ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 3, March 2017

A Review on: Person Re identification and Summarization

Kajal P. Waykole[1], S. S. Vasekar[2]

PG Student, VLSI and Embedded System, SKNCOE, Vadgoan (BK), Pune, India[1]

Assistant Professor, VLSI and Embedded System, SKNCOE, Vadgoan (BK), Pune, India[2]

ABSTRACT: Tracking and re-identification in wide-region camera systems is a challenging problem because of non-covering visual fields, differing imaging conditions, and appearance changes. The issue of person re-identification and tracking is considered and propose a novel clothing context-aware color extraction strategy that is robust to such changes. Annotated samples are used to learn color drift patterns in a non-parametric way utilizing the irregular forest distance (RFD) function. The color drift patterns are automatically transferred to associate objects across various perspectives utilizing a unified graph matching structure. A hyper graph representation is utilized to connect related articles for search and re-identification. A different hyper graph positioning system is proposed for person-focused network summarization. The proposed algorithm is validated on a wide-region camera organize comprising of ten cameras on bike paths. Also, the proposed algorithm is compared with the state of the art person re-identification algorithms on the VIPeR dataset.

KEYWORDS: Camera network, person re-identification, search, summarization.

I. INTRODUCTION

Distributed camera networks represent a few difficulties to information investigation, including the huge amount of video that should be conveyed. Given the transmission capacity and network constraints, de centralized methodologies taking full utilization of restricted computational power in individual cameras have been gaining prominence.

Person tracking over different cameras is one of the preparatory steps in many distributed camera applications. However, because of differing lighting, complex shape and appearance changes, the performance of person tracking is still far from the ideal. To this extent, there have been several works previously that perform garments based re-distinguishing proof to relate a man over different cameras to take care of the following issue.

In surveillance videos it is hard to parse the clothing information for reliably associating a person across various camera views due to unknown transformation in color and overall appearance of the person (people occupy few pixels relative to the entire image frame). Additionally, frequent exchange of crude picture information to the focal server imposes network bottlenecks and is not adaptable for substantial systems. Existing methodologies handle the issue either with appearance based features alone or in combination with spatial-temporal information. Be that as it may, appearance based features are insufficient for matching due to viewpoint, pose and lighting changes. Spatial-temporal information helps to an extent, and it requires complete knowledge of the system. Moreover, existing methodologies expect that worldwide are available and accurate. With the state-of-the-art person detectors and trackers, this is not a reasonable assumption.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 3, March 2017

II. LITERATURE SURVEY

The search and retrieval system for a distributed camera network. Raw videos are processed locally at the remote camera nodes, and only the computed information is passed to the central server. Based on the collected information, a graph is built to model the network observations to facilitate search and retrieval. A central philosophy of our work is to generate an informative and diverse set of visual results based on a user query. This is achieved by amanifold ranking on the constructed graph, followed by a greedy diversification algorithm. Experimental results using real world data has demonstrated the effectiveness of the proposed system [1].

A novel context-aware graph based system to assist human analysts to efficiently browse and search objects in a large wide area camera network. A novel strategy to encode contextual information such as appearance, scene and spatial-temporal contexts is provided to increase the overall accuracy of retrieval. The proposed methodology is validated with extensive experiments on some real-world large scale camera network dataset. With contextual information, the accuracy of the proposed system is increased by approximately 50% (in terms of F-measure) [2].

A vision-based adaptive control approach for a two-WIP robot to track a moving target person. To combine an OptiTrack camera and a Kinect camera for more efficient and robust tracking performance, a new algorithm is proposed. Leader–follower control, dynamic balance control, and visual tracking have been effectively developed together such that desired tracking and balancing performance has been achieved. Extensive experiment studies have been performed to verify the effectiveness of the proposed approach and to show that robust human tracking has good performance within indoor environments by the mobile inverted pendulum robot [3].

A new pose prior technique that can effectively leverage the correlation between images from different viewpoints, significantly improving re-identification performance. Moreover, with the discriminative feature approach, the distinctiveness of a person can be better highlighted. Experimental results on challenging datasets suggest that the proposed algorithm can significantly improve performance and robustness in real-world re-identification problems with lighting and viewpoint changes. Some newly proposed metric learning algorithm claim higher performance than the ones used for comparison. To combine our proposed algorithm with such metric learning algorithms, with the hope that the re-id performance can be further elevated. In a more general scenario, the proposed algorithms can be extended by using two pose priors corresponding to the front view and back view of a subject. For each person, both the front and back views can be estimated, along with measures of confidence in the two views based on the estimated pose. As the target is tracked and observe more images, the descriptors of the front and/or back views can be updated if the incoming image has a higher confidence than the current confidence index [4].

A robust method for detection and tracking human poses in videos by matching video trajectories to a 3D motion capture model. STM matches trajectories to a 3D model, and hence it provides intrinsic view-invariance. The main novelty of the work resides in computing the correspondence between video and motion capture data. Although it might seem computationally expensive and difficult to optimize at first, using an l1-formulation to solve for correspondence results in an algorithm that is efficient and robust to outliers, missing data and mismatches. How STM outperforms state-of-the-art approaches to object detection based on deformable parts models in the Berkeley MHAD is shown, the Human3.6M and the CMU MAD dataset. A major limitation of our current approach is the high computational cost for calculating the joints' response, which is computed independently for each frame [5].

A new feature, SLTP, for people detection in depth images has been introduced in this paper. Experimental results have shown that SLTP can take advantage of depth images and gives better detection rates than HOG and LTP with low computational cost. Besides, the selection of the threshold in feature extraction and the detection window size is also suggested empirically by a series of experiments. SLTP is a kind of statistical feature, in which only depth differences within a local region are considered. The human body size and some other body geometries are not yet taken as features here. If the statistical features are combined with some geometric features, higher detection rate with lower false positive rate would be achieved[6].

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 3, March 2017

Use semantic color names in the person re-id problem. Compared to commonly used color histogram, the visual matching results by color names is much closer to what human operators would consider intuitive. Also propose to apply the Rank Boost algorithm to learn the weights of similarity measurements of the corresponding local image descriptors. The experiment on challenging dataset shows the effectiveness of our proposed system compared to the state-of-the-art methods [7].

The distance metric is critically important for effective person re-identification in surveillance tasks. Thus, it is rational to find a suitable distance metric learning algorithm to boost the performance of person re-identification. In recent years, many distance metric learning algorithms have been developed, such as ITML and metric learning for LMNN. However, these algorithms are not suitable for person re-identification, because there are only limited training image pairs to learn a metric in person re-identification. Although KISS metric learning is considered state-of-the-art, it shares a similar problem. Given a small number of training samples, observe that covariance matrices estimated by KISS are biased. Therefore, present the MCE-KISS. The proposed MCE-KISS algorithm exploits the smoothing technique to enlarge the small eigenvalues of the estimated covariance matrix, and discriminative learning based on MCE which is more reliable than classical ML estimation. The employed two statistical techniques effectively enlarge the underestimated small eigenvalues and better estimate the covariance matrix. Therefore, MCE-KISS significantly improves KISS for person re-identification [8].

The system effectively integrates the human detectors and V-SLAM framework to relocate the humans in 3-D space, followed by an innovative 3-D based CMK tracking, which not only locally associates the targets but also globally optimizes the associations according to the 3-D information. Such system can be regarded as a key component for high level applications, such as video analysis in a large scale of mobile network. Besides, the proposed framework can also be further applied to other class of on-road objects if the corresponding detectors are available [9].

Proposed a semi-supervised framework for human parsing which leverages video contexts without extra annotation. It contains two components: the contextual video parsing and the non-parametric human parsing. Extensive experiments on two benchmark human parsing datasets as well as a newly collected FI dataset well demonstrate the effectiveness of our proposed framework. More images can optionally label to train a better human parser. However, the great burden of human labeling may considerably limit the scalability of the human parser. Since our method only needs the unsupervised videos which can be easily crawled from the web, the solution can easily scale up to new videos with even more challenging poses and views, e.g. lying on the floor or sitting in the chair. If the human labeling are available, it requires labelers to check the video pose co-estimation results and then train a better human parser [10].

III. SYSTEM OVERVIEW

The below figure is the block diagram of the proposed system which is further divided into various parts. The block diagram shows the identification of person.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 3, March 2017

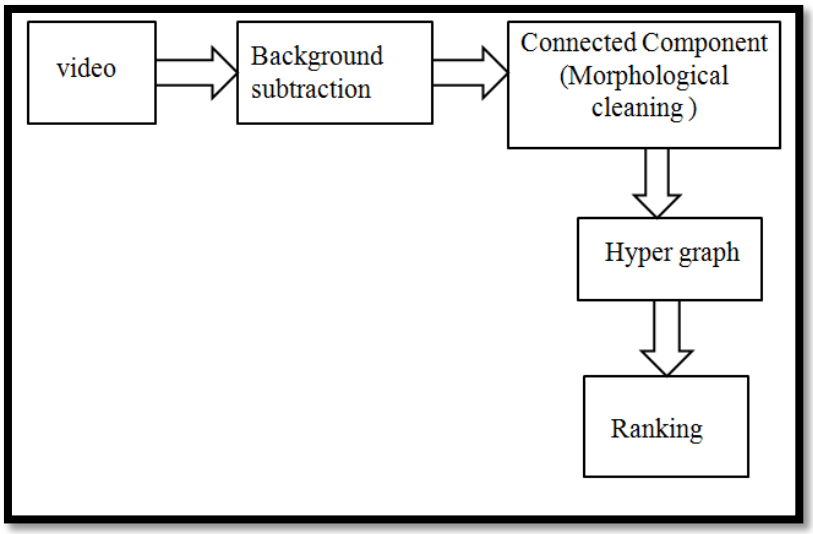

Fig1. Block Diagram Of Proposed System

Background Substraction :

Background subtraction, also known as Foreground Detection, is a technique in the fields of image processing and computer vision wherein an image's foreground is extracted for further processing (object recognition etc.). Generally an image's regions of interest are objects (humans, cars, text etc.) in its foreground. After the stage of image preprocessing (which may include image denoising, post processing like morphology etc.) object localization is required which may make use of this technique.

Background subtraction is a widely used approach for detecting moving objects in videos from static cameras. The rationale in the approach is that of detecting the moving objects from the difference between the current frame and a reference frame, often called "background image", or "background model". Background subtraction is mostly done if the image in question is a part of a video stream.

Background subtraction provides important cues for numerous applications in computer vision, for example surveillance tracking or human poses estimation. However, background subtraction is generally based on a static background hypothesis which is often not applicable in real environments. With indoor scenes, reflections or animated images on screens lead to background changes. In a same way, due to wind, rain or illumination changes brought by weather, static backgrounds methods have difficulties with outdoor scenes.

Connected component (morphological)

Consider a camera network with *M*distributed and static cameras. It is assume that the entire network is time-synchronized and no calibration information is available. At each camera node, a background subtraction based tracker is used to segment foreground pixels and automatically detect moving objects (pedestrians and bikers) using connected component analysis. Each camera assigns a unique local ID to every person and sends an abstracted record consisting of: camera ID, timestamp, person's bounding box on the image plane and the person's image data (obtained from the frame with the maximum person

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 3, March 2017

area) to the central server. The central server builds a time-evolving hypergraph with a decaying memory using the observations obtained from remote cameras.

Let$\omega_{ij}^{\text{Appearance}}$ and $\omega_{ij}^{\text{Spatial-Temporal}}$be appearance and spatial-temporal topology based similarity matrices where i and$j$ are tracklet/node indices. In the following appearance based similarity computations by explicitly modelling color drift patterns is described. Then,spatial-temporal weighting using network topology. Given this setup, it is envision that the central server could be queried to infer network events and provide a smart summarization. An example query would be "Find people wearing green shirt and grey shorts in camera 8between time 9:32am and 9:34am". Superpixels are leveraged defined over clothing regions to improve the robustness of appearance based matching. Most importantly, existing approaches do not explicitly model color drift patterns and they are not applicable for multiple views. A unified framework for matching person's appearance between multiple views is proposed by explicitly modeling color drifts in a discriminative manner.

Spatial-temporal topology is modelled using ground truth associations and build a hypergraph representation for modeling relationships. The problem of person re-identification is posed as hypergraph ranking. Finally, propose a diverse hypergraph ranking algorithm for summarization.

A. Appearance Modeling

1) Dense Color Histogram:

For every moving person, a multi-dimensional ($D = 288$) LAB color histogram is extracted. A patch size of10×10 is sampled on a grid with a step size of 4 pixels. 32-bins in L, A and B channels are used respectively. Also, the image is down-sampled into 3-levels with scaling factors of 0.5, 0.75 and 1.0. Finally, the $L2$-normalized dense color histogram is averaged to represent regions. Each person bounding box is divided into three regions (1/8 for head, 3/8 for torso and 4/8 for legs) and dense color histogram for torso and legs are extracted for learning color drift pattern. In addition to dense color histogram, culture color histogram is computed for indexing top and bottom portions of the person. Finally, color drifts patterns are automatically transferred for appearance matching.

2) Learning Multi-View Color Drift Patterns:

During the training stage, it is assumed that ground truth associations between tracklets in two different views are available. Let $z_i^{\text{torso}}$and z_j^{torso} be dense color histograms computed for torsos of people i and$j$ respectively (similarly,z_i^{legs} and $z_j^{\text{legs}}$are computed for legs of people). Euclidean distance provides a simple method for computing distances and it is often insufficient for capturing the underlying data representation.

Traditionally, Mahalanobis based distance metric is used for learning a globally optimal metric based on the equivalence constraints It is limited in its capacity to capture complex data representation and it is computationally inefficient as the dimension of data increases. To overcome this, several multi-metric techniques have been proposed and they are often affected by the memory storage complexity since the matrixes need to be stored per point or the subset of points. In contrast, Random Forest Distance function provides a non-parametric and non-linear distance metric that achieves the efficiency of both global and multi-metric techniques. Given the set of training pairs with equivalence constraints i.e.$\{(z_i, z_j)\}$, a distance metric is learn using the Random Forest Distance function: A set of similar (label 1) and

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 3, March 2017

dis-similar (label 0) samples are generated from the training set. Dissimilar samples are generated by randomly pairing color descriptors from different persons. A random forest classifier is trained using the training samples from similar and dis-similar sets and the color drift probability is given by

$$p^{color-drift}(z_i, z_j) = \frac{1}{U}\sum_{U} f_u(z_i, z_j)$$

WhereUis the number of trees (in this experiments $U = 800$ is set and $f_u$is the classification output of the tree (similar vs dis-similar). The above-mentioned strategy provides a simple yet powerful technique for modeling color drift patterns in surveillance scenarios wherein pixel-wise multi-view correspondences not possible. Also, the RFD method scales well for increasing dimensionality compared to Kernel Density Estimation and they are inherently non-linear.

3) Salient Super pixel Graph*:*

During the testing stage, torso and legs regions of tracklets are over-segmented into several non-overlapping segments using SLIC super pixel algorithm. Let $s_i = \{s_{ik}\}$be the set of super pixels for the *i*-thtracklet where*k* is the superpixel index. represent the super pixels using dense color histograms and choose top K superpixels based on the saliency. Pair wise feature distance between superpixels is computed and order the median-th distance in the increasing order and choose the top K super pixels. A complete graph $G_i = (V_i, E_i)$ is defined over the salient super pixels whereV_i is the set of$n_i$super pixel nodes such that $n_i \leq k$ andE_i is the set of edges defined between super pixels. Similarly, a complete graph $G_j = (V_j, E_j)$ is defined for the tracklet in a different view with $n_j$super pixel nodes ($s_j = \{s_{jk'}\}$ where$k'$is the super pixel index).

4) Clothing Context-Aware Appearance Matching:

Propose a novel clothing context-aware appearance based association using a graph matching framework. Let$Q \in R^{ninj \times ninj}$ be the global affinity matrix that defines node and edge affinities such that and otherwise

$$Q_{k1k1',k2k2'} = \begin{cases} \emptyset(k1, k1'), & k1 = k2 \text{ and } k1' = k2' \\ \varphi(e_{k1,k2}, e'_{k1',k2'}), & k1 \neq k2 \text{ and } k1' \neq k2' \\ 0, & \text{otherwise} \end{cases}$$

Where$\emptyset(k1, k1')$encodes node-to-node similarity. For computing node-to-node similarity, color drift pattern is taken into account and it is given by

$$\emptyset(k1, k1') = p^{color-drift}(z_{k1,k2}) \overbrace{\sum_{d=288}^{d} \min(z_{k1}(d), z_{k1'}(d))}^{\text{Histogram Intersection}}$$

Wherez_{k1} and $z_{k1'}$are LAB space color histograms extracted over superpixel regions. $\varphi(e_{k1,k2}, e'_{k1',k2'})$ encodes edge-to-edge similarity and takes clothing context into account. Edge $e_{k1k2}$is represented by a feature difference vector computed

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 3, March 2017

using culture color histograms of nodes $k1\ and\ k2$ i.e. $f = |z_{k1} - z_{k2}|$. Similarly, edge $e'_{k1',k2'}$ is represented by feature difference vector f'. By this, clothing context aware edge representation is captured i.e. edge between torso and leg superpixels captures clothing difference $\varphi(e_{k1,k2}, e'_{k1',k2'})$ is given by

$$\frac{1}{D}\sum_{d}^{D=288}\begin{cases}1, & \text{if } f(d), f'(d) = 0\\ \max\left[I\left(f(d) = f'(d)\right), \dfrac{\min\left(f(d), f'(d)\right)}{\max\left(f(d), f'(d)\right)}\right]\end{cases}$$

Where I is the indicator function. Intuitively, φ equals one when the difference histogram matches exactly.

5) Computational Complexity:

The total complexity of graph matching is proportional to the number of non-zero elements in Q. Let m_i and m_j be the number of edges in graphs G_i and G_j respectively. The complexity is given by $O(m_i m_i$ for full matching. For surveillance videos, people appear relatively small and hence number of superpixels is also small.

A. Hypergraph

A pair wise simple graph is insufficient to represent relationship between vertices. In multi-graph based representation, multiple edges are constructed between two vertices. In hyper graph based person re identification, multiple edges are constructed between three or more vertices. Also, a hyper graph accounts for relationship between three or more vertices containing local grouping information and also models higher order relationship between vertices. Compared, implicitly introduce group relationship between tracklets using hyper graph representation. A hyper graph contains a set of vertices defined as a weighted hyper edge; the magnitude of hyper edge weight denotes the probability of a vertex belonging to the same cluster.

IV. ACKNOWLEGEMENT

I would like to express my sincere thanks to the Head of Department **Dr. S. K. Shah** for her support and valuable guidance throughout the work. Also, I would be grateful to Principal, **Dr. A.V. Deshpande** for his encouragement throughout the course.

V. CONCLUSION

Person re-identification handles pedestrian matching and positioning crosswise over non-covering camera sees. It has numerous essential applications in video surveillance by saving a lot of human efforts on exhaustively looking for a man from a lot of video groupings. In any case, this is likewise an extremely difficult assignment; here a novel color drift and clothing context-aware person search and re-identification strategy is proposed for a wide-range camera with an emphasis on summarization. The proposed logic is affirmed with wide trials on some genuine tremendous scale camera arrange dataset with 10 cameras along the college bicycle way. We demonstrated critical changes over best in class remove metric based appearance planning and graph situating methodology. Moreover, the proposed garments setting mindful individual re-recognizable proof calculation is differentiated and the best in class singular re-conspicuous confirmation computations in VIPeR dataset.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 3, March 2017

REFERENCES

[1] J. Xu, V. Jagadeesh, Z. Ni, S. Sunderrajan, and B. Manjunath, "Graphbasedtopic-focused retrieval in a distributed camera network," IEEETrans. Multimedia, vol. 15, no. 8, pp. 2046–2057, Dec. 2013.
[2] S. Sunderrajan, J. Xu, and B. Manjunath, "Context-aware graph modelingfor object search and retrieval in a wide area camera network,"in Proc 7th Int. Conf. IEEE Distrib. Smart Cameras, Oct.–Nov. 2013,pp. 1–7.
[3] Weiquan Ye and Zhijun Li, "Vision-Based Human Tracking Control of a Wheeled Inverted Pendulum Robot" 2015 IEEE transactions.
[4] Ziyan Wu and Yang Li," Viewpoint Invariant Human Re-identification inCamera Networks Using Pose Priors andSubject-Discriminative Features" IEEE transactions on pattern analysis and machine intelligence, October 2013.
[5] Feng Zhou and Fernando De la Torre" Spatio-temporal matching for Human Pose Estimation in Video" 2015 IEEE.
[6] Shiqi Yu, Shengyin Wu and Liang Wang "SLTP: A Fast Descriptor for People Detection in Depth Images" 978-0-7695-4797-8/12 $26.00 © 2012 IEEE.
[7] Cheng-HaoKuo and SamehKhamis "Person Re-identification using Semantic Color Names and RankBoost" 978-1-4673-50542-95/132/$31.00 ©20132 IEEE.
[8] Dapeng Tao and Lianwen Jin "Person Reidentification by Minimum ClassificationError-Based KISS Metric Learning" 2168-2267© 2014 IEEE.
[9] Kuan-Hui Lee and Jenq-Neng Hwang "Ground-Moving-Platform-Based Human Tracking
Using Visual SLAM and ConstrainedMultiple Kernels" 1524-9050 © 2016 IEEE.
[10] Si Liu, Member and Xiaodan Liang "Fashion Parsing With Video Context" IEEE transactions on multimedia, vol. 17, no. 8, august 2015.