ONLINE PERSON ORIENTATION ESTIMATION BASED ON CLASSIFIER UPDATE

Hong Liu, Liqian Ma

Engineering Lab on Intelligent Perception for Internet of Things (ELIP),
Key Laboratory for Machine Perception (KLMP), Shenzhen Graduate School, Peking University, China
hongliu@pku.edu.cn, maliqian@sz.pku.edu.cn

ABSTRACT

Person orientation estimation is valuable for intelligent video surveillance. Although much progress has been made in recent years, it still faces challenges such as varying poses, illuminations and viewpoints. Most existing approaches merely use appearance information or combine it with motion information. Appearance-based classifiers are trained offline without updating in real time, which can not adapt to unknown scenes. To fix it, a novel orientation estimation approach based on online appearance-based classifier update is proposed. Reliable motion direction is determined acting as pre-estimated person orientation to update the appearance-based classifier. Moreover, a novel criterion based on motion reliability is proposed to determine the motion direction. Experimental results show that the proposed approach achieves more competitive performances especially for unknown scenes.

Index Terms— Person orientation estimation, Online learning, Multi-class classification

1. INTRODUCTION

Person orientation information is very useful for surveillance applications, such as tracking, action analysis and pose estimation. Generally, appearance information and motion information can be used to estimate person orientation. Appearance information is a valid cue for orientation estimation since different orientations may generate corresponding body appearances. However, appearance information is seriously influenced by variations of poses, illuminations and viewpoints. Motion information can also be used to estimate person orientation since they are generally consistent. Motion information will become unreliable in some particular scenes such as turning around and walking in a crowd [1], while in most cases it is reliable and useful. Our important observation is that the probably true orientation can be obtained from reliable motion information. Compared to appearance information, motion information is more robust to various poses, illuminations and viewpoints. Therefore, reliable motion information and appearance information can be combined to increase the performances of orientation estimation.

Most conventional approaches are merely appearance-based and corresponding orientation classifiers are trained offline [1–11], while overlooking different poses, illuminations and viewpoints which result in sensitivity. So that they cannot adapt to new scenes which haven't been used as training data. Later works combine contributions of online motion information and offline appearance-based classifiers using decision fusion [12–14]. Performances are improved while appearance-based classifiers are still trained offline without updating in real time. Actually, it is effective to learn online appearance-based orientation classifiers which have self-adaptability for unknown scenes. Therefore, a novel orientation estimation approach based on online appearance-based classifier update is proposed. Reliable motion information is used to determine the probably accurate person orientation which is then used together with corresponding appearance information to update the appearance-based classifier in real time. Finally, person orientation is estimated based on current appearance information using the updated classifier.

Relation to prior work: Baltieri $et\ al.$ [1] proposed an offline orientation classifier based on extremely randomized trees only using appearance information. However, this approach does not have self-adaptability for unknown scenes with different poses, illuminations and viewpoints. Chen $et\ al.$ [13] estimate orientation utilizing a soft coupling of appearance and motion information in a particle filtering framework. Ichim $et\ al.$ [14] use cues like HOG descriptors, velocity direction and the presence of face to estimate the person orientation and combine three different classifiers to do the classification. These approaches merely use online motion information, ignoring the great effects of online appearance information. In this paper, an online framework based on online appearance-based classifier update is proposed. Our approach effectively combines the contributions of online motion and appearance information, and has great self-adaptability for different scenes.

This work is supported by National Natural Science Foundation of China (NSFC, No.61340046, 60875050, 60675025), National High Technology Research and Development Program of China (863 Program, No.2006AA04Z247), Science and Technology Innovation Commission of Shenzhen Municipality(No.JCYJ20120614152234873, No.JCYJ20130331144631730, No.JCYJ20130331144716089), Specialized Research Fund for the Doctoral Program of Higher Education (No. 20130001110011).

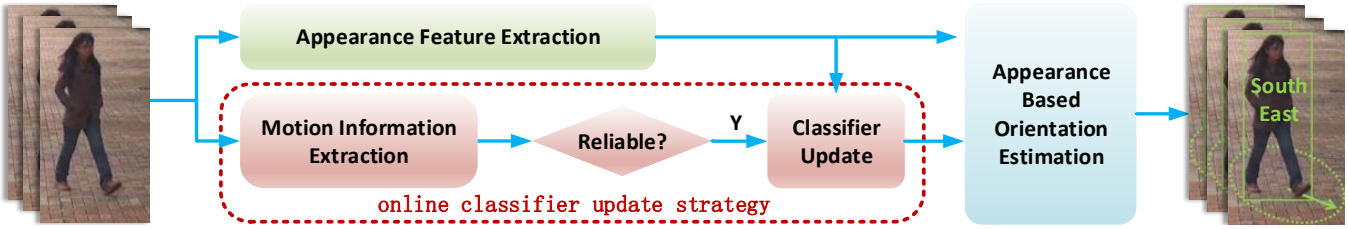

Fig. 1. The framework of the proposed approach containing three major modules in three different colors. **Best viewed in color**.

2. ALGORITHM DESCRIPTION

2.1. Method overview

In this paper, person orientation estimation is treated as a classification problem. Primarily, classifier performance depends on the training samples. When the online testing samples greatly differ from the training ones, a conventional classifier will become invalid since it's trained offline. Therefore, an online person orientation estimation approach based on classifier update is proposed, which can adapt to different scenes. As depicted in Fig.1, our framework contains three major modules: appearance feature extraction, online orientation classifier update and appearance-based orientation estimation. Appearance information is used to estimate the person orientation and update the classifier with reliable motion direction. The online classifier update strategy extracts reliable motion information to update the appearance-based classifier, which can greatly enhance self-adaptation ability. Here, eight discrete directions are considered as shown in Fig.2(a).

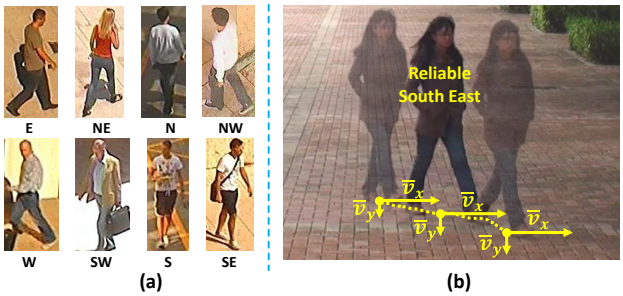

Fig. 2. (a) Eight directions. (b) Reliable motion direction.

2.2. Appearance feature extraction

A single scale HOG descriptor [15] is adopted to describe the appearance information, which is widely used in pedestrian detection task. Human body is resized to 100×200 and divided into 10×20 blocks. For each block, a 9-bin histogram of oriented gradients is extracted and normalized over 2×2 sets of blocks. The obtained 1800-dimensional feature vector is then used to represent person appearance.

2.3. Online classifier update strategy

A. Motion infomation extraction.

Velocity is extracted as the motion information and used to calculate motion reliability. In order to extract motion information from 2D images without camera calibration, the camera is assumed to be in a downward angle without any rotation. For 2D images, person position (p_X, p_Y) is adjusted by a linear transform in order to reduce the perspective effect,

$$p_X = \frac{p_x}{h}, \quad p_Y = \frac{p_y}{h} \tag{1}$$

where (p_x, p_y) is the coordinate of person's center and h denotes person's pixel height. Generally, motion information obtained from 2D image is unstable due to varying poses such as alternating legs and arms, hence temporal smooth strategy is very important. Chen $et\ al.$ [16] use temporal filtering to exploit temporal smoothness only using previous information. In order to obtain more reliable smoothed information, a simple temporal smooth strategy proposed in this paper which uses both previous and future information. The smoothed position $\bar{p}$ and velocity $\bar{v}$ are calculated as follows,

$$\begin{aligned} \bar{p}(t) &= \frac{1}{2\delta_{\bar{p}}+1} \sum_{i=-\delta_{\bar{p}}}^{\delta_{\bar{p}}} p(t+i) \\ \bar{v}(t) &= \frac{1}{2\delta_{\bar{v}}+1} \sum_{i=-\delta_{\bar{v}}}^{\delta_{\bar{v}}} v(t+i) \end{aligned} \tag{2}$$

where $\bar{p}(t)$ denotes the position at time t smoothed in span of $2\delta_{\bar{p}}+1$ and $p(t)$ means the (p_X, p_Y) at time t. Variable $\bar{v}(t)$ denotes the velocity at time t smoothed in span of $2\delta_{\bar{v}}+1$ and $v(t)$ means the velocity calculated by linear fitting $\bar{p}$ using least squares approximation in span of $2\delta_v+1$. Although this strategy may lead to some delay, it is acceptable since the motion information is used to update the appearance-based classifier instead of directly estimating current orientation.

B. Reliable motion direction determination.

Motion reliability is first calculated to determine reliable motion direction. Since there is a transition phase in velocity when turning happens, motion information of δ_r fore-and-aft frames should all be considered to determine the reliability,

$$r(t) = \begin{cases} +1 & \text{for } \forall \bar{v}(t+i) > \tau \\ -1 & \text{for } \forall \bar{v}(t+i) < -\tau \\ 0 & \text{for otherwise} \end{cases} \tag{3}$$
$$i = -\delta_r, -\delta_r+1, ..., \delta_r-1, \delta_r$$

where $r(t)$ denotes the motion reliability calculated in span of $2\delta_r+1$ and τ is the velocity threshold which can be obtained according to the statistics of velocity distribution on the X/Y axis. The total number of frames used to calculate $r(t)$ is

Table 1. Motion direction determination based on motion reliability

$r_X(t)$	+1	+1	0	-1	-1	-1	0	+1
$r_Y(t)$	0	-1	-1	-1	0	+1	+1	+1
Motion Direction	E	NE	N	NW	W	SW	S	SE

$2\delta + 1$, and $\delta = \delta_{\bar{p}} + \delta_v + \delta_{\bar{v}} + \delta_r$. Lacking 3D information, velocity on the X and Y axis can't be merged together. Thus a simple but effective determinant criterion based on motion reliability is presented as shown in Table 1. When neither $r_X(t)$ nor $r_Y(t)$ is zero, motion direction is reliable and acts as pre-estimated orientation as shown in Fig.2(b).

C. Appearance-based orientation estimation.

The pre-estimated person orientation and corresponding appearance information are used to update the appearance-based orientation classifier with LaRank algorithm. Additionally, considering the horizontal spatial symmetry of human body, flipped test data is also utilized to update the classifier in order to make full use of the reliable new data. As mentioned in B, the update will delay δ frames. However, for the video sequences with 25 fps, this delay can be ignored. Since person motion is a continuous process, once a frame with reliable motion information is captured, the updated classifier can handle the similar appearance correctly afterwards. Therefore, the proposed approach is more suitable for the pedestrian surveillance video sequence data, where some reliable motion information can be obtained to update the online classifier.

2.4. Orientation estimation

Considering of the real-time performance, low cost linear LaRank [17] is adopted which is a highly successful sequential minimal optimization based multi-class support vector machine algorithm. For offline learning, LaRank performs one or more learning epochs over the randomly reordered training set. For online learning, LaRank reaches its optimal performance in a single pass over the training examples, which is competitive with those of the full optimization. LaRank learns a function f which maps patterns $x \in \mathcal{X}$ to discrete class labels $y \in \mathcal{Y}$,

$$f(x) = \arg\max_{y \in \mathcal{Y}} S(x, y) \tag{4}$$

where $S(x, y)$ is the discriminant function that measures the correctness of the association between pattern x and class label y. LaRank learns $S(x, y)$ in dual programs and optimizes with an adaptive schedule. However, for LaRank algorithm, more training examples generate more support vectors, which may lead to high computation cost and storage since they are linearly consistent. Thus, a budget maintenance procedure [18] is adopted to bound the number of support vectors.

Algorithm 1 shows the self-explanatory pseudo code of our approach. Firstly, an initial classifier is trained using labeled data. Then the classifier is updated according to the strategy mentioned in Sec.2.3.

Algorithm 1 Online Orientation Estimation

Input: Labeled Data S_{tr}; Test Data S_{ts}
Output: Person orientation PO

```
Extract S_tr appearance feature HOG_tr
Initialize LaRank classifier with HOG_tr
Define MD as the motion direction
loop
    Extract S_ts(t) appearance feature HOG_ts(t)
    Estimate PO(t) ← LaRank(HOG_ts(t))
    if t > 2δ then
        t̃ ← t − δ
        Calculate motion reliability r_X(t̃), r_Y(t̃)
        if r_X(t̃) ≠ 0 or r_Y(t̃) ≠ 0 then
            Determine MD(t̃) by r_X(t̃) and r_Y(t̃)
            Flip data Ŝ_ts(t̃) ← Flip(S_ts(t̃))
            Get ĤOG_ts(t̃), M̂D(t̃) from Ŝ_ts(t̃)
            Update classifier by HOG_ts(t̃), MD(t̃)
            Update classifier by ĤOG_ts(t̃), M̂D(t̃)
            Estimate PO(t) ← LaRank(HOG_ts(t))
        end if
    end if
end loop
```

3. EXPERIMENTS AND DISCUSSIONS

Datasets and evaluation protocol. Performance is reported by two measures: Accuracy 1 considers exact hits only, and Accuracy 2 also considers adjacent classes [1]. For LaRank, linear kernel is chosen to obtain a fast implementation and the number of support vectors is bounded to 5000. Three datasets are used as follows,

TUD Multiview Pedestrian dataset[1] contains 5331 samples of pedestrians. In our experiments, only the training set of 4732 samples is used.

3DPeS dataset[2] contains 1012 samples which are randomly selected from the provided surveillance videos. And the orientation labels are provided.

PKU Person Orientation dataset[3] contains 4 video sequences which are captured in campus surveillance scenes with 25 fps containing eight discrete orientations. The orientation labels and position information are both annotated.

Evaluations and Analysis. The effectiveness of LaRank is first evaluated, comparing with some state-of-the-arts online (row1-4) and offline (row5-6) methods as shown in Fig.3. The classifiers are all trained on TUD Multiview Pedestrian dataset and tested on 3DPeS dataset. LaRank$\times$1 and LaRank$\times$10 mean that classifier is trained for 1 epoch and 10 epochs respectively. Fig.3 reports that the single pass over the training examples is sufficient to reach the optimal performance which significantly outperforms ORF [19] as well as

[1] https://www.d2.mpi-inf.mpg.de/node/428
[2] http://www.openvisor.org/3dpes.asp
[3] https://github.com/mlq513773348/PKU-Person-Orientation.git

Table 2. Comparisons of online orientation estimation on PKU Person Orientation dataset

Method	Seq 1		Seq 2		Seq 3		Seq 4		Average	
	Acc1	Acc2	Acc1	Acc2	Acc1	Acc2	Acc1	Acc2	Acc1	Acc2
MultiSVM (from [13])	0.63	0.96	0.56	0.96	0.57	0.93	0.47	0.79	0.56	0.91
MultiSVM (from [13]) + RMD	0.84	0.98	0.77	0.98	0.77	0.98	0.72	0.98	0.78	0.98
MoAWG [1]	0.70	0.98	0.63	0.97	0.59	0.94	0.53	0.84	0.61	0.93
MoAWG [1] + RMD	0.86	0.98	0.79	0.98	0.78	**0.99**	0.73	0.98	0.79	0.98
LaRank [17]	0.63	0.96	0.54	0.95	0.57	0.93	0.48	0.80	0.56	0.91
RMD	0.71	0.85	0.69	0.89	0.66	0.92	0.69	0.90	0.69	0.89
LaRank + RMD	0.84	0.98	0.77	0.97	0.77	0.98	0.73	0.99	0.78	0.98
LaRank + Update	0.91	0.98	**0.90**	0.98	0.90	0.97	0.90	**0.99**	0.90	0.98
LaRank + Update + Flip	**0.92**	**0.99**	**0.90**	**0.99**	**0.92**	0.97	**0.91**	**0.99**	**0.91**	**0.99**
LaRank + Update + Flip + RMD	0.88	**0.99**	0.80	0.98	0.80	0.97	0.75	0.98	0.81	0.98

OMCLP [20] and is competitive with other offline learning methods. This indicates the great scalability of LaRank.

Comparisons and discussion For a fair comparison, all classifiers are first trained on TUD Multiview Pedestrian dataset for initialization and then tested on PKU Person Orientation dataset. The detailed parameters are set as follows, $\tau_X = 0.01, \tau_Y = 0.00125, \delta_{\bar{p}} = 5, \delta_v = 3, \delta_{\bar{v}} = 5, \delta_r = 5$. The comparison is a little bit difficult since few previous works focus on the online orientation estimation. Therefore, the comparison is organized in two aspects as shown in Table 2: self-comparisons (row5-10) and comparisons with the state-of-the-arts offline learning method (row1-4).

Firstly, self-comparison (row5-10) is conducted by evaluating the performance of four parts in the framework: (1) LaRank, linear LaRank algorithm, (2) Update, update classifier with reliable motion direction, (3) Flip, update classifier using horizontal flipped data, (4) RMD, estimate orientation using reliable motion direction. It can be seen from row 2 in Table 2 that the accuracy of RMD is considerable. This result accords with the hypothesis that person orientation generally consists with reliable motion direction. When appearance and motion information are combined simply (LaRank+RMD), both Accuracy1 and Accuracy2 attain a certain amount of improvement, which indicates that the appearance and motion information are complementary. Additionally, when the classifier is updated online with reliable motion direction (LaRank+Update), Accuracy1 attains another 12 percent increase averagely than simple combination. If the horizontal flipped data is also used to update the classifier (LaRank+Update+Flip), the accuracy only gets one more percent increase. The probable reason is that the contribution of flipped data may be more prominent when reliable motion information is less. However, PKU Person Orientation dataset contains much reliable motion information, which weakens the effect of flipped data. Moreover, when RMD and online classifier are combined (LaRank+Update+Flip+RMD), Accuracy1 decreases by 10 percent. The main reason may be that for LaRank+Update+Flip+RMD, decision-level fusion is used to combine the outputs of online updated appearance-based classifier and motion-based classifier. The probable

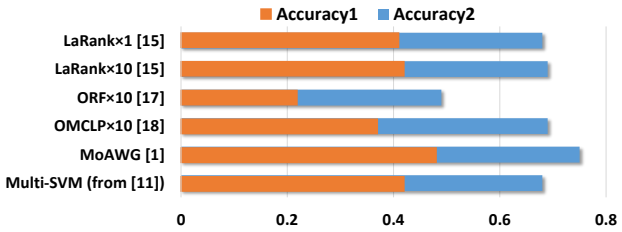

Fig. 3. Evaluation of LaRank algorithm.

Fig. 4. Qualitative results on PKU Person Orientation dataset

mistake in motion direction determination may significantly affect the final results. While for LaRank+Update+Flip, orientation is estimated only based on online appearance-based classifier, which has a memory of the historical information and may be slightly affected by the false motion direction.

Secondly, our approach is also compared with the state-of-the-arts offline learning method (row1-4). As we have expected, combining the appearance-based classifier and motion information will bring a certain promotion to accuracy. But the accuracy is still less than that of ours due to no updates of the appearance-based classifier.

4. CONCLUSIONS

This paper introduces an online orientation estimation approach in which the appearance-based classifier is updated in real time by reliable motion information. Motion information is a quite valuable cue for updating the appearance-based classifier, which can make the classifier adapt to unknown scenes. Experimental results show that our approach achieves better performance than offline learning methods indicating that it is more suitable for unknown surveillance applications.

5. REFERENCES

[1] D. Baltieri, R. Vezzani, and R. Cucchiara, "People orientation recognition by mixtures of wrapped distributions on random trees," *in Proceedings of ECCV*, pp. 270–283, 2012.

[2] M. Andriluka, S. Roth, and B. Schiele, "Monocular 3d pose estimation and tracking by detection," *in Proceedings of CVPR*, pp. 623–630, 2010.

[3] M. Enzweiler and D. M. Gavrila, "Integrated pedestrian classification and orientation estimation," *in Proceedings of CVPR*, pp. 982–989, 2010.

[4] C. Weinrich, C. Vollmer, and H.M. Gross, "Estimation of human upper body orientation for mobile robotics using an svm decision tree on monocular images," *in Proceedings of IROS*, pp. 2147–2152, 2012.

[5] N. Noceti and Odone F., "Semi-supervised learning of sparse representations to recognize people spatial orientation," *in Proceedings of ICIP*, pp. 3382–3386, 2014.

[6] C. Chen and J. Odobez, "We are not contortionists: coupled adaptive learning for head and body orientation estimation in surveillance video," *in Proceedings of CVPR*, pp. 1544–1551, 2012.

[7] A. Heili, J. Varadarajan, B. Ghanem, N. Ahuja, and J.M. Odobez, "Improving Head And Body Pose Estimation Through Semi-supervised Manifold Alignment," *in Proceedings of ICIP*, pp. 1912–1916, 2014.

[8] W. Liu, Y. Zhang, S. Tang, J. Tang, R. Hong, and J. Li, "Accurate estimation of human body orientation from RGB-D sensors," *in IEEE Transactions on Cybernetics*, vol. 43, no. 5, pp. 1442–1452, 2013.

[9] M.C. Liem and D.M. Gavrila, "Person appearance modeling and orientation estimation using spherical harmonics," *in Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition Workshops*, pp. 1–6, 2013.

[10] G. Zhao, M. Takafumi, K. Shoji, and M. Kenji, "Video based estimation of pedestrian walking direction for pedestrian protection system," *Journal of Electronics (China)*, vol. 29, pp. 72–81, 2012.

[11] H. Shimizu and T. Poggio, "Direction estimation of pedestrian from multiple still images," *in Proceedings of IEEE Intelligent Vehicles Symposium*, pp. 596–600, 2004.

[12] O. Ozturk, T. Yamasaki, and K. Aizawa, "Estimating human body and head orientation change to detect visual attention direction," *in Proceedings of ACCV Workshops*, pp. 410–419, 2010.

[13] C. Chen, A. Heili, and J. M. Odobez, "Combined estimation of location and body pose in surveillance video," *in Proceedings of AVSS*, pp. 5–10, 2011.

[14] M. Ichim, R. T. Tan, N. Aa, and R. Veltkamp, "Human body orientation estimation using a committee based approach," *in Proceedings of VISAPP*, pp. 515–522, 2014.

[15] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," *in Proceedings of CVPR*, vol. 1, pp. 886–893, 2005.

[16] C. Cheng, A. Heili, and J. Odobez, "A Joint Estimation of Head and Body Orientation Cues in Surveillance Video," *in Proceedings of ICCV Workshops*, pp. 860–867, 2011.

[17] A. Bordes, L. Bottou, P. Gallinari, and J. Weston, "Solving multiclass support vector machines with Larank," *in Proceedings of ICML*, pp. 89–96, 2007.

[18] S. Hare, A. Saffari, and P.H. Torr, "Struck: Structured output tracking with kernels," *in Proceedings of ICCV*, pp. 263–270, 2011.

[19] A. Saffari, C. Leistner, J. Santner, M. Godec, and H. Bischof, "On-line random forests," *in Proceedings of ICCV Workshops*, pp. 1393–1400, 2009.

[20] A. Saffari, M. Godec, T. Pock, C. Leistner, and H. Bischof, "Online multi-class lpboost," *in Proceedings of CVPR*, pp. 3570–3577, 2010.