IRSTI 28.23.15
UDC 004.93

ANALYSIS OF METHODS FOR SOLVING THE PROBLEM OF 3D FACE DETECTION

P. Komada[1], A. Sadykova[2]
[1] Lublin University of Technology, Lublin, Poland
[2]al-Farabi Kazakh National University, Almaty, Kazakhstan
[2]s_akmanat@mail.ru
[1]ORCID ID: https://orcid.org/0000-0002-9032-9285
[2]ORCID ID: https://orcid.org/0000-0002-5512-2297

Abstract. This article considers the problem of face detection in the case of 3D and discusses the current state of the problem, describes the principles of operation of existing 3D detection systems. The fact is that the methods for reconstructing a three-dimensional face model have limitations and drawbacks, which do not allow using them to effectively solve the problem of detecting faces from a video sequence. The methods of restoring the shape from motion, based on matrix factorization, restore the three-dimensional coordinates of only some points of the object, therefore, the problem of interpolating its surface between the restored points of the model arises. Reconstruction of a three-dimensional scene from a pair of images taken at different angles can give an accurate three-dimensional image for almost all points of the original scene, but requires high accuracy of data on the relative position of the cameras. Methods for obtaining a shape from shadows do not allow correctly recovering three-dimensional information about an object in real conditions, when the nature of the illumination is unknown and can be arbitrarily changed.

In research process, the common methods of face recognition in 3D at a given point in time were initially considered. Based on the analysis and research of existing methods, methods for solving a key problem in the field of three-dimensional face recognition - obtaining three-dimensional information about a face are presented.

Keywords: face detection, SFM, PCA, restoration of 3D model of the face.

Introduction

Considered theme must be a grandiose event, most of which have not arrived yet, but are still rising through the present asphalt. Face detection technology is one such occurrence. By the observation, face detection and recognition was mass manufactured in 2017, also now already continued to demonstrate brilliant effectiveness. And terrify whole states to shaking knees.

It is clear that face detection is one of the most promising methods of biometric contactless identification of a person by face. The first face detecting and recognition systems were implemented as programs installed on a computer. Nowadays, facial recognition technology is most often used in video surveillance systems, access control, on a variety of mobile and cloud platforms. The Massachusetts Institute of Technology Journal - MIT Technology Review included facial recognition technology in its 2017 Top 10 Breakthrough Technologies [1].

It is known that one of the main tasks in the field of visual detection is still face detection. A lot of research and development is devoted to this problem, however, the effectiveness of existing face detection systems is still far from human capabilities.

Currently, a relevant and intensively developed direction of research in the field of detection of visual images, especially faces, is the direction associated with obtaining three-dimensional information about an object.

3D detection algorithms [2] use information about the depth and curvature of the surface, unlike 2D detection systems [3] that traditionally use features based on the brightness of image pixels. Consequently, 3D descriptors are more accurate in describing surface features; better suited for describing the properties of the face in the cheeks, forehead and chin; invariant to angle and lighting.

At the moment, there is a task of creating a face detection system based on extracting 3D information from a video stream, which consists in the following: a person sequentially turns his head in three degrees of freedom (tilts forward-backward and left-right, turns left-right) in front of the camera, according to the resulting sequence of frames is built a three-dimensional model of the head and face detection is performed (comparison of the resulting model with the available models in the database).

The main goal of this article and research is to consider modern methods of 3D face detection to solve the problem of developing an effective face detection system.

The current state of the problem of 3D face detection

Currently existing 3D face detection systems use special equipment to reconstruct a three-dimensional face model (sensory systems). Touch 3D detection technologies fall into three categories:

1. Stereo. Two cameras with a known relative position are used to obtain a stereopair of object images; the corresponding points are found on the obtained images and the position of the matched points in three-dimensional space is calculated.

2. Structured light. This approach uses a camera and a light projector: the structural light projects onto the face a special texture, and the camera registers the distortion of this texture on the volumetric object. Using recovery methods shape by texture is calculated the location of points in three-dimensional space.

3. Laser scanning. Laser scanners use light as a source to detect the distance to the scanned object. They measure the reflection time of the laser from the object and receive information about the depth of the points on its surface.

Despite the fact that such technologies give a very high result (detection error within one percent), even with ideal illumination, sensory systems are subject to drawbacks: a frequent case is the appearance of artifacts in the model in the form of "holes" and "protrusions" due to missing data and recovery errors. Another disadvantage of sensor systems is a small depth of field for obtaining the necessary information, for stereo systems - about 0.3 m, for systems with structured lighting - about one meter. Another disadvantage is complex and often expensive equipment.

Nowadays, the following companies that are involved in the development of 3D sensor detection technology can be noted: Geometrix (USA), Genex Technologies (USA), Bioscrypt (Canada), L-1 Identity Solutions (England). In Russia, the Artec Group is working in this direction.

3D face detection systems that do not use additional equipment exist only as experimental developments and do not yet have commercial applications.

Some ways to obtain 3D face information

In order to obtain three-dimensional information about an object, algorithms are used, united in the English-language literature under the name "shape from X" (obtaining a shape using X), where X means a variety of methods [4]. Let us consider those of them that are most promising from the point of view of solving the problem posed in the introduction.

Restoration of the shape by shadows (shape from shading, SFS). The task of restoring the shape of an object by changing the brightness of the image pixels is based on a person's ability to determine the shape of an object using visual information about the nature of light reflection on its surface. This task is a task inverse to the task of visualization (rendering), when the brightness of a point of the simulated scene depends on a number of factors and is calculated according to a given mathematical lighting model [5].

Among the factors affecting the brightness of a point on the surface of an object, the following can be distinguished:

1. Properties and location of light sources.
2. Characteristics of a surface that determine its reflective properties.
3. Orientation of a surface area corresponding to a given point in space.
4. The point of view of the observer

As a mathematical model of the interaction of light and a surface, the Lambert scattering model is usually used, which describes the function of the dependence of the brightness of an image point on the intensity of a single light source, albedo (reflection coefficient) of the surface and the scalar product of the unit normal to the surface and the direction vector to the light source.

Since this mathematical model contains a large number of unknown parameters, in order to reduce the SFS problem to a solvable form, various simplifications are applied, mainly regarding the direction of illumination.

Accordingly, the disadvantages of "shape from shading" algorithms are: the need for a priori knowledge of the law of scattering, too general assumptions about the reflective properties of the surface, leading to incorrect reconstruction of the surface shape. The problem of restoring the shape from filling from the point of view of face detection is most fully covered in [6].

Reconstructing a shape from a stereo pair

Construction of 3D models from a stereopair is traditionally considered as two sequential tasks: stereo comparison and construction of 3D models from a set of points. The task of the algorithm is to obtain data on the distance to objects in the scene, on the basis of which a disparity map is built.

Most of the existing stereo matching algorithms can be divided into two categories of solutions: local and global. Local methods are based on finding feature points and matching them between two frames. Global methods look for a correspondence between images for each pair of pixels, and since there are areas where any texture is missing, smoothness constraints are imposed. Good algorithms look for the displacement map as a piecewise smooth function, with a limited number of discontinuity lines, and take into account that some points are visible in only one image. Generally speaking, local algorithms are less computationally expensive, while global algorithms generate more accurate displacement maps.

Reconstruction of a three-dimensional scene from a stereopair is capable of giving a high-quality result and restoring a three-dimensional image for almost all points of the original image, however, it requires high accuracy of calibration of the stereopair cameras.

With a lack of the spatial structure of an object (the absence of pronounced characteristic points and texture), stereo restoration algorithms find only rough details of the object's shape.

Shape from motion (SFM) restoration

This method of reconstruction of three-dimensional scenes uses the relative movement between the camera and the scene in a sequence of images [7]. As in stereo reconstruction, the SFM problem can be divided into two subtasks: finding a one-to-one correspondence of characteristic points on successive frames and scene reconstruction. But there are also some important differences. The difference between successive frames is much less than between images in a typical stereo pair, since the video is shot at a frequency of several tens of frames per second. Also, unlike stereo, in motion the relative displacement between camera and scene is not necessarily caused by the same 3D transformation.

In terms of matching, the SFM algorithm provides many closely related images (video frames) for analysis, and this is an advantage of this approach. First, tracking techniques can be used here that use the history of movements to predict differences in the next frame. Second, the matching problem can also be viewed as the problem of assessing the apparent motion in an image (optical flow).

As a rule, two types of methods are used to determine correspondences. Differential methods use time derivatives estimates and therefore require a dense sample of sequential images. This method works with every pixel in the image and results in dense measurements. Other methods use a Kalman filter to match and track point characteristics. These methods work with a small number of image points and result in sparse measurements.

In contrast to the problem of finding matches, the problem of reconstruction in this approach is more complicated than that of stereo restoration. Restoring motion and structure frame by frame is more sensitive to noise. This is because the baseline between successive frames is very small.

In the problem of shape restoration from motion, matrix factorization algorithms are used, with the help of which it is possible to restore the position and orientation of cameras, the internal parameters of cameras (focal length), i.e. parameters that are very often unknown. In addition, a large number of frames makes it possible to check the correctness of the matching. Additionally, in a number of cases, you can also obtain estimates of the reconstruction accuracy that correspond to a given scene.

SFM methods based on matrix factorization do not work directly with images, but require as input the coordinates of the characteristic points of the images in pixels and the presence of a marker (number) for each characteristic point, and the same point of the real scene must correspond to the same marker. Thus, the three-dimensional coordinates of only some points of the scene are restored; therefore, the problem arises of interpolating the scene surfaces between the restored points of the model.

3D face detection algorithms

Among the various approaches of 3D detection, three main ones can be distinguished: analysis of the shape of the 3D surface of the face, statistical approaches and the use of a parametric model of the face [8].

Analysis of the shape of a 3D surface. Methods based on the analysis of the shape of a three-dimensional image of a face use local or global characteristics of the surface that describes the face, for example, curvature, line profiles, metrics of distances between two surfaces.

Surface curvature is used to segment the surface of a face into features that can be used to compare surfaces. Another approach is based on 3D face surface descriptors in terms of mean and Gaussian curvature or in terms of distances and angle ratios between feature points of surfaces. Another locally-specific approach is the signature-point approach. The idea of the method is to form a representation-description of a selected point by neighboring points around a given point on the surface. These point signatures are used to compare surfaces.

Global methods use all information about a 3D face image as input to the detection system. For example, a face model is aligned based on its mirror symmetry, and then the face profiles are selected and compared along the alignment plane. It also uses the method of comparing face models based on the maximum and minimum values and the direction of the curvature of the profiles.

Another approach is based on the method of comparing the distances between the detection surfaces. Some methods are based on calculating the metrics of the smallest distances between the surfaces of the models, while others are based on measuring the distance not only between surfaces, but also the texture on the surface. However, a significant limitation of these methods is that the face should not be deformed and its surface is rigid.

The third approach is based on the extraction and analysis of 3D profiles and outlines highlighted on the face.

There are also hybrid methods based on combining local information about the surface in the form of local moments with a global three-dimensional mesh that describes the surface of the entire face. In one of these methods, the value of the function Z (x, y), which describes the depth map of the face in the aligned coordinate system, is decomposed into Fourier components. Decomposition of the function into moments (basis functions) makes it possible to smooth out fine high-frequency “noise on the face” and random outliers. In addition to the Fourier expansion, other basis functions are also used: power series, Legendre polynomials, and Zernike moments.

Statistical methods, in particular the Principal Component Analysis (PCA), were previously widely used in 2D detection. The PCA method was also implemented for 3D detection and was simultaneously extended to a combination of depth and color maps. An alternative for PCA is the linear discriminant analysis method, in which, unlike PCA, one object (a given person) is specified not by one person, but by a set of models (3D faces).

Another statistical method also borrowed from 2D detection is the Hidden Markov Model (HMM) method. The theory of Markov random fields allows one to construct estimates of various spatially variable quantities from images, while imposing certain a priori restrictions on these quantities. Such space-variable quantities can be, for example, offset values in the problem of stereo reconstruction. In the literature on 3D detection, this method is known as pseudo 3D hidden

Markov models and is used, in particular, for recognizing facial expressions.

Using a parametric face model. The key idea of detection by models is based on the so-called parametric 3D models, when the face shape is controlled by a set of parameters (coefficients) of the model. These coefficients describe the 3D shape of the face and can also set the color (texture) on its surface. This method uses one or more facial images as input, mainly photographs taken from the front and profile [9].

The algorithm for solving the problem is built on an iterative principle. As the initial iteration, a certain averaged three-dimensional model of the human head is selected, and its step-by-step improvement is performed. This uses a set of anthropometric points of the face, extracted from the photograph, which is deformed to a given three-dimensional surface. Deformation parameters are calculated during 3D reconstruction using an elastic model. These parameters are then used to recognize this face as a vector.

Conclusion

The above methods of reconstructing a three-dimensional model of a face have limitations and drawbacks that do not allow using them to effectively solve the problem of face detection from a video sequence. The methods of shape recovery from motion, based on matrix factorization, recover the three-dimensional coordinates of only some points of the object, therefore, the problem of interpolating its surface between the reconstructed points of the model arises. Reconstruction of a three-dimensional scene from a pair of images obtained from different angles can give an accurate three-dimensional image for almost all points of the original scene, but it requires high accuracy of data on the relative position of cameras. The methods of obtaining the shape from the shadows are not able to correctly restore the three-dimensional information about the object in real conditions, when the nature of the illumination is unknown and can be arbitrarily changed.

Thus, the most reasonable is the combination of different approaches, proposed in [10]. Algorithms based on matrix factorization are able to provide the necessary information about the three-dimensional coordinates of cameras, their orientation in space and the accuracy with which these values are known. In this case, the methods of stereo matching can give a dense reconstruction of the three-dimensional surface of the object (for each pixel of the image).

In addition, the possibility of determining and comparing the corresponding points of the face in a sequence of video frames based on the construction of a scene illumination model is interesting for further research

References

[1] Top 10 Breakthrough Technologies. URL https://www.technologyreview.com/10-breakthrough-technologies/2017/html (accessed December 12, 2020).

[2] Santos, O. Emotions and personality in adaptive e-learning systems: an affective computing perspective. In Emotions and Personality in Personalized Services. Springer International Publishing. 2016. 263-285.

[3] Wagner, J. and Lingenfelser F. The Social Signal Interpretation Framework (SSI) for Real Time Signal Processing and Recognitions. Proceedings of the INTERSPEECH Florence, Italy (2011). 152-158.

[4] Binali, H. and Potdar, V. Emotion detection state of the art. Proceedings of the CUBE International Information Technology Conference (CUBE '12). ACM, New York, NY, USA (2012). 501-507.

[5] Vezhnevec, V. Graphics and multimedia. URL http://cgm.graphicon.ru/issue2/Paper_vvp2/ (accessed December 15, 2020).

[6] William, A. and Smith, P. Statistical Methods For Facial Shape-from-Shading and Recognition. PhD thesis. University of York. 2007.

[7] Casale S., Russo A., Scebba G., Serrano S. Speech Emotion Classification Using Machine

Learning Algorithms. Proceedings of the IEEE International Conference on Semantic Computing. Santa Clara, CA (2008). 158-165.

[8] Kleinsmith A., Bianchi-Berthouze N. Affective Body Expression Perception and Recognition: A Survey. Proceedings of the IEEE Transactions on Affective Computing. (2013). 4(1). 15-33.

[9] Rosalind, W. Future affective technology for autism and emotion communication. Philosophical Transactions of The Royal Society. B: Biological Sciences. 2009. 364(1535). 3575–3584.

[10] Tao, J and Tan, T. Affective computing: a review. In Proceedings of the First international conference on Affective Computing and Intelligent Interaction (ACII'05), Springer-Verlag, Berlin, Heidelberg (2005). 981-995.