Transductive Transfer Learning for Action Recognition in Tennis Games

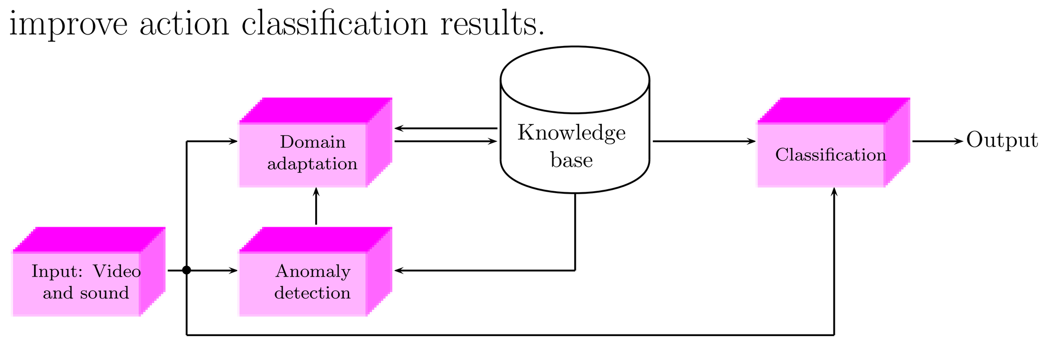
Nazli FarajiDavar, Teo de Campos, Josef Kittler, Fei Yan

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

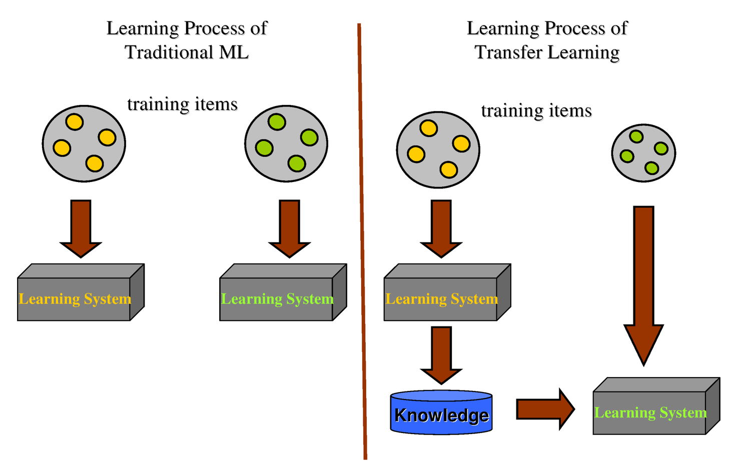
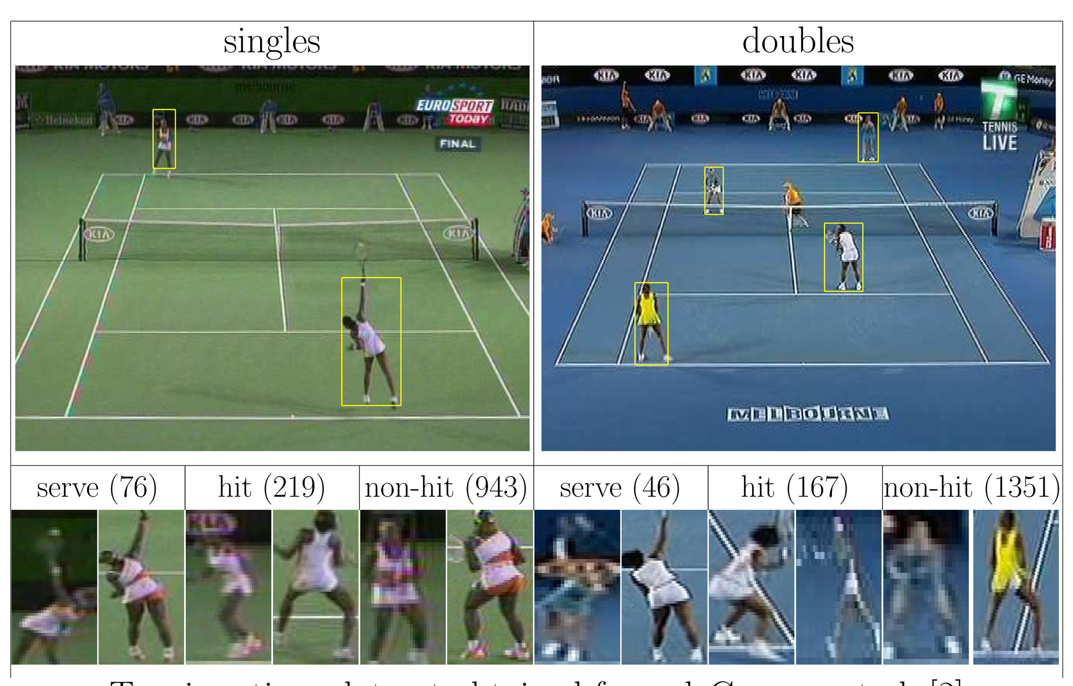


Figure obtained from Dai et al., ©authors of [1]

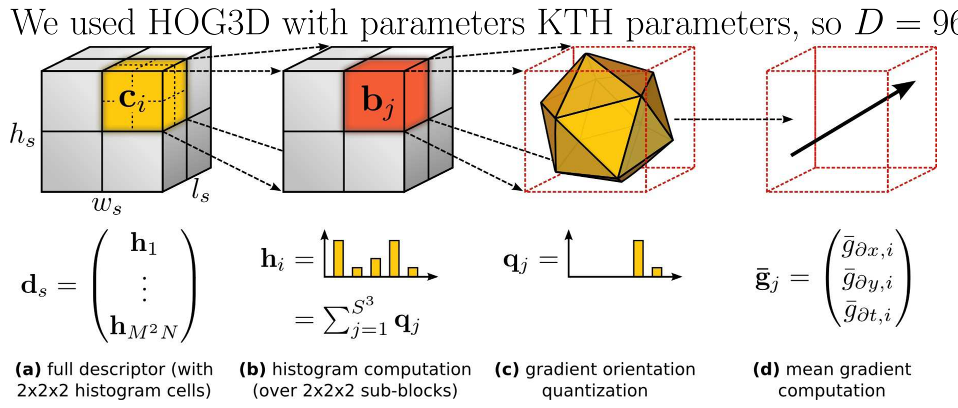
Datasets



Tennis actions dataset obtained from deCampos et al. [2]

Feature extraction

We used HOG3D with parameters KTH parameters, so $D = 960$



Kläser et al.'s HOG3D diagram [5]

Methodology

Notation

Domain: $\mathcal{D} = \{\mathcal{X}, P(\mathbf{x})\}$, where

- $\mathcal{X}$ is the feature space (e.g. $\mathbb{R}^D$)
- $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^N$ is the set of samples, where $\mathbf{x}_i = (x_1^i, \cdots, x_D^i)$
- $P(\mathbf{x})$ is its marginal distribution

Task: $\mathcal{T} = P(\mathbf{Y}|\mathbf{X})$, where

- $\mathbf{Y} = \{y_i\}_{i=1}^N \in \mathcal{Y}$ is the set of labels and
- $P(\mathbf{Y}|\mathbf{X})$ is a prediction function.

$\mathcal{D}$ and $\mathcal{T}$ are defined for source (src) and target (trg) domains.

Problem definition

In transductive transfer learning (and domain adaptation),

- $\mathcal{X}^{src} = \mathcal{X}^{trg}$ and $\mathcal{Y}^{src} = \mathcal{Y}^{trg}$
- but $P(\mathbf{x}^{src}) \neq P(\mathbf{x}^{trg})$

Outline of the Method

Input:

- Data set: source $\{(\mathbf{X}^{src}, \mathbf{Y}^{src})\}$ and target $\{\mathbf{X}^{trg}\}$
- Classification function with the model Λ_{src} trained using $\mathbf{X}^{src}$. We used the KLDA classifier [3] with ℓ_1-norm for the RBF kernels.

Algorithm:

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

Estimating $G(\mathbf{X})$

Reweighting features (a modification of [4])

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

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

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

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

Translating and scaling

$$G(x_j^i) = \frac{x_j^i - E^{src}[x_j, y_i]}{\sigma_{j,y_i}^{src}} \sigma_{j,y_i}^{trg} + E_{\Lambda_{src}}^{trg}[x_j, y_i] \text{ ,} \forall i = 1\colon N_{train}^{src} \text{ ,}$$

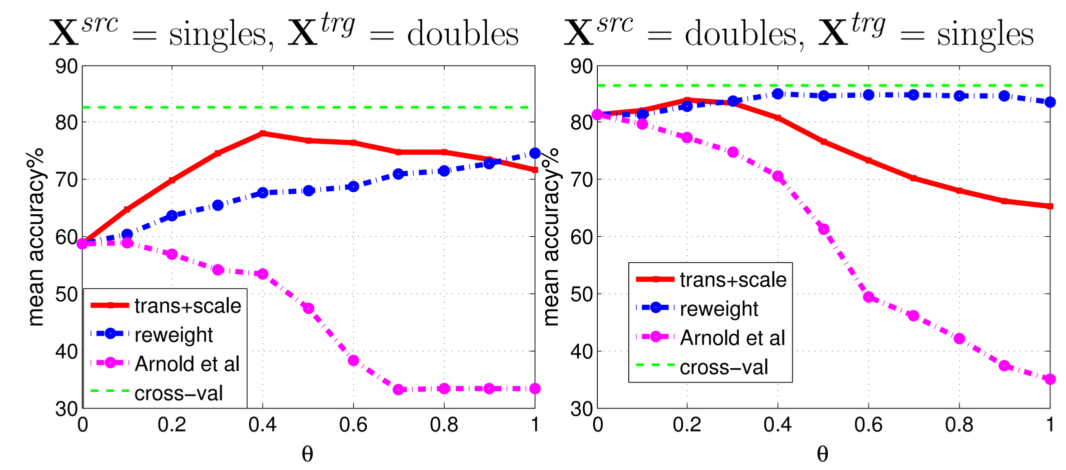
where σ_{j,y_i}^{src} is the standard deviation of feature x_j of the source samples labeled as y_i and

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

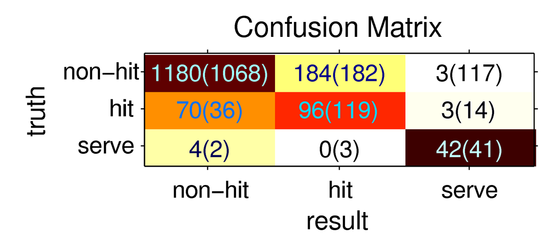
A **smoothing factor** θ is used to control transfer rate:

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

Results and discussion



Mean accuracy vs transfer rate θ



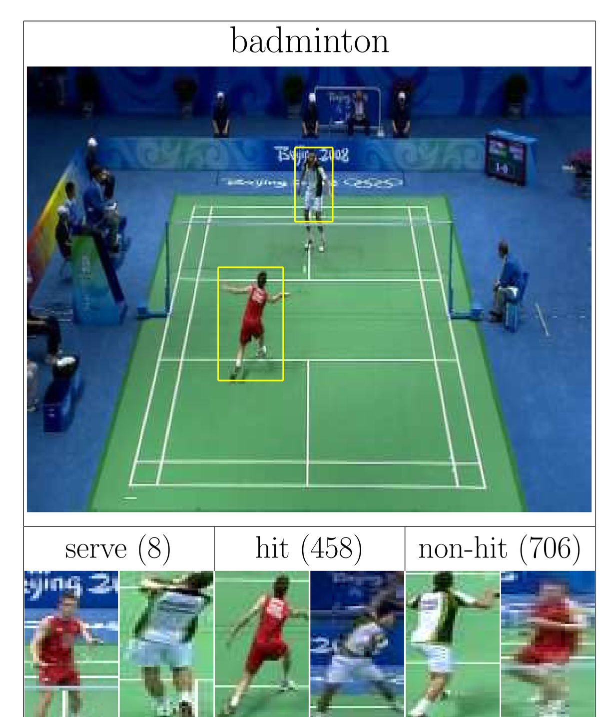
Confusion matrix obtained with *trans+scale* and $\theta = 0.4$ (in brackets: results of [2])

- A significant performance boost was obtained with TL
- The proposed *trans+scale* method approaches the upper bound in performance, set by cross-validation on the test data set.
- Both methods show performance improvement with conservative values of θ
- Results with the original method of [4] present negative transfer, we have rectified it as the *reweight* method.

New experiments

In [6], we describe experiments with the following new datasets, which include data from a badminton match:

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



Results

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**

Conclusion and future work

- We investigated a novel application of transductive transfer learning for sport video annotation.
- We introduced a method based on translating and scaling the feature space and evaluated it in comparison to a modification of the reweighting method of Arnold et al. [4].
- We presented experiments on an action classification dataset and showed that, in one scenario, the proposed method can lead to an increase of nearly 20% in mean accuracy.
- Our new results on different sports [6] show that these methods can lead to performance improvement even if the target set is not the same 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: W. Christmas, D. Windridge and A. Khan.

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, T. deCampos, D. Windridge, J. Kittler, W. Christmas. Domain Adaptation in the Context of Sport Video Action Recognition. In *Domain Adaptation workshop, in conj. NIPS*, Dec 2011.