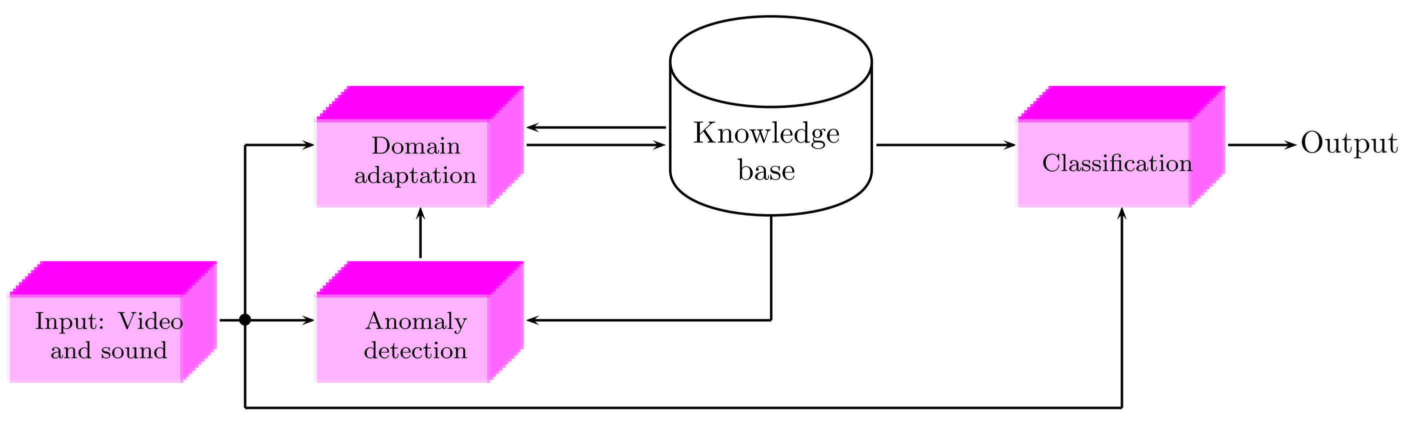
Domain Adaptation in the Context of Sport Video Action Recognition

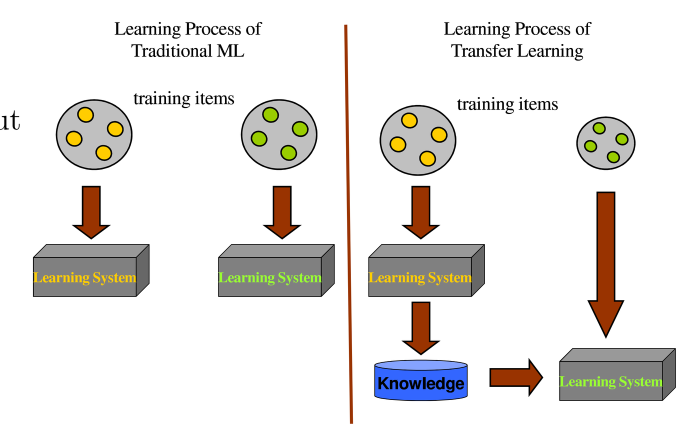
N. FarajiDavar T. deCampos D. Windridge J. Kittler W. Christmas
Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford, UK
{N.Farajidavar, T.Decampos}@surrey.ac.uk http://www.ee.surrey.ac.uk/CVSSP/

Introduction

Our ultimate goal is to investigate adaptive methods for sports video annotation. An anomaly detection system can detect domain change, enabling us to gather samples from the new domain and use methods of **transductive transfer learning** to adapt the models and improve action classification results.



Traditional ML vs. TL

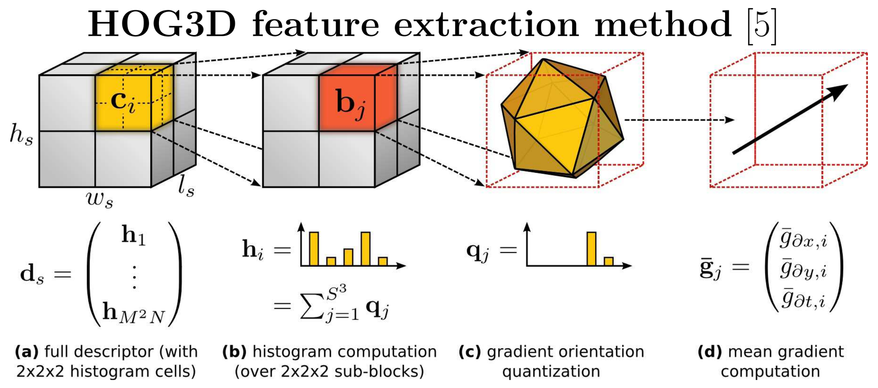


Datasets

Datasets information

label	sport	gender	number	competition	year	non-hit	hit	serve
TWSA03	Tennis	Women	Singles	Australian	2003	944	214	72
TMSA03	Tennis	Men	Singles	Australian	2003	1881	469	123
TWDA09	Tennis	Women	Doubles	Australian	2009	1064	135	36
BMSB08	Badminton	Men	Singles	Beijing	2008	706	458	8

HOG3D feature extraction method [5]



Methodology

Algorithm

- Estimate $P_{\Lambda_{src}}(\mathbf{Y}^{trg}|\mathbf{X}^{trg})$
- Estimate a transformation function $G(\mathbf{X})$ such that $P(\mathbf{Y}, G(\mathbf{X}^{src})) \approx P(\mathbf{Y}, X^{trg})$
- Re-train the classifiers (including kernel computation) using $\{(G(\mathbf{X}^{src}), \mathbf{Y}^{src})\}$

Estimating $G(\mathbf{X})$

We used two different methods to estimate the transformation function $G(\mathbf{X})$:

- Reweighting the features, a modification of [4]

and

- Translating and scaling features [6]

Reweighting features

$$G(x_j^i) = x_j^i \frac{E^{trg}_{\Lambda_{src}}[x_j, y_i]}{E^{src}[x_j, y_i]}, \ \forall i = 1: N^{src}_{train} \text{ , where}$$

$$E^{src}[x_j, y] = \frac{\sum_{i=1}^{N^{train}_{src}} x_j^i \mathbb{1}_{[y]}(y_i)}{\sum_{i=1}^{N^{train}_{src}} \mathbb{1}_{[y]}(y_i)},$$

$$E^{trg}[x_j, y] \approx E^{trg}_{\Lambda_{src}}[x_j, y] = \frac{\sum_{i=1}^{N^{trg}} x_j^i P_{\Lambda_{src}}(y|\mathbf{x}_i)}{\sum_{i=1}^{N^{trg}} P_{\Lambda_{src}}(y|\mathbf{x}_i)},$$

and $\mathbb{1}_{[y]}(y_i)$ is an indicator function

Translating and scaling

$$G(x_j^i) = \frac{x_j^i - E^{src}[x_j, y_i]}{\sigma^{src}_{j,y_i}} \sigma^{trg}_{j,y_i} + E^{trg}_{\Lambda_{src}}[x_j, y_i] \text{ , } \forall i = 1: N^{src}_{train} \text{ ,}$$

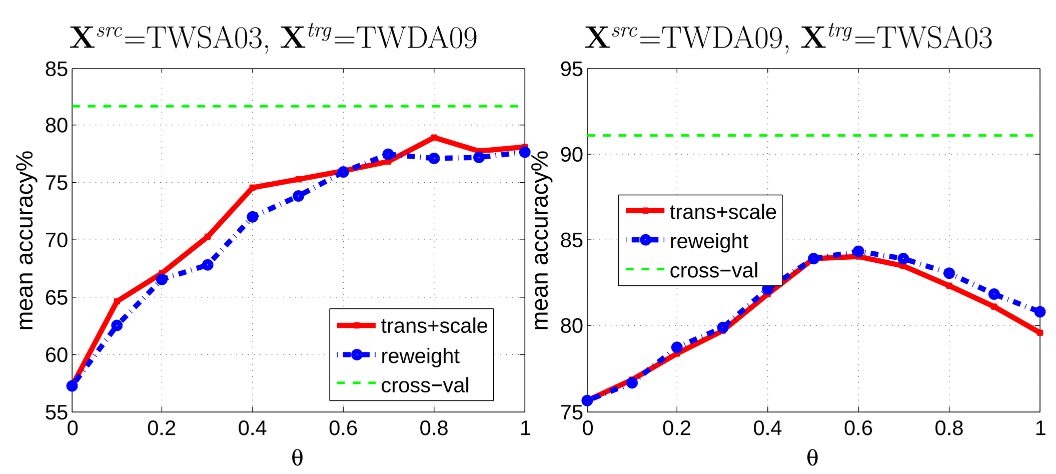
where σ^{src}_{j,y_i} is the standard deviation of feature x_j of the source samples labeled as y_i and

$$\sigma^{trg}_{j,y_i} = \sqrt{\frac{\sum_{k=1}^{N^{trg}} (x_j^k - E^{trg}_{\Lambda_{src}}[x_j, y_i])^2 P_{\Lambda_{src}}(y_i|\mathbf{x}_k)}{\sum_{k=1}^{N^{trg}} P_{\Lambda_{src}}(y_i|\mathbf{x}_k)}}.$$

A **smoothing factor** θ is used to control transfer rate:

$$G'(x_j^i) = (1 - \theta)x_j^i + \theta G(x_j^i)$$

Results



Mean accuracy vs transfer rate θ

Mean accuracy (in ‰) obtained with the baseline (no transfer) and with the two transfer learning methods (*reweight*|*trans+scale*) for $\theta = 0.5$

	source	target: transfer	target: test	accuracy per class (‰): serve	hit	non-hit	macro average
a	TWSA03	–	TWDA09	571	149	996	572
b	TWSA03	test set	TWDA09	857|886	418|433	939|939	738|752
c	TWDA09	–	TWSA03	986	305	978	756
d	TWDA09	test set	TWSA03	972|972	676|634	870|912	839|839
e	TWSA03	–	TMSA03	549	248	981	592
f	TWSA03	test set	TMSA03	852|902	427|442	975|973	751|772
g	BMSB08	–	TMSA03	0	779	359	**379**
h	BMSB08	test set	TMSA03	0|0	886|852	327|393	404|415
i	BMSB08+TWSA03	–	TMSA03	500	357	940	599
j	BMSB08+TWSA03	test set	TMSA03	852|942	427|547	975|917	767|802
k	BMSB08	TWSA03	TMSA03	0|0	983|908	245|330	394|**413**

Results (accuracy in ‰) obtained by swapping TMSA03 and TWSA03, i.e., using the men's game for validation or adaptation and the women's game for test.

	source	target: adaptation	target: test	accuracy per class (‰): non-hit	hit	serve	macro average
a	TMSA03	–	TWSA03	971	427	931	776
b	TMSA03	test set	TWSA03	931|917	610|671	972|986	838|858
c	BMSB08	–	TWSA03	391	883	0	**425**
d	BMSB08	test set	TWSA03	362|440	930|873	0|0	431|438
e	BMSB08+TMSA03	–	TWSA03	966	488	887	781
f	BMSB08+TMSA03	test set	TWSA03	921|851	709|803	958|972	862|875
g	BMSB08	TMSA03	TWSA03	300|369	945|939	0|0	416|**436**

- A significant performance boost was obtained with TL
- Both TL methods showed performance improvement
- Results with the original method of [4] presented negative transfer, we have rectified it as the *reweight* method.

Conclusion and future work

- We have presented an evaluation of domain adaptation techniques for action classification in court sports.
- We complemented the experiments presented in [6] using more video sequences and introduced experiments with different sports (Badminton).
- By applying domain adaptation, we obtained an improvement in classification results, even if the video used to compute adaptation parameters was not the same as that in the test set.
- For future work, we plan to evaluate these methods on other popular domain adaptation datasets and to examine techniques based on classifier model adaptation.

Acknowledgments

This project is sponsored by EPSRC/UK through grant EP/F069421/1 (ACASVA - Adaptive Cognition for Automated Sports Video Annotation). We are also grateful for the support of the EU PASCAL2 network of excellence.
We thank other members of the ACASVA project at Surrey for the useful discussions: A. Khan and F. Yan.

References

[1] W. Dai, Y. Chen, G. rong Xue, Q. Yang, and Y. Yu. Translated learning: Transfer learning across different feature spaces. In *NIPS*, pages 353–360, 2008.
[2] T. de Campos, M. Barnard, K. Mikolajczyk, J. Kittler, F. Yan, W. Christmas, and D. Windridge. An evaluation of bags-of-words and spatio-temporal shapes for action recognition. In *IEEE Workshop on Applications of Computer Vision (WACV)*, Kona, Hawaii, January 2011.
[3] D. Cai, X. He, and J. Han. Efficient kernel discriminant analysis via spectral regression. In *International Conference on Data Mining*, 2007.
[4] A. Arnold, R. Nallapati, and W. W. Cohen. A comparative study of methods for transductive transfer learning. In *Proceedings of the Seventh IEEE International Conference on Data Mining Workshops*, ICDMW '07, pages 77–82, Washington, DC, USA, 2007.
[5] A. Kläser, M. Marszałek, and C. Schmid. A spatio-temporal descriptor based on 3D-gradients. In *British Machine Vision Conference*, pages 995–1004, Sep 2008.
[6] N. FarajiDavar and T. deCampos and J. Kittler and F. Yan. Transductive Transfer Learning for Action Recognition in Tennis Games. 3rd International Workshop on Video Event Categorization, Tagging and Retrieval for Real-World Applications (VECTaR), in conjunction with ICCV, 2011.