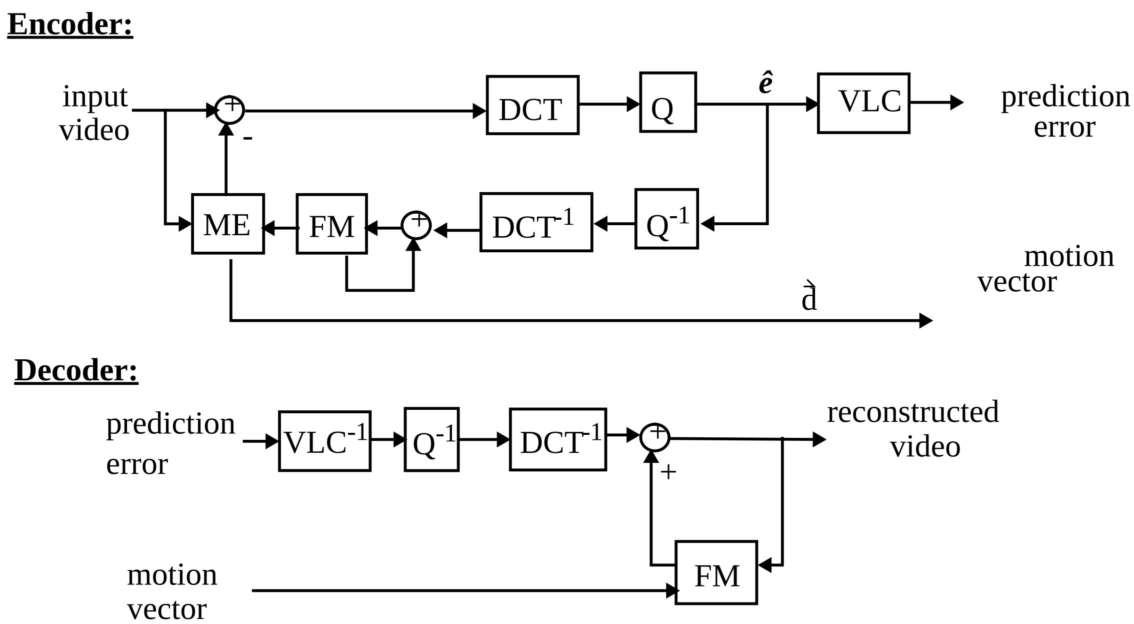


FIGURE 2 Block diagram for hybrid MC-DCT based compression systems. ME: Motion Estimation, FM: Frame Memory, VLC: Variable Length Code.

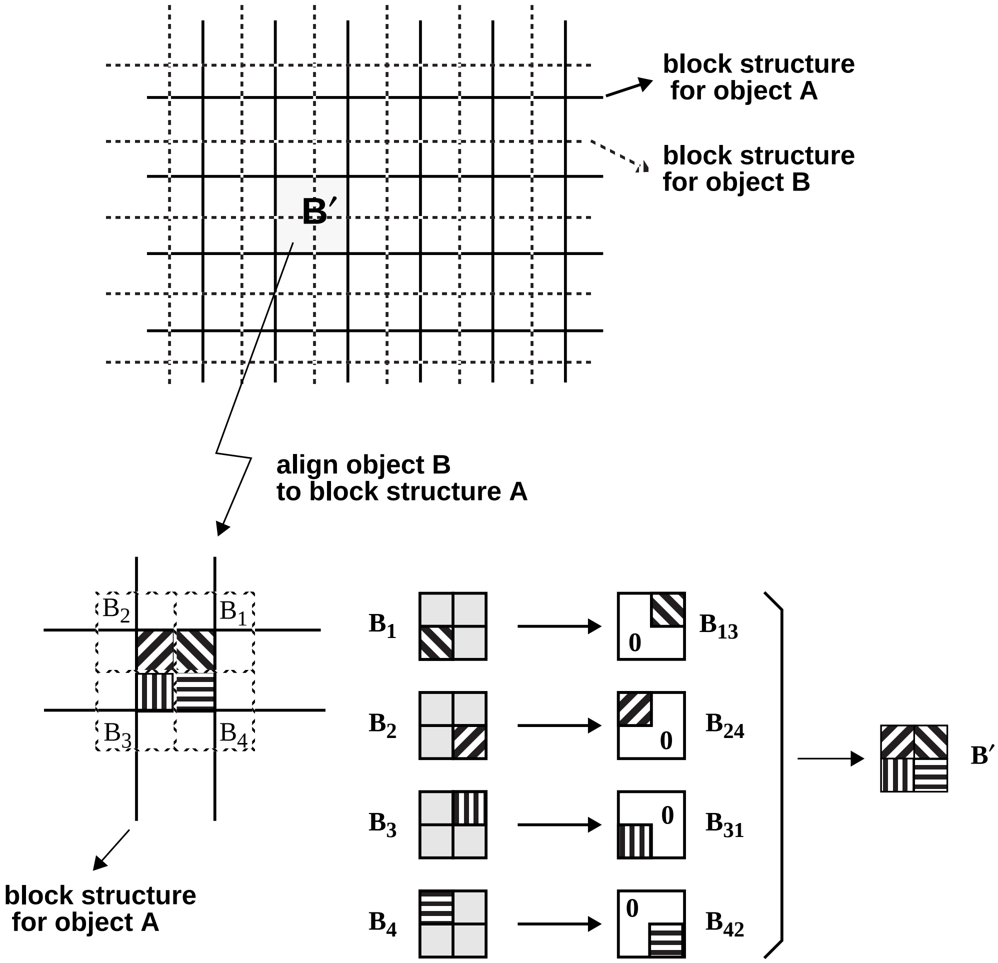


FIGURE 3 Re-assembling the image blocks of video object B based on a new block structure, which mismatches object B′s original block structure. After re-alignment, a new image block, B′, consists of contributions (B_{13}, B_{24}, B_{31}, and B_{42}) from four original neighboring blocks (B_1—B_4).

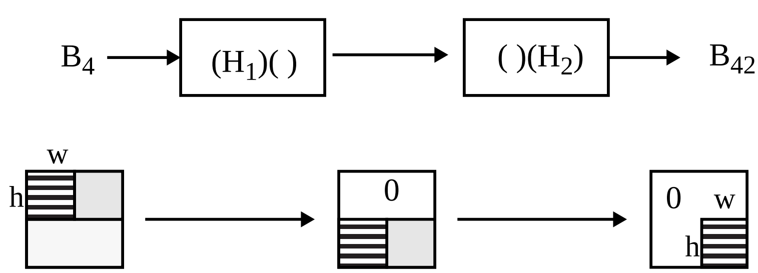


FIGURE 4 Using matrix multiplication to extract a subblock and translate it to the opposite corner.

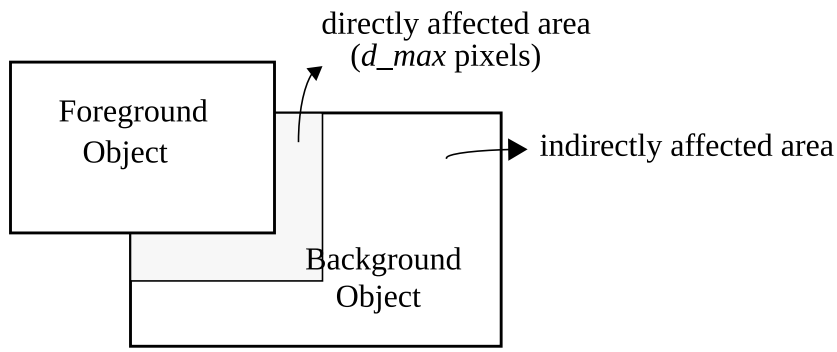


FIGURE 5 An example showing the problem with video compositing directly in the MC domain. *d_max* is the maximum search distance.

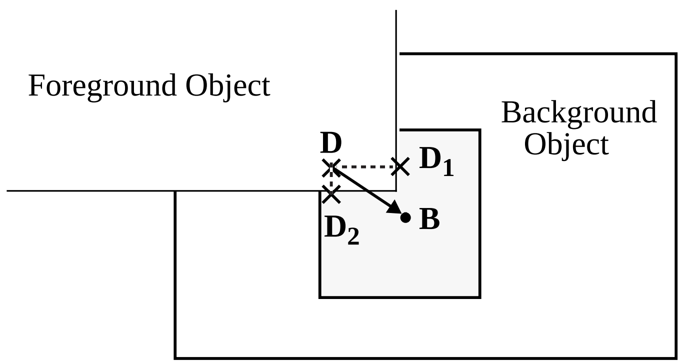


FIGURE 6 Reducing the number of search positions to two by using Jain and Jain's assumption. B is the location of the current image block. D is the optimal reference location. The shaded area represents the search area.

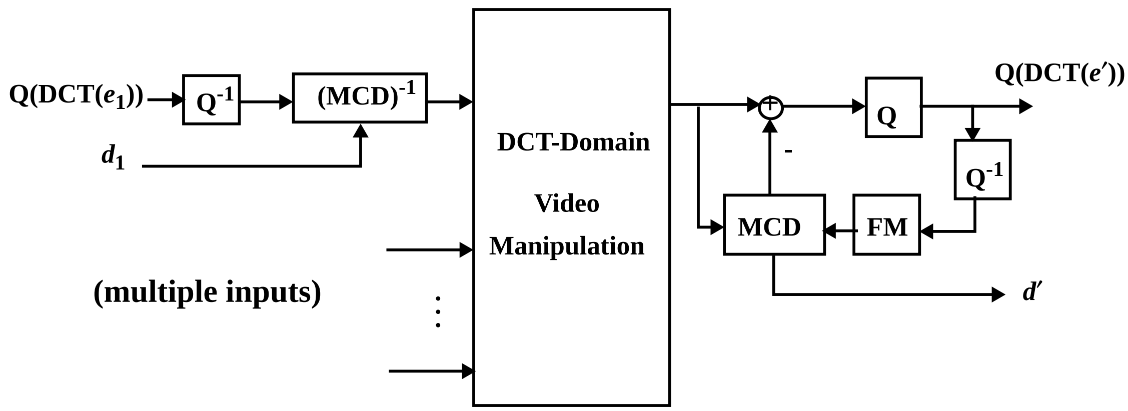
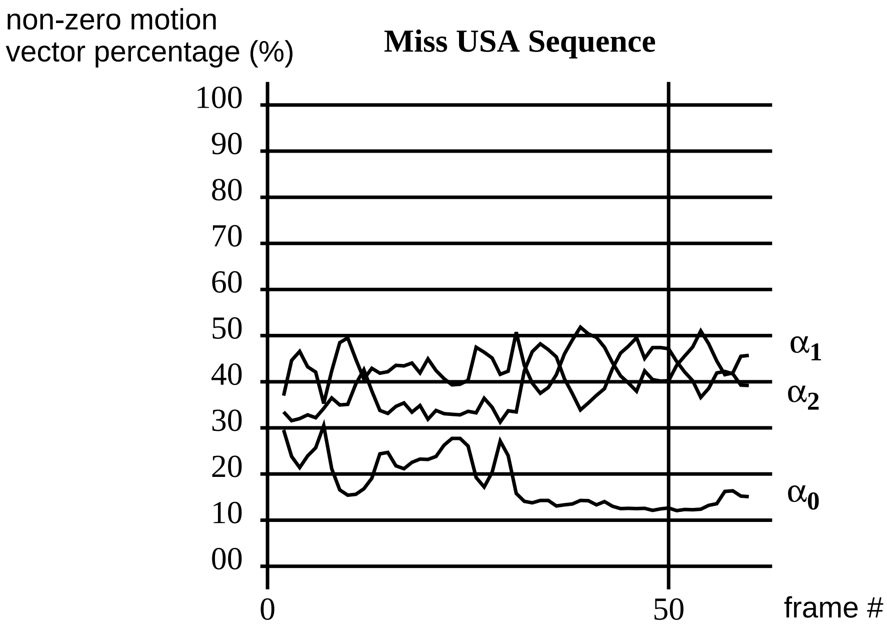
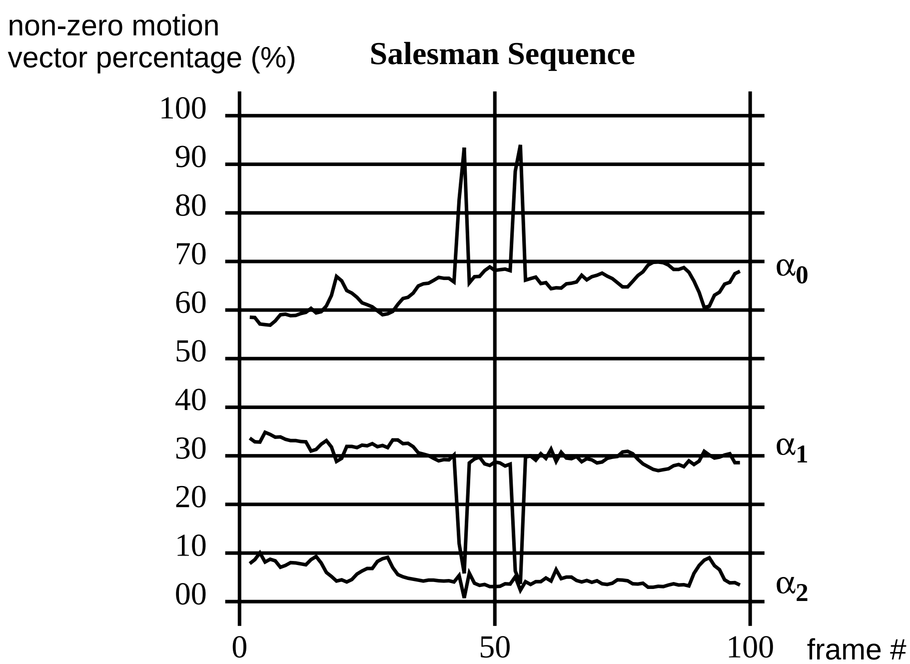


FIGURE 7 Manipulation of MC-DCT compressed video in the DCT domain. The output video is MC-DCT encoded as well.



(a)



(b)

FIGURE 8 Experimental results of non-zero motion vectors. α_2, α_1 **are defined in Table 1,** $\alpha_0 = 1 - \alpha_2 - \alpha_1$.

Table 1 Computational complexity of major video manipulation functions[a]

	operation	# of multi. / pixel	# of add. / pixel
DCT Domain:	MCD, MCD^{-1}	$(4/\beta + 2/\sqrt{\beta}) \cdot N \cdot \alpha_2 + (2/\beta) \cdot N \cdot \alpha_1$	$[(4/\beta + 2/\sqrt{\beta}) \cdot N + 3] \cdot \alpha_2 + [(2/\beta) \cdot N + 1] \cdot \alpha_1 + 1$
	pixel-wise translation	$(2/\beta+2/\sqrt{\beta}) \cdot N$	$(2/\beta+2/\sqrt{\beta}) \cdot N + 3$
	scale 1/2×1/2	$(1/\beta+1/(2\sqrt{\beta})) \cdot N$	$(1/\beta+1/(2\sqrt{\beta})) \cdot N + 3/4$
	scale 1/3×1/3	$(1/\beta+1/(3\sqrt{\beta})) \cdot N$	$(1/\beta+1/(3\sqrt{\beta})) \cdot N + 8/9$
	pixel multiplication	$N^2/(\beta_1 \cdot \beta_2)$	$N^2/(\beta_1 \cdot \beta_2)$
	semi-transparent block-wise overlapping	$< (1/\beta_1 + 1/\beta_2)$	$< 2 \cdot (1/\beta_1 + 1/\beta_2)$
Spatial Domain:	FDCT, $FDCT^{-1}$ [b]	$2 \cdot \log_2 N - 3 + 8/N$	$3 \cdot (\log_2 N - 1) + 4/N$
	MC, MC^{-1}	0	1
	scale 1/2×1/2	1/4	3/4
	scale 1/3×1/3	1/9	8/9
	pixel multiplication	1	0
	semi-transparent block-wise overlapping	1	2
Common:	opaque block-wise overlapping	0	0
	Inverse Quantization	$1/\beta$	0
	Quantization	1	0

a. Notations:
α_2: the percentage of image blocks which need block boundary adjustment in both directions, i.e., both d_x and d_y are not integral multiples of the block size.
α_1: the percentage of image blocks which need block boundary adjustment in only one direction, i.e., only one of d_x and d_y is an integral multiple of the block size.
N: the block width (8 in our experiments).

b. We use the fast DCT algorithm of Chen & Smith [8]. If we use the 2N-point FFT approach, the computational complexity will be doubled.

**Table 2 Comparison of computational complexity for MC-DCT coded video manipulations
(the DCT domain vs. the spatial domain)**

	Spatial Domain		DCT Domain	
	# op/pixel	CPU time[a]	# op/pixel	CPU time
Scene 1	33.28 mul. 57.57 add	40.72 sec.	26.63 mul. (***1.25***[b]) 38.96 add. (***1.48***)	31.46 sec. (***1.29***)
Scene 2	17.99 mul. 32.53 add	22.53 sec.	16.10 mul. (***1.12***) 23.89 add. (***1.41***)	20.29 sec. (***1.11***)
Scene 3	13.41 mul. 23.39 add	16.57 sec.	13.76 mul. (***0.97***) 20.39 add. (***1.15***)	14.69 sec. (***1.13***)

a. CPU time on a SPARC I machine.

b. Each highlighted score in the parenthesis represents the ratio between the DCT-domain result and the corresponding spatial-domain result. A score greater than one means that the DCT-domain implementation is more efficient than the spatial-domain approach.

**Table 3 Comparison of computational complexity for DCT-coded video manipulations (*without MC*)
(the DCT domain vs. the spatial domain)**

	Spatial Domain	DCT Domain
	# op/pixel	# op./pixel
Scene 1	33.28 mul, 50.35 add	9.66 mul (***3.45***), 11.23 add (***4.48***)
Scene 2	17.99 mul, 28.53 add	7.36 mul (***2.44***), 8.01 add (***3.56***)
Scene 3	13.41 mul, 20.39 add	3.3 mul (***4.06***), 2.89 add (***7.06***)

Table 4 Impact of Video Manipulation on Video Quality
(the DCT-domain approach vs. the spatial-domain approach) [a]

	After Manipulation (before the 2nd coding)	After the 2nd coding	
		Spatial-domain Compositing	DCT-Domain Compositing
Scene 1	31.5 dB	28.8 dB (-***2.7***[b] dB)	28.5 dB (-***3.0*** dB)
Scene 2	34.7 dB	27.0 dB (-***7.7*** dB)	26.3 dB (-***8.4*** dB)
Scene 3	30.7 dB	29.5 dB (-***1.2*** dB)	29.5 dB (-***1.2*** dB)

a. all input video streams are MC-DCT encoded.

b. these highlighted scores show the extra quality degradation introduced in the 2nd coding pass.