Multimedia Content Development as a Facial Expression Datasets for Recognition of Human Emotions

N E Mamonto*, H Maulana, D Y Liliana, and T Basaruddin
State Polytechnic of Jakarta, Depok, West Java, Indonesia

*Nicolaus.ernest.mamonto@hotmail.com

Abstract. Datasets that have been developed before contain facial expression from foreign people. The development of multimedia content aims to answer the problems experienced by the research team and other researchers who will conduct similar research. The method used in the development of multimedia content as facial expression datasets for human emotion recognition is the Villamil-Molina version of the multimedia development method. Multimedia content developed with 10 subjects or talents with each talent performing 3 shots with each capturing talent having to demonstrate 19 facial expressions. After the process of editing and rendering, tests are carried out with the conclusion that the multimedia content can be used as a facial expression dataset for recognition of human emotions.

1. Introduction
Humans are given the ability to express emotions he was experiencing. Emotions can be expressed through various things such as posture or gesture, voice, and facial expressions. A psychologist named Mehrabian in his research says that facial expressions accounted for 55% in message delivery. While the language and voice of each donate 7% and 38%. Current technological developments had a profound effect on all aspects of human life. Including face detection technology which has been widely used by devices or mobile computers that are used by humans. Face detection or facial recognition is one of biometric technologies that has been widely applied in security systems in addition to the introduction of the retina of the eye, the introduction of fingerprints and iris. In Indonesia until 2016, no researcher has done the research to the recognition of human emotions through facial expressions. Previous research conducted by New York University in 2013 in New York, USA. In the study, the subjects selected were local residents. And the study focuses only on images of human facial expressions and not videos. As a research material, the team needs datasets that contain video and facial images with certain expressions and emotions. In particular, the videos and images required are videos and pictures with Indonesian facial expressions. Talents are needed are 10 people, with each talent would be shot 3 times with emotion predetermined basic emotions are 6 types, 12 types of mixed emotions, and one facial expression without emotion. Shooting results must be edited into videos of each type of emotion and each video will create an image sequence from the first frame to the end. Therefore, authors carry out Multimedia Content Development as A Facial Expression Datasets for Recognition of Human Emotions to support the research team at the Laboratory Machine Learning and Computer Vision (MLCV) Faculty of Computer Science University of Indonesia so that new technologies can be generated that can be used by the public.
2. Literature Study

2.1. Video
According to Kamus Besar Bahasa Indonesia (KBBI) online the word video has meant the recording of live images or television programs to be aired on television. Meanwhile, the video is a sequence of many processing technology of moving pictures produced by the camera [1]. The video consists of a collection of pictures that are displayed in a sequence and in a certain speed so as to give the impression the images are moving. The set of images called frames and a measure of speed is called the frame rate as well as the units called frames per second (fps). Video is an electronic signal processing technology that represents moving images. A common application of video technology is television. Videos can also be used in engineering, scientific, production, and security applications. The term video is also used as an abbreviation videotape, video recorder, and video player. There are currently two categories of video, namely analog video and digital video.

2.2. Image sequence
Image sequences are videos that are rendered into each frame as individual images. Video with 200 frames will be exported as 200 individual pictures, each picture shows the process of the movement [2]. Meanwhile, image sequence is a collection of still images that form a sequence of movements [3]. In Multimedia Content Development as Datasets of Human Emotion Recognition Through Facial Expressions, image sequences generated from the rendering or exporting process of the raw video that has been edited. Image sequences contain the movement of talent expression of each frame.

2.3. Human emotion
Emotions are feelings or affections that can include passion of psychology, conscious experience, and behavioral experience [4]. Emotions are complex. Body, mind, and face play an important role in emotion. Emotions are divided into two categories: basic emotions and mixed emotions. Basic emotions can be defined as six, happy, sad, angry, surprise, disgust, and fear. Mixed emotions are defined as affective experiences that have the characteristic of activating two basic emotions, usually the two opposing valence emotions eg happy and sad.

2.4. Camera settings
The way the camera sees the object based on several things: the aperture (aperture), shutter speed (aperture), ISO or sensitivity [5]. Aperture or diaphragm is a window of light that enters from the lens to the film or sensor camera. Shutter speed is the speed of opening or closing the shutters of light to the sensor. The term ISO in the world of photography is the sensitivity of film or camera sensors. ISO is an acronym of the International Organization of Standardization. Exposure or lighting is the intensity of a light on a photo. Exposure is determined by aperture, shutter speed, and ISO.

2.5. Adobe Premiere Pro CC 2015
Adobe Premiere Pro CC 2015 is a video editing tool favored by video or professional fans and can enhance creativity and creative freedom. Adobe Premiere Pro is the most scalable, efficient and appropriate tool to edit a video. Adobe Premiere Pro CC 2015 can support various video formats including AVCHD, HDV, XDCAM, P2 DVCPRO HD, AVC-Intra, Canon XF, RED, ARRIRAW, QuickTime, and others [6].

2.6. Adobe Media Encoder CC 2015
Adobe Media Encoder is a stand-alone application that can run by itself or can be launched from Adobe Premiere Pro. Can work from Adobe Premiere Pro, Adobe Media Encoder adds the export queue files when the queue button is pressed on the setting of export in export panel settings. Aside from the Adobe Premiere Pro, Adobe Media Encoder can also work from some other source [7].
3. Design and Realization
Multimedia Content Development as A Facial Expression Datasets for Recognition of Human Emotions using the method of multimedia Development-Molina Villamil version. Multimedia content built in the form of video and images.

3.1. Design of multimedia products
Based on Villamil-Molina's multimedia development method, the design is divided into 2 stages, development and preproduction

3.1.1. Development. The idea came from the discussion and interview of the author with a research team at the Machine Learning and Computer Vision (MLCV) Laboratory of the Faculty of Computer Science, University of IndonesiaThe initial goal of this research is to produce multimedia content containing Indonesian facial expressions with 6 basic emotions: happy, sad, disgust, angry, and surprise, 12 mixed emotions are happily surprised, happily disgusted, sadly surprised, sadly fear, sadly disgusted , Sadly angry, fearfully surprised, disgusted surprised, fearfully disgusted, angrily disgusted, and an emotionless facial expression.

3.1.2. Preproduction. Preproduction is the stage of preparation of all kinds of needs required for production processes such as activity schedules, tool preparation, talent preparation, software preparation, and so on.

The required tools are camera and lighting set. The camera used was a Nikon D5100 and Canon EOS M3 obtained by way of rent. While the lighting set obtained by way made by researchers in accordance with the needs of the writer is for video shooting in the room and take done without the movement of the camera. The place required is a studio or location for shooting. Talent required is 10 people with the provisions of 5 men and 5 female gender. The talent must be Indonesian nationals. Talent is determined by the judgment method of the researcher.

3.2. Realization of multimedia products (production)
The production process according to Villamil-Molina involves the process of shooting and editing.

3.2.1. Process video shooting. Video shooting is done using pre-prepared equipment that is rented camera and lighting set that has been made. The position of the camera, lighting set, and talent position is determined in such a way as to get maximum image results, Placement plan of camera and lighting set can see in figure 1.

![Figure 1. Placement plan of camera and lighting set.](image)

In the first week, the camera used is Nikon D5100. This camera is used to take pictures of 3 talents. Settings performed on this camera are f2.3, ISO 100, and exposure +1.5. While in the second and third weeks, the camera used is Canon EOS M3. This camera is used to take pictures of 7 talents with details of 5 talents in the second week and 2 talents in the third week. Settings performed on this camera are
f2.3 and ISO 100 and exposures +1.5. Each talent takes 3 shots. In each shot, talent demonstrates 19 specified emotions. The process of shooting or video shooting is done for 3 days in 3 different weeks. In the first week there are 3 talents or subjects who do the shooting. In the second week there are 5 talents that do take pictures. And in the third week there are 2 talents that do take pictures.

3.2.2. Editing process. The editing process is done to process raw video into a video and image sequences, as well as labeling in accordance with the talent, take, and the expression of the video. In the process of video editing, Adobe Premiere Pro CC 2015 is used. For the queue rendering process, Adobe Media Encoder CC 2017 is used.

a. Raw video editing
   At this stage, the video that has been generated from the video shooting process is edited to become a clean video and can be an image sequence. Adobe Premiere Pro CC 2015 is used to perform video editing process. Each sequence is named according to subject and take video shooting. For subject 1 take 1, the sequence is named S1_T1_Master. For Subject 2 take 2, the sequences are named S2_T2_Master, and so on. Then raw video is inserted into the workspace sequence that has been made before. Then the video is trimmed using the Razor Tool according to the tag or marker of the start of the scene seen in the video preview. After the trim process for the video master finished, then made 19 new sequences for each video master. The 19 sequences are named according to the Subject, take, and expression shown. For subject 1, take one, happy expression sequence named S1_T1_Happy, to subject 2, take 2, sadly angry expression sequences named S2_T2_Sad-Birds, and so on. The naming of this sequence matches the label name to be given to the end result of the video and the image sequence. After a new sequence has been created, each previously trimmed video is inserted into a new sequence in accordance with the sequence name and marker on the video. After the video into the sequence has been determined, then do trim technique to remove parts that are not needed are part of the marker at the beginning of the video and transfer section at the end of the video scene. Then after the trim process is complete, the light settings are done by adding Brightness & Contrast effect. The brightness and contrast values for the resulting light are not much different from the light in a stable position in the middle of the video. The process of raw video editing is done to all raw video video shoot results. The total sequence required for the raw video editing process is 600 sequences.

b. Queue rendering process
   Once the editing process has finished, the video needs to be rendered to produce video files that can be played on a universal video player. But due to the vast quantity of video, the queue rendering process is done using Adobe Media Encoder CC 2015. Queue rendering is done through Adobe Premiere Pro CC 2015 by setting export presets and then clicking queue button to queue queue to Adobe Media Encoder CC 2015. The format used for video is H.264, this format is selected to generate file with mp4 format that is currently being popular. As for image sequence, the selected format is JPEG.

c. Rendering process through Adobe Media Encoder CC 2015
   Once all rendering queues have entered the render queue list, the rendering process is ready. The rendering process is done by clicking the start queue button in Adobe Media Encoder CC 2015 in the case of figure 2. Then automatically, Adobe Media Encoder CC 2015 will perform the rendering process in accordance with the settings that have been made before and in accordance with the queue that has been made.
The total file generated from this rendering process is 67,389 files consisting of 570 videos and 66,819 image sequences.

4. Discussion
After the production process is completed, then the next process is on the method of Villamil-Molina is post production which is testing phase.

4.1. Alpha testing
Alpha testing is done by looking at sample video and image sequences that have been completed in rendering and and will be used in beta testing [8]. The purpose of testing is to make sure no unneeded parts are still present in the rendered video.

4.2. Beta testing
Beta testing is done by testing the video and image sequences that have been produced [8]. Video testing is done by selecting video samples from 2 subjects and selected the same emotional sample. Each emotional sample is selected from 3 take that has been taken. Selected emotional samples are 1 emotionless expression, 1 basic emotion, 1 mixed emotion. While testing the image sequence is done by selecting 2 subject and selected the same emotional sample with the video. The selected expression is 1 take each. Then from the selected expression, the image is selected by selecting 2 images for 1 second of the video with a consistent frame interval or distance of 6 frames. The selected image is the first image and the 7th image in seconds running. In this test, the author chooses subject 8 and subject 9 that will be used as testing material. The selection of this subject is based on the judgment of the researcher with the assumption of using the best and worst subject according to the researcher. While the expression used is the expression without emotion, surprise expression (surprise), and sadly angry expression (sadly angry). As for the image sequence testing, there are 78 images consisting of 41 images for the first subject and 37 images for the second subject. Testing is done by showing the videos to the respondent. Each respondent was asked to fill in the questionnaire enclosed with 2 possible answers, is "Yes" and "No". Total respondents who did beta testing were 25 respondents for video testing and 25 respondents for image sequence testing.

4.3. Results and analysis of alpha testing
From the results of tests that have been done, it can be concluded that the alpha test, video and image sequence sample selected has been produced well and has been in the storage folder should be. The file size of each video is also in line with expectations, which is below 10 MB. All unneeded sections at the beginning of the video as well as at the end of the video are missing on the video's end result.
4.4. Results and analysis of beta testing

| No | Subject Expression                                      | Answer |
|----|---------------------------------------------------------|--------|
| 1  | Subject 9; Take 1; Facial expression without emotion (flat / neutral) | 88%    | 12%   |
| 2  | Subject 9; Take 1; Surprise expression                  | 96%    | 4%    |
| 3  | Subject 9; Take 1; Sadly angry expression               | 80%    | 20%   |
| 4  | Subject 8; Take 1; Facial expression without emotion (flat / neutral) | 96%    | 4%    |
| 5  | Subject 8; Take 1; Surprise expression                  | 36%    | 64%   |
| 6  | Subject 8; Take 1; Sadly angry expression               | 46%    | 54%   |
| 7  | Subject 9; Take 2; Facial expression without emotion (flat / neutral) | 48%    | 52%   |
| 8  | Subject 9; Take 2; Surprise expression                  | 96%    | 4%    |
| 9  | Subject 9; Take 2; Sadly angry expression               | 88%    | 12%   |
| 10 | Subject 8; Take 2; Facial expression without emotion (flat / neutral) | 96%    | 4%    |
| 11 | Subject 8; Take 2; Surprise expression                  | 64%    | 36%   |
| 12 | Subject 8; Take 2; Sadly angry expression               | 32%    | 68%   |
| 13 | Subject 9; Take 3; Sadly angry expression               | 92%    | 8%    |
| 14 | Subject 9; Take 3; Facial expression without emotion (flat / neutral) | 28%    | 72%   |
| 15 | Subject 9; Take 3; Surprise expression                  | 96%    | 4%    |
| 16 | Subject 8; Take 3; Sadly angry expression               | 48%    | 52%   |
| 17 | Subject 8; Take 3; Facial expression without emotion (flat / neutral) | 84%    | 16%   |
| 18 | Subject 8; Take 3; Surprise expression                  | 44%    | 56%   |

a. Question number 1, 7, and 14, which is intended to assess the expression without emotion of the subject 9 showed 88%, 48%, and 28% answered yes. Based on these results, the average score is 54.67%. This result shows that the emotionless expression of subject 8 is not perfect, and needs improvement.

b. Questions 2, 8 and 15, which were aimed at assessing the surprise expression of subject 9, obtained 96%, 96%, and 96% of the results answered Yes. Based on this result, the average score is 96%. This result indicates that the surprise expression of subject 8 is very good and deserves to be used as the next research data.

c. Questions 3, 9, and 13 aimed at assessing the sadly angry expression of subject 9 resulted in 80%, 88%, and 92% answered Yes. Based on this result, the average score is 86.67%. These results indicate that the sadly angry expression of subject 8 has been very good and feasible to serve as further research data.

d. Questions 4, 10, and 17 aimed at testing the emotionless expression of subject 8 resulted in 96%, 96%, and 84% answered Yes. Based on this result, the average score is 92%. This result shows that the emotionless expression of subject 8 is very good and deserves to be used as the next research data.

e. Questions 5, 11, and 18 aimed at testing the surprise expressions of subject 8 resulted in 36%, 64%, and 44% answered Yes. Based on this result, the average score is 48%. This result indicates that the surprise expression of subject 8 is still not perfect and needs improvement.

f. Questions 5, 11, and 18 aimed at testing the sadly angry expressions of subject 8 resulted in 46%, 32%, and 48% answered Yes. Based on this result, the average score is 42%. This result indicates that the sadly angry expression of subject 8 is still not perfect and needs improvement.

These are the analysis of beta image sequence testing.

a. Testing image sequences for emotionless or neutral expressions get high percentage responses above 50% for all images of subject 1 or subject 2. This shows for flat expression can be shown from the first frame of the video to the last frame of the video. It is pronounced emotionless expression is an expression that does not require a stimulus to be expressed so that the appearance of this expression can be seen from the beginning to the end of the video.
b. In the first subject the peak of the surprise expression is in the 30th frame until the 54th frame of the test video is visible from the 90% and above values obtained for the frame. While the transition occurs in the 24th frame. In the second subject, the survey results show that the 18th and 24th frames have reached the peak of the surprise emotions from the second subject from the 12th frame to the 60th frame.

c. In the first subject, the expression of the sadly-angry mixture occurs on the 42nd to 66th frames and the 78th frame up to the 108th frame. The transition occurs on the 72nd frame. In the second subject, the peak of the expression occurs on the 24th to the 36th frame with a value of 92%. Expression continues until the 66th frame with the value decreases from 88% to 68%. This shows that the second subject raises an extreme sadly-angry expression at the beginning of the video and begins to lower the emotional level up to the 66th frame.

Analysis was also conducted to recap back results by dividing respondents into two groups based on education midwives’ technology and non-technology [9]. The field of technology is better able to provide a firm assessment of the subject's expression although the image still shows the transitional expression from without emotion to expression with emotion. While the group of respondents with non-technical education field still shows doubt on some expressions. But in general, for other questions, the value obtained is not much different. This shows that the image sequences tested to the respondents can be well received by the respondents in general in accordance with expectations.

4.5. Delivery
Multimedia products are distributed to the research team at the Machine Learning and Computer Vision (MLCV) Laboratory of the Faculty of Computer Science University of Indonesia by the author through in the form of an external hard disk. Furthermore, the video and image sequences will also be distributed through the website to be accessible to the public at large and can be used to conduct similar research.

5. Conclusion
Based on the results of paper creation entitled “Multimedia Content Development as A Facial Expression Datasets for Recognition of Human Emotions” it can be concluded that:
1. Video and image sequences have been successfully created by involving 10 talents or subjects, with 3 shots for each subject, and each subject shoot should display 19 facial expressions.
2. The process of shooting is done for 3 weeks with the prepared equipment and talent that has been determined by the author. After the shooting process is complete, the editing process is done to remove the unnecessary parts of each video, labelling, and rendering process to produce video and image sequences according to the label specified.
3. After going through alpha testing, the entire video and image sequences are in the folder according to the subject, take, and the displayed scenes. Each video has also removed the scene marker at the beginning of the video and the scene transfer section at the end of the video. After going through beta testing, it can be concluded that video with subject 9 is better than video with subject 8, but overall video can be used as research data of human emotion recognition through facial expressions. Testing the image sequence done to 25 different respondents with the respondents on the video test resulted in the conclusion that in the neutral expression, the expression can be found from the first frame to the last frame, but for basic expression and mixed expression can be found on different frames but can be seen Transition expression more clearly on image sequence testing. The analysis was continued by dividing the respondents based on the education field i.e. technology and non-technology. Respondents in the field of technology can provide a more assertive assessment of facial expression shown compared to respondents from non-technological areas.
References

[1] Kurniawan E 2014 *Analisis kualitas real time video streaming terhadap bandwidth jaringan yang tersedia* (Medan: Universitas Sumatra Utara)

[2] Smith C 2013 *Video in Photoshop for Photographers and Designers. 1st ed.* (Berkeley, Calif.: Peachpit Press)

[3] Santosa B dan Purnama B 2010 Perancangan Studio Mini Berbasis Multimedia Universitas Surakarta *Journal Speed – Sentra Penelitian Engineering dan Edukasi* 2 2 17-25

[4] King L A 2011 *The Science of Psychology An Appreciative View (2nd ed.)* (New York: McGraw-Hill)

[5] Kurniawan H 2013 *DSLR Untuk Pemula 1st ed* (Jakarta: Mediakita) 8-15

[6] Jago M 2015 *Adobe Premiere Pro CC Classroom in a Book (2015 Release). 1st ed.* (Indianapolis: Adobe Press)

[7] Adobe Creative Team 2010 *Adobe Premiere Pro CS5 Classroom in a Book* (California: Peachpit)

[8] Windiyani T 2012 Instrumen Untuk Menjaring Data Interval. Nominal, Ordinal Dan Data Tentang Kondisi, Keadaan, Hal Tertentu Dan Data Untuk Menjaring Variabel Kepribadian *Jurnal Pendidikan Dasar* 3 5 203-208

[9] Darma A A G B 2013 Analisis Sistem Pengadaan Bahan Dan Peralatan Pada Proyek Kontruksi Jembatan Tukad Penet Di Badung Bali *Jurnal Ilmiah Elektronik Infrastruktur Teknik Sipil* 2 2 XIII-XII7