Exploring Bag of Words Architectures for Facial Expression Recognition

?

Neutral or Sad

Karan Sikka, Tingfan Wu, Joshua Susskind and Marian Bartlett

Machine Perception Lab
University of California, San Diego

Motivation

- Advancements in BoW model
- Advantages over other methods
  - *Ex. Gabor, Local binary patterns*
- Recently applied to subordinate level classification problems
- Few previous studies and/or systematic evaluations

Goals

- Compare BoW to current approaches
  - Ex. LBP and Gabor
- Identify differences in BoW model for AFER vs. object (or scene) recognition
- Propose a BoW pipeline tailored to requirements of AFER
- Evaluate the contribution of each component of the proposed BoW pipeline

*AFER- Automatic Facial Expression Recognition

Challenges

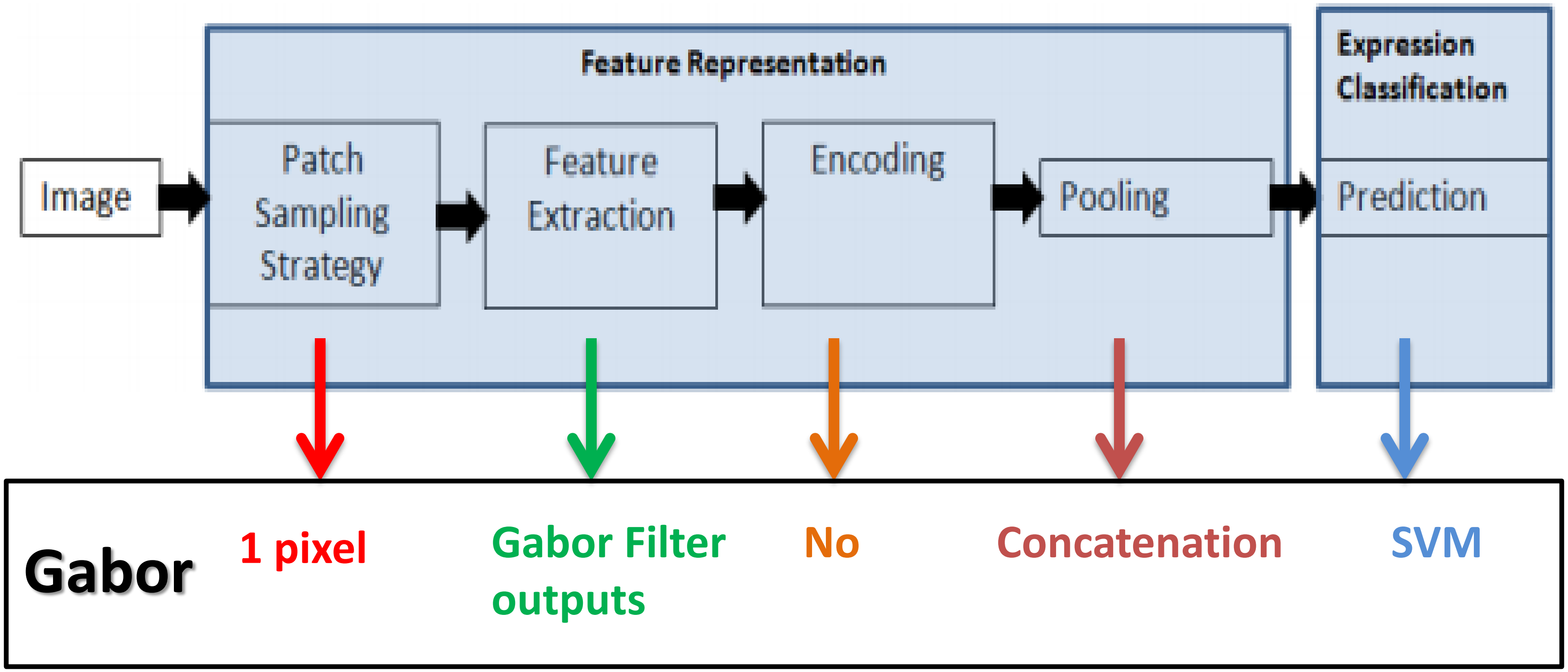
- Fundamental differences described between faces and objects*



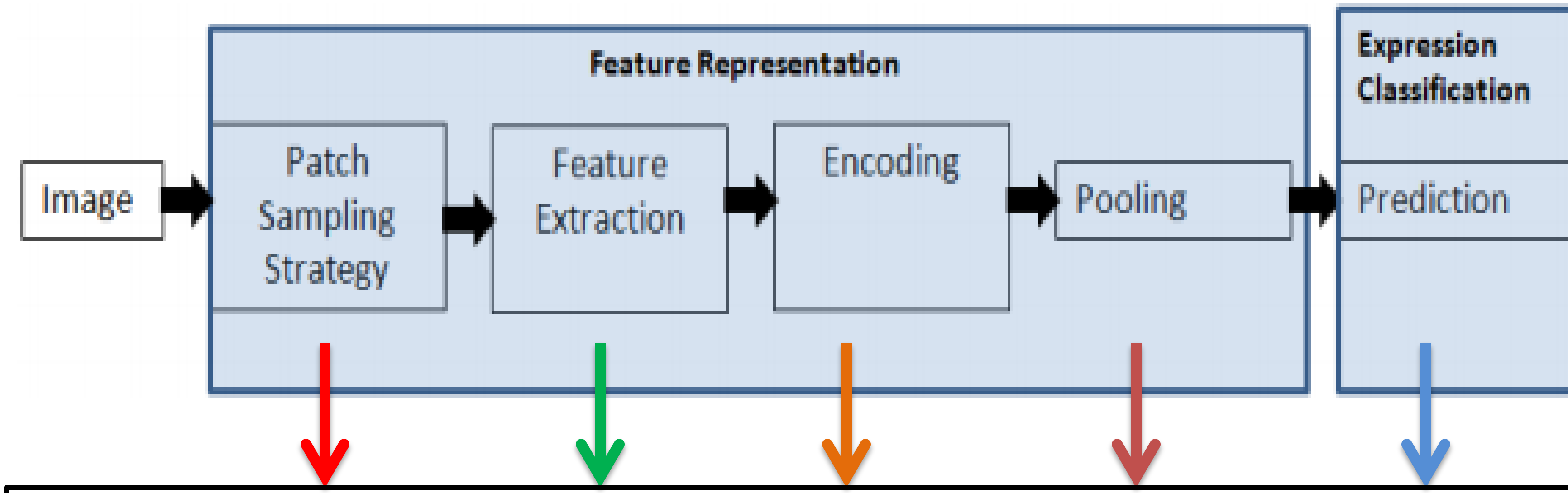
- BoW pipeline suited for objects may differ for faces

* Biederman et.al, Neurocomputational bases of objects and bases, Neurocomputational bases of objects and face recognition (1997).

Components of AFER Approach



Components of AFER Approach



LBP	**1 pixel**	**LBP**	**No**	**Sum-Pooling Over regular grids**	**SVM**

LBP: Local Binary Patterns

Related Works

Appearance based Discriminative Approaches

- Gabor wavelets:
  - Multi-scale-orientation features extracted densely at every pixel.
  - ❖ Lower spatial invariance relative to other features.

- LBP: Local Binary Patterns
  - Binary Histograms encoding local texture.
  - Features pooled over Rectangular region of support achieving higher spatial invariance.
  - ❖ Selecting grid-pattern is a non-trivial problem.

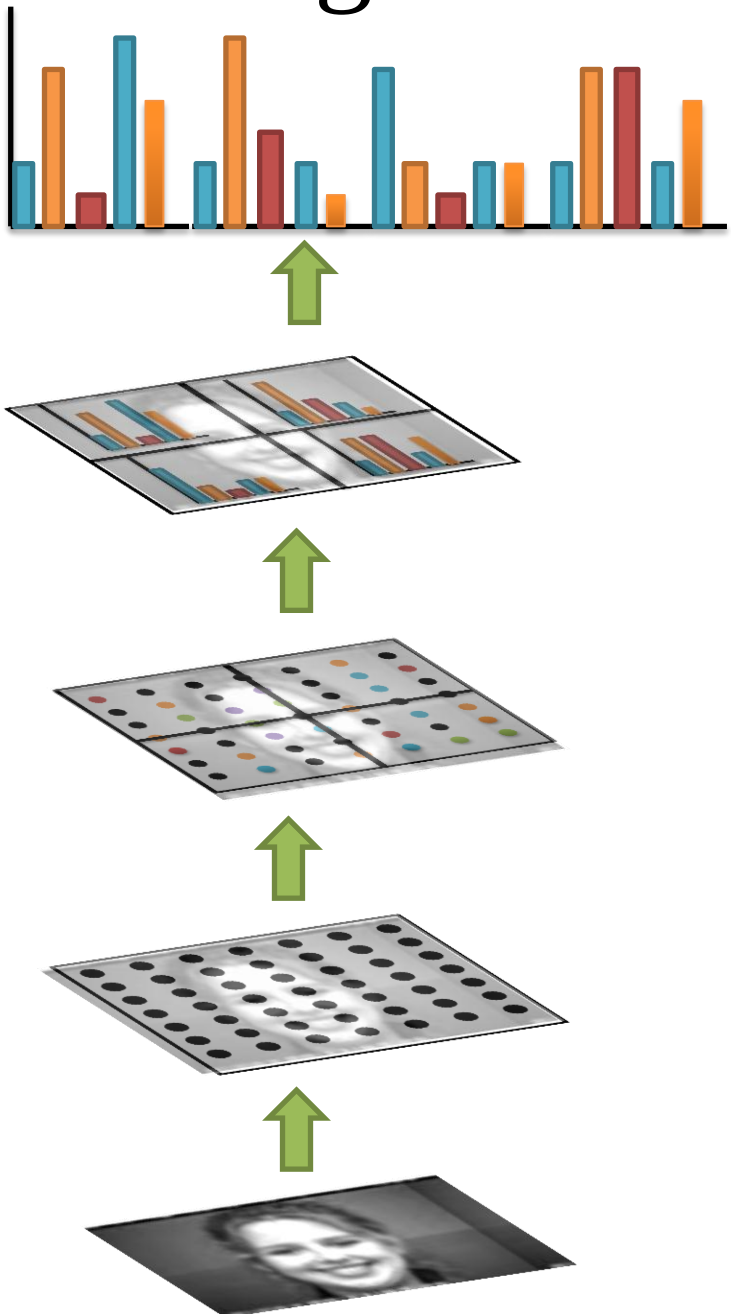
Related Works

- BoW + PHOG*:
  - Visual words pooled over 4 facial regions obtained via segmentation.
  - Fused PHOG features at classifier level.
  - ❖ BoW representation didn't give good performance alone.

- Unanswered question: if BoW has any coding advantages?

*Imai et.al, Facial-component-based bag of words and phog descriptor for facial expression recognition, IEEE Systems, Man and Cybernetics, 2009

Revisiting BoW Model



Final Feature

Pooling over
Spatial Pyramids

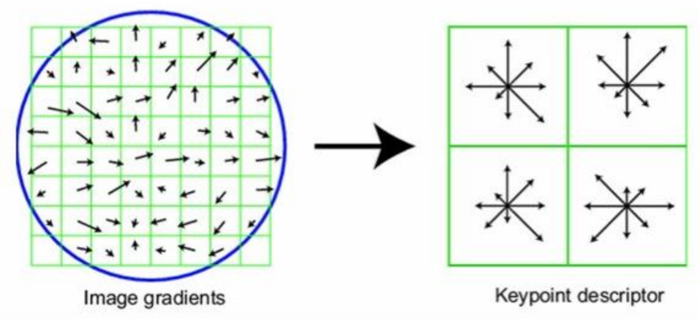
Encoding

Local
Descriptors

Image

Proposed Approach

- **Features**
  - SIFT- Scale invariant Feature Transform.
  - Histograms of gradient.



- **Sampling**
  - Dense or sparse (interest points) sampling.
  - (1) Interest point based features saturate* (2) Patches at fine-scales are most informative*.
  - Multi-scale dense SIFT- MSDF features.

*Nowak et.al, Sampling strategies for Bag of feature image classification, ECCV 2006

Proposed Approach

- **MSDF**
  - Dense: Features extracted every **2 pixels**.
  - Multi-scale: SIFT spatial bin set to 4, 8, 12, 16, 24.

- **Codebook**
  - Approximate k-means clustering.
  - Codebook size set to 800 (empirically).

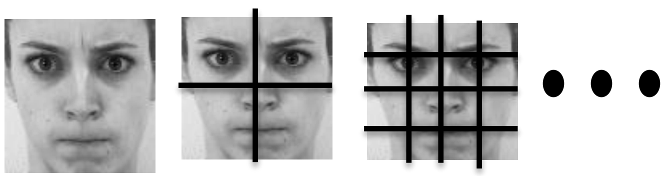
- **Encoding and Pooling**
  - Encoding and pooling is important for good classification*.
  - Employ LLC with max-pooling.

* Chatfield et.al, Devil is in the details, BMVC 2011

Proposed Approach

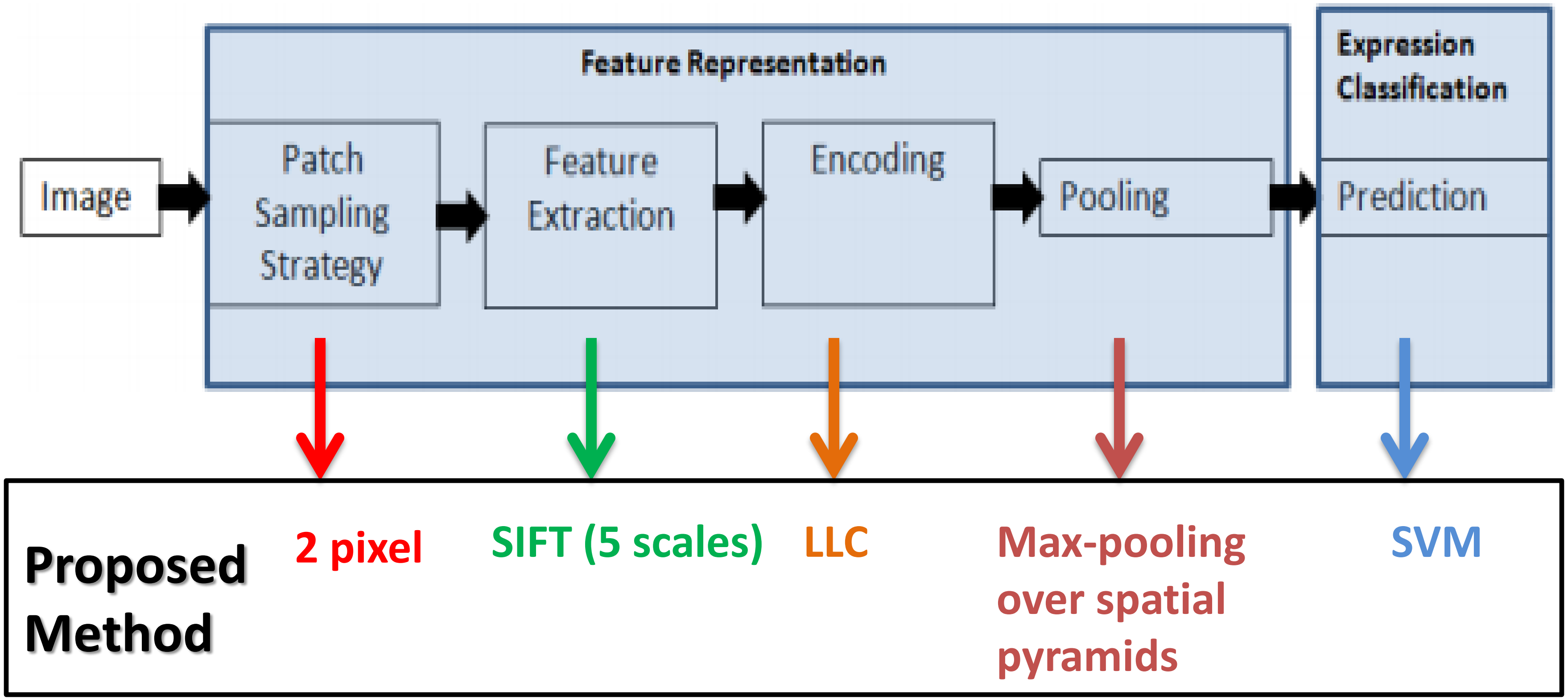
- LLC- Locality Constrained Linear Encoding*.
- Projects each descriptor to a subspace spanned by few codewords.

- **Spatial information**
  - Spatial Pyramid Matching (SPM) framework.
  - Advantage: Standard way to pool features. (vs LBP and BoW+PHOG).
  - Shown to work well and eliminates need to find the **best grid pattern**.

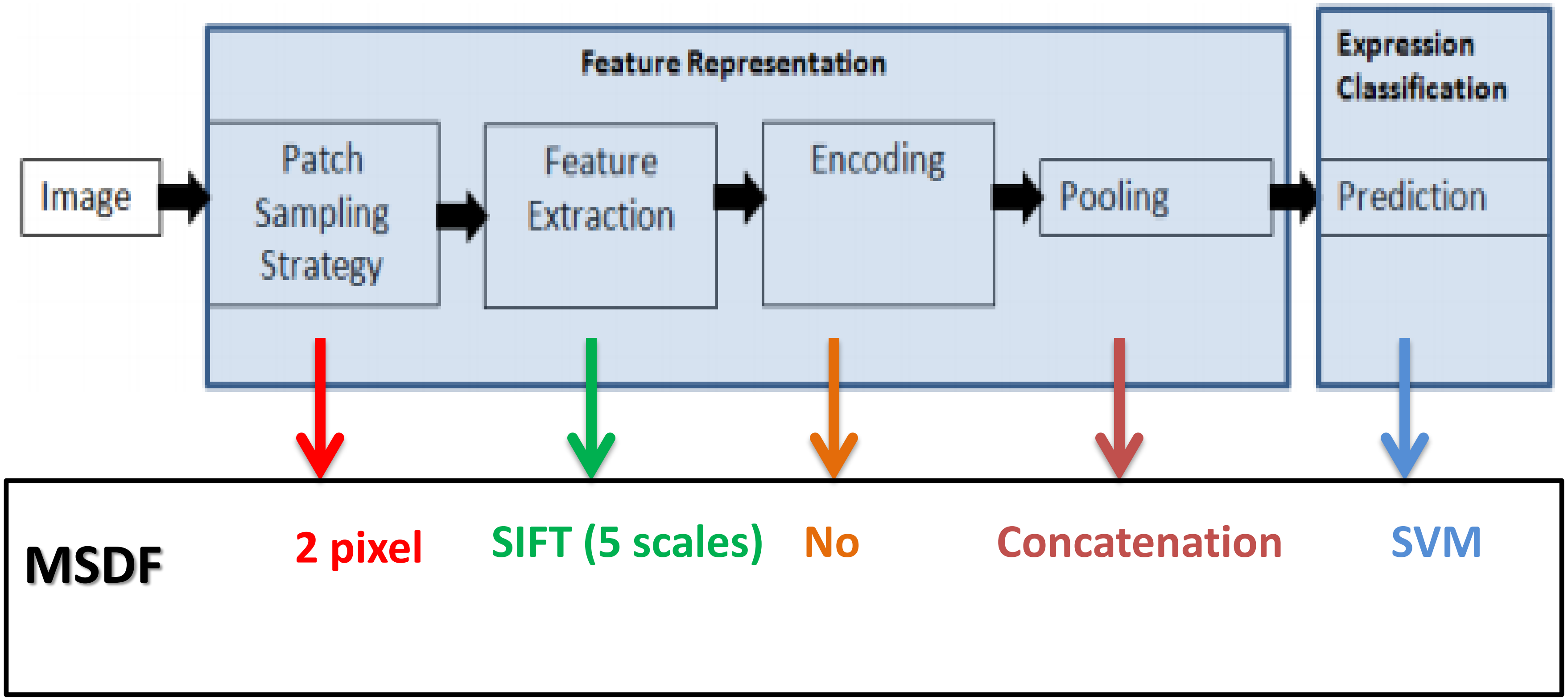


* J. Wang et.al, Locality-constrained linear coding for image classification, CVPR, 2010

Components of AFER Approach



Components of AFER Approach



Datasets

- **CK+**
  - 123 subjects.
  - Seven facial expressions.
  - 327 Samples (peak-frame).
  - Leave-one subject out validation
- **Adfes**
  - 22 subjects
  - Six basic emotions
  - 216 Samples (peak-frame).
  - 5 fold cross validation on subjects. (Balanced training set)

Comparison Architectures

- Pre-Processing:
  - Variant of Viola Jones detector.

- Gabor:
  - Gabor* (72 Filters) + Linear SVM.

- LBP:
  - Uniform LBP histograms
  - Best performing parameters selected for fair comparison.
  - Polynomial kernel SVM.

* Littlewort et.al, The computer expression recognition toolbox (cert), FG 2011

Results

DATASET	ADFES	CK+
Gabor	94.59 ± 2.61	91.81 ± 1.94
LBP	94.96 ± 1.96	82.38 ± 2.34
Proposed Method	96.30 ± 1.08	95.85 ± 1.40

- How does BoW compare to previous approaches?

Results

DATASET	ADFES	CK+
Gabor	94.59 ± 2.61	91.81 ± 1.94
LBP	94.96 ± 1.96	82.38 ± 2.34
Proposed Method	96.30 ± 1.08	95.85 ± 1.40

- BoW outperforms previous state of the art approaches.
- Thus BoW provides performance benefits for AFER.

Results

DATASET	ADFES	CK+
MSDF	92.59 ± 3.41	94.34 ± 1.62
Simple BoW	94.09 ± 2.32	92.67 ± 1.93
SS-SIFT + BoW	93.30 ± 1.13	93.28 ± 1.76
Proposed Method	96.30 ± 1.08	95.85 ± 1.40

- Does BoW gives any performance advantages over MSDF features.
  - Employed MSDF features without encoding and pooling (similar to Gabor).

Results

DATASET	ADFES	CK+
MSDF	92.59 ± 3.41	94.34 ± 1.62
Simple BoW	94.09 ± 2.32	92.67 ± 1.93
SS-SIFT + BoW	93.30 ± 1.13	93.28 ± 1.76
Proposed Method	96.30 ± 1.08	95.85 ± 1.40

- "BoW provides performance benefits beyond MSDF features"
  - MSDF has lower performance compared to **proposed method** involving **BoW**.

Results

DATASET	ADFES	CK+
MSDF	92.59 ± 3.41	94.34 ± 1.62
Simple BoW	94.09 ± 2.32	92.67 ± 1.93
SS-SIFT + BoW	93.30 ± 1.13	93.28 ± 1.76
Proposed Method	96.30 ± 1.08	95.85 ± 1.40

- How does Multi-scale SIFT (MSDF) compare to single scale SIFT (SS-SIFT)"
  - Employed SS-SIFT instead of MSDF with the proposed pipeline.

Results

DATASET	ADFES	CK+
MSDF	92.59 ± 3.41	94.34 ± 1.62
Simple BoW	94.09 ± 2.32	92.67 ± 1.93
SS-SIFT + BoW	93.30 ± 1.13	93.28 ± 1.76
Proposed Method	96.30 ± 1.08	95.85 ± 1.40

- "Multi-scale SIFT (MSDF) are better than single scale SIFT (SS-SIFT)"
  - MSDF features give 3% advantage over single scale features.

Results

DATASET	ADFES	CK+
MSDF	92.59 ± 3.41	94.34 ± 1.62
Simple BoW	94.09 ± 2.32	92.67 ± 1.93
SS-SIFT + BoW	93.30 ± 1.13	93.28 ± 1.76
Proposed Method	96.30 ± 1.08	95.85 ± 1.40

- Is ***LLC + max-pooling*** better than ***simple voting + sum-pooling*** (simple BoW) for AFER.

Results

DATASET	ADFES	CK+
MSDF	92.59 ± 3.41	94.34 ± 1.62
Simple BoW	94.09 ± 2.32	92.67 ± 1.93
SS-SIFT + BoW	93.30 ± 1.13	93.28 ± 1.76
Proposed Method	96.30 ± 1.08	95.85 ± 1.40

- "***LLC + max-pooling*** is better than ***simple voting + sum-pooling*** (simple BoW)".
  - LLC with max-pooling lead to significant improvement.

Results

DATASET	ADFES	CK+
MSDF	92.59 ± 3.41	94.34 ± 1.62
Simple BoW	94.09 ± 2.32	92.67 ± 1.93
SS-SIFT + BoW	93.30 ± 1.13	93.28 ± 1.76
Proposed Method	96.30 ± 1.08	95.85 ± 1.40

- “ Most substantial benefit by Spatial Pyramids”

Results

DATASET	ADFES	CK+
MSDF	92.59 ± 3.41	94.34 ± 1.62
Simple BoW	94.09 ± 2.32	92.67 ± 1.93
SS-SIFT + BoW	93.30 ± 1.13	93.28 ± 1.76
Proposed Method	96.30 ± 1.08	95.85 ± 1.40

- " Most substantial benefit by Spatial Pyramids"
  - Without SPM performance dropped from 95.9% to 83.1% for CK+.

Conclusion

- Explored application of BoW for AFER.
- Spatial information provided by SPM
  - Performance drops significantly without it.
- Employed highly discriminative MSDF features
  - Multi-scale SIFT better than single-scale SIFT.
  - Non-linearities introduced in BoW provide performance benefit beyond MSDF features.
- Application of novel encoding and pooling strategies for AFER
  - Better than traditional histogramming techniques.

Questions?

Thanks

Karan Sikka

Tingfan Wu

Machine Perception Lab
UCSD

Dr. Joshua
Susskind

Dr. Marian S. Bartlett