On-Line Table 1: Demographics of research subj	ects
Demographics/Brain Tumor Type	Value
Gender (M:F)	36:30
Mean age (SD)	58 (10.2)
Tumor type (GBM:AA:LGG:other)	52:7:6:1
Surgical treatment (GTR:STR:Bx)	22:24:20

On-Line Table 2	Sensitivity compariso	on of the 6 displ	ay methods for	each of the 3	raters, and whe	n requiring all raters to	agree
	Display					Sensitivity	Specificity
Reader	Mode	TP	TN	FP	FN	(%)	(%)
Reader 1	Cp	38	17	1	10	79	94
Reader 1	C+F ^c	41	17	1	7	85	94
Reader 1	Ν	29	16	2	19	60	89
Reader 1	N+F	32	16	2	16	67	89
Reader 1	Sc	38	17	1	10	79	94
Reader 1	S+F ^c	39	17	1	9	81	94
Reader 2	С	38	16	2	10	79	89
Reader 2	C+F	38	17	1	10	79	94
Reader 2	Ν	44	14	4	4	92	78
Reader 2	N+F	34	15	3	14	71	83
Reader 2	S	39	16	2	9	81	89
Reader 2	S+F	36	16	2	12	75	89
Reader 3	Cp	37	18	0	11	77	100
Reader 3	$C + F^{b}$	39	17	1	9	81	94
Reader 3	Ν	36	17	1	12	75	94
Reader 3	N+F	36	15	3	12	75	83
Reader 3	S	40	16	2	8	83	89
Reader 3	S+F	38	17	1	10	79	94
All 3 agree	Ca	34	15	3	14	71	83
All 3 agree	C+F ^a	35	15	3	13	73	83
All 3 agree	Ν	29	13	5	19	60	72
All 3 agree	N+F	26	13	5	22	54	72
All 3 agree	Sa	35	14	4	13	73	78
All 3 agree	S+F	32	15	3	16	67	83

Note:—When requiring all 3 raters to agree, automated change detection with or without flicker is significantly better than traditional side-by-side mode. There is a nonsignificant trend for it to be better than image subtraction. ^a P < .05 vs N; ^b P < .01 vs N; ^c P < .001 vs N.

On-Line Table 3: Statistical significance between methods, for cases where all 3 raters agree

	С	CF	Ν	NF	S	SF
С	*	0.6547	0.1317	0.0209	0.1429	0.4142
CF		*	0.0339	0.0027	1.0000	0.2568
Ν			*	0.1797	0.0339	0.3173
NF				*	0.0027	0.0339
S					*	0.1797
SF						*
С	*	1.0000	0.1573	0.3173	0.5637	1.0000
CF		*	0.1573	0.3173	0.5637	1.0000
Ν			*	1.0000	0.3173	0.1573
NF				*	0.3173	0.1573
S					*	0.3173
SF						*

Note:—For *P* values from the comparison of sensitivities (all 3 agree), the table (*top*) shows that the sensitivity of both change detection and subtraction are significantly better than the "normal display" mode with and without flicker in most cases. There is no difference detected between change detection and subtraction. For *P* values from the comparison of specificities (all 3 agree), the table (*bottom*) shows there is no significant difference detected between change detection.

On-Line Table 5: κ agreement statistic between raters for the 6 display formats

		С	CF	Ν	NF	S	SF
Reviewer 1	Reviewer 2	0.78	0.78	0.44	0.60	0.68	0.69
Reviewer 1	Reviewer 3	0.81	0.74	0.76	0.60	0.78	0.78
Reviewer 2	Reviewer 3	0.84	0.78	0.58	0.63	0.71	0.72
Overall ĸ		0.82	0.77	0.60	0.62	0.73	0.73

Note:—Automated change detection without flicker has the highest mean agreement score, followed closely by change detection with flicker. Image subtraction with and without flicker is in the middle, and nonenhanced methods fares the worst. If κ is <0.4, agreement is considered poor; if κ is 0.4–0.75, agreement is considered good; and if κ is <0.0.75, agreement is considered excellent. Five of 6 κ ratings are in the excellent range for change detection (with or without flicker). For the traditional side-by-side mode (with or without flicker), be the Good and 2 in the excellent range.

On-Line T	able 4: Recei	ver operato	or characte	ristic curve a	nalysis
Reviewer	Modality	AUC	SE	95% LCL	95% UCL
1	С	0.8870	0.0463	0.7963	0.9777
1	CF	0.9000	0.0429	0.8159	0.9841
1	Ν	0.7440	0.0639	0.6188	0.8692
1	NF	0.8230	0.0453	0.7342	0.9118
1	S	0.8870	0.0460	0.7968	0.9772
1	SF	0.8780	0.0429	0.7939	0.9621
2	С	0.8450	0.0512	0.7446	0.9454
2	CF	0.8560	0.0494	0.7592	0.9528
2	Ν	0.8460	0.0624	0.7237	0.9683
2	NF	0.8060	0.0657	0.6772	0.9348
2	S	0.9030	0.0340	0.8364	0.9696
2	SF	0.8380	0.0501	0.7398	0.9362
3	С	0.8540	0.0399	0.7758	0.9322
3	CF	0.9170	0.0433	0.8321	1.0000
3	Ν	0.8500	0.0226	0.8057	0.8943
3	NF	0.8180	0.0547	0.7108	0.9252
3	S	0.8700	0.0496	0.7728	0.9672
3	SF	0 8820	0 0464	0 7911	0 9729

Note:—The table shows a trend for side-by-side display to perform inferior to change detection and subtraction, but there is significant overlap.

Rater	Modality	Median	Minimum	Maximum	Rank
1	С	39	14	68	2
1	CF	41	14	137	4
1	Ν	41.5	14	97	5
1	NF	42.5	18	84	6
1	S	39	12	83	2
1	SF	38	19	92	1
2	С	85	27	129	5
2	CF	85	38	176	5
2	Ν	73.5	22	246	3
2	NF	77	33	286	4
2	S	71	33	138	1
2	SF	73	37	260	2

On-Line Table 6: Rater review times in seconds for nonprogressing

cases

3

3

3

С

CF

Ν

148

97

95

3 50 NF 164.5 343 5 353 3 S 161 32 4 3 SF 184 56 440 6 Note:--These values represent the time spent reviewing cases that were ultimately judged to be negative. We excluded cases with markings indicating progression because there was a difference in the number of markings produced between methods and that would necessarily affect total time. Two of the raters were more efficient with subtraction. The traditional side-by-side mode tended to be the slowest, though for 1 rater it was the fastest. Flicker mode was slower than its nonflicker companion mode for 7 (we are not sure that it is 8) of the 9 rater-mode comparisons.

43

31

43

261

353

326

3

2

1