Supplementary Materials – Explicit Partial Cycle Consistency in Partial Multi-View Matching

Fedor Taggenbrock[1,2][0009−0002−6166−0865], Gertjan Burghouts[2][0000−0001−6265−7276], and Ronald Poppe[1][0000−0002−0843−7878]

[1] Utrecht University, Utrecht, Netherlands
[2] TNO, The Hague, Netherlands

This document serves as the supplementary material to "Self-Supervised Cycle-Consistency for Multi-View Matching". Our theoretical contribution is outlined in Sections 1 and 2. Section 3 provides additional results which did not fit in the original paper.

1 Partial Multi-View Matching Setting

We analyze the Multi-View-Matching setting in greater detail, with an emphasis on partial overlap.
Consider N viewpoints containing different feature representations of the same set of real world objects. Specifically, each $V_i = \{f_1^i, \dots f_{n_i}^i\},\ i \in \{1, \dots, N\}$, contains n_i, D-dimensional features. The features are representations of some real world objects in the set $U = \{o_1, \dots o_{|U|}\}$, so that $n_i \leq |U|,\ \forall i \in \{1, \dots, N\}$. In our case, the features f^i are obtained through inference with some trained feature network Φ, applied to the detections seen from view V_i. Each detection corresponds to an object in U, the unique people in an underlying scene.

Given the features in V_i and V_j, we can compute pairwise similarity matrices S_{ij} using the cosine similarities between features. Given N views, this provides a multi-view similarity matrix $\mathbf{S} \in \mathbb{R}^{\sum_{i=1}^N n_i \times \sum_{i=1}^N n_i}$:

$$\mathbf{S} = \begin{bmatrix} S_{11} & S_{12} & \dots & S_{1N} \\ S_{21} & S_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ S_{N1} & \dots & \dots & S_{NN} \end{bmatrix}. \tag{1}$$

The goal is to obtain the optimal partial multi-view matching matrix

$$\mathbf{P}^* \in \{0,1\}^{\sum_{i=1}^N n_i \times \sum_{i=1}^N n_i} :$$

$$\mathbf{P}^* = \begin{bmatrix} P_{11}^* & P_{12}^* & \dots & P_{1N}^* \\ P_{21}^* & P_{22}^* & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ P_{N1}^* & \dots & \dots & P_{NN}^* \end{bmatrix}, \tag{2}$$

with $P^*_{ij} \in \mathcal{P}$, being partial permutation matrices:

$$P_{ij} \in \mathcal{P} \iff \begin{cases} P_{ij} \in \{0,1\}^{n_i \times n_j}, \\ P_{ij} \cdot \mathbf{1}_{n_j} \leq \mathbf{1}_{n_i}, \text{ and } P_{ij}^T \cdot \mathbf{1}_{n_i} \leq \mathbf{1}_{n_j}. \end{cases} \tag{3}$$

This means that each row in P_{ij} can be matched to one or none of the columns in P_{ij} and the other way around. The optimal $\mathbf{P}^*$ combines the optimal P_{ij} that do not break the cycle-consistency constraint together. The optimal matchings are made given the similarities $\mathbf{S}$ and some parameter s_{min} to quantify the amount of overlap. This means we are dealing with the following optimization problem:

Definition 1 (Partial multi-view matching problem).
Given a multi-view similarity matrix $\mathbf{S} \in \mathbb{R}^{\sum_{i=1}^N n_i \times \sum_{i=1}^N n_i}$*, we obtain* $\bar{\mathbf{S}} \in \mathbb{R}^{\sum_{i=1}^N n_i \times \sum_{i=1}^N n_i}$ *by setting all elements in* $\mathbf{S}$ *lower than* s_{min} *to some negative value. The optimal multi-view matching matrix* $\mathbf{P}^*$ *is then given by:*

$$\mathbf{P}^* = argmax_{\mathbf{P} \in \mathbb{P}} \langle \mathbf{P}, \bar{\mathbf{S}} \rangle_F, \tag{4}$$

where $\langle \cdot, \cdot \rangle_F$ *denotes the matrix dot product, and* $\mathbb{P}$ *is the set of globally cycle-consistent partial multi-view permutations matrices.*

An optimal unconstrained P^*_{ij} would be given by

$$P^*_{ij} = \operatorname{argmax}_{P_{ij} \in \mathcal{P}} \langle P_{ij}, \bar{S_{ij}} \rangle_F, \tag{5}$$

namely the partial matching in $\mathcal{P}$ for which the total sum of matched similarities is maximal, with the negative elements in $\bar{S_{ij}}$ creating non-matches. Maximizing Equation 4 can thus be understood as optimizing the individual pairwise matchings under the global cycle-consistency constraint.
The optimization problem from Definition 1 is NP-Hard to solve, due to the combinatorial nature of the cycle consistency constraint $\mathbf{P} \in \mathbb{P}$. Theoretical works provide approximate algorithmic solutions, often using some relaxation of the cycle-consistency constraint [4,5]. The fastest and simplest solution is to just obtain the optimal pairwise matchings from Equation 5 using the well known Hungarian algorithm, disregarding cycle-consistency during inference [2,1].

2 Explicit Partial Cycle-Consistency

As mentioned in Section 1, cycle-consistency is a constraint in the theoretical partial multi-view matching optimization problem. Global cycle-consistency is defined by requiring that each of the pairwise matches P_{ij} can be decomposed into two matches, namely one from view V_i to an abstract set of real world objects U and one from U back to view V_j [3]. This constraint is often written as $\mathbf{P} = \mathbf{P}_u \mathbf{P}_u^T$, where $\mathbf{P}_u \in \mathbb{R}^{\sum_{i=1}^N n_i \times |U|}$ contains the N matchings between views V_i with the abstract set U.

Global cycle-consistency is equivalent to the more practical Equations 6 to 8, which we will refer to as implicit partial cycle-consistency[3,6]:

$$P_{ii} = I_{n_i \times n_i} \quad \forall i \in \{1, \ldots, N\}, \tag{6}$$

$$P_{ij} = P_{ji}^T \quad \forall i, j \in \{1, \ldots, N\}, \tag{7}$$

$$P_{ij} P_{jk} \subseteq P_{ik} \quad \forall i, j, k \in \{1, \ldots, N\}, \tag{8}$$

where $I_{n_i \times n_i} \in \mathbb{R}^{n_i \times n_i}$ is the identity matrix. Equation 6 says objects should be matched to themselves. Equation 7 encodes that the partial matches should be symmetric, so that the matches made from view V_i to V_j should exactly equal the matches made from view V_j to V_i. Finally Equation 8 models the transitivity constraint: Combining the matchings from V_i to V_j with V_j to V_k should give the matching from V_i to V_k, with the exception of matches that are lost through view V_j.

This work extends Equation 8 by explicitly modelling the subset relation, so that it becomes explicitly known which of the matches are lost through the transitive views:

Definition 2 (Explicit partial cycle-consistency).
A partial multi-view matching matrix $\mathbf{P} \in \{0, 1\}^{\sum_i n_i \times \sum_j n_j}$ *belongs to the set* $\mathbb{P}$ *of partially cycle-consistent multi-view matching matrices, if and only if:*

$$P_{ii} = I_{n_i \times n_i} \quad \forall i \in \{1, \ldots, N\}, \tag{9}$$

$$P_{ij} = P_{ji}^T \quad \forall i, j \in \{1, \ldots, N\}, \tag{10}$$

$$P_{ij} P_{jk} = P_{ijk} \quad \forall i, j, k \in \{1, \ldots, N\}, \tag{11}$$

where $P_{ijk} \in \{0, 1\}^{n_i, n_k}$, $P_{ijk} \subseteq P_{ik}$, *is the pairwise matchings matrix from* i *to* k, *filtering out all matches that are not seen in view* j:

$$P_{ijk}[a, c] = \begin{cases} 1 & \textit{if } P_{ik}[a, c] = 1 \ \& \ \exists b : \\ & \quad \textit{s.t. } P_{ij}[a, b] = P_{jk}[b, c] = 1. \\ 0 & \textit{else.} \end{cases} \tag{12}$$

The notation $P[\cdot, \cdot]$ *is used for indexing a matrix* P.

Proposition 1 (Explicit = Implicit = Global)**.**
Equation 11 is equivalent to Equation 8, so that explicit partial cycle consistency is indeed equivalent to implicit and global partial cycle-consistency.

Proof. It holds by Definition that $P_{ijk} \subseteq P_{ik}$. To prove the other way around, assume $P_{ij} P_{jk} \subseteq P_{ik}$. This gives us that

$$P_{ij} P_{jk}[a, c] = 1, \tag{13}$$

$$\iff P_{ik}[a,c] = 1\,\&\sum_{b'=1}^{n_j} P_{ij}[a,b']P_{jk}[b',c] = 1. \tag{14}$$

Now because P_{ij} and P_{jk} are partial permutation matrices, we know that each row and column has at most a single 1, with the rest of the elements being 0:

$$\iff P_{ik}[a,c] = 1\,\&\,\exists b:\text{ s.t. } P_{ij}[a,b] = P_{jk}[b,c] = 1, \tag{15}$$

$$\iff P_{ijk}[a,c] = 1. \tag{16}$$

This in turn proves that $P_{ij}P_{jk} = P_{ijk}$, because the elements in both matrices can only be 0 or 1.

When using cycle-consistency as a learning signal, it helps to formulate it as a combination of matches that start and end in the same view, so that their product should be (a subset of) the identity matrix. For partially overlapping views, we provide Corollaries 1 and 2 that make the identity matrix subset relation explicit, and show that an identity matrix formulation can only be implied.

Corollary 1 (Explicit pairwise cycles).
Equation 10 implies that

$$P_{ij}P_{ji} = I_{iji} \quad \forall i,j \in \{1,\dots,N\}, \tag{17}$$

where $I_{iji} \subseteq I_{n_i \times n_i}$, is the identity map from view i back to itself, filtering out matches that are not seen in view V_j:

$$I_{iji}[a,c] = \begin{cases} 1 & \textit{if } a = c, \\ & \&\ \exists b \textit{ s.t. } P_{ij}[a,b] = 1. \\ 0 & \textit{else.} \end{cases} \tag{18}$$

Proof. We prove that Equation 10 implies Equation 17. Equivalence trivially holds for full overlap cycles, but with partial overlap it could be that either $P_{ij}P_{ji} = I_{iji}$ or $P_{ij}P_{ji} \subset I_{iji}$, even though different non-matches are made between P_{ij} and P_{ji}, so that $P_{ij} \neq P_{ji}.T$

Using a similar argument as in the proof Proposition 1, we get that:

$$P_{ij}P_{ji}[a,c] = 1, \tag{19}$$

$$\iff \exists b \text{ s.t. } P_{ij}[a,b] = P_{ji}[b,c] = 1. \tag{20}$$

Now our assumption that $P_{ij} = P_{ji}^T$ tells us that $a = d$.

Corollary 2 (Explicit triplewise cycles).
Equations 10 and 11 together imply that:

$$P_{ij}P_{jk}P_{ki} = I_{ijki} \quad \forall i,j,k \in \{1,\dots,N\}, \tag{21}$$

where $I_{ijki} \subseteq I_{n_i \times n_i}$*, is the identity mapping from view* i *back to itself, filtering out all matches that are not seen in views* V_j *and* V_k*:*

$$I_{ijki}[a,d] = \begin{cases} 1 & \textit{if } a = d, \\ & \& \ \exists b, c \textit{ s.t. } P_{ij}[a,b] = P_{jk}[b,c] = P_{ki}[c,d] = 1. \\ 0 & \textit{else.} \end{cases} \tag{22}$$

Proof. We again get that the other way around also holds for full overlap cycles, but that a counter example can be constructed given partial overlap with inconsistencies in the matches that are **not** made, so that $P_{ij}P_{jk} \nsubseteq P_{ik}$. This means that any form of $P_{ij}P_{jk}P_{ki} \subseteq I_{ii}$ can only be implied.

Now to show Equation 11 implies Equation 21: Using Equation 11 gives

$$P_{ij}P_{jk}P_{ki} = P_{ijk}P_{ki}, \tag{23}$$

we need to show that:

$$P_{ijk}P_{ki}[a,d] = 1 \iff I_{ijki}[a,d] = 1. \tag{24}$$

Writing out the left hand side gives

$$P_{ijk}P_{ki}[a,d] = 1 \tag{25}$$

$$\iff \exists c \text{ s.t. } P_{ijk}[a,c] = P_{ki}[c,d] = 1, \tag{26}$$

where a similar argument as in Proposition 1 is used. Continuing, we get that

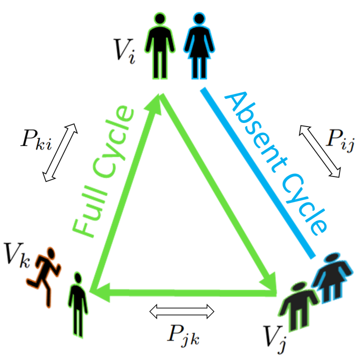
$$\iff \exists b, c \text{ s.t. } P_{ij}[a,b] = P_{jk}[b,c] = P_{ik}[a,c] = P_{ki}[c,d] = 1. \tag{27}$$

Now using Equation 10, we know that $P_{ik}[a,c] = P_{ki}[c,a]$, so that $P_{ki}[c,d] = P_{ki}[c,a] = 1$ gives that $a = d$. This gives that

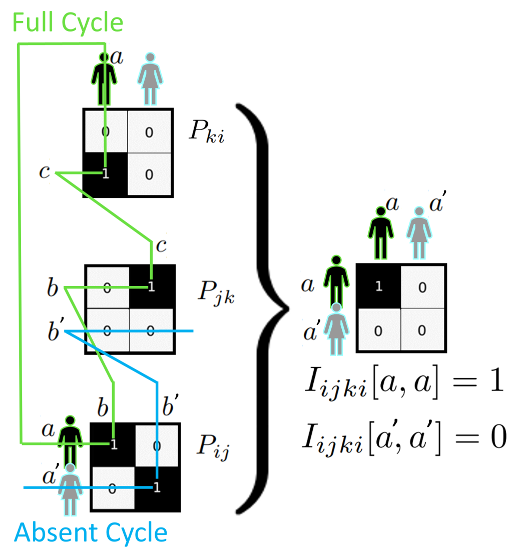
$$P_{ijk}P_{ki}[a,d] = 1 \iff I_{ijki}[a,d] = 1, \tag{28}$$

concluding the proof.

A visual interpretation of Equation 21 from Corollary 2 is also provided in Figure 1b, and an overview of the cycle-consistency formulations and their relations is provided in Figure 2.



(a) Multiple views with partial overlap, only the green person should be matched in a cycle.



(b) A visual interpretation of Equation 22. $I_{ijki}[a, a] = 1$ because a is matched to a b, matched to a c which is then matched back to a. The same cannot be said for a', so this cycle is absent.

Fig. 1: Partial cycle-consistency and the explicit identification of existing and absent cycles between multiple views visualized.

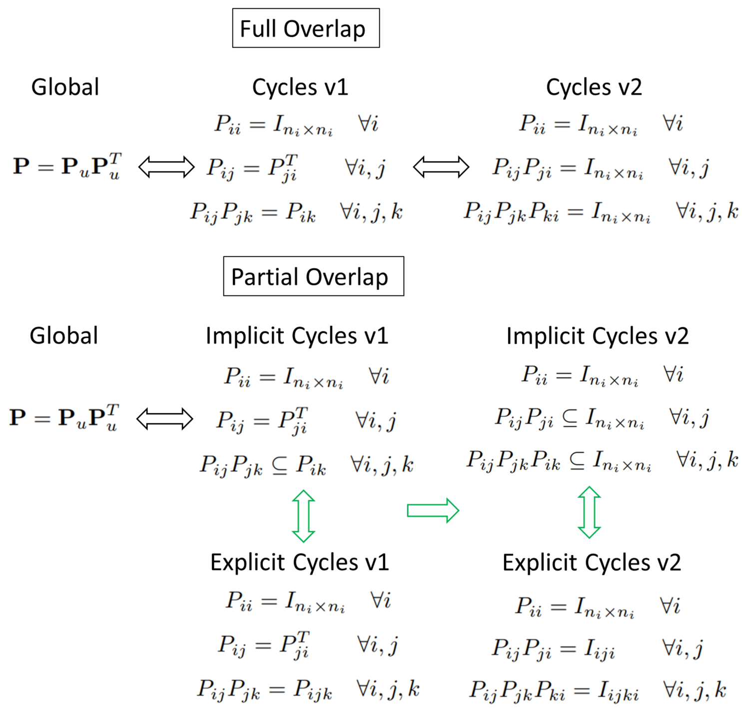


Fig. 2: Cycle-consistency definitions and their relations, both with and without partial overlap. The green arrows denote newly proven relations.

3 Additional Results

We report the amount of overlap and average number of people to be matched in Table 1, providing insight into why certain scenes in the testset provide a harder matching problem. The overlap in the corresponding train data for each set also influences the learning signal obtained from this set during training, making scenes such as shop especially hard, because not much could be learned from this scene, and the matching problem during evaluation requires finding correct cross-camera matches between 32 people per frame on average.

Jaccard Index	**Gate2**	**Square**	**Moving**	**Circle**	**Gate1**
Two Cameras	0.74|0.56	0.27|0.44	0.47|0.29	0.40|0.38	0.44|0.47
Three Cameras	0.65|0.44	0.14|0.28	0.31|0.15	0.26|0.23	0.29|0.33
Num People	4.7|6.6	19.1|21.5	7.7|13.4	15.2|14.0	16.4|17.2

Jaccard Index	**Floor**	**Park**	**Ground**	**Side**	**Shop**
Two Cameras	0.44|0.35	0.35|0.38	0.48|0.48	0.29|0.32	0.14|0.16
Three Cameras	0.25|0.17	0.17|0.24	0.32|0.30	0.12|0.16	0.07|0.05
Num People	17.2|16.3	13.4|16.4	29.6|32.1	24.6|23.6	30.5|32.2

Table 1: The average overlap (Jaccard Index) and number of people in the **train|test** data for each scene, sorted from easy to hard based on the matching difficulty [1].

We visualize the variation between runs for the main models and contributions in Figure 3, where it can be seen how Time Divergent Scene Sampling improves performance overall, and how combining cycle variations with partial masking not only boosts performance but also improves the consistency of the approach.

References

1. Gan, Y., Han, R., Yin, L., Feng, W., Wang, S.: Self-supervised multi-view multi-human association and tracking. In: Proceedings of the 29th ACM International Conference on Multimedia. pp. 282–290 (2021)
2. Han, R., Wang, Y., Yan, H., Feng, W., Wang, S.: Multi-view multi-human association with deep assignment network. IEEE Transactions on Image Processing **31**, 1830–1840 (2022)
3. Huang, Q.X., Guibas, L.: Consistent shape maps via semidefinite programming. In: Computer graphics forum. vol. 32, pp. 177–186. Wiley Online Library (2013)
4. Pachauri, D., Kondor, R., Singh, V.: Solving the multi-way matching problem by permutation synchronization. Advances in neural information processing systems **26** (2013)

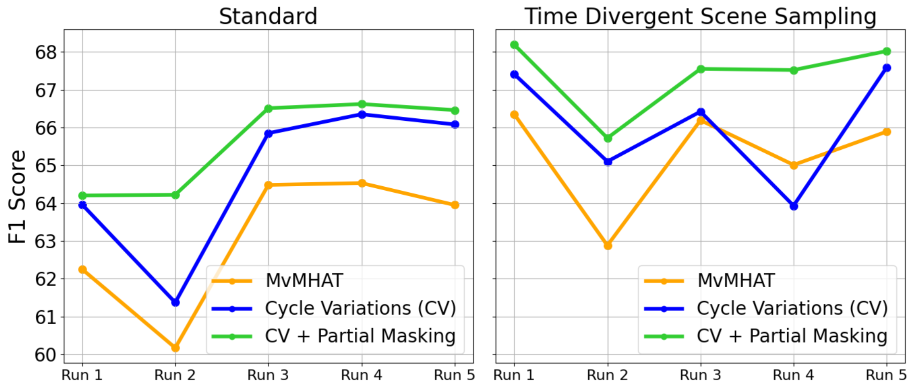


Fig. 3: Variation of the runs with and without Time Divergent Scene Sampling. Combining all contributions provides the best performing and most consistent model.

5. Swoboda, P., Mokarian, A., Theobalt, C., Bernard, F., et al.: A convex relaxation for multi-graph matching. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11156–11165 (2019)
6. Ye, Z., Yenamandra, T., Bernard, F., Cremers, D.: Joint deep multi-graph matching and 3d geometry learning from inhomogeneous 2d image collections. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 36, pp. 3125–3133 (2022)