2011 10th International Conference on Machine Learning and Applications

Avatar Face Recognition using Wavelet Transform and Hierarchical Multi-scale LBP

Abdallah A. Mohamed, Darryl D'Souza, Naouel Baili and Roman V. Yampolskiy
Computer Engineering and Computer Science
University of Louisville
Louisville, USA
{aamoha04, darryl.dsouza, n0bail02, roman.yampolskiy}@louisville.edu

***Abstract*—Recognizing avatars in virtual worlds is a very important issue for law enforcement agencies, terrorism and security experts. In this paper, a novel face recognition technique based on wavelet transform and Hierarchical Multi-scale Local Binary Pattern (HMLBP) is presented and shown to increase the accuracy of recognition of avatar faces. The proposed technique consists of three stages: preprocessing, feature extraction and recognition. In the preprocessing and feature extraction stages, the wavelet decomposition is used to enhance the common features of the same class of images and the HMLBP is used to extract representative features from each avatar face image without a need for any training. In the recognition stage, the Chi-Square distance is used to achieve a robust decision and to indicate the correct class to which the input image belongs. Experiments conducted on two manually cropped avatar image datasets (Second Life and Entropia Universe) show that the proposed technique performs better than traditional (single scale) LBP, Wavelet Local Binary Pattern (WLBP) and HMLBP in terms of accuracy (78.57% and 67.50% recognition rates for Second Life and Entropia Universe datasets respectively).**

Keywords-Avatar; face recognition; HMLBP; LBP; Second Life; wavelet transform

I. INTRODUCTION

A virtual world is a three-dimensional, computer-based simulated environment comprised of online communities connected over the Internet. These worlds are rapidly gaining momentum as they possess the capability to enrich society. They are gaining popularity across the globe and are moving towards being an integral part of it. Second Life [6], Active Worlds [7], and Entropia Universe [8] are a few popular virtual worlds with millions of registered users.

Virtual worlds bring a sense of a "personal" digital space for users by mirroring real world activities. The users create their own "avatars" and navigate the world. An "avatar" is the user's virtual identity within these worlds whose appearance can be altered as per the user's choice. This provides them with a lot of flexibility and adaptability. Communities, social groups, enterprises and institutions are all present within the virtual worlds. An avatar can navigate the world by moving around buildings, fly and swim as well as teleport to different locations. Communication between users ranges from text, visual gestures, sound and occasionally touch and voice-commands. Virtual money which can be purchased using real money could be exchanged for goods and services.

Given the complex nature of these adaptable and personal worlds they possess potential for doing a lot of good or bad things. Education is fostered with a strong emphasis on teaching and learning. Interactivity and collaboration between users helps achieve tasks and bond people together. However, destructive purposes involving traditional crimes like identity thefts, fraud, tax evasion, illegal gambling and terrorist activities are reportedly on the rise in virtual worlds [9]. Quick investigation of Second Life reveals that it is populated by numerous terrorist organizations associated with Al-Qaeda who can train in such simulated environments using weapons identical to their real-world counterparts [10].

These criminal activities are posing a grave problem to the law enforcement agencies in these lawless virtual worlds. Forensic experts are expressing interest in accurate and automatic tracking of users and their avatars. Authenticating humans (biological entities) is essential and a well-developed science utilized to determine one's identity in today's modern society. However, avatar authentication (non-biological entities) is an issue that needs to be highlighted and addressed [11]. Profiling avatars is a novel and a challenging approach contributing towards a new research direction in face detection and recognition. To address the concerns for an affordable, automatic, fast, secure, reliable and accurate means of identity authentication Yampolskiy et al. define the concept of *Artimetrics* – a field of study that will allow identifying, classifying and authenticating robots, software and virtual reality agents [12, 13].

In the context of investigating criminal and terrorist activity in virtual worlds four scenarios requiring a face recognition algorithm [9]:

a. *Matching a Human face to an Avatar face*
   Generally many users have the tendency to use their real face as their online avatar which helps represent them well.
b. *Matching one avatar face with another*
   This capability helps to continuously track an avatar through cyberspace at different places at different times.
c. *Matching an Avatar's face from one virtual world to the same avatar in a different virtual world*
   A recent development within virtual communities is to interconnect different virtual worlds. This will help in uniquely identifying and tracking records of the avatars.
d. *Matching an Avatar sketch to the Avatar face*
   Just like the traditional methods of matching the forensic sketch of human faces provided by the description of the victim or witness to their real faces, it is equally important

978-0-7695-4607-0/11 $26.00 © 2011 IEEE
DOI 10.1109/ICMLA.2011.41

to map this scheme within virtual worlds to match the virtual criminal with its avatar identity.

Prior work on authentication of avatars is limited. Klare et al. [9] emphasize the significance of face recognition within virtual worlds. An approach for parameterized generation of avatar face datasets was reported in [14]. The facial biometric authentication of avatars using wavelet transform for feature extraction and SVM for classification is discussed in [11]. Yampolskiy & Gavrilova have applied biometric principles towards avatar recognition and outlined future directions and potential applications [12]. Boukhris et al. [15] have presented an approach for applying face recognition to avatars as part of security framework for virtual worlds. Mohamed & Yampolskiy have applied wavelet transform with Local Binary Pattern (LBP) to recognize avatar faces [16].

In this paper, we propose a new avatar face recognition algorithm by using the idea of wavelet decomposition in preprocessing the avatar face images and then extracting the face features using Hierarchical Multi-scale Local Binary Pattern (HMLBP). The experimental results demonstrate the efficiency of the proposed algorithm.

The remainder of this paper is organized as follows:
In Section 2 we provide an introduction to Wavelet transformation, its role and benefits in the field of image processing. Section 3 describes the LBP operator and histogram. Section 4 explains the wavelet hierarchical multi-scale LBP (the proposed algorithm). Experiments implemented on avatar face datasets are presented in section 5. The comparisons of the proposed algorithm with various other methods are also given in Section 5. Finally, useful conclusions are given in section 6.

II. WAVELET DECOMPOSITION OF AN IMAGE

Wavelet Transform (WT) or Discrete Wavelet Transform (DWT) is a very popular tool for image analysis in the field of image processing. It helps to view and process digital images at multiple resolutions. Its mathematical background and advantages have been discussed in many research articles [17]. The chief advantages of using Wavelet Transforms are listed below:

- It decomposes an image by reducing the resolutions of its sub-images and helps reduce the computational complexity of the system. Harmon [18] demonstrated that the image with 16 x 16 resolution is sufficient to recognize human faces.
- It decomposes images into sub-bands corresponding to different frequency ranges. They easily meet the input requirements for the next major step thus, minimizing the computational overhead in the proposed system.
- Wavelet decomposition provides local information in both spatial and frequency domains, in comparison with Fourier decomposition, which supports only global information in the frequency domain [19]. Thus, they provide spatial and frequency characteristics of the image at the same time.

The main characteristic of wavelets is that they provide multi-resolution analysis of the image in the form of co-efficient matrices. Strong arguments for the use of this multi-resolution decomposition in psychovisual research support evidence that humans process images in a multi-scale way.

The computational complexity of wavelets is linear with the number (N) of computed coefficients (O(N)), while other transformations in their fast implementation have a N $\log_2(N)$ complexity. Thus, wavelets are adapted towards dedicated hardware designs.

The basic functions of wavelet transform are obtained from a single prototype (mother) wavelet by dilation and translation.
The mother wavelet function for the 1-D signal f(t) is shown below:

$$\psi_{m,n}(t) = \frac{1}{\sqrt{m}}\psi\left(\frac{t-n}{m}\right) \qquad (1)$$

This equation [17] can be discretized by restraining m and n to a discrete lattice where $m=2^s$ and $n \in \mathbb{Z}$ with s being the scale. The mother wavelet ψ has to satisfy the admissibility criterion to ensure it is a localized zero-mean function. Typically, some more constraints are imposed on ψ to ensure the transform is non-redundant, complete and constitutes a multi-resolution representation of the original signal.

Fig. 1 gives an idea about the structure of the wavelet coefficient and the first and the second level of wavelet decomposition for one of the avatar face images used in the experiments.

The two-dimensional wavelet transform is performed by applying a one-dimensional wavelet transform to the rows and columns of the two-dimensional data. The one-level wavelet decomposition of an image resulted in an approximation image (LL_1) and three detail images in horizontal (HL_1), vertical (LH_1) and diagonal (HH_1) directions respectively.

The approximation image, obtained by a low-pass filter,

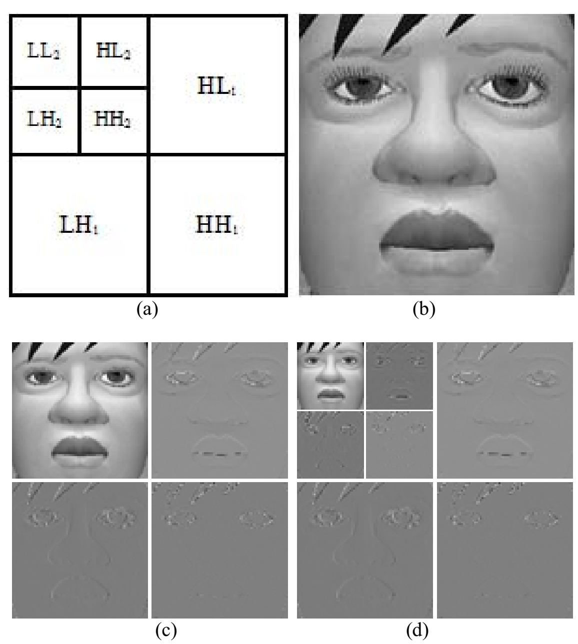


Figure 1. (a) Wavelet coefficient structure [1-3] (b) A sample image of one of the avatar face images in the dataset (c) One level wavelet decomposition for the avatar face image in b (d) Two levels wavelet decomposition for the avatar face image in b.

contains the low-frequency information of the face image and it is used for the next level of decomposition. The detail images contain most of the high frequency information of the face image such as illumination and facial expressions which are called the local changes of the face image [20]. The original image is thus represented by a set of sub-images at several scales [17].

III. LOCAL BINARY PATTERN

The local binary pattern (LBP) operator, introduced by Ojala et al. [21], is a powerful local descriptor for describing image texture and has been used in many applications such as industrial visual inspection, image retrieval, automatic face recognition and detection. The LBP operator labels the pixels of an image by thresholding the value of the central pixel against its surrounding 8 pixels (for a given size of 3x3 neighborhood of each pixel) and considering the result as a binary value [22]. The binary value will be converted to the decimal value to get the LBP value. The output value of the LBP operator can be defined as follows [5, 22]:

$$LBP(x_c, y_c) = \sum_{i=0}^{7} 2^i S(g_i - g_c) \qquad (2)$$

where g_c corresponds to the gray value of the central pixel, (x_c, y_c) are its coordinates, g_i $(i = 0,1,2,..,7)$ are the gray values of its surrounding 8 pixels and $S(g_i - g_c)$ can be defined as follows:

$$S(g_i - g_c) = \begin{cases} 1, g_i \geq g_c \\ 0, otherwise \end{cases} \qquad (3)$$

So we can say that LBP is an ordered set of binary comparisons between the central pixel value and the values of its neighborhood pixels [5]. Fig. 2 gives an illustration of the basic LBP operator and how to compute the LBP value.

The LBP operator can be extended to use pixels from neighborhoods of different sizes [3, 5, 20]. Fig. 3 gives us some examples of different LBP operators where R is the radius of the neighborhood and P is the number of pixels in that neighborhood.

The neighborhood can be either in a circular or square order. Using the circular order neighborhood allows any radius and number of the pixels in the neighborhood [23].

One of the most important and successful extensions to the basic LBP operator is called uniform LBP (ULBP). An LBP is called uniform when it contains at most two different conversions from 0 to 1 or 1 to 0 when the binary string is

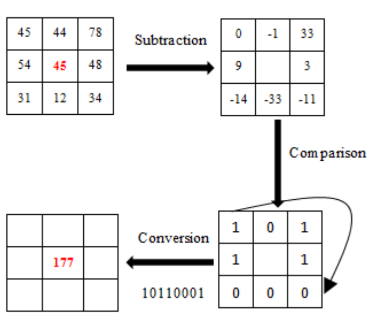


Figure 2. The basic LBP operator [3, 5]

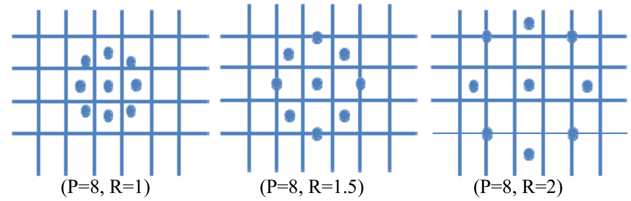


Figure 3. Three different LBP operators [3]

viewed as a circular bit string [3, 23]. For example, 11111111, 00011000 and 11110011 are uniform patterns. Ojala reported that with P = 8 and R = 1 neighborhood, uniform patterns account for around 90% of all patterns and with P =16 and R = 2 neighborhood, uniform patterns account for around 70% of all patterns [3]. So only a little amount of information will be lost when using uniform patterns [5].

After labeling an image using the LBP operator, the histogram of the labeled image can be defined as follows [5]:

$$H_i = \sum_{x,y} I(f(x, y) = i), i = 0,1,.., n-1 \qquad (4)$$

where 'n' is the number of different labels produced by the LBP operator, $f(x, y)$ is the labeled image and $I(A)$ is a decision function with value 1 if the event A is true and 0 otherwise.

LBP histogram has very useful information about the distribution of the local microstructures, such as spots and edges, over the whole image and so can be used to describe and represent the global characteristics of the image [5, 20].

IV. WAVELET HIERARCHICAL MULTI-SCALE LBP (WHMLBP)

The proposed algorithm has three steps: preprocessing, feature extraction and recognition or classification.

A. *Preprocessing Face Image*

To improve the efficiency of extracting the face features we have to apply a set of preprocessing operations. First, we manually cropped the input images to pure face images by removing the background which is not useful in recognition. Second, these pure face images have to be normalized and then decomposed using the first level of wavelet decomposition to obtain pure facial expression images (See Fig. 4).

Detailed images resulting from applying wavelet decomposition contain changes which represent the difference of face images. So considering only the approximation images will enhance the common features of the same class of images and at the same time the difference will be reduced. For this reason, our experiments were concerned only with the approximation images resulting from the first level of wavelet

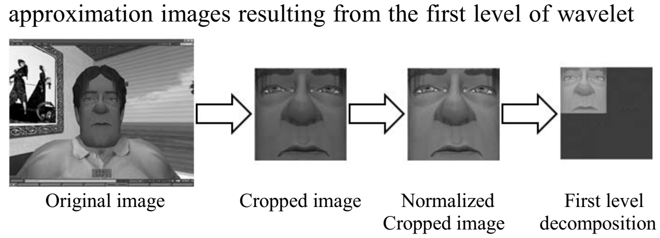


Figure 4. Face image preprocessing

decomposition and which we used in testing to evaluate the performance of the proposed algorithm.

B. HMLBP Feature Extraction

The performance of the multi-scale or multi-resolution LBP operator is better than the performance of a single scale LBP operator for many reasons, such as:

a- Multi-scale operator can help to extract more image features under different settings [4]. Calculating features based on a limited size neighborhood in single scale LBP may lead to inadequate capture of dominant features of an image.

b- As a result of single scale LBP operator "non-uniform" patterns are clustered into one non-uniform pattern. As the radius of the LBP increases, the cluster size of the "non-uniform" patterns increases as well, leading to a substantial loss of information [4].

Some work [4] was carried out towards extracting more useful features from the image by digging out information from the "non-uniform" patterns. Such methods are based on a training step to learn the useful patterns and so the training samples have a great effect on the accuracy of recognition [4].

In HMLBP algorithm the LBPs for the biggest radius is extracted first. The new LBPs of "non-uniform" patterns have to be extracted further using a smaller radius to extract "uniform" patterns. This process continues until the smallest radius is processed. This hierarchical scheme does not have a training step and thus it is insensitive to training samples [4].

Fig. 5 shows an example of the hierarchical multi-scale LBP scheme. The LBP histogram for R=3 is first built. For those "non-uniform" patterns of the R=3 operator, a new histogram is built by the R=2 operator. Then, the "non-uniform" patterns of R=2 lead to the histogram building process for the R=1 operator. Finally the three histograms are concatenated into one multi-scale histogram to form the feature histogram of an image [4].

C. Dissimilarity Measure

The last stage of our proposed algorithm is to classify each facial image to its class by computing the dissimilarity between training samples and a test (input) sample. To do that we apply Chi-Square distance as follows [3]:

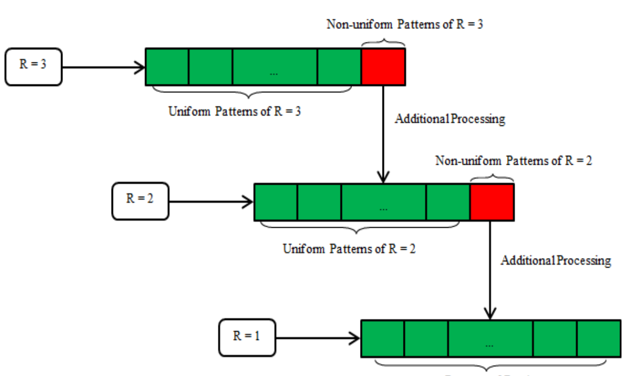


Figure 5. An example of hierarchical multi-scale LBP Scheme [4]

$$D(X,y) = \sum_{n=1}^{N} \frac{(X_n - Y_n)^2}{X_n + Y_n} \tag{5}$$

where X is the tested image (sample), Y is the training sample(s) or image(s) and N is the sum dimension.

V. EXPERIMENTAL RESULTS AND ANALYSIS

To ensure the efficiency of the proposed method, two virtual world datasets are used to test the performance of the proposed method. This is the first time given algorithm is used on the gray scale images and consequently there is no baseline results available for direct comparison.

The first dataset, from Second Life virtual world, contains 581 (1280 x 1024 pixels) gray scale images of 83 avatars. The second dataset, from Entropia virtual world, consists of a total of 490 (407 x 549 pixels) gray scale images representing 98 avatars.

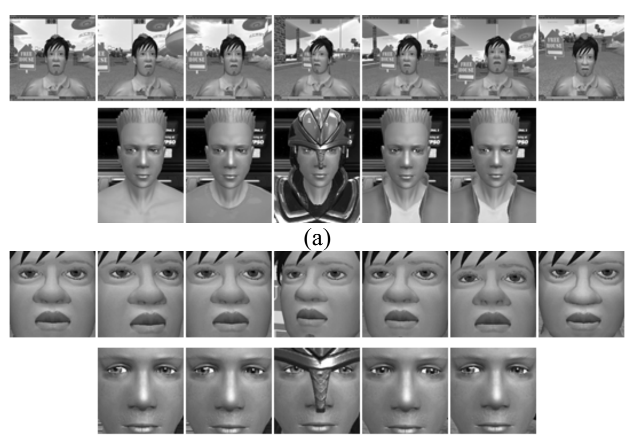
We tested these datasets with three well-known algorithms (LBP, WLBP and HMLBP) and compared their result with the results coming from the proposed method.

A. Experimental setup

All images in the Second Life dataset are manually cropped to 260x260 pixels while images in Entropia dataset are manually cropped and resized to 180x180 pixels. The resulted 581 Second Life avatar face images dataset is organized into 83 classes each of which has 7 face images of the same avatar with different frontal angles (front, far left, mid left, far right, mid right, top and bottom). So we can say that the Second Life avatar face images dataset focuses on pose angle and facial expression.

The resulted Entropia avatar face dataset is organized into 98 classes each of which has 5 avatar face images. In one of them the avatar is wearing a mask while in the others the avatar has different facial expressions and eye angles. See Fig. 6 for an example of two classes of avatars (one from each dataset) before and after cropping.

The resolution of images used in the experiments is changed from 260x260 to 130x130 pixels (for Second Life dataset) and from 180x180 to 90x90 (for Entropia dataset) using the first level of wavelet decomposition. The avatar face images in both datasets are preprocessed and prepared for feature



(a)

(b)

Figure 6. a. Two classes of unprocessed avatar images. b. The same two classes after cropping the avatar faces.

extraction step. HMLBP is used to extract the best descriptive features and then at the end the Chi-Square measure is applied to accomplish classification. The experiments are performed on the condition of a single training image. Each time one image is used as a trainer. The Chi-Square distance computes the dissimilarity between this image and all other images in the dataset. These distances will then be ordered in an ascending order. The 6 images (for Second Life dataset) associated to the least 6 distances in the ascending order will be checked if they are from the same class of the trained image or not. The same will be done but with only 4 images for the Entropia dataset. Based on the number of corrected classified images we can compute the accuracy for each dataset using the following formula: classification accuracy (CA) or recognition rate (RR) equation:

$$RR = \frac{number\ \ of\ \ corrected\ \ classified\ \ images}{total\ \ number\ \ of\ \ samples\ \ in\ the\ dataset}\ x100\,\%$$

B. *Comparing WHMLBP with HMLBP and other algorithms*

In order to gain better understanding on whether using wavelet transform with HMLBP is advantageous or not we compared WHMLBP with HMLBP, WLBP and LBP with several experiments. First we got the performance of WHMLBP with different block size with R = [3, 2, 1] and P = [16, 16, 16] as we can see in Fig. 7.

We can see that changing the block size affects the result of the recognition rate. In Fig. 7, the recognition rate is increased as the block size is larger, and the performance is dropped as the block size is larger than 42x42 on the two datasets, that is because dense blocks obscure the image features.

As a result we compare the performance of WHMLBP and HMLBP using 42x42 block size with the same radius R = [3, 2, 1] and a different neighborhood size for the two datasets as in Fig. 8. The experimental results showed that the recognition rate of WHMLBP increases about 4% to 5% in Second Life dataset and the greatest accuracy is about 80.03% when the neighborhood size is 24*24*24. And in the Entropia dataset, almost all the cases are better than using HMLBP while the accuracy rate increases about 1%. The average of the recognition rate of the two methods for both datasets using different neighborhood sizes can be seen in table I.

To compare the performance of WHMLBP method with other methods, we applied WLBP and LBP methods on the same two datasets. We applied both methods with R = 1, 2, 3 and P = 8, 16, 24 and we got the average of the recognition

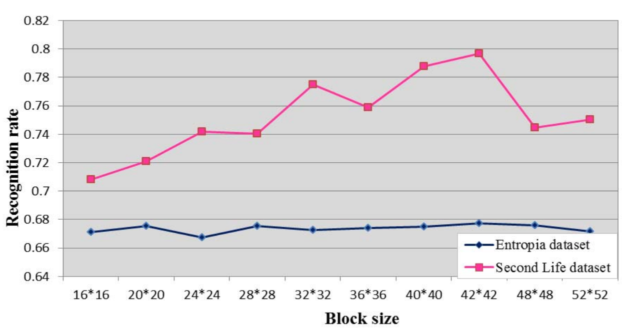


Figure 7. Performance of WHMLBP with different block size.

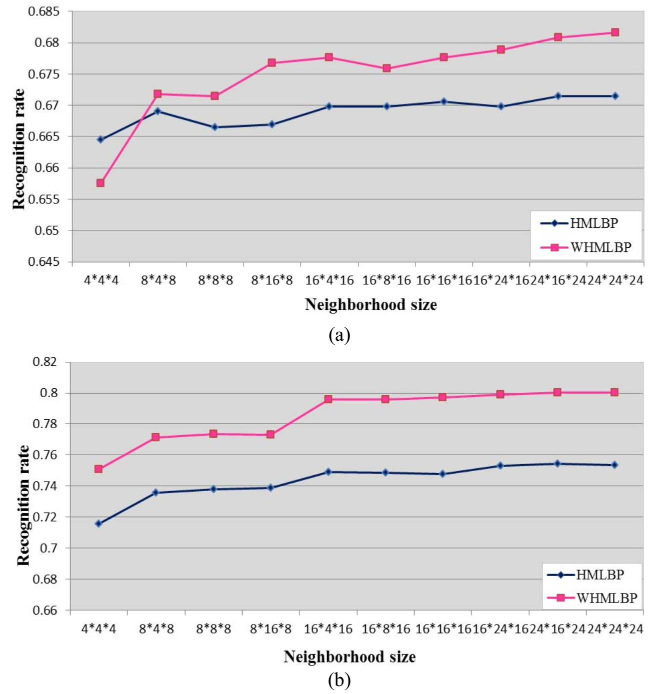


Figure 8. The RR of WHMLBP and HMLBP on: (a) Entropia dataset (b) Second Life dataset.

Table I. Average Recognition rate for different algorithms

Dataset	LBP	WLBP	HMLBP	WHMLBP
Second Life	67.42%	77.27%	74.34%	78.57%
Entropia	66.45%	65.78%	66.90%	67.50%

rate for both datasets as in table I.

The results we obtained demonstrate the effectiveness of our algorithm in comparison to other algorithms.

VI. CONCLUSION

In this paper, to improve the efficiency of the HMLBP in extracting useful features from an image we applied wavelet transform to the normalized manually cropped images. The effectiveness of this proposed method is shown in two avatar face datasets. Compared with HMLBP method, the proposed method gets more than 4% improvement in the first dataset and about 1% improvement in the second one. Compared with two other well-known methods (LBP and WLBP) the proposed method gets higher recognition rate.

Applying other classifiers may lead to better results and this is what we intend to attempt in the future using larger datasets from different virtual worlds. The final goal is to build a complete automatic system for avatar face detection and recognition.

REFERENCES

[1] Wang, H., S. Yang, and W. Liao, *An improved PCA face recognition algorithm based on the discrete wavelet transform*

and the support vector machines, in *International Conference on Computational Intelligence and Security Workshops* 2007: Harbin, Heilongjiang, China. p. 308-311.

[2] Luo, B., Y. Zhang, and Y.-H. Pan, *Face recognition based on wavelet transform and SVM*, in *IEEE International Conference on Information Acquisition* 2005: Hong Kong and Macau, China. p. 373-377.

[3] Ahonen, T., A. Hadid, and M. Pietikainen, *Face description with local binary patterns: application to face recognition.* IEEE Transactions on Pattern Analysis and Machine Intelligence 2006. **28**(12): p. 2037-2041.

[4] Zhenhua, G., Z. Lei, D. Zhang, and M. Xuanqin, *Hierarchical multiscale LBP for face and palmprint recognition*. in *International Conference on Image Processing (ICIP)* 2010. Hong Kong. p. 4521-4524.

[5] Wang, W., F. Chang, J. Zhao, and Z. Chen, *Automatic facial expression recognition using local binary pattern*, in *8th World Congress on Intelligent Control and Automation* 2010: Jinan, China. p. 6375-6378.

[6] *Second Life*. Available from: www.secondlife.com.

[7] *Active Worlds*. Available from: www.activeworlds.com.

[8] *Entropia Universe*. Available from: www.entropiauniverse.com.

[9] Klare, B., R.V. Yampolskiy, and A.K. Jain, *Face recognition in the virtual world: recognizing avatar faces*, MSU Technical Report, MSU-CSE-11-2, 2011.

[10] O'Brien, N. *Spies watch rise of virtual terrorists*. 2007; Available from: http://www.news.com.au/top-stories/spies-watch-rise-of-virtual-terrorists/story-e6frfkp9-1111114075761.

[11] Ajinal, S., R.V. Yampolskiy, and N.E.B. Amara, *Avatar facial biometric authentication*, in *2nd International Conference on Image Processing Theory, Tools and Applications* 2010: Paris. p. 2-5.

[12] Gavrilova, M.L. and R.V. Yampolskiy. *Applying biometric principles to avatar recognition*. in *International Conference on Cyberworlds (CW)*. 2010. Singapore. p. 179-186.

[13] Yampolskiy, R.V. and V. Govindaraju, *Behavioral biometrics for verification and recognition of malicious software agents*, in *Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense VII. SPIE Defense and Security Symposium* 2008: Orlando.

[14] Oursler, J.N., M. Price, and R.V. Yampolskiy, *Parameterized generation of avatar face dataset*, in *14th International Conference on Computer Games: AI, Animation, Mobile, Interactive Multimedia, Educational & Serious Games* 2009: Louisville, KY. p. 17-22.

[15] Boukhris, M., A.A. Mohamed, D. D'Souza, M. Beck, N.E.B. Amara, and R.V. Yampolskiy, *Artificial human face recognition via daubechies wavelet transform and SVM*, in *16th International Conference on Computer Games: AI, Animation, Mobile, Multimedia, Educational and Serious Games* 2011: Louisville, KY, USA. p. 18-25.

[16] Mohamed, A.A. and R.V. Yampolskiy, *An improved LBP algorithm for avatar face recognition*, in *23th International Symposium on Information, Communication and Automation Technologies* 2011: Sarajevo, Bosnia & Herzegovina. in press.

[17] Christophe, G., Z. Giorgos, and T. Giorgos, *A wavelet-based framework for face recognition*, in *5 th European Conference on Computer Vision (ECCV'98)* 1998: Freiburg, Germany. p. 84-92.

[18] Harmon, L., *The recognition of faces.* Scientific American, 1973. **229**(5): p. 71-82.

[19] Mazloom, M. and S. Ayat. *Combinational method for face recognition: wavelet, PCA and ANN*. in *Digital Image Computing: Techniques and Applications*. 2008. Canberra. p. 90-95

[20] Liu, X., M. Du, and L. Jin, *Face features extraction based on multi-scale LBP*, in *2nd International Conference on Signal Processing Systems (ICSPS)* 2010: Dalian. p. 438-441.

[21] Ojala, T., M. Pietikainen, and D. Harwood, *A comparative study of texture measures with classification based on feature distributions.* Pattern Recognition, 1996. **29**(1): p. 51-59.

[22] Meng, J., Y. Gao, X. Wang, T. Lin, and J. Zhang, *Face recognition based on local binary patterns with threshold*, in *IEEE International Conference on Granular Computing* 2010: San Jose, CA. p. 352-356.

[23] Tan, N., L. Huang, and C. Liu, *Face recognition based on a single image and local binary pattern*, in *2nd International Symposium on Intelligent Information Technology Application* 2008: Shanghai. p. 366-370