Age Estimation in Make-up (Cosmetics) Using Color and Depth Images

Seyed Muhammad Hossein Mousavi [1], Vyacheslav Lyashenko [2]

[1] Department of Computer Engineering, Bu Ali Sina University, Hamadan-Iran
mosavi.a.i.buali@gmail.com

[2] Department of Informatics (INF), Kharkiv National University of Radio Electronics, Kharkiv-Ukraine
lyashenko.vyacheslav@gmail.com

ABSTRACT

Age estimation in makeup (cosmetics) has application in surveillance and security. Traditional estimation systems used to employ just color images to do this. Also, traditional methods use an algorithm to remove makeup first and then perform estimation process, but this paper uses a novel way by a new technique. It is possible to use depth images along with color images to estimate even stronger (even in absolute darkness). Kinect Version 2 sensor is employed in most of the experiments in this paper. Using depth sensor, it is possible to make 3-dimentional model of the face and extract wrinkles of the face which is hidden under makeup or cosmetics. This feature helps in age estimation process by estimating the number of wrinkles in the face. It is tried to use depth image along with color image to increase experiment accuracy in faces under makeup. Proposed method does not need training process and just one image is enough for estimation. It is 100 times faster in some databases but lower accuracy just as double in max. There is no research done before using depth data for age estimation in makeup, which makes this research unique. Also, a new face detection and extraction method out of depth images is proposed in the paper. A simple color-depth makeup-based dataset is presented which is recorded using Kinect V.2. Proposed method is tested with some benchmark makeup datasets and returned satisfactory and promising results.

Keywords
Age estimation, Depth images, Kinect Version 2, Makeup (cosmetics)

1. INTRODUCTION

With extracting wrinkles and structure of the face using depth sensors as a 3-dimentional model, identity and age is clear even under the most intense makeups (cosmetics), except stretching the skin. Also, depth sensors could record and recognize even in absolute darkness. These features have high of importance in surveillance and security. Paper consists of 5 sections as follow. Section I explains fundamentals of the paper require for next sections. Section II pays to some of the prior researches on age estimation under makeup. Section III demonstrate proposed method in details. Section IV is explained all experiments on different color and depth datasets. And section V consists of conclusion, discussion and future works.

1.1 FACE DETECTION AND RECOGNITION

The process of distinguishing a specific object in image out of many other objects is called object detection. Now if this object is a face, then the process is called face detection. In other hand, the process of distinguishing one object out of many other similar objects is called object recognition. Similar to detection method, if the object is a face and wants to distinguish it from other faces, then it is called face recognition [1]. Now the process of estimating the age of that recognized face is called age estimation. Figure 1 represents difference between face detection, recognition and age estimation using samples from MIFS dataset [2].

1.2 AGE ESTIMATION

Aging effect is determined by gender and race, and it is distinct in different weather, and also it is depended on other factors like job, family environment, diet, sicknesses and hereditary effects. So, we can’t say, “Age detection” and should use “age estimation”, because it is not a precise subject. The only way that we can use age detection, is when using the face patterns during a person's life time, and that is just belonged to that specific person. However, using image processing techniques, there is high possibility to estimate different race people ages. With extracting proper patterns from digital image and using appropriate pre and post processing methods, age estimation is possible with good accuracy. Figure 2 shows the aging effect on two people from the FG-Net aging database [3] in different ages.

1.3 DEPTH IMAGES AND SENSORS

An image which shows the distance between the subject and the sensor is called depth image. These types of images mostly used for extracting 3d model of environment as they could projectile infrared light spectrum toward subject. There are different sensors to this purpose such as Kinect [4], Intel realsense and LEAP motion controller [5]. They use different technology to achieve this goal. Kinect is one of the cheapest and the best sensors for researchers. Kinect V.2 could projectile infrared light in the form of point cloud in maximum range of 5 meters using Times of Flight (TOF) technology. As depth sensors could make 3-d model of the face, they could extract wrinkles of the face which are one of the main signs of aging effect. This data increases significant amount of details to color image. Figure 3 presents different depth sensors and their specifications.

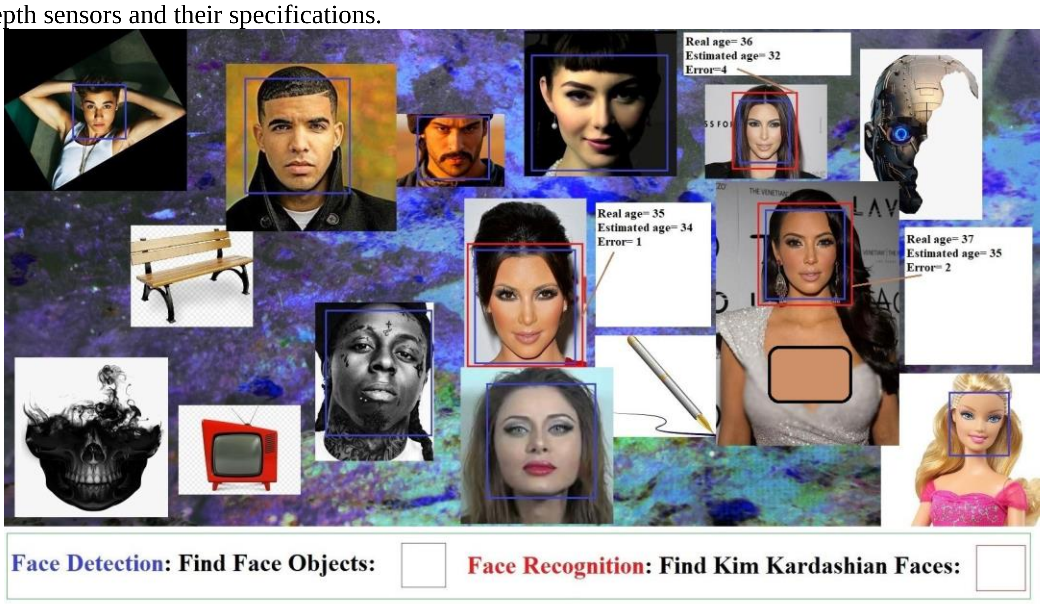

Figure 1. Face detection, face recognition and age estimation using samples from MIFS dataset [2]

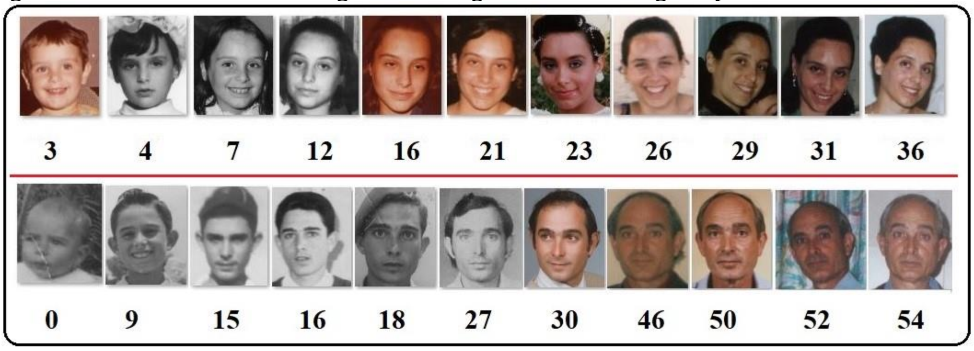

Figure 2. Aging effect on two subjects from the FG-Net aging database [3] in different ages

	Microsoft Kinect v1	Microsoft Kinect v2	Intel Realsense (F200)	LEAP Motion Controller
Technology	Structured Light	Time-of-Flight	Time-of-Flight	Stereo Cameras
RGB camera	640x480 @30fps	1920x1080 @30fps	1920x1080 @30fps	undisclosed
Depth sensors	640x480 @30fps	512x424 @30fps	640x480 @60fps	undisclosed
Microphone	Quad-array microphone	Quad-array microphone	Dual array microphone	-
Range	0.4 to 4.5 m	0.5 to 4.5 m	0.2 to 1.2 m	0.03 to 0.6 m
Skeletons tracked	2	6	-	-
Horizontal Field of View (FOV)	57°	70°	70°	150°
Vertical FOV	43°	60°	43°	120°
USB	2.0	3.0	3.0	3.0
Gestures tracking	Yes	Yes	Yes	Yes
Body joints	20	25	-	-
SDK	Yes	Yes	Yes	Yes
Portability	No	No	Yes	Yes

Figure 3. Specification of different most common depth sensors [5]

2. PRIOR RELATED RESEARCHES

There is no research done on using depth data for age estimation in makeup or non-makeup before and this is the first research recognizable. But for having the history of age estimation in makeup using color images and non-makeup using depth images a brief section is considered. Also, for better readability and saving number of pages, this history is represented in Table 1. Also researches on face detection using color and depth images is placed at the bottom of the table as proposed paper offers a new face detection and extraction method out of depth images. Figure 4 shows wrinkles of the face in color and 3-d model [38]. Also, some of the nice researches in age estimation in makeup using color images are [30], [31], [32], [33], [34], [35], [36] and [37].

Figure 4. Wrinkles of face in color and 3-d model of face [38]

Table 1- The history of age estimation in makeup (color) and non-makeup (depth)

Application	Features and descriptions	Classifier	Database	Accuracy/MAE	Ref	Year
Age Estimation	PCA	SVM	FG-NET	Mean Absolute Error (MAE): 6.77	[6]	2007
Age Estimation	Patch-based Gaussian Mixture Models (GMM) and Hidden Markov Models (HMM) Supervector	Nearest Centroid Classifier (NCC)	Novel	MAE: Male= 5.397 Female= 6.333	[7]	2008
Age Estimation	Preprocessing = Active Appearance Model (AAM) Feature = Least Angle Regression (LAR)	Support Vector Regression (SVR)	FG-NET	MAE: 5.70	[8]	2009
Age Estimation	Ordinal discriminative aging features and Preserving Locality and Ordinal Information (PLO) method		FG-NET	MAE: 4.82	[9]	2012
Age Estimation	Facial Component Localization using Active Shape Model (ASM) and Biologically Inspired Feature (BIF)	SVM	FG-NET PCSO	MAE: 4.7 MAE: 7.2	[10]	2013
Age Estimation	convolutional neural network (CNN)		MORPH	MAE: 3.63	[11]	2014
Age Estimation	convolutional neural network (CNN)		Novel	84.7% +,- 2.2	[12]	2015
Age Estimation	convolutional neural network (CNN)		MORPH AFAD	MAE: 3.27 ± 0.02 MAE: 3.34 ± 0.08	[13]	2016
Age Estimation	group-aware deep feature learning (GA-DFL)		FG-NET MORPH Chalearn-Challenge	MAE: 3.93 MAE: 3.25 MAE: 4.21	[14]	2017
Age Estimation	Entropy of depth image and summation of edges and entropy pixels		FG-NET Curtin Face FEEDB	Male and Female MAE: 20.55 MAE: 12.69 MAE: 28.99	[15]	2018
Face Detection	RGB images (Weak in Depth images)				[16]	2001
Face Detection	RGB images (Using Ellipse fitting in the paper, page 6)				[17]	2015
Face Detection	Depth images				[18]	2008

3. PROPOSED METHOD

This paper presents a novel age estimation method based on color and depth images on faces in makeup. There is no any research done before on this subject using depth images and this is the first research as author's knowledge. One of the features of the research is that using just one sample from dataset is enough, so the algorithm is so fast in compare with other similar algorithms. The speed of the method is due to its non-learning-based structure. It is correct that accuracy decrease double but speed increases 100 times faster in some databases as validation section indicates.

3.1 PROPOSED FACE DETECTION AND EXTRACTION METHOD FROM DEPTH IMAGE

Depth images normally have range of values which as the value of the pixel increases the range increases too. So, for instance and in Kinect V.2, value 1500 means that subject is in 1.5 meter of the sensor. In this method, closest sample from subject to the sensor which is noise tip will be found. In the second step local range filter [19] applies to the face to clarifying the shape of the head. Ellipse fitting on the face takes place to extracting the elliptical shape of the face. And finally, percentage of four sided of the face removes to have just the main muscles of the face. Final step decreases the final error.

3.2 PROPOSED AGE ESTIMATION METHOD FROM COLOR AND DEPTH IMAGES

First step is acquiring color and depth data from datasets (one sample is enough). For color face detection and extraction, viola and jones [20] algorithm is use and for depth images, proposed method which is described in previews section is employed (second step). Third step is choosing lower side of the image (mouth section) in both color and depth images. Fifth step is consisting of applying proposed edge detection method [29] on color image and local range filter [19] + entropy filter [21] on depth image. Final step is to sum the gray values of both color and depth Images and normalization between youngest and oldest age in dataset. Figure 5 represents proposed method in flowchart form and Figure 6 represents it using a sample from proposed dataset in visual format.
Viola and Jones algorithm [20] is one of the best face detection algorithms, which is using for a long period of time in face image processing. This algorithm is so fast and robust, and it could be used for depth images too, but with lower accuracy. This algorithm is an object detection algorithm and could be used for any learned object but mainly uses for face. The algorithm has four stages:

1. Haar Feature Selection
2. Creating an Integral Image
3. Adaboost Training
4. Cascading Classifiers

Image entropy is a quantity which is used to describe the business of an image, i.e. the amount of information which must be coded for by a compression algorithm. An image that is perfectly flat will have an entropy of zero. Consequently, they can be compressed to a relatively small size. On the other hand, high entropy images such as an image of heavily cratered areas on the moon have a great deal of contrast from one pixel to the next and consequently, cannot be compressed as much as low entropy images [21]. Image entropy formula is as (1):

$$Entropy = -\sum P_i Log_2 P^i \qquad (1)$$

In the above expression, Pi is the probability that the difference between two adjacent pixels is equal to i, and Log2 is the base 2 logarithm.

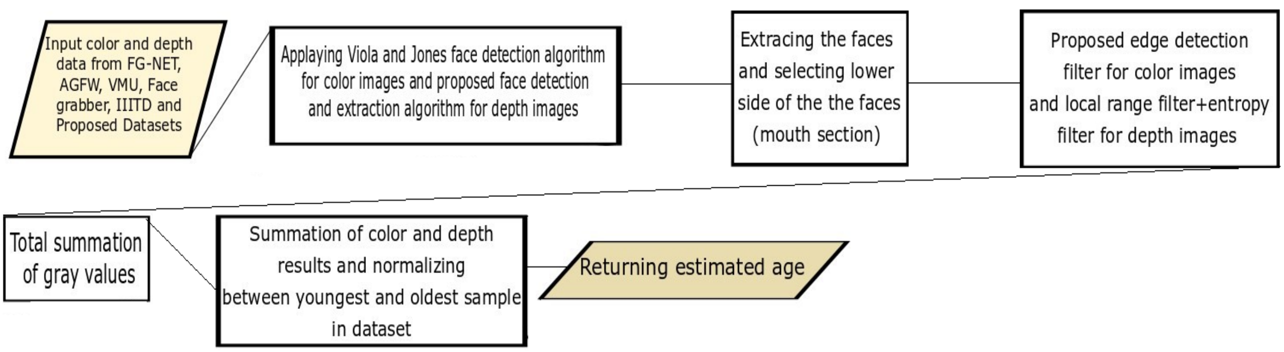

Figure 5. Proposed method flowchart

4. VALIDATION AND RESULTS

In this section and for validation 6 datasets (3 color based and 3 color + depth based) is used. Some databases are clearly used cosmetics in their faces and some used mild amount of these materials. For those mild makeups, real age of subjects is asked from dataset provider or it was in the metadata of the mentioned dataset. Also due to the lack of face dataset in makeup (depth based), a simple dataset consists of 10 subjects is made. Kinect V.2 sensor is employed to make this dataset. The dataset is called Iranian Depth Makeup Dataset (IDMD). Figure 7 presents dataset acquisition environment and system which is used to it.

4.1 DATASETS

It is considerable that from each datasets, just those images with makeup is selected such as FG-NET [5], Face Grabber [24] and IIITD_Kinect_RGBD [25-26] as not all of subjects are in makeup, except those datasets which specifically are designed for makeup researches such as AGFW [22], MIFS [23] and Proposed IDMD. Also, all researches are just on female subjects so this research is female makeup-based paper. Table 2 represents the specifications of employed datasets.

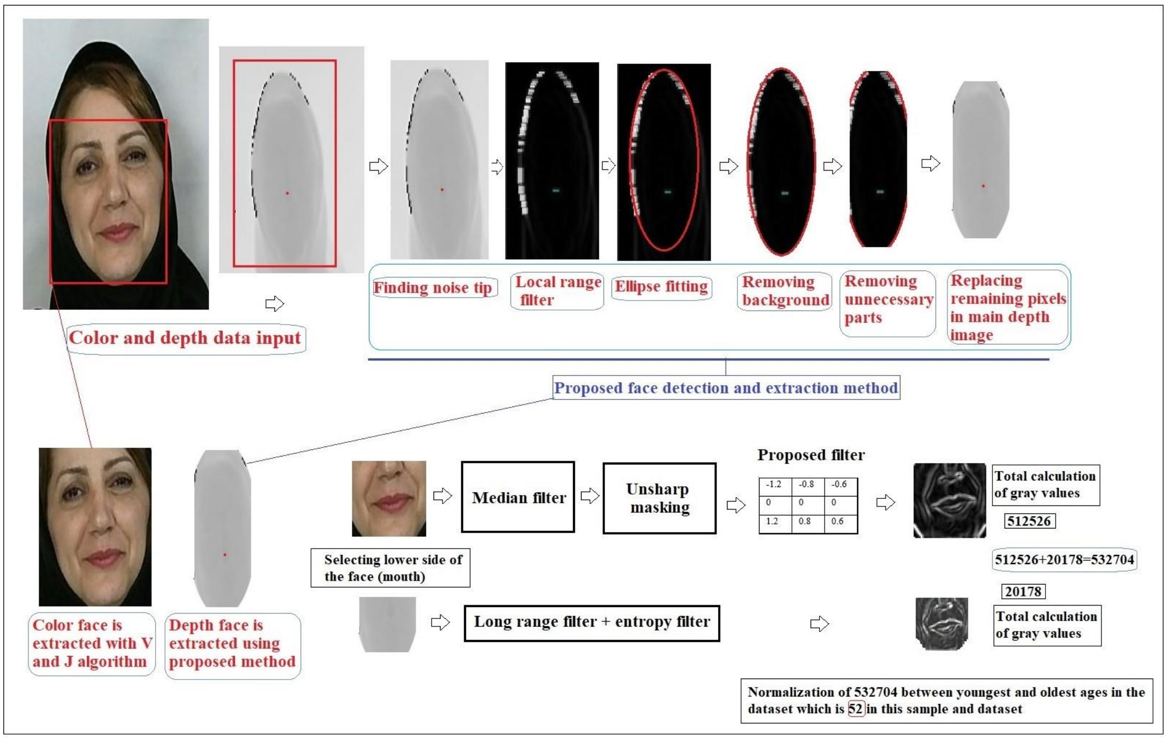

Figure 6. Proposed method in visual form

4.2 MAE AND RESULTS

Mean Absolute Error (MAE) [27] is a measure of difference between two continuous variables. Assume X and Y are variables of paired observations that express the same phenomenon. Examples of Y versus X include comparisons of predicted versus observed, subsequent time versus initial time, and one technique of measurement versus an alternative technique of measurement. Consider a scatter plot of n points, where point i has coordinates (xi, yi)... Mean Absolute Error (MAE) is the average vertical distance between each point and the Y=X line, which is also known as the One-to-One line. MAE is also the average horizontal distance between each point and the Y=X line. The mean absolute error is a common measure of forecast error in the time-series analysis, [28] where the terms "mean absolute deviation" is sometimes used in confusion with the more standard definition of mean absolute deviation. Table 3 shows acquired error (MAE) results from each dataset in four different ranks using proposed method. Also Figure 10 represents graphical form of MAE for each dataset.

$$MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} = \frac{\sum_{i=1}^{n} |e_i|}{n} \quad (2)$$

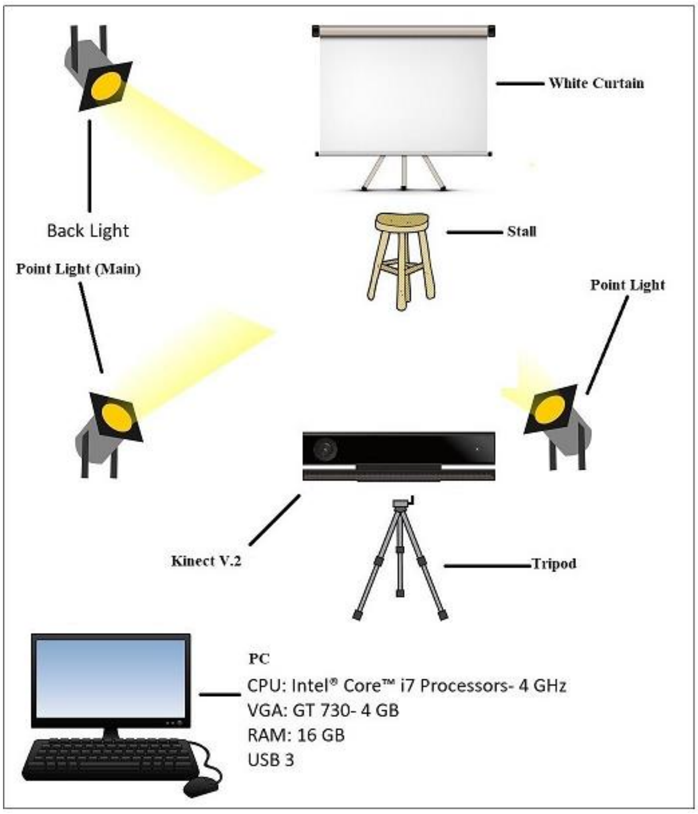

Figure 7. Dataset acquisition environment and system

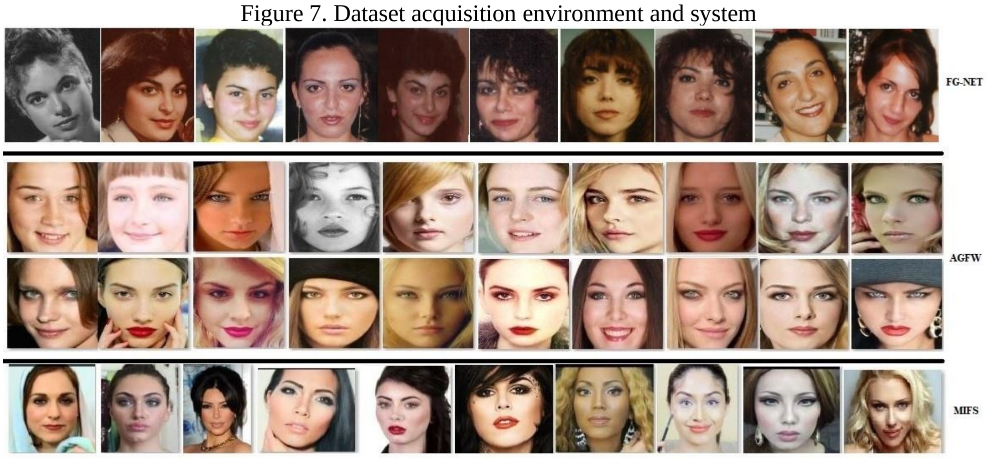

Figure 8. Samples from makeup color based datasets

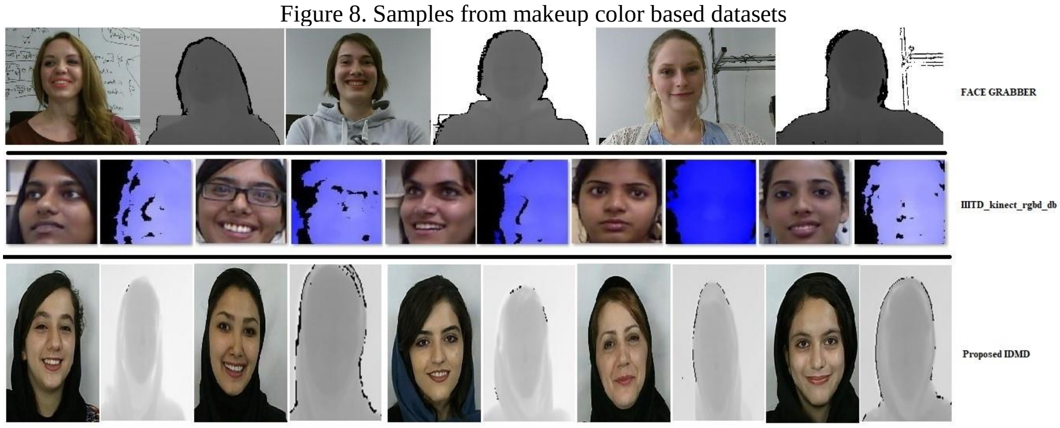

Figure 9. Samples from makeup color-depth based datasets

Table 2- Specifications of employed datasets

Database	**FG-NET**	**AGFW**	**MIFS**	**FACE GRABBER**	**IIITD_kinect_rgbd_db**	**Proposed IDMD**
Authors	FG-NET Group	Minear, Meredith, and Denise C. Park	Chen, Cunjian, et al	Merget, Daniel, et al	Goswami, Gaurav, et al	S.M.H. Mousavi, Vyacheslav Lyashenko
Samples	82	36299 images from web	107	40	106	10
Sensor	Different types	Different types	From YouTube	Kinect V.2	Kinect V.1	Kinect V.2
Function	Facial expression recognition, Face recognition, Gender recognition, Age estimation	Face recognition and age estimation with makeup	Face recognition and age estimation with makeup	Facial expression recognition, Face recognition, Gender recognition, Age estimation	Facial expression recognition, Face recognition, Gender recognition, Age estimation	Age estimation in makeup
Data type	1002 color images	36299 images from web	642 color images	67159 color and depth images	46050 color and depth images	20 color and depth images
Resolution	Different resolutions and quality	128*128	Different resolutions and quality	Color 1920*1080 and depth 512*424	Different resolutions	Color 1920*1080 and depth 512*424
Year	2004	2004	2017	2016	2013	2019
Size	43 Mega bytes	560 Mega bytes	43 Mega bytes	8.5 Giga byte	470 Mega bytes	2 Mega bytes
Reference	[5]	[22]	[23]	[24]	[25-26]	-

Table 3- Acquired MAE from each dataset in four different ranks

Database	**FG-NET**	**AGFW**	**MIFS**	**FACE GRABBER**	**IIITD_kinect_rgbd**	**Proposed IDMD**
Rank 1 (25 %)	8.9	6.9	6.3	8.1	5.3	4.7
Rank 2 (50 %)	10.2	8.0	7.7	8.5	5.9	5.2
Rank 3 (75 %)	12.7	8.8	8.1	8.9	6.8	5.8
Rank 4 (100 %)	15.5	9.2	9.5	9.6	7.0	6.1

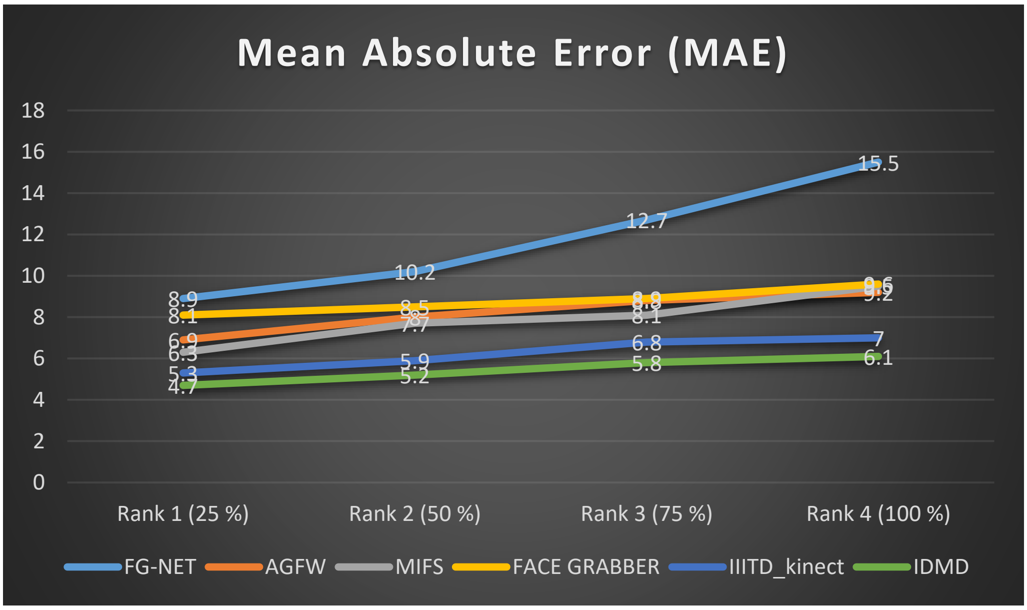

Figure 10. MAE for all datasets in different ranks

4.3 EVALUATION RESULTS AND DISCUSSION

As it is clear in Figure 10, highest error belongs to FG-NET dataset with MAE of 15.5 in rank four and the lowest to proposed IDMD with MAE of 4.7 in rank 1. Each rank is number of used samples in each dataset. For instance, 50 % or rank 2 means just half of the admissible data from that specific dataset is used. Also, it clear that as rank gets higher, accuracy gets lower. It seems color-depth based datasets have better results in total (except face grabber dataset). As it mentioned before, due to non-learning-based structure of the proposed method, it is so fast method and doesn't need strong systems to work with. This high-speed lead to low accuracy, but still worth it.

5. CONCLUSION AND FUTURE WORKS

This paper presents the performance of a non-learning base makeup detection system for color-depth images. System has significant amount of speed, but lower accuracy. The speed in almost all experiments was less than 0.5 second which is considerable. Due to high speed of this method and need to less amount of calculation. It is possible to make even smartphones version of it. It is suggested to use this method on other color-depth makeup datasets (if exists). Depth based datasets used in this research did not made for age estimation for makeup (specifically), but it was possible to use them in that way. So, making a strong makeup depth-based age estimation dataset is a brilliant idea as it is needed in a lot of scientific fields and lack of it. Also adding more samples to IDMD is of future works.

REFERENCES

[1] Jain, Anil K., and Stan Z. Li. Handbook of face recognition. New York: springer, 2011.

[2] Chen, Cunjian, et al. "Spoofing faces using makeup: An investigative study." 2017 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA). IEEE, 2017.

[3] The FG-NET Aging Database, http://sting.cycollege.ac.cy/~alanitis/ fgnetaging/index.htm, 2002.

[4] https://www.slideshare.net/SugiuraTsukasa/kinect-v2-introduction-and-tutorial

[5] Leong, Chee Wee, et al. "Utilizing depth sensors for analyzing multimodal presentations: Hardware, software and toolkits." Proceedings of the 2015 ACM on International Conference on Multimodal Interaction. ACM, 2015.

[6] Geng, Xin, Zhi-Hua Zhou, and Kate Smith-Miles. "Automatic age estimation based on facial aging patterns." IEEE Transactions on pattern analysis and machine intelligence 29.12 (2007): 2234-2240.

[7] Zhuang, Xiaodan, et al. "Face age estimation using patch-based hidden markov model supervectors." Pattern Recognition, 2008. ICPR 2008. 19th International Conference on. IEEE, 2008.

[8] Ricanek, Karl, et al. "Generalized multi-ethnic face age-estimation." Biometrics: Theory, Applications, and Systems, 2009. BTAS'09. IEEE 3rd International Conference on. IEEE, 2009.

[9] Li, Changsheng, et al. "Learning ordinal discriminative features for age estimation." Computer vision and pattern recognition (cvpr), 2012 ieee conference on. IEEE, 2012.

[10] Han, Hu, Charles Otto, and Anil K. Jain. "Age estimation from face images: Human vs. machine performance." Biometrics (ICB), 2013 International Conference on. IEEE, 2013.

[11] Yi, Dong, Zhen Lei, and Stan Z. Li. "Age estimation by multi-scale convolutional network." Asian Conference on Computer Vision. Springer, Cham, 2014.

[12] Levi, Gil, and Tal Hassner. "Age and gender classification using convolutional neural networks." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2015.

[13] Niu, Zhenxing, et al. "Ordinal regression with multiple output cnn for age estimation." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016.

[14] Liu, Hao, et al. "Group-aware deep feature learning for facial age estimation." Pattern Recognition 66 (2017): 82-94.

[15] Mousavi, Seyed Muhammad Hossein. "A New Way to Age Estimation for RGB-D Images, based on a New Face Detection and Extraction Method for Depth Images." International Journal of Image, Graphics and Signal Processing 10.11 (2018): 10.

[16] Viola, Paul, and Michael Jones. "Rapid object detection using a boosted cascade of simple features." Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on. Vol. 1. IEEE, 2001.

[17] Carcagnì, Pierluigi, et al. "Facial expression recognition and histograms of oriented gradients: a comprehensive study." SpringerPlus 4.1 (2015): 645.

[18] ter Haar, Frank B., and Remco C. Veltkamp. "3D face model fitting for recognition." European Conference on Computer Vision. Springer, Berlin, Heidelberg, 2008.

[19] Bailey, Donald G., and Robert M. Hodgson. "Range filters: Localintensity subrange filters and their properties." Image and Vision Computing 3.3 (1985): 99-110.

[20] Viola, Paul, and Michael Jones. "Rapid object detection using a boosted cascade of simple features." Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on. Vol. 1. IEEE, 2001.

[21] Gonzalez, R.C., R.E. Woods, S.L. Eddins, Digital Image Processing Using MATLAB, New Jersey, Prentice Hall, 2003, Chapter 11.

[22] Minear, Meredith, and Denise C. Park. "A lifespan database of adult facial stimuli." Behavior Research Methods, Instruments, & Computers 36.4 (2004): 630-633.

[23] Chen, Cunjian, et al. "Spoofing faces using makeup: An investigative study." 2017 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA). IEEE, 2017.

[24] Merget, Daniel, et al. "Capturing facial videos with Kinect 2.0: A multithreaded open source tool and database." 2016 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2016.

[25] G. Goswami, M. Vatsa, and R. Singh, RGB-D Face Recognition with Texture and Attribute Features, IEEE Transactions on Information Forensics and Security, 2014.

[26] G. Goswami, S. Bharadwaj, M. Vatsa, and R. Singh, On RGB-D Face Recognition using Kinect, International Conference on Biometrics: Theory, Applications and Systems, 2013.

[27] Willmott, Cort J.; Matsuura, Kenji (December 19, 2005). "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance". Climate Research. 30: 79–82.

[28] Hyndman, R. and Koehler A. (2005). "Another look at measures of forecast accuracy"

[29] Mousavi, S. Muhammad Hossein, and Marwa Kharazi. "An Edge Detection System for Polluted Images by Gaussian, Salt and pepper, Poisson and Speckle Noises, 4th National Conference on Information Technology,Computer & Tele-Communication, Iran-Torbat-e Heydarieh, 2017."

[30] Takimoto, Hironori, et al. "Robust gender and age estimation under varying facial pose." Electronics and Communications in Japan 91.7 (2008): 32-40.

[31] Sörqvist, Patrik, and Mårten Eriksson. "Effects of training on age estimation." Applied Cognitive Psychology 21.1 (2007): 131-135.

[32] Huerta, Ivan, et al. "A deep analysis on age estimation." Pattern Recognition Letters 68 (2015): 239-249.

[33] Chen, Cunjian, Antitza Dantcheva, and Arun Ross. "Impact of facial cosmetics on automatic gender and age estimation algorithms." 2014 International Conference on Computer Vision Theory and Applications (VISAPP). Vol. 2. IEEE, 2014.

[34] Vairavan, Tamilb, and Kaliyaperumal Vani. "An Efficient Age Estimation System with Facial Makeover Images Based on Key Points Selection." International Arab Journal of Information Technology (IAJIT) 14.1 (2017).

[35] Liu, Li, Jianming Liu, and Jun Cheng. "Age-group classification of facial images." 2012 11th International Conference on Machine Learning and Applications. Vol. 1. IEEE, 2012.

[36] Gao, Peter Xiang. "Facial age estimation using clustered multi-task support vector regression machine." Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012). IEEE, 2012.

[37] Iga, Ryotatsu, et al. "A gender and age estimation system from face images." SICE 2003 Annual Conference (IEEE Cat. No. 03TH8734). Vol. 1. IEEE, 2003.

[38] Suwajanakorn, Supasorn, Ira Kemelmacher-Shlizerman, and Steven M. Seitz. "Total moving face reconstruction." European Conference on Computer Vision. Springer, Cham, 2014.