Transferring Knowledge across Robots: a Risk Sensitive Approach

Gabriele Costante[a], Thomas A. Ciarfuglia[a], Paolo Valigi[a], Elisa Ricci[a,b]

[a]*Department of Engineering, University of Perugia, via Duranti 93, 06125, Perugia, Italy*
[b]*Fondazione Bruno Kessler, via Sommarive 18, 38123, Povo, Trento, Italy*

Abstract

One of the most impressive characteristics of human perception is its domain adaptation capability. Humans can recognize objects and places simply by transferring knowledge from their past experience. Inspired by that, current research in robotics is addressing a great challenge: building robots able to sense and interpret the surrounding world by reusing information previously collected, gathered by other robots or obtained from the web. But, *how can a robot automatically understand what is useful among a large amount of information and perform knowledge transfer?* In this paper we address the domain adaptation problem in the context of visual place recognition. We consider the scenario where a robot equipped with a monocular camera explores a new environment. In this situation traditional approaches based on supervised learning perform poorly, as no annotated data are provided in the new environment and the models learned from data collected in other places are inappropriate due to the large variability of visual information. To overcome these problems we introduce a novel transfer learning approach. With our algorithm the robot is given only some training data (annotated images collected in different environments by other robots) and is able to decide whether, and how much, this knowledge is useful in the current scenario. At the base of our approach there is a transfer risk measure which quantifies the similarity between the given and the new visual data. To improve the performance, we also extend our framework to take into account multiple visual cues. Our experiments on three publicly available datasets demonstrate the effectiveness of the proposed approach.

Keywords:
Autonomous robot navigation, Visual place recognition, Domain adaptation, Unsupervised learning, Multiple cues

1. Introduction

In recent years robotics research has focused on the integration of visual information to improve autonomous systems performance in many tasks, such as robot localization, mapping or manipulation. Many vision and learning techniques have been exploited to build robotic systems able to face challenging and unknown scenarios. Furthermore, recent research activities in the computer vision and robotics fields have also led to the creation and diffusion of a vast number of image and video collections publicly available on the web. A robot can now access a lot of information and potentially can take advantage of these data to improve its performances in many important tasks such as navigation or manipulation. The challenge here is that visual information obtained from the web has typically been collected in very different situations and scenarios with respect to those the robot is operating in. Thus, it is clear that knowledge transfer approaches are fundamental in this context. In general, building autonomous systems that can easily adapt their internal models when environmental conditions change and can exploit the information collected by other robots is very important. In this work we deal with the knowledge transfer problem in the specific context of semantic place recognition.

Many works have been proposed in the literature to solve the place categorization problem [1, 2]. However, while these works represent the state-of-the-art methods for place classification, two fundamental problems arise. First, if the domain changes, *i.e.* the robot is moved to another environment, learning must be performed again from scratch. Second, these approaches are not designed to take advantage of any other available source of information, *e.g.* visual data downloaded from the web.

Transfer learning is the answer to these problems. With transfer learning the robot can reuse the previously learned classification models by adapting them. In this way the human labeling effort is greatly reduced as no image annotation is required in the new scenario. In [3] and [4] a transfer learning framework is introduced in a place classification context. However, these approaches make a strong assumption about the relation between the old and the new scenario: they have to share the same set of place categories. Unfortunately this assumption does not hold for many real world applications where even very similar environments may contain at least one or two place specific classes. Moreover, no previous methods have proposed a strategy to integrate multiple visual sources of information

Email addresses: gabriele.costante@studenti.unipg.it (Gabriele Costante), thomas.ciarfuglia@unipg.it (Thomas A. Ciarfuglia), paolo.valigi@unipg.it (Paolo Valigi), elisa.ricci@unipg.it (Elisa Ricci)

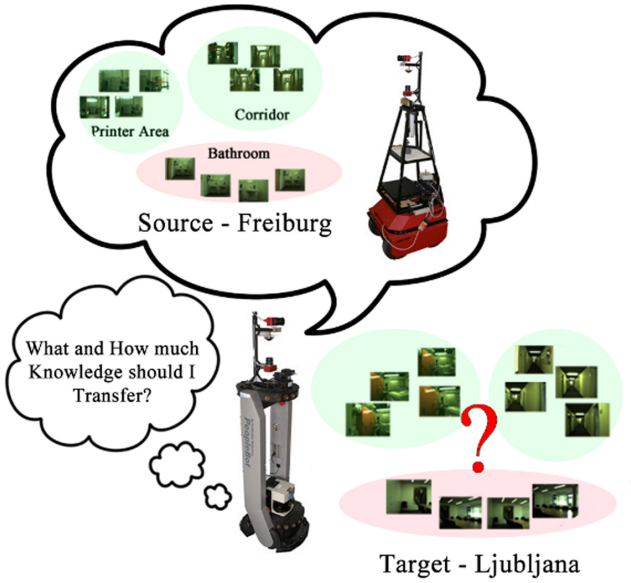


Figure 1: Illustration of the idea behind the proposed method: the robot performs visual place categorization using images collected in a new unknown environment (*target* data) and reasons about transferring knowledge available from a different scenario (*source* data).

while performing knowledge transfer.

In this work we present a novel transfer learning framework for place recognition targeting the more general situation when different categories are considered in the new and the old scenario. Figure 1 illustrates the intuition behind our approach.

Consider the simple case of a robot that is asked to recognize places in a unknown environment. No labels are provided for visual images collected in this scenario. Can the robot take advantage of other sources of information, *i.e.* images available from the web, or collected during its own past paths or by other robots? Two main questions must be answered in this context: *Is knowledge transfer convenient?* and *How much information should be transferred?* Intuitively, if the new scenario shares many room categories with the old one, then operating knowledge transfer is very helpful. If this condition is not met, transferring information could be potentially harmful. To answer the questions above in this paper we propose a risk sensitive transfer learning framework. Initially we compute a divergence measure between the visual data distributions associated to the old and the new environments. Then we introduce the concept of transfer risk, defined in terms of this divergence measure. Specifically we compare two risk measures, based respectively on the well known Kullback-Leibler (KL) divergence and on the Earth Mover's Distance (EMD) [5]. Place categorization is performed by considering a spectral decomposition problem that incorporates the notion of transfer risk. Here, the risk measure aims to balance the influence of the past experience with that of the visual data collected in the current scenario. To further improve the place recognition performance we also propose to use multiple complementary visual cues and we extend the proposed optimization framework accordingly. Our approach has been evaluated extensively on three publicly available datasets. Our results demonstrate that (i) our risk sensitive transfer learning framework outperforms the considered baselines in most of the cases and that (ii) combining multiple cues is greatly beneficial in this context.

To summarize, the main contributions of this work are:

- We cast the problem of place categorization in an unknown scenario within a *transfer learning* framework.
- We propose a method to quantify the similarity between the data from the old and the new scenario to understand *how much* to transfer. We compare two divergence measures, *i.e.* the *Kullback-Leibler (KL)* divergence and the *Earth Mover's Distance (EMD)*.
- We set up a spectral decomposition problem which effectively integrates information about past knowledge defining a notion of transfer risk.
- We extend our transfer learning approach to fuse *multiple visual cues*.

This paper extends our previous work in [6]. However, with respect to the conference paper, a more detailed discussion about related works is presented, the notion of transfer risk computed with EMD is introduced and the results of a more extensive experimental evaluation are shown.

The rest of the paper is organized as follows: Section 2 reviews related work. Section 3 introduces the proposed transfer learning framework, illustrating the details of the computation of the transfer risk, our risk sensitive domain adaptation approach and its extension to the multi-cue setting. Results and conclusions are presented respectively in Sections 4 and 5.

2. Related Works

2.1. Sharing Knowledge across Robotic Platforms

In recent decades we have witnessed to a rapid and impressive growth of data publicly available all over the web. This "big data revolution" has reshaped the traditional way in which people learn: learning new concepts and tasks has become faster and easier having access to this large amount of information. Inspired by this phenomenon, in the last few years robotics researchers have envisioned a similar process for developing the new generation of autonomous systems. It would be very useful if the robots could share their experiences and benefit from publicly available data.

An important contribution in this direction has been given by the RobotEarth platform [7] which has inspired our current work. Its aim is to provide a unified framework to share information and experiences among robots, in a World Wide Web fashion. The goal of RoboEarth is to allow robotic platforms to take advantage of the experience of their fellows, providing a giant network and a database repository. Following [7] many other works have recently been proposed. In [8] the authors introduced an approach for processing web resources to help a

robot perform everyday manipulation tasks. Instead of building specific models for each platform, the robot can process information gathered from some websites and acquire the knowledge needed to perform specific actions and tasks, *e.g.* cleaning a room or repairing machinery. Similarly in [9] Samadi *et al.* proposed a method where the web is searched to learn a model reflecting the probability of finding objects in specific rooms. The importance of building a shared knowledge system is also exploited in [10]: an approach to create a modular knowledge representation shared among robot platforms is proposed. The idea is that a robot can infer the optimal control decision taking into account the shared information.

2.2. Transfer Learning

The majority of machine learning algorithms develop from a common assumption: the training and test data are drawn from the same probability distribution. While for many applications this assumption holds, there are many real world scenarios where it does not. Since collecting labels in the new domain often requires a massive annotation effort, approaches of transfer learning are the only possible solution. In the transfer learning scenario, one has access to many labeled data from a *source* domain, while few data are available in the *target* domain. The idea is to learn a classification/regression model on target data by leveraging useful information from the source.

In the last few years several transfer learning approaches have been developed (see [11] for a survey). In [12] Shi *et al.* addressed the domain adaptation problem across different feature spaces. This approach is based on three fundamental steps. First spectral embedding is used to unify the feature spaces of the source and the target sets. Then a sampling strategy is adopted to select the source samples most related with the target data. Finally, a Bayes approach is used.

The problem of designing a method able to handle the situation where source and target data have different feature representation is also addressed in [13]. The authors first introduce a linear transformation that maps features from the target to the source domain and then propose learning the transformation and the classifier parameters jointly. Another interesting work is presented in [14] where the concept of dual transfer learning is introduced. In [15] a metric learning approach for domain adaptation is proposed.

Two interesting methods are described in [16] and [17]. In [16] the authors introduce the Domain Adaptation Machine approach which takes advantages from a set of classifiers, learned from different source domains, to improve the classification performances on the target domain. Conversely, in [17] a geodesic flow kernel strategy is proposed to model domain shift from the source to the target domain. Moreover, they also introduce a metric to measure how to automatically select the optimal source domain for adaptation and avoid the less desirable ones. In fact, a key issue in transfer learning is the ability to avoid negative transfer, *i.e.* the harmful situation where integrating knowledge from source data actually degrades classifier performance on the target domain. The large majority of previous approaches, *e.g.* [18, 19], have addressed this problem in a supervised setting, *i.e.* when some labeled target samples are available. Instead, in this work we deal with the negative transfer problem in an unsupervised setting.

2.3. Semantic Place Classification

In the last few years many works have addressed the problem of semantic place categorization. In [2] a probabilistic framework is built to combine heterogeneous sources of information. To deal with uncertainties provided by data from different sensors and allow spatial reasoning, a conceptual map representation in terms of a probabilistic chain graph model is introduced. The work in [1] also describes a multi-modal place classification system operating in an indoor environment. In [20] a recognition algorithm measuring its own level of confidence on classification results is proposed. An incremental learning approach to place recognition is proposed in [21]. With this method the same classification performance of the batch algorithm is obtained while the algorithm runs online. Furthermore, the algorithm is able to keep the memory requirements low while the system updates its internal representation. In [22] a low-dimensional global image descriptor is proposed, to provide robust and strong contextual information about an image to be used for scene categorization tasks.

However most of these previous works assume that the robot operates in the same scenario where data used to learn its classification model are collected, *i.e.* training and test data are supposed to be drawn from the same distribution.

In the context of place recognition, only a few knowledge transfer approaches have been proposed. In [23] a transfer learning framework is introduced for place classification robust to illumination and environment changes. The domain adaptation problem is also addressed in [3, 4, 24]. In [3] a SVM-based method for knowledge transfer across robots is presented. The proposed algorithm can handle new information continuously for incremental model adaptation. However, the robots are all assumed to perform the same tasks. Another issues arises in the model update step: old data are discarded without comparing them with the new ones, while in the ideal case old data should be eliminated only if they are significantly different from the most recent ones. In other words, no measure to quantify the transfer risk is implemented. In [4] a transfer learning algorithm based on least square SVM is proposed. This approach embeds a notion of transfer risk in the learning process, *i.e.* allows the system to automatically understand *how much* to transfer. On the other hand, like previous methods, it relies on supervised learning. In many situations this may be too restrictive an assumption. Robots are often moved into environments that can significantly differ from the previous ones, *e.g.* with different room categories. Moreover, labeling data in the new scenario usually requires some human annotation effort, an annoying procedure which it is desirable to avoid. In [24] transfer learning is adopted for object classification. Objects are described through a combination of different visual cues (*e.g.* color, texture, shape), then the learned object models are shared among multiple robot platforms.

In contrast to all these previous works, in this paper we consider a more challenging scenario: we focus on domain adaptation when different visual categories are considered in the

source and in the target data, and while we assume access to labeled source data, no labels are provided in the target domain. Moreover, to our knowledge there are no works that address the problem of multiple cues integration in a transfer learning context.

3. Transfer Learning for Place Recognition

In this section we describe the proposed knowledge transfer approach for place categorization. We consider the situation where a robotic platform has to perform scene recognition in a completely unknown scenario. Since it has no a priori information about the new environment, it can rely only on knowledge gathered through its past experiences or by other robots. More specifically, in this work we assume that, while the robot does not have at its disposal annotated data for the new scenario (the *target* domain), it can access other sources of information, *i.e.* image datasets or videos recorded in different places for which labels are provided (the *source* data). Can the robot use these visual data and decide autonomously these are useful for the current place recognition task, *i.e. how much* the past video sequences are similar to those it observes in the new scenario? More importantly, can it avoid negative transfer, *i.e.* discard the source data which will degrade recognition accuracy in the new environment?

Our approach is based on two main phases. In the first phase, the similarity between the source and the target data distributions is assessed. This step aims to understand whether the two locations have similar visual appearances and sets the transfer risk accordingly. In the second step, if the transfer risk is small, the information provided by the labeled source data is used in the form of constraints to learn our place recognition model. On the contrary, if the transfer risk is high, we rely only on the images from the new location and learning is reduced to cluster target data.

More formally, we are given a set (the *source data*) $\mathcal{S} = \{(\mathbf{x}_1^s, y_1^s), (\mathbf{x}_2^s, y_2^s), \ldots, (\mathbf{x}_{N_s}^s, y_{N_s}^s)\}$ where $\mathbf{x}_i^s \in \boldsymbol{R}^D$ are visual features extracted from video frames and $y_i^s \in \{1, 2, \ldots, K_C^S\}$ are the corresponding labels indicating the room types (*e.g.* corridor, office, etc), and a set $\mathcal{T} = \{\mathbf{x}_1^t, \mathbf{x}_2^t, \ldots, \mathbf{x}_{N_t}^t\}$ (the *target* data), where $\mathbf{x}_i^t \in \boldsymbol{R}^D$ are visual features extracted in the new scenario for which labels are not available. We are interested in learning a model in order to classify the target data. Note that the categories of the target data are not the same as the K_C^S classes in $\mathcal{S}$. As the target and the source data belong to different probability distributions, respectively $\mathcal{P}_\mathcal{S}$ and $\mathcal{P}_\mathcal{T}$, we would also like to measure the distance between them in order to quantify the risk of knowledge transfer, *i.e.* of using the source data to build a suitable model for the target data.

3.1. Clustering-based Transfer Risk

As the target and the source data correspond to video sequences recorded under different conditions, it is reasonable to assume that they belong to different distributions. Therefore, in order to build a robust learning model, a first step is to measure the distance between them in order to quantify the risk of knowledge transfer.

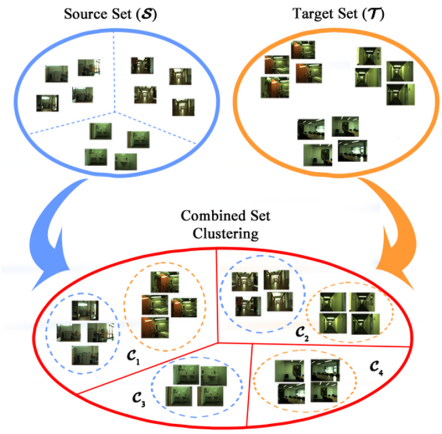


Figure 2: KL divergence computation. $|\mathcal{S}|$ and $|\mathcal{T}|$ represent the cardinality of the source and the target data set respectively, while $|C_c|$ indicates the size of cluster c. The term $|\mathcal{S} \cap C_c|$ ($|\mathcal{T} \cap C_c|$) represents the cardinality of the intersection between the source (or target) set and the cluster c.

A popular approach to computing the distance between distributions is the Kullback-Leibler (KL) divergence, defined as:

$$KL(\mathcal{S},\mathcal{T}) = \sum_x P_\mathcal{T}(x) \log \frac{P_\mathcal{T}(x)}{P_\mathcal{S}(x)} \tag{1}$$

As calculating the KL divergence directly from the data can be time consuming, in [25] a more practical solution is proposed, where an approximation is computed based on the output of a clustering algorithm operating on the combined data (source and target data together). More specifically the following definition of *Clustering-based KL divergence* is proposed:

$$D_c^{KL}(\mathcal{S},\mathcal{T}) = \frac{2}{|\mathcal{T}|}\sum_{c=1}^{|\mathcal{C}|}(\frac{|\mathcal{T}\cap C_c|}{|C_c|}\log\frac{|\mathcal{T}\cap C_c|}{|\mathcal{S}\cap C_c|}) + \log\frac{|\mathcal{S}|}{|\mathcal{T}|} \tag{2}$$

where $|\mathcal{S}|$ and $|\mathcal{T}|$ represent the cardinality of the source and the target data sets respectively, $|\mathcal{C}|$ is the number of clusters, while $|C_c|$ indicates the size of cluster c. The term $|\mathcal{S} \cap C_c|$ ($|\mathcal{T} \cap C_c|$) represents the cardinality of the intersection between the source (or target) set and the cluster c. The computation of the clustering-based KL divergence is illustrated in Fig.2 (see [25] for details on the derivation of (2)).

To measure the distance between the source and the target distributions $\mathcal{P}_\mathcal{S}$ and $\mathcal{P}_\mathcal{T}$ we also propose a different approach based on Earth Mover's Distance [5]. Similarly to the KL divergence computation, this approach is also based on using clustering algorithms to calculate the distance between two distributions. More specifically, by running Normalized-Cut [26] sepa-

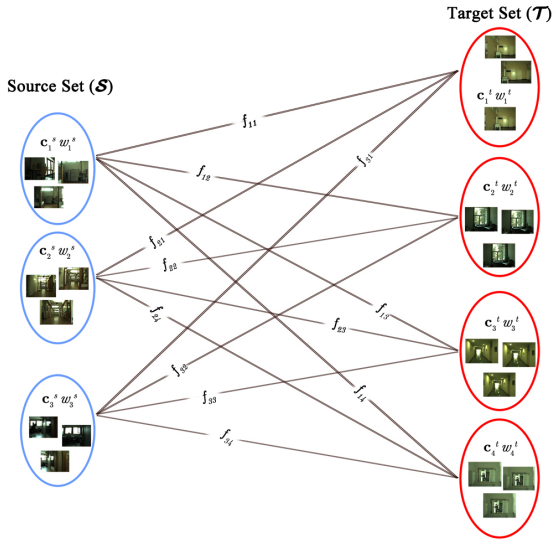


Figure 3: EMD distance computation. First clustering is performed on both the source (blue ellipses) and the target sets (red ellipses) to compute the centroids $\mathbf{c}_i^s$ and $\mathbf{c}_i^s$. Then the weights w_i^s and w_i^t are set counting the number of images in each cluster. Finally the flows f_{ij} are computed solving the transportation problem (3)

rately on the source and the target data, we compute the signatures $S = \{(\mathbf{c}_1^s, w_1^s), \dots (\mathbf{c}_M^s, w_M^s)\}$ and $T = \{(\mathbf{c}_1^t, w_1^t), \dots (\mathbf{c}_M^t, w_M^t)\}$, where $\mathbf{c}_i^s$, $\mathbf{c}_i^t$ are the cluster centroids, respectively computed on the source and the target data and w_i^s, w_i^t denotes the weights associated to each cluster. In this paper for thr sake of simplicity we consider the same number of clusters M for the source and the target data, and the cardinality of each cluster is used as cluster weight.

Given two signatures S and T, the Earth Mover's Distance (EMD) between the associated data distribution $\mathcal{P}_S$ and $\mathcal{P}_\mathcal{T}$ is defined by the following transportation problem:

$$D_c^{EMD}(\mathcal{S},\mathcal{T}) = \min_{f_{ij}\geq 0} \sum_{i,j=1}^{M} d_{ij} f_{ij} \qquad (3)$$
$$\text{s.t.} \sum_{i=1}^{M} f_{ij} = w_j^s \quad \sum_{j=1}^{M} f_{ij} = w_i^t$$

where f_{ij} are flow variables and d_{ij} is the ground distance $d_{ij} = \|\mathbf{c}_i^s - \mathbf{c}_j^t\|^2$. In a nutshell, the EMD represents the minimum cost needed to transform one distribution into another. A representation of its computation is depicted in Fig. 3. The motivation of using EMD to measure the distance between distributions lies mainly in its computational efficiency and on the possibility of handling partial matches between sets in a very natural way and to reflect the notion of nearness between clusters properly, thanks to ground distance computation.

The risk of transferring source data information while learning from target data is defined as follows:

$$R_{\mathcal{S},\mathcal{T}} = \frac{1}{1 + e^{(\gamma - D_c(\mathcal{S},\mathcal{T}))}} \qquad (4)$$

where $D_c(\mathcal{S},\mathcal{T})$ is the distance between the source and the target distributions computed with (2) or (3) and γ is a fixed parameter. This exponential form allows the risk to be normalized between [0,1]. The algorithms to compute the transfer risk are illustrated in Algorithm 1 and Algorithm 2.

Algorithm 1 Transfer risk computation with Kullback Leibler divergence

Input: source data $\mathcal{S}$, target data $\mathcal{T}$, total number of categories K_C

$\mathbf{W}$ = computeSimilarityMatrix($\mathcal{S}$, $\mathcal{T}$)
Set $\mathbf{D}$ with $\mathbf{D}_{ii} = \sum_j \mathbf{W}_{ij}$
Set $\mathbf{L}$ = $\mathbf{D}$ - $\mathbf{W}$
$\mathbf{U}$ = eig($\mathbf{D}^{-\frac{1}{2}}\mathbf{L}\mathbf{D}^{-\frac{1}{2}}$, K_C)
$\mathbf{U} = \mathbf{D}^{-\frac{1}{2}}\mathbf{U}$
Normalize $\mathbf{U}$ by row where $\mathbf{U}_{ij} = \mathbf{U}_{ij} / \sqrt{\sum_{l=1}^{K_C} \mathbf{U}_{il}^2}$
C = kmeans($\mathbf{U}$, K_C)
Compute $KL(\mathcal{S},\mathcal{T})$ with (2) and $R_{\mathcal{S},\mathcal{T}}$ using (4)

Output: risk $R_{\mathcal{S},\mathcal{T}}$

3.2. *Transfer Learning with Different Class Labels*

The transfer learning approach we adopt in this paper is an extension of the Normalized-Cut algorithm [26]. It amounts to solving the following optimization problem:

$$\min_{\mathbf{U}} \frac{\mathbf{U}^T\mathbf{L}\mathbf{U}}{\mathbf{U}^T\mathbf{D}\mathbf{U}} + \beta((1 - R_{\mathcal{S},\mathcal{T}})\|\mathbf{M}_\mathcal{S}\mathbf{U}\|^2 + R_{\mathcal{S},\mathcal{T}}\|\mathbf{M}_\mathcal{T}\mathbf{U}\|^2) \qquad (5)$$

where $\mathbf{L} = \mathbf{D} - \mathbf{W}$ is the Laplacian matrix, $\mathbf{W}$ is the similarity matrix computed on the entire dataset $\mathcal{S}\cup\mathcal{T}$, $\mathbf{D}$ = diag($\mathbf{We}$) and $\mathbf{e}$ is a vector with all the coordinates set to 1. The matrix $\mathbf{M}_\mathcal{S} = [\mathbf{m}_1\mathbf{m}_2 \dots \mathbf{m}_{N_c^s}]^T$ where $\mathbf{m}_i \in \boldsymbol{R}^{N_s+N_t}$ is a vector with 1 in the i-th position and -1 in the j-th position if the source data points $\mathbf{x}_i$ and $\mathbf{x}_j$ have the same labels. The matrix $\mathbf{M}_\mathcal{T}$ is similarly defined on the target data. However, as for the target data labels are not provided, a preprocessing phase where the target data are clustered with Normalized-Cut [26] is performed. The matrix $\mathbf{M}_\mathcal{T}$ is then defined using as labels the vectors indicating the cluster membership.

The objective function in (5) is the sum of two terms. The first term simply aims to cluster the entire dataset using Normalized - Cut, while the second term force the learned clustering structure to satisfy some constraints. More specifically two sets of constraints are imposed. One guarantees that the learned projection matrix leads to clusters consistent with the labels of the source data. The second set of constraints imposes some consistency between the new clustering results and those that are obtained grouping only the target data. The trade-off between transferring source data information and not using it is regulated by the risk $R_{\mathcal{S},\mathcal{T}}$.

Fig. 4 illustrates the intuition behind the proposed approach. Analyzing the anti-diagonal sub-blocks of the matrix $\mathbf{W}$ it is possible to detect cross-domain similarities. High similarity values are observed when source and target data correspond to places having similar visual appearance, *e.g.* the same categories 'Corridor' or 'Printer Area'. In these cases transferring the source information, *i.e.* considering source constraints,

Algorithm 2 Transfer risk computation with EMD

Input: source data $\mathcal{S}$, target data $\mathcal{T}$, total number of categories M

$\mathbf{W}^s$ = computeSimilarityMatrix($\mathcal{S}$)
Set $\mathbf{D}$ with $\mathbf{D}_{ii} = \sum_j \mathbf{W}^s_{ij}$
Set $\mathbf{L}$ = $\mathbf{D}$ - $\mathbf{W}^s$
$\mathbf{U}$ = eig($\mathbf{D}^{-\frac{1}{2}}\mathbf{L}\mathbf{D}^{-\frac{1}{2}}$, M)
$\mathbf{U} = \mathbf{D}^{-\frac{1}{2}}\mathbf{U}$
Normalize $\mathbf{U}$ by row where $\mathbf{U}_{ij} = \mathbf{U}_{ij} / \sqrt{\sum_{l=1}^{M} \mathbf{U}_{il}^2}$
$[\mathbf{c}^s_1, w^s_1, \ldots, \mathbf{c}^s_M, w^s_M]$ = kmeans($\mathbf{U}$, M)

$\mathbf{W}^t$ = computeSimilarityMatrix($\mathcal{T}$)
Set $\mathbf{D}$ with $\mathbf{D}_{ii} = \sum_j \mathbf{W}^t_{ij}$
Set $\mathbf{L}$ = $\mathbf{D}$ - $\mathbf{W}^t$
$\mathbf{U}$ = eig($\mathbf{D}^{-\frac{1}{2}}\mathbf{L}\mathbf{D}^{-\frac{1}{2}}$, M)
$\mathbf{U} = \mathbf{D}^{-\frac{1}{2}}\mathbf{U}$
Normalize $\mathbf{U}$ by row where $\mathbf{U}_{ij} = \mathbf{U}_{ij} / \sqrt{\sum_{l=1}^{M} \mathbf{U}_{il}^2}$
$[\mathbf{c}^t_1, w^t_1, \ldots, \mathbf{c}^t_M, w^t_M]$ = kmeans($\mathbf{U}$, M)

Compute $D_c(\mathcal{S}, \mathcal{T})$ with (3) and $R_{\mathcal{S},\mathcal{T}}$ using (4)

Output: risk $R_{\mathcal{S},\mathcal{T}}$

leads to a great improvement in accuracy when performing place recognition in the target scenario.

Defining:

$$\mathbf{A} = \mathbf{L} + \beta((1 - R_{\mathcal{S},\mathcal{T}})\mathbf{M}_{\mathcal{S}}^T\mathbf{M}_{\mathcal{S}} + R_{\mathcal{S},\mathcal{T}}\mathbf{M}_{\mathcal{T}}^T\mathbf{M}_{\mathcal{T}})$$
$$\mathbf{Y} = \mathbf{D}^{\frac{1}{2}}\mathbf{U} / \|\mathbf{D}^{\frac{1}{2}}\mathbf{U}\| \tag{6}$$

the optimization problem (5) can be reformulated as follows:

$$\begin{aligned} \min_{\mathbf{Y}} \quad & \mathbf{Y}^T\mathbf{D}^{-\frac{1}{2}}(\mathbf{D} - \mathbf{W})\mathbf{D}^{-\frac{1}{2}} \\ & +\beta((1 - R_{\mathcal{S},\mathcal{T}})\|\mathbf{M}_{\mathcal{S}}\mathbf{D}^{-\frac{1}{2}}\mathbf{Y}\|^2 + R_{\mathcal{S},\mathcal{T}}\|\mathbf{M}_{\mathcal{T}}\mathbf{D}^{-\frac{1}{2}}\mathbf{Y}\|^2) \\ = \min_{\mathbf{Y}} \quad & \mathbf{Y}^T\mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}\mathbf{Y} \\ = \min_{\mathbf{U}} \quad & \frac{\mathbf{U}^T\mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}\mathbf{U}}{\mathbf{U}^T\mathbf{U}} \end{aligned} \tag{7}$$

The resulting transfer learning method is presented in Algorithm 3 and Fig. 5 depicts its fundamental steps.

3.3. Transfer Learning with Complementary Visual Cues

Integrating multiple complementary cues has been shown to be beneficial for many different visual tasks [1, 27, 28]. In this section we describe the extension of the transfer learning framework introduced in the previous section in order to operate with two complementary visual cues. In particular in the context of indoor place recognition we adopt two different descriptors: the spatial pyramid matching kernel (SPMK) originally proposed in [29] and the Spatial Principal component Analysis of Census Transform histograms (SPACT) descriptor [30].

The SPMK representation has been shown to be very effective and has been widely used for place recognition applications in the context of robotic systems. Specifically the pyramid

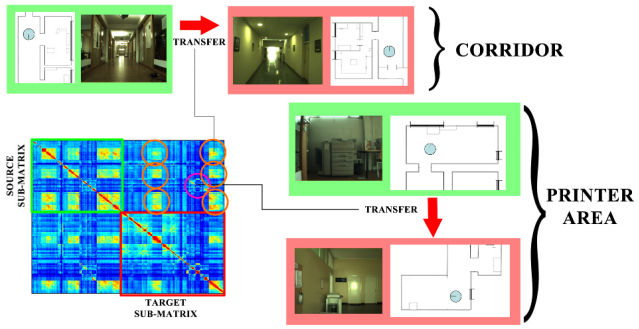


Figure 4: The similarity matrix $\mathbf{W}$ computed on source and target data. The anti-diagonal sub-blocks indicate cross-domain similarities. Orange circles highlight the similarity between source and target data corresponding to the class 'Corridor', while the purple one refers to the similarities among printer areas.

matching strategy works by dividing the image into a set of increasingly coarser grids and computing a weighted sum of the matches that occurs at each level. Two points are said to match if they are in the same cell, given a certain resolution. According to this scheme the matching kernel is computed calculating the histogram intersection between the vectors formed by concatenating the weighted histograms at all resolutions. More specifically we use the SIFT descriptors [31] to extract interest points from images.

The CENTRIST descriptor [32] was originally proposed for scene classification tasks and has been shown to be very effective as it captures the structural properties of the scenes. The Census Transform is a nonparametric local transform introduced to compare local patches. It compares the intensity of a pixel with its eight neighbors and the binary values obtained replaces the pixel itself. The CENTRIST descriptor has 256 bins where each bin counts the occurrences of a value in the range [0 255] after the application of the Census Transform to the entire image. Following [32], to obtain the final descriptor, we also apply the spatial-pyramid [29] to capture the global structure of the image at a large scale and Principal Component Analysis (PCA) to reduce the dimensionality of histograms and obtain a more compact representation. The final descriptor is called SPACT (Spatial Principal component Analysis of Census Transform histograms).

Given $\mathbf{L}_S = \mathbf{D}_S^{-\frac{1}{2}}\mathbf{W}_S\mathbf{D}_S^{-\frac{1}{2}}$ and $\mathbf{L}_C = \mathbf{D}_C^{-\frac{1}{2}}\mathbf{W}_C\mathbf{D}_C^{-\frac{1}{2}}$, where $\mathbf{W}_S$ and $\mathbf{W}_C$ are respectively the SPMK and the SPACT kernels and $\mathbf{D}_S = \text{diag}(\mathbf{W}_S\mathbf{e})$, $\mathbf{D}_C = \text{diag}(\mathbf{W}_C\mathbf{e})$, the problem of transfer learning can be formulated as follows:

$$\begin{aligned} \max_{\mathbf{U}_S,\mathbf{U}_C} \quad & \textstyle\sum_{i\in\{S,C\}} tr(\mathbf{U}_i^T\mathbf{B}_i\mathbf{U}_i) + \lambda\mathcal{A}(\mathbf{U}_S, \mathbf{U}_C) \\ \text{s.t.} \quad & \mathbf{U}_S^T\mathbf{U}_S = I\,, \mathbf{U}_C^T\mathbf{U}_C = I \end{aligned} \tag{8}$$

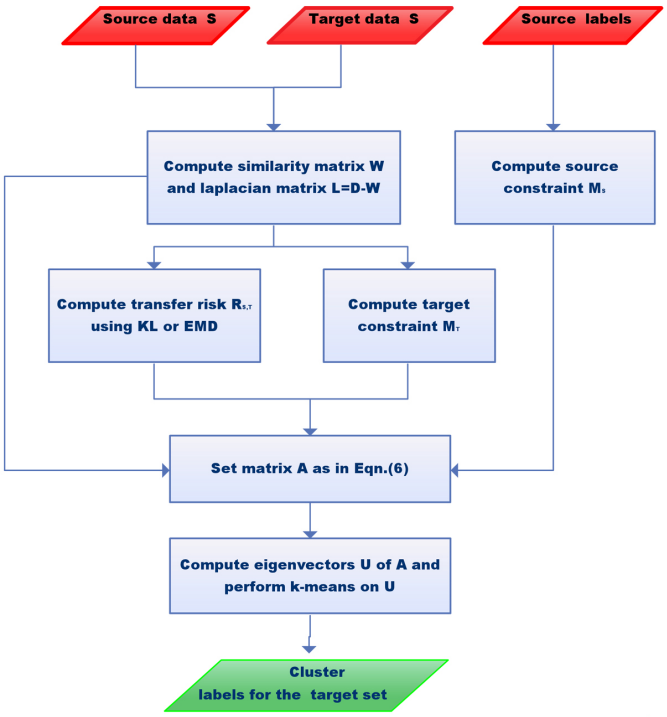


Figure 5: Block diagram illustrating our transfer learning approach.

with:

$$\mathbf{B}_S = \mathbf{L}_S - \beta_S(1 - R^S_{\mathcal{S},\mathcal{T}})\mathbf{M}^T_{\mathcal{S}}\mathbf{M}_{\mathcal{S}} + R^S_{\mathcal{S},\mathcal{T}}\mathbf{M}^T_{\mathcal{T}_S}\mathbf{M}_{\mathcal{T}_S} \quad (9)$$

$$\mathbf{B}_C = \mathbf{L}_C - \beta_C(1 - R^C_{\mathcal{S},\mathcal{T}})\mathbf{M}^T_{\mathcal{S}}\mathbf{M}_{\mathcal{S}} + R^C_{\mathcal{S},\mathcal{T}}\mathbf{M}^T_{\mathcal{T}_C}\mathbf{M}_{\mathcal{T}_C} \quad (10)$$

where λ is an appropriate regularization parameter and $\mathcal{A}(\mathbf{U}_S, \mathbf{U}_C)$ is the agreement term between the two views defined as follows:

$$\mathcal{A}(\mathbf{U}_S, \mathbf{U}_C) = tr(\mathbf{U}_S\mathbf{U}_S^T\mathbf{U}_C\mathbf{U}_C^T) \quad (11)$$

In practice the proposed optimization problem (8) is a sum of two main terms. The first aims to reason about transferring knowledge from source data separately for each modality, the second is meant to impose consistency between the two projected eigenspaces.

To solve this problem efficiently an alternating optimization approach is adopted, solving separately for $\mathbf{U}_S$ and $\mathbf{U}_C$. In particular for a given $\mathbf{U}_C$ we get:

$$\max_{\mathbf{U}_S} \quad tr\left\{\mathbf{U}_S^T(\mathbf{B}_S + \lambda\mathbf{U}_C\mathbf{U}_C^T)\mathbf{U}_S\right\} \quad (12)$$
$$\text{s.t.} \quad \mathbf{U}_S^T\mathbf{U}_S = I$$

which can be easily solved using spectral decomposition methods. Similarly when $\mathbf{U}_S$ is fixed, an analogous problem must be solved with respect to $\mathbf{U}_C$. The main steps of the proposed multi-cue transfer learning method are shown in Algorithm 4.

Algorithm 3 Transfer Learning algorithm

Input: source data $\mathcal{S}$, target data $\mathcal{T}$, number of target categories K_C^T, total number of categories K_C, β
$\mathbf{W}$ = computeSimilarityMatrix($\mathcal{S}, \mathcal{T}$)
$\mathbf{M}_{\mathcal{S}}$ = computeSourceConstraints(y^s)
$R_{\mathcal{S},\mathcal{T}}$ = computeRisk with Algorithm 1 or 2
Set $\mathbf{D}$ with $\mathbf{D}_{ii} = \sum_j \mathbf{W}_{ij}$
Set $\mathbf{L} = \mathbf{D} - \mathbf{W}$
$\mathbf{M}_{\mathcal{T}}$ = computeTargetConstraints($\mathbf{W}$, K_C^T)
$\mathbf{A} = \mathbf{L} + \beta((1 - R_{\mathcal{S},\mathcal{T}})\mathbf{M}^T_{\mathcal{S}}\mathbf{M}_{\mathcal{S}} + R_{\mathcal{S},\mathcal{T}}\mathbf{M}^T_{\mathcal{T}}\mathbf{M}_{\mathcal{T}})$
$\mathbf{U}$ = eig($\mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}$, K_C^T).
$\mathbf{U} = \mathbf{D}^{-\frac{1}{2}}\mathbf{U}$
Normalize $\mathbf{U}$ by row where $\mathbf{U}_{ij} = \mathbf{U}_{ij}/\sqrt{\sum_{l=1}^{K_C^T}\mathbf{U}_{il}^2}$
C = kmeans($\mathbf{U}$, K_C^T)
Output: Target set clusters C

4. Experimental Results

4.1. Datasets

To evaluate the effectiveness of the proposed approach we ran several experiments selecting as source and target data images gathered in different real world environments. We consider sequences from three different datasets: the COLD [33], the KTH-IDOL2 [21] and the VPC [34] datasets. Figure 6 shows some sample images for all the place categories of the considered sequences.

The COLD dataset consists of several video sequences from university laboratories in three different European cities: the Visual Cognitive Systems Laboratory at the University of Ljubljana, the Autonomous Intelligent System Laboratory at the University of Freiburg and the Language Technology Laboratory at the German Research Center for Artificial Intelligence in Saarbrücken. The video sequences have been collected using three different robotic platforms (an ActivMedia People Bot, an ActiveMedia Pioneer-3 and an iRobot ATRV-Mini) with two Videre Design MDCS2 digital cameras to obtain perspective and omnidirectional views. Each frame is registered with the associated absolute position recovered using laser and odometry data and annotated with a label representing the corresponding place. The acquisition was performed under different weather and illumination conditions and across different days. Moreover each dataset has some sequences containing rooms with similar functionalities also shared by the other two. For each lab there are place categories in common with the other datasets, *e.g.* Corridor (CR), Printer Area (PA) or Bathroom (TL), but also dataset specific rooms, *e.g.* the Robotics Lab in the Saarbrücken sequences or the Stairs Area in the Freiburg data. Moreover rooms of different datasets associated with the same labels may have very different appearance. An example is the Corridor (CR) class: the separating walls between offices in the Freiburg data are made of glass, while in the Saarbrücken and Ljubljana sequences concrete walls are depicted. Therefore, transfer learning is very challenging.

The IDOL2 dataset is similar to COLD: it contains several image sequences recorded under various weather and illumination conditions. The acquisition was performed in an indoor environment that contains five types of rooms: One-Person Office (OO), Two-person Office (TO), Corridor (CR), Kitchen (KT) and Printer Area (PA). The robotic platforms used were a Mo-

Algorithm 4 Multi-Cue Transfer Learning

Input: source data $\mathcal{S}$, target data $\mathcal{T}$, number of target categories K_C^T, total number of categories K_C, β_S, β_C, λ, number of iteration T

$\mathbf{W}_S$ = computeSPMKernel($\mathcal{S}, \mathcal{T}$)
$\mathbf{W}_C$ = computeCENTRISTKernel($\mathcal{S}, \mathcal{T}$)
$\mathbf{M}_S$ = computeSourceConstraints(y^s)
$R^S_{\mathcal{S},\mathcal{T}}$ = computeRisk with Algorithm 1 or 2
$R^C_{\mathcal{S},\mathcal{T}}$ = computeRisk with Algorithm 1 or 2
$\mathbf{M}_{\mathcal{T}_S}$ = computeTargetConstraints($\mathbf{W}_S$, K_C^T)
$\mathbf{M}_{\mathcal{T}_C}$ = computeTargetConstraints($\mathbf{W}_C$, K_C^T)
Compute $\mathbf{B}_S$ and $\mathbf{B}_C$ using (9) and (10)
$\mathbf{U}_S$ = eig($\mathbf{B}_S$, K_C^T).
for $t = 1, \ldots, T$ **do**
  $\mathbf{U}_C$ = eig($\mathbf{B}_C + \lambda \mathbf{U}_S \mathbf{U}_S^T$, K_C^T).
  $\mathbf{U}_S$ = eig($\mathbf{B}_S + \lambda \mathbf{U}_C \mathbf{U}_C^T$, K_C^T).
endfor
Normalize $\mathbf{U}_S$ and $\mathbf{U}_C$
C = kmeans($[\mathbf{U}_S\ \mathbf{U}_C]$, K_C^T)
Output: Target set clusters C

bileRobots PeopleBot and a PowerBot equipped with a Canon VC-C4 camera.

Finally the VPC dataset consists of several sequences collected in six houses with different room categories. The dataset was recorded using a camera (JVC GR-HD1) mounted on a mobile tripod. We chose the VPC dataset to prove the effectiveness of our method in scenarios where knowledge transfer may be harmful.

4.2. Experimental Setup

To properly evaluate the performance of our method, we chose sequences extracted from all the datasets and recorded at different illumination conditions. In every experiment we select sequences where the source and the target data have different place categories. We only require that they have at least one specific room type (one class) in common. This is meant to show the validity of our method which operates in the realistic situation where transferring knowledge across different scenarios and determining automatically *how much* to transfer is essential. The labels of the source data provided in the ground truth files of each dataset are used to specify the source constraints and define $\mathbf{M}_S$.

We build the similarity matrices using the SPMK scheme and the SPACT descriptors introduced in the previous section. Specifically to compute the SPMK feature set we create a vocabulary of 400 visual words following the standard bag-of-words approach using 800 images as training set. Finally the histograms for each image are constructed projecting the extracted SIFT in the vocabulary at each level of resolution and for each cell. We choose $L = 3$ as the number of pyramid levels. The similarity matrix $\mathbf{W}_S$ is then obtained computing histogram intersection. For the SPACT descriptors we set the number of

Figure 6: Sample images for all the place categories extracted in the three datasets used. The red box contains room categories shared only by a subset of datasets, the orange boxes highlight specific rooms in each database and the green ones shows the categories which are common to all the datasets.

pyramid levels to $L = 3$ and the number of principal components equal to 40. After computing the SPACT histograms, we use the RBF kernel to calculate the similarity matrix $\mathbf{W}_C$.

In our experiments we first tested the proposed transfer learning approach using a single visual cue. We perform experiments both for the SPMK and the SPACT descriptors. We also compare our approach against two baselines: a *No-Transfer* method which applies a clustering algorithm (specifically Normalized Cut [26]) without any information transferring and a *Full Transfer* algorithm where the knowledge gathered from source is completely transferred to the target *i.e.* without considering the risk of transferring potentially harmful information. This situation is obtained setting $R(\mathcal{S}, \mathcal{T}) = 0$.

A second series of experiments aim to test the proposed multi-cue approach. As baselines we again consider the *No-Transfer* and *Full Transfer* methods. In this case the two visual cues are simply combined taking the average of the two computed kernels. In both single-cue and multi-cues tests the parameters β in (5) and β_S and β_C in (8) are set to 1.

The output of our algorithm consists in a set of clusters representing place categories, thus we measure the performance in

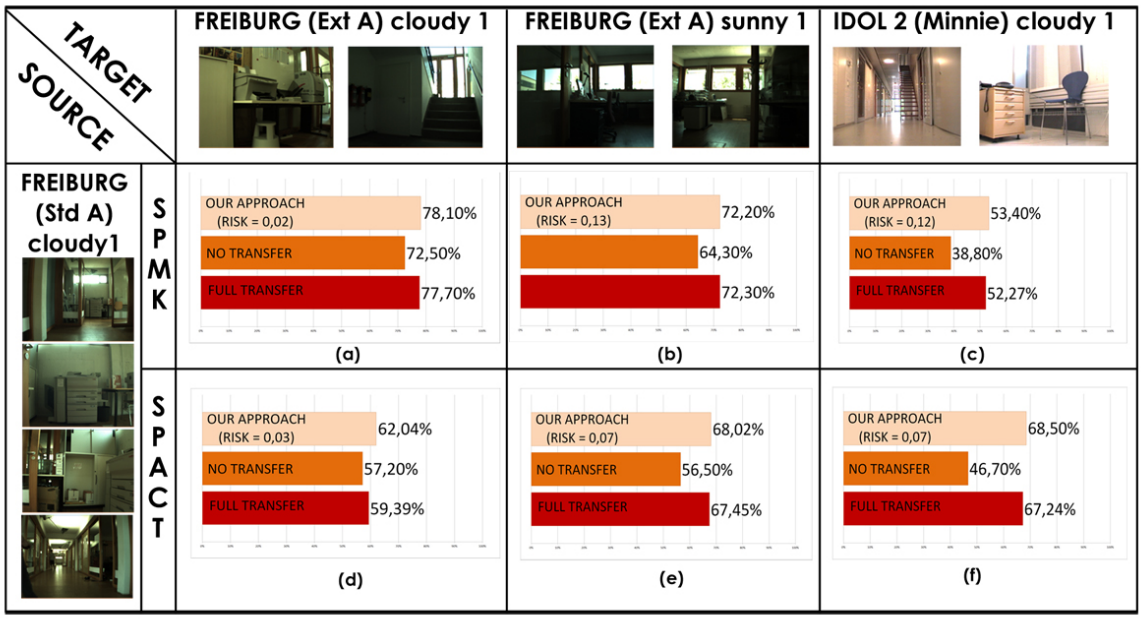


Figure 7: Single-cue place recognition experiments using the KL divergence measure.

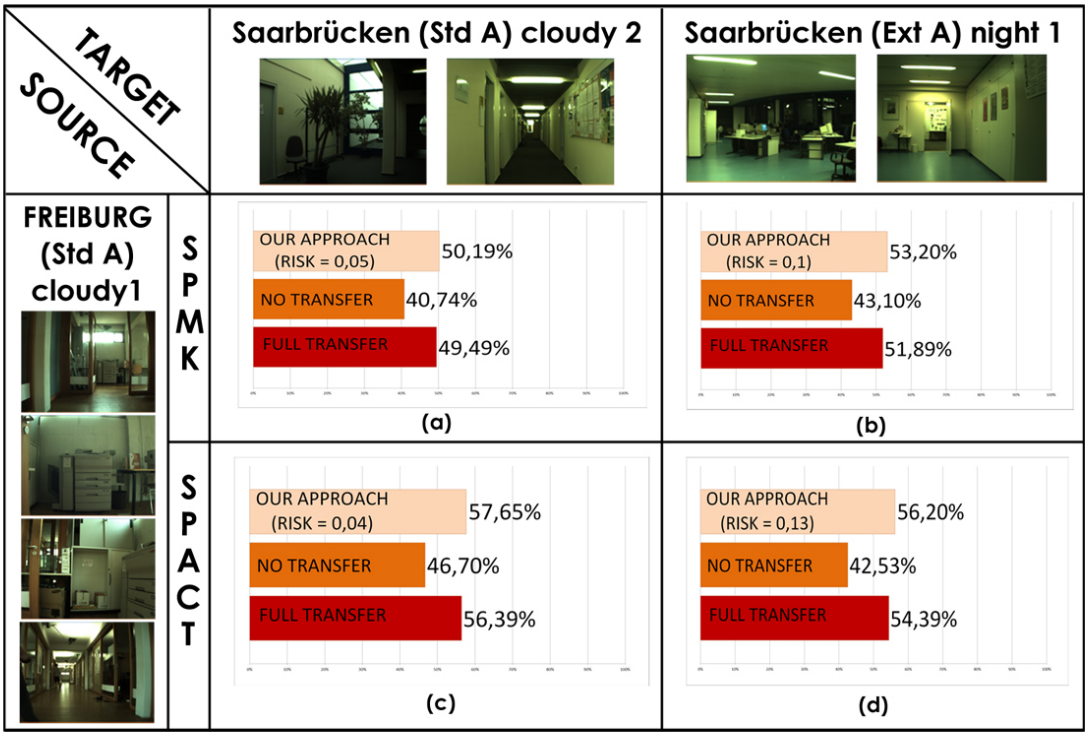


Figure 8: Single-cue place recognition experiments using the KL divergence measure.

terms of clustering accuracy [35]:

$$\text{Accuracy} = \frac{\sum_{i=1}^{N_T} \delta(y_i, map(c_i))}{N_T} \quad (13)$$

where N_T is the total number of images on target data, y_i is the true label for the i-th image, c_i is the cluster label. $\delta(y, c)$ is a function that is 1 if true label and cluster label are the same and 0 otherwise and $map(\cdot)$ is a permutation function that maps cluster labels to true labels. The optimal matching is found using the Hungarian algorithm [36]. Due to the variability introduced by the k-means algorithm, we repeat the clustering step after the spectral decomposition 10 times. The resulting average accuracy is considered. Since in our scenario we are interested in performing place categorization in the new scenario the clustering accuracy is evaluated on the target set.

4.3. Quantitative Evaluation

4.3.1. Single-cue Experiments

In a first series of experiments we show some place recognition results on target sequences using a single visual cue comparing the KL and the EMD divergence measure for risk computation. Our aim here is to demonstrate the capability of the proposed method to understand what to transfer, avoiding negative transfer and maximizing the use of information gathered from the source data.

Figures 7-10 depict the results obtained for the SPMK and SPACT experiments based on the KL risk measure. In Tables 1 and 2 the performances obtained using the EMD divergence are also reported. In cases where transferring knowledge is helpful, both the KL divergence and the EMD are high, *i.e.* the transfer risk is correctly set to a value close to zero. For example in the Freiburg cloudy - Freiburg cloudy experiments the No-Transfer algorithm achieves an accuracy of 72.5% while the Full-Transfer approach reaches 77.7% with the Spatial Pyramid Matching Kernel features. Similar results are obtained the in case of the SPACT features, respectively 57.2% and 59.39%. In both cases, since the source and the target distributions are similar, *i.e.* rooms have similar visual appearance, our strategy correctly determines that the source knowledge helps clustering the target data: we get 78.1% (KL divergence) and 78.3% (EMD distance) with the SPMK features and 62.04% (KL divergence) and 62.13% (EMD distance) with SPACT. Similarly in the Freiburg cloudy - Freiburg sunny experiment we demonstrate that our approach is robust to changes in environmental conditions. The sequences considered in this experiment have been gathered under different light conditions and share most of the room categories. Even in this more challenging situation our risk sensitive method correctly notices the similarity between distributions: in particular we get 72.3% (KL) and 74.55% (EMD) with SPMK and 68.02% (KL) and 69.56% (EMD) with SPACT while No-Transfer reaches respectively only 64.3% and 56.5%.

In the Freiburg-IDOL2, Freiburg-Saarbrücken cloudy 2 and Freiburg-Saarbrücken night 1 experiments our approach still outperforms both No-Transfer and Full Transfer (see Fig. 7.c, Fig.7.f and Fig. 8). The results show how knowledge transfer leads to a great improvement in place categorization performance when source and target scenarios are similar, even under different illumination conditions. Comparing the results obtained with our approach using KL divergence with those we get computing the risk based on EMD, similar performances are observed. Both divergence measures correctly detect domain similarities, thus we can effectively take advantage of the past knowledge.

The tests on the Ljubljana-Freiburg sequences consider another scenario. Here the source and the target sequences share some categories, while the others are rather different. The transfer risk, computed with both KL and EMD measures, is about $0.5 - 0.6$. In this case, both the No-Transfer and the Full-Transfer methods do not represent optimal approaches for handling this situation. As shown in Fig. 9.a and 9.c we get 79.2% with SPMK and 59.3 % with SPACT using the KL divergence, while both the baselines reach lower performances. Moreover from Table 1 and Table 2 we can observe that with the EMD measure we get similar results.

Tests on the Ljubljana-Saarbrücken and VPC-Freiburg sequences show how our distribution sensitive method avoids negative transfer. Here, the place categories in the source and the target data are very different so transferring knowledge from the source is potentially harmful. Hence the computed risk is close to 1, both with the EMD distance and the KL divergence. This is consistent with the fact that the No-transfer method outperforms the Full-transfer approach, both with the SPMK and the SPACT kernels. Our approach also outperforms the No-Transfer baseline, meaning that a small amount of information from the source data could be effectively used for improving the performance on clustering target data. This is reflected by the value of the risk which is slightly lower than 1.

Fig.10 shows an example where our method fails. The computation of both the clustering-based KL divergence and the EMD distance are not accurate and a risk close to 1 is obtained despite the Saarbrücken and the Freiburg sequences sharing some similar patterns. In this situation the Full-transfer method outperforms both our approach and the No-transfer algorithm. We believe that this is due to the fact that it is challenging to correctly compute the distance between distributions when the number of categories is large, *e.g.* 10-12. How to further improve our approach when several classes are considered will be addressed in future works.

The choice of the parameters K_C and M, *i.e.* the numbers of cluster to compute the divergence between distribution, requires a further discussion. In all the previous experiments we set the aforementioned parameters equal to the total number of different categories in the source and the target domains. Although this choice gives in most of the cases good results, there are some situations where it is not optimal. In Figure 11 we report the classification performances varying the number of clusters used to compute the KL divergence in three different experiments. Analyzing the results it is possible to observe that the optimal K_C value is related to the distribution of points of the source and the target sets. In the Ljubljana-Freiburg test we can capture the divergence by setting K_C equal to the total number of categories, *i.e.* 6, suggesting that the points are distributed in well-defined and separated chunks in the representation space. Conversely, we believe that in the Freiburg-

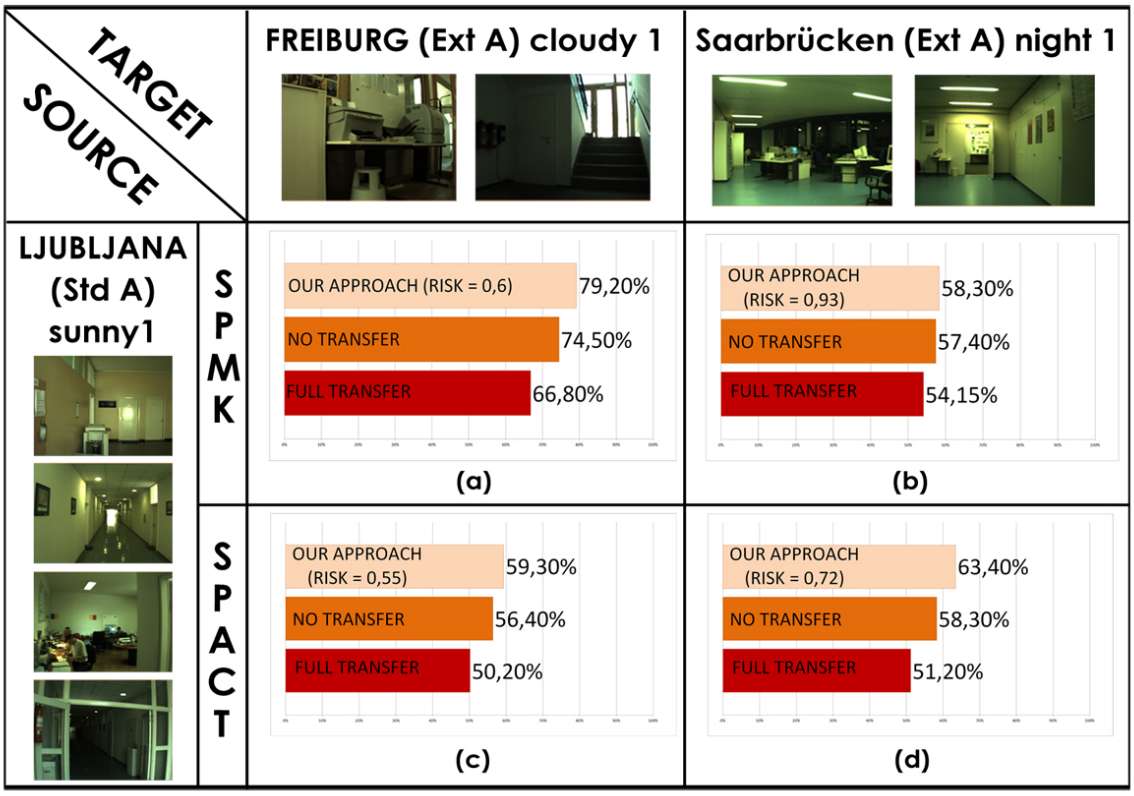


Figure 9: Single-cue place recognition experiments using the KL divergence measure.

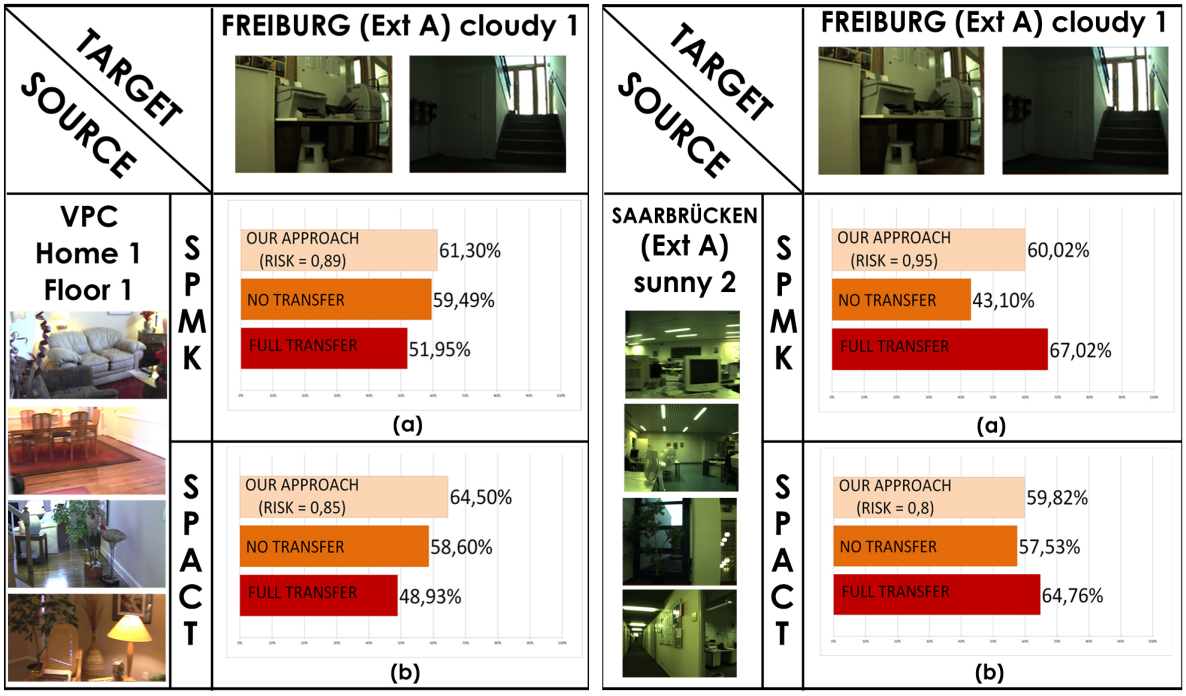


Figure 10: Single-cue place recognition experiments using the KL divergence measure.

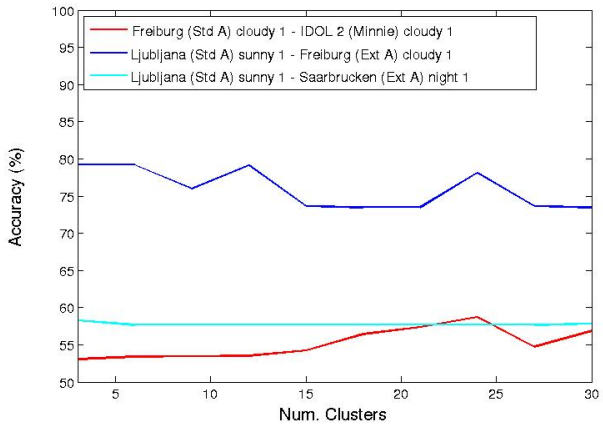


Figure 11: Classification accuracy comparison when varying the number of clusters used to compute the KL distance.

Ljubljana experiments it is more difficult to group the points in few large clusters, thus a higher value for K_C is required. Finally, the Ljubljana-Saarbrücken test shows a case where the number of clusters does not affect the performances. The automatic optimal selection of the parameters K_C and M will be considered in future works.

4.3.2. Multi-cue Experiments

In the second series of experiments we test our transfer learning approach when multiple visual cues are combined. Since both the SPMK and the SPACT features have different strengths and weaknesses, we combine the resulting similarity matrices in order to improve performances with respect to the single cue case. Results are shown in Fig. 12 where we compare the multiple cue approach with the single features one, considering both KL and EMD divergence measures. The value of the parameter λ in (8) is set to 0.5 in all our experiments. It is evident that the multi-cue strategy is beneficial for place recognition accuracy. For example in the Ljubljana-Freiburg experiment the single cue tests reach 80.35% (SPMK) and 62.12% (SPACT) with the EMD measure and 79.2% (SPMK) and 59.3% (SPACT) with the KL one, while combining cues we get 88.5% and 85.65% with EMD and KL respectively. Similarly in the Freiburg-IDOL2, where SPACT achieves lower performance with respect to SPMK, the multiple-cue approach outperforms both of them, *i.e.* we obtain an accuracy of 76.25% (EMD) and 73.13% (KL).

4.4. Comparison with Transfer Learning Methods

In this section we compare our risk sensitive approach with the Domain Adaptation Machine (DAM) method in [16], where the authors proposed a multiple source domain adaptation strategy that exploits a set of classifiers learned from multiple source domains. Although the method can handle multiple source sets and can benefit from labeled data in the target set, in our tests we consider only one source domain and no labeled data on the new domain to be consistent with the experiments in the previous sections. We use the code of the unsupervised version of DAM publicly available [1].

Table 3: Comparison with DAM: Classification Accuracy (%)

	F → L	F → S	L → F	L → S	S → F	S → L
Full Transfer	67.04	58.02	64.02	60.04	62.55	65.04
No Transfer	74.56	54.35	54.35	55.89	65.06	55.08
DAM [16]	**77.23**	**62.02**	56.75	60.03	54.09	65.45
Our Approach Risk - (KL)	76.45	61.55	64.03	61.51	68.86	**66.07**
Our Approach Risk - (EMD)	77.01	60.43	**64.45**	**62.04**	**69.76**	65.67

In [16] the algorithm assumes that the source and target categories are the same. Thus, despite the flexibility of our approach, that addresses scenarios where source and target domains can have different place categories, to achieve a fair comparison we restrict the experiments to sequence segments that share the same semantic places. In particular we choose three sequences, Freiburg(Std A) cloudy 1 (**F**), Ljubljana(Std A) sunny 1(**L**) and Saarbrücken(Std A) cloudy 2(**S**), selecting only the images that belong to place common to all of them. For each of the six possible source-target combinations we compute the SPMK kernel. Results are shown in Table 3.

It is clear that we outperform the DAM baseline in L→F, L→S, S→F and S→L, while the DAM reaches better performances in F→L and F→S. However, it should be noticed that while the DAM is specifically suited for application where the domains share the same categories, our approach can handle the more general and challenging scenario where the target set contains different image classes with respect to the source.

4.5. Computational Complexity

In the last set of experiments we discuss the computational complexity of the proposed approach. A single run of the overall pipeline requires the execution of Algorithms 1 or 2 to compute the transfer risk and, afterwards, the computation of cluster assignments following Algorithm 3 or Algorithm 4, respectively in case of single-cue or multi-cues transfer learning.

It is straightforward to notice that the most expensive operation, in terms of computational complexity, is the eigenvalues computation routine which requires $O(N^3)$ operations, where N is the total number of images. Therefore, in our MATLAB implementation, a single-cue transfer learning run needs to evaluate the **eig** function twice, while a multi-cue test requires $T + 1$ evaluations. The computational times associated to different dataset sizes are depicted in Figure 13. The tests are evaluated using a 2.4 GHz Intel core i7 processor.

The results show that a single eigenvalues decomposition takes 59.74 seconds when processing 3400 images and reaches 12160.23 seconds with 20000 elements in the dataset. When the complete single-cue transfer learning algorithm is evaluated, the required time is doubled, while in the multi-cues case it is multiplied by the number of iteration of the alternate optimization procedure.

However, in our experiments we observed that when the number of categories is not large, a dataset consisting of 5000-6000 images is a good trade-off between computational time

[1] http://vc.sce.ntu.edu.sg/transfer_learning_domain_adaptation/domain_adaptation_home.html

Table 1: Place recognition accuracy (%) obtained with SPMK features

	Our approach	
Source: Freiburg(Std A) cloudy 1 Target: Freiburg(Ext A) sunny 1	Risk (KL) = 0.08 Risk (EMD) = 0.01	72.3 ± 0.15 74.55 ± 0.11
Source: Freiburg(Std A) cloudy 1 Target: Freiburg(Ext A) cloudy 1	Risk (KL) = 0.02 Risk (EMD) = 0.003	78.1 ± 0.1 78.3 ± 0.05
Source: Freiburg(Std A) cloudy 1 Target: IDOL2 Minnie cloudy 2	Risk (KL) = 0.12 Risk (EMD) = 0.38	53.4 ± 0.2 56.1 ± 0.11
Source: Freiburg(Std A) cloudy 1 Target: Saarbrücken(Std A) cloudy 2	Risk (KL) = 0.05 Risk (EMD) = 0.001	50.19 ± 0.2 51.03 ± 0.11
Source: Freiburg(Std A) cloudy 1 Target: Saarbrücken(Ext A) night 1	Risk (KL) = 0.1 Risk (EMD) = 0.07	53.2 ± 0.1 52.13 ± 0.08
Source: Ljubljana(Std A) sunny 1 Target: Freiburg(Std A) cloudy 1	Risk (KL) = 0.6 Risk (EMD) = 0.57	79.2 ± 0.19 80.35 ± 0.12
Source: Ljubljana(Std A) sunny 1 Target: Saarbrücken(Ext A) night 1	Risk (KL) = 0.93 Risk (EMD) = 0.95	58.3 ± 0.16 58.63 ± 0.12
Source: VPC Home 1 Floor 2 Target: Freiburg(Std A) cloudy 1	Risk (KL) = 0.89 Risk (EMD) = 0.98	61.3 ± 0.09 59.15 ± 0.08
Source: Saarbrücken(Ext A) sunny 2 Target: Freiburg(Ext A) cloudy 1	Risk (KL) = 0.95 Risk (EMD) = 1.0	60.2 ± 0.09 58.2 ± 0.07

Table 2: Place recognition accuracy (%) obtained with SPACT features

	Our approach	
Source: Freiburg(Std A) cloudy 1 Target: Freiburg(Ext A) sunny 1	Risk (KL) = 0.07 Risk (EMD) = 0.012	68.02 ± 0.16 69.56 ± 0.09
Source: Freiburg(Std A) cloudy 1 Target: Freiburg(Ext A) cloudy 1	Risk (KL) = 0.03 Risk (EMD) = 0.002	62.04 ± 0.11 62.13 ± 0.09
Source: Freiburg(Std A) cloudy 1 Target: IDOL2 Minnie cloudy 2	Risk (KL) = 0.07 Risk (EMD) = 0.32	68.50 ± 0.25 73.12 ± 0.14
Source: Freiburg(Std A) cloudy 1 Target: Saarbrücken(Std A) cloudy 2	Risk (KL) = 0.04 Risk (EMD) = 0.006	57.65 ± 0.1 59.01 ± 0.08
Source: Freiburg(Std A) cloudy 1 Target: Saarbrücken(Ext A) night 1	Risk (KL) = 0.13 Risk (EMD) = 0.04	56.20 ± 0.12 54.35 ± 0.08
Source: Ljubljana(Std A) sunny 1 Target: Freiburg(Std A) cloudy 1	Risk (KL) = 0.55 Risk (EMD) = 0.49	59.30 ± 0.14 62.12 ± 0.11
Source: Ljubljana(Std A) sunny 1 Target: Saarbrücken(Ext A) night 1	Risk (KL) = 0.72 Risk (EMD) = 0.57	63.40 ± 0.13 66.12 ± 0.13
Source: VPC Home 1 Floor 2 Target: Freiburg(Std A) cloudy 1	Risk (KL) = 0.85 Risk (EMD) = 0.96	64.50 ± 0.08 61.34 ± 0.06
Source: Saarbrücken(Ext A) sunny 2 Target: Freiburg(Ext A) cloudy 1	Risk (KL) = 0.8 Risk (EMD) = 0.98	59,82 ± 0.05 55.4 ± 0.05

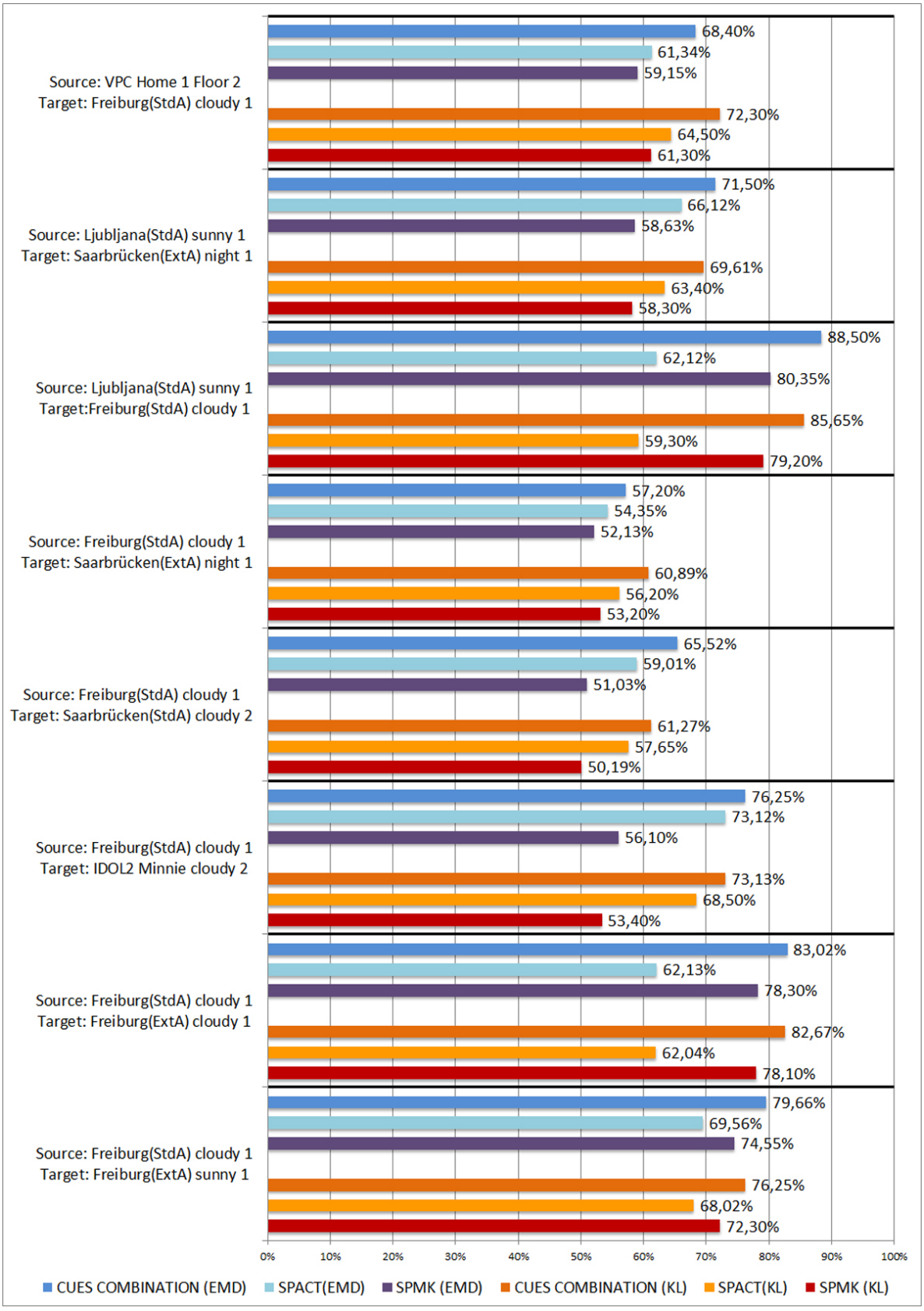


Figure 12: Cues combination results. The performance obtained with both KL and EMD transfer risk measures are reported.

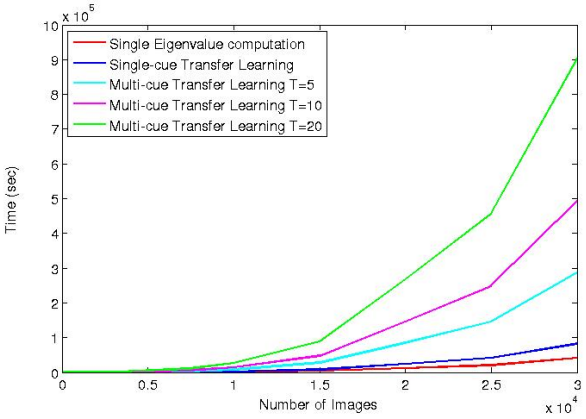


Figure 13: Computational time when running the proposed single-cue and multi-cue transfer learning approach.

and classification accuracy. Furthermore, as we do not require to compute all the eigenvalues of the kernel matrix, iterative methods can be applied to considerably reduce the computational time, making the problem feasible even for very large image sets.

5. Conclusions and Future Works

We have presented a novel approach for place recognition based on a risk sensitive transfer learning framework. We faced the challenging problem of domain adaptation when source and target data have different categories. This is meant to model our scenario of interest where the robot is moved in a new unknown location and it has only access to visual data collected in other environments. In this situation it is reasonable to assume that place categories in the current and in the past locations may differ significantly. In our approach a transfer risk measure is introduced to automatically quantify *how much* to transfer. It is based on the computation of the distance between the source and the target distributions. We compared the Kullback-Leibler divergence and the Earth Mover's Distance reaching similar performances in most of the tests. While other measures can be used to compute the distance among probability distributions, the proposed clustering based solutions represent a very computationally effective approach. Finally, we extended the proposed adaptation framework to merge multiple visual modalities, *i.e.* SPACT and SPMK, to benefit from different sources of information and further improve recognition accuracy.

Future works will include extending the proposed algorithm to operate in an incremental fashion and the development of a class-specific transfer risk measure to integrate into our learning framework. Finally the multi-cue approach can be modified to include more than two sets of features, *e.g.* adding depth sensor information.

6. Acknowledgments

This work has been partly supported by IIT funds under the project HARNESS coordinated by ENEA.

References

[1] A. Pronobis, Ó. M. Mozos, B. Caputo, P. Jensfelt, Multi-modal semantic place classification, International Journal of Robotics Research, IJRR 29 (2-3) (2010) 298–320.

[2] A. Pronobis, P. Jensfelt, Large-scale semantic mapping and reasoning with heterogeneous modalities, in: Proc. of the International Conference on Robotics and Automation, ICRA, 2012, pp. 3515–3522.

[3] J. Luo, A. Pronobis, B. Caputo, Svm-based transfer of visual knowledge across robotic platforms, in: Proc. of the International Conference on Computer Vision Systems, ICVS, 2007.

[4] S. P. Elango, T. Tommasi, B. Caputo, Transfer learning of visual concepts across robots: a discriminative approach, Tech. Rep. Idiap-RR-06-2012, Idiap (2012).

[5] Y. Rubner, C. Tomasi, L. J. Guibas, The earth mover's distance as a metric for image retrieval, International Journal of Computer Vision, IJCV 40 (2) (2000) 99–121.

[6] G. Costante, T. A. Ciarfuglia, P. Valigi, E. Ricci, A transfer learning approach for multi-cue semantic place recognition, in: Proc. of the International Conference on Intelligent Robots and Systems (IROS), 2013.

[7] M. Waibel, M. Beetz, J. Civera, R. D'Andrea, J. Elfring, D. Galvez-Lopez, K. Haussermann, R. Janssen, J. Montiel, A. Perzylo, B. Schiessle, M. Tenorth, O. Zweigle, R. van de Molengraft, Roboearth, IEEE Robotics Automation Magazine 18 (2) (2011) 69–82.

[8] T. Moritz, K. Ulrich, P. Dejan, B. Michael, Web-enabled robots – robots that use the web as an information resource, Robotics & Automation Magazine 18 (2) (2011) 58–68.

[9] M. Samadi, T. Kollar, M. Veloso, Using the web to interactively learn to find objects., in: Proc. of the International Conference on Artificial Intelligence, AAAI, 2012.

[10] M. Tenorth, K. Kamei, S. Satake, T. Miyashita, N. Hagita, Towards a networked robot architecture for distributed task execution and knowledge exchange, in: Proc. of the International Workshop on Standards and Common Platforms for Robot, SCPR, 2012.

[11] S. J. Pan, Q. Yang, A survey on transfer learning, IEEE Transactions on Knowledge and Data Engineering 22 (10) (2010) 1345 –1359.

[12] X. Shi, Q. Liu, W. Fan, P. Yu, Transfer across completely different feature spaces via spectral embedding.

[13] J. Hoffman, E. Rodner, J. Donahue, K. Saenko, T. Darrell, Efficient learning of domain-invariant image representations, arXiv preprint arXiv:1301.3224.

[14] M. Long, J. Wang, G. Ding, W. Cheng, X. Zhang, W. Wang, Dual transfer learning, in: Proc. of the International Conference on Data Mining (SDM), 2012, pp. 540–551.

[15] P. Dhillon, P. Talukdar, K. Crammer, Metric learning for graph-based domain adaptation, Tech. Rep. Technical Report No. MS-CIS-12-17, University of Pennsylvania Department of Computer and Information Science (2012).

[16] L. Duan, I. W. Tsang, D. Xu, T. S. Chua, Domain adaptation from multiple sources via auxiliary classifiers, in: International Conference on Machine Learning, 2009, pp. 289–296.

[17] B. Gong, Y. Shi, F. Sha, K. Grauman, Geodesic flow kernel for unsupervised domain adaptation, in: Proc. of the International Conference on Computer Vision and Pattern Recognition, CVPR, IEEE, 2012, pp. 2066–2073.

[18] E. Eaton, M. Desjardins, T. Lane, Modeling transfer relationships between learning tasks for improved inductive transfer, in: Proc. of the European Conference on Machine Learning and Knowledge Discovery in Databases, ECML PKDD, 2008, pp. 317–332.

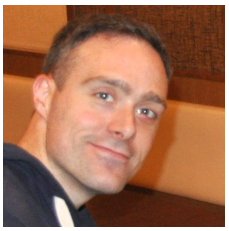
[19] X. Shi, W. Fan, J. Ren, Actively transfer domain knowledge, in: Proc. of the European conference on Machine Learning and Knowledge Discovery in Databases, ECML PKDD, 2008, pp. 342–357.

[20] A. Pronobis, B. Caputo, Confidence-based cue integration for visual place recognition, in: Proc. of the International Conference on Intelligent Robots and Systems, IROS, 2007.

[21] J. Luo, A. Pronobis, B. Caputo, P. Jensfelt, Incremental learning for place recognition in dynamic environments, in: Proc. of the International Conference on Intelligent Robots and Systems, IROS, 2007.
[22] A. Torralba, K. Murphy, W. Freeman, M. Rubin, Context-based vision system for place and object recognition, in: Proc. of the International Conference on Computer Vision, ICCV, 2003.
[23] M. M. Ullah, A. Pronobis, B. Caputo, J. Luo, P. Jensfelt, H. I. Christensen, Towards robust place recognition for robot localization, in: Proc. of the International Conference on Robotics and Automation, ICRA, 2008, pp. 530–537.
[24] Z. Kira, Inter-robot transfer learning for perceptual classification, in: Proc. of the International Conference on Autonomous Agents and Multi-agent Systems, AAMAS, 2010, pp. 13–20.
[25] X. Shi, W. Fan, Q. Yang, J. Ren, Relaxed transfer of different classes via spectral partition, in: Proc. of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, ECML-PKDD, 2009.
[26] J. Shi, J. Malik, Normalized cuts and image segmentation, IEEE Transaction on Pattern Analysis and Machine Intelligence, PAMI 22 (8) (2000) 888–905.
[27] F. Orabona, J. Luo, B. Caputo, Online-batch strongly convex multi kernel learning, in: proc. of the International Conference on Computer Vision and Pattern recognition, CVPR, 2010, pp. 787–794.
[28] S. Duffner, J. Odobez, E. Ricci, Dynamic partitioned sampling for tracking with discriminative features, in: Proc. of the British Machine Vision Conference, BMVC, 2009.
[29] S. Lazebnik, C. S. J., Ponce, Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories, in: Proc. of the International Conference on Computer Vision and Pattern Recognition, CVPR, 2006.
[30] J. Wu, J. M. Rehg, Centrist: A visual descriptor for scene categorization, IEEE Transaction on Pattern Analysis and Machine Intelligence, PAMI 33 (8) (2011) 1489–1501.
[31] D. G. Lowe, Object recognition from local scale-invariant features, in: Proc. of the International Conference on Computer Vision, ICCV, 1999.
[32] J. Wu, J. Rehg, Centrist: A visual descriptor for scene categorization, IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI 33 (8) (2011) 1489 –1501.
[33] A. Pronobis, B. Caputo, COLD: COsy Localization Database, International Journal of Robotics Research, IJRR 28 (5) (2009) 588–594.
[34] J. Wu, H. I. Christensen, J. M. Rehg, Visual place categorization: Problem, dataset, and algorithm, in: Proc. of the International Conference on Intelligent Robots and Systems, IROS, 2009, pp. 4763–4770.
[35] M. Wu, B. Schlkopf, A local learning approach for clustering, in: Proc. of the Interntional Conference on Neural Information Processing Systems, NIPS, 2006.
[36] C. H. Papadimitriou, K. Steiglitz, Combinatorial optimization: algorithms and complexity, Prentice-Hall, Inc., 1982.

Biographies

Gabriele Costante received the B.Sc. *magna cum laude* degree in Electronic and Information Engineering and the M.Sc. *magna cum laude* degree in Information and Automation Engineering from the University of Perugia respectively in 2010 and 2012. He then joined the Service and Industrial Robotics and Automation Laboratory (SIRALab) in 2012 and he is currently a Ph.D. student there. His research interests are mainly robotics, computer vision and machine learning.



Thomas A. Ciarfuglia received the M.Sc. *magna cum laude* degree in Electronics Engineering from the University of Perugia in 2004. He worked as HW/FW/SW designer engineer for various companies from 2004 to 2006. He then got an M.Sc. in Mechatronics and a Ph.D. degree in Robotics from the University of Perugia in 2008 and 2011 respectively. He joined the Service and Industrial Robotics and Automation Laboratory (SIRALab) in 2008 and he is currently working as a PostDoc there. His research interests are machine learning and computer vision applied to robotics.

Paolo Valigi received the Laurea degree in 1986 from University of Rome "La Sapienza" and the Ph.D. degree from University of Rome "Tor Vergata" in 1991. From 1990 to 1994 he worked with the Fondazione Ugo Bordoni. From 1998 to 2004 he was associate professor at the University of Perugia, Department of Electronics and Informatics Engineering, where since 2004 he has been full professor of System Theory and Optimization and Control. His research interests are in the field of robotics and systems biology. He has authored or co-authored more than 130 journal and conference papers and book chapters.

Elisa Ricci is an assistant professor at University of Perugia and a researcher at Fondazione Bruno Kessler. She received her Ph.D. from the University of Perugia in 2008. During her Ph.D. she was a visiting student at University of Bristol. After that she has been a post-doctoral researcher at Idiap, Martigny and the Fondazione Bruno Kessler, Trento. Her research interests are mainly in the areas of computer vision and machine learning.