Real-time Human Age Estimation based on Facial images using uniform Local Binary Patterns

Mohamed Selim[1], Shekhar Raheja[2] and Didier Stricker[1]
[1]*German Research Center for Artificial Intelligence, DFKI, Kaiserslautern, Germany*
[2]*Technical University Kaiserslautern, Germany*
{*mohamed.selim, didier.stricker*}*@dfki.de, s_raheja12@informatik.uni-kl.de*

Keywords: Age-Group Estimation, Local Binary Patterns, Extended Local Binary Patterns, Real-Time, K-Nearest Neighbours

Abstract: This paper summarizes work done on real-time human age-group estimation based on frontal facial images. Our approach relies on detecting visible ageing effects, such as facial skin texture. This information is described using uniform Local Binary Patterns (LBP) and the estimation is done using the K-Nearest Neighbour classifier. In the current work, the system is trained using the FERET dataset. The training data is divided into five main age groups. Facial images captured in real-time using the Microsoft Kinect RGB data are used to classify the subjects age into one of the five different age groups. An accuracy of 81% was achieved on the live testing data. In the proposed approach, only facial regions affected by the ageing process are used in the face description. Moreover, the use of uniform Local Binary Patterns is evaluated in the context of facial description and age-group estimation. Results show that the uniform LBP depicts most of the facial texture information. That led to speeding up the entire process as the feature vector's length has been reduced significantly, which optimises the process for real-time applications.

1 INTRODUCTION

In recent years, interest toward age-group estimation using facial images has increased. Age information can be very useful in various applications such as Human-Computer Interaction, digital signage displays in shopping stores, and security surveillance. This can be done using complex approaches that require long processing time. In this paper, we present work on age-group estimation from facial images using features that are not computationally expensive, and can be done in real-time.

In the current work, we aim at developing a real-time age-group estimation system where we track and capture faces in real-time. Estimation is carried out using tracked facial images and extracting skin texture information from them using Local binary patterns (LBP) operator. This is followed by classification of a person's age into five groups: *13-20, 21-30, 31-40, 41-50* and *51-above*. We use images from FERET (Phillips et al., 2000) database for training the system. Due to insufficient number of samples for persons in age-group 1-12, this group is not involved in the system.

Classification is carried out using the k-Nearest Neighbours algorithm. Finally, we look into reducing the dimensionality of the feature vectors used for classification to optimise the system for real-time applications. This is achieved by considering only specific regions of face which are affected the most due to ageing. This helps in reducing the processing time and speeding up the age-estimation process significantly and is discussed more in detail in section 3.3.

The system uses frontal facial images from the FERET dataset in the training process. Various training criterion has been performed throughout the work. First, the data was divided into classes based on the subject age. Later, more in depth separation of the data was performed where each age group was divided into sub-classes based on different criterion. For example, the senior age-group was divided into two classes based on the gender (males-seniors, females-seniors). This will be discussed later in further details in section 4.

Preliminery results show that the system trained using FERET dataset, and tested on live data(RGB data) captured by the Microsoft Kinect camera reached accuracy of roughly 81% of correct age-group classification. The processing time including the face and feature extraction was on average 15 milliseconds, which is suitable for real-time applications.

This paper is organized as follows, in section 2, we present related work and background information related to this work. In section 3, we present our approach in details for human age-group estimation. Following that we discuss the approach in section 5. In section 4 the results of various training and testing are presented. Finally, in section 6 conclusions are presented with ideas for future work.

Our contribution is estimating the age group of a subject from its facial image based on simple LBP features applied on specific regions on the face that are affected by the aging process. The time needed to process a frame is around 15 milliseconds, thus our approach is also suitable for real time applications.

2 RELATED WORK

The basic approach of almost all work in the area of age estimation till now is more or less the same as shown in the age-estimation pipeline in Fig. 1.

As shown in Fig. 1 an image containing face is fed into the system. Some pre-processing operations like cropping, smoothing and resizing are carried out on the input image. The pre-processed image is then passed for feature extraction. Subsequently, machine learning techniques are then used to classify these images into age-groups.

(Lanitis et al., 2004) carried out a survey to compare various age-estimation techniques like, quadratic functions, shortest distance classifier and neural networks. Their results showed that classifiers based on quadratic functions and neural networks obtained the best results. However, quadratic functions do not always map the ageing process perfectly since each individual ages in a different way and hence we decided to work with neural networks.

Work of (Hewahi et al., 2010) relies on use of Multilayer Perceptrons (MLPs) and facial landmarks along with gender information. They focus on classification of ages into four age-groups and sub-dividing each group into two for more fine classification. (Kohail, 2012) uses the same approach with a little modification. This modification basically dealt with combining individual *x,y*-coordinates into one meaningful unit using Singular Value Decomposition (SVD), before feeding them to neural network. This approach relies on facial landmark detection on the subject face which might not be always detected in live camera feed due to image quality or face size.

Another approach which doesn't rely on specific landmark points on the face, but takes into consideration the facial skin texture which varies in accordance with age. If this variation information can be described, it can serve as a very good feature for age-estimation. We use Local Binary Patterns (Zhao and Pietikäinen, 2007) as a descriptor for the facial texture information.

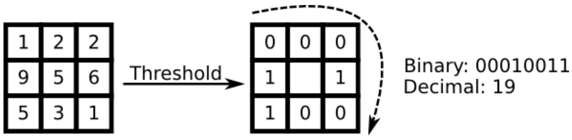

Figure 2: Local Binary Patterns [1]

LBP (*see* Fig. 2) is an operator that works on pixels. It considers the neighbourhood of each pixel and assigns a label to this pixel after thresholding with the neighbouring pixel values. It is based on the following equation:

$$LBP_{P,R}(x_c) = \sum_{p=0}^{p-1} u(x_p - x_c)2^p, \quad u(y) = \begin{cases} 1 & \text{if } y \geq 0 \\ 0 & \text{if } y < 0 \end{cases} \tag{1}$$

where, x_c is pixel under consideration, x_p represents neighbouring P pixels, R is a radius and $u(.)$ is the step function. Since we consider 8 neighbours, this results in an 8-bit binary pattern. In total there can be 2^8 or 256 different binary patterns. Finally, the pixel under consideration is set to the decimal equivalent value of its corresponding binary pattern.

(Gunay and Nabiyev, 2008) used this approach to classify ages into 10-year intervals. However, they also used some weighting mechanism on the image according to importance of the information in a specific region of the face. The weights on different regions of image are decided approximately. (Karthigayani and Sridhar, 2011) also focus on implementing Local binary patterns (LBP) with artificial neural networks. (Fard et al., 2013) implemented Local binary patterns (LBP) on 3 regions of face that are more affected by ageing and showed considerable improvement. Also, (Ylioinas et al., 2012) applied LBP and LBP with contrast measure on 9 regions on the face. However, in our approach, we consider more regions with finer granularity that are also affected by ageing effects.

As before, the face is divided into 8×8 patches and histograms are generated for each patch and concatenated together. The final histogram is the input feature vector. However, this time we focus on reducing dimensionality of this feature vector and we implement the concept of uniform and non-uniform local binary patterns.

A local binary pattern is called uniform if the binary pattern contains at most two bitwise transitions from 0 to 1 or vice versa when the bit pat-

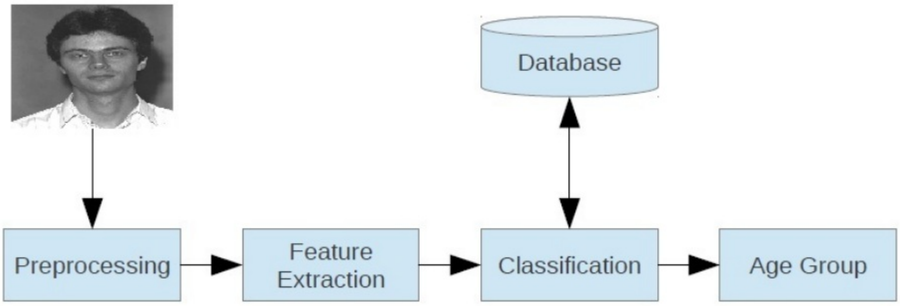

Figure 1: General Age-Group Estimation Pipeline

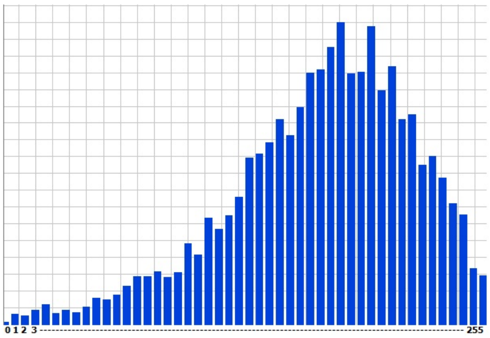

Figure 3: 256-bin histogram(not exact count)

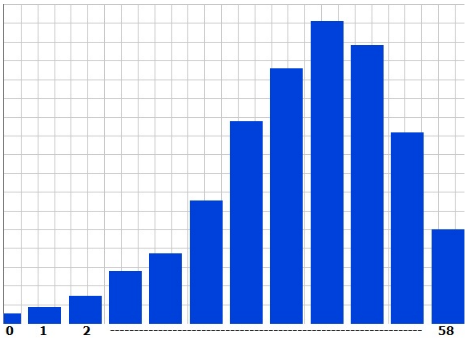

Figure 4: 59-bin histogram(not exact count)

tern is traversed circularly. For example, the patterns 00000000 (0 transitions), 01110000 (2 transitions) and 11001111 (2 transitions) are uniform whereas the patterns 11001001 (4 transitions) and 01010010 (6 transitions) are not. In the computation of the LBP labels, there is a separate label for each uniform pattern and all the non-uniform patterns are labelled with a single label. For example, when using (8,R) neighbourhood, there are a total of 256 patterns, 58 of which are uniform, which yields in 59 different labels.

3 PROPOSED APPROACH

Our approach relies on detecting features that can depict age-variations (e.g. skin texture, wrinkles, age-spots, etc.). These features are described using Local Binary Patterns (Zhao and Pietikäinen, 2007). The face is divided into 8x8 regions.

3.1 Image Preprocessing and Face Extraction

Preprocessing the input image is a vital step. Every image that is fetched from the database undergoes various preprocessing steps as shown in Fig. 5.

After the image is input to the system, the face is detected using (Viola and Jones, 2001) haar-cascades. Following the face detection, the two eyes are detected using also haar-cascades. The face is then rotated so that the two eyes lie on the same horizontal line.

Cropping is an important step in the pipeline as the face is the input to the LBP deescriptor. The face detected by the haar-cascade face detector is not always consistent (same cropping), thus, the face is cropped using the distance between the eyes. The width and height of the face is defined as in figure 6. Those parameters were chosen to crop the face excluding the ears and any background part behind the face. The cropped out image is then smoothed by adding little Gaussian blur and re-sized finally to a predetermined size of 150x150.

3.2 LBP Features

In order to optimize the approach for real-time age-group estimation, the size of the computed feature vector was reduced. First, the size of the LBP histogram is reduced from 256 bins to 59 bins. Moreover, the local binary pattern operator is applied to selected facial regions. Details on how this is done are explained in the following paragraphs.

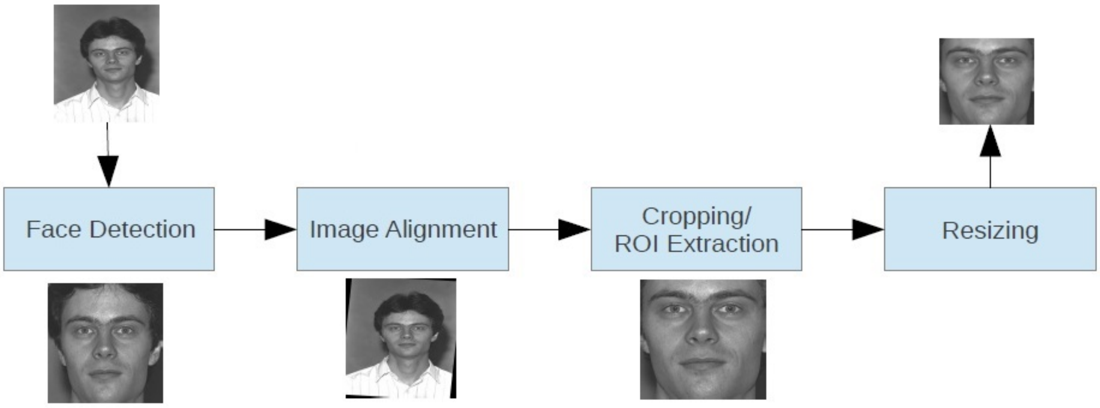

Figure 5: Image Preprocessing and Face Extraction Pipeline

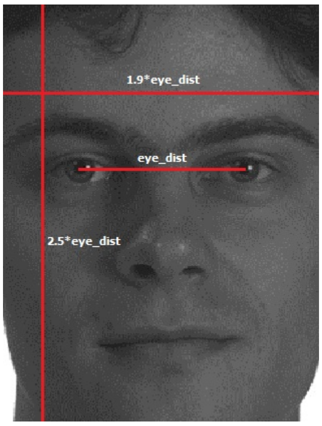

Figure 6: Face cropping using distance between eyes as metric

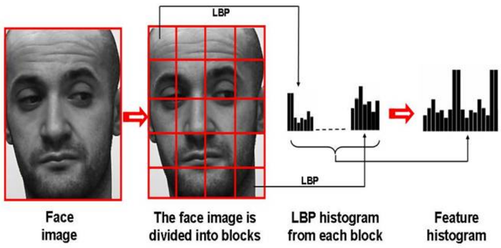

Figure 7: Histogram generation from 8x8 patches [2]

Instead of using normal Local Binary Patterns, where the histogram consists of 256 bins, uniform Local Binary Patterns are used. As mentioned earlier, the uniform LBP histogram consists of only 59 bins, where 58 bins are for the 58 uniform patterns, and 1 bin for all non uniform ones. Using the uniform LBP as descriptor for the facial images must be evaluated in the domain of facial images and for the application of age-group estimation. A comparison is carried out in the results section (see section 4) to evaluate the accuracy of the system using normal LBPs vs. uniform LBPs.

The facial image is divided into 8×8 patches and LBP operator is applied on each patch one by one. For every patch, a 59-bin histogram is generated. These histograms are concatenated and yields one large histogram of size (64×59) 3776. This resulting histogram contains entire texture information for one facial image. Histograms are generated for all the images in training and test data set.

As described by (Fard et al., 2013), certain regions of face are much better indicators of age as compared to others. Regions like, forehead, cheeks, eye-corners, etc. reflect ageing effects the most. On the other hand, regions like eyes, eyebrows and lips hardly give any information about age. Also as the face is not in a rectangular shape, there is always some area besides the chin which is also part of detected face. This is not useful information and this region is eliminated as well. Considering this fact, we decide to work with only selected patches as shown in Fig. 8.

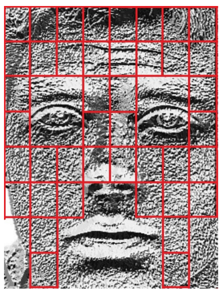

Figure 8: Selective Facial Regions on a LBP image

The regions which are covered by a square are the ones considered for age-estimation. This approach eliminated 24 patches, leaving behind only 40 patches as compared to 64 patches used earlier. So, now the histograms are of (40x59) 2360 bins. Earlier this size was (64x59) 3776 bins. This results in histograms or the input feature vectors which are 62.5% of previous size. This also made sure that the processing time came down which is a crucial factor for real-time performance.

In the classification phase, minimum distance criteria is used for matching and classification between histogram of test image and training images.

3.3 K-Nearest Neighbour Classifier

Using the nearest neighbour classifier might results in misclassification. However, if we can make a decision by considering more number of neighbours, then classification accuracy is expected to go higher. Hence, we work with k-nearest neighbours. The system follows a voting strategy and assigns the class with maximum votes to the current sample. More details about choosing the appropriate K value is discussed in section 4.

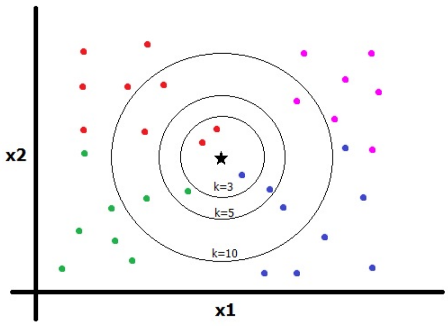

Figure 9: k-Nearest Neighbour classifier with different values of *k*

3.4 Training with FERET Dataset

FERET (Phillips et al., 2000) database consists of facial images from people of different ages and different ethnicities. It has two sets of frontal facial images, *fa* set and *fb*. For each subject the *fb* image was taken seconds after the *fa* one.The *fa* images are of neutral expression whereas, in the *fb* the subjects have slight facial expression.

[2]source: www.scholarpedia.org/article/Local_Binary_Patterns

Figure 10: fa image

Figure 11: fb image

The FERET dataset is used for training and testing. Prior to age-estimation, some images are discarded in preprocessing stage for which faces are not detected properly. Remaining images in the database are used for training and testing purpose. For uniform distribution of images into training and test sets, we carry out 10-fold cross-validation. With each iteration, 90% of images are taken as training set and remaining 10% for test set. A point to note here is that, we apply 10-fold cross-validation to all the four age-groups. Hence, it is ensured that, with every iteration, training set contains 90% samples from all the groups and test set contains 10% samples from each age-group. It is also ensured that no single image is simultaneously present in both the sets. The age groups are divided as shown in table 1.

Age-group 1	13-20
Age-group 2	21-30
Age-group 3	31-40
Age-group 4	41-50
Age-group 5	51-above

Table 1: Age groups.

3.5 Live Feed from Microsoft Kinect

We tried our new system for age-estimation in real-time. We used Microsoft Kinect (Microsoft, 2012) to grab images in real-time. We chose to work on higher resolution of Kinect, 1280x960 so that fine details of facial texture can be captured. Some rotation is required to align the images with the horizontal and for this we need to first calculate the tilt. This tilt or angle of rotation is determined with the help of eye coordinates. Haar cascades are used to find the eye locations in these images. The grabbed images are then rotated by the angle found earlier to make sure it is aligned with the horizontal. After this, the distance between eyes is used as a metric for cropping the face. Finally, this extracted face image is then passed on for

Figure 12: Sample images from FERET dataset with age-group classification and age ground truth.

Local Binary Patterns (LBP) and corresponding histogram computation. Once these are computed, the histogram is passed to the kNN classifier to predict its age-group.

4 RESULTS

Various tests were carried out in this work. First, the results of testing the simple LBP operator -that yeilds a histogram of 256 bins- over the data without selective regions on the face are shown. During those tests, the data was divided into four groups only (Kids, Young, Adults, Seniors), as shown in table 2.

Age-Group	Kids	Young	Adults	Seniors
Age Range	1-10	11-25	26-50	50-above
Count	18	525	726	52

Table 2: Age-Groups in the first tests.

We tested the system on mixed dataset i.e, datasets containing both male and female subjects and also on datasets containing only males or only females. Some sample images of the tests are shown in figure 12. It was noticed, that the system was not able to classify samples from *Kids* and *Seniors* age-group properly. This is mainly because the FERET database contains few number of images for these two age groups compared to *Young* and *Adults* group. Following are results obtained (*see* Table 3) for training and testing using Local Binary Patterns (LBPs).

Accuracy	Mixed	Males	Females
Average	61.7453%	59.6787%	71.0657%
Kids	-	-	-
Young	63.6538%	60.3846%	76.1538%
Adults	65.4167%	64.8%	71.4286%
Seniors	24.0%	22.5%	-

Table 3: Age-group estimation results for mixed and gender specific datasets

In table 4 the error-distribution for above results are shown. Here, *Error of 1* means the error by nearest age-group or error by 1-class gap. Similarly, *Error of 2* means the error by second-nearest age-group or error by 2-class gap. And, *Error of 3* means the error by farthest age-group or error by 3-class gap.

Error %	Mixed	Males	Females
Error of 1	96.0396%	96.0843%	97.2028%
Error of 2	3.9604%	3.9156%	2.0979%
Error of 3	0%	0%	0.6993%

Table 4: Error Distribution for Approach 2

One thing that we noticed from the results above was that, considering only misclassified samples, more than 96% are misclassified by nearest class. Only around 3% are misclassified by second nearest age-group and negligible amount for the farthest age-group. We strongly believe that, the system is not able to differentiate properly for boundary values.

We also tried, to test our system on images free from eye-glasses and facial hair like beard and moustache. Since, Local Binary Patterns operate at the pixel level, presence of eye-glasses and facial hair can affect the performance as it changes the texture. Results obtained for this test are shown in Table 5 and Table 6.

Accuracy	Mixed	Males	Females
Average	63.108%	59.5125%	72.7419%
Kids	0%	0%	-
Young	68.4444%	64.2857%	82.0833%
Adults	65.625%	65.0%	71.0526%
Seniors	10.0%	5.0%	-

Table 5: Age-group estimation results for images without eye-glasses and facial hair

The results of the constrained images (no glasses and no facial hair) are slightly higher than the unconstrained ones. However, in order to improve the results of the age-group classification, more divisions were carried on the Young and Adults age group as shown in table 1. Also, uniform patterns were used.

Error %	Mixed	Males	Females
Error of 1	96.4674%	95.3917%	96.8%
Error of 2	3.5326%	4.6082%	2.4%
Error of 3	0%	0%	0.8%

Table 6: Error Distribution

Finally, selected face regions were used in the training and testing of the following results. The results have the average accuracy of the whole data with different K-Value.

Using normal LBP Histograms, the results are shown in table 7

Results	Unconstrained Data				
	K=1	K=5	K=10	K=20	K=30
All LBP	63.2%	69.8%	**70.0%**	69.5%	67.6%
Uniform	62.1%	69.3%	**70.6%**	69.3%	68.2%
	Constrained Data				
	K=1	K=5	K=10	K=20	K=30
All LBP	63.2%	68.9%	**69.3%**	68.1%	65.2%
Uniform	61.0%	68.0%	**68.5%**	67.1%	66.8%

Table 7: Accuracy achieved with LBP with different parameter values

Clearly, the results are higher than those of the 4-groups data division that was carried out before. For more in-depth analysis, more test were carried out using the Extended Local Binary Patterns with different values for the radius (1 and 2 pixels). The results are shown in the following tables.

Results	Unconstrained Data				
	K=1	K=5	K=10	K=20	K=30
All LBP	62.9%	69.4%	**70.3%**	69.0%	67.8%
Uniform	54.7%	66.9%	**66.4%**	67.5%	65.5%
	Constrained Data				
	K=1	K=5	K=10	K=20	K=30
All LBP	62.4%	68.1%	**68.9%**	67.3%	65.9%
Uniform	57.3%	66.2%	**65.9%**	65.9%	64.0%

Table 8: Accuracy achieved with ELBP (1,8) Radius 1, 8 neighbouring pixels.

Looking at the results, The maximum accuracy reached was at K of value 10. Consequently, the K value was fixed to 10. Also, using the 59 uniform patterns yields nearly the same accuracy as using the full Local Binary Patterns. Thus, the uniform patterns contribute the most towards the correct classification, and at the same time, reduces the size of the feature vector dramatically. Consequently, uniform patterns were chosen for the real time system.

Results	Unconstrained Data				
	K=1	K=5	K=10	K=20	K=30
All LBP	64.4%	69.3%	**71.1%**	68.5%	67.6%
Uniform	65.4%	69.3%	**70.9%**	68.3%	68.5%
	Constrained Data				
	K=1	K=5	K=10	K=20	K=30
All LBP	64.1%	68.6%	**68.2%**	66.1%	65.0%
Uniform	64.9%	68.7%	**70.4%**	67.6%	66.2%

Table 9: Accuracy achieved with ELBP (2,8) Radius 2, 8 neighbouring pixels.

Real-time testing Real-time test was carried out at the DFKI. We tested with 20 subjects (3 females and 17 males). The subjects were facing the Kinect(Microsoft, 2012) camera at a distance between 1 meter to 1.5 meters. The subjects were all between or nearby to age-group $21-30$. The system performed quite well on those subjects, with correct age-group estimation for nearly 81% cases.

5 DISCUSSION

Considering the results in Table 7, Table 8 and Table 9, the following discussion points can be drawn:

- Accuracy is equally good even when we work with only 59 uniform patterns instead of all the 256 local binary patterns. This means, that its the 58 unique uniform patterns that contribute maximum towards classification. All the remaining 198 non-uniform patterns are put into the same bin in histogram and still it does not affect the results.
- The accuracy of the system improved with increased number of neighbours in the classifier that are contributing towards classification.
- We achieved equally good or better accuracy with significantly reduced dimensionality.
- Smaller histograms means processing time is less and hence the system is fast.

Based on the results, the proposed approach of using uniform Local Binary Pattern is suitable for human age-group estimation in real time. Since the time taken on average for a single prediction with is 15 milliseconds, which is appropriate in the context of real-time performance.

6 CONCLUSION

As it is clear from experimental results, using uniform Local Binary Patterns (LBP) as a descriptor is suitable for the age-group estimation based on facial images. That is due to the face that skin texture is directly affected by ageing, and the texture is described using the LBP. Since, time taken by for single prediction using Local binary patterns is in the order of few milliseconds, this was found suitable for age-group estimation in real-time.

Also, the reduced dimensionality of input feature vector by using uniform patterns and selected facial regions has significantly reduced the processing time which is a major source of concern for real-time performance.

Regarding future work, it can be summarized in the following points:.

- Include more datasets that have more samples in Kids and Seniors age groups.
- Use higher resolution camera that can enable the functionality of the system at further distances(>2m).
- Use regression to deal with the values of the age at the boundaries of the age-groups
- Test with different poses of the face.
- Normalize the lighting of the faces using Self-Quotient Images (Wang et al., 2004).

ACKNOWLEDGEMENTS

This work has been partially funded by the European project Eyes of Things under contract number GA 643924.

REFERENCES

Fard, H. M., Khanmohammadi, S., Ghaemi, S., and Samadi, F. (2013). Human age-group estimation based on anfis using the hog and lbp features.

Gunay, A. and Nabiyev, V. (2008). Automatic age classification with lbp. In *Computer and Information Sciences, 2008. ISCIS '08. 23rd International Symposium on*, pages 1–4.

Hewahi, N., Olwan, C., Tubeel, N., EL-Asar, S., and Abu-Sultan, Z. (2010). Age estimation based on neural networks using face features. *Journal of Emerging Trends in Computing and Information Science*, 1(2):61 – 67.

Karthigayani, P. and Sridhar, S. (2011). A novel approach for face recognition and age estimation using local binary pattern, discriminative approach using two layered back propagation network. In *Trendz in Information Sciences and Computing (TISC), 2011 3rd International Conference on*, pages 11–16.

Kohail, S. (2012). Using artificial neural network for human age estimation based on facial images. In *Innovations in Information Technology (IIT), 2012 International Conference on*, pages 215–219.

Lanitis, A., Draganova, C., and Christodoulou, C. (2004). Comparing different classifiers for automatic age estimation. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 34(1):621–628.

Microsoft (2012). Kinect. http://www.microsoft.com/en-us/kinectforwindows/.

Phillips, P. J., Moon, H., Rizvi, S. A., and Rauss, P. J. (2000). The feret evaluation methodology for face-recognition algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(10):1090–1104.

Viola, P. and Jones, M. (2001). Robust real-time object detection. In *International Journal of Computer Vision*.

Wang, H., Li, S., and Wang, Y. (2004). Face recognition under varying lighting conditions using self quotient image. In *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on*, pages 819–824.

Ylioinas, J., Hadid, A., and Pietikainen, M. (2012). Age classification in unconstrained conditions using lbp variants. In *Pattern Recognition (ICPR), 2012 21st International Conference on*, pages 1257–1260.

Zhao, G. and Pietikäinen, M. (2007). Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(6):915–928.