Revista Cubana de Ciencias Informáticas
Vol. 14, No. 2, Abril-Junio, 2020
ISSN: 2227-1899 | RNPS: 2301
http://rcci.uci.cu
Pág. 70-82

Artículo original

Single-Shot Re-identificación de persona basada en información saliente no supervisada

Single-Shot Person Re-Identification based on Unsupervised Saliency Information

Reynolds León Guerra [1*] https://orcid.org/0000-0003-3433-1740
Edel B. García Reyes [1] https://orcid.org/ 0000-0002-6426-1264

[1] Advanced Technologies Application Center (CENATAV), Havana, Cuba.

*Autor para la correspondencia. rleon@cenatav.co.cu

RESUMEN

La re-identificación de personas es una tarea importante en video protección para mejorar la seguridad en áreas públicas. En los últimos años existe un gran incremento en las investigaciones sobre este tema. Sin embargo, el desempeño de estos algoritmos es afectado por diferentes problemas presentes en las escenas, por ejemplo, fondos complejos, condiciones atmosféricas y otros. Algunos métodos como el aprendizaje profundo y descriptores de saliencia han sido usados para contrarrestar estos problemas en el mundo real. En el presente artículo, es desarrollado un método basado en la combinación de redes neuronales por comvolución sin aplicar *fine-tuning* y un descriptor de saliencia para ponderar toda la información presente en la imagen de la persona. Los mapas de rasgos son extraídos desde la última capa de convolución de una red neuronal y combinado con otro mapa de saliencia obtenido en el dominio espacial. Finalmente, diferentes rasgos son generados basados en un histograma de color y patrones

binarios locales. Para verificar el desempeño del método propuesto, es validado en la base de datos VIPeR y comparado con otros algoritmos del estado del arte. Los resultados muestran que el método propuesto es fácil de implementar y es comparable con otros métodos usando la curva de correspondencia cumulativa.
Palabras clave: Aprendizaje profundo; Re-identificación de personas; Saliencia; Mapas; Rasgos.

ABSTRACT

Person re-identification task is important in video surveillance to improve security in public place. In recent years there is a lot of investigation about this thematic. However, the performance in these algorithms is affected by different problems in the scenes, for example, complex background, atmospheric conditions, etc. Some methods as deep learning and saliency descriptor have been used to solve these problems in the real world. In this paper, we developed a method based on the combination of convolutional neural network without fine-tuning and a saliency descriptor to weight all the information present into a person image. Feature maps are extracted from the last convolutional layer of a neural network and merged with other salient map obtained in spatial domain. Finally, different features are generated based on color histograms and local binary patterns. To verify the effectiveness of our proposal, the method is validated using VIPeR dataset and compared with others state of the art algorithms. The results shown that our proposal is easy to implement and is comparable with other approach using the Cumulative Matching Characteristic curve.
Keywords: Deep learning; Person re-identification; Saliency; Maps; Features.

Recibido: 17/02/2020

Revista Cubana de Ciencias Informáticas
Vol. 14, No. 2, Abril-Junio, 2020
ISSN: 2227-1899 | RNPS: 2301
http://rcci.uci.cu
Pág. 70-82

Aceptado: 31/03/2020

INTRODUCTION

Person re-identification (**Re-id**) aims to identify a person within camera networks, with non-overlapping viewpoints and in different moments. Also, the **Re-id** is categorized as single-shot (only a person image per camera is used) or multiple-shot (if multiple person image per camera are used). In real world, there is a great variety of smart video surveillance centers using **Re-id** algorithms to increase security in train stations, markets, public places, etc. However, in these places or scenes there are different conditions that affect the performance of the computer vision algorithms. General speaking, the **Re-id** is a challenge task. The problems are usually illumination changes (see Figure 1), pose, occlusion and low resolution (Kansal et al, 2019).

To face these issues aforementioned, in recent years the research has been focused in three aspects. First, to look for hand-crafted features that are robust to the illumination or pose changes (Liao et al, 2015). Second, to learn a metric (no euclidean) such as Mahalanobis distance where the difference inter-class increase and decrease the relation intra-class (Jia et al, 2017). Third, the features are automatically learned using deep learning based on convolutional neural networks (Wang et al, 2017).

Nowadays, in the **Re-id** algorithms have been applied different saliency detection methods. Saliency detection task aims to detect relevant information and reject the redundant information (Álvarez et al, 2018). Zhao (Zhao et al, 2013) use two independent methods to obtain salient region on person image based on KNN (K-Nearest Neighbor) and SVM (Support Vector Machine) of one class. Here, it is used different soft-features, as color histogram and SIFT (Scale Invariant Features Transformation). Niki (Niki et al, 2014) proposed to learn a distance function from a sub-

set of multiple metric learning for features. A salient map is obtained based on graph theory and a color histogram is weighted. Huo (Huo et al, 2015), proposed a weight for the saliency direction trained with a SVM. Li (Li et al, 2018) extract optimal regions with high similarity from person image. After, the salient maps is built into these regions.

On the other hand, the use of the salient information obtained by traditional methods together with deep learning method has good results in the **Re-id** performance. Li (Li et al, 2018b) proposed to learn a Harmonious Attention Convolutional Neural Networks for salient maps with global and local features. Rahimpour (Rahimpour et al, 2017) use a triplet architecture to deep learning where first block is used to obtain a salient map of the person and in second block is obtained the feature representation.

This paper is different to other works, because is developed a method for person re-identification based on detection of salient regions combined with deep learning without fine-tuning. It permits to obtain a final unsupervised salient map. For feature extraction are used the traditional methods. The major contribution of this work is a salient map to obtain a weighted person images (**WPI**) by the combination of two saliency maps, one from deep learning (VVG-F) and another from FqSD algorithm.

COMPUTATIONAL METHODOLOGY

Our Approach

The method is shown in Fgure 2; a convolutional neural network is used to obtain a filter that represent a region salient in the image. This filter is combined with other salient map obtained from FqSD algorithm. Finally, a final salient map is used to weight the information in a person image.

Revista Cubana de Ciencias Informáticas
Vol. 14, No. 2, Abril-Junio, 2020
ISSN: 2227-1899 | RNPS: 2301
http://rcci.uci.cu
Pág. 70-82

Fig. 1 - Samples of different pedestrians captured by two cameras on the VIPeR (Gray et al, 2007) and PRID2011 (Hirzer et al, 2011) datasets. First row is camera A and second row is camera B. Each column indicate the images of the same person.

Weighted Person Image

In deep learning is very known that last convolutional layer represents salient information of the image (Zeiler et al, 2011). The VVG-F is a model (Chatfield et al, 2014) from the Return of the Devil (CNN-F), which were trained with ImageNet dataset (1.2 million of images). During the step of training the filters are learned and named coefficient or weight. For general, these weights learned tend to show semantic information present in an image. In person re-identification, the datasets are homogeneous because only has person images. For the aforementioned the filters in last convolutional layer (in ours experiment are the filters in layer five) generally represent the salient regions of a person.

The VVG-F in layer five has 64 filters. We observe that specifically the filter 38 has the best representation of saliency for person image. The feature map (filter 38) is improved using the transforms, as follows:

Revista Cubana de Ciencias Informáticas
Vol. 14, No. 2, Abril-Junio, 2020
ISSN: 2227-1899 | RNPS: 2301
http://rcci.uci.cu
Pág. 70-82

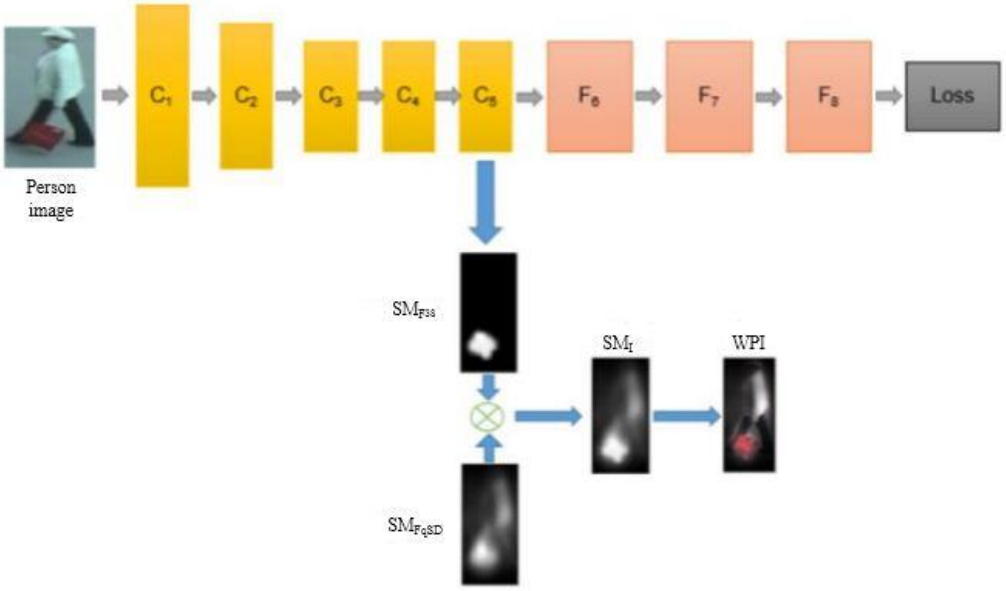

Fig. 2 - Illustration of the different steps used to obtain the WPI.

$$F_n = F_{38}/\max(F_{38}) \quad (1)$$

$$F_s = F_n \star \Gamma \quad (2)$$

$$SM_{F_{38}} = \max(0, F_s) \quad (3)$$

Where, F_{38} is the filter 38 in layer five using the VVG-F, Γ is a radial filter and ($\star$) convolutional product. SM_{F38} is a salient map using filter 38.

Other method to obtain salient map used is the FqSD algorithm (Guerra et al, 2018). This algorithm is based on following aspect: a Gaussian pyramid is applied to get several images with low resolutions. All images are processed to build salient maps using spatial and frequency information in the domain of the quaternions. This algorithm has as output a salient object, we

Revista Cubana de Ciencias Informáticas
Vol. 14, No. 2, Abril-Junio, 2020
ISSN: 2227-1899 | RNPS: 2301
http://rcci.uci.cu
Pág. 70-82

modified it to obtain a salient map of whole person image. In step of the image fusion of the FqSD algorithm is applied a normalization as follows:

$$SM_{FqSD} = IF/\max(IF) \quad (4)$$

Where, SM_{FqSD} is salient map obtained, **IF** is the salient map obtained in image fusion step from FqSD algorithm. To obtain a salient map improved, is necessary to make a fusion between each element obtained in SM_{F38} and SM_{FqSD} as show the expression (5).

$$SM_{I_{(m,n)}} = \begin{cases} 0, & if\ (SM_{F_{38}(m,n)} = 0)\ \ and\ \ (SM_{FqSD(m,n)} = 0) \\ \max\left(SM_{F_{38}(m,n)}, SM_{FqSD(m,n)}\right), & otherwise \end{cases} \quad (5)$$

Where, m and n are spatial coordinates of the image. Finally, $SM_{I(m,n)}$ is multiplied with each channel of the original image to obtain a weighted person image (WPI), see Figure 3.

Experimental design

Our goal is to validate the performance of the developed method on a complex person dataset, applying different mechanics in the extraction of features. VIPeR dataset: It has 1264 images of 632 persons, which were captured by two cameras in an outdoor academic environment. Only one image of the same person appear by camera (see Figure 1). Some characteristic challenges into dataset are: different pose, high illumination changes, background clutter and all person images have size 48x128. Also the images captured have variations from 0 degree to 90 degree (camera A) and from 90 degree to 180 degree (camera B).

Revista Cubana de Ciencias Informáticas
Vol. 14, No. 2, Abril-Junio, 2020
ISSN: 2227-1899 | RNPS: 2301
http://rcci.uci.cu
Pág. 70-82

To validate the performance of our method is applied the next protocol: The Cumulative Matching Characteristic curve (CMC) is used because provide rank-k recognition rate. The dataset is divided to apply the metric learning in training (316)/test (316) and all experiments were run 10 times. To learn a distance based on Mahalanobis are used Keep It Simple and Straightforward Metric (**KISSME**) (Koestinger et al, 2012) and Cross-view Quadratic Discriminant Analysis (**XQDA**) (Liao et al, 2015) approaches. KISSME is a method to obtain a distance from a statistical inference with equivalence constraints and XQDA is an extension of KISSME based cross-view quadratic discriminant analysis, where the metric is learned.

Implementation details: A feature vector (FV_1) is formed from features map in VVG-F using the layer five. Before building the FV_1 is realized a pre-processing to improve these maps as follows:

Fig. 3. Visual samples of the WPI in the second row and in the first row the original person image.

Revista Cubana de Ciencias Informáticas
Vol. 14, No. 2, Abril-Junio, 2020
ISSN: 2227-1899 | RNPS: 2301
http://rcci.uci.cu
Pág. 70-82

$$F_n = F_n + SM_I \quad (6)$$

$$SM_{F_n} = F_n / \max(F_n) \quad (7)$$

Where, F_n are the feature maps present in layer five, n :{ 1,…,64} and SM_{Fn} salient maps improved. After, each SM_{Fn} is divided into eight horizontal rectangular strips and a score (value of saliency) of this area is obtained. Finally, FV_1 has a dimension of 512, note 8x64 = 512.

A second feature vector FV_2 is built using color histograms and Local Binary Patterns (LBP). The color space used are RGB, HSV, normlizedRGB, Lab and Ycbcr. Note, that there are 15 channel, in each one is built a histogram of 256 bins. Moreover, to work with LBP is only used the color space RGB, HSV, normlizedRGB where is built a histogram by color channels but with 64 bins. Equal to FV_1 the images are divided in eight regions, but are only used six regions (first and eighth regions are reject because there is not important information, for example, head and foot). The FV_2 has a dimension of 35 328. In the methods (WIP + Original) is applied the FV_1 and FV_2 an original image and WIP.

RESULTS AND DISCUSSION

We can observe in table 1, using only FV_1 that represent the salient information in each one feature maps of the layer five in VVG-F is obtained a value of 10.28% (KISSME) and 17.03% (XQDA) in rank-1. This result is possible by the intrinsic characteristics that there are in last layer where a similarity with the visual attention mechanism of the human brain has. When the FV_2 is concatenated together with FV_1 the best result is obtained applying XQDA metric with value of 21.68%, but using KISSME the increase in the values is not significant. This result is because the

Revista Cubana de Ciencias Informáticas
Vol. 14, No. 2, Abril-Junio, 2020
ISSN: 2227-1899 | RNPS: 2301
http://rcci.uci.cu
Pág. 70-82

XQDA has major stability for large dimension vectors of features against KISSME especially for the first rank. On the other hand, is not always possible to ensure that the salient region can be visible in each camera viewpoint. We solve this using the combination (WIP + Original). Table 1 shows how the results increase up 19.03% (KISSME) and 24.34% (XQDA). However, the best results to ranks 1, 5, 10 and 20 is to KISSME metric and only rank 1 is to XQDA. In other words, is necessary work with original and WIP person images information to obtain robust features in different cameras.

The comparison with others state of the art algorithms shown that our proposal is competitive in terms of the CMC. Our results are among the last reports in **Re-id**. However, the best results are obtained with deep learning algorithms for person re-identification using training. Note that our proposal is used deep learning without training. The main advantage of the proposed method is to avoid a lot of training time and high dependence of amount data used. But, the disadvantage is the decrease in accuracy in the first rank of the CMC.

Table 1 - Comparison results among feature vectors and state-of-the-arts algorithms on the VIPeR dataset. The values are expressed in percent in the ranks 1, 5, 10 and 20 of the CMC.

Methods	Algorithms	Rank – 1	Rank – 5	Rank – 10	Rank – 20
Ours	WIP + FV_1 + KISSME	10.28	35.54	49.05	65.66
	WIP + FV_1 + XQDA	17.03	37.63	49.08	62.18
	WIP + FV_1 + FV_2 + KISSME	11.55	33.86	48.73	65.98
	WIP + FV_1 + FV_2 + XQDA	21.68	45.22	57.85	70.06
	WIP + Original + KISSME	19.03	**54.91**	**68.04**	**76.27**
	WIP + Original + XQDA	24.34	48.58	60.22	72.85
Others	Hessian (Feng et al ,2019)	20.31	41.35	52.16	68.18
	MSHF (Fang et al, 2018)	20.01	43.67	55.43	68.65
	ZHANG (Zhang and Liu, 2018)	**31.4**	49.2	62.4	74.7
Deep Learning	Cross-GAN (Zhang et al, 2019)	49.28	-	91.66	93.47
	MC-PPMN (Mao et al, 2018)	50.13	81.17	91.46	-

CONCLUSIONS

Without the need to perform a training or fine-tuning for use the VVG-F architecture was possible to extract features to implement a person re-identification algorithm. A final salient map of person image is obtained using a combination between the features maps in layer five of the VVG-F and the improved FqSD algorithm. The extraction features in original and WIP image is the best option to increase performance of the algorithm in VIPeR dataset. In future works, a new combination and feature descriptors will be to increase robust in different scenes. Further, an experiment will be developed with other datasets and different deep learning architectures.

REFERENCES

Kansal, Kajal; Subramanyam, A. V. Hdrnet: Person Re-Identification Using Hybrid Sampling in Deep Reconstruction Network. IEEE Access, 2019, vol. 7, p. 40856-40865.

Liao, Shengcai, et al. Person re-identification by local maximal occurrence representation and metric learning. En Proceedings of the IEEE conference on computer vision and pattern recognition. 2015. p. 2197-2206.

Jia, Jieru, et al. Multiple metric learning with query adaptive weights and multi-task re-weighting for person re-identification. Computer Vision and Image Understanding, 2017, vol. 160, p. 87-99.

Wang, Jiabao; Li, Yang; Miao, Zhuang. (2017). Siamese cosine network embedding for person re-identification. En CCF Chinese Conference on Computer Vision. Springer, Singapore. p. 352-362.

Álvarez-Miranda, Eduardo; Díaz-Guerrero, John. (2018). Multicriteria saliency detection: a (exact) robust network design approach. *Annals of Operations Research*, p. 1-20.

Zhao, Rui; Ouyang, Wanli; Wang, Xiaogang. (2013). Unsupervised salience learning for person re-identification. En Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. p. 3586-3593.

Martinel, Niki; Micheloni, Christian; Foresti, Gian Luca. (2014). Saliency weighted features for person re-identification. En European Conference on Computer Vision. Springer, Cham. p. 191-208.

Huo, Zhonghua; Chen, Ying; Hua, Chunjian. (2015). Person re-identification based on multi-directional saliency metric learning. En International Conference on Computer Vision Systems. Springer, Cham. p. 45-55.

Li, Tiezhu, et al. (2018a). Person re-identification using salient region matching game. Multimedia Tools and Applications, vol. 77, no 16, p. 21393-21415.

Li, Wei; Zhu, Xiatian; Gong, Shaogang.(2018b). Harmonious attention network for person re-identification. En Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. p. 2285-2294.

Rahimpour, Alireza, et al. (2017).Person re-identification using visual attention. En 2017 IEEE International Conference on Image Processing (ICIP). IEEE. p.4242-4246.

Gray, Douglas; Brennan, Shane; Tao, Hai. (2007). Evaluating appearance models for recognition, reacquisition, and tracking. En Proc. IEEE International Workshop on Performance Evaluation for Tracking and Surveillance (PETS). Citeseer. p.1-7.

Hirzer, Martin, et al. (2011) .Person re-identification by descriptive and discriminative classification. En Scandinavian conference on Image analysis. Springer, Berlin, Heidelberg. p. 91-102.

Zeiler, Matthew D., et al. (2011). Adaptive deconvolutional networks for mid and high level feature learning. En ICCV. p. 6.

Chatfield, Ken, et al. (2014) .Return of the devil in the details: Delving deep into convolutional nets. arXiv preprint arXiv:1405.3531.

Guerra, Reynolds León, et al. (2018) . FqSD: Full-Quaternion Saliency Detection in Images. En Iberoamerican Congress on Pattern Recognition. Springer, Cham. p. 462-469.

Koestinger, Martin, et al. (2012). Large scale metric learning from equivalence constraints. En 2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE. p. 2288-2295.

Feng, Guanhua, et al. (2019) .Hessian Regularized Distance Metric Learning for People Re-Identification. Neural Processing Letters, p. 1-14.

Fang, Wen, et al. (2018) . Perceptual hash-based feature description for person re-identification. Neurocomputing, vol. 272, p. 520-531.

Zhang, Chen; LIU, Qiaoling. (2018). Region constraint person re-identification via partial least square on Riemannian manifold. IEEE Access, vol. 6, p. 17060-17066.

Zhang, Chengyuan, et al. (2019) .Crossing generative adversarial networks for cross-view person re-identification. Neurocomputing,.

Mao, Chaojie, et al. (2018) .Multi-channel pyramid person matching network for person re-identification. En Thirty- Second AAAI Conference on Artificial Intelligence.

Conflicto de interés

Los autores autorizan la distribución y uso del presente artículo.

Contribuciones de los autores

Reynolds León Guerra: Su contribución es asociada al desarrollo e implementación de la ida general del artículo materializado en el algoritmo propuesto.

Edel B. García Reyes: Su contribución es en la supervisión y mejoras del algoritmo propuesto en el presente artículo.