Software for 3D reconstruction of objects and faces using 2D photographs

Automatic 3D reconstruction is an important area of computer visualization, which in the last two decades has an important application in computer graphics, virtual reality, medicine, communications, and similar . The reconstruction process is inverse to the process of obtaining a 2D image based on a 3D model or scene. For the input set of images, which represent 2D projections of parts of a three-dimensional model or parts of a scene, it is possible to perform an automatic reconstruction of that model, or scene. There must be at least two images in the set, since on the basis of one image, information about the depth of any pixel can not be obtained. However, in practice, the number of images is usually higher, from several tens to several hundred, which depends on the size of the model, terrain or scene that is being reconstructed. The key to this process is the way in which the images from which it is necessary to generate the model are made. In practice, due to various external factors, the model can not always be reliably reconstructed. For example, images may have different exposures or not all parts of the model must be visible. The process of automatic reconstruction consists of the following four phases:

1. Determining the position of the cameras - at this stage, on the basis of the initial set of images, an assessment of the position of the camera in the space is made;

2. Determining the depth of the points - based on the initial images and certain positions, the cameras determine their depths for all significant points, which can approximately determine their position in the space;

3. Reconstruction of the model - based on the depth of the points and the position of the camera, a precise or approximate reconstruction of the model is carried out;

4. Texturing - at this stage, the generated unexhausted model is dyed.

In Figure 5.1, three phases of generating a 3D model are shown, assuming that the position of the camera is determined based on the input images. On the left part of the image there is a set of significant points with its depths. In this case, it is monochrome, but it is in practice near the depth, sometimes known information and the color of the dot. The next phase represents the reconstruction of the model, which represents an unpainted surface in the 3D space (part of the image in the middle). Finally, on the right-hand side of the picture is a final textured model.

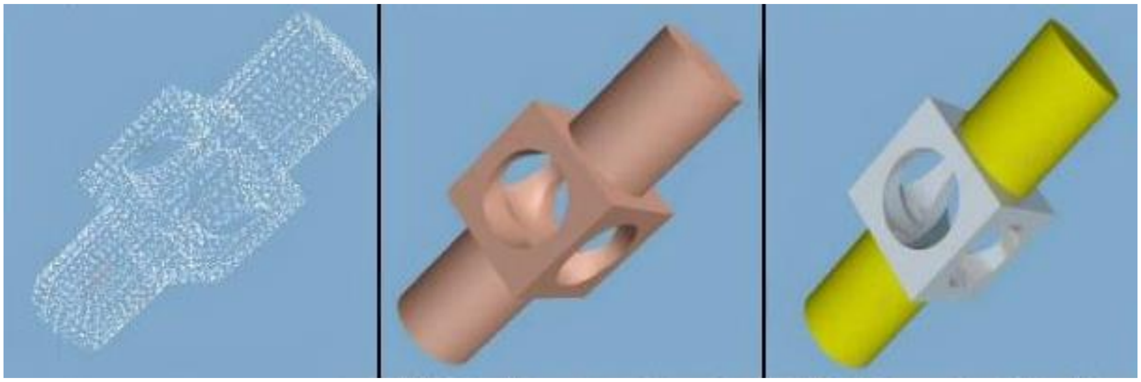


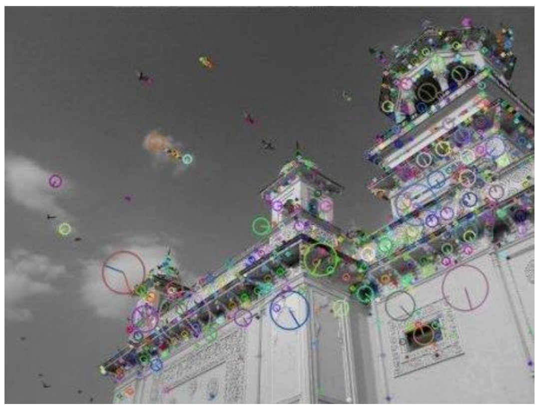
Figure 1. A set of points with depths, an unexpressed and textured model

Determination of camera positions

The first part in the process of automatic reconstruction of the 3D model is calculating camera sensor parameters, as well as computing camera positions in space. Camera sensor parameters consist of focal length, distortion coefficients, and points of projection of the center of the perspective to the plane of the image. The sensor parameters serve to calculate the point position in the 3D space of the camera.

The position of the camera in the space consists of rotation and translation. These parameters serve to calculate the position of points in the 3D space of the world. The positioning process of the camera begins by determining the key points in the images. Detection of key points is done using the SIFT algorithm (Figure 5.2).

Figure 2. A set of key points detected using the SIFT algorithm



After detection of key points, it is necessary to pair the key points from different images. Points for which a matching pair of two or more images are found are used in the reconstruction of the scene and the position of the camera. Pairing can be improved by giving initial assumptions about camera positions. If a picture is made by phone, as in the case here, it is possible to determine the rotation of the phone based on the sensor. This can reduce the scope of searches of key point pairs in other images, and thus the appearance of inadequate points being paired. There are several important tests on pairing. One of them is a mirror test. If a point is found at a certain point in the same image that is closer to the parameters than the point from the other image, that potential pair is declared invalid. This is important in situations where there are plenty of repetitive textures. In addition to this test, an angular test and a key point size test were implemented. All these tests reduce the chance of pairing inadequate points.

After pairing, a set of cameras and paired points are generated, whose exact positions in the space are not known. It is necessary to optimize these positions based on the current knowledge of the scene itself. Here is the optimization of the graph where the points and cameras move in each iteration. For each point, a set of cameras is known from which it is visible. On this basis, such a point can be projected to the plane of the image based on the current positions and calculate the distance from its potential position. Generated set of points in space is rare, but with precise depths. In the next phase, the depths of a large number of points are calculated, on the basis of which the model can be reconstructed.

Determine the depth of the points

A reconstruction of the model requires a much larger set of points than a set generated after the camera positioning phase. At this stage, known camera positions, which remain fixed, are known. To determine the depth of each point, two cameras are needed. One serves as a camera for whose points the depth is sought, and with the help of the other the depth of the individual points is determined. First, it is necessary to determine the set of pairs of cameras. For each camera, the best pair from a set received after the positioning phase is found. This set of pairs can be determined based on camera positions, or the number of key points seen by both cameras. After that, for each pair it is necessary to determine as many points as possible for which a certain depth value with a sufficient probability can be assumed.

There are several ways to determine the depth of the points from the first image. The majority is based on the assumption of the minimum and maximum depths seen in the first image. This data is drawn from a set of points that have been reconstructed in the first phase. For each point from the first image, a reproduction on the second image from the pair is found for the corresponding minimum and maximum depth. This creates the length of the second image. All points at that length represent a potential pair of the corresponding point from the first image. Defining the function that determines the similarity of the pixels is of great importance and is a key part of finding the depth.

Implementation of the determination of the depth of points in the space is done on the graphic card, because for the large number of points it is necessary to do the same calculations. Due to the possibility of parallelization on the graphics card, the algorithm is accelerated to several dozen times.

Reconstruction of the model

The reconstruction of the model presents the problem of obtaining 3D models based on a set of points that describe it. If the set of points is precise and sufficiently dense, it is about the exact reconstruction. However, in practice, often the set of points is incomplete or unevenly distributed, which is why the model is approximated by the existing set of points.

An initial set of points can be linear or thick. It is clear that the mere approximation of the model is more precise for a larger set of points. Also, in practice points may be unevenly distributed. Thus, in certain places, there are sufficient points on the basis of which reconstruction can be convincing, while in other parts of the point it may be missing. In this way, more or less holes can be created, which is why the lack of information and the approximation of the model is much more difficult. A hole can also appear in a set of points in the event that a certain part of the 3D model was not found in a single image from the starting set. In practical situations, the position of the points can not always be precisely measured, which leads to the appearance of noise. The method that successfully performs the reconstruction process must be robust to the occurrence of forests, in the sense that such points during the reconstruction of the model are inverted. Finally, there are frequent occurrences of isolated points, which are far from most of the other points. For these points it is assumed that they do not belong to the model, so they are usually ignored. Another option is that after generating the model, parts containing such points are simply eliminated.

Reconstruction methods take into account the way in which the set of starting points is presented. Usually, without additional assumptions about points, the problem is not well defined, since many models can be generated based on such a set of points. Most commonly, at each point, it is uniquely defined its normal, the rights that are normal to the tangent space. The Tagnent space represents the localized approximation of the model at a given point. Norms that do not have direction information are called unordered. Such kinds of normals can often be obtained from the very set of points. On the other hand, oriented norms carry information on the direction of delivery, which in most cases is to the interior or to the exterior of the model. Finally, besides the set of points, there are sometimes known information about the position of the cameras with their exact coordinates or coordinates obtained in the first phase of the automatic 3D reconstruction. In this paper, the used 3D model of the camera is determined in the first phase.

In this software, a method for reconstructing a model is used, which is a combination of the application of Delunius triangulation and the cutting of graphs. The method used assumes that there are known information about the position of the cameras. Delunian triangulation of a set of three-dimensional points represents a convex contour that is divided into tetrahedrons, so that no tetrahedron contains no additional point beside its four ends. The entire contour consists of a set of triangles representing the sides of the tetrahedron, and the final model will be a subset of this set of triangles. The method of cutting the graph has the same name as the well-known method in the theory of graphs, since it is based on it. Here, tetrahedrias represent nodes, and the triangles of the edges between two tetrahedra. The graph is oriented, since all triangles in triangulation are directed. As a result of the algorithm, it is necessary to determine which of the tetrahedrars is a visible space, and which space can not be seen by any camera. The graphs of the graph are assigned the initial values of the parameters that indicate the dimension of visibility, smoothness, surface, and the like. In this sense, most frequent are the cross sections of the tetrahedron with the rays that connect the positions of the cameras and the points from the set. Such tetrahedrers are expected to be visible in the resulting model. On such a controled graph, a method of maximum flow is used, on the basis of which visible and sheltered tetrahedrias are determined. The final model is a set of triangles (so-called interfaces) located between two tetrahedra, one of which is visible and the other is hidden

Texturing

Texturing is the latest phase of automatic 3D reconstruction. It is very important for many practical situations, where, besides the shape of the model itself, the corresponding texture is important. In some situations, texturing is not necessary, because it is only relevant to the shape of a 3D model (for example, the model itself can be constructed to determine the panorama). Textures sometimes provide pre-defined textures or textures can be generated based on images. If, for example, it is texturing the head model, it is possible to automatically recognize shapes such as the nose, mouth, eyes or eyebrows and, accordingly, paint the model. In the case of the triangulated model used in this paper, each triangle is painted with a unique texture. One way of obtaining a specific triangle texture is design from the image to the model, with the most commonly taken texture from the one from which this part of the model is most visible

Development Plan:

1. Implementation of the basic algorithm. Testing the solution obtained, and comparing the results obtained by different algorithms.

2. Optimization of results obtained in the first stages of the project. Implementation of additional algorithms for each of the steps and testing their effect on different photo sets

3. Optimization of performance and precision algorithms. A database of objects (forms) is created that is of interest to the user, and tests the solution obtained.

4. Creating a graphical user interface with an adequate set of commands.

Planned applications of software:

1. Integration into a video surveillance system of heavily accessible or inaccessible large territories (Subproject 1)

2. Integration with a database on terrorism and organized crime (Subproject 3)

3. Generating a panorama of interiors of objects of interest based on photographs