IT'S ALL ABOUT HABITS: EXPLOITING MULTI-TASK CLUSTERING FOR ACTIVITIES OF DAILY LIVING ANALYSIS

Yan Yan[1] *Elisa Ricci*[2,3] *Negar Rostamzadeh*[1] *Nicu Sebe*[1]

[1] Department of Information Engineering and Computer Science, University of Trento, Italy
[2] Fondazione Bruno Kessler, Trento, Italy
[3] Department of Engineering, University of Perugia, Italy

ABSTRACT

Motivated by applications in areas such as patient monitoring, tele-rehabilitation and ambient assisted living, analyzing activities of daily living is an active research topic in computer vision and image processing. In this paper we address the problem of everyday activity recognition from unlabeled data proposing a novel multi-task clustering (MTC) approach. Our intuition is that, when analyzing activities of daily living, we can take advantage of the fact that people tend to perform the same actions in the same environment (*e.g.* people working in an office environment use to read and write documents). Thus, even if labels are not available, information about typical activities can be exploited in the learning process. Arguing that the tasks of recognizing activities of specific individuals are related, we resort on multi-task learning and rather than clustering the data of each individual separately, we also look for clustering results which are coherent among related tasks. Extensive experimental results show that our method outperforms several state-of-the-art approaches by up to 11% on the Rochester activities of daily living dataset.

Index Terms— Multi-Task Clustering, Activities of Daily Living Analysis

1. INTRODUCTION

Activities of daily living (ADL) are defined as "the things we normally do on a daily basis for self-care such as feeding ourselves, bathing, dressing, grooming, work, homemaking, and leisure"[1]. In the last few years, automatic analysis of ADL has received an increasing interest in the computer vision and image processing community [1, 2]. The problem of everyday activity recognition poses several challenges, mostly implying the engineering of discriminative features and scalable recognition algorithms. A further problem arises as, independently of the considered scenario, several hours of videos are usually collected. This generates a large amount of data for which annotation is typically not available requiring a great human labeling effort.

[1] http://en.wikipedia.org/wiki/Activities_of_daily_living#cite_note-MN-2

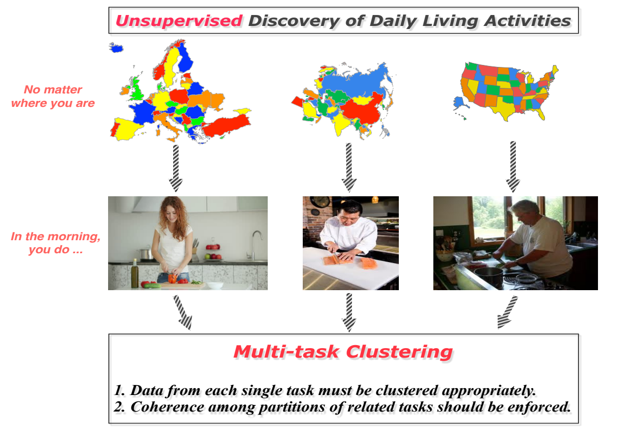

Fig. 1. Overview of the considered problem: *no matter where you are, in the morning you probably have breakfast and use a knife to cut food into pieces*. In this paper we exploit this and other information about people habits to perform ADL analysis proposing a novel multi-task clustering approach.

In this paper, we consider the problem of everyday activity recognition from unlabeled data under a novel perspective (see Fig.1). People working in an office environment tend to perform the same kind of activities (*e.g.* working in front of a personal computer, reading documents). Similarly, most people when they wake up in the morning use to drink coffee and brush their teeth. In other words, the tasks of recognizing everyday activities of different individuals are related. In this paper we treat ADL analysis as a multi-task clustering problem, and we show that exploiting the relations between data associated to different people more accurate recognition models can be obtained with respect to considering each user separately or to naively combining data from different people together. In particular if there are limited data for a single person, traditional clustering methods applied to this data fail to discover the correct clusters and using data from other people as an auxiliary source of information is beneficial, inducing the correct space partitioning. On the other hand, simply combining data from different people and clustering them also does not lead to accurate results as the data distributions associated to some targets can be very different. Oppositely,

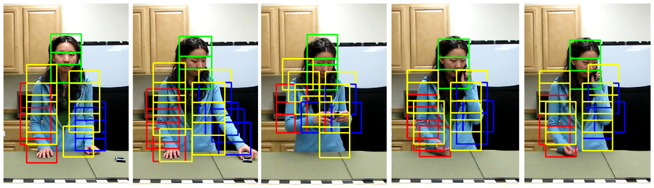

Fig. 2. Rochester ADL dataset: a sequence depicting the activity 'answering phone' and the computed body parts.

with the proposed MTC solutions the degree of similarity between different tasks (data from different people) is taken into account during learning.

To summarize, the contributions of this paper are the following. We address the problem of everyday activity recognition proposing a MTC approach. A novel MTC optimization problem is introduced and an efficient algorithm is proposed for solving it. We demonstrate the effectiveness of our approach on the Rochester ADL dataset, comparing it with several single task and MTL methods.

2. RELATED WORKS

ADL Analysis. In the last few years, several works have considered the problem of everyday activity recognition, not only in computer vision and image processing but also in other related research areas, *e.g.* ubiquitous computing [3, 4]. Many of these recent works are based on the use of RFID tags or inertial sensors. However, systems based on cameras still have an important role being generally cheap and easy to deploy. A survey of recent works on activity recognition is presented in [5]. Some recent publications have addressed specifically the task of ADL analysis [1, 3]. Messing *et al.* [1] proposed an approach based on features computed from the velocity history of tracked keypoints for recognizing complex everyday activities performed in a kitchen environment. In [2], Rohrbach *et al.* also considered a kitchen scenario but focused on a more difficult problem of fine-grained activity recognition. Other works (see *e.g.* [3]) exploited the use of the novel RGB-D sensors showing improved performance with respect to the use of traditional cameras alone. In this paper we address the problem of analysing activities of daily living under the perspective of MTC. However, while multi-task learning have already been exploited for visual based activity recognition [6, 7, 8], we are not aware of works which address simultaneously the problem of lack of annotated data.

Multi-task Learning. Multi-task learning (MTL) approaches have received considerable attention in the last few years. Learning from data of multiple related tasks simultaneously is greatly advantageous in terms of performance with respect to learning on every single task independently. The effectiveness of MTL has been demonstrated in several applications in computer vision and image processing [9, 10, 11, 12]. Most existing works on MTL methods tackle classification and regression problems. Only few works have considered unsupervised approaches to MTL [13, 14], *i.e.* the scenario where the data of each task are unlabeled and the aim is to predict the cluster labels in each task. In [13] the authors proposed to learn a subspace shared by all the tasks, through which the knowledge of one task can be transferred to all the others. Zhang and Zhang [14] introduced a MTC approach based on a pairwise agreement term which encourage coherence among clustering results of multiple tasks. Our approach is inspired by [14], but it is based on another objective function and thus on a radically different optimization algorithm. Furthermore, in the considered application, it provides superior accuracy with respect to [14].

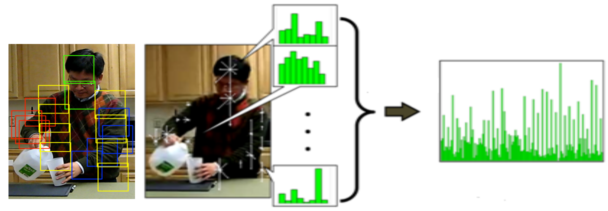

Fig. 3. Rochester ADL dataset: feature representation for a single frame.

3. MULTI-TASK CLUSTERING FOR ADL ANALYSIS

The proposed approach articulates in two main phases: first features are extracted from video sequences, then our MTC algorithm is used for recognizing each individual's activities.

3.1. Features Extraction

In this paper we consider the Rochester ADL dataset [1]. It consists of a set of pre-segmented video clips, each depicting 10 different activities performed 3 times by 5 different people. Typical activities are answering a phone, drinking water, eating a snack, or peeling a banana. The recorded people have different appearance, genders and ethnicity. Each video clip is on average 3-50s long. The frame size is 1280×720 and the frame rate is 30 frames/s.

We follow one of the most recent works on this dataset [15] and we first extract features on a frame-basis (at rate of one frame/s) considering a combination of both low-level and high-level cues. Specifically to compute high-level cues we adopt the pictorial deformable model for body pose estimation proposed in [16] and detect the location of 18 body-parts. Fig.2 shows an example of body parts extracted on a sequence of the activity 'answering phone'. To extract low-level cues, we compute the optical flow using the Lucas-Kanade algorithm and we quantize it into 8 possible directions. Then we construct a descriptor for each body part, represented by an eight bin histogram computed from the optical flow information. Finally, we concatenate all the histograms and create a 144 bin histogram for each frame (Fig.3). To compute the video clips descriptors we adopt two different approaches as

Algorithm 1 Algorithm for solving (2)

Input: Data matrices $\mathbf{X}_1, \mathbf{X}_2$; numbers of clusters k_1, k_2; λ.

Initialize $\mathbf{F}$ as an identity matrix.

Initialize $\mathbf{W} > 0$ with l_1 normalized columns and $\mathbf{P} > 0$ with l_1 normalized rows.

Repeat until convergence

Compute $\mathbf{F}$ using a linear programming solver.

Compute $\mathbf{W}$ using a projected gradient method: $\mathbf{W}^{k+1} = \max(0, \mathbf{W}^k - \alpha_k \nabla_{\mathbf{W}} \Delta(\mathbf{P}^k, \mathbf{W}^k, \mathbf{F}^k))$.

Compute $\mathbf{P}$ using a projected gradient method: $\mathbf{P}^{k+1} = \max(0, \mathbf{P}^k - \alpha_k \nabla_{\mathbf{P}} \Delta(\mathbf{P}^k, \mathbf{W}^k, \mathbf{F}^k))$.

Normalize $\mathbf{P}$ by $\mathbf{P}_{ij} \leftarrow \frac{\mathbf{P}_{ij}}{\sum_j \mathbf{P}_{ij}}$.

Output: The optimized matrices $\mathbf{W}$, $\mathbf{P}$.

suggested in [15]: one consisting in accumulating frame features, the other in using a Fisher-Kernel representation.

3.2. EMD Regularized Multi-task Clustering

As the tasks of recognizing activities of each person are related, we propose a MTL algorithm. Formally, we are given T related data sources (corresponding to different individuals), each one consisting of data samples in the set $X_t = \{\mathbf{x}_1^t, \mathbf{x}_2^t, ..., \mathbf{x}_{N_t}^t\}$, where $\mathbf{x}_j^t \in I\!R^d$ is a d-dimensional feature vector extracted from a video clip, N_t is the number of samples associated to the t-th data task (person). We want each data source to be partitioned into k_t clusters, where the number of required partitions can be different in different tasks. As we assume the tasks to be related, we also require that the resulting partitions are consistent with each other. Defining $N = \sum_{i=1}^T N_i$, $k = \sum_{i=1}^T k_i$ we consider the data matrix $\mathbf{X} \in I\!R^{N \times d}$, $\mathbf{X} = [\mathbf{X}_1; \ldots; \mathbf{X}_T]$, obtained by concatenating the individual matrices $\mathbf{X}_t = [\mathbf{x}_1^t;\ \mathbf{x}_2^t;\ \ldots;\mathbf{x}_{N_t}^t]$ associated to each task t. We are interested in finding the centroid matrix $\mathbf{C} = [\mathbf{C}_1; \ldots; \mathbf{C}_T]$, $\mathbf{C} \in I\!R^{k \times d}$, $\mathbf{C}_t \in I\!R^{k_t \times d}$, and the cluster indicators matrix $\mathbf{W} = \text{blkdiag}(\mathbf{W}_1, ..., \mathbf{W}_T)$, $\mathbf{W} \in I\!R^{N \times k}$, $\mathbf{W}_t \in I\!R^{N_t \times k_t}$, by solving the following optimization problem:

$$
\begin{aligned}
&\min_{\substack{\mathbf{C}_1,\ldots,\mathbf{C}_T,\\ \mathbf{W}_1,\ldots,\mathbf{W}_T, f_{ij} \geq 0}} \sum_{t=1}^{T} \|\mathbf{X}_t - \mathbf{W}_t \mathbf{C}_t\|_F^2 \\
&+\lambda \sum_{t,s=1}^{T} \sum_{i=1}^{k_t} \sum_{j=1}^{k_s} f_{ij} [(\mathbf{C}_t)_{i.} - (\mathbf{C}_s)_{j.}]' [(\mathbf{C}_t)_{i.} - (\mathbf{C}_s)_{j.}] \\
&s.t. \begin{cases} \sum_{j=1}^{k_s} f_{ij} = \sum_{n=1}^{N_t} (\mathbf{W}_t)_{ni} \quad (1 \leq i \leq k_t) \\ \sum_{i=1}^{k_t} f_{ij} = \sum_{n=1}^{N_s} (\mathbf{W}_s)_{nj} \quad (1 \leq j \leq k_s) \\ \sum_{i=1}^{k_t} \sum_{j=1}^{k_s} f_{ij} = 1 \quad (1 \leq i \leq k_t, 1 \leq j \leq k_s) \end{cases}
\end{aligned} \tag{1}
$$

where $(\cdot)'$ denotes the transpose operator, $(\mathbf{C}_t)_{i\cdot}$ and $(\mathbf{C}_s)_{j\cdot}$ denote the i-th row of $\mathbf{C}_t$ and j-th row of $\mathbf{C}_s$ respectively. The first term in the objective function is a relaxation of the traditional k-means objective function for T separated data sources. The second term is added to explore the relationships between clusters of two different data sources and it consists of the popular Earth Mover's Distance (EMD) [17] computed considering the signatures $\mathcal{S}$ and $\mathcal{T}$ obtained by clustering the data associated to task t and s separately, *i.e.*, $\mathcal{T} = \{((\mathbf{C}_t)_{1\cdot}, w_t^1), \ldots ((\mathbf{C}_t)_{k_t\cdot}, w_t^{k_t})\}$, $w_t^i = \sum_{n=1}^{N_t} (\mathbf{W}_t)_{ni}$, and $\mathcal{S} = \{((\mathbf{C}_s)_{1\cdot}, w_s^1), \ldots ((\mathbf{C}_s)_{k_s\cdot}, w_s^{k_s})\}$, $w_s^i = \sum_{n=1}^{N_s} (\mathbf{W}_s)_{ni}$. In practice $(\mathbf{C}_t)_{i\cdot}$ and $(\mathbf{C}_s)_{j\cdot}$ are the cluster centroids and w_i^s, w_i^t denote the weights associated to each cluster (reflecting somehow the number of datapoints of each cluster). In practice the second term represents a distance between two distributions and minimizing it we impose the found partitions between two related tasks to be consistent. The variables f_{ij} are a set of EMD flows.

In (1) there are no constraints on the $\mathbf{C}$ values. In this paper we also impose that the vectors defining $\mathbf{C}$ lie within the column space of $\mathbf{X}$, *i.e.* the columns of $\mathbf{C}$ are a weighted sum of certain data points. In other words, we define $\mathbf{C} = \mathbf{PX}$ where $\mathbf{P} = \text{blkdiag}(\mathbf{P}_1 \ldots \mathbf{P}_T)$, $\mathbf{P} \in I\!R^{k \times N}$. In the following, for the sake of simplicity and easy interpretation, we consider only a two tasks problem. The extension to T tasks is straightforward. As typical in many learning algorithm, we also introduce the mapping $\mathbf{X} \to \phi(\mathbf{X})$ and the associated kernel matrix $\mathbf{K_X} = \phi(\mathbf{X})\phi(\mathbf{X})'$. The objective function of (1) becomes:

$$
\begin{aligned}
\Delta = \quad & \|\phi(\mathbf{X}) - \mathbf{WP}\,\phi(\mathbf{X})\|_F^2 + \lambda \text{tr}(\mathbf{MP}\phi(\mathbf{X})\phi'(\mathbf{X})\mathbf{P}'\mathbf{M}'\mathbf{F}) \\
= \text{tr}(\ & \mathbf{K_X} - \mathbf{2K_X P'W'} + \mathbf{WPK_X P'W'} + \lambda \mathbf{MPK_X P'M'F}\ )
\end{aligned}
$$

Defining $\mathbf{A} = \mathbf{WPK_X} - \mathbf{2K_X}$, the proposed optimization problem is:

$$
\begin{aligned}
\min_{\mathbf{P}>\mathbf{0},\mathbf{W}>\mathbf{0},\mathbf{F}>\mathbf{0}} \quad & \text{tr}(\ \mathbf{K_X} + \mathbf{AP'W'} + \lambda \mathbf{MPK_X P'M'F}\ ) \\
\text{s.t.} \quad & \|(\mathbf{P}_t)_{i.}\|_1 = 1, \ \ \forall i \quad \forall\, t = 1, 2 \\
& \text{tr}(\mathbf{I}_j \mathbf{F}) = \sum_{i=1}^{N_1+N_2} \mathbf{W}_{ij}, \ \ j = 1, ..., k_1 + k_2 \\
& \text{tr}(\mathbf{F}) = 1
\end{aligned} \tag{2}
$$

with $\mathbf{F} = \text{diag}(f_{11} \ldots f_{k_1 k_2})$, $\mathbf{F} \in I\!R^{k_1 k_2 \times k_1 k_2}$, and

$$
\mathbf{I}_j = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ & & & \\ 0 & 0 & \ddots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}, \mathbf{M} = \begin{bmatrix} \underbrace{\begin{matrix} 1 & 0 & \cdots \\ 1 & 0 & \cdots \\ 1 & 0 & \cdots \\ 0 & 1 & \cdots \\ \vdots & \vdots & \vdots \\ 0 & 0 & \cdots \end{matrix}}_{1:k_1} & \underbrace{\begin{matrix} -1 & 0 & \cdots \\ 0 & -1 & \cdots \\ 0 & \cdots & -1 \\ -1 & 0 & \cdots \\ \vdots & \vdots & \vdots \\ 0 & \cdots & -1 \end{matrix}}_{k_1+1:k_1+k_2} \end{bmatrix},
$$

$\mathbf{I}_j \in I\!R^{k_1 k_2 \times k_1 k_2}$ and $\mathbf{M} \in I\!R^{k_1 k_2 \times (k_1 + k_2)}$, are appropriately defined selection matrices.

To solve (2), we first note that the optimal solution can be found adopting an alternating optimization scheme, *i.e.* optimizing separately (2) first with respect to $\mathbf{P}$ and then with respect to $\mathbf{W}$ and $\mathbf{F}$ jointly. In both cases, a non-negative least square problem with constraints arises, for which standard solvers can be employed. However, due to computational efficiency, in this paper we consider an approximation of (2), replacing the constraints (3) with $\text{tr}(\mathbf{I}_j \mathbf{F}) = \mathbf{e}$, where

Table 1. Clustering results on Rochester ADL dataset: comparison of different methods using *accumulation* features.

	Acc			NMI		
	Task 1	Task 2	Avg	Task 1	Task 2	Avg
KM	0.523	0.513	0.518	0.671	0.646	0.659
KKM	0.545	0.537	0.541	0.689	0.672	0.681
SemiNMF [18]	0.556	0.526	0.541	0.604	0.637	0.621
SemiEMD-MTC [14]	0.580	0.533	0.557	0.658	0.655	0.657
KSemiEMD-MTC [14]	0.602	0.561	0.581	0.686	0.689	0.688
LSMTC [13]	0.480	0.503	0.492	0.598	0.621	0.610
CNMF [18]	0.607	0.647	0.627	0.746	0.772	0.759
CEMD-MTC	0.693	0.627	0.660	0.827	0.842	0.835
KCEMD-MTC	**0.700**	**0.770**	**0.735**	**0.853**	**0.883**	**0.868**

Table 2. Clustering results on Rochester ADL dataset: comparison of different methods using *fisher kernel* features.

	Acc			NMI		
	Task 1	Task 2	Avg	Task 1	Task 2	Avg
KM	0.533	0.537	0.535	0.682	0.656	0.669
KKM	0.555	0.552	0.554	0.704	0.694	0.699
SemiNMF [18]	0.581	0.531	0.556	0.634	0.639	0.637
SemiEMD-MTC [14]	0.595	0.567	0.581	0.678	0.675	0.677
KSemiEMD-MTC [14]	0.621	0.584	0.603	0.699	0.702	0.701
LSMTC [13]	0.501	0.525	0.513	0.602	0.634	0.618
CNMF [18]	0.621	0.644	0.633	0.755	0.782	0.769
CEMD-MTC	0.713	0.653	0.683	0.833	0.852	0.843
KCEMD-MTC	**0.741**	**0.765**	**0.753**	**0.874**	**0.888**	**0.881**

$\mathbf{e} \in I\!R^{k_1 k_2}$, $(\mathbf{e})_i = \frac{1}{k_1}$, if $i \leq k_1$, $(\mathbf{e})_i = \frac{1}{k_2}$ otherwise. This approximation implies that for each task the same number of datapoints is assigned to all the clusters. In this way an efficient solver can be devised. Specifically, we adopt an alternating optimization strategy, *i.e.* we optimize (2) separately with respect to $\mathbf{F}$, $\mathbf{W}$ and $\mathbf{P}$ until convergence. The algorithm for solving (2) is summarized in Algorithm 1.

4. EXPERIMENTAL RESULTS

We perform a series of experiments randomly selecting two targets out of five from the Rochester ADL dataset. Thus, two tasks are considered. Experiments are run 10 times and the average results are reported. Table 1 and 2 show the results of different clustering methods applied on the the *accumulation* and the *fisher kernel* representations respectively. We compare our approach (EMD Regularized Multi-task Clustering with linear and rbf kernel denoted as CEMD-MTC, KCEMD-MTC respectively) with single task clustering approaches, *e.g.* the k-means (KM), kernel k-means (KKM), convex (CNMF) and semi (SemiNMF) nonnegative matrix factorization [18]. We also consider recent MTC approaches such as the SemiEMD-MTC proposed in [14], its kernel version KSemiEMD-MTC and the LSMTC method in [13]. Ten runs are performed corresponding to different initializations conditions for all the methods. For each experiment the average results are considered. To evaluate the clustering results, we adopt the popular clustering accuracy (Acc) and normalized mutual information (NMI) metrics. The value of the regularization parameters λ of our approach is set in the range $\{10^{-2}, 10^{-1}, ...10^{2}\}$. The reported results correspond to the best clustering performance.

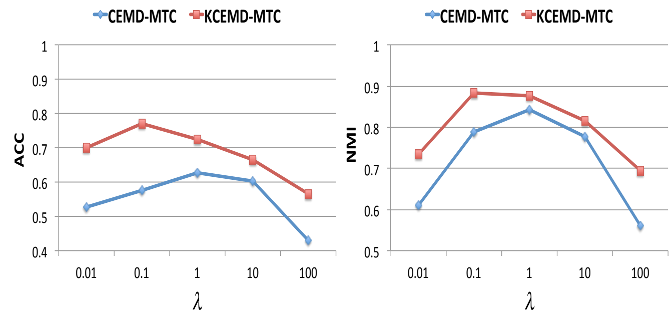

Fig. 4. Performance variation at different value of λ for Task 2 of the Rochester ADL dataset.

From Table 1 and 2 several observations can be made. First of all, independently on the adopted features representation, MTC approaches always perform better than single task clustering methods (*e.g.* SemiEMD-MTC outperforms SemiNMF, CEMD-MTC provide higher accuracy than CNMF). An exception is represented by the LSMTC proposed in [13] which performs quite poorly (worse than k-means) in the considered application. Confirming the findings reported in [15], we also observe that the Fisher Kernel representations is advantageous with respect to features computed based on a simple accumulation scheme. Noticeably, our methods are among the best performers, with KCEMD-MTC reaching the higher values of accuracy and NMI. This is somehow expected probably due to both the use of kernels and the adoption of the multi-task paradigm.

Finally, we investigate the effect of different values of the regularization parameter λ in (2) on clustering performance when Fisher Kernel features are used. As shown in Fig.4, both accuracy and NMI values are sensitive to varying λ. The best performance for CEMD-MTC and KCEMD-MTC are obtained when $\lambda = 1$ and $\lambda = 0.1$ respectively. This clearly confirms the advantage of using a MTC approach.

5. CONCLUSIONS

In this paper we consider the task of everyday activity recognition from unlabeled data as a MTC problem. A novel MTC algorithm has been proposed and evaluated extensively on Rochester ADL dataset. Our results clearly demonstrate the advantage of using a MTC approach (in particular KCEMD-MTC) for ADL analysis. Future works include exploiting the suitability of the proposed MTC algorithms for other vision applications as well as investigating how to improve our MTC methods (*e.g.* by detecting outlier tasks).

6. ACKNOWLEDGEMENTS

This work was partially supported by EU FP7 xLIME and PERTE project.

7. REFERENCES

[1] R. Messing, C. Pal, and H. Kautz, "Activity recognition using the velocity histories of tracked keypoints," in *ICCV*, 2009.

[2] M. Rohrbach, S. Amin, M. Andriluka, and B. Schiele, "A database for fine grained activity detection of cooking activities," in *CVPR*, 2012.

[3] J. Lei, X. Ren, and D. Fox, "Fine-grained kitchen activity recognition using rgb-d," in *UBICOMP*, 2012, pp. 208–211.

[4] E. M. Tapia, S. S. Intille, and K. Larson, "Activity recognition in the home using simple and ubiquitous sensors," in *Pervasive Computing*, 2004, pp. 158–175.

[5] P. Turaga, R. Chellappa, V. S. Subrahmanian, and O. Udrea, "Machine recognition of human activities: A survey," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 18, no. 11, pp. 1473–1488, 2008.

[6] B. Mahasseni and S. Todorovic, "Latent multitask learning for view-invariant action recognition." in *ICCV*, 2013.

[7] Y. Yan, G. Liu, E. Ricci, and N. Sebe., "Multi-task linear discriminant analysis for multi-view action recognition." in *ICIP*, 2013.

[8] C. Yuan, W. Hu, G. Tian, S. Yang, and H. Wang, "Multi-task sparse learning with beta process prior for action recognition." in *CVPR*, 2013.

[9] R. Salakhutdinov, A. Torralba, and J. Tenenbaum, "Learning to share visual appearance for multiclass object detection," in *CVPR*, 2011.

[10] S. J. Hwang, F. Sha, and K. Grauman, "Sharing features between objects and their attributes," in *CVPR*, 2011.

[11] Y. Yan, E. Ricci, R. Subramanian, O. Lanz, and N. Sebe, "No matter where you are: Flexible graph-guided multi-task learning for multi-view head pose classification under target motion," in *ICCV*, 2013.

[12] Y. Yan, E. Ricci, G. Liu, R. Subramanian, and N. Sebe, "Clustered multi-task linear discriminant analysis for view invariant color-depth action recognition," in *ICPR*, 2014.

[13] Q. Gu and J. Zhou, "Learning the shared subspace for multi-task clustering and transductive transfer classification," in *ICDM*, 2009.

[14] J. Zhang and C. Zhang, "Multitask bregman clustering," *Neurocomput.*, vol. 74, no. 10, pp. 1720–1734, May 2011.

[15] N. Rostamzadeh, G. Zen, I. Mironica, J. Uijlings, and N. Sebe, "Daily living activities recognition via efficient high and low level cues combination and fisher kernel representation." in *ICIAP*, 2013.

[16] Y. Yang and D. Ramanan, "Articulated human detection with flexible mixtures-of-parts," *IEEE Trans. on PAMI*, 2013.

[17] Y. Rubner, C. Tomasi, and L. J. Guibas, "A metric for distributions with applications to image databases." in *ICCV*, 1998.

[18] C. Ding, T. Li, and M. I. Jordan, "Convex and semi-nonnegative matrix factorizations," *IEEE Trans. on PAMI*, vol. 32, no. 1, pp. 45–55, 2010.