

Real-time Full-body Visual Traits Recognition from Image Sequences

Christoph Jung¹, Reimar Tausch¹ and Christian Wojek^{2,3}

¹Fraunhofer-Institut für Graphische Datenverarbeitung IGD, Darmstadt, Germany

²Department of Computer Science, TU Darmstadt, Germany

³MPI Informatics, Saarbrücken, Germany

Abstract

The automatic recognition of human visual traits from images is a challenging computer vision task. Visual traits describe for example gender and age, or other properties of a person that can be derived from visual appearance. Gathering anonymous knowledge about people from visual cues bears potential for many interesting applications, as for example in the area of human machine interfacing, targeted advertisement or video surveillance. Most related work investigates visual traits recognition from facial features of a person, with good recognition performance. Few systems have recently applied recognition on low resolution full-body images, which shows lower performance than the facial regions but already can deliver classification results even if no face is visible. Obviously full-body classification is more challenging, mainly due to large variations in body pose, clothing and occlusion. In our study we present an approach to human visual traits recognition, based on Histogram of oriented Gradients (HoG), colour features and Support Vector Machines (SVM). In this experimental study we focus on gender classification. Motivated by our application of real-time adaptive advertisement on public situated displays, and unlike previous works, we perform a thorough evaluation on much more comprehensive datasets that include hard cases like side- and back views. The extended annotations used in our evaluation will be published. We further show that a hierarchical classification scheme to disambiguate a person's directional orientation and additional colour features can increase recognition rates. Finally, we demonstrate that temporal integration of per-frame classification scores significantly improves the overall classification performance for tracked individuals and clearly outperforms current state-of-the-art accuracy for single images.

Categories and Subject Descriptors (according to ACM CCS): I.4.8 [Image Processing and Computer Vision]: Scene Analysis

1. Introduction

During the last years researchers have started to investigate visual properties of people in images, mostly motivated by interesting new applications in the area of video surveillance, advanced human machine interfaces, robotics or content-based search. Our work addresses the application of targeted advertisement in digital signage. Such systems use cameras to observe and analyse the audience in front of a public situated display in order to present appropriate content. Most relevant information to be extracted from the audience is usually gender, age, eye gaze, level of attention or dwell time.

In our study we investigate trait recognition for pedestrians

based on full-body low-resolution images, as they could be provided by a single camera in the above mentioned scenario. We conduct the independent analysis of single images as well as the analysis of image sequences of the same individual, as they could be provided by a pedestrian tracking system. Our approach to human visual traits recognition is based on Histogram of oriented Gradients (HoG), colour features and Support Vector Machines (SVM). Unlike previous work we provide a thorough evaluation on a much more comprehensive datasets that does not exclude hard cases like side- and back views. We further show that a two-step classification approach with respect to a person's direction and colour features can increase recognition rates. Finally, we demonstrate that temporal integration of

per-frame classification scores significantly improves the overall classification performance for tracked individuals.

2. Related Work

Recognition of visual traits in terms of visual human properties has been addressed by various authors during the last years. Most studies consider facial information to classify. The authors in [KVL99, ALC04] perform age classification on human faces. These approaches usually require frontal faces including some prior automatic or manual alignment. In [KDLC07, WM08] the classification of human attractiveness is investigated, which is modelled as a complexion of averageness, smoothness and symmetry. Common approaches make use of manual landmarks, colour, texture to train classifiers such as SVM or kNN. Some work on recognition of people's clothing [CXLZ06, GC08] has also been published. The classification is used as means to the end when solving the actual problem of people segmentation. The authors exploit body parts models or graph-cuts to perform segmentation. Ethnicity (or race, nationality) has shown to be related to gender recognition. The authors of [SVM02] present an approach to ethnicity recognition using Haar-like features and boosting, based on the prior published seminal face detection paper. They also mention temporal integration, but do not provide experimental results. Most publications in the last years focused on gender recognition. To our knowledge the work of Lawrence *et al.* [GLS90] was the first to mention gender recognition from face images. It was followed by many other approaches e.g. using SVM on thumbnail images [MY02], PCA and LDA on some selected facial characteristics (length of nose, presence of hair, etc.) [BDGF05] or Active Appearance Model (AAM) on automatically aligned faces [MR08], resulting in high recognition rates of 86% and more. Non-aligned faces are addressed in [TA09] by using scale- and rotation-invariant SIFT features for recognition. However, when observing people over time in front of a display, as it is the case in our application, exploitation of facial information is not always possible.

To our best knowledge there have been only two publications so far studying full-body recognition of visual traits (gender only), which are most related to our paper. Cao *et al.* [CDFH08] claim to be the first to investigate gender classification on full-body images. They limit their evaluation to frontal and back views from the MIT pedestrian dataset [OPS*97] and introduce a body parts model based on HoG classification of sub-patches, achieving 75% of accuracy. Accordingly, Collins *et al.* [MCW09] introduced colour features and demonstrated improved accuracy of 80%, but limited their evaluation to frontal images only and reduced influence of background clutter by tight image cropping.

3. Approach

Our approach aims to classify single images as well as sequences of images showing individual persons in full body arbitrary upright postures. Similar to [CDFH08, MCW09] the trait classification is based on HoG as appearance feature and local colour features learned by SVMs as classifier. Prior to classification of human attributes, such as gender, age group and body physique, we apply a classification into 4 different directional orientations of people (front, left, right, back). For the case of image sequences, we integrate the classification scores of all frames to form reliable and more robust decisions.

The following subsections give an overview on our approach's building blocks including the image descriptors, classifiers and temporal integration.

3.1. Descriptors

We investigate two different histogram-based descriptors, one representing appearance and the other representing colour information.

Appearance: Originally proposed by Dalal and Triggs [DT05] histograms of oriented gradients (HoG) are used to describe the appearance of objects and were successfully applied for people detection. The descriptor discards colour information by exclusively working on the gradient image or edge map, which is usually retrieved from the intensity or grey image. After the gradients are extracted the image is divided by a grid into connected cells of fixed size (8×8 pixels was found to be optimal for people detection on 64×128 pixel full-body windows). On each cell a histogram of oriented (sign-less) gradients is computed, each histogram bin representing one direction of orientation. The sign-less half circle is typically subdivided into 9 bins, covering 20° each. Therefore, HoG is robust to small alignment errors. The cells are grouped by overlapping blocks of usually 2×2 cells and each block is normalized. The resulting descriptor vector consists of the concatenated histograms from each block. This study focuses on the use of 2×2 blocks with L2 normalization, while different cell sizes between 4×4 and 8×8 pixels were investigated. For more information on HoG please refer to [DT05].

Colour: Humans dress in different colors with respect to their gender or age group. Some of the commonly known dress codes can therefore discriminate people into classes, e.g., young people tend to dress more colourful than the elderly, boys or men usually prefer muted, blueish colours instead of bright pink colours. While these trends would not be robust enough as a discriminating factor alone, Collins *et al.* [MCW09] showed that supplementing HoG with local colour features slightly enhanced the gender recognition accuracy. We extended the problem with age group recognition and use a simple approach, in which colour histograms are extracted on three local image patches representing a person's head, torso and legs, in order to capture clothing, skin

and hair colours. The fixed position and size of these patches was set based on the dataset average image. As histogram basis, different colour spaces were investigated, e.g., RGB (red, green, blue), HSV (hue, saturation, value), which separates the color hue in only one dimension, and normalized RG, which discards intensity and codes color using only two chromaticity coordinates.

3.2. Classifiers

We use linear SVMs to classify attributes and learn the discriminating visual traits (figure 1). In practice this combination of linear SVMs and histogram based descriptor types proved to work well [DT05] [MCW09].

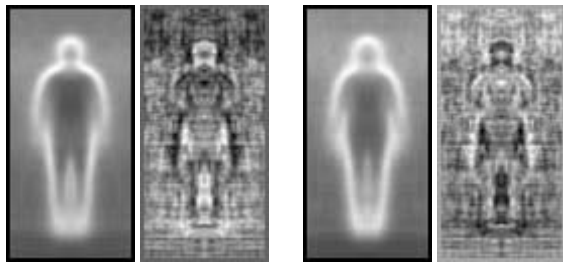


Figure 1: Standard and SVM-weighted gradient average images from our training dataset of classes male (left) and female (right). The weighted images illustrate the traits learned by the HoG-trained classifier, i.e., where class distinctive gradients should appear (white areas) and where not (black areas). Cues, such as long open hair, wider hips and tighter stance for females were found.

To achieve multi-class classification a binary one-vs-all SVM is trained for each attribute class. In order to disable biasing the examples of under-represented classes in the dataset must relatively gain a stronger weight during the training process. Therefore we adapted these samples weight according to their ratio in the dataset.

We found that knowledge about a person’s directional orientation improved the classification results on the remaining attributes (gender, age group, body physique), since their corresponding distinctive visual traits (e.g., hair lines, shoulders, hips) can significantly differ with respect to the view direction (front, back, side). Thus, we argue for a two-stage classification hierarchy to achieve better sample registration. On the first level a person is classified by 4 SVMs according to directions (front, back, right, left) as well as classified by direction-dependently trained SVMs for the other attributes (e.g., front male vs. front female, side male vs. side female). The resulting scores are then combined on the second level by SVMs trained on these scores to form the final decision (e.g., male or female). Here, the classifiers can benefit from explicit direction classification and thus can weight the direction-dependent classification scores of the other attributes.

3.3. Temporal Integration

We found that the full-body based classification can be severely interfered, e.g., by partial occlusion or irritating background. In urban video scenes with moderate people density, these issues can be easily induced, but often last for only a short moment. Thus, if a person is detected and tracked in video, it is expedient to take all frames of the full observation time (i.e., track) into account, rather relying on single frames. Moreover, attributes, such as gender, age, or physique, are considered invariant over the observation time. Therefore, temporal integration strategies can be applied to fuse classifier outputs at each time frame in order to accumulate evidence and thus form more confident and reliable decisions.

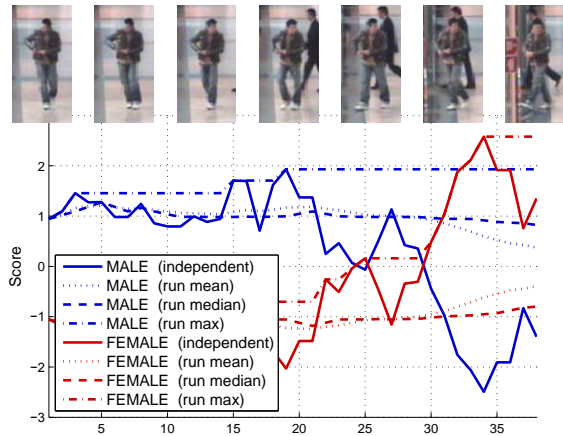


Figure 2: A person track (images on top) and its gender classifier outputs per frame. A decision can be made at each time frame, whereby the winning class has the maximum score. (Red and blue curves are vertically mirrored, since the attribute gender consists of only two complementary classes.)

Figure 2 visualizes the investigated methods, such as the running arithmetic mean, median and maximum on consecutive classification scores of a track. While the running mean and median tend to smooth the score curve, the maximum focuses on peaks, i.e., the most discriminating classification score / confidence value that occurred during the observation time. As shown in the according image frames of the track, a person in the background temporarily confuses the classification by accidentally adding female-like gradients to the person in the front. Regarding the generated score magnitude, these traits seem to be even more discriminating for the class female than the male-voting traits recognized before on the track. Thus, the running maximum immediately reacts to the peak and the person will be classified as woman from that time point on, while a decision based on the inert mean and median curves would still be correct.

3.4. Real-time Implementation and Setup

We have implemented the approach presented above in a real-time capable application (see figure 3). A GPU-implementation of the HoG pedestrian detector (for example using [PR, WDSS08]) provides bounding boxes that can be tracked using a simple Kalman filtering. Classification of traits is performed using SVM^{light} [Joa99], a widely used C++ SVM implementation. The application is running on a Quad-core Intel CPU and an nVIDIA GTX 260 GPU in real-time.



Figure 3: Screenshots of our full-body based real-time visual traits classification: no high resolution faces available, varied postures, illumination and occlusion.

4. Evaluation

This section is subdivided into the evaluation for static images and track classification. First 10-fold cross-validation on a dataset that consists of 8948 static images is conducted to measure the classification accuracies for the attributes: direction, gender, age group, body physique. Second a video dataset, which contains 322 people traces of aligned, temporally consecutive full-body frames, is used to measure the improvement on gender recognition that can be achieved by introducing temporal integration strategies.

4.1. Evaluation on Static Images

4.1.1. Static Image Dataset

In order to cover the high variety of body poses and clothing styles of pedestrians, three sets were fused to one large testing and training database containing 8948 samples. All images have a relatively low resolution of 64×128 pixels with the person centered taking approx. 100 pixels in height (figure 4). Table 1 separately shows the annotated attribute classes with their cardinal numbers of all three utilized datasets: MIT CBCL Pedestrian database [OPS*97], INRIA Person dataset [DT05], and TUD-MotionPairs dataset [WWS09], which all were originally designed for the pedestrian detection or tracking task.

In contrast to the other studies [CDFH08] [MCW09] that investigated gender classification on full body, the datasets were completely labelled including challenging samples, such as side views. To further increase the number of training samples and because of the symmetry of the human body

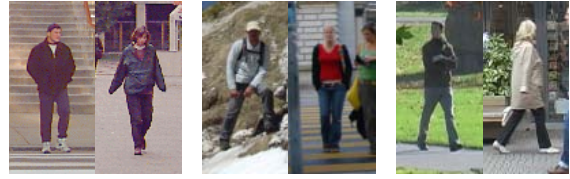


Figure 4: Samples from the three utilized datasets.

Attribute Class	Dataset Samples			
	MIT	INRIA	TUDMP	Union
direction				
front	874	1168	500	2542
back	974	1082	572	2628
right	0	649	1240	1889
left	0	649	1240	1889
gender				
male	1224	2142	1748	5114
female	624	1406	1804	3834
age group				
young	32	308	92	432
middle	1658	2550	3068	7276
old	158	690	392	1240
physique				
thin	352	848	860	2060
medium	1124	1938	2044	5106
broad	372	762	648	1782
total	1848	3548	3552	8948

Table 1: Annotation of the individual image dataset.

all images were horizontally mirrored after the annotation. We found that data quantity is crucial for the classification success in order to cover the large variety of possible full-body appearances. By adding the TUD-MotionPairs dataset the number of total side views tripled and likewise the gender classification accuracy achieved only on side views increased from 55% to 63%. Unlike the discrete attribute gender, which was relatively easy to label, the other attributes (direction, age, physique) contain classes placed on a continuous scale. This makes their discrete annotation task much harder. Adequate borders have to be defined before separating the continuous samples into discrete classes. The attribute direction is labeled 90°-wise into front, right, back and left views. Regarding the three age classes (young, middle, old) boundaries of approx. 15 and 60 years were chosen. Additionally, the persons body physique was labeled into three classes, where the medium class is considered as a buffer zone between the two extremal, clearly differing classes thin and broad.

The main focus on our work is the attribute gender. The attribute direction is used to support the recognition on the other attributes.

4.1.2. Appearance-based Classification Results

The focus in this section is on HoG features describing person silhouettes. We investigated different cell sizes between 4×4 and 8×8 pixels, but conclude that HOG with 6×6 pixel cells performed best for gender classification. In the following this configuration is fixed and used to compare the results achieved by direction-dependent and hierarchical classification.

By direction-dependent evaluation on subsets of the dataset (front, back, right views separately) we observed that classification accuracies achieved on the more dynamic side views were exceeded by those achieved on front and back views (table 2). The best recognition rates were obtained on back views, which apparently reveal more vital cues for the classification process (e.g., as for the attribute gender the long hair for females) (figure 1). The inclusion of the side view samples, when evaluating on arbitrary views, resulted in a drop of gender recognition accuracy. Compared to the standard approach that classifies all arbitrary views with only one trained SVM, the hierarchical classifier that involves explicit information about a person's directional orientation achieved higher accuracies by about 2% for all attributes.

View	Overall Accuracy		
	gender	age	physique
front	69.7%	42.5%	49.2%
back	71.9%	48.0%	52.5%
right	63.4%	45.2%	45.0%
arbitrary standard	67.3%	48.2%	48.2%
arbitrary hierarchic	68.9%	50.1%	50.1%

Table 2: Direction-dependent classification results separately achieved on data subsets: front, back and side views. Descriptor: HoG 6x6 pixel cells.

Here, the direction classification on the first stage attained an attribute overall accuracy of 73.3%. The corresponding ROC-curves (figure 5) show that the recognition of the side views performed best. As indicated by the confusion matrix, if only relying on body silhouettes the confusion between front and back views is much higher, than with the side views.

The ROC-curves for the direction-supported classification of the other attributes (gender, age group, body physique) can be observed in figures 6, 7 and 8 and are briefly commented as follows. The gender recognition shows equal results for both classes (male, female). In contrast, the recognition rates for the three age groups slightly differ, with best results for the class young. Here, HoG descriptors proved useful to distinguish a young child's appearance by height and proportions (shorter legs, longer arms, bigger head). Concerning the attribute physique, the recognition of the medium class performed poor emphasizing its role as a buffer zone, while there is less confusion between the ex-

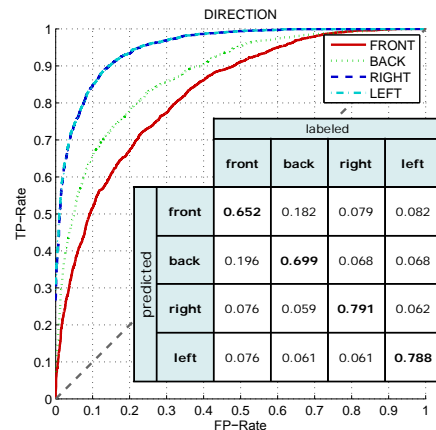


Figure 5: ROC curves of the prior directional orientation recognition applied for the hierarchical classification. Attribute overall accuracy 73.3%. Descriptor: HoG 6x6 pixel cells.

tremal classes (thin, broad). Figure 11 indicates how appearance of the different attributes is reflected by the SVM weights.

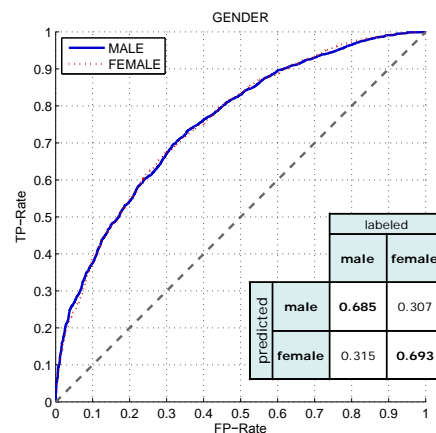


Figure 6: ROC curves of the hierarchical gender classification. Attribute overall accuracy 68.9%. Descriptor: HoG 6x6 pixel cells.

4.1.3. Colour-based Classification Results

We compared different colour spaces for the histograms on the three patches (head, torso, legs), such as RGB, HSV, normalized RG. Based on our experiments we conclude that the original RGB histograms perform best. Normalized RG disposes to much intensity information (e.g., grey values) that proved valuable for classification. In order to measure the significance of colour for the classification process,

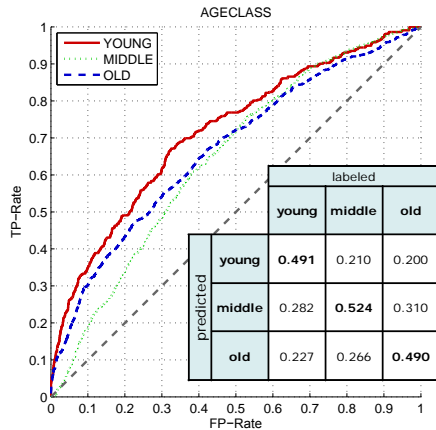


Figure 7: ROC curves of the hierarchical *age group* classification. Attribute overall accuracy **50.1%**. Descriptor: HoG 6x6 pixel cells.

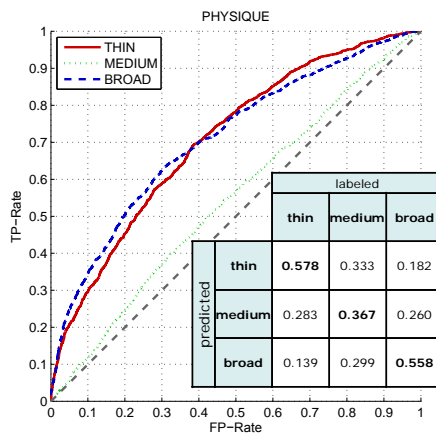


Figure 8: ROC curves of the hierarchical *body physique* classification. Attribute overall accuracy **50.1%**. Descriptor: HoG 6x6 pixel cells.

we conducted an evaluation with colour features alone. In contrast to [MCW09] colour proved not valuable for gender recognition on our larger dataset (accuracy 51.4% is slightly above chance level). However, colour information can help to better distinguish age groups (figure 9). Especially elderly people were recognized by grey hair or unobtrusive, monotonous and often grey beige clothing. Against the appearance-based classification, the recognition of the class young performed worst, since younger people tend to dress in a more colourful way and this variety could not be sufficiently captured and learned from the few dataset samples. Generally, the performance of the colour histograms compared to appearance features was found relatively low, and thus integrating the colour descriptors by early fusion

with above evaluated HoG descriptors resulted only in a small accuracy improvement of maximal 2%-3% for the age group recognition.

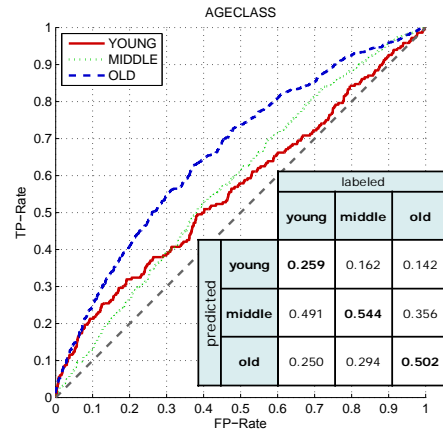


Figure 9: ROC curves of *age group* classification. Attribute overall accuracy **43.5%**. Descriptor: 3 RGB histograms (head, torso, legs) each with 10^3 bins.

4.2. Evaluation on Video Tracks

4.2.1. People Tracks Dataset

To evaluate classification performance while leveraging temporal cues from video sequences we employed the AVSS'07 i-Lids bag dataset [AVS10] (figure 3). In order to be not affected by detector or tracker errors the dataset was manually annotated with bounding boxes surrounding and tracking the people. The dataset was further extended by tracks from the CAVIAR dataset [CAV10]. Both video sources were recorded in urban environment (subway, shopping center) with moderate people density. Altogether, the dataset (table 3) consists of 322 (only) gender labeled person tracks, with an average of 315 frames at the uniform size of 64×128 pixels.

Classes	Dataset Samples	
	tracks	frames
gender		
male	186	59593
female	136	41971
total	322	101564

frames per track	
minimum	20
maximum	2214
average	315

Table 3: Annotation of the video tracks dataset.

4.2.2. Temporal Integration Results

The classification improvement gained by introducing the above described three simple temporal integration methods,

mean, median and maximum, can be measured in two different ways, either frame-based or track-based. For the frame-based evaluation at each time frame in the dataset (101564) a decision is made on basis of the current frame's classifier scores and, when using temporal integration, of all priorly observed scores on the same track. As visualized in figure 2, running average methods can be applied to smooth and stabilise the score curves of tracks. For the track-based evaluation for each track in the dataset (322) only one decision is made based on the track's fused scores that are calculated once from all contained frames by one application of mean, median, or maximum.

Temporal Integration		Overall Accuracy	
		HOG8x8	HOG6x6
frame-based	Non (indep.)	73.3%	74.1%
	Run. Mean	81.8%	86.6%
	Run. Median	80.7%	86.0%
	Run. Maximum	80.7%	85.8%
track-based	Mean	82.5%	86.6%
	Median	81.7%	86.4%
	Maximum	80.1%	85.8%

Table 4: Gender classification accuracies on image tracks.

Table 4 gives an overview of performed evaluations. Here, two different HoG descriptors (8×8 and 6×6 pixel cells) were investigated, while the classifier in both cases consists of linear SVMs trained on the individual images dataset (table 1) for each attribute class (male, female). The table indicates that if handling each frame separately, without applying temporal integration, a gender recognition rate of 74.1% is achieved. This is about 5% better than the corresponding cross-validation results on the more challenging static image dataset (table 2). Best results of 86.6% were achieved in combination with a simple arithmetic mean as temporal score averaging on tracks as well as on running frames, while the maximum method proved less robust. The corresponding ROC-curves for the frame-based evaluation (figure 10) illustrate the significant improvement. The curves appear smooth, since they are based on 101564 decisions. These results emphasize the importance of expanding the observation period from single frames to whole tracks when classifying from video. In doing so, incorrect classification often caused by temporary issues can be effectively reduced.

5. Conclusion

In this study we presented a thorough evaluation of full-body visual traits classification based on SVM, HoG and colour features. Motivated by our specific application scenario audience observation, we used larger image and video datasets, comprising arbitrary upright postures of pedestrians, with many hard cases such as severe occlusions and cluttered background, as well as image sequences from person tracks. We showed that a person's orientation can be reasonably

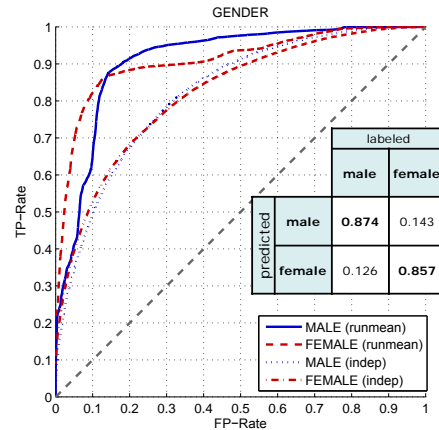


Figure 10: Measured effect of temporal integration on image sequences: independent vs. mean score averaging.

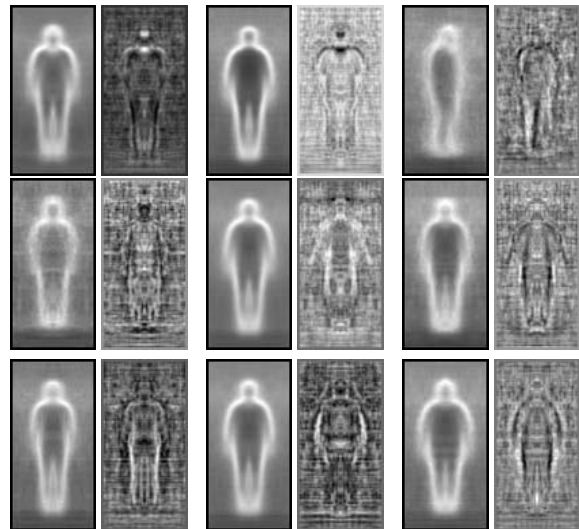


Figure 11: Standard and SVM-weighted gradient average images for attributes - Top row: **direction** (front, back, right), middle row: **age** (young, middle, old), bottom row: **physique** (thin, medium, broad).

classified, which can be used to improve attribute classification with a hierarchical classifier setup. An additional simple colour feature, represented by a coarse layout configuration, further improved age classification. Besides recognition of a person's gender we have shown experiments with age and physique recognition from full-body. However, these two attributes achieved substantially lower performance compared to gender classification. Our approach is integrated in a real-time capable application, using standard PC hardware and GPU processing.

In the future we expect to improve results when using more

sophisticated observation strategies for tracked individuals. When classifying from video, it is conceivable to involve detector and tracker confidence outputs as an additional classification weighting and perform temporal integration on multiple levels of the hierarchical classifier. We further plan to integrate classification of facial features in an active camera setup, as these still outperform full-body classification for single images. We would also like to investigate additional person related scene information, such as pose, motion, clothing style, social relations or emotion.

6. Acknowledgements

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n°248893.

References

- [ALC04] A. LANITIS C. D., CHRISTODOULOU C.: Comparing different classifiers for automatic age estimation. *IEEE Transactions on Systems, Man, and Cybernetics* 34, 1 (2004), 621–628. 2
- [AVS10] Advanced video and signal based surveillance – i-lids bag dataset 2007. http://www.elec.qmul.ac.uk/staffinfo/andrea/avss2007_d.html, accessed July 2010. 6
- [BDGF05] BUCHALA S., DAVEY N., GALE T. M., FRANK R. J.: Principal component analysis of gender, ethnicity, age, and identity of face images. In *IEEE ICMI* (2005). 2
- [CAV10] Ec funded caviar project/list 2001 37540 – caviar test case scenarios - clips from shopping center in portugal. <http://homepages.inf.ed.ac.uk/rbf/CAVIAR/>, accessed July 2010. 6
- [CDFH08] CAO L., DIKMEN M., FU Y., HUANG T. S.: Gender recognition from body. In *MM '08 Proceedings* (New York, NY, USA, 2008), ACM, pp. 725–728. 2, 4
- [CXLZ06] CHEN H., XU Z. J., LIU Z. Q., ZHU S. C.: Composite templates for cloth modeling and sketching. In *CVPR '06 Proceedings* (Washington, DC, USA, 2006), IEEE Computer Society, pp. 943–950. 2
- [DT05] DALAL N., TRIGGS B.: Histograms of oriented gradients for human detection. In *CVPR '05 Proceedings - Volume 1* (2005), IEEE Computer Society, pp. 886–893. 2, 3, 4
- [GC08] GALLAGHER A. C., CHEN T.: Clothing cosegmentation for recognizing people. In *CVPR '08 Proceedings* (2008), pp. 1–8. 2
- [GLS90] GOLOMB B. A., LAWRENCE D. T., SEJNOWSKI T. J.: Sexnet: A neural network identifies sex from human faces. In *NIPS-3 Proceedings* (1990), Morgan Kaufmann Publishers Inc., pp. 572–577. 2
- [Joa99] JOACHIMS T.: Making large-scale support vector machine learning practical. In *Advances in kernel methods: support vector learning*. MIT Press, Cambridge, MA, USA, 1999, pp. 169–184. 4
- [KDLCo07] KAGIAN A., DROR G., LEYV T., COHENOR D.: A humanlike predictor of facial attractiveness. *NIPS* (2007). 2
- [KVL99] KWON Y. H., VITORIA LOBO N. D.: Age classification from facial images. *Computer Vision and Image Understanding* 74, 1 (1999), 1–21. 2
- [MCW09] MATTHEW COLLINS JIANGUO ZHANG P. M., WANG H.: Full body image feature representations for gender profiling. In *VS '09 ICCV: IEEE Workshop on Visual Surveillance* (2009), pp. 1235–1242. 2, 3, 4, 6
- [MR08] MAKINEN E., RAISAMO R.: Evaluation of gender classification methods with automatically detected and aligned faces. *IEEE TPAMI* 30, 3 (2008), 541–547. 2
- [MY02] MOGHADDAM B., YANG M.-H.: Learning gender with support faces. *IEEE TPAMI* 24, 5 (2002), 707–711. 2
- [OPS*97] OREN M., PAPAGEORGIOU C., SINHA P., OSUNA E., POGGIO T.: Pedestrian detection using wavelet templates. In *CVPR* (1997), pp. 193–99. 2, 4
- [PR] PRISACARIU V., REID I.: *fastHOG - a real-time GPU implementation of HOG*. Tech. Rep. 2310/09, Department of Engineering Science, Oxford University. 4
- [SVM02] SHAKHAROVICH G., VIOLA P. A., MOGHADDAM B.: A unified learning framework for real time face detection and classification. In *FGR '02 Proceedings* (2002), IEEE Computer Society, p. 16. 2
- [TA09] TOEWS M., ARBEL T.: Detection, localization, and sex classification of faces from arbitrary viewpoints and under occlusion. *IEEE TPAMI* 31 (2009), 1567–1581. 2
- [WDSS08] WOJEK C., DORKÓ G., SCHULZ A., SCHIELE B.: Sliding-windows for rapid object class localization: A parallel technique. In *Proceedings of the 30th DAGM symposium on Pattern Recognition* (Berlin, Heidelberg, 2008), Springer-Verlag, pp. 71–81. 4
- [WM08] WHITEHILL J., MOVELLAN J. R.: Personalized facial attractiveness prediction. *FGR* (2008). 2
- [WWS09] WOJEK C., WALK S., SCHIELE B.: Multi-cue onboard pedestrian detection. In *CVPR* (2009), pp. 1–8. 4