EFFICIENT PERSON RE-IDENTIFICATION IN VIDEOS USING SEQUENCE LAZY GREEDY DETERMINANTAL POINT PROCESS (SLGDPP)

Gaurav Kumar Nayak, Utkarsh Shreemali, R Venkatesh Babu, Anirban Chakraborty

Department of Computational and Data Sciences, Indian Institute of Science, Bangalore

PERSON RE-ID IN VIDEOS

Input Query: 

Retrieved from Gallery

- Challenge: To handle **high correlation** between temporally adjacent frames
- **High redundancy** between the frames make them non-informative
- Necessary to identify important observations in a given sequence and **remove the non-informative ones** for an efficient and accurate re-id

PROBLEM DESCRIPTION

- Selecting important frames can be viewed as a **subset selection problem**.
- Given a sequence of frames of a person captured within a camera FoV, the goal is to find out an **optimal subset of frames** relevant to the task of person re-id.
- Using an optimal/diverse subset will **reduce memory and computation time**.
- We use Determinantal Point Process (**DPP**) [1,2] to achieve this objective which is a statistical tool for subset selection with maximum diversity.

DPP FOR SUBSET SELECTION

- First to propose the use of DPP for the application of video based person re-id
- Let $\beta = \{1, 2, \ldots, M\}$ denote a set containing M frames of a video tracklet.
- Then, total number of possible subsets = 2^M
- Let $\alpha \subseteq \beta$ be any subset whose probability of being selected is given by:

$$P(\alpha; L) = \frac{det(L_\alpha)}{det(L+I)}$$

$$L_{jk} = \phi(\mathbf{x}_j)^T \phi(\mathbf{x}_k)$$

I: Identity matrix L: Kernel matrix

L_α is the principal minor (sub-matrix) with rows and columns from indexed by the integers in α

- The most likely subset given by **MAP** estimate is :

$$\boxed{\alpha^* = \arg\max_\alpha P(\alpha; L) = \arg\max_\alpha \det(L_\alpha)}$$ [**NP Hard** problem]

- We take $\phi(.)$ to be output of neural network which is function of frame feature
- We solve it through LGDPP approach based on lazy greedy technique [3] which is usually accurate for most practical applications
- To utilize the sequential nature of videos we propose **Sequential Lazy Greedy DPP (SLGDPP)** which divides the video into equal sets & performs LGDPP on it.

PROPOSED METHOD

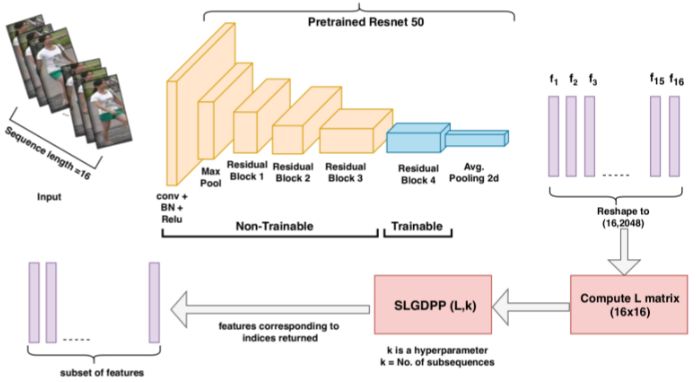


Fig: 1: Proposed Architecture (Subset Selection Module)

- The proposed DPP based subset selection method is generic as it can be easily coupled with any video based pretrained model and any fusion technique can be used atop it.

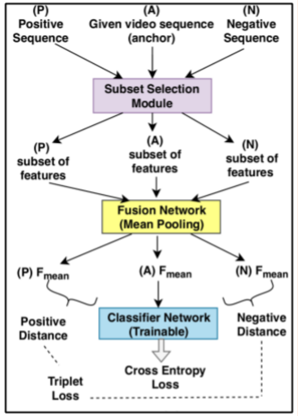
Algorithm 1: SLGDPP

Input: matrix L and count of subsequences k (Sequence of length N, with index starting from 1, is partitioned into k subsequences.)

```
y_prev ← ∅ ;
y_cur ← ∅ ;
X ← ∅;
i ← 1;
while i <= k do
    y_cur ← y_prev ∪ indices of i^th subsequence;
    A ← submatrix of L corresponding to y_cur;
    S ← sorted(LGDPP(A));
    Ś ← S \ y_prev;
    if Ś ≠ ∅ then
        X ← X ∪ Ś;
        y_prev ← Ś;
    end
    i++;
end
```

Output: subset indices X

Proposed Algorithm



Total Loss= Triplet + CrossEntropy

RESULTS

- For evaluating our method, we use “rank1” accuracy and “mean average precision”(mAP) on two popular video re-id datasets : MARS and PRID 2011

Effect of Transformation on proposed LGDPP method

$\phi()$	MARS Rank 1	MARS mAP	PRID 2011 Rank 1	PRID 2011 mAP
without	81.0	73.6	89.0	93.07
with	**82.2**	**75.3**	**90.69**	**93.94**

Effect of Transformation on proposed SLGDPP method

$\phi()$	MARS Rank 1	MARS mAP	PRID 2011 Rank 1	PRID 2011 mAP
without	81.1	73.8	88.99	92.98
with	**82.6**	**76.0**	**91.13**	**93.98**

Comparison with other video re-id methods

Method	MARS Rank 1	MARS mAP	PRID 2011 Rank 1	PRID 2011 mAP
Song et al.	77.17	71.17	-	-
DuATM	78.74	62.26	-	-
Dai et al.	80.50	69.10	87.80	-
Li et al.	82.30	65.80	**93.20**	-
Resnet 50	81.00	74.00	90.33	93.74
LGDPP	82.20	75.30	90.69	93.94
SLGDPP (Ours)	**82.60**	**76.00**	91.13	**93.98**

Fig. 2: Visualization of the subset returned by proposed method (highlighted in red color)

CONCLUSION

- Video based re-id problem can be framed as subset selection problem where important frames of a person need to be identified for better & efficient re-id.
- We propose the use of two variants of the classical DPP technique, LGDPP and our proposed SLGDPP along with the task specific transformation function ϕ
- SLGDPP based re-id approach performs better than the LGDPP variant as it captures the sequential nature of frames.

REFERENCES

1. Odile Macchi, “The coincidence approach to stochastic point processes,” Advances in Applied Probability, 1975.
2. Alex Kulesza and Ben Taskar, “Determinantal point processes for machine learning,” Foundations and Trends® in Machine Learning, 2012
3. Michel Minoux, “Accelerated greedy algorithms for maximizing submodular set functions,” in Optimization techniques, 1978.

Acknowledgement: We thank IEEE SPS Bangalore Chapter for providing a travel grant.