Research on Motion Tracking using Facial Features

Se In Jung, Shin Dong Ho
Graduate and Professor, My Paul School
12-11, Dowontongmi-gil, Cheongcheon-myeon, Goesan-gun
Chungcheongbuk-do, Republic of Korea
eavatar@hanmail.net

Jeongwon Kim
Department of Economics, College of Economics, Nihon University
3-2 Kanda-Misakicho, 1-chome, Chiyoda-ku, Tokyo, Japan
eavatar@hanmail.net

Abstract

In order to transmit real-time video, a CCD camera was implemented as the transmission medium. To process the input facial image, an MPEG file was created using DirectShow SDK, which supports multimedia streaming in MicroSoft Windows. Based on the 20 input feature points, the location of the feature point in each frame was based on the location of the same feature point in the previous image. The two-dimensional facial features were tracked using a template matching method using a coarse-to-fine tracking method that changes the template size in the search area of this particular region. The coarse-to-fine tracking method is one of the registration strategies that performs registration while increasing the resolution. When performed in real time, a transition image with appropriate resolution can be obtained within the desired time. However, if the data is acquired incorrectly at low resolution, there is the potential for errors to propagate. Matching refers to the task of selecting candidate feature points in the second image that can correspond to a particular feature point in the first image. Using the 2D coordinates tracked from the right and left images, 3D coordinates could be obtained using stereo image tracking points. The 3D data obtained allowed the creation of a compatible file in 3D Studio MAX, a 3D package.

Keywords
artificial intelligence, facial features, Face recognition, DirectShow structure and feature points

1. Introduction
In face recognition and expression recognition, a lot of research is being done to extract facial parts from images and extract areas and features such as eyes and mouth from the extracted facial parts. The study of facial expression has been studied in various scientific fields and significant achievements have been achieved. Based on this knowledge, many technical challenges need to be overcome in order to display 3D animations naturally, and facial expression recognition, which can estimate a person's internal state from facial images, is also being actively researched. In this study, we used two CCD cameras that input facial movements to generate MPEG files using the DirectShow SDK. We have developed a system that tracks 3D positional movements based on multiple facial features and maps the resulting coordinate values to a graphics package. The camera used for location determination has a wealth of information that allows a versatile response to stored image analysis and can extract features from stored data. We created a package compatible program to convert the data for use in 3D graphics.

2. Body
The DirectShow framework defines how multimedia data streams are controlled and processed using modular components called filters. Filters have inputs and outputs or all pins and are connected together in a configuration called a filter graph. The application uses an object called the Filter Chart Manager to assemble and move data through the filter chart. The Filter Graph Manager automatically manages the data flow. The Filter Graph Manager provides a COM interface to allow applications to access the Filter Graph. An application can call the Filter Graph Manager interface directly to control media streams or use the Media Player control to play media files. Therefore, you can access DirectShow through the media player control, which is a COM interface, or you can access DirectShow through

the Media Control Interface (MCI), as shown in Figure 1. Due to the modular structure of the DirectShow architecture, the graphics filter is very valuable.

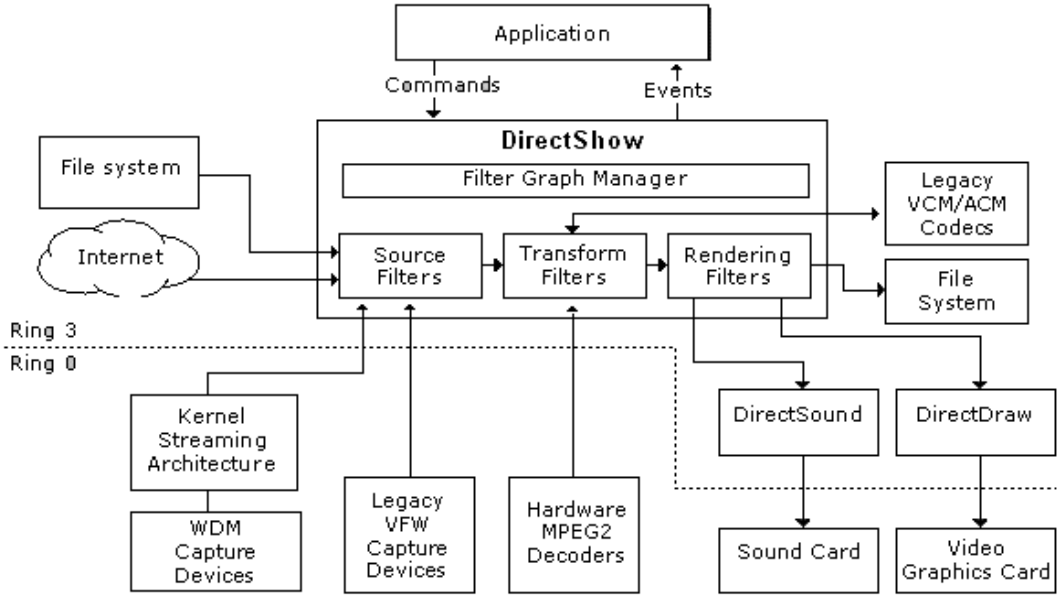

Figure 1. DirectShow Structure

With today's visual processing technology, it is not possible to extract much 3D information from 2D images and a few 3D images. After many characteristic points are extracted from the input image, they are transformed according to the characteristic points, and points that are not characteristic points are transformed with reference to the degree of change of the characteristic points compared to the general model. And we use images from angles that allow us to get a lot of 3D information. Most existing systems use front and side images, which provide most of the information about the face. The roles of specific filters are explained, e.g. B. Video or audio recording filters, AVI-MUX filters (multiplexers) and file writer filters. Filters with input pins are the most important part as above. Capture Filter Graph explains the filter graph, which is created by combining basic Capture filter graphs to provide multiple functions. Video preview filter charts allow you to view videos from a video player, a camera, or videos on your computer screen. These include a video recording filter and video rendering (Figure 2).

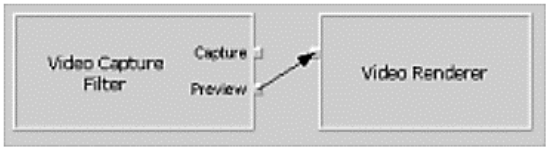

Figure 2. Video preview filter graph

Video capture filter charts capture captured video data and save it as a file. The term "video capture filter chart" often refers to both video capture and preview functions, but here it refers only to capture to a file.
The simplest video capture filter diagram includes a video capture filter, a multiplexer filter, and a file capture filter.

Figure 3. Video capture filter graph structure

Methods for creating a specific person's face using 2D images (Figure 3) have several similarities. Current visual processing technologies cannot extract much 3D information from some 2D images or some 3D images. After many characteristic points are extracted from the input image, they are transformed according to the characteristic points, and points that are not characteristic points are transformed with reference to the degree of change of the characteristic points compared to the general model. And we use images from angles that allow us to get a lot of 3D information. Most existing systems use front and side images, which provide most of the information about the face.
In this paper, information can be obtained using right and left images with 15 feature points.

A feature point is a point such as a corner point of an object in an image or a pixel with a large change in brightness compared to surrounding pixels, and these points are different from other points in the image. These function points can be defined in different ways, e.g. B. as a point with high difference values in different directions or as a point with properties that are invariant to a particular transformation.

Feature points are commonly used in computer vision areas such as stereo vision, 3D detection and motion estimation, and their role is particularly important in areas such as image registration and camera calibration.

The feature point matching method based on direction difference is a method that calculates direction difference values for each pixel in an image and uses points with sufficiently large values as feature points. A representative method is the Plessey function point extractor proposed by Harris. The method based on comparing brightness values has a feature point extractor based on comparing the brightness values of each pixel with the surrounding pixels.

Methods that use directional differentiation have disadvantages. First, local measurements of derivatives are very sensitive to noise. And when smoothing is used, the accuracy of feature localization decreases. In addition, compensation calculations, calculations of difference values and calculations of function points require a lot of computing time. In contrast to the former, methods that use the comparison of brightness values have the advantage of being simple and quick to calculate and stable against noise.

As one of the registration strategies, the coarse-to-fine technique is used to perform registration in low-resolution images while gradually increasing the resolution. This technique is also called the pyramid technique. For example, if there is a 512×512 image, the method is to first downsize to a 64×64 image and perform registration, then gradually increase the resolution and gradually display the distance difference in more detail. The matching is done based on the values obtained in the previous step. In this case, not only can the search area be reduced, but errors in the event of sudden variations can also be minimized. In addition, when carried out in real time, there is the advantage that a transition image with a suitable resolution can be obtained within the desired time. But. If the transition image obtained at the lowest resolution is obtained incorrectly, it is disadvantageous to propagate the error to the next step.

It takes advantage of function-based stereo technology. An edge is a point where the brightness of an image changes from a low value to a high value or from a high value to a low value. There are many applications that use edge detection and it is also used for a variety of special effects. In addition, edge segments or surfaces are also used as features.
Feature matching has the advantage of being more reliable than contrast-based matching for the matching part. The specification-based stereo method also performs a faster adjustment than the contrast-based stereo method. Additionally, when the edge is represented as a line segment, accuracy in subpixel units can be achieved.
However, the problem with the distance difference map is that it is too coarse. Typically, a dense distance difference map is created from the distance difference obtained through an interpolation process. However, there is also a problem that if an edge extracted from one image is not extracted from another image, a large error may occur due to edge mismatch.

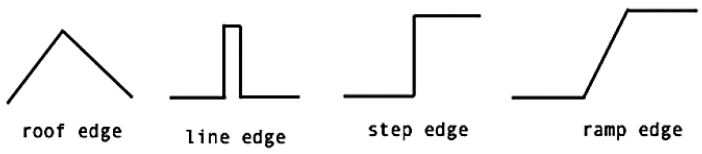

Figure 4. Side view of various types of edges

Keith Waters saw the source of this force in the muscles buried in the facial skin and performed facial animation using the muscles. If you look at the structure of the muscle, one part is attached to the facial skeleton and the other part is buried in the skin tissue (Figure 4)

For each muscle, the area of influence of the muscle, the maximum movement and the point at which the muscle movement begins and ends are defined, and as each muscle moves, the skin points associated with it move.
Anatomical models were used to recognize facial expressions. Methods based on anatomical structures not only enable realistic animations, but also enable a natural understanding of the motivation behind the movements used in animations. In other words, the degree of muscle relaxation and contraction is standardized and used to represent facial expressions.

This method not only allows you to create realistic animations, but also allows you to naturally understand the motivation behind the movements used in the animation. However, it is not only difficult to model the face close to the actual anatomy. This method cannot express detailed differences such as wrinkles or swelling of the skin.
In this study, we use a stereo matching method that is not difficult to implement and allows precise measurements over a relatively large area.

3D STUDIO MAX features an all-new real-time interface that supports a dual-processor system and a graphics hardware accelerator, supporting both desktop functions and workstation-level performance. Therefore, some users suggested that it is not an upgrade concept to 3D STUDIO but a completely new tool. The R1.0 version includes features such as track display, spatial deformation, support for spline and polygon modeling, and inverse kinematics, which offers a new user interface and improved time editing controls including unlimited time curves and notation tracks.

3D Studio MAX can support various files such as .max, *.3ds, *.al, *.dxf and *.shp, including *.max and *.3ds files that support animation. In this compatible file, feature points have been created as boxes and animations have been created using keyframes. Keyframes are a method of automatically generating values between frames to match actual behavior by specifying each object in just a few frames of the entire frame. However, since the purpose of this article is to obtain the location data of each frame, a compatible file has been created by specifying the location in each frame.

In this study, MPEG files were created from videos using DirectShow SDK. In the first frame of the created MPEG file, the first search value for each feature point was entered using the mouse. The saved MPEG file is decoded to create a BMP file. Perform a pre-processing process to get a grayscale image using the created BMP file. After performing the preprocessing process, the position of the feature point is tracked for each frame and x, y coordinate values are extracted. With the coordinate values extracted for each image, the three-dimensional coordinate values x, y, and z are calculated using the stereo image tracking points. The extracted 3D coordinates were used to create a compatible file that could be animated.

The size of the search area was Δx and Δy each 5 pixels, and the total search area was 11×11 pixels per feature point. If optimal search is not achieved, the search area is expanded to 20×20 pixels per feature point. The position error of the tracked coordinates was calculated by comparing the results of manual search and automatic search using 10 images. It was confirmed that the average absolute error value of 10 images for 15 feature points was accurately tracked with 0.15 pixels in the x-axis direction and 0.11 pixels in the y-axis direction. The following Figures 5 and 6 show the right and left images, and Figure 7 shows the results of feature point registration for a specific frame.

Figure 5. Image taken to the right

Figure 6. Image taken to the left

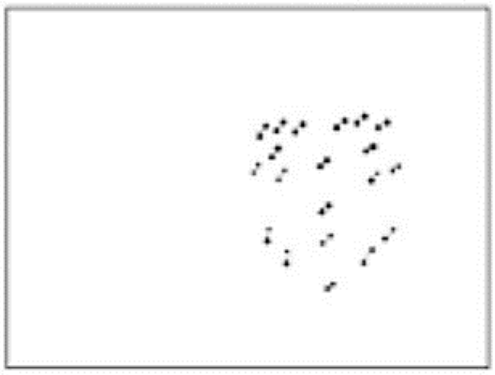

Figure 7. Result of matching feature points of a specific frame

In this article, two CCD camera devices were used to save the left and right images of an MPEG file using the DirectShow SDK.

As a result of tracking with the template matching method using the coarse-to-fine method that changes the template size, two-dimensional coordinates were created. This method allowed us to create a file compatible with 3D Studio MAX, a 3D graphics package, using a 3D tracking algorithm.
In this study, it became possible to move function points, such as expressing different facial expressions, using 3D data. However, this study had
limitations in installing light-blocking film, fixing the camera, and using marked feature points in facial images.
As a result of the research of this work, methods of connecting two CCD cameras using a personal computer and methods of tracking feature points using unlabeled facial images should be investigated.
Through this study, it is expected that it can be applied to various fields such as natural composition and real-time animation, similar to existing characters rather than virtual ones.

3. Conclusion

The problems of visualizing and analyzing key issues using R-program word cloud techniques are first caused by the omission of technical terms and new words in the Korean dictionary (KoNLP), secondly, the analyst's poor use of R-program and poor interpretation of third visualization. Heuristic preprocessing and post-processing processes are very important, which must be extracted from low-frequency but high-importance words, and which are not included in the commercial Korean dictionary are manually added. Therefore, the proposed problem solving and refining model are evaluated to be very useful for increasing the reliability of the verification result data and analysis results, and these measures are meaningful as practical application guidelines for word cloud techniques.

Future research should be conducted on commercial word cloud techniques that can improve the reliability of unstructured text analysis, not interest-oriented or text-design word cloud, by complementing the problems of the development and research results for big data analysis.

References

M. Han, Y. Kim, C. Lee, Analysis of News Regarding New southeastem Airport Using Text Mining Techniques, Smart Media Journal, Vol. 6, No. 1, 2017.

Jong Suk Lee and 3 others, Big data analysis of civil complaint texts using R language, 2020.

Kwon H., "Efficient feature point matching technique using unique matching pairs," Pohang University of Science and Technology, 1998.

Medioni G. and Yasumot Y.,, "Corner detection and curve representation using cubic b-splines," CVGIP, Vol.39, pp.267-278, 1987.

Han T., "3D facial model system for realistic facial expression animation," KAIST, 1997.

Berdsley P., Torr P. and Zisserman A., "3D model acquisition from extended image sequences," in ECCV, pp. 683-695, 1996.

Harris C.G., "Determination of ego-motion from matched points," in Proc. Alvey Vison Conf., 1987.

Haralick R.M. and Shapiro L.G., Computer and robot vision(volume 2), Addison Wesley publishing, 1993

Park Y., Yook Y., An S. and Cho C., "3D facial motion capture system for automatic graphic animation production," Korean Multimedia Society Fall Conference Proceedings, pp.653-657, 1999.

Park D., Shim Y. and Byeon H., "3D face synthesis and animation based on facial movement tracking," Journal of Information Science and Technology, Vol. 27, No. 6, pp.618-630, June 2000.

Koo B., "3D Shape Information Restoration Using Stereo Registration," '98 Spring Conference, pp.151-154, 1998.

Minagawa T., Saito H., Ozawa T., "Face-Direction Estimating System Using Stereo41 Vision", Proceeding of the 23rd International Conference on Industrial Electronics, Controls, and Instrumentation, Vol.3, 1997, pp.1454～1459.

http://www.microsoft.com/DirectX/dxm/.

Qiong L., Charissa L. and Thomas H., "Facial motion Tracking from Fine to Coarse", Proceedings of the 14th International Conference on Pattern Recognitionm. Vol.1, pp.725～727, 1998.

http://members.tripood.lycos.co.kr/jhcsang/seminar/survery/ sld001.htm.

Biographies

Se In Jung is Graduate in My Paul School. She is interested in artificial intelligence, deep learning, cryptography, robots, mechanical engineering, automotive engineering, architectural engineering, block chains, drones, autonomous vehicles, etc., and is conducting related research.

Jeongwon Kim is Graduate in College of Economics, Nihon University. She is interested in artificial intelligence, deep learning, cryptography, robots, block chains, drones, autonomous vehicles, etc., and is conducting related research.

Shin Dong Ho is Professor and Teacher in MY PAUL SCHOOL. He obtained his Ph.D. in semiconductor physics in 2000. He is interested in artificial intelligence, deep learning, cryptography, robots, block chains, drones, autonomous vehicles, mechanical engineering, the Internet of Things, metaverse, virtual reality, and space science, and is conducting related research.