ACTIVE IMAGE PAIR SELECTION FOR CONTINUOUS PERSON RE-IDENTIFICATION

Abir Das, Rameswar Panda, Amit Roy-Chowdhury

Electrical and Computer Engineering Department, University of California, Riverside, USA

ABSTRACT

Most traditional multi-camera person re-identification systems rely on learning a static model on tediously labeled training data. Such a framework may not be suitable for situations when new data arrives continuously or all the data is not available for labeling beforehand. Inspired by the 'value of information' active learning framework, we propose a continuous learning person re-identification system with a human in the loop. In brief, we term this 'continuous person re-identification'. The human in the loop not only provides labels to the incoming images but also improves the learned model by providing most appropriate attribute based explanations. These attribute based explanations are used to learn attribute predictors along the way. The overall effect of such a stratgey is that starting with a few annotated images, the system begins to improve via a symbiotic relationship between the man and the machine. The machine assists the human to speed the annotation and the human assists the machine to update itself with more annotation so that more and more distinct persons are re-identified as more and more images come in. Using a benchmark dataset, we validate our approach and compare with state-of-the-art methods.

Index Terms— Person re-identification, Active learning, Attributes.

1. INTRODUCTION

Person Re-identification addresses the task of identifying and monitoring people across a number of non-overlapping cameras. Traditional re-identification methods are static and involve an intensive supervised training phase [1, 2, 3, 4, 5, 6] where it is assumed that all the training examples are labeled. Apart from high cost of labeling, all the data may not be available at the very outset. In this work, we focus on the fundamental challenges that need to be overcome in order to address the largely unaddressed problem of continuously updating a person re-identification model starting with a small pool of labeled images. In short, we term this as 'continuous person re-identification'.

In presence of a continuous inflow of unlabeled images containing both previously seen and unseen persons, inputs from a human is necessary. A scalable approach to reduce the labeling effort requires a small number of questions (label requests) to be asked without compromising the performance. Towards this goal, the system can use feedback from the human so that human knowledge is transferred and reflected in the fewer and better questions asked to the human subsequently. This paper proposes such an active learning based continuous person re-identification framework. Traditional active learning settings involve tedious comparisons with all the classes by a human. The incorporation of the domain knowledge from the human to the process can help in reducing the subsequent effort in labeling. A recent line of work [7, 8, 9] draws inspiration from the way human experts simplify the task of discrimination by using mid level semantic features, called *attributes*. Attributes define a richer language to convey the domain knowledge from the expert to the model. Inspired by the recent success of using attributes as feedback in face recognition and scene classification [7, 10, 11], we combine attribute feedback with 'value of information' [12] based active learning strategy to select a small but informative set of images for labeling.

Though some recent works in re-identification [13, 14, 15, 16] have studied the use of attributes, they used it as a replacement to low level features. Unlike these works where preannotated data with a predefined vocabulary of attributes were assumed, the proposed framework uses the attribute feedback to learn attribute predictors on the way. Active learning methods with or without involving attributes, often, depend on the assumption of having training examples from all possible classes at the start. A simple way of alleviating this restriction is by setting up a threshold of maximum number of comparisons before giving a new label [12]. These assumptions are unrealistic for real life re-identification problems where new persons come in continuously. Such re-identification systems, handling large number of previously unseen people, may incorrectly assign separate labels to different instances of the same person. Instead of relying on the user set threshold, we propose an optimization framework with the possibility of encountering previously unseen persons.

To the best of our knowledge, this is the first work where a person re-identification system tackles a continuous inflow of data with the help of a human in the loop. The human is queried for labels as well as discriminating attributes to update itself and build a knowledge base about the attributes. Starting with absolutely no attribute information, the system uses the incrementally built knowledge base to reduce the burden on the human as time progresses. We experiment using a publicly available benchmark dataset and compare with state-

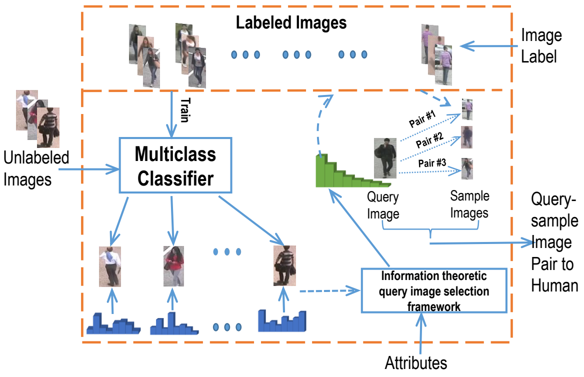


Fig. 1: Image pair selection. Samples are presented in order of decreasing probabilities obtained from the sorted class membership distribution of the query image. Next the query-sample pair is examined by the expert for match.

of-the-art re-identification methods.

Contributions of the paper: To summarize, the contributions of the proposed approach to the problem of person re-identification are the followings. (1) Person re-identification has been formulated as a continuous learning system which is able to learn and update from new data arriving in batches. (2) The proposed continuous person re-identification system uses a value of information based strategy for active image pair selection capable of handling a large number of new classes in each batch. (3) Apart from giving just the label, the role of the human is also to give attribute label explanations which is used to reduce the number of active interrogations.

2. METHODOLOGY

The person re-identification system is based on a low level feature based multi-class classifier where each class corresponds to a separate person. To get started, the classifier is trained on a small amount of training data, labeled only with the person ids without any attribute information. As the next batch of unlabeled images arrives, the feature based classifier chooses the most informative unlabeled image (query image) and a list of candidate images (sample images) from the labeled set. The query-sample pair is chosen so that subsequent misclassification risk is maximally reduced upon getting the label of the query image from the human expert. The human expert labels the query image by answering ‘yes’ or ‘no’ to the question whether the query and sample image are of the same person or not. The probable sample images for a particular query image are presented to the expert in decreasing order of the sample image’s class membership probabilities. The class membership probability distribution, given by the classifier, expresses the probability of the query image to belong to one of the already labeled persons from where the sample images are chosen.

Let the number of labeled classes at a certain moment be K and the K length class membership distribution of an unlabeled image x be $\mathbf{p}_x = \{p_x^1, p_x^2, \cdots, p_x^K\}$. x can belong to a previously seen or an unseen class. For the sake of simplicity let us assume equal risk if x is misclassified into any of the classes other than its true class. Given $P_n(x)$ be the probability that x is a previously unseen person, the expected misclassification risk is given by,

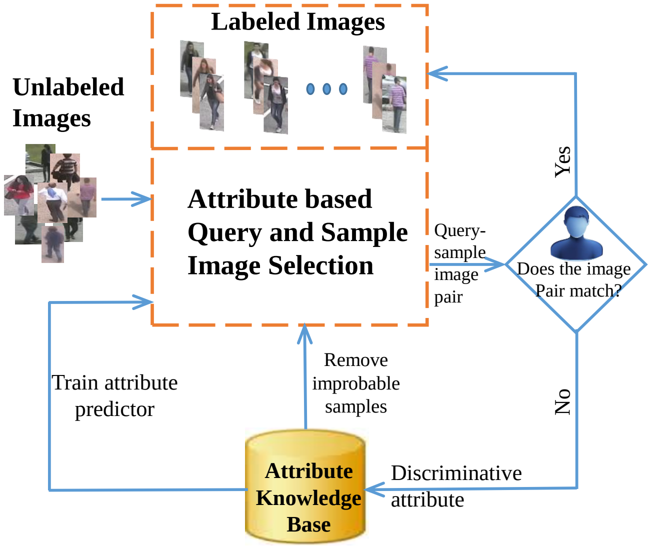

Labeled Images
Unlabeled Images
Attribute based Query and Sample Image Selection
Query-sample image pair
Does the image Pair match?
Yes
No
Train attribute predictor
Remove improbable samples
Attribute Knowledge Base
Discriminative attribute

Fig. 2: Attribute feedback in image pair selection. As unlabeled images come, the classifier along with an attribute predictor, learned on the way, selects a query image which is presented to the human along with candidate matches from the labeled pool. The human does the labeling and gives attribute based explanation of the mismatches, that, in turn, is used to learn and improve the attribute predictors.

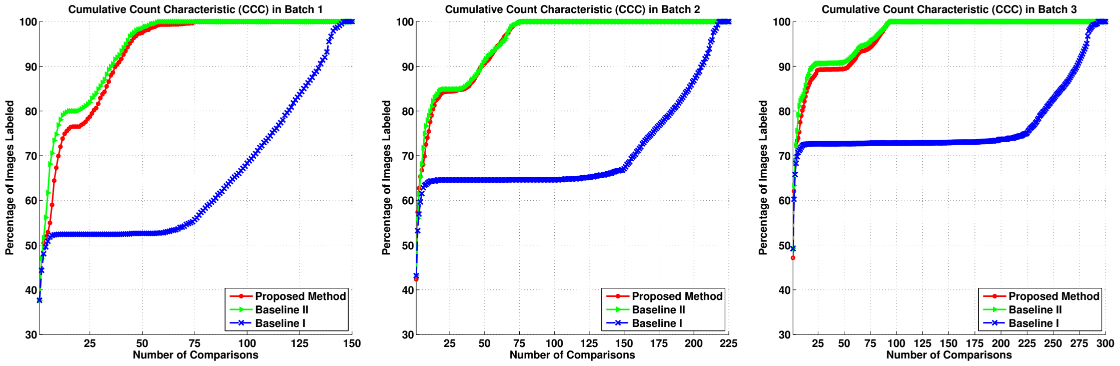
$$\begin{aligned}\mathbf{R}(x) &= \big(1 - P_n(x)\big) \sum_{i=1}^{K} \sum_{\substack{j=1 \\ j\neq i}}^{K} p_x^i \cdot p_x^j + P_n(x) \sum_{j=1}^{K} p_x^j \\ &= \sum_{i=1}^{K} \sum_{\substack{j=1 \\ j\neq i}}^{K} p_x^i \cdot p_x^j + P_n(x)\big(1 - \sum_{i=1}^{K} \sum_{\substack{j=1 \\ j\neq i}}^{K} p_x^i \cdot p_x^j\big)\end{aligned} \tag{1}$$

Ideally, the class membership distribution of an image of a previously unseen person will be more uncertain than an image of a person seen previously. *Shannon entropy* is a measure of uncertainty of an event characterized by its probability distribution. For any image x with class membership distribution $\mathbf{p}_x$, the entropy is given by $H(x) = -\sum_{i=1}^{K} p_x^i \ln p_x^i$. The probability of being a new class can be estimated as a fraction of its entropy compared to the maximum entropy which occurs when the class membership distribution is most uncertain. The maximum value of entropy of an event is characterized by an uniform distribution and the entropy is given by $\ln K$. Thus, $P_n(x)$ is given by,

$$P_n(x) = \frac{-\sum\limits_{i=1}^{K} p_x^i \ln p_x^i}{\ln K} = \frac{\sum\limits_{i=1}^{K} p_x^i \ln \frac{1}{p_x^i}}{\ln K} \tag{2}$$

Using this value of $P_n(x)$ in eqn. (1), the misclassification risk of x can be expressed as,

$$\mathbf{R}(x) = \sum_{i=1}^{K} \sum_{\substack{j=1 \\ j\neq i}}^{K} p_x^i \cdot p_x^j + \frac{\sum\limits_{i=1}^{K} p_x^i \ln \frac{1}{p_x^i}}{\ln K} \Big(1 - \sum_{i=1}^{K} \sum_{\substack{j=1 \\ j\neq i}}^{K} p_x^i \cdot p_x^j\Big) \tag{3}$$



(a) (b) (c)

Fig. 3: CCC curves for the i-LIDS-VIDS dataset. 3(a), (b) and (c) show the comparison count performance for batch 1, 2 and 3 respectively.

Once a query is obtained, the samples from the labeled set of images are presented to the human according to their chances of match to the query. To avoid notational complexity, let us assume that the class membership distribution $\{p_x^1, p_x^2, \cdots, p_x^K\}$ is sorted in order of decreasing value. Sample image from class 1 will be presented to the expert first, then, the sample image from class 2 and so on. Thus, p_x^i also gives the probability of getting a match for x in *exactly* i comparisons. As the expert has to either accept or deny the chosen sample image, the cost of labeling the query, essentially, is proportional to the number of sample images presented before a match is found. Since p_x^i denotes the probability of getting a match in exactly i comparisons, the expected number of comparisons $\mathbf{C}(x)$ is given by $\sum_{i=1}^{K} p_x^i \cdot i$.

The optimum query x^* is to be selected such that on labeling x^*, misclassification risk is reduced maximally at the cost of minimum number of comparisons. Mathematically,

$$x^* = \underset{x}{\operatorname{argmax}} \big(\mathbf{R}(x) - \mathbf{C}(x)\big) \tag{4}$$

Fig. 1 summarizes the proposed active image pair selection framework for continuous person re-identification.

The role of the human in the loop is further extended in the sense that our model learns the way human uses different traits or attributes (*e.g.*, 'having long hair' or 'wearing green colored shirt or not') to discriminate between persons. The attribute information about the already labeled persons is used to choose or discard sample images on top of the order determined by the class membership probabilities. Fig. 2 shows the high level scheme of such use of attributes with active learning. To keep the burden on the human expert to minimum, only the attributes which distinctly discriminates the query and the sample are sought. For a match, finding attributes which differentiates the person from all others is harder. As a result, the human expert is asked to give attribute feedback only for non-matches.

Assume, for a mismatch, the expert identifies the attribute a_q as not present in the query image x while it is present in the sample image from class k. This information is stored against the respective classes in an attribute knowledge base. A short term advantage of the knowledge base is that, before choosing the next image, classes having the same trait as class k with respect to the attribute a_q are removed from being a match to x. This reduces the annotation cost as the expert does not have to judge repetitively on sample images with similar attributes.

Another advantage of the attribute feedback is that it helps in reducing the number of comparisons by building attribute predictors on the way based on this acquired knowledge. Let a set of M binary attribute predictors are trained on M different attributes. Each of the predictors gives a $\{1, 0\}$ output where 1 implies the presence of the attribute and 0 implies otherwise. Let $\mathbf{A}_k = \{a_1^k, a_2^k, \cdots a_M^k\}$ be the set containing the attribute labels for images of class k with an index set $\mathbf{I}_k \subset \{1, 2, \cdots M\}$. $\mathbf{I}_k$ contains the attribute indices which got labels from the expert at any moment for this class. Elements of the set $\mathbf{A}_k$ is defined as,

$$a_i^k \in \begin{cases} \{\phi\} & \text{if } i \notin \mathbf{I}_k, \quad [\{\phi\}\text{denotes a null set}] \\ \{0, 1\} & \text{otherwise} \end{cases} \tag{5}$$

Let the number of such labeled attributes be m_k (*i.e.*, $|\mathbf{A}_k| = m_k$). Let the class membership probabilities of the unlabeled image x provided by the re-identification system be denoted as $\{p_{x,r}^1, p_{x,r}^2, \cdots p_{x,r}^K\}$. These probabilities are modified by running the attribute predictors on x for the attributes in $\mathbf{A}_k$. Let m_x of the predicted attribute values of x match with the corresponding attributes of class k. We employ a majority voting strategy to refine these class membership probabilities to get p_x^k as,

$$p_x^k = \begin{cases} p_{x,r}^k \cdot e^{\frac{m_x}{m_k}} & \text{if } m_x > m_k - m_x \\ p_{x,r}^k \cdot e^{-\frac{m_k - m_x}{m_k}} & \text{if } m_x < m_k - m_x \\ p_{x,r}^k & \text{otherwise} \end{cases} \tag{6}$$

The refined class membership probability values are used to select the most informative query for labeling.

3. EXPERIMENTS

The approach was validated using a new benchmark dataset iLIDS-VID [5] consisting of images of 300 people at an airport arrival hall captured by 2 non-overlapping cameras. Some of the popular datasets (*e.g.*, VIPER), though, have more persons, the number of images per person is too few to suit a continuous framework. We compare the performance

Table 1: Total and average number (per image) of query-sample pair comparisons made by the human to get all the images labeled. The proposed method is close to baseline II. Baseline I requires far more number of comparisons to get all the images labeled than the other two methods.

		batch 1	batch 2	batch 3
i LIDS -VID	Baseline I	16046.8	29868.8	42437.4
	Baseline I (avg)	148.6	184.4	202.1
	Proposed	**3517.2**	**5123.6**	**6620**
	Proposed (avg)	**32.6**	**31.6**	**31.5**
	Baseline II	2960.4	4920.2	5797.4
	Baseline II (avg)	27.4	30.4	27.6

Table 2: Comparison of the proposed method with the state-of-the-art in terms of re-identification accuracy (%).

	Batch 1	Batch 2	Batch 3
Proposed	15.87	24.8	31.07
MS-Color&LBP+DVR [5]	-	-	34.5
MS-Color+DVR [5]	-	-	32.7
MS-SDALF [1]	-	-	6.3
MS-SDALF+DVR [5]	-	-	26.7

with the following two baselines. **Baseline-I** assumes that no attribute information is fed back by the expert. **Baseline-II** assumes that information about every attribute of every unlabeled image is provided as feedback. These two baselines are two extremes where the former assumes no attribute information and the later assumes perfect attribute information for all labeled images. The proposed framework, on the other hand, uses attribute predictors which is incrementally built based on the attribute feedback. This scenario lies in between the two baselines and is validated by the experimental results.

Results are evaluated in terms of the number of comparisons to get all images in each batch labeled. This is shown as a Cumulative Count Curve (CCC) which gives the number of images (%) getting labeled within a certain number of comparisons. As an example, say the number of unlabeled images getting labeled after exactly the first and second comparison be 10 and 5 respectively. So cumulatively the number of images getting labeled within a maximum of 2 tries is 10+5 = 15. The CCC plot, in that case, has 1 and 2 in the x axis corresponding to 10 and 15 in the y axis. As the number of classes vary in each batch we express the y axis in percentage.

Low level features are extracted following the same preprocessing steps as done in [17]. The dataset is divided into 4 batches so that 25% of the total persons are seen for the first time in each batch. Along with the new persons, each batch also contains images from 50% of the persons seen till the previous batch. As a concrete example, the first batch contains images of 75 (25% of 300) people. The second batch contains images of new 75 persons as well as 37 old persons. Similarly, the third and fourth batch contain images of 75 new persons each. At the same time, they have 74 and 111 randomly chosen previously seen persons. The initial training is done on the first batch assuming labeled data but no attribute information. We use a linear Support Vector Machine (SVM) throughout, as the multi-class classifier for person re-identification and the binary classifier for attribute prediction. The toolbox LIBSVM [18] is used for the experimentations.

Fig. 3(a), (b) and (c) compare the percentage cumulative count. It can be seen that as more and more data comes, more and more images are labeled with smaller number of comparisons by the human. In both the baselines 37.65%, 43.14% and 49.19% of the unlabeled images are presented with the true class image as the very first sample image in batch 1, 2 and 3 respectively. That is, 37.65%, 43.14% and 49.19% unlabeled images get their labels within the first comparison. For the rest of the images the number of comparisons increase gradually for baseline I such that it takes upto 144, 217 and 288 comparisons per image to get 99% of the images labeled. These numbers are 53, 70 and 90 when all attribute information are known (baseline II) while the same numbers for the proposed method are 57, 71 and 90 for batch 1, 2 and 3 respectively. It should be noted that the slightly better performance of baseline II comes at the cost of much more human effort. In batch 3, the number of images getting labeled within the first comparison for the proposed method is little less than both the baselines (47.5% vs 49.19%). This is due to the fact that the uncertainty in the attribute predictor affects the class membership distribution of some of the unlabeled images badly such that the probability of true class decreases. But, the catching up of the proposed method with baseline II suggests that the attribute information helps to get more number of images labeled with little effort while affecting a few by increasing the number of comparisons.

Table 1 gives a comparative analysis of the total and average number (per image) of comparisons to label all the unlabeled images for each batch. Table 2 gives a comparative analysis of the test accuracy. The disjoint test set was created using 2 images per person per camera. Though re-identification accuracy of the proposed approach have been reported after each batch of data have been labeled, the comparison with the state-of-the-art can only be done after the framework sees all the persons. It can be seen that the test accuracy increases gradually to reach the state-of-the-art even after starting with a few labeled images.

4. CONCLUSIONS

In this work, we addressed the problem of continuously re-identifying persons starting with a small set of labeled data in an active learning set up. We also showed that mid level attribute based explanations from the human help in reducing the effort of getting labels for unlabeled images. The future directions of our research will be to apply the framework to bigger networks with large numbers of cameras, and cope with wider space-time horizons in a continuous setting.

Acknowledgements: This work was partially supported by NSF grants CNS-1330110 and IIS-1316934.

5. REFERENCES

[1] Loris Bazzani, Marco Cristani, and Vittorio Murino, "Symmetry-Driven Accumulation of Local Features for Human Characterization and Re-identification," *CVIU*, vol. 117, no. 2, 2012. 1, 4

[2] Dong Seon Cheng, Marco Cristani, Michele Stoppa, Loris Bazzani, and Vittorio Murino, "Custom Pictorial Structures for Re-identification," in *BMVC*, 2011, pp. 68.1–68.11. 1

[3] Mert Dikmen, Emre Akbas, Thomas S Huang, and Narendra Ahuja, "Pedestrian Recognition with a Learned Metric," in *ACCV*, 2010, pp. 501–512. 1

[4] Martin Hirzer, Peter M Roth, Martin Kostinger, and Horst Bischof, "Relaxed Pairwise Learned Metric for Person," in *ECCV*, 2012. 1

[5] Taiqing Wang, Shaogang Gong, Xiatian Zhu, and Shengjin Wang, "Person Re-Identification by Video Ranking," in *ECCV*, 2014. 1, 3, 4

[6] Yang Yang, Jimei Yang, Junjie Yan, Shegcai Liao, Dong Yi, and Stan Z. Li, "Salient Color Names for Person Re-identification," in *ECCV*, 2014, pp. 536–551. 1

[7] Ali Farhadi, Ian Endres, Derek Hoiem, and David Forsyth, "Describing Objects by their Attributes," in *CVPR*, 2009. 1

[8] Devi Parikh and Kristen Grauman, "Interactively Building a Discriminative Vocabulary of Nameable Attributes," in *CVPR*, 2011. 1

[9] Devi Parikh and Kristen Grauman, "Relative Attributes," in *ICCV*, 2011, number Iccv. 1

[10] Amar Parkash and Devi Parikh, "Attributes for Classifier Feedback," in *ECCV*, 2012. 1

[11] Arijit Biswas and Devi Parikh, "Simultaneous Active Learning of Classifiers & Attributes via Relative Feedback," in *CVPR*, 2013. 1

[12] Ajay J Joshi, Fatih Porikli, and Nikolaos P Papanikolopoulos, "Scalable Active Learning for Multiclass Image Classification," *PAMI*, vol. 34, no. 11, pp. 2259–2273, 2012. 1

[13] Ryan Layne, Timothy Hospedales, and Shaogang Gong, "Person Re-identification by Attributes," in *BMVC*, 2012. 1

[14] Ryan Layne and Timothy M Hospedales, "Re-id : Hunting Attributes in the Wild," in *BMVC*, 2014. 1

[15] Daniel A Vaquero, Rogerio S Feris, Duan Tran, Lisa Brown, Arun Hampapur, and Matthew Turk, "Attribute-Based People Search in Surveillance Environments," in *WACV*, 2009. 1

[16] Xiao Liu, Mingli Song, Qi Zhao, Dacheng Tao, Chun Chen, and Jiajun Bu, "Attribute-restricted latent topic model for person re-identification," *Pattern Recognition*, vol. 45, no. 12, pp. 4204–4213, Dec. 2012. 1

[17] Abir Das, Anirban Chakraborty, and Amit K. Roy-Chowdhury, "Consistent re-identification in a camera network," in *European Conference on Computer Vision*. 2014, vol. 8690 of *Lecture Notes in Computer Science*, pp. 330–345, Springer. 4

[18] Chih-chung Chang and Chih-jen Lin, "LIBSVM : A Library for Support Vector Machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 27:1—-27:27, 2011, Software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm. 4