Utilizing Semantic Information for Color Histogram-Based Person Re-Identification

Sun, Shanqian
Kyushu University

Inoue, Kohei
九州大学

出版情報：2021-07-29. MIRU2021実行委員会
バージョン：

Utilizing Semantic Information for Color Histogram-Based Person Re-Identification

Shanqian Sun[1,a)] Kohei Inoue[1,b)]

Abstract

Deep learning based Person Re-Identification method has recently attracted a lot of research interests. But it requires a great number of high quality images when sometimes can not be satisfied. In this paper, we transform the RGB color space to our color space based on human's visual experience and calculate histograms of low quality images. Then we use 4 distance metric methods to calculate histogram similarities between each pair of images and match. Moreover, we use semantic information into our matching method. The experimental results show that semantic information is able to improve the accuracy of color histogram based re-id method.

1. Introduction

Person Re-Identification(re-id) refers to matching the same person appearing in different images captured by unrelated cameras. Since it was put forward as a terminology in 2005 [1] and set as an independent research topic [2], re-id has become one of the most popular areas of computer vision, machine learning and artificial intelligence. And it can be widely used in intelligent video surveillance and security assurance.

The basic flow of re-id may proceed as follows: (1) detect person from an image, (2) extract feature, and (3) calculate similarity. Especially, steps (2) and (3) are the core of re-id. The existing re-id methods can be divided into three ways: handcrafted features and metric learning based traditional method, deep learning based method and unsupervised learning based method. The basic idea of handcrafted features based method is to design distinguishable feature with strong robustness, i.e., is able to resist the influence of illumination change, perspective change, occlusion, etc. MJ Swain and DH Ballard proposed histograms based on RGB, HSV and Lab color spaces [3]. Local binary patterns [4], Gabor filters [5] and histograms of oriented gradients [6] are proposed to discriminate texture. The basic idea of metric learning based method is to obtain a metric function, which can calculate the similarity between two images, making the result of same person as small as possible while different person as large as possible. Bryan Prosser et al. proposed a ranking method by Support Vector Machine [7]. Davis et al. proposed a algorithm called Information-Theoretic Metric Learning [8]. And Weinberger K Q et al. proposed Large Margin Neighbor Learning algorithm [9]. In 2012, as AlexNet [10] won the first place in image recognition challenges with a great advantage, Convolutional Neural Networks(CNN) have aroused extensive attention. And Li et al. proposed the Deep Filter Pairing Neural Network [11], which is one of the earliest re-id methods. According to the task requirements, CNN based re-id methods can be divided into two ways: one is to extract feature through classification model; the other is based on verification model. The former takes the same person's images as a category and then trains the networks through loss function, e.g., Li et al. proposed Deep Joint Learning of Multi-Loss Classification [12] and Chen proposed Deep Pyramid Feature Learning [13]. Sililar to the metric learning, the latter train the networks through Contrastive Loss [14], Triplet Loss [15], [16], etc. to calculate the distance between images. Since supervised learning based method requires large-scale annotated data sets, some scholars pay attention to unsupervised learning based methods [17], [18].

The rest of this paper is organized as follows: Section 2 considers an approach to transform the RGB color space to human intuitive sense based color space. Section 3 utilizes semantic information into color histogram based re-id method. Finally, Section 5 concludes this paper.

2. Visual Experience Based Color Space Transform Method

Color images are typically recorded in the format of RGB color channels. However, the RGB color space is not intuitive for human vision. Therefore, we offen need to use the alternative intuitive color spaces such as HSV (hue, saturation, value), HSI(hue, saturation, intensity).

In the HSV model, Fairchild defines hue as the attribute of a visual sensation according to which an area appears to be similar to one of the perceived colors: red, yellow, green, and blue, or to a combination of two of them. And saturation is defined as the colorfulness of a stimulus relative to its own brightness as value is defined as the attribute of a visual sensation according to which an area appears

[1] Kyushu University
[a)] sunshanqian@outlook.com
[b)] k-inoue@design.kyushu-u.ac.jp

to emit more or less light. Figure 1 shows the HSV color model.

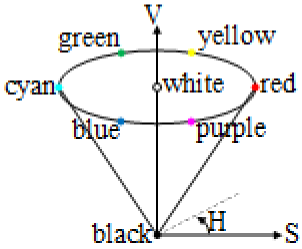

Fig. 1 HSV color model

Then, Androutsos divides the HSV color model into four parts, experimentally [19]. The pixels satisfying value $> 75\%$ and saturation $\geq 20\%$ are classfied as bright chromatic pixels. Then, colors with value $< 25\%$ can be classified as black and colors with saturation $< 20\%$ and value $> 75\%$ can be classified as white. All remaining pixels fall in the chromatic region of the HSV cone. Figure 2 shows the divided HSV color model.

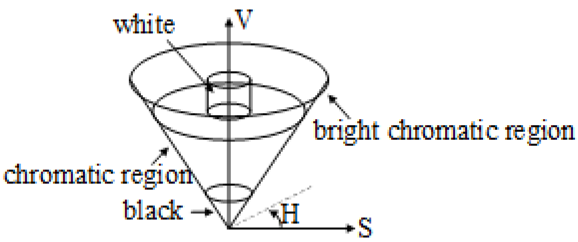

Fig. 2 Androutsos's HSV color model

Finally, we divide bright chromatic region and chromatic region into seven parts(red, orange, yellow, green, cyan, blue, purple), respectively, i.e., the divided HSV color model is divided into 16 parts. Table 1 shows the relationship between HSV and color. It should be noted that the range of H, S and V should be $[0, 360°]$, $[0, 100\%]$ and $[0, 100\%]$, respectively. But different libraries save the data in different ways, e.g., opencv we used range H as integer between 0 to 180, and the ranges of S and V are $[0, 255]$.

According to Table 1, we are able to change the RGB format images into HSV images.

Let $\omega \equiv (\omega_r, \omega_g, \omega_b)$ be an RGB format image with $M \times N$ pixels. Then

$$f_H(\omega) \equiv \begin{cases} \theta, & \text{if } \omega_b \geq \omega_g \\ 360° - \theta, & \text{if } \omega_b > \omega_g \end{cases} \tag{1}$$

with

$$\theta \equiv \cos^{-1}\left\{ \frac{\frac{1}{2}[(\omega_r - \omega_g) + (\omega_r - \omega_b)]}{[(\omega_r - \omega_g)^2 + (\omega_r - \omega_b)(\omega_g - \omega_b)]^{\frac{1}{2}}} \right\} \tag{2}$$

be the hue [20] of ω. Then, define

$$\begin{aligned} m(\omega) &\equiv \min(\omega_r, \omega_g, \omega_b) \\ M(\omega) &\equiv \max(\omega_r, \omega_g, \omega_b) \end{aligned} \tag{3}$$

Table 1 Human visual experience based HSV color space

color	H_m	H_M	S_m	S_M	V_m	V_M
black	0	180	0	255	0	63
white	0	180	0	50	192	255
red	0 156	10 180	0	255	64	191
bright red	0 156	10 180	51	255	192	255
orange	11	25	0	255	64	191
bright orange	11	25	51	255	192	255
yellow	26	34	0	255	64	191
bright yellow	26	34	51	255	192	255
green	35	77	0	255	64	191
bright green	35	77	51	255	192	255
cyan	78	99	0	255	64	191
bright cyan	78	99	51	255	192	255
blue	100	124	0	255	64	191
bright blue	100	124	51	255	192	255
purple	125	155	0	255	64	191
bright purple	125	155	51	255	192	255

we can get

$$f_S(\omega) \equiv 1 - \frac{3m(\omega)}{\omega_r + \omega_g + \omega_b} \tag{4}$$

as the saturation of ω and

$$f_V(\omega) \equiv M(\omega) \tag{5}$$

as the value of ω. Define $\hat{\omega} \equiv (f_H(\omega), f_S(\omega), f_V(\omega))$, we have finished the color space transform from RGB to HSV. The shape of transformed images are $M \times N$, obviously.

We use CUHK02 Dataset to test the transform algorithm. The images in the CUHK02 Dataset are taken in the campus of The Chinese University of Hong Kong. This dataset cover 7264 images captured by 5-pair of cameras with the size of 160×60, including 1816 identities. And every camera takes two photos of one identity. Moreover, most of the people take backpack, umbrella, etc. with them. Figure 3 shows 4 people's images. The left image in every pair is the original image and the right is the transformed image. Actually, the transformed images are also RGB images and set every RGB channels artificially, for observation directly. Table 2 show the relationship between color and RGB value in every channels.

Table 2 Set 16 typical colors' RGB values

color	R channel	G channel	B channel
black	0	0	0
white	255	255	255
red	139	0	0
bright red	255	0	0
orange	255	140	0
bright orange	255	165	0
yellow	255	215	0
bright yellow	255	255	0
green	0	100	0
bright green	0	128	0
cyan	0	139	139
bright cyan	0	255	255
blue	0	0	139
bright blue	0	0	255
purple	148	0	211
bright purple	128	0	128

Fig. 3 Comparison of original images and transformed images

3. Utilize Semantic Information into Re-Id

As what we talk in Section 2, we transform original images into 16 typical colors. And in this section, we try to use color histogram form this color space to calculate the similarities bwtween every two images.

As the color histogram is directly determined by the image, it would vary with the change of shooting angle and selected area. Therefore, we only use the images captured by the first camera of the second cameras-pairs and mark the identity areas manually due to the images in this sub-dataset are the front side or back side of people. And manually marked images have more accuracy than the images obtained through algorithms.

We calculate the histograms of the areas we marked in every images, and after the Histogram Normalization, we obtain normalized histograms in which each bin stands for the occupation of one typical color and ranges $[0, 1]$. And the sum of bins is 1, obviously.

Let $H \equiv (h_1, h_2, ..., h_{n-1}, h_n)$ stands as the normalized histogram of every images, h_k stands as each bin for $k = 1, 2, ..., n-1, n$, and $H(j) \equiv h_j$. Define $\bar{H} \equiv \frac{1}{n}\sum_{j=1}^{n} h_j$, let

$$d(H_1, H_2) \equiv \frac{\sum_j (H_1(j) - \bar{H}_1)(H_2(j) - \bar{H}_2)}{\sqrt{\sum_j (H_1(j) - \bar{H}_1)^2 \sum_j (H_2(j) - \bar{H}_2)^2}} \quad (6)$$

be Correlation Distance, ranging from -1 to 1, the larger the value, the higher the similarity. And let

$$d(H_1, H_2) \equiv \sum_j \frac{(H_1(j) - H_2(j))^2}{H_1(j)} \quad (7)$$

be Chi-Square Distance, result in nonnegative number, and the smaller the value, the higher the similarity. Then, let

$$d(H_1, H_2) \equiv \sum_j \min(H_1(j), H_2(j)) \quad (8)$$

be Intersection Distance also result in nonnegative number, but the larger the value, the higher the similarity. Finally, let

$$d(H_1, H_2) \equiv \sqrt{1 - \frac{1}{\sqrt{\bar{H}_1 \bar{H}_2 n^2}} \sum_j \sqrt{H_1(j) H_2(j)}} \quad (9)$$

be Bhattacharyya Distance ranging from 0 to 1, while the smaller the value, the higher the similarity.

We divide 306 pair images into 6 groups and calculate the similarities between each pair images and match by the sort of similarities. Table 3 shows the accuracy, in which the number stands as the index of image pairs. We find that the Bhattacharyya Distance based re-id method have the highest accuracy in every groups and with the expansion of images' amount, the accuracy is decreasing.

Table 3 Matching accuracy of different groups with different distance metric methods

group	number	Cor. D.	Chi. D.	I. D.	B. D.
1	1-100	42.00%	46.00%	43.00%	54.00%
2	101-200	37.00%	42.00%	42.00%	51.00%
3	201-306	48.11%	53.77%	45.28%	56.60%
4	1-200	31.00%	37.00%	35.50%	43.50%
5	101-306	32.04%	41.26%	32.52%	44.66%
6	1-306	26.47%	33.33%	30.07%	37.91%

Let's take a look at the result of the Correlation Distance based matching method. As the image quantity become bigger and bigger, the space of similarity between image pairs become smaller and smaller, leading the accuracy decreasing. Table 4 shows the similarity of image pairs(part).

Table 4 Similarities of image pairs in group 6(Part)

rank	image pair	match or unmatch	similarity
1	(251,252)	match	0.999851
2	(61,97)	unmatch	0.999849
3	(285,521)	unmatch	0.999791
4	(62 437)	unmatch	0.999720
5	(209,210)	match	0.999667
6	(291,325)	unmatch	0.999656
7	(103,472)	unmatch	0.999454
8	(283,284)	match	0.999399
9	(139,257)	unmatch	0.999295
10	(469,470)	match	0.999209

Figure 4 shows some unmatch image pairs, i.e., one of the person in the top left image pair is male but the other is female. Moreover, people can be separate from shooting angle(top right), belongings such as bag(bottom left) or umbrella(bottom right).

So, we divide each of the 6 groups into 3 tiny groups according to semantic information, i.e., we separate the images according to sex, side(front or back) and both. Table 5 shows the result. It is obvious that comparing with the original groups, semantic information used groups have higher accuracy, and the more the information is, the higher the accuracy becomes.

Fig. 4 unmatch image pairs

Table 5 Matching accuracy combining while utilizing semantic information

group	sem. info.	Cor. D.	Chi. D.	I. D.	B. D.
1	sex	44.00%	50.00%	47.00%	60.00%
	side	45.00%	53.00%	47.00%	57.00%
	both	54.00%	57.00%	59.00%	62.00%
2	sex	46.00%	50.00%	51.00%	54.00%
	side	40.00%	52.00%	51.00%	62.00%
	both	57.00%	58.00%	60.00%	63.00%
3	sex	51.89%	57.55%	51.89%	62.26%
	side	52.83%	64.15%	52.83%	66.04%
	both	57.55%	69.81%	57.55%	73.58%
4	sex	39.50%	41.50%	43.50%	50.00%
	side	34.00%	42.50%	39.50%	50.00%
	both	43.00%	48.00%	48.00%	57.00%
5	sex	34.95%	41.75%	35.92%	47.57%
	side	35.92%	49.03%	42.23%	53.40%
	both	44.66%	51.46%	45.15%	59.22%
6	sex	33.66%	38.56%	35.29%	43.14%
	side	32.03%	38.89%	35.95%	44.44%
	both	40.52%	42.81%	43.14%	50.33%

4. Conclusion and Future Work

This work proves the effectivity of using semantic information for improving the accuracy in person re-id. According to human's intuitive sense, we replace the original colors with 16 typical colors and match images due to color histogram in this color space. Moreover, by utilizing the semantic information, we improve the accuracy at last.

Many open questions have been raised that we want to answer in our future research. In this research, we mark the person edge manually, as it affects the color histogram directly. So, we consider a method to separate person area from the background. Moreover, our research aims at the situation in which deep learning based method can not work well due to the bad image quality, so selecting and detecting the semantic information is really important.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number JP21K11964.

References

[1] Zajdel W , Zivkovic Z , Krose B . Keeping Track of Humans: Have I Seen This Person Before?[C] // Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on. IEEE, 2005.

[2] N. Gheissari, T. B. Sebastian and R. Hartley, "Person Reidentification Using Spatiotemporal Appearance," 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), 2006, pp. 1528-1535, doi: 10.1109/CVPR.2006.223.

[3] Swain M J , Ballard D H . Color indexing[J]. International Journal of Computer Vision, 1991, 7(1):11-32.

[4] Ojala, Timo, Pietikainen, et al. Multiresolution Gray-Scale and Rotation Invariant Texture and Classification with Local Binary Patterns.[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2002, 24(7):971-971.

[5] Fogel I , Sagi D . Gabor filters as texture discriminator[J]. Biological Cybernetics, 1989, 61(2):103-113.

[6] Dalal N , Triggs B . Histograms of Oriented Gradients for Human Detection[C] // IEEE Computer Society Conference on Computer Vision & Pattern Recognition. IEEE, 2005.

[7] Bryan Prosser, Wei-Shi Zheng, Shaogang Gong, and Tao Xiang. Person Re-Identification by Support Vector Ranking. In Frédéric Labrosse, Reyer Zwiggelaar, Yonghuai Liu, and Bernie Tiddeman, editors, Proceedings of the British Machine Vision Conference, pages 21.1-21.11. BMVA Press, September 2010. doi:10.5244/C.24.21.

[8] Davis, Jason Kulis, Brian Jain, Prateek & Sra, Suvrit Dhillon, Inderjit. (2007). Information-theoretic metric learning. ACM International Conference Proceeding Series. 227. 209-216. 10.1145/1273496.1273523.

[9] Weinberger K Q , Saul L K . Distance Metric Learning for Large Margin Nearest Neighbor Classification[J]. Journal of Machine Learning Research, 2009, 10(1):207-244.

[10] Krizhevsky A , Sutskever I , Hinton G . ImageNet Classification with Deep Convolutional Neural Networks[J]. Advances in neural information processing systems, 2012, 25(2).

[11] W. Li, R. Zhao, T. Xiao and X. Wang, "DeepReID: Deep Filter Pairing Neural Network for Person Re-identification," 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 152-159, doi: 10.1109/CVPR.2014.27.

[12] Li W , Zhu X , Gong S . Person Re-Identification by Deep Joint Learning of Multi-Loss Classification[C] // Twenty-Sixth International Joint Conference on Artificial Intelligence. arXiv, 2017.

[13] Chen Y , Zhu X , Gong S , et al. Person Re-identification by Deep Learning Multi-scale Representations[C] // 2017 IEEE International Conference on Computer Vision Workshops (ICCVW). IEEE, 2017.

[14] Varior, Rahul Haloi, Mrinal & Wang, Gang. (2016). Gated Siamese Convolutional Neural Network Architecture for Human Re-Identification.

[15] F. Schroff, D. Kalenichenko and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 815-823, doi: 10.1109/CVPR.2015.7298682.

[16] D. Cheng, Y. Gong, S. Zhou, J. Wang and N. Zheng, "Person Re-identification by Multi-Channel Parts-Based CNN with Improved Triplet Loss Function," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 1335-1344, doi: 10.1109/CVPR.2016.149.

[17] Farenzena M , Bazzani L , Perina A , et al. Person re-identification by symmetry-driven accumulation of local features[J]. Proceedings / CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2010:2360-2367.

[18] Liao S , Yang H , Zhu X , et al. Person re-identification by Local Maximal Occurrence representation and metric learning[C]// 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2015.

[19] Androutsos D , Plataniotis K N , Venetsanopoulos A N . Vector angular distance measure for indexing and retrieval of color[J]. Proc Spie, 1998(3656):604-613.

[20] R. C. Gonzalez and R. E. Woods, Digital Image Processing, 3rd ed., Prentice Hall, p. 411, 2007.