http://dx.doi.org/10.5392/IJoC.2014.10.1.001

Construction of a Video Dataset for Face Tracking Benchmarking Using a Ground Truth Generation Tool

Luu Ngoc Do
Department of Computer Engineering
Chonnam National University, Gwangju 500-757, South Korea

Hyung Jeong Yang*
Department of Computer Engineering
Chonnam National University, Gwangju 500-757, South Korea

Soo Hyung Kim
Department of Computer Engineering
Chonnam National University, Gwangju 500-757, South Korea

Guee Sang Lee
Department of Computer Engineering
Chonnam National University, Gwangju 500-757, South Korea

In Seop Na
Department of Computer Engineering
Chonnam National University, Gwangju 500-757, South Korea

Sun Hee Kim
Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA USA

ABSTRACT

In the current generation of smart mobile devices, object tracking is one of the most important research topics for computer vision. Because human face tracking can be widely used for many applications, collecting a dataset of face videos is necessary for evaluating the performance of a tracker and for comparing different approaches. Unfortunately, the well-known benchmark datasets of face videos are not sufficiently diverse. As a result, it is difficult to compare the accuracy between different tracking algorithms in various conditions, namely illumination, background complexity, and subject movement. In this paper, we propose a new dataset that includes 91 face video clips that were recorded in different conditions. We also provide a semi-automatic ground-truth generation tool that can easily be used to evaluate the performance of face tracking systems. This tool helps to maintain the consistency of the definitions for the ground-truth in each frame. The resulting video data set is used to evaluate well-known approaches and test their efficiency.

***Key words**: Face Tracking, Ground-truth, Face Video Dataset.*

1. INTRODUCTION

The face is one of the most important components of the human body, for visually discriminating one human subject from another. Therefore, face detection, face recognition and face tracking are essential in many human-machine interaction systems. Face tracking is a difficult task in computer vision and in other fields, because the appearance of a face is affected by a number of factors, including identity, face pose, illumination, facial expression, age, occlusion, and facial hair [1].

* *Corresponding author, Email: hjyang@jnu.ac.kr*
Manuscript received Aug. 03, 2013; revised Dec. 28, 2013; accepted Jan. 09, 2014

Thus far, many face tracking techniques have been proposed, and shown to be successful in several video clips [2]-[4]. However, previously proposed tracking algorithms usually use their own datasets to show their performance. Therefore, it is very difficult to find out the most powerful approach when their performances were evaluated by different video data, collected from different environments and conditions. This leads to the necessity of developing such a database that can cover various conditions.

The background color, the movements of camera and subjects always affect the performance of tracking algorithms directly. The illumination absolutely has many effects on the conventional image processing techniques. Even though nowadays the illumination is easy to handle by using some powerful preprocessing algorithms, it cannot be missed in any kind of image or video dataset. In the face tracking process, especially in the tracking process for face recognition system, the performance of tracking algorithms for one subject and multiple subjects are quite different in both accuracy and processing time. Therefore, we proposed a dataset that can cover the conditions of illumination changes, background color, camera movements, subject movements and number of faces.

Currently no diverse dataset has been established to compute the accuracy precisely, and compare the performance of different approaches, in regards to face tracking. Many tracking algorithms only exhibit good performance in simple conditions, and cannot deal with specific circumstances [5]-[8]. However, most of the well-known datasets only include videos with simple motion of the head, or focus on only one condition, corresponding to some features for specific algorithms [9]-[11]. This limitation makes it hard to evaluate how well a tracking algorithm can perform. A face tracking application must present an invariant ability against some variant features, such as illumination, subject movement, number of faces and background complexity. For the evaluation of the face tracking algorithms, we need to define the ground-truth of a face. In the previous datasets, only a few of them included ground-truth information, which defines the face location. Since ground-truth generation requires many manual expressions, and every video contains thousands of images, it is very time-consuming and error-prone. Furthermore, all manual specifications are also user-dependent: two individuals analyzing the same scene may (and probably will) produce different ground-truth data [12]. Therefore, a ground-truth generation tool is necessary, for the consistency of evaluation of face tracking algorithms. Most of the ground-truths for the facial video database are generated manually, without any frameworks. We believe that a semi-automatic ground-truth generation tool incorporating a face detector can help resolve the problem of being time-consuming and error-prone.

In this paper, we propose a new face video database that contains more than 90 videos. Our database is clearly classified into different conditions of illumination, background complexity, human subject movement and camera movement. By using this data, we can easily identify the advantages and disadvantages of a face tracking algorithm. We also construct a tool to generate ground-truth information for every video frame, to easily evaluate the performance of face tracking systems. The tool can be integrated with a face detection algorithm, to generate answer sets for the evaluation, according to the definition of faces in a video frame.

The remainder of this paper is organized as follows. Section 2 briefly describes the main approaches of face tracking, the requirements for a facial video database, and well-known datasets. Section 3 presents how the proposed database is classified according to conditions. Section 4 describes the data collection, and the characteristics of face videos. Lastly, we present the generation of ground-truth in section 5, and the conclusion in section 6.

2. RELATED WORK

Currently, there are two main approaches for face tracking techniques. The first one is tracking the facial features, such as lips, eyebrows and eyelids, to construct a model of the head pose [5], [6], [18]. However, this type of tracking cannot achieve a good performance, since some facial features are occluded, when only one side of the face is shown. Furthermore, various types of head motions will affect the accuracy of these tracking algorithms. The second approach is to define the location of the facial area, based on the skin color of faces [7], [8], [19]. This can deal with fast movements, occlusion and scale variation. However, the tracking result will fail when the background color is similar to the color of facial skin. For example, Fig. 1 illustrates the failure cases of the CAMSHIFT algorithm, which uses skin color as the features of faces [7]. Nowadays, several robust particle filter-based tracking algorithms were proposed to deal with various kinds of conditions and solve the problems of conventional approaches well [21], [22]. Therefore, many conditions, such as subject movement, camera movement, background and illumination should be considered, due to the fact that the performances of many face tracking algorithms are affected by these conditions.

Fig. 1. Two failure cases of CAMSHIFT

Many databases have been produced for object tracking. The PETS'2001 dataset was constructed to evaluate the tracking of moving people and vehicles [10]. The M2VTS

includes hundreds of face videos from 37 human subjects, without challenges of illumination changes and background color [11]. The UO face video database contains videos collected from different illumination conditions and different cameras [9]. However, it does not deal with the challenge of camera movements, the various movement styles of human subjects, or the background color. The data set proposed by FIPA was collected from various sources, such as TV series, webcam recordings, Youtube and surveillance camera networks [13]. However, this data was not clearly classified into conditions for evaluating the performance of face tracking algorithms in some specific cases. Therefore, we proposed a dataset that can cover the conditions of illumination changes, background color, camera movements and subject movements. The ground-truth data are also provided, to evaluate the performance of tracking algorithms, and find out their limitations.

In recent years, there have been several efforts to evaluate object tracking. Li et al. attempted to solve the problem of evaluating a football players tracking system [14]. They introduced three measurements of "spatio-temporal evaluation of Identity tracking", "spatial evaluation of Category tracking" and "temporal evaluation of Category tracking" to evaluate the spatial and temporal accuracy of the tracker's output. However, some basic types of error in multiple object tracking, such as the false positive tracking and overlap, were not considered. Nghiem et al. presented 8 metrics, to evaluate object detection, localization and tracking performance [15]. There are many dependencies between separate metrics, so that only a combination of several metrics can be used to evaluate the performance of an algorithm. For example, they used the "tracking time" metric, "object ID persistence" metric and "object ID confusion" metric for evaluating object tracking algorithms, since any one of these three metrics cannot be meaningful by itself. Smith et al. attempted to provide several metrics to measure multiple objects tracking performance: 5 for measuring object configuration errors, and 4 for measuring inconsistencies in object labeling over time [16]. Bernardin and Stiefelhagen made an enhancement of Smith's work, by introducing only two metrics, to allow for the objective comparison of tracker characteristics [17]. However, their metrics are complicated, and only accordant with multiple object tracking. Our dataset with a large number of single face videos, is consistent with only one metric, and is similar to the "Coverage Test" metric provided by Smith et al. in [16].

3. CLASSES OF CONDITIONS

There are many factors that affect the performance of a face tracking system. These factors need to be considered as the conditions of face videos, to investigate the advantages and disadvantages of a system. Therefore, we propose a dataset based on the number of faces in a frame, subject movement, camera movement, background complexity and illumination.

3.1 Number of faces

We first divide the dataset into single face videos and multiple face videos. In single face videos, only one human subject is captured. The multiple face videos can contain more than one human subject in every video frame. Most of the techniques for face tracking focus on the problem of tracking a single detected face. Multiple face tracking is much more difficult, because it needs to track many targets simultaneously, and ensure that the processing time is acceptable for a real time mobile application. We limit the number of faces up to 4, to avoid misdetection and confusion during the tracking process. If there are too many faces, each face size will be too small, and this makes detection difficult. It will also not have enough space for each subject movement, making it hard to evaluate tracking performance.

3.2 Movements

The movements of subjects and camera are the most important factors in any object tracking system. Also, a different style and speed of movements can affect the performance of the same tracking algorithm. In our database, the subject movements include scaling, rotation-out-plane and simple movement. Scaling is the movement of a face from near to far from the camera, while keeping it parallel to the camera, so that the face size is changed. Rotation-out-plane is yaw and pitch movement, as described in Fig. 2. The rotation of the head following axis x is pitch, and y is yaw. In the simple movement case, human subjects move their head in a plane parallel to the camera, and keep the distance between that plane and the camera unchanged. This kind of movement was described as a small degree of rotation or translation less than 30 degree of angle. The camera movement includes translation and rolling (rotation-in-plane). Translation is the movement of the camera from left to right, and up and down, while keeping it parallel to the faces. Rolling is the rotation of the camera in the range of 90 degrees.

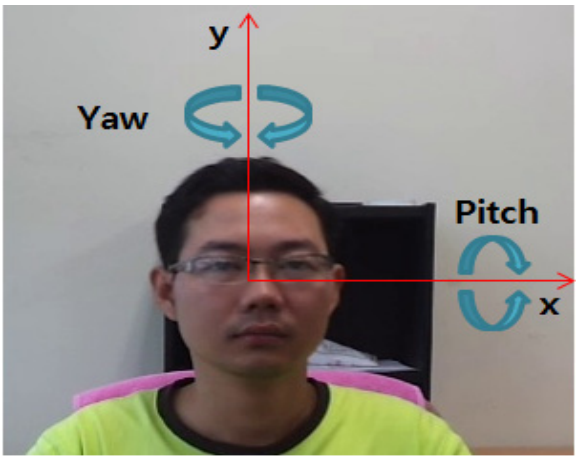

Fig. 2. Rotation-out-plane: Yaw and Pitch

3.3 Background

We divide the background condition into simple background and complex background. Due to the fact that many face tracking algorithms use the information of skin color, we define a complex background as a background having a similar color to the skin color of the human face. This complex background is totally different from a simple background, which can easily be discriminated from the human face.

3.4 Illumination

Illumination changes not only affect the performance of face tracking but also affect face detection and face recognition. In our data, we consider backlit, variant illumination and simple illumination. Backlit is the case where the background illumination is too bright, and the subject illumination is too dark. In the variant illumination videos, the illumination changes from dark to bright, and vice versa. The simple illumination is demonstrated by an indoor environment that is not too bright or dark.

3.5 Structure of conditions

For the single face videos that have complex background, we consider only the normal illumination condition, because it is very difficult to track a face with special luminance on a complex background. The main videos of our dataset are the videos that have a single face, simple background and normal illumination. The video clips with these conditions are the easiest case for any tracking algorithm, so that the number of these video clips need to be much more than any other cases, to validate consistent performance. There is no priority for 5 types of movements (3 types of head movement and 2 types of camera movement). Fig. 3 and Fig. 4 describe the classification of conditions for single face video and multiple faces video, respectively.

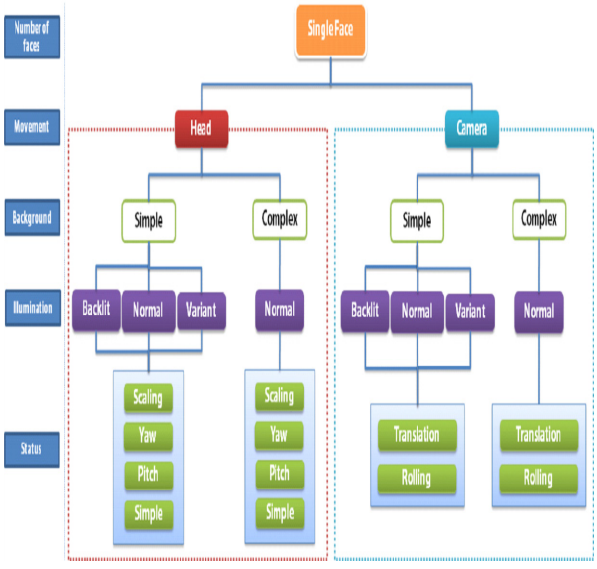

Fig. 3. Classes of conditions for single face video clips

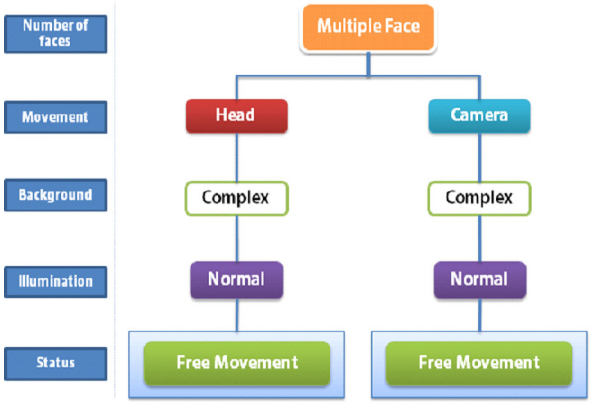

Fig. 4. Classes of conditions for multiple face video clips

4. VIDEO DATA COLLECTION

In this section, we explain some features of the collected video data, and show some examples of conditions. The video data were collected in mp4 format with a resolution of 640x480, by using the 8M camera of a Samsung Galaxy S II device. The length of each video is 10 to 15 seconds, with 30 frames per second. The whole video dataset was captured from 26 volunteer students in Chonnam National University. Most of the volunteers are Asian students with similar skin color. There are 4 conditions that need to be considered in this database, including the number of faces, subject and camera movement, background, and illumination. However, it is not necessary to include all of these conditions in a video clip. We only consider the 14 combinations of conditions listed in Table 1.

The video clips we collected according to the conditions are described in Table 2. Since the videos with conditions of single face, simple background and normal illumination are the main part of our dataset, we collected 60 videos to cover this combination of conditions, and 12 videos for each type of movement. This dataset can be accessed at http://sclab.cafe24.com/publications.jsp.

Table 1. Video Types

	Conditions	***Names***
1	Single face, simple background, normal illumination, simple movement	Simple
2	Single face, simple background, normal illumination, rotation-out-plane	Simple ROT
3	Single face, simple background, normal illumination, scaling	Simple Scale
4	Single face, simple background, normal illumination, translation	Simple Translation
5	Single face, simple background, normal illumination, rotation-in-plane (rolling)	Simple RIT
6	Single face, simple background, backlit, simple movement	Backlit

7	Single face, simple background, variant illumination, simple movement	Variant
8	Single face, complex background, normal illumination, simple movement	Complex
9	Single face, complex background, normal illumination, rotation-out-plane	Complex ROT
10	Single face, complex background, normal illumination, scaling	Complex Scale
11	Single face, complex background, normal illumination, translation	Complex Translation
12	Single face, complex background, normal illumination, rotation-in-plane (rolling)	Complex RIT
13	Mutiple face, simple background, normal illumination, head movement	Multiple Head
14	Mutiple face, simple background, normal illumination, camera movement	Mutiple Camera

Table 2. Single Face Video

	Simple Background			***Complex Background***
Illumination	Normal	Backlit	Variant	Normal
Scaling	12	N/A	N/A	3
Yaw-Pitch	12	N/A	N/A	3
Simple Movement	12	3	3	3
Translation	12	N/A	N/A	3
Rolling	12	N/A	N/A	3
Total	***60***	***3***	***3***	***15***

Fig. 5 presents a video example of a Simple video. Fig. 6 illustrates the Simple ROT video. Fig. 7 shows an example of Simple Scale. Fig. 8 is Simple Translation, and Fig. 9 is Simple RIT.

Fig. 5. Simple Video

Fig. 6. Simple ROT

Fig. 7. Simple Scale

Fig. 8. Simple Translation

Fig. 9. Simple RIT

For the backlit and variant illumination, we only consider the condition of simple background and simple movement, to avoid abnormal cases. 3 videos for each case are collected. Fig. 10 describes an example of the backlit case. Fig. 11 is an example of the variant illumination case. In these kinds of videos, a human subject moves from a dark region to a bright region, where the illumination is slowly changed; or the light was turned on and off, where the illumination is changed suddenly.

Fig. 10. Backlit

Fig. 11. Variant Illumination

Fig. 12. Complex Background

Table 3. Multiple faces video

	Subject Movement	***Camera Movement***
Background Illumination	Simple Normal	Simple Normal
Number of videos	***6***	***4***

The complex background only appears with single face and normal illumination, but all of the movement styles are included. Fig. 12 shows an example of complex background video with scaling movement. We collected 15 videos that have a complex background, and 3 videos for each type of movement, as shown in Table 2. We collected 13 videos for multiple faces with the conditions of simple background and normal illumination, as described in Table 3. In these multiple face videos, the human subjects are allowed free movements. Fig. 13 presents an example of Multiple Head video, and Fig. 14 is a Multiple Camera video.

Fig. 13. Multiple Head

Fig. 14. Multiple Camera

We defined a rule for name tagging, to easily access the video database. Every video file name has to follow this tagging rule:

[number of faces]_[type of background]_[type of illumination]_[type of movement][video number].mp4

where, [number of faces] can be 'single' for single face video, and 'multiple' for multiple faces video. [type of background] can be 's' for simple background, and 'c' for complex background. [type of illumination] can be 'n' for normal illumination, 'b' for backlit, and 'v' for variant illumination. For example, the first video which has a single face, simple background, normal illumination and simple movement, will have the name: 'single_s_n_simple1.mp4'.

Table 4. Video Name Tagging Rule

Conditions	*Notations*
Single face	single
Multiple faces	multiple
Simple background	s
Complex background	c
Normal illumination	n
Backlit	b
Variant illumination	v

5. FACE TRACKING EVALUATION

We generated ground-truth data for every frame of videos, to compute the accuracy of the face tracking algorithm. The ground-truth is a bounding box containing the face area as accurately as possible, and must be largely independent of hairstyles, hats and glasses. Therefore, in our dataset, we defined ground-truth as a bounding box in which the lower limit should be the chin, the upper limit should be the end of the forehead, and the limits on each side should be the cheeks excluding the ears. Fig. 15 describes an example of ground-truth for a face.

The ground-truth of different images can be of different size and shape, due to the different poses of the human face. However, we believe that our definition of ground-truth is objective, so that it can guarantee that all of the most informative features of the human face are included in the ground-truth, for all cases. It is a very time consuming and error-prone process to generate this kind of ground-truth manually. The manual method also generates an unstable result, and it is hard to evaluate the performance of a face tracking system precisely. Therefore, a semi-automatic tool is necessary, to redress the weak points of the manual method. This kind of tool also helps users to analyze exactly which cases are difficult for tracking.

Fig. 15. Example of ground-truth

We developed a ground-truth generation tool to detect faces frame by frame, and adjust the bounding boxes manually in mis-detected frames, as shown in Fig. 16. Users can handle a mouse, to draw a new bounding box on the image. A face detector can be collaborated to generate the ground-truth, according to the user's definition of ground-truth. For example, for our experiments we employed a Haar features-based face detector [20]. The ground truth tool first detects faces in frames, and generates information of the bounding box of faces. This bounding box can be adjusted, and the modified information is updated. The information of ground-truth in each video is stored automatically in a text file, with the following format:

Frame number: xmin, ymin, width, height

where, xmin and ymin are the coordinates of the top-left point of the ground-truth rectangle.

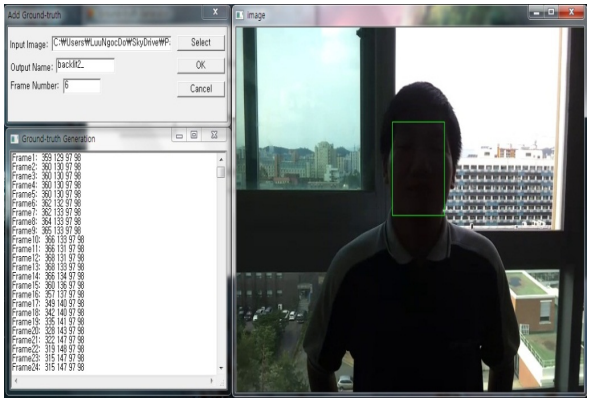

Fig. 16. Adjusting Ground-truth

Fig. 17 describes the whole process of this tool. At first, an input video is split into frames. After that, the ground-truth of the face area is detected automatically, frame-by-frame. The detected ground-truth is drawn on the original frame image, and stored separately in another image file. Ground-truths of the same video are saved in the same information text file, and they are sorted in ascending order from frame number 1. Another text file, called failed-file, is generated to identify misdetection frames. If a frame does not have enough ground-truth in the information text file, its frame number will appear in failed-file. Ground-truths are adjusted or modified, by re-drawing the bounding box. When saving a re-drawn image, this program also deleted the old ground-truth in the information text file (if necessary), and recorded the new information, following the re-drawn bounding box.

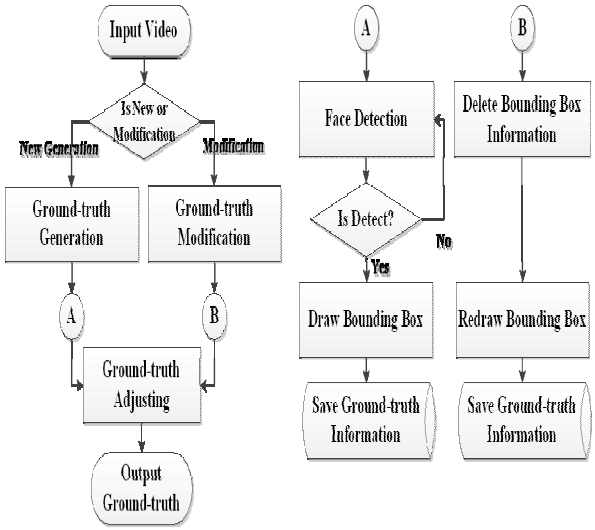

Fig. 17. Ground-truth generation tool's processing stream

We used our dataset for evaluating two of the most common face tracking algorithms: CAMSHIFT, and Color-based Particle Filter. CAMSHIFT is a well-known technique that iteratively shifts the tracked window up to the peak of a back projected probability distribution image, computed from the face region's colors [7]. Color-based Particle Filter generates object hypotheses, and verifies them using a color histogram [8].

To measure the accuracy of the tracking algorithms, we used the value *f*, which is a combination of precision and recall. Given S_R is the area of the tracking result, S_G is the area of ground-truth and *M* is the intersection of S_R and S_G, the precision, recall and *f* value are computed as below

$$\text{Precision} = \frac{M}{S_R} \quad (1)$$

$$\text{Recall} = \frac{M}{S_G} \quad (2)$$

$$f = \frac{1}{\frac{\text{one}}{\text{Precision}} + \frac{\text{one}}{\text{Recall}}} \quad (3)$$

Recall measures how much the ground-truth is covered by the tracker's output. It takes value from 0 (no overlap), to 1 (full overlap). Precision measures how much the tracker's output is covered by the ground-truth. A high value of *f* determines a good tracking result. It is consistent that *f* is only high, when both precision and recall are high.

Table 5 and Table 6 present the average accuracy of these methods for each type of video in our dataset, by computing the value *f* in equation (3).

Table 5. Tracking performance of CAMSHIFT

	Simple Background			***Complex Background***
Illumination	Normal	Backlit	Variant	Normal
Scaling	0.82	N/A	N/A	0.69
Yaw-Pitch	0.73	N/A	N/A	0.62
Simple Movement	0.81	0.58	0.52	0.67
Translation	0.86	N/A	N/A	0.71
Rolling	0.78	N/A	N/A	0.66

Table 6. Performance of Color-based Particle Filter

	Simple Background			***Complex Background***
Illumination	Normal	Backlit	Variant	Normal
Scaling	0.85	N/A	N/A	0.76
Yaw-Pitch	0.79	N/A	N/A	0.65
Simple Movement	0.88	0.62	0.6	0.7
Translation	0.88	N/A	N/A	0.76
Rolling	0.82	N/A	N/A	0.69

Since CAMSHIFT and Color-based Particle Filter used the color information of face skin for tracking, their results are very similar. By using the complex background videos in our dataset, we can find out the drawback of these two methods. Fig. 18 and Fig. 19 demonstrate that they failed when the background color is similar to the color of facial skin. The red rectangle is the tracking result, and the green one is the ground-truth.

Fig. 18. Failure cases of CAMSHIFT

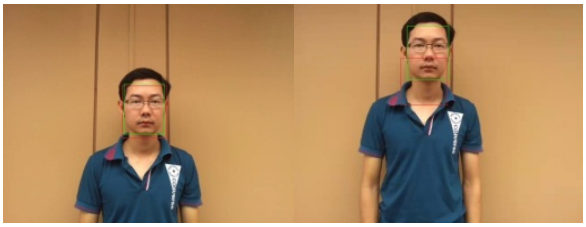

Fig. 19. Failure Cases of Color-based Particle Filter

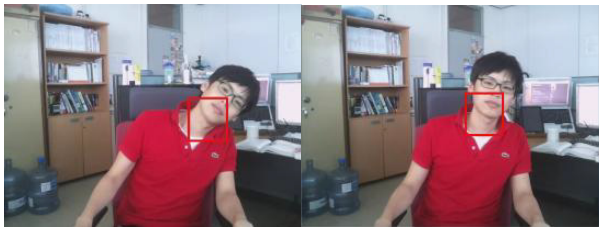

Fig. 20. Tracking performance of CAMSHIFT

The performances of CAMSHIFT and Color-based Particle Filter are not good, even on a simple background. Sometimes, they shifted to the neck region of the human subject, as shown in Fig. 20, because this region has the same color as the face.

Fig. 21. Tracking performance of Particle Filter

As described in Table 5 and Table 6, the accuracy of CAMSHIFT and Particle Filter for Yaw-Pitch movement is always the lowest, compared to other types of movement with the same condition. Therefore, we can consider that tracking the side view of face is also a drawback of these conventional color-based tracking algorithms. Fig. 22 describes the tracking performance of Particle Filter for Yaw-Pitch movement.

None of the checked algorithms can handle illumination conditions. It seems that a pre-processing step is necessary to improve the image quality, before applying tracking algorithms. Fig. 22 demonstrates the performance of the Color-based Particle Filter in a backlit video.

Fig. 22. Tracking performance of Particle Filter

Table 7. Comparison between different databases

	Proposed Dataset	*UO Database [9]*	*FIPA Database [13]*	*M2VTS Database [11]*
Illumination Conditions	Yes	Yes	No	No
Background Conditions	Yes	No	No	No
Ground-truth	Yes	Yes	Yes	No
Multiple Subjects	Yes	No	Yes	No

Table 7 demonstrates the characteristics of several well-known face databases [9], [13], [11], and our proposed dataset. Only UO database and our dataset contain the videos collected from different illumination conditions. The background complexity was not ever considered as a video condition in the other databases, even though it can affect the performance of tracking algorithms, as we showed. Our proposed dataset and the FIPA database can be used for evaluating multiple face tracking algorithms.

6. CONCLUSION

In this paper, we proposed a new dataset for face tracking applications, and showed the accuracy of the dataset, using some well-known face tracking algorithms. The main purpose of this paper is to provide a database for evaluating different face tracking algorithms, in different conditions of background, illumination and movement. Our dataset is diverse enough to find out the advantages and drawbacks of tracking systems. We also developed a ground-truth generation tool that produces answer sets frame-by-frame, to establish an easy way for computing the accuracy of tracking algorithms.

For future work, we will generate additional videos with the condition of movement speed, to check the power of some new algorithms. The 3D pose is also an important factor of face tracking. We will include further video clips considering 3D pose, with other features for the advanced algorithms to evaluate and analyze. In addition, we will propose a new algorithm, to compensate for the drawbacks of the previously proposed tracking algorithms.

ACKNOWLEDGEMENTS

This research was supported by the MKE (The Ministry of Knowledge Economy), Korea, under the ITRC (Information Technology Research Center) support program, supervised by the NIPA (National IT Industry Promotion Agency) (NIPA-2012-H0301-12-3005). This work was supported by the National Research Foundation of Korea (NRF) grant, funded by the Korea government (MEST) (2012-047759).

REFERENCES

[1] R. Gross, *Handbook of Facial Recognition*, NY, USA, Springer, 2005.
[2] D. Comaniciu and V. Ramesh, "Robust detection and tracking of human faces with an active camera," The Third IEEE International Workshop on Visual Surveillance, 2000, pp. 11-18.
[3] M. Kim, S. Kumar, V. Pavlovic, and H. Rowley, "Face tracking and recognition with visual constraints in real-world videos," Conference on Computer Vision and Pattern Recognition, no. 1, 2008, pp. 1-8.
[4] E. Maggio, E. Piccardo, C. Regazzoni, and A. Cavallaro, "Particle PHD filtering for multi-target visual tracking," ICASSP, vol. 1, 2007, pp. 1101-1104.
[5] F. Dornaika and J. Orozco, "Real time 3D face and facial feature tracking," Journal of real-time image processing, vol. 2, 2007, pp. 35-44.
[6] M. D. Cordea, E. M. Petriu, and D. C. Petriu, "Three-Dimensional Head Tracking and Facial Expression Recovery Using an Anthropometric Muscle-Based Active Appearance Model," Transactions on Instrumentation and Measurement, 2008, pp. 1578-1588.
[7] G. R. Bradski, "Computer Vision Face Tracking for Use in a Perceptual User Interface," Intel Technology Journal, 2(2), 1998, pp. 13-27.
[8] K. Nummiaro, E. Koller-Meier, and L. J. Van Gool, "An adaptive color-based particle filter," Image Vision Computing, 21(1), 2003, pp. 99-110.
[9] B. Martinkauppi, M. Soriano, S. Huovinen, and Laaksonen, "Face video database," Proc. First European Conference on Color in Graphics, Imaging and Vision, 2002, pp. 380-383.
[10] http://www.cvg.cs.rdg.ac.uk/PETS2001/pets2001-dataset.html
[11] http://www.tele.ucl.ac.be/PROJECTS/M2VTS/m2fdb.html
[12] T. List, J. Bins, J. Vazquez, R. B. Fisher, "Performance evaluating the evaluator," ICCCN, 2005, pp. 129-136.
[13] http://fipa.cs.kit.edu/507.php
[14] Y. Li, A. Dore, and J. Orwell, "Evaluating the performance of systems for tracking football players and ball," Proceedings of the IEEE International Conference on Advanced Video and Signal Based Surveillance, 2005, pp. 632-637.
[15] A. T. Nghiem, F. Bremond, M. Thonnat, and V. Valentin, "ETISEO, performance evaluation for video surveillance systems," Proceedings of the IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS '07), London, UK, September 2007, pp. 476-481.
[16] K. Smith, D. Gatica-Perez, J. Odobez, and S. Ba, "Evaluating multi-object tracking," Proceedings of the IEEE Workshop on Empirical Evaluation Methods in Computer Vision (EEMCV '05), San Diego, Calif, USA, vol. 3, June 2005, p. 36.
[17] K. Bernadin and R. Stiefelhagen, "Evaluating Multiple Object Tracking Performance The CLEAR MOT Metrics," EURASIP Journal on Image and Video Processing, 2008, pp.1-11.
[18] J. Xiao, S. Baker, I. Matthews, and T. Kanade, "Real-Time Combined 2D+3D Active Appearance Models," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2004, pp. 535-542.
[19] R. Stolkin, I. Florescu, and G. Kamberov, "An adaptive background model for camshift tracking with a moving camera," Proc. International Conference on Advances in Pattern Recognition, 2007, pp. 147-151.
[20] P. Viola and M. J. Jones, "Robust Real-Time Face Detection," International Journal of Computer Vision, 57(2), 2004, pp. 137-154.
[21] C. Bao, Y. Wu, H. Ling, and H. Ji, "Real time robust L1 tracker using accelerated proximal gradient approach," CVPR, 2012, pp. 1830-1837.
[22] Z. Xiao, H. Lu, and D. Wang, "Object tracking with L2-RLS," ICPR, 2012, pp. 1351-1354.

Luu Ngoc Do
He received the B.S from Chonnam National University, South Korea in 2011. He is currently a M.S student at Dept. of Electronics and Computer Engineering, Chonnam National University, South Korea. His main research interests include Data Mining, Pattern Recognition, Machine Learning and Bioinformatics.

Hyung-Jeong Yang
She received her B.S., M.S. and Ph. D from Chonbuk National University, Korea. She is currently an associate professor at Dept. of Electronics and Computer Engineering, Chonnam National University, Korea. Her main research interests include multimedia data mining, pattern recognition, artificial intelligence, e-Learning, and e-Design.

Soo-Hyung Kim

He received the B.S. degree in Computer science from Seoul National University, Korea in 1986. He received the M.S. and Ph. D. degrees in Computer Science from KAIST, Korea in 1988 and 1993, respectively. He worked in Samsung Electronics as a senior researcher from 1990 to 1996 years. Since 1997, he has been a professor in the Electrical & Computer Engineering at Chonnam National University in Korea. His research interests include Artificial Intelligence, Pattern Recognition, Document Images Information Retrieval and Ubiquitous Computing.

Guee-Sang Lee

He received the B.S. and M.S. degrees in Electric Engineering from Seoul National University, Korea in 1980 and 1982, respectively. He received the Ph. D. degrees in Computer Science from University of Pennsylvania, in 1991. He has been a professor in the Electrical & Computer Engineering at Chonnam National University in Korea in 1994. His research interests include Multimedia Communication, Image Processing and Computer vision.

Sun-Hee Kim

She received M.S in Dongguk University, Korea. She received Ph. D degree at Dept. Electronics and Computer Engineering, Chonnam National University, Korea. She is currently a Post-doc researcher at School of Computer Science, Carnegie Mellon University, USA. Her research interests focus on data mining, sensor mining and stream mining.

In Seop Na

He received his B.S., M.S. and Ph.D. degree in Computer Science from Chonnam National University, Korea in 1997, 1999 and 2008, respectively. Since 2012, he has been a contract professor in Department of Computer Science, Chonnam National University, Korea. His research interests are image processing, pattern recognition, character recognition and digital library.