



HAL
open science

Matching correspondence between images and 3D model in a reconstruction process

Mau Le Tien, Khoi Nguyen Tan, Romain Raffin,

► To cite this version:

Mau Le Tien, Khoi Nguyen Tan, Romain Raffin,. Matching correspondence between images and 3D model in a reconstruction process. *Journal of Science and Technology: Issue on Information and Communications Technology*, 2016, 2 (1), pp.64-69. 10.31130/jst.2016.29 . hal-02113975

HAL Id: hal-02113975

<https://hal.science/hal-02113975>

Submitted on 29 Apr 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Matching correspondence between images and 3D model in a reconstruction process

Mau Le Tien, Khoi Nguyen Tan, and Romain Raffin

Abstract—3D reconstruction from photographs is an active research trend. The resolution of the sensors is increasing and the data processing is more accurate (not only restricted to calibrated stereo vision). In archaeological research it becomes a common way to safeguard some views of an ancient site, coupled with a manner to describe in 3D the artifacts with the same set of photographs. Archaeological scientists are now facing a complex problem to handle these digital data. An important usage is to describe semantically the artifacts. It is generally made "by hand", supplied by the knowledge of the scientists. We propose a solution that can perform a part of this work automatically, to generate descriptions of the obtained geometry. It combines image processing, geometry processing, 3D reconstruction. This paper aims at presenting an algorithm for 2D/3D point matching. The 3D reconstruction process of model from multiple views based on SIFT algorithm. The matching process uses a 2D mask pattern to lookup the 3D corresponding point. Experimental results show that our matching algorithm is precise, highly flexible, and can be successfully applied to a variety of 3D shapes.

Index Terms—Geometric Modeling, reconstruction, geometric characterization



1. Introduction

In recent years the 3D reconstruction usage is increasing as reconstruction algorithms get flexible and hardware get cheaper and accurate. The process of 3D reconstruction of model from multiple views consists of three following stages: preprocessing, features points extraction and identification the 3D points corresponding with 2D points. At first, the segmentation is performed based on the informations in each images. In second stage, these attributes are used to compute the characteristics, which describe the general and detail information of 3D points. They are robust features since they can exactly represent the information of each images, which greatly affects the 3D point generation.

Finding point correspondences plays an important role in identifying semantic shape models from a mask of 2D images. In this paper, we propose an algorithm for identifying the relationship of 2D-3D points between image mask and 3D model, in which the 3D points are extracted based on the geometrical properties of the images (via Scale Invariant Feature Transform – SIFT algorithms), to detect and describe local features in images [3].

The rest of the paper is organized as follows: Section 2 describes some related work; the details of the proposed approach are presented in Section 3; the ex-

perimental results are shown in Section 4 and Section 5 gives the conclusion and discussion.

2. Related Work

Methods to obtain accurate 3D models were initially based on laser scanner and range images. [11] presents a real-time scanner based on a video camera and a beamer to display pattern on the object. Patterns differentiate the object's shape and a 3D point cloud is obtained. As usual in these methods, the colors of points are not obtained, and another method is needed to reconstruct the topology of this point cloud: connecting the points by triangles for example. [5] proposed a method based upon photographs only. Similarly to [6], [12] the 3D model is obtained by pairing image couples.

David G. Lowe [3] presents a method for extracting distinctive invariant features from images. It can be used to perform reliable matching between different views of an object or scene. Dmitriy Bespalov [5] shows how a Scale-Space technique can extract features that are invariant, with respect to the global structure of the model.

Concerning archaeological objects such reconstruction is an important trend. Scientists can quickly save a digital copy of the artifact they work on (on the field), and manipulate it later (in the lab). In [1] for example an excavation is reconstructed and the archaeological scientists annotate the 3D model directly in a 3D environment (Meshlab). As described in [12] the next needs for 3D cultural heritage preservation are computational analysis tools and the enhancements of the metadata used to describe an artifact.

Despite that works have been done in 2D image matching, operating on 3D points achieved from different perspective remains to be a challenging problem.

Mau Le Tien is with the Quangngai College of Industrial Technology, Vietnam (Email: mau.letien@gmail.com).

Khoi Nguyen Tan is with the University of Danang, Vietnam (Email: khoi.nguyentan@gmail.com).

Romain Raffin is with Computer-Science at Aix-Marseille University, LSIS UMR 7296 (Email: romain.raffin@univ-amu.fr).

Manuscript received February 1, 2016; revised April 10 and June 25, 2016; accepted July 28, 2016.

Digital Object Identifier 10.31130/jst.2016.29

Partial 3D matching is a special case of 3D object reconstruction that involves finding correspondences on parts of 3D models. One of the most challenging problems is then to find correspondences between 3D objects, when the information describing the objects is not complete.

The work we describe in this paper is to provide the matching of pixel to 3D point correspondence and establish the first tool that can provides a characterization of a part of the 3D model.

3. Proposed Approach

These are main stages in our approach, namely pre-processing, feature points extraction, and 3D generation are presented in Fig.1. To ensure a correct organization of our approach, we have inherited some functions of Micmac toolkit [9] in our program to implement SIFT algorithm. Our goal is to generate a 3D data points from an 2D image set, and then from a mask pattern with 2D points, so that we propose an algorithm to compute all 3D points corresponding to the 2D points of this mask pattern.

3.1. Digitization process

3D reconstruction from multiple images is the creation of three-dimensional model from a set of images. It is the reverse process of obtaining 2D images from 3D scenes. The essence of an image is a projection from a 3D scene onto a 2D plane, during which process the depth is lost. The 3D point corresponding to a specific image point is constrained to be on the line of sight. From a single image, it is impossible to determine which point on this line corresponds to the image point. If two images are available, then the position of a 3D point can be found as the intersection of the two projection rays. This process is referred to as triangulation [3]. The key for this process is the relations between multiple views, which convey the information that corresponding sets of points must contain some structure, and informations on how the structure is related to the poses and the calibration of the camera.

From a set of images, the digitization process is built with various angles of sight. The focus is manual (and locked for the session), the white balance is made before the session and is kept unchanged, last the zoom level is also constant.

3.2. Pairing the SIFT markers

To solve the corresponding problem between two images, we choose the SIFT algorithm [8,10]. The main idea of SIFT algorithm is a local feature extraction algorithms, in the scale space looking for extrema points, extract the location, scale and rotation invariant. There are 4 basic steps in SIFT method:

- 1) Scale-space extrema detection
- 2) Feature point matching
- 3) Orientation assignment
- 4) Local key point descriptor generation

Once the SIFT feature vectors of the key points are created, the Euclidean distances between the feature

vectors are exploited to measure the similarity of key points in numerous digital images. A feature point from one image is chosen, and then another two feature points are found by traversing all the feature points in another corresponding image which have the shortest and next-shortest Euclidean distances. Among these two feature points, if the divisor between the shortest and next-shortest Euclidean distances is less than a threshold value, then they are judged to be paired feature points.

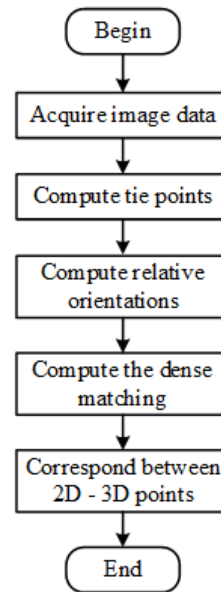


Fig. 1: Overview of our processing stages

3.3. 3D Point cloud generation

3D reconstruction is addressing the problem of recovering depth information from images. A physical point in space is projected onto different locations on images, if the viewpoint for capturing the images is changed. The depth information is inferred from the difference in the projected locations.

The 3D construction by multiple views is based on the two-views 2D construction. The most important problems to be solved are the feature matching among a set of images, and the transformation of the 2D coordinates into a set of 3D points.

In order to perform 3D points generation, we inherit the SIFT algorithm in the IGN Micmac software [9], to detects images features and correspondences. The images orientation and the 3D point cloud are also outputs of this software. The main steps in the SIFT algorithm to detect 3D points include: computing the tie points, detecting the relative orientations and the depth map from the oriented images [12].

3.3.1. Compute tie points between all the pairs of a set images using SIFT

This algorithm automates the workflow of feature matching between multiple images. The process is done as follows:

- Scale-space extrema detection: the first stage of computation searches over multiple scales and image locations. It is implemented efficiently by using a difference of Gaussian functions to identify potential interest points that are invariant to scale and orientation.
- Key point localization: for each candidate location, fit a detailed model to determine location and scale. Key points are selected based on measures of their stability.
- Orientation assignment: compute best orientations for each key point region based on local image gradient directions. This stage provide invariance to the assigned orientation, scale, and location for each feature. Therefore, it achieves invariance to image rotation.
- Key point descriptor: use local image gradients at selected scale and rotation to describe each key point region. These are transformed into a representation that allows significant levels of local shape distortion and changes in illumination.
- Choose an orientation and position of camera from previous process and computes image matching
- Calculate for each point in this space a measure of likelihood (which is the sum of correlation coefficients for each images pair).
- Extract the surface by using energetic approach to minimize, on the whole considered space, a sum made up of a likelihood measure and a smoothing term in order to homogenize the surface.

At this stage, we get the 3D point cloud created from 2D image. These 3D points are stored in a 3D data file (XYZ file) and associated with the master image. It allows to link with all pixels in the master image.

3.4. Correspondence between 2D points and 3D points

Once the three first stages one completed by MicMac process [9], we obtain a 3D point cloud of the complete artifact and the depth images files of the key images. A 3D point cloud is a set of data points in three-dimensional coordinate system. These points are defined by X, Y, and Z coordinates.

Using a black and white image mask, we can show/hide portions of a display object or display group according to the mask composition. For example, a mask image can define with white portions (unused points) and black portions (interesting points). We will now propose the algorithm to match interesting pixels in 2D images with corresponding 3D points of object. Imagine we wish to find a semantic data ("eye", "mouth" or "nose" for example), we must construct first an image processing algorithm to characterize pixels in the photos. It is based upon well known methods like denoising, finding contours and for this article the last step is human-assisted. It is applied on photos with their associated depth map images.

According to the 3D models created with the 3D points set, we present a novel 2D-3D matching algorithm, which is applicable to the pose detection and the correspondence between 2D points of mask pattern with 3D points on the model. Our method matches 2D black points against their corresponding 3D points directly as follows:

- 1) Read all pixel of 2D mask pattern in a sequential way, each pixel corresponds to a 2D point. When reading a black pixel, we note the row and column position (i, j) in the image and mark this pixel.
- 2) Perform the matching process for each black pixel (i, j) with all coordinates (x, y) of the 2D data point set (XY file). If it exist the value $i = x, j = y$, then put this value into a queue.
- 3) Find the corresponding 3D points of 2D points in the mask pattern: with each 2D point value in the queue Q, match this 2D point with all point in 3D data file to determine the correspondence between this point.

3.3.2. Compute relative orientations from tie points

In this stage, we need to determine the orientation, position and calibration of a set of images compatible with a set of tie points, because the set of images is obtained from different position and orientation of each image. Tie points are useful in the image orientation and they provide an approximation in image matching. The step to compute relative orientations based on the tie points is done as follows:

- Choose the first image, in arbitrary position.
- Find the best images to add by computing stability estimator on the cloud of tie points;
- Use the tie points to compute the initial value for an orientation of a new image with direct algorithms.
- Make one round of a bundle adjustment to avoid error accumulation.
- Repeat the steps from 1 to 3 for all images.
- Take out the internal parameters in this order: distortion coefficient, focal length, distortion center, principal point. Each time a parameter is released, process a bundle adjustment round.
- Construct several rounds of bundle adjustment with more and stricter weighting on the projection residual.

The result of this stage is the orientation and position of camera (internal calibration).

3.3.3. Compute the dense matching from the oriented images

We continue to compute the depth maps from a set of images, their poses estimation and a depth interval. This step provides a map of correlation coefficients that are corresponding in order to obtain the depth for each pixel in depth map. The steps for this process are the following:

- 4) Extract the corresponding 3D data to a new file and display these points.

4. Experiment Results

Our system is implemented in C++ language helped by the MicMac library [9] in order to compute SIFT markers, cameras orientations and 3D point cloud production. In our experiment, the proposed algorithm was tested on sample data sets that are provided with MicMac software [9].

Figure 2 presents Buddha images from various view-points.



Fig. 2: The 2D image data set of Buddha statue

The Figure 3 shows the complete 3D reconstruction (as points cloud) of the Buddha model (left image).

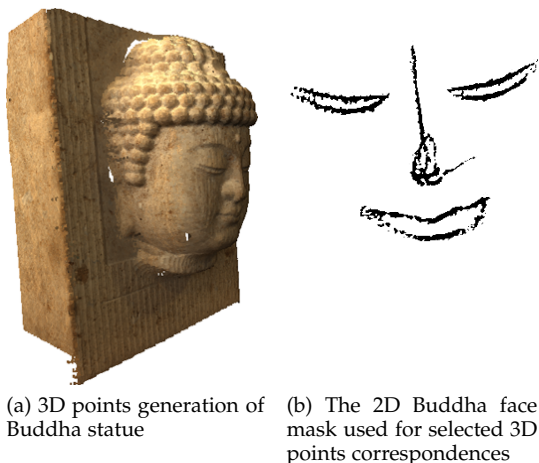


Fig. 3: Result of 3D Buddha model reconstruction a) and the 2D Buddha face mask used for select 3D points correspondences



Fig. 4: 3D points correspondences from 2D face mask on Buddha model

As seen on Figure 3b, we computed a blank and white mask with the recognized parts of the Buddha face. With the generated point cloud of master images, pixel coordinates are calculated. The coordinate of 3D points (X, Y, Z) are stored in 3D data files (XYZ files). The 2D coordinate (X, Y) of the point in XY file data is the resulting computation of the 3D point and the image orientation. By back projecting these 3D points into 2D image space, we can determine the relation between a 3D point at row i in XYZ file with a 2D point at row i in XY file. Each (X, Y) value in the 2D data file is corresponding with a pixel having (X, Y) coordinate on any image. If we have characterized pixels in the starting images, we obtain the corresponding 3D points.



Fig. 5: 2D Buddha head mask

In the Figure 3b, we use the marks to select a region of image and identify the 3D point correspondence with all 2D point in the mask. These 3D points correspondences are displayed in Figure 4. This result is consistent with the proposed method, and it will be used to define the semantics in 3D model.



Fig. 6: 3D points correspondence with 2D points of Buddha mask pattern

TABLE 1: 2D and 3D point matching of Buddha model

2D point of mask	3D points (X Y Z R G B)
871 247	-0.0059287 0.0621612 0.0520009 255 0 0
872 247	-0.0056792 0.0618721 0.0521558 255 0 0
870 248	-0.0056525 0.0621339 0.0521400 255 0 0
844 248	-0.0056223 0.0623959 0.0521611 255 0 0
869 251	-0.0053944 0.0618453 0.0523872 255 0 0
841 249	-0.0053696 0.0621069 0.0523515 255 0 0
842 249	-0.0053439 0.0623685 0.0523252 255 0 0
843 249	-0.0051256 0.0615575 0.0527489 255 0 0
844 249	-0.0045656 0.0636235 0.0535181 255 0 0
869 252	-0.0045413 0.0638847 0.0534775 255 0 0
...	...

Figure 4 illustrates the computing of black and white mask with the recognized parts of the Buddha face. Table 1 shows the correspondence of 2D points in Buddha face mask and 3D points of Buddha model reconstruction from multiple images (last three columns are color values).



Fig. 7: The 2D image data set of ancient Egypt statue



(a) 3D points generation of Buddha model (b) 2D Egypt face mask of Buddha model

Fig. 8: Result of 3D Egypt model reconstruction a) and the 2D Egypt face mask used for select 3D points correspondences

TABLE 2: Statistic number of 2D points and 3D points of Buddha model

Buddha data set	2D point number	3D point number
Data point number	1.456.954	1.456.954
Points correspondences of 2D Buddha face mask	4.632	4.632
Points correspondences of 2D Buddha head mask	65.768	65.768

TABLE 3: 2D and 3D point matching of Egypt model

2D point of mask	3D points (X Y Z R G B)
440 272	-1.2092670 0.7654080 -8.3441038 255 0 0
448 272	-1.1762609 0.7652940 -8.3434048 255 0 0
448 280	-1.1539530 0.7186260 -8.1854992 255 0 0
456 280	-1.1131790 0.7131350 -8.1234465 255 0 0
456 288	-1.1033241 0.6751960 -8.0518312 255 0 0
432 296	-1.1928380 0.6406020 -8.0135336 255 0 0
448 296	-1.1296279 0.6405250 -8.0135336 255 0 0
...	...

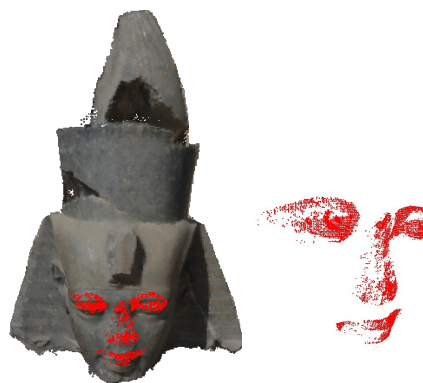


Fig. 9: 3D points correspondence with 2D head mask of Egypt model (in red)

The Figure 8a shows the complete 3D reconstruction of the Egypt model from the image data set (Figure 7). By using the marks to select a region of image (Figure 8b), we identify the 3D points corresponding with all 2D point in the mask. Figure 9 illustrates the 3D points correspondences from various orientations.

Based on the result from Figure 9, Table 3 shows the correspondence of 2D points in Egypt face mask and 3D points in the model reconstructed from multiple images. We can see that the points number of mask pattern is always equal to the 3D points number of model (Table 4). This can demonstrate the accuracy of the our solution.

5. Conclusion and Discussion

In this paper, we proposed a solution to find the correspondence between images and 3D model. Our algorithm achieves robust matching. The matching process is fully automatic, based on a mark as an input pattern. In our experiment, we have ran the algorithm on various data set samples (Buddha, Egyptian face statue...),

TABLE 4: Statistics of the number of 2D points and 3D points

	Buddha face mask	Buddha head mask	Egypt face mask
The 2D points number of mask pattern	4.632	65.768	5.490
The corresponding 3D points number	4.632	65.768	5.490

the results illustrate that our method has robust performance on 3D data, under various transformations and noise values.

The experimental result shows that the described features can represent the relation between 2D and 3D points correspondences by using a 2D mark from image data set. The advantage of the proposed approach is that it can utilize a 2D mask of 2D images to find all 3D points of model generation. It overcomes the limitations of solving a system of linear equations which is the most common problem when dealing with large number of images.

As further work, our approach will be improved for specific problems, such as image processing for feature point's detection in 2D images. The feature lines extracted from 3D model will be also considered to increase the characterization of mesh parts and the semantic knowledge of 3D areas. Finally, we wish to reconstruct the semantic identity of 3D model using both 2D to 3D points correspondences.

References

- [1] Adeline Manuel, Livio De Luca and Philippe Véron, "A Hybrid Approach for the Semantic Annotation of Spatially Oriented Images", *International Journal of Heritage in the Digital Era*, 2014.
- [2] Christian Lindequist Larsen, "3D Reconstruction of Buildings From Images with Automatic Façade Refinement", Master's Thesis, *Vision, Graphics and Interactive Systems*, 2010
- [3] David G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", *International Journal of Computer Vision*, Volume 60 Issue 2, November 2004.
- [4] Ding-Yun Chen and Ming Ouhyoung, "A 3D Object Retrieval System based on Multi-Resolution Reeb Graph". *Proc. of Computer Graphics Workshop*, 2002.
- [5] Dmitriy Bespalov , William C. Regli , Ali Shokouf, "Local feature extraction and matching partial objects" *Computer-Aided Design* 38(9), pp. 1020–1037, 2006.
- [6] G. Stavropoulos, P. Moschonas, K. Moustakas, D. Tzovaras and M.G. Strintzis, "3D Model Search and Retrieval from Range Images using Saliency Features", *IEEE Transactions on Multimedia*, vol. 12, no.7, pp. 692-704, November 2010
- [7] JC. Torres, G. Arroyo, C. Romo, "3D Digitization using Structure from Motion", *CEIG-Spanish Computer Graphics Conference*, 2012.
- [8] Jebara, Tony, Ali Azarbayejani, and Alex Pentland. "3D structure from 2D motion, *Signal Processing Magazine*", *IEEE* 16.3, pp. 66-84, 1999

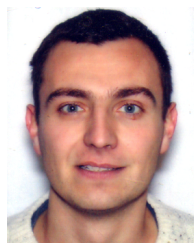
- [9] MicMac, Apero, Pastis and Other Beverages in a Nutshell, 2015.
- [10] SF. El-Hakim, JA. Beraldin, M. Picard, "Detailed 3D reconstruction of large-scale heritage sites with integrated techniques", *Computer Graphics and Applications*, *IEEE* Volume 24, Issue 3, pp. 21-29, 2004.
- [11] Park, Hyun Soo, et al, "3D Trajectory Reconstruction under Perspective Projection", *International Journal of Computer Vision*, pp. 1-21, 2015.
- [12] C. Baillard, C. Schmid, A. Zisserman, and A. Fitzgibbon, "Automatic line matching and 3D reconstruction of buildings from multiple views", In *ISPRS Conference on Automatic Extraction of GIS Objects from Digital Imagery*, volume 32, pp. 69–80, 1999.
- [13] R. Berthilsson, K. Astrom, and A. Heyden, "Reconstruction of general curves, using factorization and bundle adjustment", *International Journal of Computer Vision*, 41(3), pp. 171–182, 2001.



Mau Le Tien received the B.S. degree in Information Technology from the Quynhon University, Vietnam, in 2007, the M.Sc. degree in Computer Science from the University of Danang, in 2011. He is now a lecture at Quangngai College of Industrial Technology, Vietnam. His current research field is Computer vision.



Khoi Nguyen Tan received the B.S. degree in Information Technology from the Danang University of Technology, Vietnam, in 1997, the M.Sc. degree in Computer Science in 2003, and the Ph.D. degree in Computer Science from the Aix-Marseille University in 2010. He now is lecturer at University of Science and Technology, Danang, Vietnam. His current research interests include Geometric modeling, Computer Graphics.



Romain Raffin received the Ph.D. degree in Computer Science 2000 from the University of Burgundy, France, on constrained free-form deformations. He obtains the accreditation to supervise research (HDR) in 2013 from Aix-Marseille University. He is assistant-professor in Computer-Science at Aix-Marseille University since 2001. His current research interests in Computer Graphics are subdivision surfaces, mesh characterization and semantic reconstruction (LSIS UMR7296 lab).