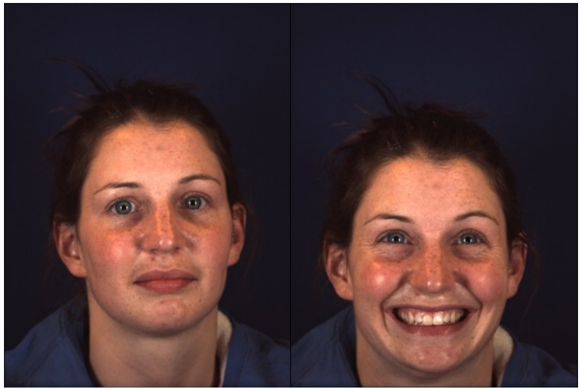
11th IEEE International Conference on
Automatic Face and Gesture Recognition
FG2015

BINGHAMTON
UNIVERSITY
State University of New York

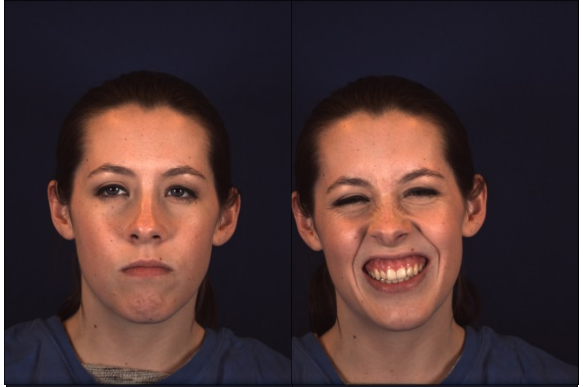
Spontaneous Facial Expression Analysis Based on Temperature Changes and Head Motions

Peng Liu and Lijun Yin
State University of New York-at Binghamton

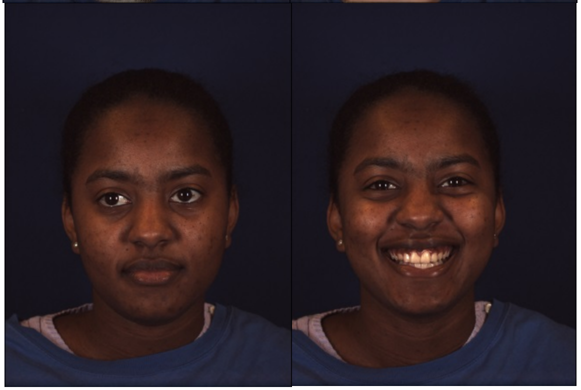
Motivation



Fear

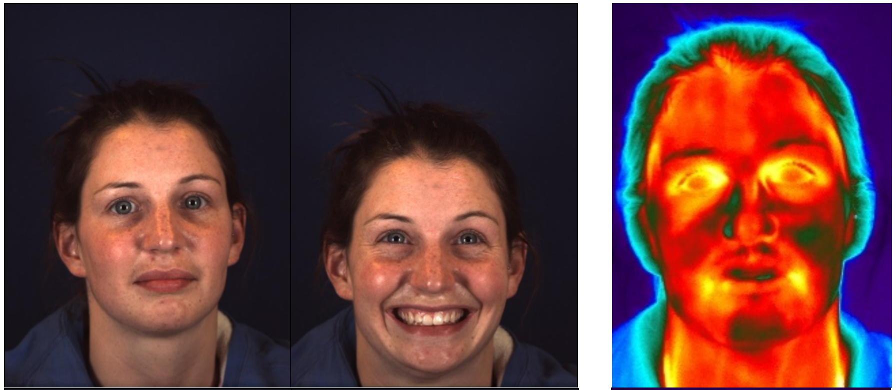


disgust

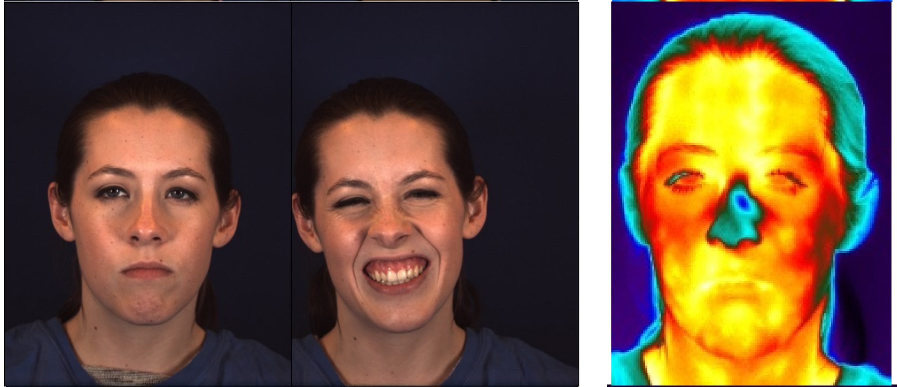


Happy

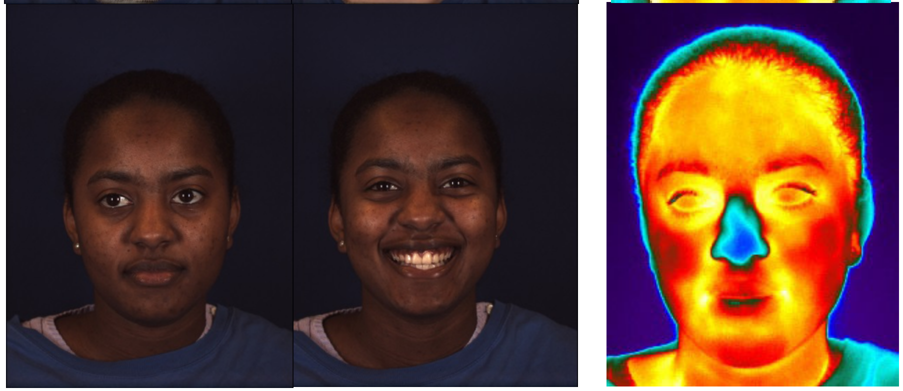
Motivation



Fear

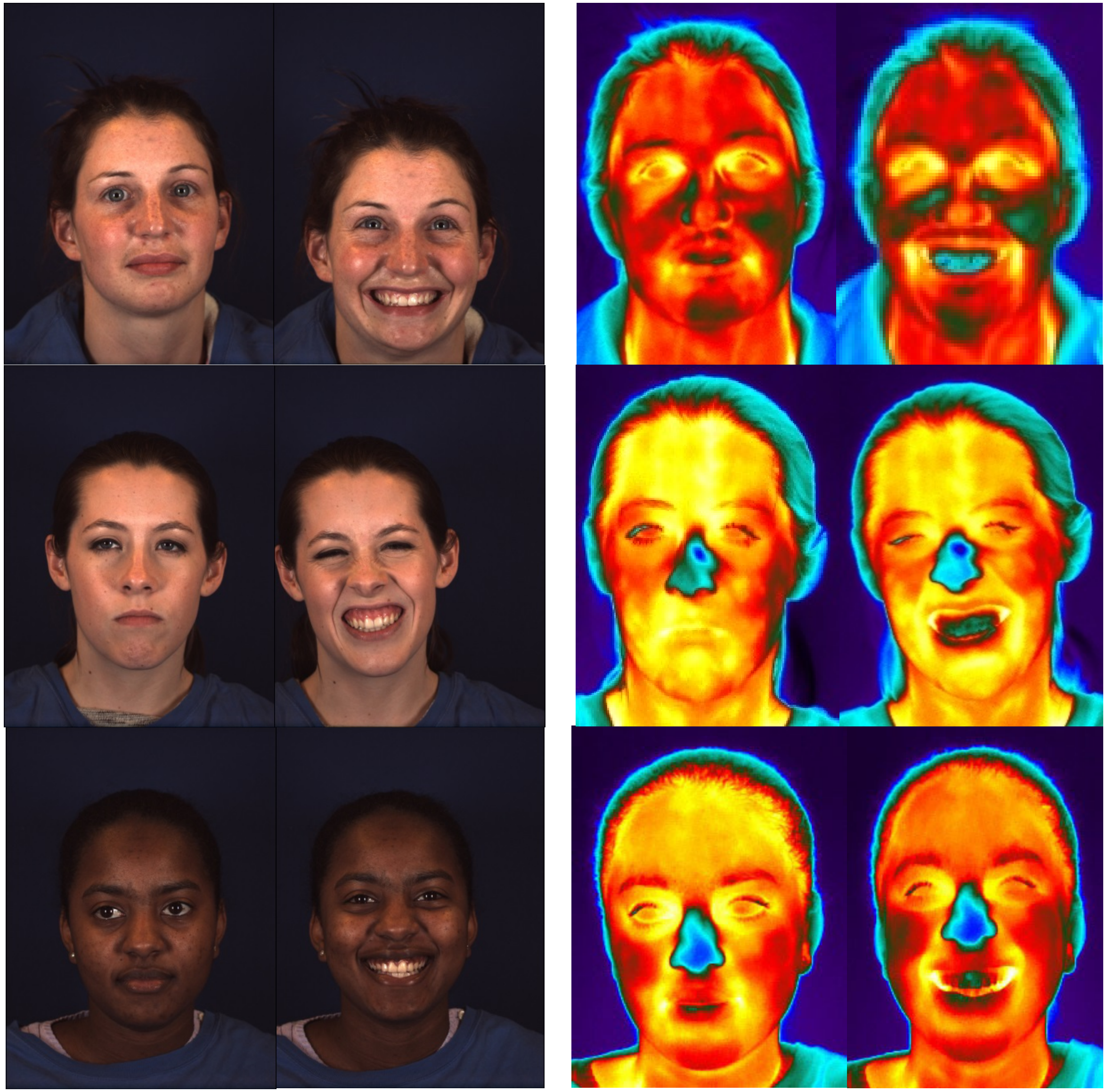


disgust



Happy

Motivation

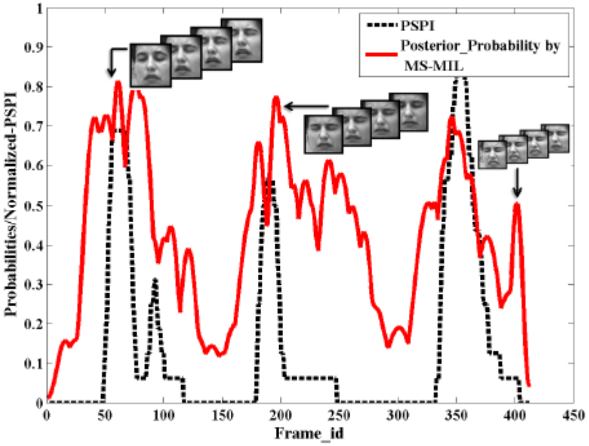


Fear

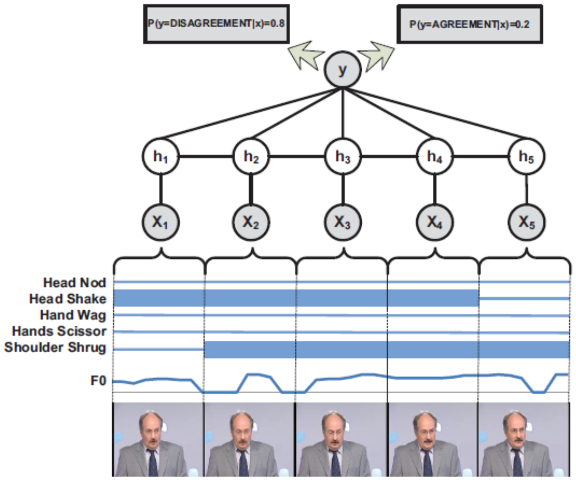
disgust

Happy

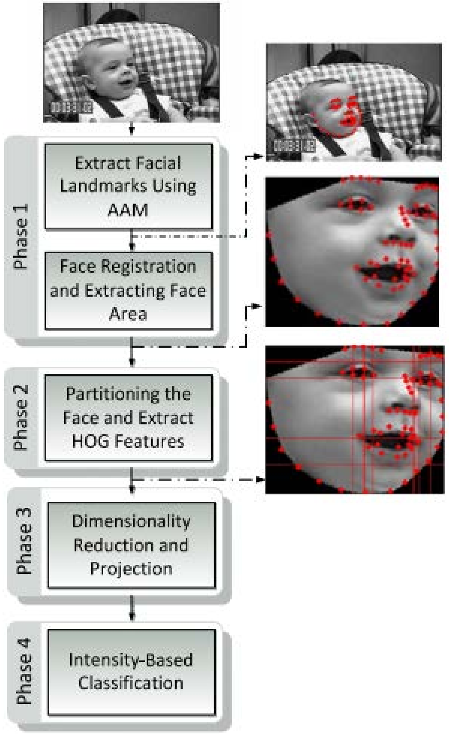
Prior works



Video clips,
Multiple instance learning
[K. Sikka et.al. *FG 2013*]



head nod, head shake and hand wag,
Hidden Conditional Random Field
[K. Bousmalis et.al. *FG 2011*]

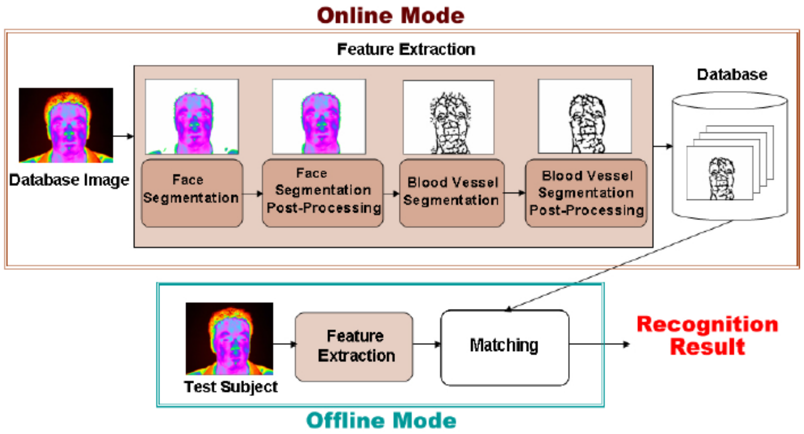


Phase 1
Extract Facial Landmarks Using AAM
Face Registration and Extracting Face Area
Phase 2
Partitioning the Face and Extract HOG Features
Phase 3
Dimensionality Reduction and Projection
Phase 4
Intensity-Based Classification

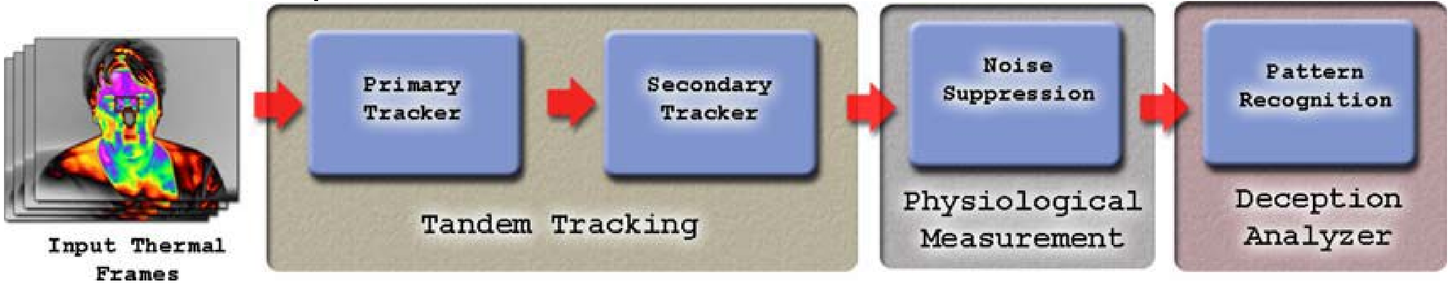
facial expression interaction of infants with their mothers, specific Action Units [N. Zaker et.al. *FG 2013*]

Prior works

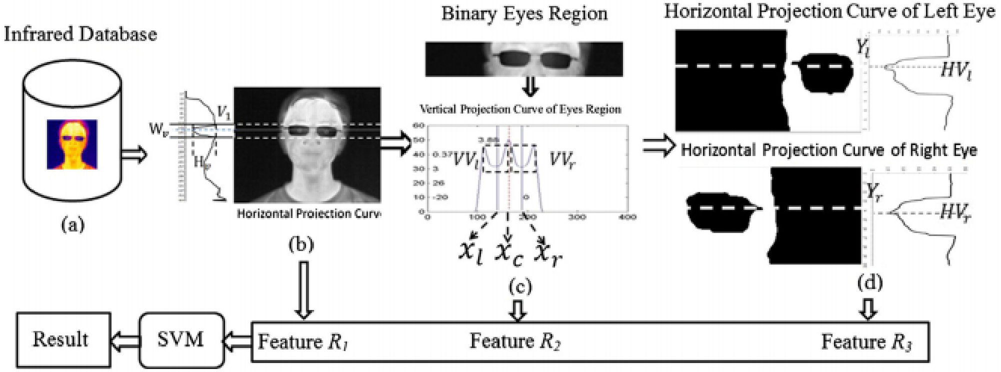
Face recognition [P. Buddharaju et.al. *CVPR 2009]*



Detection of Deceit
[P. Tsiamyrtzis et.al. *IJCV 2007*]



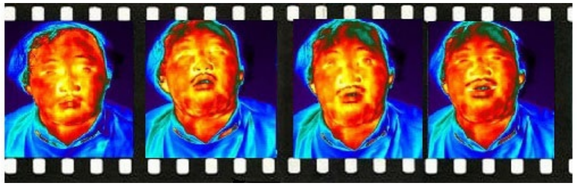
Eye localization
[S. Wang et.al.
Pattern Recognition
2013]



Overview of our approach

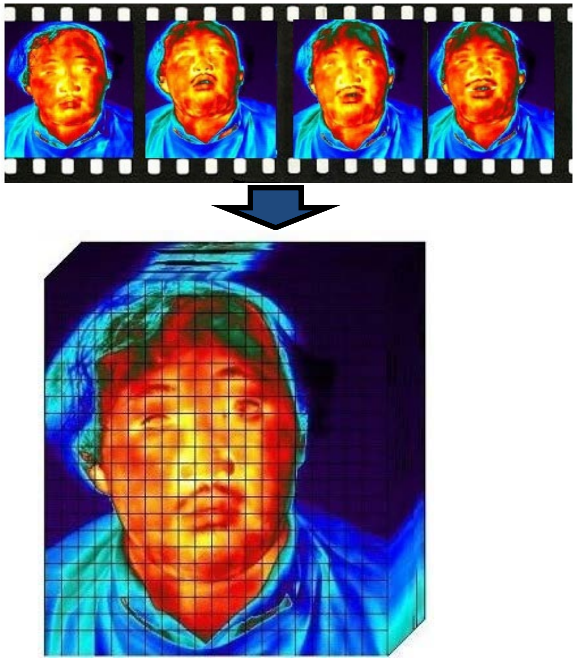
Overview of our approach

Training video



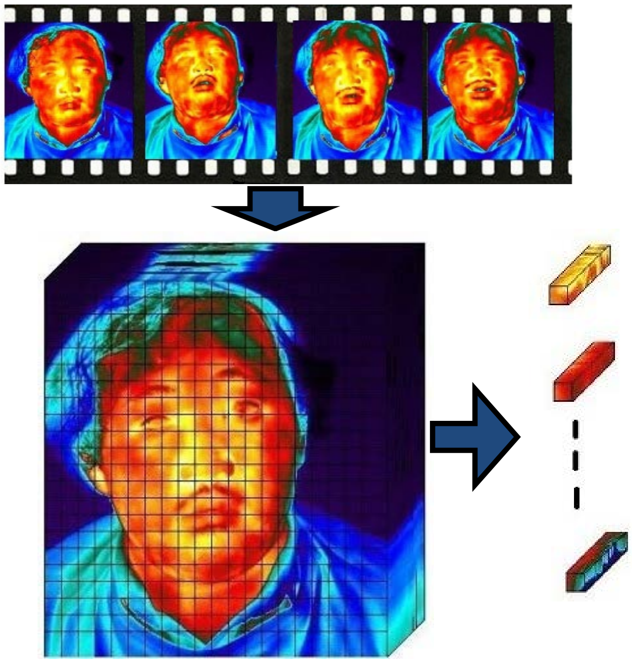
Overview of our approach

Training video

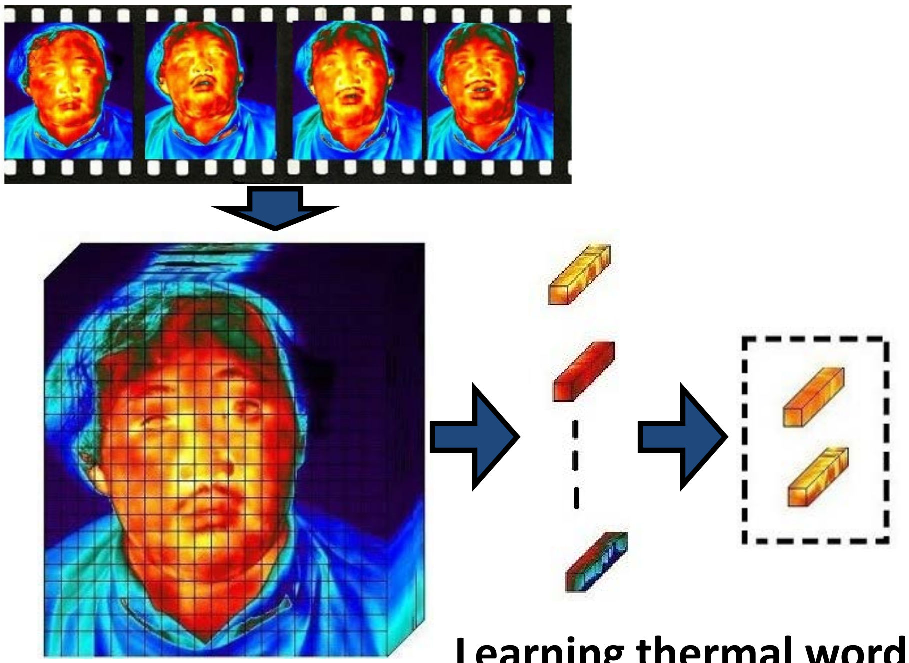


Overview of our approach

Training video



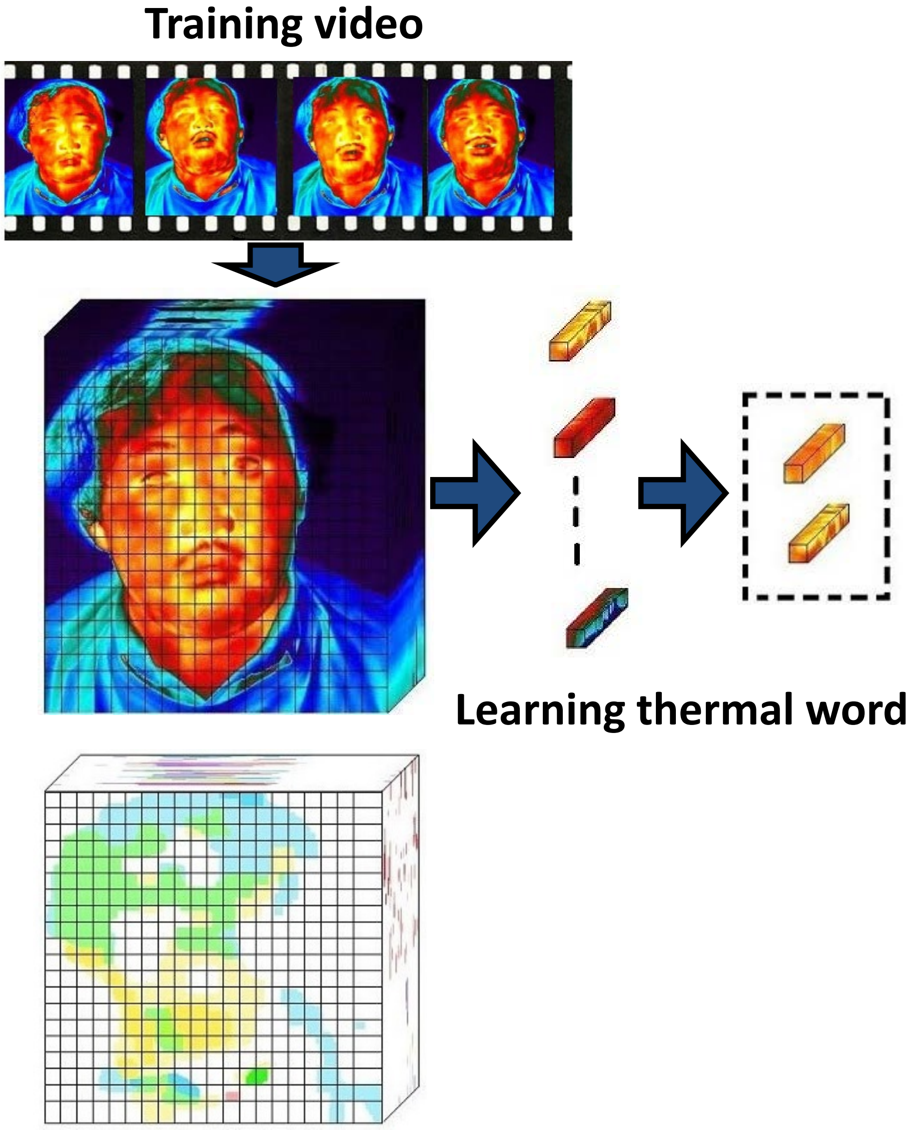
Overview of our approach



Training video

Learning thermal word

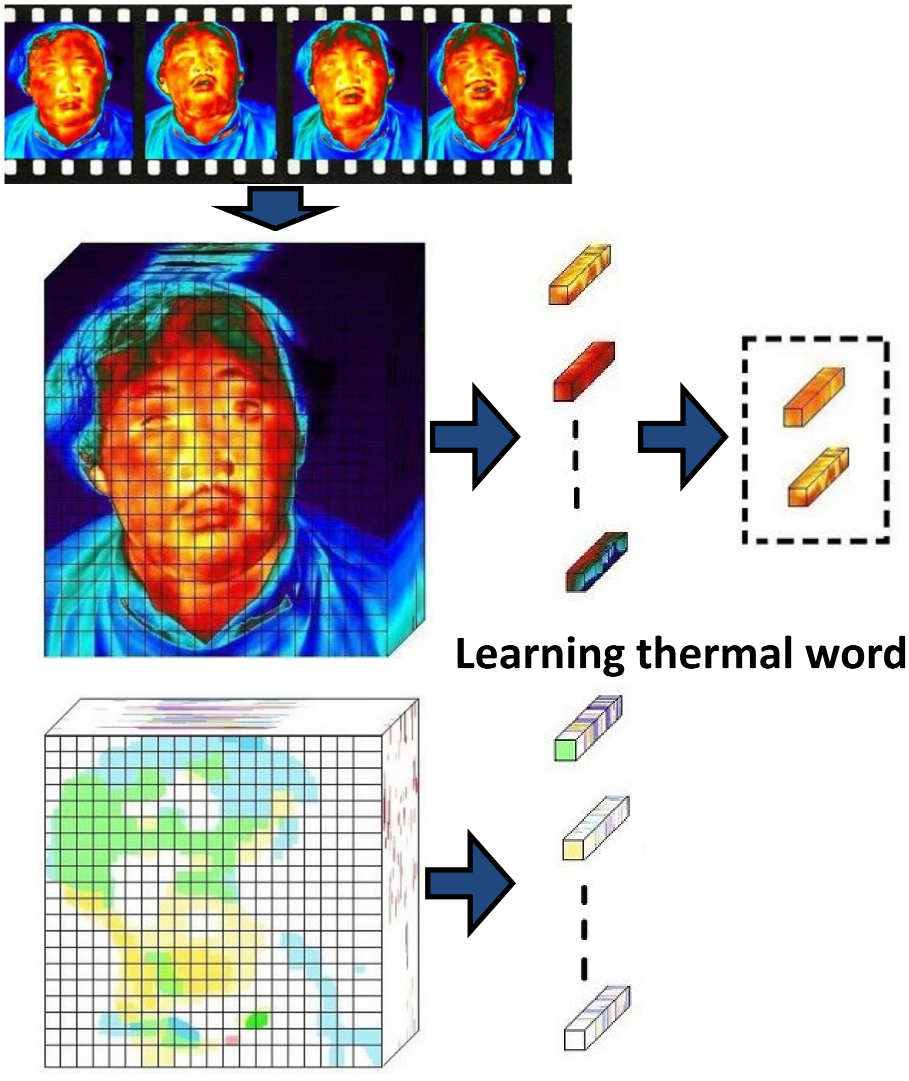
Overview of our approach



Training video

Learning thermal word

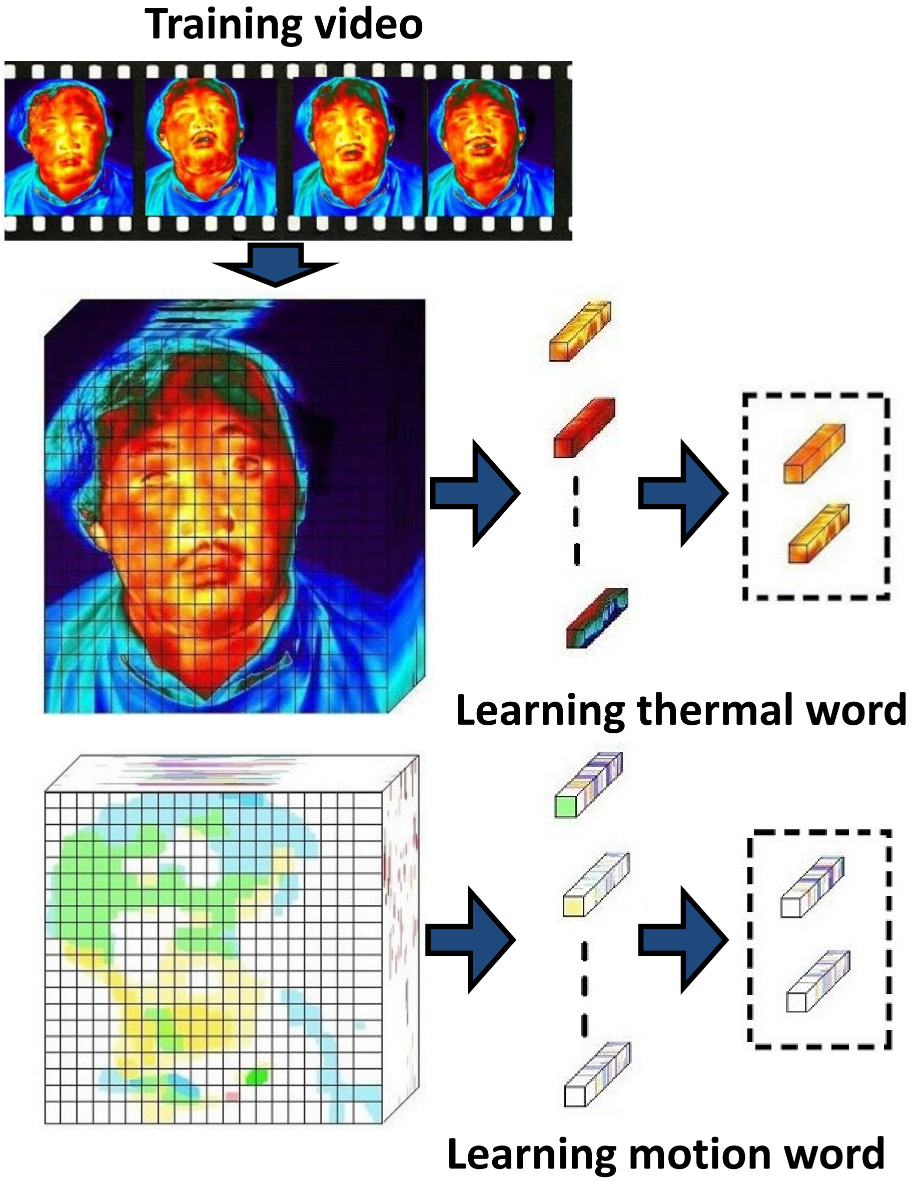
Overview of our approach



Training video

Learning thermal word

Overview of our approach

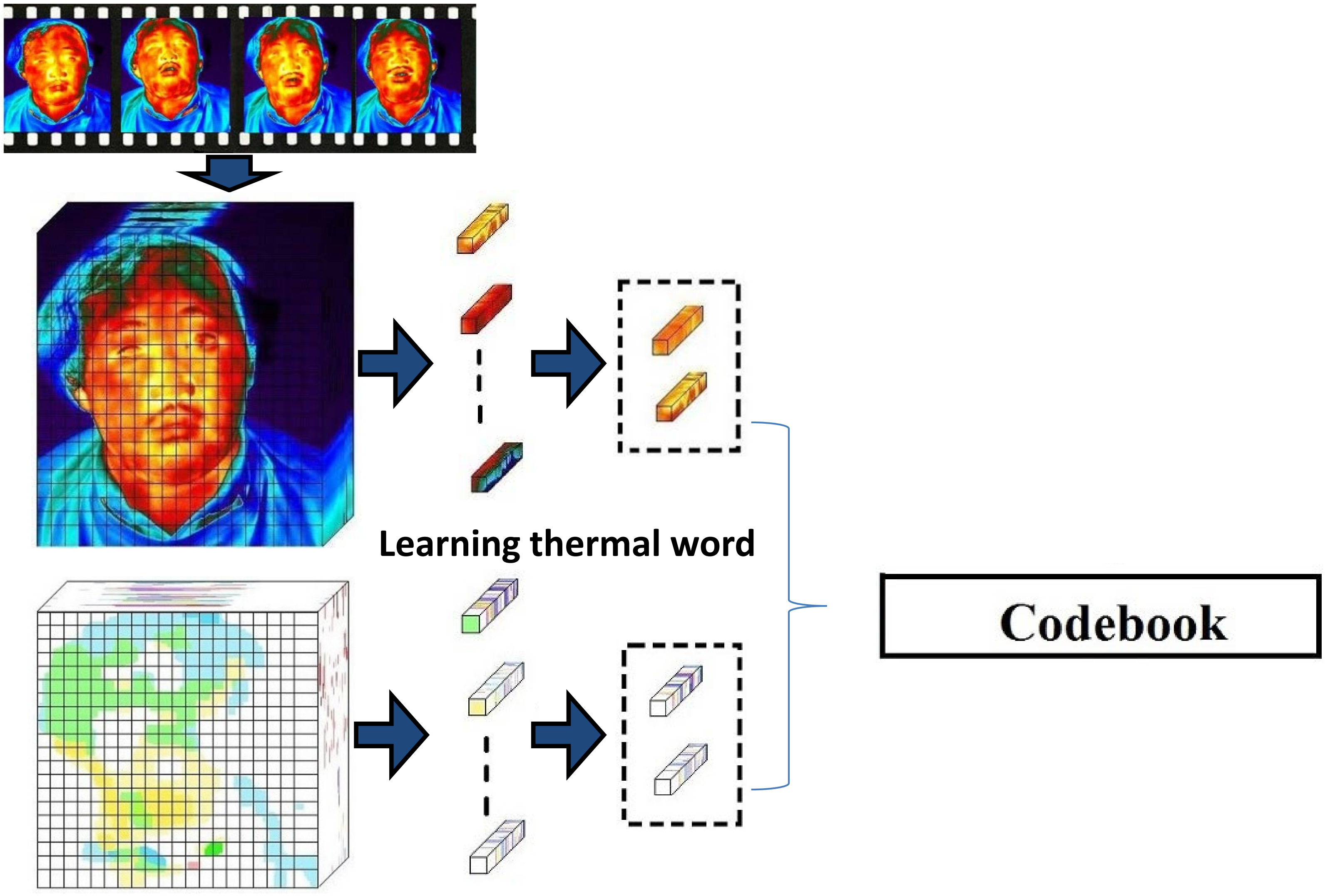


Training video

Learning thermal word

Learning motion word

Overview of our approach



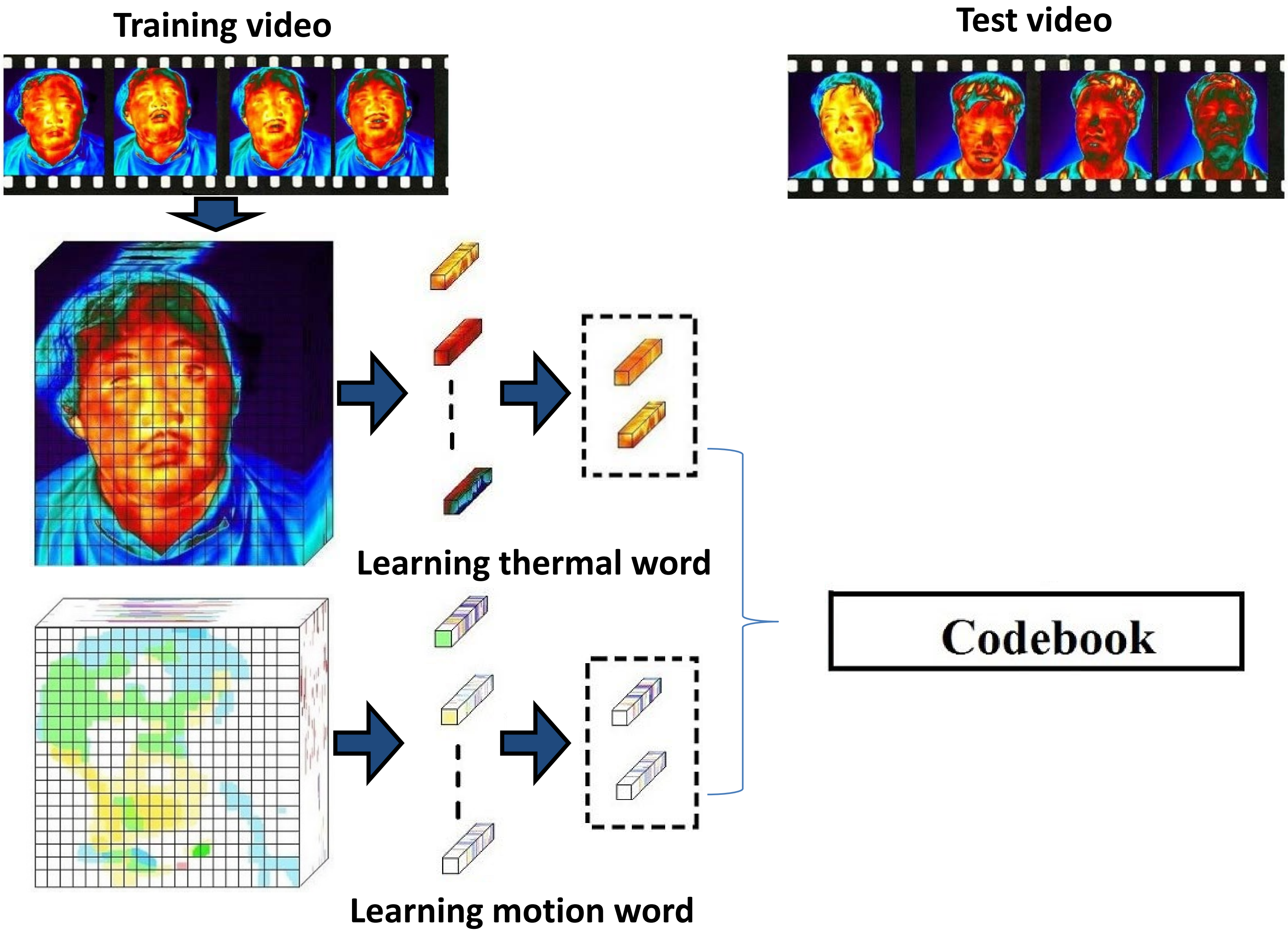
Training video

Learning thermal word

Codebook

Learning motion word

Overview of our approach



Training video

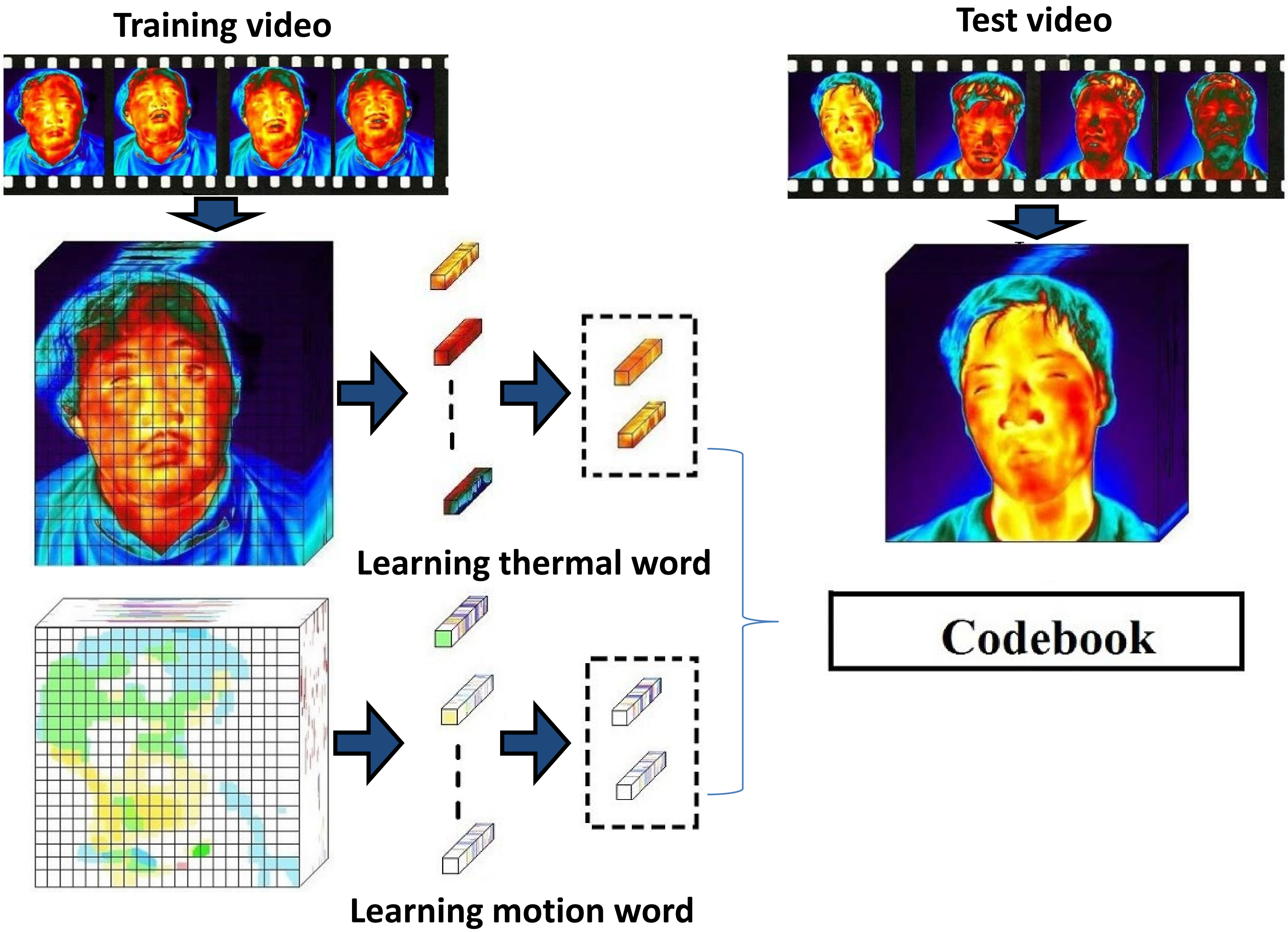
Test video

Learning thermal word

Codebook

Learning motion word

Overview of our approach



Training video

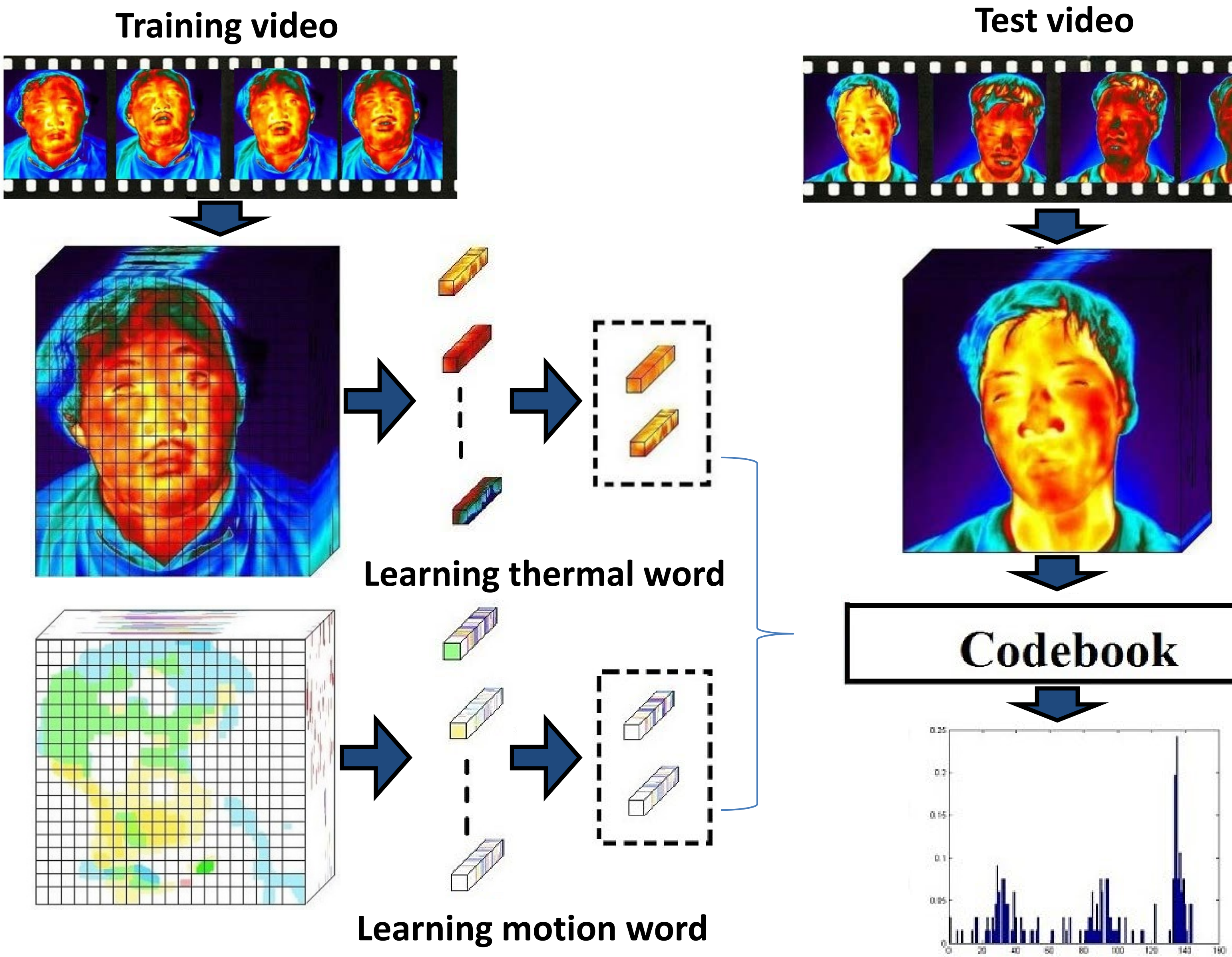
Test video

Learning thermal word

Learning motion word

Codebook

Overview of our approach



Training video

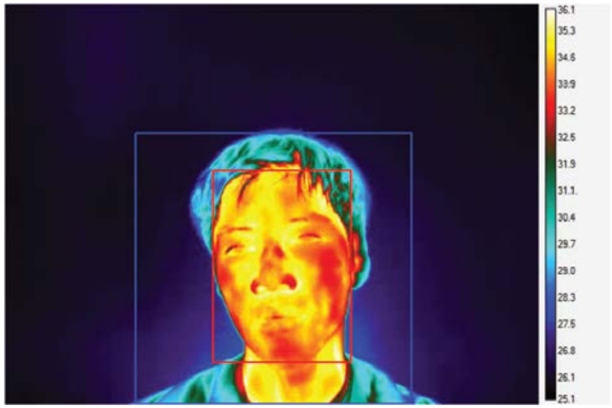
Test video

Learning thermal word

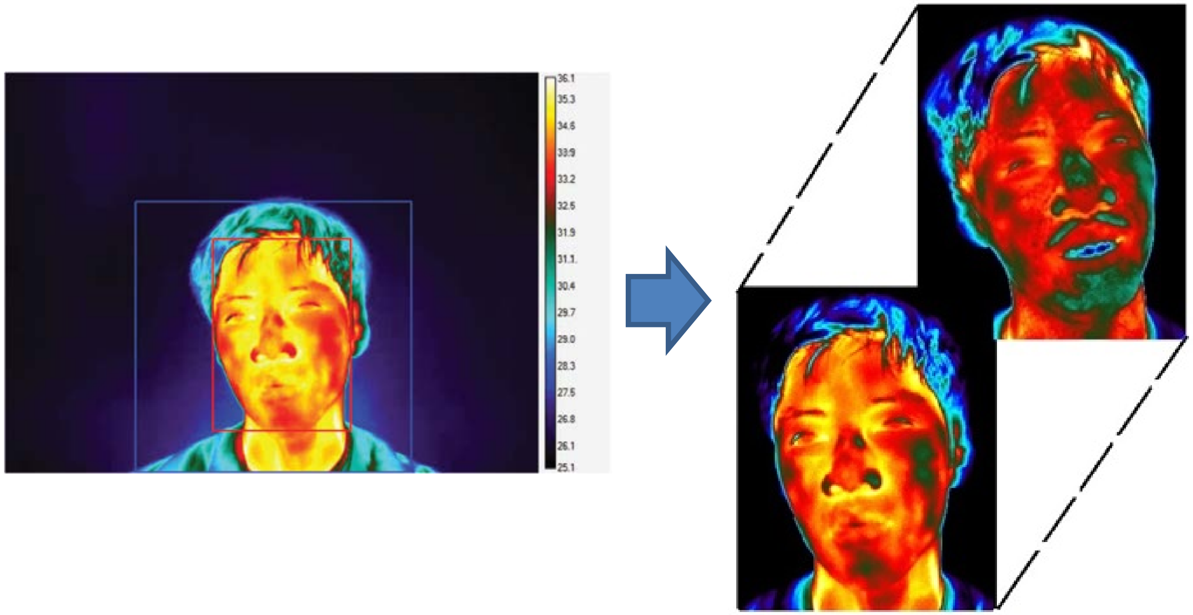
Learning motion word

Codebook

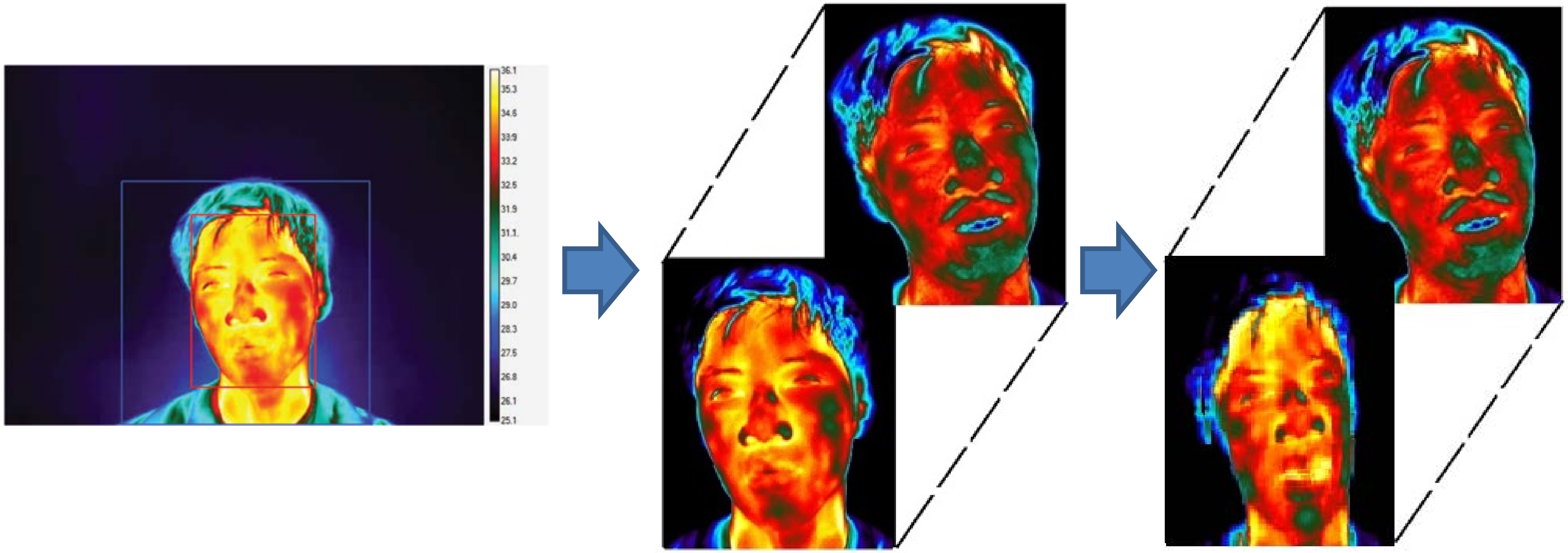
Face region alignment and warping



Face region alignment and warping



Face region alignment and warping



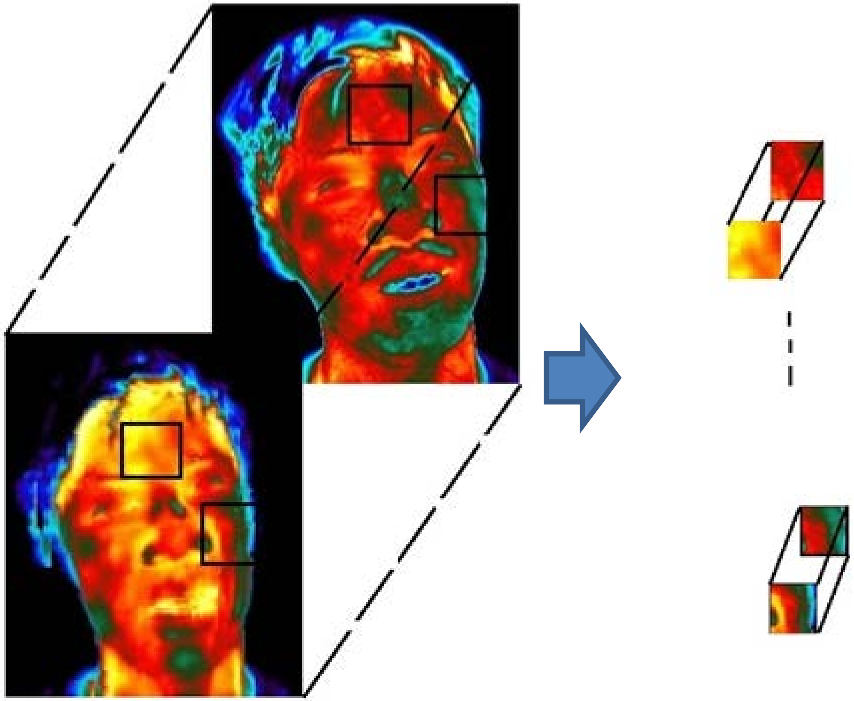
[C. Liu et.al. Trans. *PAMI 2011*]

$$E(w) = \sum_{\boldsymbol{p}} \min(||f(T_Q(\boldsymbol{p}) - f(T_R(\boldsymbol{p}' + w(\boldsymbol{p}))))||_1, t) +$$

$$\sum_{\boldsymbol{p}} \eta(|u(\boldsymbol{p})| + |v(\boldsymbol{p})|) +$$

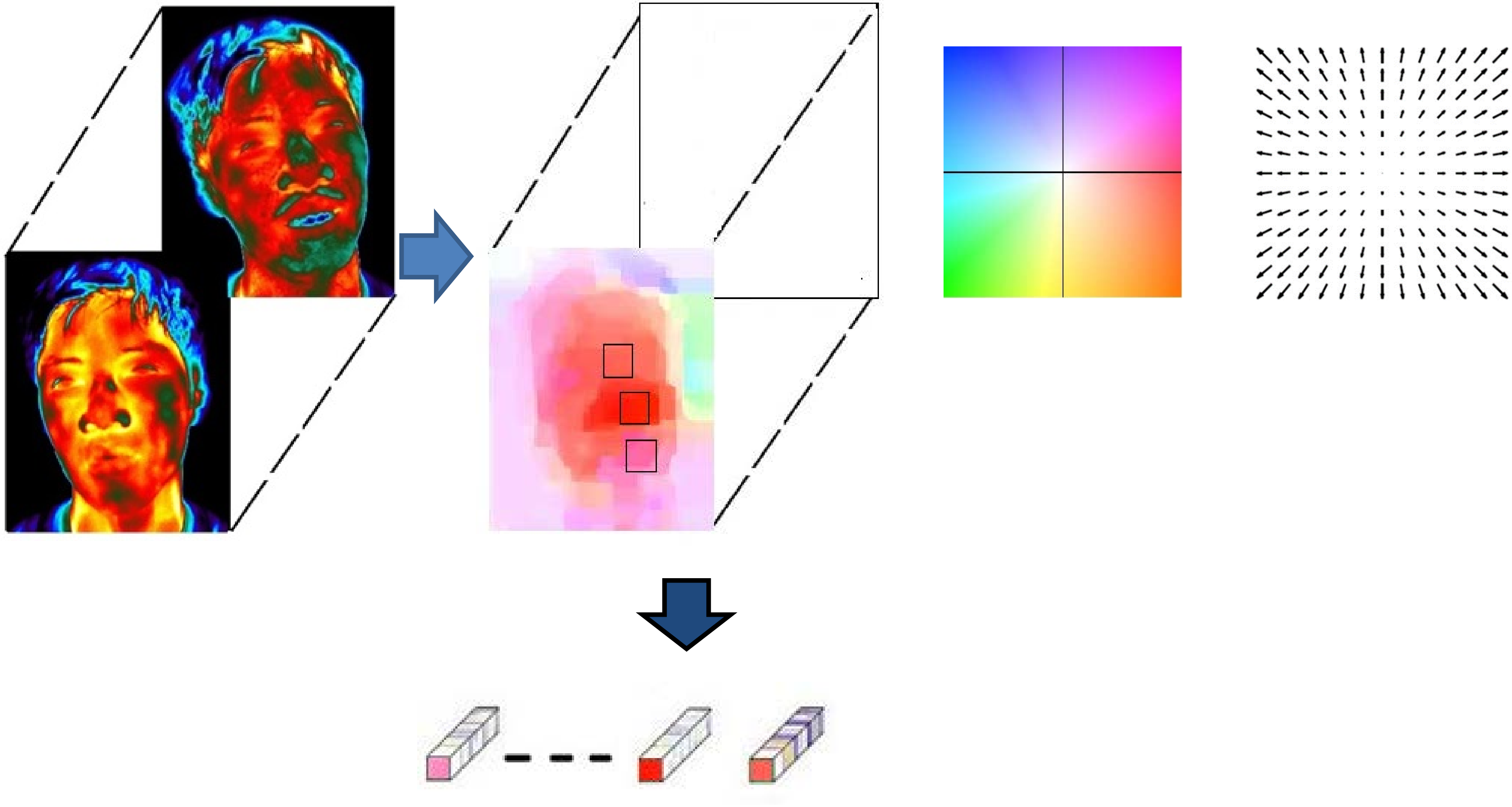
$$\sum_{(\boldsymbol{p},\boldsymbol{q})\in N} min(\alpha|u(\boldsymbol{p}) - u(\boldsymbol{q})|, d) + min(\alpha|v(\boldsymbol{p}) - v(\boldsymbol{q})|, d)$$

Max pooling the most distinguished cubic



$$y_j = \max\{|x_{ij}|, |x_{2j}|, \dots, |x_{Nj}|\}$$

Visualization of SIFT flow and motion video cubic



$$\hat{y}_j = \text{mean}\{|\hat{x}_{1j}|, |\hat{x}_{2j}|, \dots, |\hat{x}_{Nj}|\}$$

Thermal video descriptor

Based on the codebook, the test video is represented by the thermal video word and SIFT flow motion video word.

$$H^T = (\frac{N_{\Delta t1}}{N_T}, \frac{N_{\Delta t2}}{N_T}, \dots, \frac{N_{\Delta tn}}{N_T})$$

$$H^M = (\frac{N_{\Delta v1}}{N_M}, \frac{N_{\Delta v2}}{N_M}, \dots, \frac{N_{\Delta vn}}{N_M})$$

$$H = \{H^T, H^M\}$$

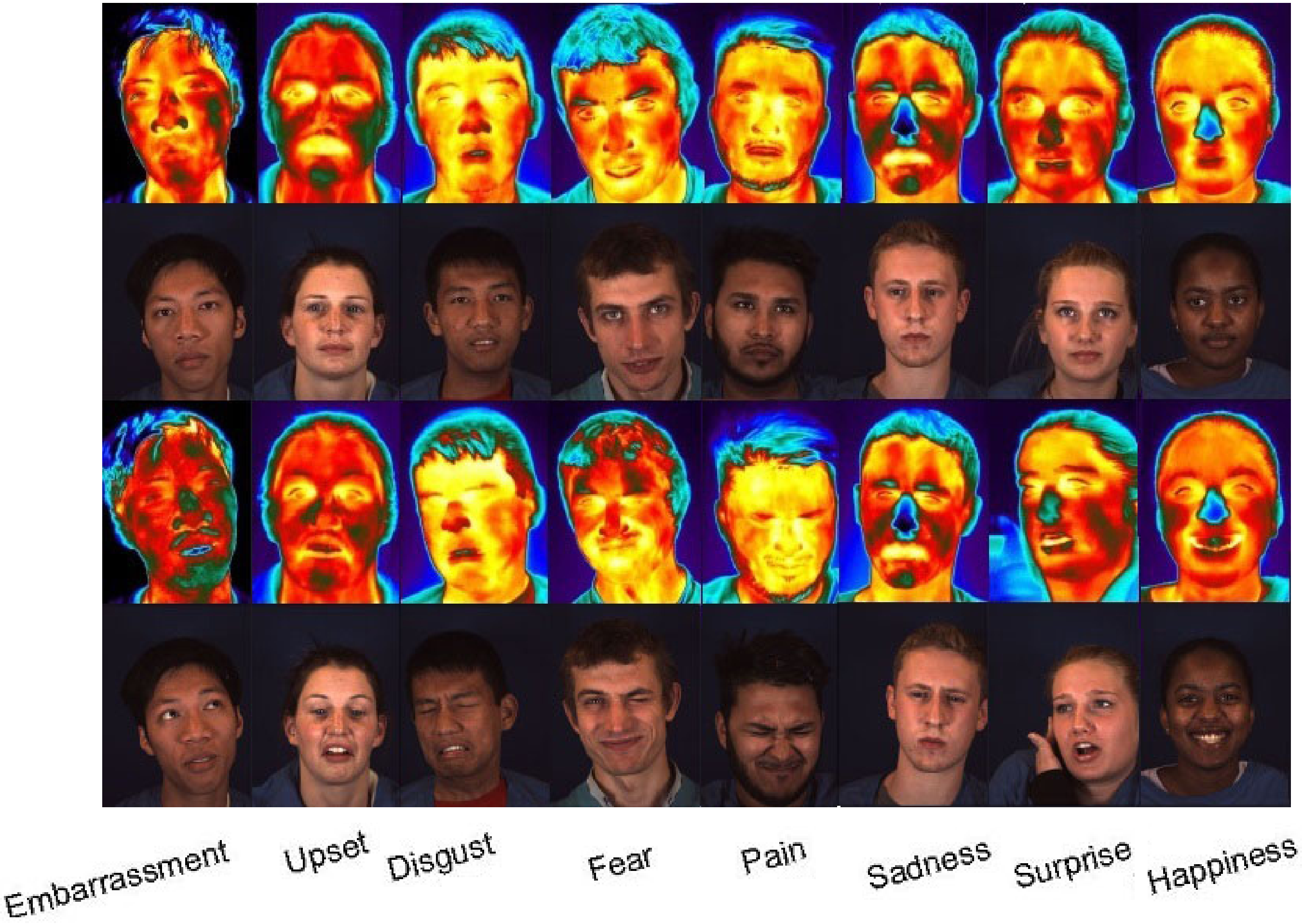
H^T is the histogram of thermal video words

H^M is the histogram of motion video words

N_T is the number of thermal video cubic extracted from thermal video clips.

N_M is the number of motion video cubic extracted from motion video clips.

Experiments and Evaluation



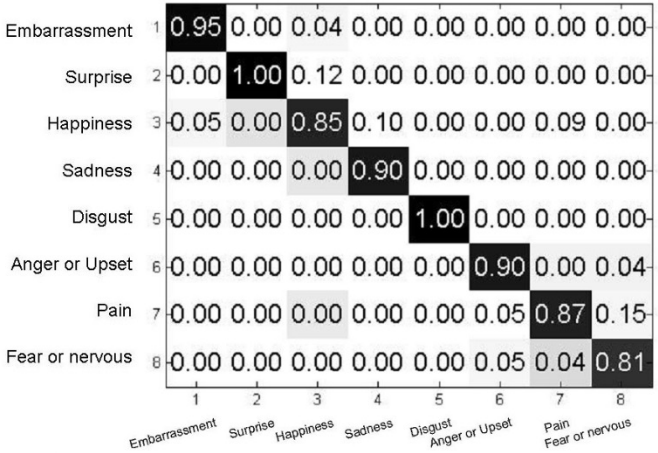
Neutral thermal

Neutral texture

Expression Thermal

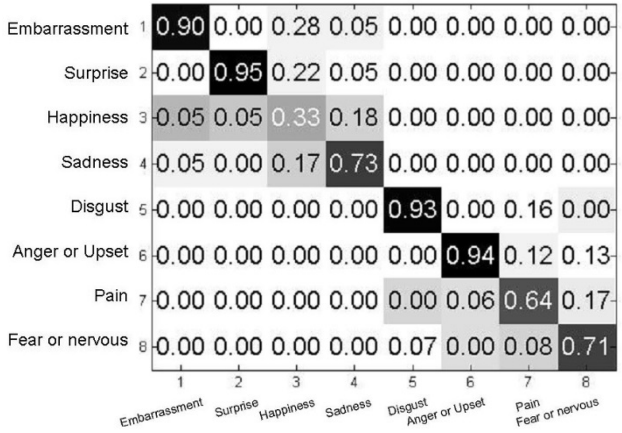
Expression texture

Evaluation of the thermal and motion descriptor



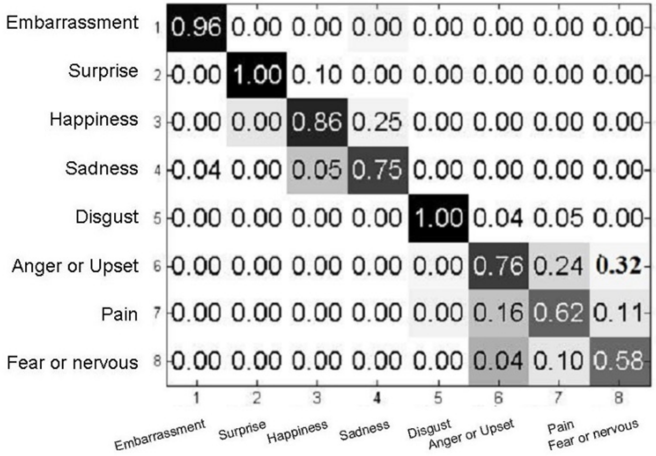
	1 Embarrassment	2 Surprise	3 Happiness	4 Sadness	5 Disgust	6 Anger or Upset	7 Pain	8 Fear or nervous
Embarrassment 1	0.95	0.00	0.04	0.00	0.00	0.00	0.00	0.00
Surprise 2	0.00	1.00	0.12	0.00	0.00	0.00	0.00	0.00
Happiness 3	0.05	0.00	0.85	0.10	0.00	0.00	0.09	0.00
Sadness 4	0.00	0.00	0.00	0.90	0.00	0.00	0.00	0.00
Disgust 5	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
Anger or Upset 6	0.00	0.00	0.00	0.00	0.00	0.90	0.00	0.04
Pain 7	0.00	0.00	0.00	0.00	0.00	0.05	0.87	0.15
Fear or nervous 8	0.00	0.00	0.00	0.00	0.00	0.05	0.04	0.81

(a)The confusion matrix of utilizing both thermal and motion video words.



	1 Embarrassment	2 Surprise	3 Happiness	4 Sadness	5 Disgust	6 Anger or Upset	7 Pain	8 Fear or nervous
Embarrassment 1	0.90	0.00	0.28	0.05	0.00	0.00	0.00	0.00
Surprise 2	0.00	0.95	0.22	0.05	0.00	0.00	0.00	0.00
Happiness 3	0.05	0.05	0.33	0.18	0.00	0.00	0.00	0.00
Sadness 4	0.05	0.00	0.17	0.73	0.00	0.00	0.00	0.00
Disgust 5	0.00	0.00	0.00	0.00	0.93	0.00	0.16	0.00
Anger or Upset 6	0.00	0.00	0.00	0.00	0.00	0.94	0.12	0.13
Pain 7	0.00	0.00	0.00	0.00	0.00	0.06	0.64	0.17
Fear or nervous 8	0.00	0.00	0.00	0.00	0.07	0.00	0.08	0.71

(b)The confusion matrix of just utilizing thermal video words.



	1 Embarrassment	2 Surprise	3 Happiness	4 Sadness	5 Disgust	6 Anger or Upset	7 Pain	8 Fear or nervous
Embarrassment 1	0.96	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Surprise 2	0.00	1.00	0.10	0.00	0.00	0.00	0.00	0.00
Happiness 3	0.00	0.00	0.86	0.25	0.00	0.00	0.00	0.00
Sadness 4	0.04	0.00	0.05	0.75	0.00	0.00	0.00	0.00
Disgust 5	0.00	0.00	0.00	0.00	1.00	0.04	0.05	0.00
Anger or Upset 6	0.00	0.00	0.00	0.00	0.00	0.76	0.24	0.32
Pain 7	0.00	0.00	0.00	0.00	0.00	0.16	0.62	0.11
Fear or nervous 8	0.00	0.00	0.00	0.00	0.00	0.04	0.10	0.58

(c)The confusion matrix of just utilizing motion video words.

Evaluation of the max pooling method and comparison

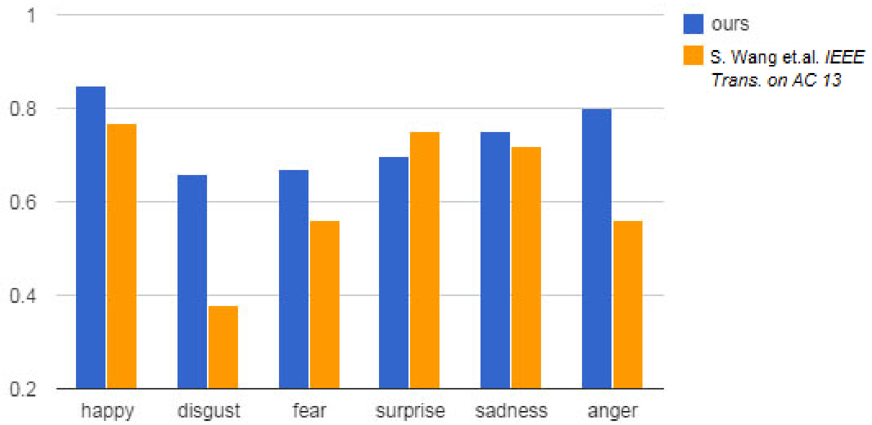
Class	Max pooling	Forehead	Left cheek	Right cheek
Embarrassment	0.96	1	1	1
Surprise	1	0.14	0.33	0.15
Happiness	0.88	0.88	0.84	0.65
Sadness	0.95	0.9	0.78	0.73
Disgust	1	1	1	1
Anger or upset	0.91	0.82	0.82	0.65
Pain	0.78	0.63	0.5	0.42
Fear or nervous	0.79	0.86	0.89	0.9
Weighted average	**0.91**	**0.77**	**0.76**	**0.68**

Comparison to traditional descriptor on thermal video

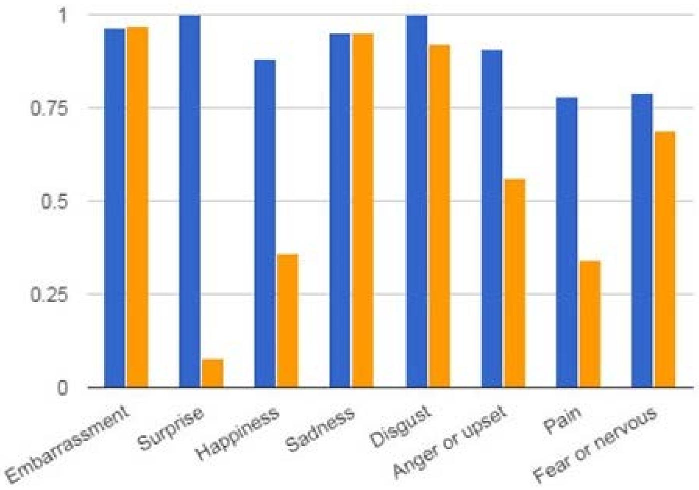
Class	Accuracy	Precision	Recall	F1 score
Ours	**0.91**	**0.88**	**0.88**	**0.88**
HOG[Laptev, CVPR’08]	0.82	0.75	0.82	0.76
HOF[Laptev, CVPR’08]	0.80	0.83	0.78	0.76
HOG+HOF[Laptev, CVPR’08]	0.88	0.88	0.85	0.83
Cuboids[Doll´ar, PETS’05]	0.32	0.31	0.29	0.30

Comparison to prior method

Comparison on USTC-NVIE database



Comparison on our new database



Comparison on USTC-NVIE database with two modalities

Approach	Accuracy	Precision	Recall	F1 score
Ours on the thermal	**0.72**	**0.70**	**0.71**	**0.68**
Ours on the texture	0.63	0.61	0.60	0.59
Method in [Wang et. al. *IEEE Trans. on AC '13*] on the thermal	0.64	0.61	0.60	0.59

Comparison on our new database with two modalities

Approach	Accuracy	Precision	Recall	F1 score
Ours on the thermal	**0.91**	**0.88**	**0.88**	**0.88**
Ours on the texture	0.73	0.72	0.72	0.71
GW-based on the texture [Zhang et. al. *FG'13*]	0.66	0.64	0.62	0.60

Conclusion

- We presented a new infrared thermal video descriptor which can compactly describe a spatio-temporal-temperature information.
- We demonstrated through many experiments that the new descriptor can be a very useful tool to spontaneous facial expression classification.

Future work

- We will utilize the spatial and temporal structural information for improving the classification accuracy.
- We will combine thermal data, texture data and physiology data to further improve the classification performance.

Acknowledgment

- This material is based on the work supported in part by the NSF under grant CNS-1205664 and the SUNY IITG.

- We thank Dr. Jeffrey Cohn and Dr. Qiang Ji and their groups for collaboration on data collection and annotation.

Thanks!