Open Journal of Applied Sciences, **2013, 3, 22-26**
Published Online March 2013 (http://www.scirp.org/journal/ojapps)

Regular Stereo Matching Improvement System Based on Kinect-supporting Mechanism

Din-Yuen Chan, Che-Han Hsu
Department of Computer Science and Information Engineering, National Chiayi University, Chiayi City, Chinese Taipei
Email: {dychan, s1000436} @mail.ncyu.edu.tw

Received 2012

ABSTRACT

In this paper, we built a stereoscopic video associated experimental model, which is referenced as Kinect-supporting improved stereo matching scheme. As the depth maps offered by the Kinect IR-projector are resolution-inadequate, noisy, distance-limited, unstable, and material-sensitive, the appropriated de-noising, stabilization and filtering are first performed for retrieving useful IR-projector depths. The disparities are linearly computed from the refined IR-projector depths to provide specifically referable disparity resources. By exploiting these resources with sufficiency, the proposed mechanism can lead to great enhancement on both speed and accuracy of stereo matching processing to offer better extra virtual view generation and the possibility of price-popularized IR-projector embedded stereoscopic camera.

Keywords: Stereo Matching; Image Registration; Kinect; IR-Projector Image

1. Introduction

Multi-view video systemization is the most emerging subject for bare eyes stereo vision. The camera array is utilized for the promising implementation in multi-view manufacture [1]. Generally speaking, the camera array applied to multi-view video photography is not suited for dynamic setups. The pre-processing filter is specified to modify the depth map for sustaining fewer holes in synthesizing another view [2]. If the filter is only designed for holes reduction [2], the side effect such as bending distortion [3] is hardly inevitable. Moreover, the one view with single-channel depth using DIBR is rather difficult to obtain adequate view angle extensions. With low-labor calibration, the binocular 3D-camera is acknowledged by proper equipment to straightforwardly capture the stereoscope two-view videos. However, when the depth decision fully leans on the stereo matching of paired color images, the suited depth acquisition of sparse-texture is quite difficult. In general, the IR-sensor emitting infrared (IR) for stably acquiring depths is usually expensive [4], and cannot completely sense materials being IR-detection unavailable. Therefore, a more valuable approach shall attempt to implement a common low-cost IR-sensor into an efficient auxiliary apparatus for generating high-quality depths. The Kinect IR-projector is not really considered to be useful for depth measurement due to its low spatial/depth resolution, distance-limited sensing and unstable depth capture, difficultly of detecting specular, transparent, and reflective objects. However, Kinect has become a very popular, low-price off-the-shelf depth detector nowadays; the stereoscope investigation based on Kinect begins emerging [5].

In [5], the work first performs the stereo matching between Kinect's IR-image and RGB image to generate a depth map. The depth map and the inner depth map are then computed using Kinect IR-projector captured patterns are fused for accomplishing a more qualified depth map. Because the captures of IR and RGB images by Kinect are in turn active, rather than synchronous, the mechanism [5] seems efficient, but at most only suited for monotonous, low-activity videos.

Theoretically, by adding an IR sensor, the handling of stereo matching problem shall become much easier. In effect, to effectively integrate two hetero images of large quality difference, Kinect IR-projector image and 3D-camera image is quite a challenge task. Therefore, our work is to identify a valuable approach of developing an inexpensive IR-sensor embedded 3D-camera for facilitating the multi-view manufacture can be obtained. Because the high-resolution depth maps can be easily yielded by operating stereo matching on the captured two views of 3D-camera, the depths captured by Kinect IR-projector only play a reference/consultation role rather than an arbiter in stereo matching operations.

Our strategy leaves the IR-projector depth purely as a stereo matching indicator, so some IR-projector problems need not to be taken care of any more. And fortunately, the add-on charge coupled device (CCD) camera in Ki-

nect can offer the necessity of inter-medium for difficult image registration from 3D camera color images to concurrently captured IR-projector depth images.

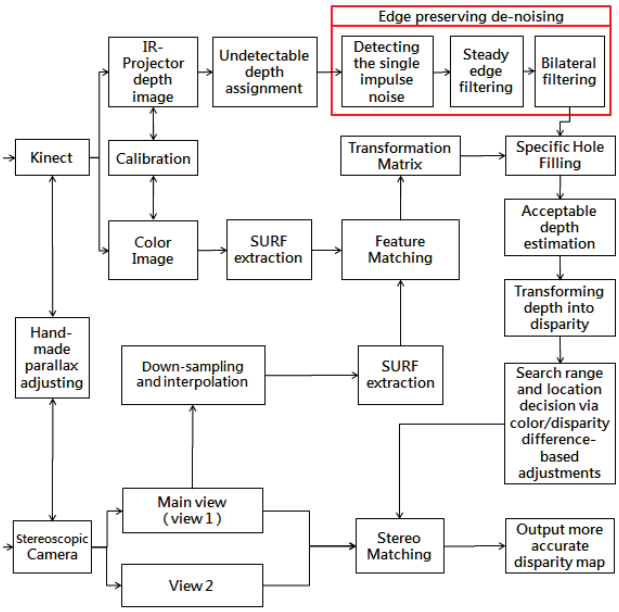


Figure 1. Processing flow of the proposed system.

The proposed system contains a series of appropriated refinement processing on the IR-projector image, the registration from the main view image to the IR-projector depth image, and the stereo matching improvement supported by the refined IR-projector image, as shown by Figure 1. The computation of homographic transform matrixes for the registration between Kinect and 3D-camera images is the most complex part in the proposed system, but only performed once at the initialization of depth generation of a 3D video sequence.

2. Cross Geometrical Image Relation Identification

The most troublesome barrier is the large content differences between 3D-camera color image and IR-projector depth image in their registration. Since the Kinect add-on color image can be easily calibrated to the IR-projector depth image synchronically captured, it is a good intermediate interface for the following registration. Specifically, the mapping for image registration will run from 3D-camera main view image to Kinect add-on color image, and then to IR-projector depth image. For well mutual images registration, the geometrical image relation identification need to be performed in advance so it includes the image localization (partition) by analyzing feature points distribution and the global transition registration by selecting a representative registration feature point based on the precedence of center location and salient response strength.

Generally speaking, the 3D camera and the Kinect sensor are not difficult to obtain adequate parallelization of photographing planes via hand-made adjustment for constrain their perspective difference as much as possible. Therefore, the remaining registration task is to find the relative transition and rotation between two frames captured by them. The implementation of relating the global center registration between 3D camera Image and Kinect Color Image is depicted as below.

For simplifying the registration of two images with large-scale resolution difference, the faster way is to pull down the resolution of high-resolution image to compromise low-resolution one. Therefore, in the study, the main-view color image of 3D camera (abbreviated as MCI) is made large-scale down-sampling and then small-scale linear interpolation to attain same low horizontal/vertical resolutions as Kinect RGB image. Before computing the image registration (homographic transform) matrix, the relative global location difference between the resolution-reduced main-view color image (RMCI) and the Kinect color image (KCI) need be estimated in advance. Such a location difference need be shift back for the translation of two targeted images during performing their registration. In this work, a center zone in RMCI is allocated to extract several speeded up robust features (SURFs), of which their response strengths can exceed a given threshold, as the candidate anchor nodes. The simultaneously sampled KCI is also extracted at the same number of SURFs as its candidate anchor nodes by the same way operated on RMCI. A double checking process is then utilized to identify the trustful reference points between two images for making the necessary global transition at the beginning of registration. The process tests the anchor nodes one by one from the place nearest the mask center to those around the mask borders in the clockwise or counter-clockwise direction by two phases shown as follows.

- Phase 1 < Collecting a couple of qualified anchor nodes >
- Step 1. Check if two qualified anchor nodes (QANs) have been acquired in the main view color image, and their corresponding feature nodes matched in the simultaneous Kinect color image. If it does, go to Step 5 (the second phase). Otherwise, go to the next step.
- Step 2. Select a candidate anchor node on the main view image as the new test one. When a previous QAN exists, the selection shall be set a shortest distance from the last qualified anchor points (QAP).
- Step 3. Compute the difference between the characteristic vectors of new test node and its corresponding feature node best-matched in Kinect color image.
- Step 4. If the difference is less than threshold, the test anchor point is denoted as a new QAN go back to Step 1. Otherwise, bypass this test node and then go

back to Step 2.

- Phase 2 < Fast Checking of geometrical similarity >
- Step 5. Line two QANs to get a straight segmental line in the main view color image, so does their matched feature nodes to get another segmental line in the simultaneous Kinect color image.
- Step 6. If the two segmental lines have similar slopes and length, both after normalizing them by the raster scanning ratios and the resolutions of these two registration-targeted images, then the QAN having the lower characteristic-vector difference among two QANs is identified as the trustful reference point, and stop the routine. Otherwise, give up the QAN of larger characteristic vectors, and then go back to Step 2.

The scheme will set the normalized coordinate difference vector between trustful reference point and its corresponding feature node as the global transition vector for the main view color image and the simultaneous Kinect color image.

3. Undetectable Depth Assignment and Edge Preserving De-noising

For better promoting the utilization confidence of IR-projector depth image, IR-projector depth image after undetectable depth assignment and edge preserving de-noising will be further refined by three proposed processes. The first step of the process is to detect the single impulse noise in the IR-projector depth image by subtracting the strength of each pixel (sensed depth value) and the strengths of its 8 neighbors. If all the 8 subtraction outcomes of a pixel are larger than a threshold, this pixel is regarded as single impulse noise. The strength of this pixel is then replaced by the mean of the 8 neighboring strengths for removing its impact on the subsequent processes, the processed image is called the impulse-noise dropped IR-projector depth-image (IDIRI). The second process is for marking the so-called steady edges. Via performing the Sobel-filter filtering on IDIRI, each edge point is examined whether its filtered intensity is similar to anyone of its 8-neighboring filtered intensities or not. If it does, this point is denoted as a steady edge point. For simply enhancing the effect of edge preserving via bilateral filtering, the procedure of bilateral filtering proposed herein will skip over the steady edge points. Except the steady edge points, of which set is grouped as set G_{ssp}, the remaining points of IDIRI are performed by an appropriated bilateral filter with adaptive piece-wise mask. Through the bilateral filtering of window size of $(2L+1)\times(2L+1)$, The filtered intensity at (x, y), denoted by $I_{SIR}(x, y)$, is given by

$$I_{SIR}(x,y) = \frac{1}{\sum \omega_{x+i,y+j}} \sum_{i=-L}^{L} \sum_{j=-L}^{L} \omega_{x+i,y+j} \cdot I_{IR}(x+i,y+j) \; for (x,y) \in G_{IDIRI} - G_{ssp}. \quad (1)$$

where $I_{IR}(x+i, y+j)$ is the pixel intensity at point (x+i, y+j) on IDIRI, of which pixels set is G_{IDRI}, and its filter weight is $\omega_{x+i, y+j}$. In (1), $\omega_{x+i, y+j}$ has two ingredients multiplied together :

$$\omega_{x+i,y+j} = 2^{-(\alpha_{i,j}-1)} \cdot u\left(\beta_{i,j}\right). \quad (2)$$

for bilateral effects on the spatial distance and the intensity difference, where u(.) is the unit step function. The spatial weighting function $2^{-(\alpha_{i,j}-1)}$ is piece- wise, where

$$\alpha_{i,j} = \max\left(|i|, |j|, 1\right). \quad (3)$$

the intensity weighting function $u(\beta_{i, j})$ exploits u(.) for displaying the characteristic of a bi-level valued mask, where

$$\beta_{i,j} = T_{IR} - \left|I_{IR}(x,y) - I_{IR}(x+i,y+j)\right|. \quad (4)$$

The bilateral filter applies several window sizes, where L and its maximum, denoted by L_{max}, are proportional to the quantized $I_{IR}(x, y)$ and the standard deviation of quantized $I_{IR}(x, y)$, respectively, on the impulse-noise dropped IR depth image. More specifically, the point closer to the camera (with the larger depth value) will be filtered by the bigger filtering window. The larger standard deviation of quantized $I_{IR}(x, y)$ will cause the smaller L_{max} applied.

4. IR-Projector Supported Stereo Matching Improvement

The candidate anchor nodes of paired RMCI and KCI are treated as targets of Random Sample Consensus (RANSAC) processing to obtain the registration matrix for the registration achievement between the pixels of KCI and that of MCI.

4.1. Adjustment of Referred Search Location and Search Region

The computed by Kinect IR-projector detected depth can offer good and fast search references in the search of two views stereo matching. Further adjustment is necessary, if two adjacent pixels having close chrominance and luminance but quite different search reference positions referred by the IR-projector image. Such a case usually happens nearby the borders of two occluded objects. In this case, the suspected search reference location shall be replaced by a higher confidence one, but in practice, the definition or measurement of confidence is not easy. For the views stereo matching from left to right, the computation of original search reference position at (x, y) is given by

$$p_{S\_ref}(x,y) = \Gamma\left(I_{IR}(x',y')\right) - x. \quad (5)$$

where $\Gamma(I_{IR}(x', y'))$ is to linearly transform the depth value $I_{IR}(x', y')$ at (x', y') in the IR-projector image to the

disparity of pixel at (x, y) in the left view image that (x', y') is registered to (x, y) by the obtained registration matrix. The search range centered at $P_{S\_ref}(x, y)$ is then set as

$$[p_{S\_ref}(x,y) - \chi(I_{IR}(x',y')), p_{S\_ref}(x,y) + \chi(I_{IR}(x',y'))] . \quad (6)$$

to find an adequately matched pixel on the right view image that the searching offset $\chi(I_{IR}(x', y'))$ is a variable according to $I_{IR}(x', y')$. In (5) and (6), $\Gamma(.)$ and $\chi(.)$ are mainly relevant to the display screen parameters and the dynamic range of IR-projector captured depths.

4.2. Pixel Extrapolation Outsides the IR-Projector Image Region

When the registered location of referred point in the IR-projector image exceeds the image region, the effective extrapolation is necessary for figuring out the depth value in that location. Since the pixel extrapolation can be considered as the extension of image size, therefore a continuity-preference predication is able to be addressed to extend the IR-projector image. The registered pixel exceeding yet still contacts the border of current extended IR-projector image will have three (or two) neighbors, which has the depth values (IR-projector detected intensities), among its 8-nrighboring locations. Assume the pixel position is (x', y'). The connection straight line among the other three connection straight line radiated from (x, y) has the lowest depth change is selected such that the difference of successive two pixels on it will be adopted to extrapolate the pixel intensity at (x, y). The continuity-preference predication to predict the pixel intensity at (x, y), denoted by $\tilde{I}_{IR}(x, y)$, is formularized by

$$\tilde{I}_{IR}(x,y) = I_{IR}(x+\Delta x, y+\Delta y) + \left(I_{IR}(x+\Delta x, y+\Delta y) - I_{IR}(x+2\Delta x, y+2\Delta y)\right) . \quad (7)$$

where ($\Delta x = 1$, $\Delta y = \gamma$), ($\Delta x = -1$, $\Delta y = -\gamma$), ($\Delta x = \gamma^{-1}$, $\Delta y = 1$) and ($\Delta x = -\gamma^{-1}$, $\Delta y = -1$) are set for extrapolating (extending) the exterior pixel along the left border, the right border, the bottom, and the top of IR-projector image, respectively. Parameter γ as the slope of e continuity preference is given by

$$\gamma = Arg\left(\underset{\rho \in \phi}{Min} \left| I_{IR}(x+\Delta x, y+\Delta y) - I_{IR}(x+2\Delta x, y+2\Delta y) \right| \right) . \quad (8)$$

where ρ expresses the slope of two-point straight line and the set of ρ's is φ, which equals to {−1, 0, 1} for extending the image outward its left/right border, and {−1, ∞, 1} outward its bottom/top border.

4.3. False-Reduced Modification for Stereo Matching Search References

After applying the homographic transform matrix to map coordinates from IR-projector image to the main view image, the necessary adjustment of original search locations will be set for rational relations rather than accurate status from the point of geometrical view. The proposed method is following and exploiting the raster-scan processing order for prompt adjustment that its procedure is depicted as follows. For the non-leftmost pixels at ($x \neq 0$, y)'s, if the criterion below is satisfied:

$$\left|\mathbf{C}(x,y) - \mathbf{C}(x-1,y)\right| < \delta_C \text{ and } \left|p_{S\_refl}(x,y) - p_{iS\_ref}(x-1,y)\right| > \delta_S . \quad (9)$$

then $P_{S\_ref}(x, y)$ is replaced by $P_{S\_ref}(x-1, y)$ that C(x, y) is the color vector at (x, y), δ_C and δ_S are empirical thresholds. Similarly, for the leftmost pixels at (0, y)'s, if the criterion given by

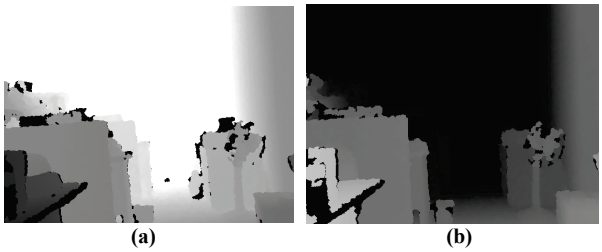
$$\left|\mathbf{C}(x,y) - \mathbf{C}(x,y-1)\right| < \delta_C \text{ and } \left|p_{S\_refl}(x,y) - p_{S\_ref}(x,y-1)\right| > \delta_S . \quad (10)$$

holds, then $P_{iS\_ref}(x, y-1)$ substitutes for $P_{S\_ref}(x, y)$.

For alleviating the miss-matching error resulting from unsuitable modification in suspected $P_{S\_ref}(x, y)$, the search region for the pixel at (x, y) is enlarged for making a protective compensation in stereo matching. The adopted straightforward way is to add a fraction of δ_S to the search distance.

4.4. Stereo Matching Acceleration by IR-Projector Depth Image

For facilitating the mapping between refined IR-projector depths and stereo matching depths as well as removing wrong or inappropriate differences in flat zones, the above refined IR-projector depth is quantized in advance. Then, the level mapping relation, which is associated with dynamic regions registration and one-on-one statistic observation, between the stereo matching depth and the above-refined IR-projector depth can be statistically estimated. Its formula is treated as a mutual mapping function of heterogeneous depths. Thus, via the heterogeneous mapping of refined IR-projector depth to 3D camera dual images, the initial search coordinate can be obtained to improve both of speed and accuracy of the stereo matching for the depth computation in testing point. This is suited for all of existing stereo matching algorithms including supporting weight [6], cross-based and census transform ones [7].



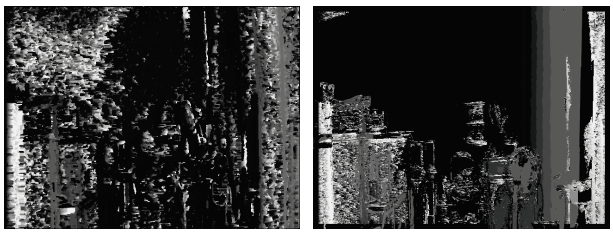
(a) **(b)**

Figure 2. (a) Original IR-projector image (b) Refined result.

5. Simulation Results

In our experiments, the stereo matching adopts AD-Census method [7] to find the best matched point in right view images from the left (main) view ones. Figure 2 displays an original IR-projector image and its refined result with the proposed processes, the refinement speed of IR- projector can obtain to 30 frames per second. It demonstrates that the refined IR-projector images are quite stable that various noise, photography artifacts, and IR-detection unavailable parts causing IR-sensed image holes can be removed. In Figure 3, the stereo matching outcomes without and with the proposed Kinect-supporting improvement are compared. By the Kinect-supporting improvement, the stereo matching accuracy can be raised especially for the fatness or sparse-texture parts.

The Kinect-supporting stereo matching against original stereo matching speed-up is 34.08%, and the frame-by-frame computational overheads acquiring the referred search points are counted in the former except for homographic transform matrixes generating, which belongs to the system-setup initialization.



(a) (b)

Figure 2. Depth Maps of Stereo matching:(a) without Kinect supporting (b) with Kinect supporting

6. Conclusion

In this study, a Kinect-supporting mechanism with regular structure is proposed for efficiently improving the stereo matching processing. Through proposed different-resolution hetero-image registration, the disparities linearly computed from those refined IR-projector depths are applied to the main view color image of 3D camera as disparity searching references. By concisely exploiting the disparity reference resources, the proposed scheme can lead to the effectiveness of promotion for the accuracy and speed of stereo matching. This investigation indicates that developing a low-cost IR-sensor embedded 3D-camera, by which the multi-view video beyond five views generation can be manufactured rapidly as soon as users (or artist) shoots a two-view video sequence.

7. Acknowledgements

The author would like to thank the fund support by NSC 101-2221-E-415-020- and 101-EC-17-A-02-S1-201.

REFERENCES

[1] Y. Taguchi, T. Koike, K. Takahashi, T. Naemura, "TransCAIP: A Live 3D TV System Using a Camera Array and an Integral Photography Display with Interactive Control of Viewing Parameters," IEEE Transactions on Visualization and Computer Graphics, vol.15, no.5, pp.841-852, Sep. 2009.

[2] W. J. Tam, G. Alain, L. Zhang, T. Martin, R. Renaud, "Smoothing depth maps for improved stereoscopic image quality," Three-Dimensional TV, Video, and Display III(Proceedings of the SPIE), Vol. 5599, pp. 162-172,2004.

[3] L. Zhang, W.J. Tam, "Stereoscopic image generation based on depth images for 3D TV," IEEE Transactions on Broadcasting, vol.51, no.2, pp.191-199, Jun. 2005.

[4] J. Zhu, L. Wang, R. Yang, J.E. Davis, Z. Pan, "Reliability Fusion of Time-of-Flight Depth and Stereo Geometry for High Quality Depth Maps," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.33, no.7, pp.1400-1414, Jul. 2011.

[5] W.C. Chiu, U. Blanke, M. Fritz, "Improving the Kinect by Cross-Modal Stereo," The 22nd British Machine Vision Conference (BMVC 2011), pp.116.1-116.10, Sep. 2010.

[6] K.J. Yoon and I.S. Kweon, "Locally adaptive support-weight approach for visual correspondence search," Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp.924-931, 2005.

[7] X. Mei, X. Sun, M. Zhou, S. Jiao, H. Wang and X. Zhang, "On Building an Accurate Stereo Matching System on Graphics Hardware," GPUCV'11: ICCV Workshop on GPU in Computer Vision Applications, 2011

