

# 1 Experiments

## 1.1 Image recognition evaluation

### 1.1.1 Data set description

### 1.1.2 Performance of spatial pyramid matching

SVM	One-level	Two-level	Three-level
Linear	74.17	80	<b>87.5</b>
Poly	70	62.5	31.67
RBF	37.5	83.33	74.17
Sigmoid	16.67	16.67	16.67
Histogram Intersection	79.17	82.5	<b>85</b>

Table 1: Recognition accuracies (percent) of different spatial pyramids using SVM with different kernels

	One-level	Two-level	Three-level
KNN	50.83	45	48.33

Table 2: Recognition accuracies (percent) of different spatial pyramids using KNN

### 1.1.3 Performance of earth mover’s distance

	RBF	LAP	ID	ISD
Aligned Distance	79.44	76.11	73.89	75.56
Unaligned Distance	78.89	75.56	73.33	75

Table 3: Comparison of recognition accuracy between aligned distance and un-aligned distance

## 1.2 Video recognition

	<b>Gaussian</b>	<b>Laplacian</b>	<b>ISD</b>	<b>ID</b>	<b>Fused kernels</b>
Level 0	$44.38 \pm 2.13$	$44.90 \pm 2.73$	$44.01 \pm 2.13$	$45.36 \pm 3.13$	$44.33 \pm 2.61$
Level 1 (Unaligned)	$43.08 \pm 3.14$	$43.85 \pm 3.84$	$43.22 \pm 3.11$	$43.85 \pm 3.56$	$43.56 \pm 3.46$
Level 1 (Aligned)	$43.61 \pm 2.97$	$43.40 \pm 3.18$	$43.46 \pm 2.97$	$43.22 \pm 3.11$	$44.08 \pm 3.25$

Table 4: Means and standard deviations (percent) of MAPs over six events at different levels using SVM with different kernels.

	<b>Gaussian</b>	<b>Laplacian</b>	<b>ISD</b>	<b>ID</b>	<b>Fused kernels</b>
bxx	$40.20 \pm 2.57$	$38.35 \pm 2.31$	$39.93 \pm 2.58$	$38.23 \pm 2.08$	$39.34 \pm 2.55$
txx	$44.28 \pm 2.14$	$44.90 \pm 2.73$	$44.01 \pm 2.13$	$45.36 \pm 3.13$	$44.33 \pm 2.61$
txc	$42.15 \pm 4.73$	$45.01 \pm 3.45$	$43.47 \pm 4.56$	$45.38 \pm 3.20$	$44.11 \pm 3.90$
tfx	$43.76 \pm 2.99$	$44.14 \pm 3.36$	$43.61 \pm 3.03$	$44.05 \pm 3.51$	$44.18 \pm 3.22$
tfc	$43.71 \pm 1.37$	$46.02 \pm 1.84$	$44.93 \pm 1.64$	$46.21 \pm 1.83$	$45.28 \pm 1.62$
straightforward soft	$43.54 \pm 2.12$	$44.77 \pm 2.41$	$43.52 \pm 2.08$	$45.24 \pm 2.47$	$44.79 \pm 2.55$
Gaussian soft	$44.77 \pm 2.80$	$45.23 \pm 2.76$	$44.90 \pm 3.01$	$45.23 \pm 2.87$	$45.20 \pm 3.04$

Table 5: Means and standard deviations (percent) of MAPs over six events using different mechanisms to build histograms

	<b>Gaussian</b>	<b>Laplacian</b>	<b>ISD</b>	<b>ID</b>	<b>Fused kernels</b>
spherical 128	$24.70 \pm 1.41$	$43.04 \pm 1.61$	$26.92 \pm 1.00$	$43.64 \pm 0.96$	$32.91 \pm 2.20$
spherical 64	$23.99 \pm 1.40$	$42.35 \pm 1.64$	$25.62 \pm 1.11$	$43.42 \pm 1.18$	$29.01 \pm 1.10$
full 128	$25.69 \pm 7.57$	$21.39 \pm 7.32$	$26.49 \pm 8.38$	$21.93 \pm 7.75$	$21.79 \pm 7.29$
full 64	$25.23 \pm 0.94$	$29.69 \pm 1.81$	$25.68 \pm 1.34$	$30.74 \pm 1.67$	$26.74 \pm 1.63$

Table 6: Means and standard deviations (percent) of MAPs over six events using different GMMs

<b>Recognition accuracy</b>	
Kodak $\rightarrow$ Kodak	$38.5 \pm 12.7$
Youtube $\rightarrow$ Kodak	$30.0 \pm 6.9$
Baseline	$41.6 \pm 11.5$

Table 7: Means and standard deviations (percent) of recognition accuracies using concept attributes

<b>Training videos</b>	<b>Testing videos</b>	<b>Original videos</b>	<b>Compressed videos</b>
60	846	$38.9 \pm 2.9$	$38.6 \pm 2.8$
120	786	$45.7 \pm 2.2$	$44.5 \pm 1.6$
180	726	$49.5 \pm 1.8$	$48.3 \pm 1.9$
240	666	$52.0 \pm 2.1$	$50.6 \pm 2.1$

Table 8: Recognition accuracies (percent) using distances calculated on original videos and distances calculated on compressed videos.

	<b>SVM_T</b>	<b>SVM_AT</b>	<b>FR</b>	<b>A-SVM</b>	<b>MKL</b>	<b>DTS</b>
L0	44.33 $\pm$ 2.61	52.21 $\pm$ 2.54	52.33 $\pm$ 2.20	47.03 $\pm$ 3.26	50.82 $\pm$ 2.16	47.14 $\pm$ 2.16
L1 Unaligned	43.55 $\pm$ 3.46	55.37 $\pm$ 2.26	55.95 $\pm$ 3.79	45.86 $\pm$ 4.39	55.08 $\pm$ 6.17	50.97 $\pm$ 3.46
L1 Aligned	44.08 $\pm$ 3.25	57.56 $\pm$ 3.02	53.91 $\pm$ 1.48	45.42 $\pm$ 3.62	53.81 $\pm$ 3.86	53.32 $\pm$ 3.25
Voc2005 Soft Assignment	44.79 $\pm$ 2.55	47.80 $\pm$ 1.67	51.89 $\pm$ 1.99	47.41 $\pm$ 3.13	47.16 $\pm$ 1.77	45.05 $\pm$ 2.55
Specialized GMM (S 128)	32.91 $\pm$ 2.20	33.15 $\pm$ 1.78	41.78 $\pm$ 3.98	37.07 $\pm$ 3.52	33.64 $\pm$ 0.74	46.61 $\pm$ 2.20
Voc1000 Level 0	44.23 $\pm$ 3.20	53.67 $\pm$ 3.29	52.16 $\pm$ 4.00	45.07 $\pm$ 4.05	52.33 $\pm$ 4.23	49.90 $\pm$ 3.20
Voc1000 Gaussain Assign	45.20 $\pm$ 3.04	56.90 $\pm$ 2.79	54.03 $\pm$ 4.02	46.62 $\pm$ 3.14	55.47 $\pm$ 4.11	53.41 $\pm$ 3.04
Aligned L1 + Gau Ass	44.69 $\pm$ 2.84	60.21 $\pm$ 1.94	55.29 $\pm$ 3.00	46.28 $\pm$ 4.23	58.36 $\pm$ 3.75	57.01 $\pm$ 2.84
Aligned L1 + Gau Ass + Voc1000 L0	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$

Table 9: Means and standard deviations (percent) of MAPs over six events for all methods

## References