COPYRIGHT NOTICE

UB ResearchOnline
<http://researchonline.ballarat.edu.au>

This the accepted version of Illicit image detection using erotic pose estimation
based on kinematic constraints

The final publication is available at link.springer.com and via

http://dx.doi.org/10.1007/978-1-4614-3535-8_41

Copyright Springer Science+Business Media

Illicit Image Detection using Erotic Pose Estimation Based on Kinematic Constraints

Mofakharul Islam¹ Paul Watters¹ John Yearwood¹ Mazher Hussain² Lubaba A Swarna³

¹Internet Commerce and Security Laboratory
School of Science, Information Technology and Engineering
University of Ballarat, Australia

²Melbourne Institute of Technology
Melbourne, Australia

³IBM Australia
Ballarat, Australia

Abstract-: With the advent of the Internet along with sophisticated digital image processing technology, the Internet quickly became the principal medium for the distribution of pornographic content favouring pornography to become a drug of the millennium. With the advent of GPRS mobile telephone networks, and with the large scale arrival of the 3G networks, along with the cheap availability of latest mobile sets and a variety of forms of wireless connections, the internet has already gone to mobile, drives us toward a new degree of complexity. The detection of pornography remains an important and significant research problem, since there is great potential to minimize harm to the community. In this paper, we propose a novel approach to investigate and implement a pornography detection technique towards a framework for automated detection of pornography based on most commonly found erotic poses. Compared to the results published in recent works, our proposed approach yields the highest accuracy in recognition.

I. INTRODUCTION

Internet access to the computers and a range of other communicating devices are increasingly a modern day must have for children and young people around the world. Not only are they establishing new cultural norms, they are also becoming mainstream within education. Majority of adolescent worldwide spend a great majority of their time online to locate their educational stuff and as a source of perfectly innocent fun and games. But every child and young people who uses the Internet will almost certainly be exposed at some stage to material that will shock and possibly harm them, or they will come into contact with organization or individuals who mean them injury. To follow potentially catastrophic consequences throughout a children's lifetime, he/she needs only one such encounter to go wrong. It is therefore, no surprise that the Internet as well as the social networking sites pose a serious threat to our children who are more or less vulnerable, either some or all of the time while they are on the Internet. In fact, potential outcome of the Internet use is questionable as a little is understood yet of the potential problems and benefits associated with it, and the overall impact on the society from social benefit point of view

that may arise. The risks posed by criminals who attempt to share, exchange, consume and produce child exploitation material, however, are fairly clear, and law enforcement is faced with the difficult task of trying to deal with the sheer volume of material (and offences) in a streamlined and systematic way.

Research [22] shows some of the statistics on pornography as shown in the pie-chart below;

Fig. 1: Worldwide Pornographic Revenue (Courtesy: TopTen Reviews)

Apart from the exposure of age inappropriate content to minors, dealing with pornography in the workplace is a serious challenge for many large organizations but employing a block-all-images Internet browsing and email policy no longer provides a viable solution. Web and email are more media-based than ever before, and it is common for business

document and mail to contain images such as logos, publicity shots etc. In a paper written by McGuire et al. found that the viewing of pornography can serve as a source of a paraphilic "vivid sexual fantasy" which, when contemplated during masturbation, may condition men into perversion [1]. Further, research also established a direct link between pornography consumption and crimes like sexual abuse, rape, violence, and child molestation [2], [3], [4].

Therefore, it is obvious that exposure of pornography and age-inappropriate materials, either legal or illegal have a devastating impact most often turns into a complex, often compulsive, psychosexual disorder with profound implications on us and our children who all use the internet or will do soon.

* Every second - \$3,075.64 is being spent on pornography
* Every second - 28,258 internet users are viewing pornography
* Every second - 372 internet users are typing adult search terms into search engines

Table 2: Internet Pornography Statistics

Research showed that most of the audience belongs to the lower levels that gradually progress towards higher levels. If we are able to put a stopping mechanism somewhere in this low level area, it will protect people from entering the exiting world of arousal and being pornography addicted.

Fig. 2. Progression Model for Pornography Consumer

Contemporary research established a link between child pornography and adult pornography. From that point of view, it will not only prevent people from viewing the most erotic and bestial images but also put an end for the potential paedophiles who are currently belongs to less severe groups to progress further towards most severe groups that eventually make them professional paedophiles at some stage. From the studied literatures, we find a progression model for pornography consumer as shown in Fig. 1.

So, there is a fundamental demand in real time classification of attached images to block inappropriate contents in a business or commercial environment. The problem therefore is the detection and prevention of certain types of images. Real-time detection of pornographic images can effectively prevents pornographic images from entering into the workplace via email and Internet browsing.

Prevention of explicit sexual content is an extreme challenge that paved the way to a growing industry aimed at blocking and filtering such contents. The software industry suggests two types of solutions to combat with the problem of pornography - collecting the adult web site addresses and collecting key words of adult web pages. While the former

allows the user to access the requested page upon verifying its IP address, the latter performs the content analysis for providing access to the user. Here, the filtering programs block the pages from viewing upon finding any related key word of the web page in user's computer. The drawback of these systems is that it is an uphill task to collect all adult web site addresses because hundreds of new web sites are launching every day. The drawback of the latter system is that it tends to block the sex education pages due to the key words used on these pages.

In order to prevent access to pornographic sites, many commercial systems are available in the market. The commercial softwares/tools like Net-Nanny, CyberSitter, CyberPatrol and ChildWebGuardian allows access upon comparison of IP addresses / URLs and key words based on a long list of pornographic site IP address/URL and key words in their system databases. Although it is effective in blocking well-known pornographic sites and pages of pornographic links, it miserably fails in blocking pages containing pornographic image galleries since these most often do not contain links to other pages or objectionable text.

Other commercial systems such as ScreenShield, Snitch, System Recon and Enologic NetFilter Home in fact, deter the viewing of objectionable images rather than blocking where these systems consider the percentage of skin within an image to classify the image as to whether or not pornographic. Such an approach has been found not very accurate in practice and requires inspection by a human observer to make conclusion.

We are interested in the computational foundations of vision that helps us designing machine vision systems with applications to pornographic image detection. Our proposed project aims to find out the solutions for some fundamental questions in computer vision. How can we recognize persons, guns, cars, boats and many other suspicious categories of objects in cluttered pictures? How can we be trained these categories in the first place? Can we provide machines with ability akin to this?

This project primarily aimed at detection of pornography using erotic pose based on kinematic constraints. Erotic images focus on themes with either indicative, erotic or sensual scenes or subjects, sometimes with depictions of human nudity and lovemaking, but not always of an extremely explicit, gratuitous or pornographic nature. These kinds of films often appeal to the emotions of the viewer, with their emphasis on pleasure, physical desire, and human companionship.

The remainder of this paper proceeds as follows: In Section II we present previous work in this specific field. In Section III, we describe our proposed pornography detection model, which estimates body part appearance models for pictorial structures using latent relationships between the appearances of different body parts. Experimental results demonstrating the accuracy and efficiency of the proposed approach are discussed in Section IV and finally in Section V, we present our conclusion and future work.

II. RELATED WORK

While significant research have already been conducted on pornography or illicit image detection most often with miscellaneous low and high level features, and other visual cues in fact, no research has looked closely at the subject.

In fact, all the pornography detection approaches are based on skin detection module that defines a decision rule based on which skin and non-skin pixels will be discriminated. A decision rule based on the distance of the pixel color to skin tone using a metric plays the crucial role in the discrimination process as this metric is defined by the skin color modelling technique. The final goal of skin-color detection is to choose a classifier that will discriminate between a skin and a non-skin. From classification point of view, skin-color detection can be viewed as a two class problem: skin pixel vs. non-skin pixel. Literature suggests different classifiers are employed by different authors for skin detection based on their respective contexts, where data distributions or data modelling played a crucial role. Our literature review suggests three specific types of skin modelling techniques - Explicitly defined skin region, Non-parametric distribution modelling, Parametric distribution modelling, and Adaptive modelling. In Explicitly Defined Skin Region techniques, threshold is a parameter that stipulates the values a pixel can be if it is to be considered as skin. In Nonparametric distribution modelling, Normalized Lookup Table (LUT), Naïve Bayes Classifier, and Self Organizing Map (SOM) are most often used as skin classifier. Single Gaussian, Mixture of Gaussians, Multiple Gaussian Clusters, and Elliptical Boundary Model are mostly used as Parametric Model for skin color detection. Adaptive skin model employs two or more skin-color models from the models as discussed above in combination or in sequence to make the technique robust against the varying conditions. In other word, it is called 'Hybrid Model'.

Most of the adult image or pornographic detectors employed these skin color models based on different color space along with some other high level features like shape, color moment, color mean, region descriptors for their classification schemes.

A mathematical approach of grouping the images using the detected skin regions and extracted high level region features from these regions that might have link to the pornography is called pornographic classification, which is mainly focussed on classification of the skin region into benign or pornographic. We can broadly categorize pornographic image classifier into three groups: Supervised Machine Learning, Geometric Classifier, and Boosting Classifier. Supervised Machine Learning can be further categorized hereunder into four groups. Supervised Machine Learning can be further categorized hereunder into four groups - Support Vector Machine [9], [10], [11], Neural Network, Decision Tree [12], and K-Nearest Neighbour [13]. Fleck et al. demonstrated an affine-invariant model that define human as a set of rules describing how to assemble possible girdles and spine-thigh groups, where both the individual geometry of the body parts and the relationships between parts are constrained by the geometry of the skeleton which eventually provides an

appropriate and effective model for human body recognition [14]. Bootstrapping induces a classifier on a small set of labelled data and a large set of unlabeled data. Lee et al. reported an increase of sensitivity from 81.74% to 86.29% with boosting algorithm [15].

Cusano et al. reported better performance of SVM than multiple decision trees [9]. Bosson et al. found that neural networks (83.9specificity) gave slightly better results to that of k-NN and SVM [16]. Bosson et al. attained 94.7% sensitivity and 95.1% specificity using NN with MPEG-7 Descriptors. Zheng et al. [11] found that DT is able to provide better accuracy than NN and SVM [16]. Xu et al. demonstrated superior performance of k-NN over NN [13].

Another approach aimed at robust pornography detection proposed by the same group of authors where pornography specific contextual constraints are employed to pornography detection but no comparisons are made as both of them are contemporary work and presented in the same conference [23].

Almost all the existing techniques proposed region descriptors that are in fact, not directly aimed at capturing contextual constraints representatives of pornography rather based on some heuristic high level statistical and geometrical features on skin region like area, size, shape, location, and their relationships in terms of some ratios, resulting poor performance in finding the actual pornographic stuff on images and thus failing to utilize its full potential in areas like detection of pornography. Further, the extracted features as employed in the existing work are found to be vulnerable under the inherent imaging artifact like intensity inhomogeneity, scale, rotation, transformation, occlusion, and camera viewpoint. These are the void in the literature that this research is attempting to fill.

III. OUR PROPOSED APPROACH

Although the identification and detection of pornographic activity in an image is easy to a human observer, automatic and accurate identification is difficult and complex. Obtaining satisfactory detection results depends mainly on the ability of the descriptors in characterizing pornographic contextual constraints. There are sufficient reasons to believe that the pornographic images would not be of highest standard as these images were neither taken in ideal conditions nor taken by any skilled photographer. In addition, fears, tension, and panic associated with these sorts of heinous activity or crime are not conducive at all to have quality photographs, resulting poor quality images that suffer numerous imaging artifacts originated from intensity inhomogeneity, shadow, occlusion and unorthodox camera manoeuvrings. As a result, noisy images are the final output at pornographer's end that would eventually turns into input in our proposed project. Pre processing i.e., noise removal, enhancement, and reconstruction are not practicable at all on multiple terabyte of image data on the Internet and being found in confiscated hard drives and other storage media by the LEAs. These noisy data poses a great challenge for the descriptors to be able to extract

contextual contexts, representative of any specific context. In order to deal with such difficulty, we will employ erotic pose estimation to detect pornography rather than applying classification purely on extracted low or high level features.

Detecting humans pose in an image is noteworthy due to numerous reasons. Pose and/or pose sequence are extremely useful in depicting human's attitude and action.

Here, we propose a novel approach for estimating part appearance models from a single image. We employ generic detector to determine an approximate location in terms of location distribution and scale reference frame on the object as done in [17, 18, 19]. The basic motivation of our proposed approach is based on two observations – (i) some parts have rather stable location relative to the reference frame, (ii) different parts are statistically related in the appearance model. This implies that the appearance of some parts can be predicted from the appearance of other parts.

Over the last few years Pictorial structure (PS) [20, 21, 22] has guided most of the research on articulated pose estimation paradigm. PS [18, 20, 21, 22, 23], are usually used for humans pose estimation though it is equally effective in any articulated object class like aeroplane, car, cow, lion etc. PS is basically a probabilistic model that portrays objects made of parts tied together by pair wise potentials encoding contextual constraints as prior. In addition, a unary potential measures part's position to generate an appearance model of the Part. Finally, MAP classifier infers spatial configuration of the parts (the pose of the object).

The relative location distribution of parts with respect to the reference frame and the dependencies between the appearances of different body parts are learned from training images with ground-truth pose annotated by a stickman. These relations are exploited generating appearance models for body parts given a new image. In determining the appearance model for more mobile parts (e.g. lower arms, lower legs), parts (torso) having higher location distribution with respect to the reference frame plays a crucial role.

A human body parts are represented by the general framework of pictorial structure, where body parts are tied together in conditional random field. Typically, parts l_i are rectangular image patches and their position is parameterized by location (x, y), orientation θ , and scale s. The posterior of a configuration of parts $L = \{l_i\}$ given an image I is

$$P(L | I, \theta) \propto \exp\left(\sum_{i,j \in E} \psi(l_i, l_j) + \sum_i \phi(l_i | I, \theta)\right) \quad (1)$$

where, the $\psi(l_i, l_j)$ is a pair wise potential term correspond to a prior on the relative position of parts based on kinematic constraints (e.g. the upper arms and upper legs must be attached to the torso). The term $\phi(l_i | I, \theta)$ is an unary potential depicts the local image evidence for a part in a particular position. How parts should look like entirely depends on appearance models θ that measure the dissimilarity between the image patch at l_i and the appearance model for part i . The appearance models are in

fact, parameters of the PS and must be provided by an external mechanism.

We extend the model (1) by adding some new low level features to make it more accurate and robust under diverse imaging artifacts. Model (1) is primarily based on edge features and RGB colour features, which appears to be vulnerable while applied on unconstrained and cluttered images. We add texture features in addition to the existing color features to extend its ability to extract similarity of body parts appear on cluttered background. So, in addition to the color, texture is another feature that can be employed as a cue for image analysis. Texture features are constructed in such a way that they characterize local variation in intensity or color within the neighbourhood surrounding the pixel. As a result, a value is assigned to the pixels with a given texture. Research suggests, color and texture based analysis are in combination capable of producing more accurate results than either the color or texture feature used independently [24].

However, mixing weight of colour and texture, which is pertinent to accuracy in capturing image appearance throws an open problem. Up to date research, computer vision and image processing in particular do not provide any authenticated relationship between these two while applied in combination, which incite us to run an experiment to find out an optimal weight ratios of these two. In our experiment, we have sourced 100 pornographic images freely available on the Internet and run an MLE (Maximum Likelihood Estimation) based segmentation scheme on these images to segment skin regions on them. Misclassification rate are plotted against both colour and texture component to find a cogent mixing weight.

Fig. 3. Mixing Weights of Colour and Texture

Percentage of colour and texture components are taken at 5% interval starting from 10% to 90% and the graph is plotted on the mean classification error based on the segmentation results as obtained for the 100 images as chosen earlier. The plotted curve showing its minimum at 65% colour and 35% texture. So, we use this mixing weight of colour and texture in our work.

While the YC_rC_b space is employed as colour feature, Haar wavelet transforms are employed as texture features. Literature suggests a clear dominance of YC_rC_b , reason why

we prefer it. Similarly, wavelet transform outperforms its other counterparts while applied in cluttered image having complex background. The wavelet transform has the capability to capture subtle texture information on objects appear on cluttered background more precisely than other texture descriptors [21].

YC_rC_b is the most popular choice, where 'Y' represents luminance computed as a weighted sum of RGB values, and chrominance values 'Cb' and 'Cr' computed by subtracting the luminance component from 'B' and 'R' values, offering us a significant advantage for utilising pure color information in the detection process. Using pure color information, a skin model or detector gets an ability to be able to adapt to the changes in the lighting and the viewing environment. The key advantage of being a lighting invariant color spaces made YC_rC_b space the most popular choice to the research community working on skin detection. Although there is a strong debate on whether or not the luminance component 'Y' playing any significant role in the detection process we discard it considering the luminance invariance approach of our proposed algorithm.

In this paper, our improvement mainly focussed on the unary potential, where we introduce a texture histogram in addition to the existing colour histogram to generate appearance model. So, unary potential takes the form;

$$\sum_i \phi(l_i | I, \theta) = \sum_i \delta * \phi_C(l_c | C, \theta_c) + \sum_i (1 - \delta) * \phi_T(l_t | T, \theta_t) \quad (2)$$

where, δ is a colour component weight, C is colour image, T is texture image, θ_c and θ_t are describing how parts should look like in terms of colour and texture respectively, l_c and l_t are image patch and texture patch respectively, and ϕ_C and ϕ_T are local evidence for a part in respect of colour and texture respectively.

A person detector's detection windows is the input to our method [5].

Two observations are the main motivation of this particular approach: (i) some parts have relatively stable location in the detection window, $W = (x; y; s)$. For example, the torso is typically in the middle of an upper-body detection window; (ii) the different body parts appearances are related (e.g. the upper-arms often have the same color as the torso).

Now, two statistical observations – location prior of the body parts and appearance transfer mechanism are learned, where location prior holds distribution of the body part locations relative to the detection window and appearance transfer mechanism improves the models derived from the location prior by combining models for different body parts. The training data consists of images with ground-truth pose annotated by a stickman, i.e. a line segment for each body part (Fig. 4). Given a detection window and the learnt location

priors, initial appearance models are estimated and then subsequently refined by the appearance transfer mechanism.

A. Learning Location Prior from the Training Images

Fig. 4. Learning location priors. From top left i) a training image. ii) detection windows (magenta), , window associated to the stickman (white), body part rectangles (green) obtained by widening the stickman line segments (red). iii) learnt location priors for different body parts.

$LP_i(x, y) \in [1, 0]$ for a pixel (x, y) to be covered by the part before considering the actual image data (Fig. 4). These are in fact, location prior of the body parts with respect to the detection window/frame.

The LPs are learnt from training images with ground-truth pose annotated by a stickman (Fig. 4 i). A generic person detector is initially run to find a person detection window. Now, association of stickmen to detection windows are made as shown in (Fig. 4 ii). Then all the training stickmen are projected with the common coordinate frame aligning the location and scale roughly. Finally, the maximum likelihood estimate is employed to learn the LPs for every pixels in the window. Example LPs are presented in (Fig. 4 iii) LPs for the head and torso are found quite sharp and localized, while LPs for the arms and legs are more diffuse. As the lower arms and lower legs can move around freely, the location of these body parts appears very uncertain a priori.

B. Transferring appearance models between body parts training phase

In our case, i.e., pornographic image, all the body parts have same colour due to existence of skin on them offering us to exploit this intrinsic relation between the appearances in prediction. Being inspired by the influence of the above relations, here we employ a transfer mechanism to combine the appearance models of different body parts, where the input appearance models are derived from estimated LPs. The final output of this transfer mechanism is a new appearance model of a part as a linear combination of the input appearance

models of all parts. The newly estimated appearance model requires a mixing weight of part i , in the combination of part p . The appearance model is defined as;

$$AM_{pc}^{TM} = \sum_i w_{pc} AM_c^{LP} \quad (3)$$

where, AM_c^{LP} is the initial colour appearance model derived earlier from the colour location prior.

Similarly, we get the texture appearance model from the texture location prior based on the similar assumptions that all the body parts have similar skin texture;

$$AM_{pt}^{TM} = \sum_i w_{pt} AM_t^{LP} \quad (4)$$

We learn the mixing weights w_{pc} and w_{pt} by minimizing the squared difference between the appearance models produced by the transfer mechanism (AM_{pc}^{TM} and AM_{pt}^{TM}) and those derived from the ground-truth stickmen (AM^{GT}) as done in [6].

C. Estimation of Appearance Model for a Test Image

Estimation of good appearance models composed of two steps - colour & texture model and soft-segmentation based on the colour & texture models.

The colour & texture model estimation has three distinct steps. First, a standard coordinate frame is transformed from the detection window W cropping out of the input image and rescaling it to a fixed size to estimate the LPs. The next step is estimation of colour & texture models from the LPs. Finally, refined colour and texture models $\Pr(c | fg)$ and $\Pr(t | fg)$ are learned applying appearance transfers as in equation (3) and (4) respectively.

Characterize the appearance of the body parts are estimated using the aforesaid colour & texture models. In addition, we also estimate here a background model $\Pr(c | bg)$ and $\Pr(t | bg)$ for each body part following Ramanan [7]. The learned foreground and background colour and texture models are employed to derive the posterior probability for a pixel to belong to a part i using Bayes theorem, assuming equal prior probability for foreground and background colour and texture.

$$P_i(fg | c) = \frac{P_i(c | fg)}{P_i(c | fg) + P(c | bg)} \quad (5)$$

$$P_i(fg | t) = \frac{P_i(t | fg)}{P_i(t | fg) + P(t | bg)} \quad (6)$$

Based on the posterior foreground probabilities for both colour and texture, a soft-segmentation for each body part is done on the image to generate a cue for the modified unary term in equation (2).

IV. EXPERIMENTAL RESULTS AND DISCUSSION

We have sourced the most common categories of erotic poses from uncontrolled pornographic images freely available on the Internet to run our experiment. This data is challenging due to the inherent imaging artifacts as mentioned in Section I. Cluttered images, noise due to electro-magnetic interference, often dark illumination, persons appearing at a wide range of scales are some of the many difficulties adhered to image data. We have annotated 400 pornographic images having erotic poses, where roughly upright and approximately frontal images are only chosen. A person is annotated by a 10-part stickman (head, torso, upper and lower arms, upper and lower legs). The whole person must be visible in image. Now, LPs and mixing weights are computed as described earlier in this section.

Once LPs and mixing weights are computed from the training images, the proposed model is ready to estimate good appearance models for new test images.

For testing purpose, we have downloaded another 200 pornographic images of erotic categories from the Internet. Another 200 non-nude images from categories as sports, including aquatic sports, arts, and others with clothed people pictures having assorted poses are also downloaded from the Internet for testing purpose. Pose estimation procedure are followed as:

- (1) Detection of person using detection windows suggested in [18];
- (2) Estimation of part-specific color and texture models as described in previous section;
- (3) Detection window along with the detected person is subjected to image parsing engine [7] using directly our color and texture models in the unary potential to estimate the person's pose.

A popular measure, called Precision and Recall (Eq. 7), based on True Positive (TP), True Negative, False Positive (FP), and False Negative (FN) is applied on both - our experimental results and the results obtained from another pornography image detector proposed by Belem & Cavalcanti [8] to evaluate the performance of our novel approach. We selected Belem & Cavalcanti [8] for performance comparison as its authors reported more than 90% accuracy. Apart from the Precision and Recall, we also compared accuracy (Eq. 8) of our approach with Belem & Cavalcanti [8]

$$\text{Precision} = \frac{TP}{TP + FP} \quad (7)$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$Accuracy = \frac{TP + TN}{P + N} \quad (8)$$

Method	TP	FP	FN	TN	PRE	REC	Acc
Bel. & Cav. [8]	161	61	39	139	0.73	0.81	0.75
Proposed	189	32	19	168	0.86	0.95	0.89

Table 2: Comparison of performance between the proposed and Bel. & Cav. [8] approach

We experienced a bit lower performances with Bel. & Cav. [8] than what claimed by the authors. The low performance might be attributed to the use of unconstrained images in this particular work

V. FUTURE WORK AND CONCLUSION

We have proposed a novel pornography detection approach, where erotic poses work in tandem with some low level features on human body images to detect pornography with maximum accuracy. We primarily base on Eichen and Ferrari's work [5], which primarily aimed at human pose estimation, but extended it with addition of texture feature on top of colour feature to foster the effectiveness of the proposed technique while dealing with unconstrained and cluttered images. While Eichen and Ferrari employed RGB colour features, we apply $YCbCr$ space to make it more appropriate for skin detection. Further, luminosity component, 'Y' is discarded considering illumination or luminosity invariance. To the best of our knowledge, this is the first of its kind which is able to recognize pornography using erotic pose effectively with highest accuracy and thus paves the way for research in this area to not only help pornography detection, but also to contribute significantly in security & surveillance, human body movement and pose identification, medical imaging and much more. The model may be extended further for detection of pornography based on other erotic poses (non-frontal and non-upright) that are not considered in this particular work.

REFERENCES

- [1] R. J. McGuire, J.M. Carlisle, and B.G. Young "Sexual deviations as conditioned behaviour: A hypothesis". *Behaviour Research and Therapy* 2: 185-90, 1964.
- [2] M. H. Silbert, and A. M. Pines, 1984, "Pornography and sexual abuse of women". *Sex Roles* 10: 857-68, 1984.
- [3] D. L. Carter, R. A. Prentky, R. A. Knight, P. L. Vanderveer, and R. J. Boucher, "Use of Pornography in the Criminal and Developmental Histories of Sexual Offenders". *Journal of Interpersonal Violence* 2: 196-211, 1987.
- [4] E. M. Alexy, W. B. Ann, and A. P. Robert "Pornography Use as a Risk Marker for an Aggressive Pattern of Behavior Among Sexually Reactive Children and Adolescents". *Journal of the American Psychiatric Nurses Association* 14 (6): 442-453, 2009.
- [5] V. Ferrari, M. Marin-Jimenez, and A. Zisserman. Progressive search space reduction for human pose estimation. In CVPR, Jun 2008.
- [6] M. Eichner and V. Ferrari, "Better appearance models for pictorial structures", British Machine Vision Conference, 2009, London, UK.
- [7] D. Ramanan. Learning to parse images of articulated bodies. In NIPS, 2006.
- [8] R.J.S. Belem and J.M.B. Cavalcanti "SNIF: A Simple Nude Image Finder", Proceedings of the Third Latin American Web Congress (LA-WEB'05), IEEE, 2005.
- [9] Ruiz-del-Solar1, V. Castañeda, R. Verschae, R. Baeza-Yates, and F. Ortiz, "Characterizing Objectionable Image Content (Pornography and Nude Images) of SpecWeb Segments: Chile as a Case Study", in *Proceedings of the Third Latin American Web Congress (LA-WEB'05, IEEE)*.
- [10] C. Jeong, J. Kim, and K. Hong, "Appearance-Based Nude Image Detection", in *Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04)*, IEEE Computer Society, Washington, DC, USA, 2004.
- [11] Q.F. Zheng, W. Zeng, W.Q.A. Wang, and W.Gao, "Shape-based Adult Image Detection", in *International Journal of Image and Graphics (IJIG)*, vol. 6(1), January 2006, pp. 115-124.
- [12] D. Chai and A. Bouzerdoum, "A Bayesian approach to skin color classification in YCbCr color space", in *IEEE TENCON'00*, vol. 2, pp. 421-424, 2000.
- [13] M. Soriano, J.B. MartinKauppi, S. Huovinen, and M. Laaksonen, "Adaptive skin color modeling using the skin locus for selecting training pixels", in *Pattern Recognition*, vol. 36(3), pp. 681-690, 2003.
- [14] M. Fleck, A. Forsyth, and C. Bregler, "Finding Naked People", in *Computer Vision - ECCV'96, 4th European Conference on Computer Vision*, Cambridge, UK, Volume II, pp. 593-602, 1996.
- [15] J.-S. Lee, Y.-M. Kuo, P.-C. Chung, and E.-L. Chen, "Naked image detection based on adaptive and extensible skin colour model", in *Pattern Recognition*, vol. 40(8), pp. 2261-2270, August 2007.
- [16] Y. Wang, W. Wang, and W. Gao, "Research On the Discrimination of Pornographic and Bikini Images", in *Proceedings of the 7th IEEE International Symposium on Multimedia (ISM'05)*, 2005.
- [17] M. Andriluka, S. Roth, and B. Schiele, "People-tracking-by-detection and people-detection-by-tracking", In *CVPR*, 2008.
- [18] V. Ferrari, M. Marin-Jimenez, and A. Zisserman, "Progressive search space reduction for human pose estimation", In *CVPR*, Jun 2008.
- [19] S. Gammeter, A. Ess, T. Jaeggli, K. Schindler, and L. Van Gool, "Articulated multi-body tracking under egomotion", In *ECCV*, 2008.
- [20] P. Felzenszwalb and D. Huttenlocher, "Pictorial structures for object recognition", *IJCV*, 61(1), 2005.
- [21] Islam, M., "Unsupervised Color Image Segmentation Using Markov Random fields Model", *Master by Research Thesis*, Graduate School of Information Technology and Mathematical Sciences, University of Ballarat, Australia, 2008.
- [22] J. Ropelato, "Internet Pornography Statistics", Retrieved March 2009, from <http://internet-filter-review.toptenreviews.com/internet-pornography-statistics.html>.
- [23] M. Islam, P. Watters, J. Yearwood, M. Hussain, and L. Swarna, "Illicit Image Detection: An MRF Model based Stochastic Approach", *CISSE'2011*, Bridgeport, USA, 2011.