Real-time three-dimensional surface measurement by color encoded light projection

S. Y. Chen[a)]
College of Information Engineering, Zhejiang University of Technology, Zhejiang 310014, China and Department of Informatics, University of Hamburg, Hamburg 22527, Germany

Y. F. Li
Department of Manufacturing Engineering and Engineering Management, City University of Hong Kong, Kowloon, Hong Kong

Q. Guan and G. Xiao
College of Information Engineering, Zhejiang University of Technology, Zhejiang 310014, China

(Received 1 April 2006; accepted 17 July 2006; published online 12 September 2006)

Existing noncontact methods for surface measurement suffer from the disadvantages of poor reliability, low scanning speed, or high cost. The authors present a method for real-time three-dimensional data acquisition by a color-coded vision sensor composed of common components. The authors use a digital projector controlled by computer to generate desired color light patterns. The unique indexing of the light codes is a key problem and is solved in this study so that surface perception can be performed with only local pattern analysis of the neighbor color codes in a single image. Experimental examples and performance analysis are provided. © *2006 American Institute of Physics*. [DOI: 10.1063/1.2352729]

Using artificial vision to obtain the three-dimensional (3D) model of an object is very important to many engineering applications. A number of 3D scanning methods have been explored by researchers in the past years. However, fast (especially real-time) acquisition of 3D surface from images is still a big challenging problem in the field.[1] The most widely known passive method for 3D acquisition is stereovision.[2,3] However, it has a correspondence problem which has been formulated as an ill-posed problem in a general context.[4] An alternative approach is the active method in which properly formatted light or laser is emitted in the direction of an object, reflected on its surface, and received by the sensor.[5–10] The correspondence problem is solved using multipass dynamic programming algorithm. However, these systems still need to take dozens of images for recovering a 3D scene. As a result, its speed is limited and applications are restricted to reconstruction of stationary objects. For dynamic use, 3D measurement using a single image is desired.

For this reason, a recent method is to use a color projector which can be controlled by a computer to generate arbitrary desired color patterns. A problem of the color encoded projection is the unique indexing of the light codes so that the coordinates in the color projection can be determined in a single image. In many cases some light grids are invisible or only partially visible due to occlusions and discontinuities of the objects. Therefore it is essential that each light grid be uniquely identified by incorporating only local neighborhoods in the pattern. Although some efforts have also been made in the coding technology[11–14] to improve the light pattern, a practical system has to consider more application requirements.

Here we consider a sensor consisting of a charge-coupled device (CCD) camera and a digital projector (Fig. 1). The idea is to design special light patterns for illumination projection and fast code matching. Based on this idea, the system combines both advantages of stereovision for fast 3D measurement from a single image and structured light system for accuracy and stability.

Let P be a set of color primitives, $P=\{1,2,\ldots,p\}$, (where the numbers $\{1,\ldots,p\}$ representing different colors, e.g., 1=white, 2=red, 3=green, 4=blue, etc.). These color primitives are assigned to an $m\times n$ matrix M to form the encoded pattern. We define a word by the color value at location (i,j) in M and the color values of its four adjacent neighbors. If x_{ij} is the assigned color point at row i and column j in M, then the word for defining this location is a substring like $w_{ij}=(x_{ij},x_{i,j-1},x_{i-1,j},x_{i,j+1},x_{i+1,j})$.

If a lookup table is maintained for all of the word values in M, each word defines a location in M. We can know that an $m\times n$ matrix M has $(m-1)\times(n-1)$ words. These words are made up of a set W. The condition is, we wish, to assign

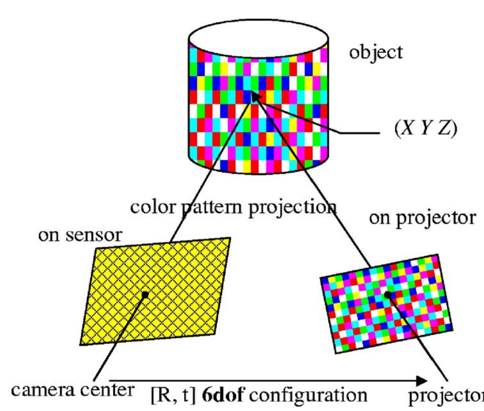

FIG. 1. (Color online) Sensor structure of color encoded vision for real-time 3D surface measurement. A light pattern with special codes is projected on to the object from the digital projector and a colored CCD camera is used to capture the scene image. 3D coordinates of each point on the object surface are computed by analyzing the codes on the color image.

[a)]Electronic mail: sy@ieee.org

the color primitives of P to the matrix M so that there are no two identical words in the matrix and every element has the different colors with its adjacent neighbors in the word.

In this way, each defined location is unique and thus correspondence will be of no problem. That is, if the pattern is projected onto a scene, and the word value for an imaged point (u,v) is determined, then the corresponding position (i,j) in M of this imaged point is uniquely defined. Of course, in addition to having each word of M be unique, we also wish to optimize the color assignments so that matrix M is as large as possible.

According to perspective transformation principle, the image coordinates and the assigned code words of a spatial point are corresponded with its world coordinates. We can establish such a mapping relation between an image point and the spatial point. X, Y, and Z are the coordinates of a world point, corresponding with the image coordinates u, v and x, y. Together with the system calibration parameters, the 3D information of surface points can be easily computed. Effectively, it can guarantee that the measurement system has a limited cost of computation since it only needs to analyze a small part of the scene and identifies the coordinates by local image processing. Therefore the acquisition efficiency is greatly improved.

Consider the code generation scheme. First, with a given color set P, we try to make a longest horizontal code sequence, $S_h=[c_1,c_2,c_3,\ldots,c_m]$, in which each color primitive is different with its adjacent ones and any triplet of adjacent colors, $T_{h3i}=[c_i\ c_{i+1}\ c_{i+2}]$, is unique in the sequence. It can be proved that the maximal length of the horizontal sequence S_h is length $(S_h)=p(p-1)(p-1)+2$. This sequence can be generated by a searching algorithm. For example, given a set with four colors, the horizontal sequence can be $S_h(4)=[12421213132312343243413423214143142412]$

Second, with a given color set P, we try to make a longest vertical color sequence, $S_v=[c_1,c_2,c_3,\ldots,c_n]$. For any adjacent color pair, it is unique in the sequence. In fact, the maximal length of the vertical sequence S_v is length $(S_v)=p(p-1)+1$. Such a sequence can be generated in this way: $S_v=[121314,\ldots,1p,2324,\ldots,2p,\ldots,(p-1)p,1]$. For example, given a set with four colors, the vertical sequence can be $S_v(4)=[1213142324341]$

Finally, the matrix for color projection can be generated by the maximal vertical and horizontal code sequences (S_h and S_v). This makes a matrix with size of $(p-1)(p-2)+2$ by $p(p-1)^2+2$, which is the maximum possible for each code word being unique in the matrix. The first row in the matrix can be defined by S_h. A second row is created by adding the first element of S_v to each element of $S_h \bmod p$, where the mod operation is only on the set $\{1,2,\ldots,p\}$. Create a third row by adding the second element of S_v to each element of $S_h \bmod p$. In this way, for a four-color set the construction is an 8×38 matrix. If we define $S=S_h\otimes S_v$, it can be proved that each word in the matrix S is uniquely located.

This method formulizes the generation of a special code matrix which satisfies the condition of uniqueness. This generation scheme is a finite automata. To increase the matrix size so that the digital projector will project a light pattern with better resolution, we have to increase the color number. In our laboratory, a set with seven colors is often used, which can generate a matrix for a 32×212 rectangle.

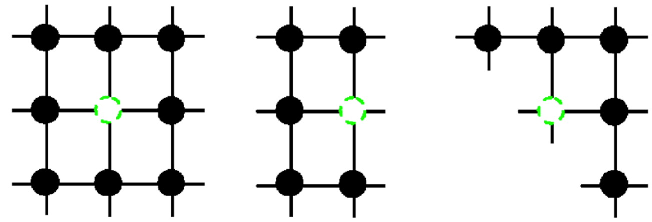

FIG. 2. (Color online) Cases of mesh amendment for holes (in which insertion operation is necessary).

Consider the image processing and 3D computation. One important step is to find a unique word (initial seed) in an unknown area of a captured image. It can be implemented in this way. Randomly generate a position in the image. The color at this position should not be black otherwise it should be regenerated. Find the square grid point at that position. A color similarity measurement is used to search a quadrangle in which colors are changing slightly compared with those outside. Based on this grid point, we try to locate its four adjacent neighbors. Simply set the offset to be the grid size estimated, the left, right, above, and nether points are initialized and the four square areas are determined. If this grid point is found not in the excepted position, or any one of the four neighbors failed to be located, another initial position should be generated. The coordinates of the seed word can be determined according to the five grid points by indexing the code word in the pattern matrix.

Then, a flood search algorithm[15,16] is used for word identification throughout the whole image. It firstly tries to search several grid points around the seed word, and then search more grid points near the known area. Each point to be added in the known partial net has to satisfy three conditions: its color, size, and regularity. Since it is a "one-pass" method, i.e., the pixels are computed only in a small local area once, the image processing can be performed very fast, promising real-time applications.

Finally, mesh amendment and grid interpolation procedures are developed for optimization of 3D results. The projection of the coded pattern should result in a regular mesh. Due to the complexity of the scene and uncertainty in image processing, the constructed grid matrix could have some faults (namely, holes and leaves). To correct these faults, this research develops a mesh amendment procedure to find and amend them, as illustrated in Figs. 2 and 3.

After all possible code words are identified from the image, it is now easy to compute the 3D world coordinates of these points according to the calibration matrices. It yields a rough 3D map of the scene. In order to improve the resolution, we may perform an interpolation algorithm on the map. Depending on application requirements, the interpolation

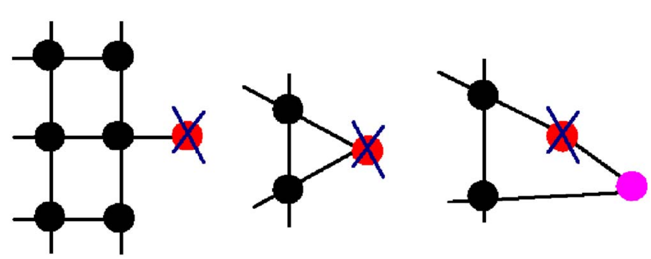

FIG. 3. (Color online) Cases of mesh amendment for leaves (in which deletion operation is necessary).

FIG. 4. (Color online) Testing system set up in the laboratory for implementation and evaluation of the proposed method.

may be only on the segment of two adjacent grid points or inside the square area formed by four regular grid points.

Practical implementation and performance analysis have been carried out in our vision system laboratory. In fact, the method has to be integrated with other techniques and algorithms for automating the modeling process, such as system calibration, image processing, 3D representation, and visualization. Thanks to many fundamental works developed in our early projects, the experimental system is not very difficult to be set up. The vision system (Fig. 4) is based on structured light principle set up with a projector and a camera. A 32×212 encoded pattern generated from a seven-color set is used to illuminate the scene. Images are captured by the camera and 3D meshes are reconstructed by the vision system after seed word initialization, flood search, mesh amendment, and correction.

To evaluate the execution performance and analyze the efficiency, this research used a Performance Analyzer to check the time spent in some important procedures. Results show that for 3D reconstruction in low-level or midlevel resolution (computing the 3D coordinates on grid points or grid edges), it takes about 10–100 ms. Such a speed is adequate for most engineering applications.

In summary, we have carried out an experimental investigation of a method for generating an encoded colored light pattern for real-time 3D data acquisition. The system combines both the advantages of stereovision for fast 3D reconstruction from a single image and structured light system for accuracy and stability. The method does not have a limit in the smoothness of object surface since it only needs analyzing a small part of the scene and identifies the coordinates by image local processing, which greatly improves the 3D measurement efficiency. The cost for setting up such a sensor is much lower than current 3D scanners. The limitation of the proposed method is that the sensor can usually only be used for measurement of objects with uniform or slightly changing colors and without much environment light disturbance. Otherwise, an adaptive color coding scheme should be applied, which is currently under consideration by the authors.

This work was supported by the Research Grants Council of Hong Kong (Project No. CityU117605) and the National Natural Science Foundation of China (NSFC-60405009, ZJNSF-Y105101, and Y104185), and a grant for Key Research Items from the Department of Science and Technology of Zhejiang Province (2006C21002). One of the authors (S.Y.C.) is a research fellow of the Alexander von Humboldt Foundation, Germany.

[1]C. Sinlapeecheewa and K. Takamasu, *Proceedings of the 2002 IEEE International Conference on Industrial Technology*, Bangkok, Thailand (IEEE, Piscataway, NJ, 2002), p. 405.
[2]H. C. Longuet-Higgins, Nature (London) **293**, 133 (1981).
[3]B. Browning and M. Veloso, *Proceedings of the 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Alberta, Canada (IEEE, Piscataway, NJ, 2005), p. 3871.
[4]Y. F. Li and S. Y. Chen, IEEE Trans. Rob. Autom. **19**, 259 (2003).
[5]G. J. Zhang and Z. Z. Wei, J. Manuf. Sci. Eng. **125**, 617 (2003).
[6]J.-P. Tardif and S. Roy, *Fifth International Conference on 3-D Digital Imaging and Modeling*, Ontario, Canada (IEEE Computer Society, Ottawa, Ontario, Canada, 2005), p. 22.
[7]T. P. Koninckx, *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Diego, CA (IEEE, San Diego, CA, 2005), Vol. 2, p. 611.
[8]F. Tsalakanidou, Real-Time Imag. **11**, 358 (2005).
[9]S. Osawa, Measurement **26**, 157 (1999).
[10]L. Zhang, B. Curless, and S. M. Seitz, *IEEE Proceedings on 3D Data Processing Visualization and Transmission*, Padova, Italy (IEEE Computer Society, 2002), p. 24.
[11]J. Batlle and E. Mouaddiba, Pattern Recogn. **31**, 963 (1998).
[12]P. M. Griffn and L. S. Narasimhan, Pattern Recogn. **25**, 609 (1992).
[13]S. Y. Chen and Y. F. Li, Meas. Sci. Technol. **14**, 33 (2003).
[14]J. Salvi, J. Pags, and J. Batlle, Pattern Recogn. **37**, 827 (2004).
[15]A. Treuenfels, C/C++Users Journal **12**, 8 (1994).
[16]A. Glassner, IEEE Comput. Graphics Appl. **21**, 78 (2001).