Journal of Southeast University (English Edition) Vol. 28, No. 4, pp. 422 - 427 Dec. 2012 ISSN 1003—7985

Integration of Lab model and EHOG for human appearance matching across disjoint camera views

Yang Biao Lin Guoyu Zhang Weigong

(School of Instrument Science and Engineering, Southeast University, Nanjing 210096, China)

Abstract: The integration of the Lab model with the extended histogram of oriented gradients (EHOG) is proposed to improve the accuracy of human appearance matching across disjoint camera views under perturbations such as illumination changes and different viewing angles. For the Lab model that describes the global information of observations, a sorted nearest neighbor clustering method is proposed for color clustering and then a partitioned color matching method is used to calculate the color similarity between observations. The Bhattacharya distance is employed for the textural similarity calculation of the EHOG which describes the local information. The global information, which is robust to different viewing angles and scale changes, describes the observations well. Meanwhile, the use of local information, which is robust to illumination changes, can strengthen the discriminative ability of the method. The integration of global and local information improves the accuracy and robustness of the proposed matching approach. Experiments are carried out indoors, and the results show a high matching accuracy of the proposed method.

Key words: human appearance matching; Lab model; extended histogram of oriented gradients (EHOG); disjoint camera views

doi: 10.3969/j. issn. 1003 - 7985. 2012. 04. 009

The need for automated surveillance tracking is rising with the increasing use of camera surveillance in public areas such as airports, shopping centers and office buildings where all the cameras share non-common areas. The key of tracking is accurate object matching across disjoint camera views. Template-based matching and appearance-based matching are two commonly used matching approaches with many successful applications in single camera monitoring. However, in disjoint camera views, the matching performance may be influenced by various perturbations, including illumination changes and different viewing angles.

Received 2012-06-04.

Biographies: Yang Biao(1987—), male, graduate; Zhang Weigong(corresponding author), male, doctor, professor, zhangwg@seu.edu.cn.

Foundation items: The National Natural Science Foundation of China (No. 60972001), the Science and Technology Plan of Suzhou City (No. SG201076).

Citation: Yang Biao, Lin Guoyu, Zhang Weigong. Integration of Lab model and EHOG for human appearance matching across disjoint camera views[J]. Journal of Southeast University (English Edition), 2012, 28 (4): 422 - 427. [doi: 10.3969/j. issn. 1003 - 7985. 2012. 04. 009]

Many matching methods have been proposed, but not all of them can be used across disjoint camera views due to such factors as lighting conditions and viewing angles. Huang and Russell[1] used an RGB color histogram with time information for object matching on a highway. However, the sensitivity to illumination changes makes it improper for appearance matching across disjoint views. HSV is also commonly used for appearance matching[2]. It is robust to illumination changes due to the separation of the luminance component from the chromaticity. However it is not reliable in dark environments. The brightness transfer functions (BTFs) are widely used for mapping a known brightness value in one camera to the corresponding observation in another camera[3-5]. But the large amount of training restricts most of the applications of the BTFs. Except for the appearance-based matching approaches mentioned above, there are also some template-based matching approaches[6-8]. However, the requirements of continuous learning make them unsuitable for our applications which are discontinuous in both temporal and spatial fields.

As discussed above, the appearance-based matching method can be used in our application. However, the selected features should guarantee the robustness to such perturbations such as illumination changes and different viewing angles. Real-time performance is another point which must be considered. The flowchart of human appearance matching based on the Lab model and the EHOG is shown in Fig. 1. Some pretreatments are done initially and then color spaces of observations are converted. The color matching rate P_{C} and the textural matching rate P_{E} are calculated, respectively. The final matching rate P_{F} can be defined as

$$P_{\mathrm{F}} = \begin{cases} 0 & P_{\mathrm{C}} < T_1 \text{ or } P_{\mathrm{E}} < T_2 \\ \alpha P_{\mathrm{C}} + (1-\alpha) P_{\mathrm{E}} & \text{otherwise} \end{cases} \tag{1}$$

where T_1 and T_2 are two given thresholds. The weight α reflects the illumination changes (e. g., a big α indicating a general illumination change, while a small one indicating a strong illumination change).

1 Pretreatments

1.1 White balancing

An automated white balancing method[9] is adopted for a strong color cast of the ambient illuminant (e. g., the

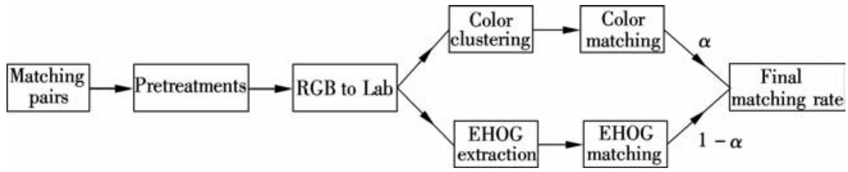

Fig. 1 Flowchart of the proposed algorithm

familiar yellow cast of indoor incandescent light). The pretreatment with white balancing alleviates the influence of color distortion, thus improving the matching accuracy.

1.2 Gamma correction

Illumination compensation is useful for appearance matching in very dark or very bright conditions. The gamma correction[10] is employed for illumination compensation in this paper. The relationship between the gamma value $\gamma(v)$ and the corresponding gray value v can be written as

$$\gamma(v)=\begin{cases}1+0.5(v_1-v)/(v_1-M) & v\in[M, v_1]\\ 1 & v\in[v_1, v_2]\\ 1-0.5(v-v_2)/(H-v_2) & v\in[v_2, H]\end{cases} \tag{2}$$

where M and H represent the lowest and the highest value of a certain color. In the low value section $[M, v_1]$, $\gamma(v)$ is greater than 1, thus the output is brightened. In the high value section $[v_2, H]$, $\gamma(v)$ is smaller than 1, thus the output is darkened. Finally the gamma correction value v_C can be defined as

$$v_C=(H-M)\left(\frac{v}{H-M}\right)^{1/\gamma(v)} \tag{3}$$

In the Lab model, the L component is significantly influenced by illumination changes among the three independent components. Thus only the L component is corrected by the gamma correction.

2 Color Appearance Matching

2.1 Color conversion from RGB to Lab model

The RGB is commonly used in pattern recognition due to its convenience in single camera views. However, in disjoint camera views, it is inaccurate to describe objects using their RGB information due to the illumination changes. The Lab model is used instead of the RGB due to its robustness to illumination changes and the property of being device-independent.

The Lab model is widely used in image segmentation. However, few works are found about the applications of the Lab model in appearance matching to the best of our knowledge. The advantages of the Lab model like homogeneity and device-independence make it a good feature for matching. The Lab model is constituted by three independent components, including the luminance channel L and two color channels a and b.

The transformation from the RGB to the Lab model is based on a XYZ color space which acts like a bridge. The XYZ values can be defined as

$$\begin{bmatrix}X\\Y\\Z\end{bmatrix}=\begin{bmatrix}0.430\,57 & 0.222\,02 & 0.020\,18\\ 0.341\,55 & 0.706\,66 & 0.129\,55\\ 0.178\,33 & 0.017\,33 & 0.939\,18\end{bmatrix}\begin{bmatrix}R\\G\\B\end{bmatrix} \tag{4}$$

After obtaining the values of X, Y, and Z, the parameter values of the Lab model can be defined as

$$L=116f\left(\frac{Y}{Y_n}\right)-16$$
$$a=500\left(f\left(\frac{X}{X_n}\right)-f\left(\frac{Y}{Y_n}\right)\right)$$
$$b=200\left(f\left(\frac{Y}{Y_n}\right)-f\left(\frac{Z}{Z_n}\right)\right)$$
$$f(t)=\begin{cases}t^{1/3} & t>0.008\,856\\ 7.787t+0.137\,93 & t\leqslant 0.008\,856\end{cases}$$

where X_n, Y_n, Z_n are three stimulating values of white light.

The chroma value C and the hue value h can be defined as

$$C=\sqrt{a^2+b^2} \tag{5}$$

$$h=\arctan\frac{b}{a} \tag{6}$$

2.2 Color clustering

Clustering methods are used to decrease computational complexity by merging similar features. In this paper, a sorted nearest neighbor clustering (SNNC) method which pays no attention to initial centers and the number of clusters is proposed for color clustering. Pixels with the same values are merged as initial clusters. We assume that O is the queue of M clusters, $O=\{O_i: i=1,2,\ldots,M\}$. O_i is the i-th cluster in O and n_i represents its pixel number. The detail is shown in Algorithm 1.

Algorithm 1 The proposed SNNC method

1) If O_i is the last cluster of the queue, the SNNC is ceased. Otherwise the cluster after O_i in the queue is chosen as O_j and step 2) is executed.

2) If O_j is the last cluster of the queue, the cluster after the original O_i is chosen as the new O_i and step 1) is executed. Otherwise, calculate the distance $d(i,j)$ between O_i and O_j. The $d(i,j)$ is defined as

$$d(i,j)=\sqrt{\left(\frac{L_i-L_j}{S_L}\right)^2+\left(\frac{C_i-C_j}{S_C}\right)^2+\left(\frac{h_i-h_j}{S_h}\right)^2} \tag{7}$$

where S_L, S_C and S_h are the weights for the lightness, chroma and hue components, respectively.

3) Compare $d(i,j)$ with a given threshold σ_1. If $d(i,j) < \sigma_1$, the cluster O_j is merged into O_i. Then O_j is removed from the queue and the updated process is executed. The updated process is defined as

$$\left.\begin{aligned} L_i &= \frac{n_i}{n_i+n_j} L_i + \frac{n_j}{n_i+n_j} L_j \\ C_i &= \frac{n_i}{n_i+n_j} C_i + \frac{n_j}{n_i+n_j} C_j \\ h_i &= \frac{n_i}{n_i+n_j} h_i + \frac{n_j}{n_i+n_j} h_j \\ n_i &= n_i + n_i \end{aligned}\right\} \tag{8}$$

4) The cluster after the original O_j is chosen as the new O_j and step 2) is executed.

All the new clusters should be sorted in a descending order according to their pixel numbers after one SNNC. The whole clustering will not stop until the pixel numbers of the first few clusters exceed 95% of the total pixel numbers.

2.3 Color matching rate

The color matching rate can be calculated using the colors after clustering. We assume that the numbers of the major colors extracted from observations A and B are K_A and K_B, respectively. Colors in observation A can be written as $H(A) = \{H(A_i): i = 1, 2, \ldots, K_A\}$. The frequencies of colors in A can be written as $p(A) = \{p(A_i): i = 1, 2, \ldots, K_A\}$, while $p(A_i)$ is calculated using the ratio of the pixel number of $H(A_i)$ to that of H_A. Similarly, H_B and $p(B)$ can be written as $H_B = \{H(B_j): j = 1, 2, \ldots, K_B\}$ and $p(B) = \{p(B_j): j = 1, 2, \ldots, K_B\}$.

The similarity between $H(A_i)$ and $H(B_j)$ can be defined as

$$p(H(A_i),\ H(B_j)) = 1 - \frac{D(i,j)}{D_M} \tag{9}$$

where D_M is the longest distance between different clusters. $D(i,j)$ is the distance between $H(A_i)$ and $H(B_j)$ and it can be defined as

$$D(i,j) = \sqrt{\left(\frac{L(A_i) - L(B_j)}{\lambda S_L}\right)^2 + \left(\frac{C(A_i) - C(B_j)}{S_C}\right)^2 + \left(\frac{h(A_i) - h(B_j)}{S_h}\right)^2} \tag{10}$$

where λ is the illumination suppressing factor and it is used to restrain illumination changes.

The similarity between $H(A_i)$ and H_B can be defined as

$$P(H(A_i), H_B) = \frac{\sum_{j=1}^{K_B} P(H(A_i), H(B_j)) P(B_j)}{\sum_{j=1}^{K_B} P(B_j)} \tag{11}$$

Finally the similarity between H_A and H_B can be written as

$$P_C(H_A, H_B) = \frac{\sum_{i=1}^{K_A} P(H(A_i), H_B) P(A_i)}{\sum_{i=1}^{K_A} P(A_i)} \tag{12}$$

A partitioned color matching method is used to calculate the final color matching rate. The observation to be matched is subdivided into two sections, including the upper part and the lower part. We assume that starting from the bottom, the first 55% is defined as the lower part and the next 30% is defined as the upper part, according to the known anatomical proportions. The remaining 15% is defined as the head part which is abandoned for convenience. Two color matching rates, P_U and P_L, are calculated, respectively. Then the final color matching rate P_C can be defined as

$$P_C = \begin{cases} 0 & \text{if } P_U < T \text{ or } P_L < T \\ \dfrac{P_U + P_L}{2} & \text{otherwise} \end{cases} \tag{13}$$

where T is the threshold which indicates the rejection rate of color matching.

3 Textural Appearance Matching

3.1 Texture extraction

The textural information is extracted based on the histograms of oriented gradients (HOG)[11]. We propose an EHOG which extends the HOG from the traditional grayscale to the Lab model. Some pretreatments should be done initially. A minimum bounding rectangle containing the detected object is rescaled to a normalized height (e.g. 128 pixels) due to the scale change under different viewing angles. The mask region is obtained based on the detected foreground. The foreground is eroded two times to alleviate the influence of the outer contours.

The HOG descriptors are extracted from the mask region in a, b components, respectively. Suppose that $a(x,y)$ and $b(x,y)$ are pixel values in a, b components. Then the gradient value $T_a(x,y)$ and the gradient direction $\theta_a(x,y)$ can be written as

$$T_a(x,y) = \sqrt{(a(x+1,y) - a(x-1,y))^2 + (a(x,y+1) - a(x,y-1))^2} \tag{14}$$

$$\theta_a(x,y) = \tan^{-1}\left(\frac{a(x,y+1) - a(x,y-1)}{a(x+1,y) - a(x-1,y)}\right) \tag{15}$$

$T_b(x,y)$ and $\theta_b(x,y)$ can be calculated in the same way.

3.2 Textural matching rate

The EHOG is obtained once $T_a(x,y)$, $\theta_a(x,y)$, $T_b(x,$

$y)$ and $\theta_b(x, y)$ are calculated. Compared with the HOG, the color information in the EHOG makes it more robust. The texture matching rate P_E is calculated based on the Bhattacharya distance, which is defined as

$$P_E = \sqrt{1 - \sum_{u=1}^{K} \sqrt{q(A)_u q(B)_u}} \tag{16}$$

where $q(A)_u$ and $q(B)_u$ are the u-th bins of observations A and B in the relevant EHOG.

4 Experiments

We did experiments indoor to test the proposed human appearance matching algorithm across disjoint camera views. Experiments were done under four different scenes, and with each scene a camera was located. The surveillance camera in our experiment is DS-2CD853F with 1 600 × 1 200 high resolution and an acquisition frame rate of 12 frame/s. The proposed matching algorithm ran in VC + + 6. 0 with P4 2. 8 GHz, with the help of OPENCV1. 0. Fig. 2 shows the camera network topology in our experiments and non-common areas are shared by four cameras.

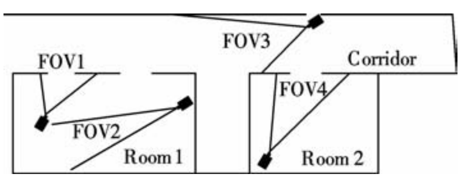

Fig. 2 Camera network topology in experiment

Several people walked in the four disjoint camera views under different illuminations. A frame difference-based tracking approach with space-time restriction was employed for tracking in single camera view. When one was detected by a camera (except the first one), a matching was done using the proposed matching algorithm to determine whether the object was an existing one or a new one. Fig. 3 shows the results of human appearance matching.

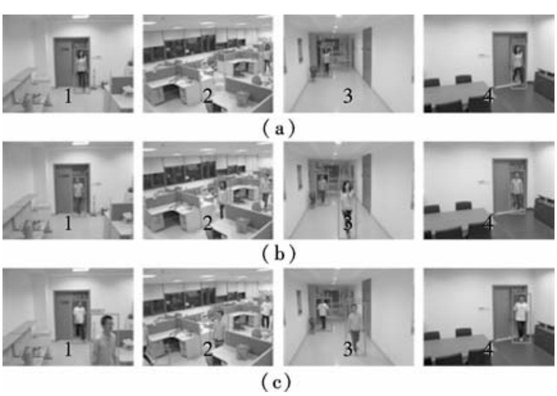

Fig. 3 Human appearance matching results. (a) Object No. 1; (b) Object No. 2; (c) Object No. 3

The advantages of the proposed method are specified in Fig. 4 and Fig. 5. Comparisons are made among our method and some commonly used methods (e. g., the HSV and the RGB methods, along with the margin and the shape methods). The Lab model, the HSV and the RGB are regarded as global features which are robust to different viewing angles. The EHOG, the margin and the shape are regarded as local features which are more robust to illumination changes. In our experiments, the RGB and the HSV are represented by the corresponding histograms. The margin is extracted using Canny operators. The shape is represented by the aspect ratio and the foreground area of the observation.

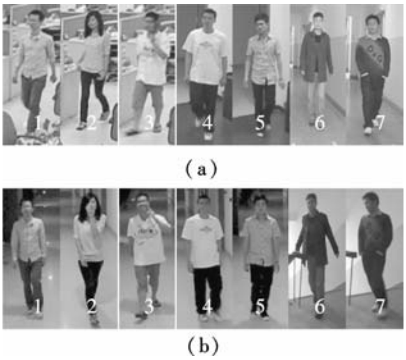

Fig. 4 The same observations. (a) Observation A; (b) Observation B

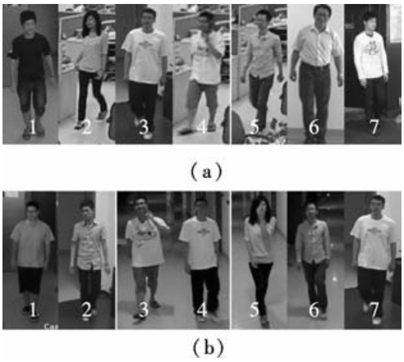

Fig. 5 Different observations. (a) Observation A; (b) Observation B

Tab. 1 and Tab. 2 show the similarities of the matched observations calculated by methods using global and local features. From the data in Tab. 1, we can find that the Lab model owns the highest similarities under different illuminations due to the separation of luminance component L. The HSV method plays worse than the Lab model due to the distortion of H and S in dark conditions. The RGB method is the most sensitive to illumination changes, thus it has the lowest similarities. For the 6th and the 7th columns, all the color matching methods play poorly due to the drastic illumination changes. We cannot distinguish the objects in color, even by eye, thus other features are necessary as an assistant. In Tab. 2, the EHOG method matches observations well as shown in the 4th and the 5th columns. It performs well even under drastic illumination changes as shown in the 6th column. The margin method is not suitable for non-rigid object matching due to the variability of the edge. The shape method can be also regarded as a candidate for its high similarities.

Tab. 1 Similarity of the same observations in global features of four different methods

Method	Observation						
	1	2	3	4	5	6	7
Proposed method	0. 806	0. 766	0. 747	0. 895	0. 864	0. 372	0. 433
Lab model	0. 858	0. 832	0. 812	0. 892	0. 854	0. 269	0. 325
HSV	0. 785	0. 796	0. 805	0. 822	0. 763	0. 152	0. 235
RGB	0. 615	0. 636	0. 605	0. 713	0. 616	0. 115	0. 218

Tab. 2 Similarity of the same observations in local features of four different methods

Method	Observation						
	7	1	2	3	4	5	6
Proposed method	0. 806	0. 766	0. 747	0. 895	0. 864	0. 372	0. 433
EHOG	0. 683	0. 612	0. 595	0. 902	0. 886	0. 613	0. 685
Margin	0. 232	0. 286	0. 213	0. 322	0. 315	0. 216	0. 256
Shape	0. 722	0. 653	0. 751	0. 812	0. 795	0. 752	0. 696

Tab. 3 and Tab. 4 show the similarities of unmatched observations owning similar color information. In Tab. 3, color information is distinguishable in most cases except for the 2nd and the 7th columns, where different observations own almost the same color information. As shown in the 2nd and the 7th columns in Tab. 4, texture information is helpful for pairs with insufficient color information to provide a definite classification. From Tab. 4 we can also find that the similarities of the shape method are high even to unmatched pairs. Thus the shape method is discarded for its low discriminative ability.

Tab. 3 Similarity of different observations in global features of four different methods

Method	Observation						
	1	2	3	4	5	6	7
Proposed method	0. 265	0. 554	0. 095	0. 425	0. 107	0. 079	0
Lab model	0. 312	0. 656	0. 496	0. 483	0. 428	0. 313	0. 735
HSV	0. 328	0. 684	0. 503	0. 452	0. 416	0. 356	0. 726
RGB	0. 298	0. 533	0. 565	0. 413	0. 395	0. 322	0. 696

Tab. 4 Similarity of different observations in local features of four different methods

Method	Observation						
	1	2	3	4	5	6	7
Proposed method	0. 265	0. 554	0. 095	0. 425	0. 107	0. 079	0
EHOG	0. 882	0. 316	0. 285	0. 288	0. 356	0. 433	0. 215
Margin	0. 312	0. 284	0. 213	0. 211	0. 285	0. 316	0. 246
Shape	0. 812	0. 713	0. 751	0. 672	0. 795	0. 712	0. 756

The proposed matching method is also compared with some other existing matching algorithms, such as the IMCSHR method[12] and the time weighted color (TWC) matching method[13]. 200 pairs of observations are matched with manual supervision. Results in Tab. 5 indicate a higher matching accuracy of the proposed method than that of the other two.

The proposed method is robust to illumination changes due to the pretreatments and the L inhibition. The importance of pretreatments is shown in Fig. 6. Twenty pairs of matched observations are used under different illuminations. Three different approaches, including using both pretreatments, using white balancing only and using gamma correction only, are represented by circle, star and rectangle, respectively. From Fig. 6, it indicates that using both pretreatments is much better than using only one pretreatment. The importance of the L inhibition is shown in Fig. 7 and Fig. 8. Fifteen pairs of observations are used with different illumination suppressing factor λ. The result of Fig. 7 indicates that a small λ decreases the matching rate of matched pairs, while the result of Fig. 8 indicates that a great λ will increase the matching rate of unmatched pairs.

Tab. 5 Accuracy rates of different methods

Method	Correct matches	Wrong matches	Accuracy rates/%
Proposed method	200	12	94
IMCSHR	200	32	84
TWC	200	28	86

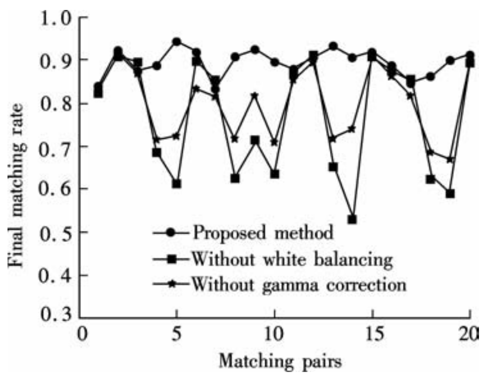

Fig. 6 Effects of pretreatments on illumination changes

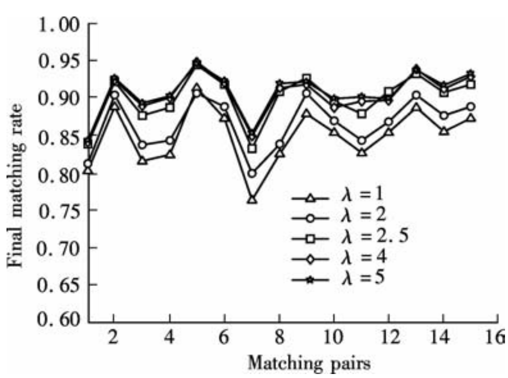

Fig. 7 Final matching rates of matched pairs with different λ

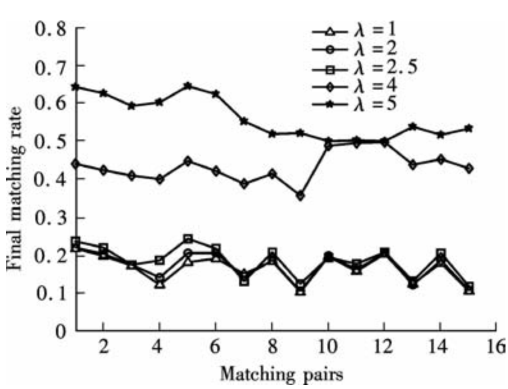

Fig. 8 Final matching rates of unmatched pairs with different λ

5 Conclusion

A new method for human appearance matching across disjoint camera views is proposed in this paper. The contribution of our work focuses on incorporating the Lab model and the EHOG together for accurate human appearance matching under illumination changes. The separation of the L component from chromaticity guarantees the robustness of the Lab model to illumination changes. Meanwhile, color information in the EHOG provides stronger discriminative ability to describe local information. Finally the experimental results show that our method can achieve a high matching accuracy with acceptable time consumption. Our future work will focus on improving the matching performance under violent illumination changes. Meanwhile, adaptive adjustment of weight α in the final matching rate calculation will be further studied.

References

[1] Huang T, Russell S. Object identification in a Bayesian context [C]//*Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence*. San Francisco, CA, USA, 1997: 1276 − 1282.

[2] Huang J G, Kong B. A new method of unstructured road detection based on HSV color space and road features [C]//*International Conference on Information Acquisition*. Seogwipo-si, South Korea, 2007: 596 − 601.

[3] Andrew G, Richard B. Tracking objects across cameras by incrementally learning inter-camera color calibration and patterns of activity [C]//*Proceedings of the 9th European Conference on Computer Vision*. Graz, Austria, 2006: 125 − 136.

[4] Pier L M, Paolo S, Tiziana D O. Object tracking by non-overlapping distributed camera network [C]//*11th International Conference on Advanced Concepts for Intelligent Vision Systems*. Bordeaux, France, 2009: 516 − 527.

[5] Omar J, Khurram S, Mubarak S. Appearance modeling for tracking in multiple non-overlapping cameras [C]//*IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Washington, DC, USA, 2005: 26 − 33.

[6] Lin Mingxiu, Liu Weijia, Xu Xinhe. Algorithm for multimode vehicle tracking based on template match [J]. *Journal of System Simulation*, 2007, **19**(7): 1519 − 1522. (in Chinese)

[7] Liu X B, Lin L, Yan S C. Adaptive object tracking by learning hybrid template online [J]. *Circuits and Systems for Video Technology*, 2011, **21**(11): 1588 − 1599.

[8] Raul M, Carlos R D, Fernando J. Robust 3D people tracking and positioning system in a semi-overlapped multi-camera environment [C]//*15th International Conference on Image Processing*. San Diego, CA, USA, 2008: 2656 − 2659.

[9] Lam E Y. Combining gray world and retinex theory for automatic white balance in digital photography [C]//*Proceedings of the 9th International Symposium on Consumer Electronics*. Macau, China, 2005: 134 − 139.

[10] Li E, Cai Lidong. An image compensation method based on the illumination component correction [J]. *Journal of Changchun University of Science and Technology*: *Natural Science Edition*, 2010, **33**(2): 137 − 139. (in Chinese)

[11] Hu Zhengping, Yang Jianxiu. Object localization algorithm based on mixture model of HOG feature and LSVM [J]. *Signal Processing*, 2011, **27**(8): 1206 − 1212. (in Chinese)

[12] Massimo P, Eric D C. Track matching over disjoint camera views based on an incremental major color spectrum histogram [C]//*IEEE Conference on Advanced Video and Signal Based Surveillance*. Teatro Sociale, Como, Italy, 2005: 147 − 152.

[13] Yang H D, Lee S W. Multiple pedestrian detection and tracking based on weighted temporal texture features [C]//*Proceedings of the 17th International Conference on Pattern Recognition*. Washington, DC, USA, 2004: 248 − 251.

结合 Lab 模型与 EHOG 特征的摄像机离散视域人物外表匹配

杨 彪 林国余 张为公

(东南大学仪器科学与工程学院, 南京 210096)

摘要:针对摄像头离散区域存在的光照变化、视角变化等干扰,提出一种结合 Lab 模型以及扩展梯度方向直方图特征的方法来改善人物外表匹配的准确率. 对于描述目标全局信息的 Lab 模型,提出一种排序最近邻聚类算法进行颜色聚类,然后使用分块颜色匹配算法计算观察值之间的颜色相似度. 对于描述目标局部信息的扩展梯度方向直方图特征,使用巴氏距离计算 2 个观察值之间的相似度. 全局信息可以很好地描述目标外形,并且能够适应摄像头视角的变化以及目标尺度上的改变. 局部信息对光照变化具有更强的鲁棒性,它能够增强模型的辨别能力. 全局信息和局部信息的结合保证了所提出算法的精确度和鲁棒性. 室内实验结果显示所提出的算法具有较高的正确匹配率.

关键词:人物外表匹配;Lab 模型;扩展梯度方向直方图;摄像机离散视域

中图分类号:TP391