Joint Probabilistic Matching Using m-Best Solutions

Seyed Hamid Rezatofighi[1] Anton Milan[1] Zhen Zhang[2] Qinfeng Shi[1] Anthony Dick[1] Ian Reid[1]
[1]School of Computer Science, The University of Adelaide, Australia
[2]School of Computer Science and Technology, Northwestern Polytechnical University, Xian, China
hamid.rezatofighi@adelaide.edu.au

Abstract

Matching between two sets of objects is typically approached by finding the object pairs that collectively maximize the joint matching score. In this paper, we argue that this single solution does not necessarily lead to the optimal matching accuracy and that general one-to-one assignment problems can be improved by considering multiple hypotheses before computing the final similarity measure. To that end, we propose to utilize the marginal distributions for each entity. Previously, this idea has been neglected mainly because exact marginalization is intractable due to a combinatorial number of all possible matching permutations. Here, we propose a generic approach to efficiently approximate the marginal distributions by exploiting the m-best solutions of the original problem. This approach not only improves the matching solution, but also provides more accurate ranking of the results, because of the extra information included in the marginal distribution. We validate our claim on two distinct objectives: (i) person re-identification and temporal matching modeled as an integer linear program, and (ii) feature point matching using a quadratic cost function. Our experiments confirm that marginalization indeed leads to superior performance compared to the single (nearly) optimal solution, yielding state-of-the-art results in both applications on standard benchmarks.

1. Introduction

Graph matching is a challenging problem that arises in many areas of computer vision including feature point matching [35], action recognition [3], multi-target tracking [31], and person re-identification (Re-ID) [26]. Whether the task is to find two different images that correspond to the same location, to associate each target to the correct measurement, or to identify the same person in two separate camera viewpoints, in its most general form it can be considered as a one-to-one assignment problem, where each element from one set should be uniquely assigned to another element in the second set. Typically, the criterion for assessing an assignment is given by a predefined distance or cost. This cost or objective is application specific. In the case of person Re-ID or multi-target tracking, for instance, the individual connections are assumed independent, leading to a linear optimization problem which can be solved optimally, *e.g.* using the Munkres (Hungarian) algorithm. When matching two sets of interest points, it is beneficial to include certain geometrical priors and consider pairwise terms within the objective function. However, solving arbitrary quadratic binary problems is NP-hard and one must resort to approximate solutions. Higher-order formulations are also possible, but are even harder to optimize.

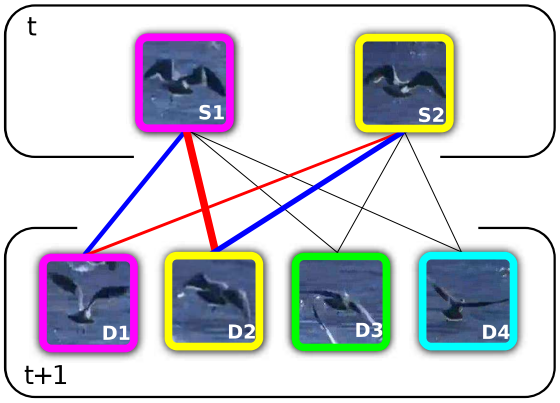


Figure 1: An example illustrating matching two objects in frame `t` to four objects in frame `t+1`. Each box color corresponds to a unique target ID, such that (S1,D1) and (S2,D2) are the correct matching pairs. However, the optimal assignment as determined by the highest joint matching score confuses the objects S1 with D2 and S2 with D1 due to their similar appearance. Here, the very strong visual similarity between S1 and D2 dominates the overall score, but ignores the fact that S2 is left without a suitable candidate. Our proposed approach makes a *collective* decision to compute the final matching criterion, leading to the correct match.

Interestingly, most existing work focuses on either designing a more suitable pairwise cost, *e.g.* by learning ap-

propriate features [17, 18, 20], or on developing better solvers [4, 44] in order to find a solution that is closer to the global optimum. What is mostly ignored, however, is the fact that the cost for each pairwise match is based solely on the two points being matched, and ignores the rest of the underlying distribution. Intuitively, this can be regarded as a "selfish" cost computation as it does not consider competing matches among other data points.

It may seem somewhat surprising, but even the globally optimal matching solution using this pairwise cost does not guarantee the best matching accuracy (see also [32]). To achieve best agreement among *all* candidates, it is desirable to follow a more "altruistic" approach and take into account all possible matching combinations for all other objects when calculating the distance (or similarity) between one particular pair of objects. This can be achieved by taking the maximum of the *marginal* distribution of the joint space for each object. An example illustrating this relationship is depicted in Fig. 2. Unfortunately, the joint matching space contains an exponential number of valid combinations, making exact marginalization intractable.

In this paper, we propose an accurate and efficient way for estimating the marginal distributions using only a tiny fraction of the entire space. To that end, we rely on the (approximate) m-best solutions. Empirically, this is sufficient to capture the majority of the mass within the joint probability distribution. We demonstrate our findings on the tasks of person re-identification, sequential target matching, and feature point matching. The first two tasks are reformulated as a binary linear program and solved using binary tree partitioning [27]. The latter involves a quadratic cost function such that only a local optimum can be achieved. Nonetheless, we show that a simple exclusion scheme is sufficient to obtain a good approximation of the underlying distribution and subsequently of the marginal distributions.

Building on existing methods, we outperform the reported results by a large margin with little computational overhead, achieving state-of-the-art performance on several tasks. It is interesting to note that marginalization not only increases the matching accuracy in the case of finding best pairs of feature points, but also improves the matching rank, as shown by examples of person re-identification.

2. Related Work

Graph matching is a fundamental problem in mathematics and computer science and has been explored in the context of various applications in computer vision [8, 35, 45]. In this section, we will review some of the approaches most relevant to our work.

Feature point matching is perhaps the most prominent example for graph matching in computer vision. It aims to find corresponding point pairs in two images, which can be used for estimating homographies in static scenes, or act as a pre-processing step for non-rigid structure from motion. In most cases, the problem is formulated as a quadratic assignment problem (QAP), which is known to be NP-hard. Therefore, the majority of the work is concerned with developing more efficient approaches to find a better approximation of the global optimum.

Early work by Gold and Rangarajan [11] combines sparsity and soft assignment constraints to escape local minima of the relaxed version of the problem. Leordeanu and Hebert [15] proposed a spectral technique to find an approximate solution to QAP. Edge weights on the matching graph model the likelihood that two objects correspond to one another. This method is highly efficient because the final assignment is recovered as the largest eigenvector of the adjacency matrix of the graph. It has been later extended to incorporate affine constraints [5], leading to better accuracy. Further, probabilistic formulations [4, 28, 36] as well as matrix factorization in combination with path-following algorithms for both undirected [45] and directed [44] graphs have been explored. Recently, hypergraph techniques [25, 34] have become popular for feature matching due to their ability to incorporate higher-order dependencies to capture the complexity of the problem more accurately.

Person re-identification, which aims to match people observed at different times by different cameras, is another example of a task that can be addressed via graph matching. Typically, a transform function that describes the photometric, geometric or other sort of transformation between cameras, needs to be established to find a pair of images that belong to the same individual. The main differences of various approaches lie in the features used and the learning algorithm. A comprehensive review and detailed discussions of many recent approaches can be found in [29].

Mignon and Jurie [24] propose a pairwise constrained component analysis (PCCA) specifically developed for dealing with high-dimensional input spaces. A projection into low-dimensional space yields good generalization while at the same time preserves desired pairwise constraints between data points. Li *et al*. [18] employ a deep neural network to learn filter pairs that encode the photometric transform between different views. Zheng *et al*. [43] learn a probabilistic relative distance comparison (PRDC) measure to discriminate true matches from wrong ones in a maximum likelihood framework. Zhao *et al*. [40] consider learning mid-level filters that strike a balance between generalization and discriminative power, without requiring tedious manual part annotations. Similarly, Liu *et al*. [20] investigate the importance of various features and propose an unsupervised learning approach to learn the corresponding weights. In their following work [21], a one-shot Post-rank optimization (POP) enables weak supervision to refine a result manually. A symmetric decision function for image pairs is learned in [19] and acts as both a distance metric and

an adaptive thresholding rule to classify the input pair as belonging to the same or to two different instances. Xiong *et al*. [30] provide a comprehensive analysis of various feature and metric learning methods. In a recent work, an effective distance learning based approach has been proposed to learn ranking of the assignments using the Cumulative Matching Characteristic (CMC) curve, a commonly used evaluation measure in person re-identification [26]. Somewhat related to our work is the network consistent re-identification (NCR) framework proposed by Das *et al*. [6]. A binary linear program is used to enforce consistent matching across the network, which at the same time improves pairwise re-identification performance between all camera pairs.

We have seen two examples of tasks that aim to find corresponding pairs in two disjoint sets. The majority of the feature point matching literature concerns the optimization, while re-identification focuses on learning better features. In this paper, we address both from a different direction. Building on any existing method, we are able to improve its performance by considering multiple solutions and using a modified similarity measure that arises from marginalization.

3. One-to-One Graph Matching

In one-to-one graph matching problems, the aim is to find the best unique match for a set of nodes, representing a set of features or objects indexed by $i \in [\mathcal{A}] = \{1, \cdots, M\}$ in a graph $\mathcal{G}_\mathcal{A}$, to another set of nodes indexed by $j \in [\mathcal{B}]_0 = \{0, 1, \cdots, N\}$ in graph $\mathcal{G}_\mathcal{B}$, using a joint matching probability $p(\cdot)$ or objective cost $f(\cdot)$. Here, 0 is a placeholder for a 'dummy' node in $\mathcal{G}_\mathcal{B}$ to allow solutions where a node from $\mathcal{G}_\mathcal{A}$ does not have a correspondence in $\mathcal{G}_\mathcal{B}$.

One-to-one matching can be represented by a bipartite graph[1] (Fig. 1) and forms a discrete combinatorial problem over the permutation space (Fig. 2). By definition, the one-to-one matching space $\mathcal{X}$ consists of all permutations where each node in $\mathcal{G}_\mathcal{A}$ is uniquely assigned to a node in $\mathcal{G}_\mathcal{B}$. This space can be defined by a set of binary vectors as follows:

$$\mathcal{X} = \Big\{ X = \left(x_i^j\right)_{i\in[\mathcal{A}],\, j\in[\mathcal{B}]_0} \;\Big|\; x_i^j \in \{0,1\}, \tag{1}$$

$$\forall j \in [\mathcal{B}] : \quad \sum_{i\in[\mathcal{A}]} x_i^j \leq 1, \tag{a}$$

$$\forall i \in [\mathcal{A}] : \quad \sum_{j\in[\mathcal{B}]_0} x_i^j = 1 \quad \Big\}, \tag{b}$$

where $X \in \mathcal{X} \subseteq \mathbb{B}^{M\times(N+1)}$ is a binary vector representing a possible solution to the entire matching problem and $x_i^j = 1$ means that node i is matched to node j. To model the situation that more than one object cannot find its counterpart, the uniqueness constraint does not hold for the dummy node, allowing multiple object-to-dummy assignments. Therefore, the one-to-one matching problem can be solved by maximizing the probability of this binary vector over the matching space $\mathcal{X}$ (MAP inference):

$$X^* = \underset{X\in\mathcal{X}}{\operatorname{argmax}} \quad p(X), \tag{2}$$

or equivalently by minimizing the binary objective:

$$X^* = \underset{X\in\mathcal{X}}{\operatorname{argmin}} \quad f(X), \tag{3}$$

where $p(X)$ or $f(X)$ represent a joint matching distribution or cost over the binary variables x_i^j, respectively.

The exact definition of $p(X)$ or $f(X)$ is application dependent and can in general take on any form. In some applications such as multi-target tracking [37] and person re-identification in surveillance cameras [6], $p(X)$ can be assumed to be statistically independent over each matching variable x_i^j, *i.e.* $p(X) \propto \prod_{\forall i,j} p(x_i^j)^{x_i^j}$. In this case, $f(X) = C^T X$ forms a binary linear program. In other problems like stereo matching [23] or iterative closest point algorithms [38], higher-order constraints, such as *e.g.* global geometric transformation consistency, can be employed.

Depending on the exact formulation of f, the globally optimal solution may or may not be easily achieved. Moreover, it turns out that in practice, even the optimal solution does not necessarily yield the correct matching assignment. This may sound counterintuitive, but we believe that this is mainly due to the dramatic simplification of the objective to unary or pairwise terms, which is necessary to keep the optimization tractable. Incorporating *all* possible combinations into the cost computation would exploit all available information. This naive approach, however, would lead to fully connected, high-order graphs that are impractical. In the next sections we first introduce a probabilistic view on the problem in the context of marginalizing the joint hypothesis space and motivate how marginalization leads to an improved matching probability (or matching cost) for a given problem. We then propose a framework to approximate the computationally prohibitive problem by considering only a fraction of the entire solution space and show its validity on challenging real-world applications.

4. Marginalizing the Joint Distribution

As discussed above, $p(X)$ is a joint distribution defined on the matching space $\mathcal{X}$. More formally, $p(\cdot)$ can be seen as M-dimensional, discrete distribution over the permutations space (Fig. 2), containing M^{N+1} elements, where the entry (i, j) corresponds to the probability (or a similarity score in the unnormalized case) that $i \in [\mathcal{A}]$ and $j \in [\mathcal{B}]_0$ should be matched. In real-world applications, $p(X)$ often

[1] Or, more generally, by a multipartite graph when more than two sets of nodes are involved.

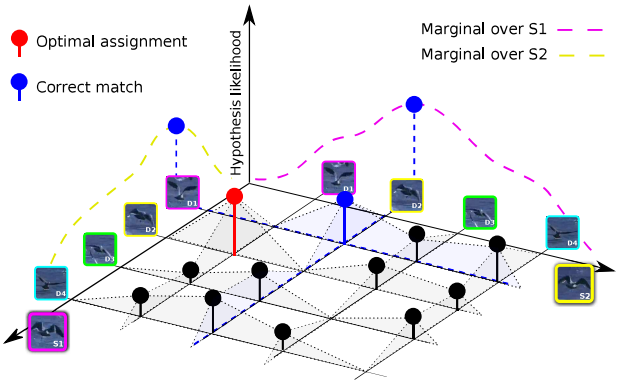


Figure 2: The joint hypothesis space from the matching example in Fig. 1. The marginals for the objects S1 and S2 (dashed) are computed by summing over the joint space, thereby gathering *all* relevant evidence to resolve ambiguities and produce "smoother" distributions with less noise.

contains numerous solutions that are close to the global optimum in terms of their objective value, for instance due to visual similarity or other ambiguities in the matching space. In other words, we usually have a number of competing solutions which are almost equally likely, but finally, MAP estimation forces us to pick only one of them as the sole winner. However, this choice is often not the correct matching configuration (*cf*. red and blue solutions in Fig. 2). Including other competing candidates provides valuable information for a final collective decision.

An alternative approach is thus to marginalize $p(X)$ over all matching solutions. If the value $p(X_k)$ for all $X_k \in \mathcal{X}$ is known, a marginalized probability distribution $\mathfrak{p}(x_i^j = 1)$ for assigning the node i in $\mathcal{G}_\mathcal{A}$ to the node j in $\mathcal{G}_\mathcal{B}$ is calculated by marginalizing $p(X)$ over all permutations that include $x_i^j = 1$:

$$\mathfrak{p}(x_i^j = 1) = \sum_{\{X \in \mathcal{X} | x_i^j = 1\}} p(X). \tag{4}$$

Similarly, a joint matching cost $\mathfrak{c}_i^j$ is obtained as

$$\mathfrak{c}_i^j = -\log \sum_{\{X \in \mathcal{X} | x_i^j = 1\}} e^{-f(X)}. \tag{5}$$

Let us briefly motivate this approach. From a probabilistic perspective, the marginalized distribution is computed using the *entire*, possibly highly complex, joint probability distribution, whereas the MAP estimate from Eq. (2) only returns one single value of that distribution, oblivious of the complexity of the problem. Another way of interpreting our approach is to consider it as a Bayesian approximation of the underlying distribution, as opposed to solely relying on the maximum a-posteriori estimation. In our example of matching, the marginals for one particular object contain *all* possible matching permutations and thus encode all the relevant information required to untangle potential ambiguities.

Another advantage of marginalization is its averaging or "smoothing" property that arises from summation (or integration in the continuous case). This typically leads to a less noisy approximation of the original joint space, which in turn means a more informative distribution with fewer matching ambiguities. This property can be directly exploited to extract a more reliable candidate ranking for matching. Finally, it is important to note that, under the assumption that the original objective is linked to the true matching accuracy, *i.e*. that a lower cost generally corresponds to a better result, the max-marginal solution will never be worse than the MAP estimate of the original cost. If the joint distribution is unimodal or contains one single strong peak, then marginalization will yield the exact same solution as the MAP one. However, the exact shape of the joint distribution is typically not known a priori, but is indeed rather complex in real-world applications. Hence, we argue that relying on the marginal distribution is always the safer choice.[2]

Despite the obvious benefits of marginalization, to the best of our knowledge it has not yet been applied to matching-related problems. We believe that this is mainly due to the computational complexity required to obtain the marginal distributions over a complex joint hypothesis space. In the following, we present a well-founded approach to approximate it by considering not all, but only few strongest matching hypotheses.

4.1. Approximation Using m-Best Solutions

The marginalized distribution $\mathfrak{p}(x_i^j = 1)$ (or the cost $\mathfrak{c}_i^j$) can be approximated by considering a fraction of the entire matching space that includes the m-highest joint probabilities $p(X)$ (or the m-lowest values for $f(X)$). Therefore, $\mathfrak{p}(x_i^j = 1)$ and $\mathfrak{c}_i^j$ are respectively approximated by

$$\mathfrak{p}(x_i^j = 1) \approx \sum_{\{X_k^* | \forall k \in [m], x_i^j = 1\}} p(X_k^*), \tag{6}$$

and

$$\mathfrak{c}_i^j \approx -\log \sum_{\{X_k^* | \forall k \in [m], x_i^j = 1\}} e^{-f(X_k^*)}, \tag{7}$$

where X_k^* is the k-th optimal solution for Eq. (2) and (3), respectively. Note that approximating $\mathfrak{p}$ with only one solution ($m = 1$) will yield the exact same final result as the MAP estimate. By increasing m, we collect more and more important samples from the joint distributions, which

[2]The marginalization in this case resembles Bayesian estimates of a distribution (with uniform prior) which has superior performance to MAP inference in the case of a multi-modal distribution.

improves upon this approximation and consequently the matching results[3]. It remains to clarify how we compute the next-best solution, given the current one.

4.2. Implementation Details

To calculate the m-th solution of Eq. (2) or (3), we follow one of two approaches: The first is a naive method that is most general and can thus be applied to any arbitrary problem and used in conjunction with most available solvers. The second is a more sophisticated approach recently introduced in [27], only applicable to a certain class of problems.

Naive exclusion strategy. Let us assume that we can obtain one (possibly globally) optimal solution X_k^* (with k=1) to the binary minimization problem in Eq. (3) respecting the constraints in Eq. (1), which corresponds to the MAP estimate. The most straightforward approach to find the next best solution is to exclude X_k^* from the search space and to rerun the optimization. This can be done by introducing an additional inequality constraint $\langle X, X_k^* \rangle \leq \|X_k^*\|_1 - 1$. Since X and X_k^* are binary, the above inequality holds if and only if $X \neq X_k^*$. This procedure can be repeated iteratively for $k = 2, \ldots, m$ to obtain a more accurate approximation of the joint distribution. This approach is very general and can be applied to any solver that can handle linear constraints. However, the number of inequality constraints increases with every iteration, which may render this strategy impractical for large values of m.

Binary Tree Partitioning. If the optimization problem in Eq. (3) can be solved optimally (or at least near optimally), the binary tree partitioning (BTP) approach introduced in [27] should be used instead. Here, redundant constraints are removed and the objective is found as a series of second-best solutions[3]. While this strategy turns out to be more efficient than the naive approach, it cannot be applied to arbitrary problems, especially when the found solution is far from the global optimum.

In this work, we rely on BTP for approximating the joint distribution of a linear objective for person re-identification, and in conjunction with a belief propagation approach for quadratic programs [39] for feature matching, as it tends to find strong optima. However, we turn to the naive approach in the context of spectral matching [15], where the quality of returned solutions is not suitable for BTP.

5. Experimental Results

To validate the strengths and generality of our approach, we perform experiments on three separate applications. First, we show how marginalization improves the ranking measure of an already globally optimal solution in the context of person re-identification (Re-ID) across camera viewpoints. Second, we show a surprising result of tracking multiple targets in challenging sequences without relying on any dynamic cues. Finally, we present our method on the application of feature point matching with a quadratic objective function.

5.1. Person Re-Identification

Datasets. We demonstrate our proposed framework on the task person re-identification on a variety of challenging public benchmarks that are commonly used in literature: RAiD [6], WARD[22], iLIDS [42], 3DPES [1], VIPeR [12], CUHK01 [17] and CUHK03 [10]. Each dataset contains pairs of images belonging to the same individual, with the number of pairs ranging from 20 to 485.

Implementation and evaluation. We employ the most commonly used cumulative matching characteristic (CMC) criterion [12] as the performance measure for evaluating person Re-ID. This evaluation measure reports the recognition rate at different ranking scores.

We achieve state-of-the-art results on all datasets, based on the assignment costs of the best reported results to date. In particular, for the RAiD and WARD datasets, we use the similarity scores from learned feature transformation (FT) and the same evaluation protocol used in [6]. For the other datasets such as iLIDS, 3DPES, VIPeR, CUHK01 and CUHK03, we use the same visual features used in [26] including SIFT/LAB, LBP/RGB, Region covariance and CNN descriptors, all of which are weighted uniformly to compute the overall similarity cost. This approach has been reported as the baseline method (Avg. Feature) in [26]. Again, for consistency, we follow the exact same experimental setup and evaluation protocol.

To apply the concept of marginalization, we first reformulate the assignment problem as the following binary linear program (BLP)[3]

$$X^* = \underset{X \in \mathcal{X}}{\operatorname{argmin}} \quad f(X) = C^\top X, \tag{8}$$

subject to constraints in Eq. (1), where C is the assignment cost. The binary problem in Eq. (8) can be solved optimally using LP relaxation [13]. To this end, we use the efficient binary tree partitioning approach to calculate the m-best solutions of the joint probability distribution. For all datasets, we chose $m = 100$, which empirically is enough to approximate the marginalized distribution. In Table 1, we report the average processing times of our proposed method for each dataset. Our code is implemented in MATLAB and the experiments were carried out on a standard desktop PC (Intel Core $i7 - 4790$, 3.60 GHz CPU with 16 GB RAM).

[3] Please refer to the supplemental material for more details.

Dataset (size)	Method	Recognition rate %			Time
		Rank-1	Rank-2	Rank-5	(Sec.)
RAiD (20×20)	FT [6]	74.0	82.0	96.0	
	mbst-FT	**85.0**	**99.0**	**100.0**	1.6
WARD (35×35)	FT [6]	50.3	70.9	88.0	
	mbst-FT	**72.0**	**81.1**	**92.6**	4.2
iLIDS (59×59)	AvgF [26]	51.9	60.7	72.4	
	mbst-AvgF	**54.7**	**63.6**	**75.4**	15.4
3DPeS (96×96)	AvgF [26]	53.6	64.1	76.9	
	mbst-AvgF	**57.5**	**67.9**	**79.5**	31.8
VIPeR (316×316)	AvgF [26]	44.9	58.3	76.3	
	mbst-AvgF	**50.5**	**63.0**	**78.0**	201.9
CUHK01 (485×485)	AvgF [26]	51.9	63.3	75.1	
	mbst-AvgF	**62.8**	**70.9**	**78.8**	485.6
CUHK03 (100×100)	AvgF [26]	57.4	71.7	85.9	
	mbst-AvgF	**74.2**	**83.1**	**90.7**	33.5

Table 1: Re-ID recognition rate on public datasets at different ranks. For each dataset, we show the reported state-of-the-art result and the improvement due to marginalization on the original cost using $m{=}100$ best solutions as an approximation. The average processing time of our method on each dataset is reported on the right.

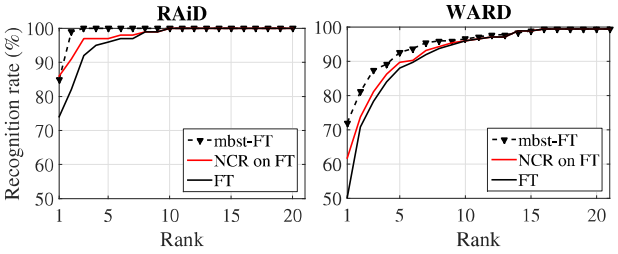


Figure 3: Comparison between CMC curves of the original cost matrix (FT) [6], the network consistent re-identification (NCR) on FT [6], which is the best known result on these datasets, and our proposed m-best marginalization (for $m{=}100$) on FT for RAiD (camera pairs 1-2) and WARD (camera pairs 2-3).

Further we employed the Gurobi ILP solver (version 5.6.3, 64bit).

Results. Table 1 lists the recognition rates at different ranks (1, 2 and 5) for state-of-the-art results using the baseline assignment costs [6, 26] and after applying our m-best marginalization approach. The results show consistent and significant improvements with respect to the original cost for all recognition rates on all datasets reaching 100% at rank 3 for RAiD.

Figures 3 and 4 show CMC curves on four exemplar datasets. Our results are plotted with a dashed black line, while the original assignment cost we used is rendered with a black solid line. Note that marginalization yields superior results when applied to the original cost compared to a network consistency constraint (NCR) [6] or a sophisticated feature learning approach (CMC^{top}) [26]. Fig. 4 additionally shows further recent results including simple metric learning (KISSME) [14], salience matching [41], learned mid-level filters [40] and a deep learning approach (FPNN) [18].

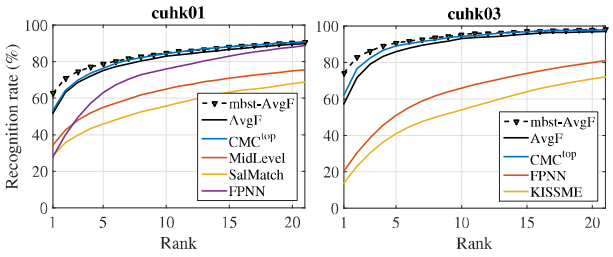


Figure 4: Comparison to state-of-the-art results on two challenging datasets. Our m-best approach (dashed black curve) is computed based on the Average Features (AvgF) baseline used in [26].

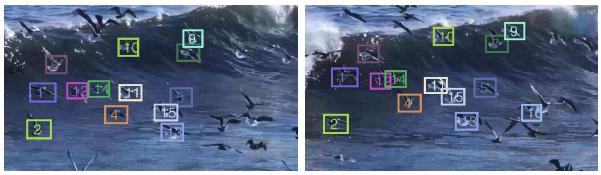


Figure 5: The results of our sequential re-identification on frames 60 and 80 of the *Seagulls* sequence. Note that target IDs are maintained merely by frame-wise visual matching, without any dynamic model.

5.2. Sequential Re-Identification.

To further challenge our proposed method, we test it on a new problem we call sequential re-identification. Here, the aim is to successfully match visually similar objects in a video sequence by considering their appearance only and without using a motion prior. In comparison with the classical Re-ID problem, all objects are captured from a single camera and their appearance does not change much from one time frame to another. However, the main difficulty is that all objects exhibit strong visual similarity and without considering motion, tracking becomes extremely challenging, even for human observers. This application is relevant for low-framerate surveillance videos, where motion information is unreliable.

To demonstrate the robustness of marginalization for visual matching, we employ several video sequences with visually similar objects used in [7]. In their work, Dicle *et al.* [7] argue that tracking multiple similarly looking objects can only be achieved by exploiting the target dynamics.

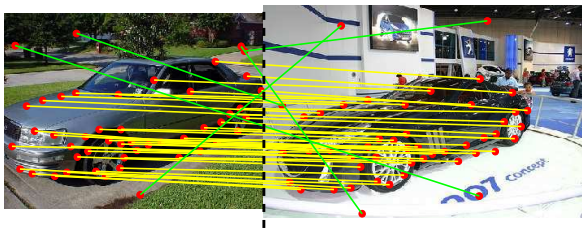


Figure 6: An example of our matching results using mbst-BP for the car dataset. Yellow lines indicate correct matches and green lines matches between outliers. There are no incorrect matches in this example.

Here, we show that even in this challenging scenario, a simple frame-by-frame matching of visual features using the marginalized score and without relying on motion achieves compatible, or even superior results.

Figures 5 shows a visual example of the temporal objects matching. Note how visually similar targets are matched correctly over long time periods. To quantify the performance, we compute the common CLEAR [2] metric *MOTA* that consists of identity switches, false positives and false negatives. Surprisingly, we achieve 5% higher MOTA score on average compared to the results of [7]. We refer the reader to the supplemental material for more details and further comparisons.

5.3. Feature Matching

As the third application, we apply our approach on the task of feature point matching which is typically formulated as a quadratic assignment problem [16]:

$$X^* = \underset{X \in \mathcal{X}}{\operatorname{argmax}} \quad J(X) = X^\top K X, \tag{9}$$

where K is a global affinity matrix. Therefore, $J(X)$ is a quadratic similarity score. Note that we can obtain an unnormalized version of the joint distribution $p(\cdot)$ from $J(\cdot)$ as $p(\cdot) \propto e^{J(\cdot)}$.

Dataset and implementation. For this application, we use the popular car and motorbikes dataset for feature point matching from [16]. This dataset consists of 30 pairs of images of cars and 20 pairs of images of motorbikes from the PASCAL VOC 2007 challenge [9] (*cf*. Fig. 6). Each image pair contains 30-60 ground-truth correspondences between pairs of interest points, followed by several outlier points. For the experimental setup and evaluation, we applied the script used in [44] to reproduce the same set of experiments, following the exact same procedure to randomly select $0 - 20$ outliers for each image pair.

To support our claim that marginalizing using m-best solutions helps to improve performance in this application, we chose two independent state-of-the-art feature matching algorithms: Integer Projected Fixed Point (IPFP) [16] and a customized Belief Propagation (BP) solver [39], specifically developed for matching problems. We applied the naive exclusion approach to the former, and the binary tree partitioning (BTP) to BP to calculate their m-best solutions, respectively. The reasoning behind this choice is as follows: The solutions found by IPFP are often rather far from the global optimum. This may have an undesirable effect when used with BTP because the search may be guided towards the wrong portion of the solution space. BP, on the other hand, returns near optimal results at each iteration. We found in our experiments that the binary tree partitioning yields better results in this case, despite the fact that global optimality cannot be guaranteed.

Results. We report the averaged matching accuracy over all image pairs in each experiment with same number of outliers and compare our results against the best solution from two aforementioned solvers. Furthermore, we show results from eight other competitive feature matching algorithms including Graduated Assignment (GA) [11], Probabilistic Matching (PM) [36], Spectral Matching (SM) [15], Spectral Matching with Affine Constraints (SMAC) [5], Reweighted Random Walks Matching (RRWM) [4], Sequential Monte Carlo Matching (SMCM) [28], Factorized Graph Matching (FGM-D) [44] and HyperGraph Matching [25].

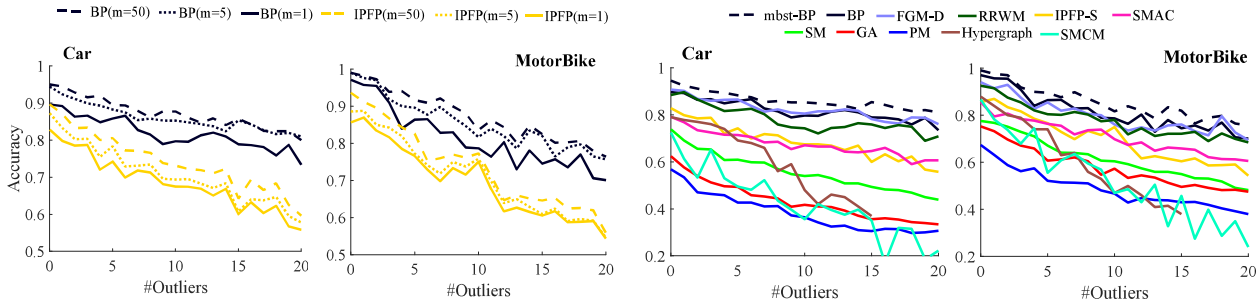
Fig. 7 (a) shows the matching results using the IPFP and BP solvers and their marginalization results over m-best solutions for two different values of m ($m = 5$ and $m = 50$). We see that marginalization consistently improves both solvers' matching accuracy. Moreover, increasing the number of solutions does improve the accuracy as well, which, of course, comes at the cost of higher processing time. We also show that using only few m-best solutions ($m = 5$) of BP, we can achieve sate-of-the-art results in this application (Fig. 7 (b)). In Table 2, we also report the matching accuracy results and processing time[4] of all approaches for the experiments without any outliers[5] .

6. Discussion and Limitations

We have seen in the previous sections that marginalization improves one-to-one matching in real-world applications. Here, we discuss further findings and point to some known limitations of this approach.

First, we examine the diversity of the m solution illustrated in Fig. 8 by their pairwise Hamming distance. In the first example, the matching accuracy increases when a new part of the solution space is discovered after about 40 iterations. However, the example on the right does not show a particular structure. Overall, we found empirically that there is no apparent correlation between the similarity of

[4]Time is reported using the same hardware configuration as in Sec. 5.1.
[5]The results for Hypergraph method are extracted from the paper [25].



(a) (b)

Figure 7: (a): Matching accuracy for IPFP (yellow) and BP (black) using different values of m ($m = 1, 5, 50$). (b): Comparison between different state-of-the-art matching approaches and mbst-BP for $m = 5$ (black dashed).

	GA	PM	SM	SMAC	SMCM	Hyper Graph	RRWM	FGM-D	IPFP-S			BP		
									$m=1$	$m=5$	$m=50$	$m=1$	$m=5$	$m=10$
Car (Acc)	0.62	0.57	0.74	0.79	0.72	0.79	0.88	0.91	0.83	0.88	0.90	0.92	0.94	**0.95**
Car (Time)	0.02	0.02	0.06	0.05	1.75	-	0.75	10.1	0.02	0.37	9.87	1.02	38.1	87.2
Motor (Acc)	0.75	0.67	0.78	0.86	0.87	0.88	0.93	0.94	0.86	0.89	0.94	0.97	**0.99**	**0.99**
Motor (Time)	0.02	0.02	0.08	0.06	4.64	-	0.80	9.20	0.02	0.25	8.70	0.25	36.3	70.9

Table 2: Matching accuracy and processing times on car and motorbike datasets for the experiment without outliers.

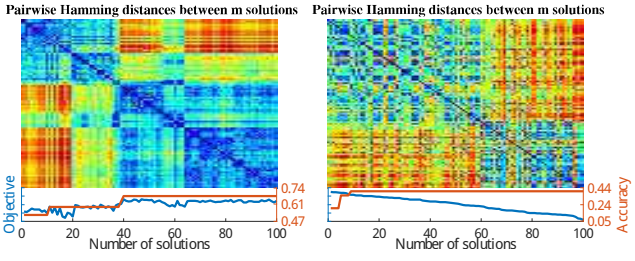


Figure 8: Pairwise Hamming distance between m solutions calculated using IPFP (left) and BP (right) solvers for two feature points matching examples. Objective values (blue) and matching accuracy (red) using marginalization for each solution is shown on the bottom.

discovered solutions and their contribution towards approximating the joint probability distribution.

Second, we would like to draw attention to the optimality of the found solutions. When dealing with unary potentials only, matching can be reformulated as a binary linear program with an assignment matrix A, which can be solved exactly. However, a solution of a quadratic program does not guarantee the global optimum. We have observed in our experiments, that by simply iterating through successive solutions, both with the naive exclusion approach and with binary tree partitioning, a solution with a better objective value can sometimes be found, which in turn may show a higher matching accuracy (*cf*. Fig. 8 (left)). Nonetheless, this is not always the case and we have found that an additional marginalization step usually leads to the best outcome (see Fig. 8 (right)).

One limitation of marginalizing the matching space is that the one-to-one constraint is no longer guaranteed. This shortcoming can still be resolved by an additional bipartite matching step (*e.g.* using the Hungarian algorithm) on the newly computed distribution. However, empirically, violation of the constraint is rare and we prefer to report the result based on the raw output of our approach.

Finally, obtaining the (approximated) marginal distribution necessarily requires a computational overhead with respect to the original problem. We believe that this is an acceptable trade-off for a significant gain in accuracy.

7. Conclusion

We have presented a novel approach to graph matching. Instead of focusing on optimization or feature learning, we consider the approximate marginal distributions of the joint hypothesis space. Our method is generic and can be applied to any existing technique. The experimental results show a clear benefit to our approach, yielding state-of-the-art performance on several exemplar tasks. Evidently, considering the m-best solutions leads to higher accuracy in one-to-one matching, as well as improving match ranking in the re-identification case. In future, we plan to explore further applications with arbitrary cost functions.

Acknowledgments. This work was supported by ARC Linkage Project LP130100154, ARC discovery projects DP140102270 and DP160100703, and ARC Laureate Fellowship FL130100102.

References

[1] D. Baltieri, R. Vezzani, and R. Cucchiara. 3dpes: 3d people dataset for surveillance and forensics. In *Proc. of Intl. Workshop on Mult. Acc. to 3D Human Objs.,*, pages 59–64, 2011. 5

[2] K. Bernardin and R. Stiefelhagen. Evaluating multiple object tracking performance: The CLEAR MOT metrics. *Image and Video Processing*, 2008(1):1–10, May 2008. 7

[3] W. Brendel and S. Todorovic. Learning spatiotemporal graphs of human activities. In *ICCV 2011*. 1

[4] M. Cho, J. Lee, and K. M. Lee. Reweighted random walks for graph matching. In *ECCV 2010*, pages 492–505. 2, 7

[5] T. Cour, P. Srinivasan, and J. Shi. Balanced graph matching. In *NIPS*2007*, pages 313–320. 2, 7

[6] A. Das, A. Chakraborty, and A. K. Roy-Chowdhury. Consistent re-identification in a camera network. In *ECCV 2014*, pages 330–345. 3, 5, 6

[7] C. Dicle, O. I. Camps, and M. Sznaier. The way they move: Tracking multiple targets with similar appearance. In *ICCV*, pages 2304–2311, 2013. 6, 7

[8] O. Duchenne, A. Joulin, and J. Ponce. A graph-matching kernel for object categorization. In *ICCV 2011*. 2

[9] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. *The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results*. 7

[10] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. *IEEE Trans. Pattern Anal. Mach. Intell. (PAMI)*, 32(9):1627–1645, 2010. 5

[11] S. Gold and A. Rangarajan. A graduated assignment algorithm for graph matching. *IEEE TPAMI*, 18(4):377–388, Apr. 1996. 2, 7

[12] D. Gray, S. Brennan, and H. Tao. Evaluating appearance models for recognition, reacquisition, and tracking. In *Proc. Intl. Workshop on Perf. Eval. of Track. and Survl.*, volume 3, 2007. 5

[13] I. Heller and C. B. Tompkins. An extension of a theorem of dantzig. *Annals of Mathematics Studies.*, 38(1), 1956. 5

[14] M. Kostinger, M. Hirzer, P. Wohlhart, P. Roth, and H. Bischof. Large scale metric learning from equivalence constraints. In *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2288–2295, June 2012. 6

[15] M. Leordeanu and M. Hebert. A spectral technique for correspondence problems using pairwise constraints. In *ICCV 2005*. 2, 5, 7

[16] M. Leordeanu, R. Sukthankar, and M. Hebert. Unsupervised learning for graph matching. *IJCV*, 96(1):28–45, Apr. 2011. 7

[17] W. Li, R. Zhao, and X. Wang. Human reidentification with transferred metric learning. In *ACCV*, pages 31–44, 2013. 2, 5

[18] W. Li, R. Zhao, T. Xiao, and X. Wang. DeepReID: Deep filter pairing neural network for person re-identification. In *CVPR 2014*. 2, 6

[19] Z. Li, S. Chang, F. Liang, T. Huang, L. Cao, and J. Smith. Learning locally-adaptive decision functions for person verification. In *CVPR 2013*. 2

[20] C. Liu, S. Gong, C. C. Loy, and X. Lin. Person re-identification: What features are important? In *ECCV 2010*, pages 391–401, 2012. 2

[21] C. Liu, C. Loy, S. Gong, and G. Wang. Pop: Person re-identification post-rank optimisation. In *ICCV 2013*. 2

[22] N. Martinel and C. Micheloni. Re-identify people in wide area camera network. In *CVPR Workshops*, pages 31–36, 2012. 5

[23] T. Meltzer, C. Yanover, and Y. Weiss. Globally optimal solutions for energy minimization in stereo vision using reweighted belief propagation. In *ICCV 2005*, pages 428–435. 3

[24] A. Mignon and F. Jurie. PCCA: A new approach for distance learning from sparse pairwise constraints. In *CVPR 2012*. 2

[25] Q. Nguyen, A. Gautier, and M. Hein. A flexible tensor block coordinate ascent scheme for hypergraph matching. In *CVPR*, pages 5270–5278, 2015. 2, 7

[26] S. Paisitkriangkrai, C. Shen, and A. v. d. Hengel. Learning to rank in person re-identification with metric ensembles. In *CVPR*, 2015. 1, 3, 5, 6

[27] H. S. Rezatofighi, A. Milan, Z. Zhang, Q. Shi, A. Dick, and I. Reid. Joint probabilistic data association revisited. In *ICCV 2015*. 2, 5

[28] Y. Suh, M. Cho, and K. M. Lee. Graph matching via sequential monte carlo. In *ECCV*, pages 624–637, 2012. 2, 7

[29] X. Wang and R. Zhao. Person re-identification: System design and evaluation overview. In *Person Re-Identification*, pages 351–370. Springer London, 2014. 2

[30] F. Xiong, M. Gou, O. Camps, and M. Sznaier. Person re-identification using kernel-based metric learning methods. In *ECCV 2014*. 3

[31] H. Xiong, D. Zheng, Q. Zhu, B. Wang, and Y. Zheng. A structured learning-based graph matching method for tracking dynamic multiple objects. *IEEE Transactions on Circuits and Systems for Video Technology*, 23(3):534–548, Mar. 2013. 1

[32] J. Yan, M. Cho, H. Zha, X. Yang, and S. Chu. Multi-graph matching via affinity optimization with graduated consistency regularization. *IEEE Trans. Pattern Anal. Mach. Intell. (PAMI)*, 2015. 2

[33] J. Yan, Y. Li, W. Liu, H. Zha, X. Yang, and S. M. Chu. Graduated consistency-regularized optimization for multi-graph matching. In *ECCV*, pages 407–422, 2014.

[34] J. Yan, C. Zhang, H. Zha, W. Liu, X. Yang, and S. M. Chu. Discrete hyper-graph matching. In *CVPR*, pages 1520–1528, 2015. 2

[35] A. Zamir and M. Shah. Image geo-localization based on multiple nearest neighbor feature matching using generalized graphs. *IEEE TPAMI*, 36(8):1546–1558, Aug. 2014. 1, 2

[36] R. Zass and A. Shashua. Probabilistic graph and hypergraph matching. In *CVPR 2008*. 2, 7

[37] L. Zhang, Y. Li, and R. Nevatia. Global data association for multi-object tracking using network flows. In *CVPR 2008*. 3

[38] Z. Zhang. Iterative point matching for registration of free-form curves and surfaces. *IJCV*, 13(2):119–152, Oct. 1994. 3

[39] Z. Zhang, Q. Shi, J. McAuley, W. Wei, Y. Zhang, and A. van den Hengel. Pairwise matching through Max-Weight bipartite belief propagation. In *CVPR 2016*. 5, 7

[40] R. Zhao, W. Ouyang, and X. Wang. Learning mid-level filters for person re-identfiation. In *CVPR 2014*. 2, 6

[41] R. Zhao, W. Ouyang, and X. Wang. Person re-identification by salience matching. In *ICCV 2013*. 6

[42] W.-S. Zheng, S. Gong, and T. Xiang. Associating groups of people. In *BMVC*, volume 2, page 6, 2009. 5

[43] W.-S. Zheng, S. Gong, and T. Xiang. Reidentification by relative distance comparison. *IEEE TPAMI*, 35(3):653–668, Mar. 2013. 2

[44] F. Zhou and F. De la Torre. Deformable graph matching. In *CVPR 2013*. 2, 7

[45] F. Zhou and F. De la Torre. Factorized graph matching. In *CVPR 2012*. 2