vincent.delaitre@ens-lyon.fr

1 K-Nearest classifier + Bag-of-features

1.1 Tested Parameters

1.1.1 Detectors

- HARRIS(S): Harris-Laplace detector with only scale invariance.
- HARRIS(SR): Harris-Laplace detector with scale and rotation invariance.
- DENSE: Dense features.
- MSDENSE: Multi-scale dense features.

1.1.2 Descriptors

- SIFT(L2): SIFT descriptor normalized with L2 norm.
- SIFT(L2T): SIFT descriptor normalized with L2 norm and truncation over 0.2.

1.1.3 Dictionnary size

Tested sizes: 128, 256, 512, 1024.

1.1.4 Histogram normalization

- NONE: no normalization.
- L1: normalized with L1 norm.
- L2: normalized with L2 norm.

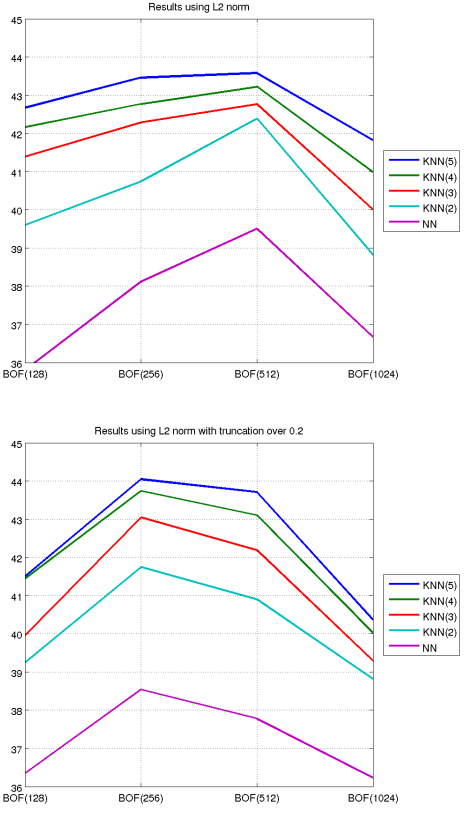
1.1.5 Classifier

Classifier is K-nearest neighbours with K from 1 to 5.

1.2 Results

1.2.1 Detector

Figure 1: KKN+BOF: Performances for $K \in \{1, \cdots, 5\}$ across different dictionnary size. Left and right graphics are respectively obtained using the L2 and L2T norm.



1.2.3 Dictionnary size

Table 3: KKN+BOF: Average performances for different dictionnary size. The mean is computed over all other combinaison of parameters.

Dic. Size	128	256	512	1024
Avg. perf.	38.0 ± 3.2	39.1 ± 3.0	$\mathbf{39.2 \pm 2.7}$	36.5 ± 4.4

1.2.4 Histogram normalization

Table 4: KKN+BOF: Average performances for different normalization of the signature histogram. The mean is computed over all other combinaison of parameters.

Histo. Norm	NONE	L1	L2
Avg. perf.	37.8 ± 4.4	38.1 ± 3.1	$\mathbf{38.6 \pm 2.9}$

However, if we consider only the Multi-scale dense detector, L1 performs better than L2 and NONE norms:

Table 5: KKN+BOF: Average performances for different normalization of the signature histogram. The detector is MSDENSE and the mean is computed over all other combinaison of parameters.

Histo. Norm	NONE	L1	L2
Avg. perf.	40.8 ± 3.5	$\mathbf{40.9 \pm 2.3}$	40.5 ± 2.5

1.2.5 Classifier

Table 6: KKN+BOF: Average performances for different values of K.

K	1	2	3	4
Avg. perf.	35.3 ± 2.9	37.8 ± 2.9	38.8 ± 3.1	39.3 ± 3.4

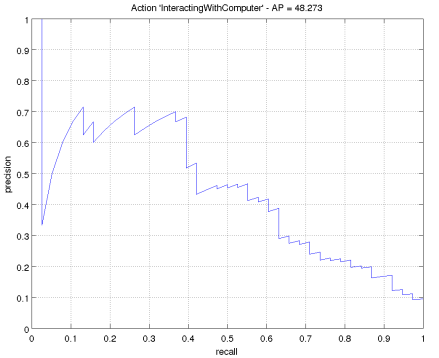
1.3 Accurracy and Precision-Recall

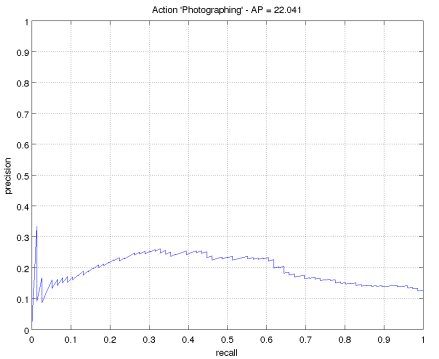
Classifier: K-Nearest Neighbours (K = 5)
Signature: Bag of features (K = 256)
Histogram normalization: None
K-means library: cpp
Channels:
(DENSE[spacing = 12-14-17-20-24-29-35-42-51-61, library: mylib]) x (SIFT[normalization: L2 (norm = 1, truncation over 0.2), library: cd])

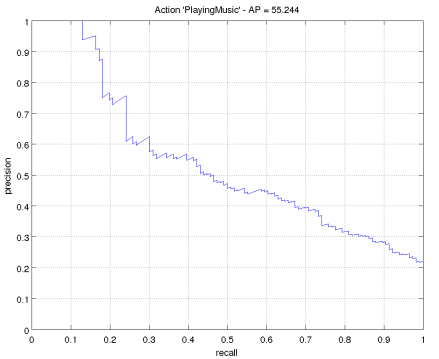
Table 7: KKN+BOF: Confusion table. The average accurracy is **42.1%**.

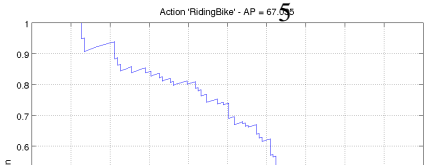
Actions	(1)	(2)	(3)	(4)	(5)	(
(1) Interacting With Computer	**39.47**	10.53	18.42	5.26	10.53	10
(2) Photographing	0.00	26.32	6.58	9.21	18.42	11
(3) Playing Music	6.90	24.14	**27.59**	6.90	12.07	6.
(4) Riding Bike	0.71	8.51	5.67	**51.06**	9.93	14
(5) Riding Horse	5.26	17.54	8.77	3.51	**38.60**	14
(6) Running	1.27	7.59	3.80	5.06	8.86	**54**
(7) Walking	0.00	8.40	0.00	5.04	10.92	18

Figure 2: KKN+BOF: Precision-Recall. The average precision is **46.5%**.









2 SVM classifier + Bag-of-features

2.1 Tested Parameters

2.1.1 Detectors

- MSDENSE: Multi-scale dense features.

2.1.2 Descriptors

- SIFT(L2): SIFT descriptor normalized with L2 norm.
- SIFT(L2T): SIFT descriptor normalized with L2 norm and truncation over 0.2.

2.1.3 Dictionnary size

Tested sizes: 256, 512, 1024.

2.1.4 Histogram normalization

- NONE: no normalization.
- L1: normalized with L1 norm.
- L2: normalized with L2 norm.

2.1.5 Kernels

- LINEAR: linear SVM.
- RBF: non linear SVM with RBF kernel.
- INTER: non linear SVM with histogram intersection kernel.
- CHI2: non linear SVM with χ^2 kernel.

2.1.6 Classifier

- 1vs1: One-versus-one classification
- 1vsA: One-versus-all classification

2.2 Results

2.2.1 Descriptor

Table 8: SVM+BOF: Average performances for different descriptors. The mean is computed over all other combinaison of parameters.

Descriptor	SIFT(L2)	SIFT(L2T)
Avg. perf.	$\mathbf{45.4 \pm 3.6}$	45.1 ± 3.3

2.2.2 Dictionnary size

Table 9: SVM+BOF: Average performances for different dictionnary size. The mean is computed over all other combinaison of parameters.

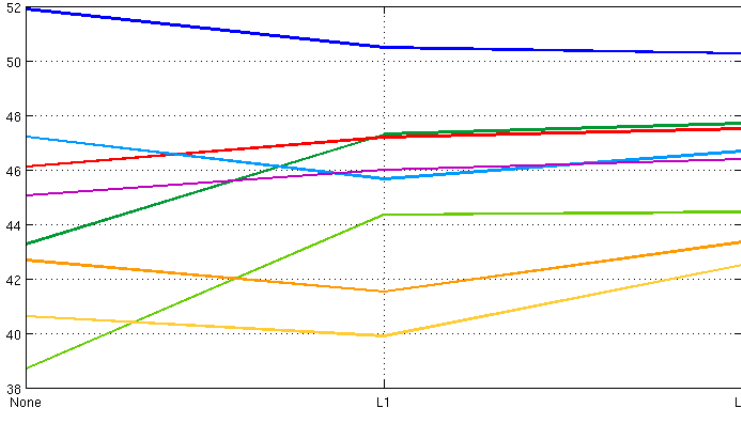
Dic. Size	256	512	1024
Avg. perf.	44.5 ± 3.5	45.2 ± 3.4	$\mathbf{46.2 \pm 3.2}$

2.2.3 Histogram normalization

Table 10: SVM+BOF: Average performances for different normalization of the signature histogram. The mean is computed over all other combinaison of parameters.

Histo. Norm	NONE	L1	L2
Avg. perf.	44.4 ± 4.0	45.3 ± 3.4	$\mathbf{46.1 \pm 2.7}$

Figure 3: SVM+BOF: Performances for several kernels and strategy across different histogram normalization. The mean is computed over all other combinaison of parameters.



2.2.4 Kernel

Table 11: SVM+BOF: Average performances for different kernels. The mean is computed over all other combinaison of parameters.

Kernel	LINEAR	RBF	INTER	CHI2
Avg. perf.	41.8 ± 1.9	44.5 ± 3.0	46.4 ± 1.3	$\mathbf{48.5 \pm 3.0}$

2.2.5 Classifier

Table 12: SVM+BOF: Average performances for different strategy of classification. The mean is computed over all other combinaison of parameters.

Strategy	1vs1	1vsA
Avg. perf.	44.2 ± 2.7	$\mathbf{46.3 \pm 3.8}$

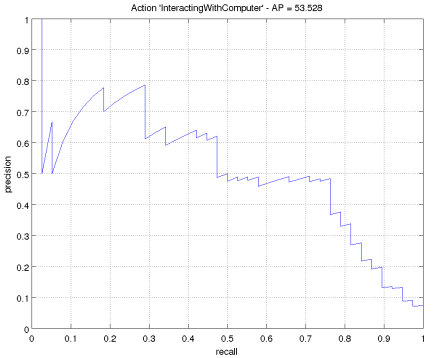
2.3 Accurracy and Precision-Recall

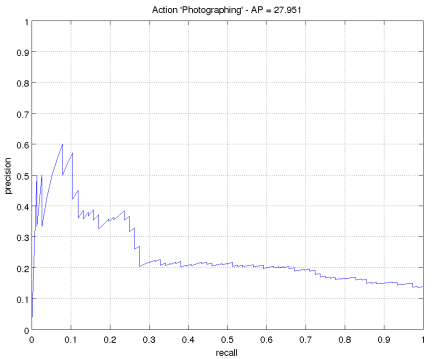
Classifier: SVM one VS all (C = 0.03389, J = 1), 5-fold cross-validation
Chi2 kernel: exp(-1/655.8768*$Chi2(X,Y)^2$)
Signature: Bag of features (K = 512)
Histogram normalization: None
K-means library: cpp
Channels:
(DENSE[spacing = 12-14-17-20-24-29-35-42-51-61, library: mylib]) x (SIFT[normalization: L2 (norm = 1, truncation over 0.2), library: colorDescriptor])

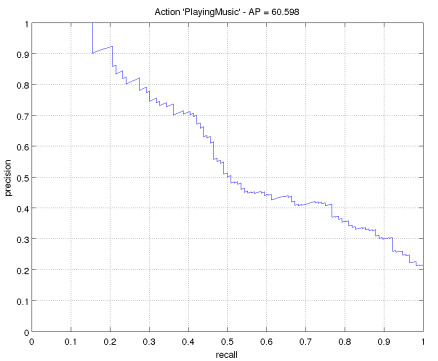
Table 13: SVM+BOF: Confusion table. The average accurracy is **48.8%**.

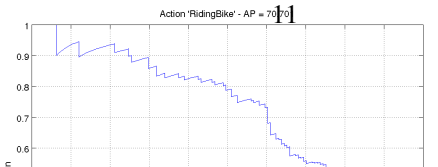
Actions	(1)	(2)	(3)	(4)	(5)	(6
(1) Interacting With Computer	**86.84**	0.00	7.89	2.63	0.00	0.
(2) Photographing	**32.89**	6.58	17.11	1.32	3.95	10
(3) Playing Music	22.41	5.17	**41.38**	13.79	5.17	0.
(4) Riding Bike	6.38	1.42	7.09	**61.70**	8.51	4.
(5) Riding Horse	12.28	8.77	15.79	7.02	**38.60**	7.
(6) Running	12.66	3.80	1.27	12.66	1.27	**46**
(7) Walking	4.20	2.52	3.36	1.68	11.76	21

Figure 4: SVM+BOF: Precision-Recall. The average precision is **51.8%**.









3 SVM classifier + Spatial Pyramid matching

3.1 Tested Parameters

3.1.1 Detectors

- MSDENSE: Multi-scale dense features.

3.1.2 Descriptors

- SIFT(L2): SIFT descriptor normalized with L2 norm.
- SIFT(L2T): SIFT descriptor normalized with L2 norm and truncation over 0.2.

3.1.3 Dictionnary size

Tested sizes: 256, 512, 1024.

3.1.4 Number of pyramid levels

Various number of levels were tested: 1, 2 and 3.

3.1.5 Histogram normalization

- NONE: no normalization.

- L1: normalized with L1 norm.
- L2: normalized with L2 norm.

3.1.6 Classifier

- 1vs1: One-versus-one classification
- 1vsA: One-versus-all classification

3.2 Results

3.2.1 Descriptor

Table 14: SVM+PYR: Average performances for different descriptors. The mean is computed over all other combinaison of parameters.

Descriptor	SIFT(L2)	SIFT(L2T)
Avg. perf.	$\mathbf{51.7 \pm 2.6}$	51.7 ± 2.7

3.2.2 Dictionnary size

Table 15: SVM+PYR: Average performances for different dictionnary size. The mean is computed over all other combinaison of parameters.

Dic. Size	256	512	1024
Avg. perf.	51.3 ± 2.6	$\mathbf{52.0 \pm 2.7}$	51.8 ± 2.8

3.2.3 Histogram normalization

Table 16: SVM+PYR: Average performances for different normalization of the signature histogram. The mean is computed over all other combinaison of parameters.

Histo. Norm	NONE	L1	L2
Avg. perf.	51.5 ± 2.7	50.8 ± 2.6	$\mathbf{52.8 \pm 2.4}$

3.2.4 Number of pyramid levels

Table 17: SVM+PYR: Average performances for different number of pyramid levels. The mean is computed over all other combinaison of parameters.

L	1	2	3
Avg. perf.	50.3 ± 2.3	52.4 ± 2.5	$\mathbf{52.8 \pm 2.6}$

3.2.5 Classifier

Table 18: SVM+PYR: Average performances for different strategy of classification. The mean is computed over all other combinaison of parameters.

Strategy	1vs1	1vsA
Avg. perf.	49.6 ± 1.5	$\mathbf{53.8 \pm 1.8}$

3.3 Accurracy and Precision-Recall

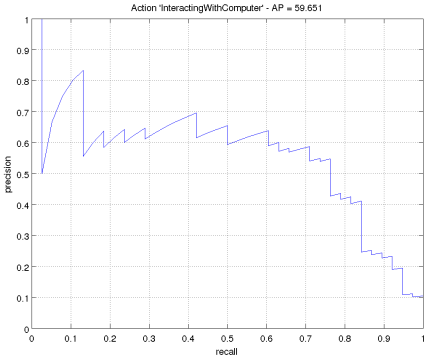
Classifier: SVM one VS all (C = 0.0625, J = 1), 5-fold cross-validation

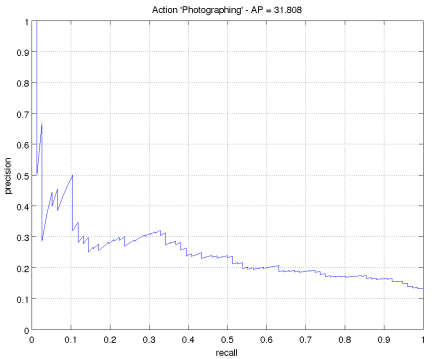
Intersection kernel: sum i(min(Xi,Yi))
Signature: Bag of features (K = 512)
histogram normalization: L2 (norm = 1)
K-means library: cpp
Channels:
(DENSE[spacing = 12-14-17-20-24-29-35-42-51-61, library: mylib]) x (SIFT[normalization: L2 (norm = 1), library: colorDescriptor])

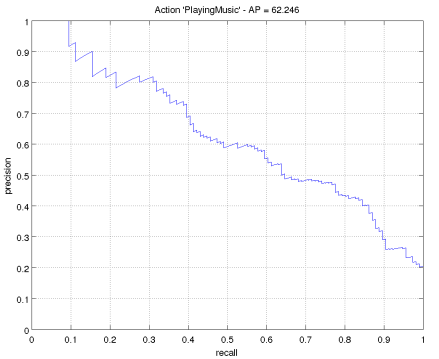
Table 19: SVM+PYR: Confusion table. The average accuracy is **55.9%**.

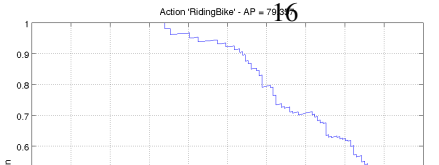
Actions	(1)	(2)	(3)	(4)	(5)	(
(1) Interacting With Computer	**81.58**	5.26	7.89	2.63	0.00	2.
(2) Photographing	15.79	18.42	18.42	5.26	7.89	14
(3) Playing Music	12.93	11.21	**46.55**	12.93	7.76	1.
(4) Riding Bike	2.13	1.42	1.42	**67.38**	12.06	3.
(5) Riding Horse	5.26	12.28	7.02	10.53	**50.88**	5.
(6) Running	6.33	3.80	1.27	10.13	0.00	**62**
(7) Walking	1.68	5.88	1.68	2.52	5.04	18

Figure 5: SVM+PYR: Precision-Recall. The average precision is **56.9%**.









4 LSVM

4.1 One component

Precision: 43.24%

Table 20: Confusion table for LSVM with one component.
The average accurracy is **45.55%**.

Actions	(1)	(2)	(3)	(4)	(5)	(6
(1) Interacting With Computer	**60.53**	0.00	2.63	7.89	2.63	15.
(2) Photographing	10.53	0.00	1.32	15.79	**30.26**	19.
(3) Playing Music	16.24	0.00	2.56	24.79	**37.61**	9.4
(4) Riding Bike	0.71	0.00	1.42	**64.54**	14.89	9.1
(5) Riding Horse	0.00	0.00	0.00	16.07	**60.71**	8.9
(6) Running	2.53	0.00	0.00	10.00	3.75	**73.**
(7) Walking	0.85	0.00	0.00	11.02	10.17	21.

4.2 Two components

Precision: 43.70%

Table 21: Confusion table for LSVM with two components.
The average accurracy is **46.18%**.

Actions	(1)	(2)	(3)	(4)	(5)	(
(1) Interacting With Computer	**34.21**	18.42	34.21	5.26	2.63	0.
(2) Photographing	3.95	**34.21**	25.00	10.53	13.16	3.
(3) Playing Music	13.68	11.11	**52.99**	9.40	7.69	1.
(4) Riding Bike	2.84	9.22	14.18	**59.57**	7.09	2.
(5) Riding Horse	0.00	7.14	23.21	10.71	**48.21**	3.
(6) Running	2.50	1.25	15.00	17.50	3.75	**50**
(7) Walking	1.68	6.78	16.10	5.93	8.47	16

Figure 6: LSVM: Precision-Recall. The average precision is **43.70%**.