

ارائه روشی نوین در تلفیق ابرنقاط حاصل از سنجنده Kinect

شهناز نخبه زعیم¹، محمد سعادت سرشت²، علیرضا صفدری نژاد³

1. کارشناس ارشد مهندسی فتوگرامتری، گروه مهندسی نقشه برداری، پردیس دانشکده‌های فنی، دانشگاه تهران

shahnaz_zaeim@ut.ac.ir

2. استادیار گروه فتوگرامتری دانشکده‌های فنی دانشگاه تهران، پردیس دانشکده‌های فنی، دانشگاه تهران

msaadat@ut.ac.ir

3. دانشجوی دکترای مهندسی عمران - نقشه‌برداری، دانشگاه صنعتی خواجه نصیرالدین طوسی

Safdari_nezhad@sina.kntu.ac.ir

چکیده

امروزه روشهای تولید آئی ابرنقطه سه‌بعدی با استفاده از ابزاری همچون دوربینهای ویدئویی استریو یا دستگاه Kinect تبدیل به واقعیتی عملی شده است که قادر است حجم بالایی از ابرنقاط رنگی را در زمان کوتاهی از محیط اخذ کند. البته تولید یک مدل سه‌بعدی کامل و باکیفیت با محدودیت‌های جدی مواجه می‌باشد که برای غلبه بر این محدودیتها، الزاماً اخذ داده‌های سه‌بعدی از مکانها و زوایای مختلف صورت می‌گیرد. از این‌رو، تعیین دقیق پارامترهای مربوط به موقعیت و وضعیت کینکت نقش کلیدی را در هم‌مرجع‌سازی ابرنقاط برای تولید مدل‌های یکپارچه ایفا می‌کند. از سوی دیگر، به دلیل ماهیت سنجنده کینکت در جمع‌آوری داده از محیط، با گذشت زمان کوتاهی حجم مشاهدات اخذ شده به شدت افزایش می‌یابد که کندی پردازش‌های پیش‌رو را فراهم می‌آورد. برای جلوگیری از این موارد، بایستی ابرنقاط سه‌بعدی در هر فریم باهم تلفیق و یکپارچه‌سازی شود. از این‌رو در مقاله حاضر، الگوریتم جدیدی به منظور خلاصه‌سازی و تلفیق ابرنقاط حاصل از سنجنده، ارائه گردید.

واژه‌های کلیدی: کینکت، داده سه‌بعدی، هم‌مرجع‌سازی، تلفیق

1- مقدمه

همزمان با توسعه سیستم‌های نقشه‌برداری همراه¹، نیاز به تولید داده‌های مکانی پرسرعت و کم‌هزینه در زمینه‌ی مدل‌سازی سه‌بعدی محیط افزایش یافته است. تسهیل در روند بازسازی سه‌بعدی، افزایش سطح اتوماسیون در فرآیند زمین‌مرجع‌سازی و کاهش نیاز به نقاط کنترل زمینی از موارد مهمی است که در این حوزه به‌عنوان موضوعات تحقیقاتی جاری قلمداد می‌شود [1]. تهیه مدل‌های سه‌بعدی از محیط‌های داخلی پیچیده که با محدودیت بکارگیری روش‌های سنتی مانند لیزر اسکنرهای زمینی روبرو است، سبب ورود روش‌های جدیدی مبتنی بر تکنیک‌های فتوگرامتری نور ساختاریافته را فراهم آورده است. سیستم‌های نور ساختاریافته‌ای مثل Mantis از شرکت Qubic و Kinect از شرکت Microsoft را می‌توان به‌عنوان نمونه‌هایی

1- Mobile Mapping Systems

از ابزارهای نوین در این حوزه برشمرد [2]، [3]. ابزارهای بیان شده قادرند به کمک روش‌های تقاطع فضایی از داده‌های تصویری استریو، حجم زیادی از موقعیت نقاط سه‌بعدی مقابل با میدان دید لحظه‌ای سنجنده را با نرخ بالایی اندازه‌گیری نمایند. از سوی دیگر، پوشش کامل محیط به‌وسیله‌ی سنجنده معمولاً نیازمند جمع‌آوری داده از نقاط دید چندگانه می‌باشد [4] که این امر به دلیل وجود پنهان‌شدگی اشیاء در راستای خط دید² سنجنده، عدم پوشش سراسری محیط به‌واسطه‌ی میدان دید محدود و تأثیر تغییرات بازتابندگی اشیاء موجود در محیط، می‌باشد [5]. از آنجایی که جهت مدلسازی سطح شی می‌بایست ابرنقاط یکپارچه‌ای از شی حاصل شود، بنابراین تمام ابرنقاط حاصل از سنجنده باید به یک سیستم مختصات واحد منتقل گردند و به‌عبارت جامع‌تر؛ ابرنقاط به‌دست آمده باید هم‌مرجع³ شوند [6]. از سوی دیگر، ایجاد ابرنقطه از محیط‌های داخلی با استفاده از یک دوربین عمقی مانند کینکت، بدون اعمال هرگونه ساده‌سازی، به ایجاد نقاط اضافی زیادی منجر می‌شود که با افزایش حجم مشاهدات با گذشت زمان کوتاهی، حافظه‌ی موردنیاز و زمان محاسبات به‌شدت افزایش یافته که باعث کندی سرعت پردازش می‌شود. برای جلوگیری از این موارد، بایستی ابرنقاط سه‌بعدی در هر فریم با هم تلفیق شود. براین اساس؛ ارائه الگوریتمی کارآمد در جهت جلوگیری از افزایش تصاعدی داده‌های ابرنقطه از طریق تلفیق ابرنقاط هم‌مرجع شده فریم‌های متوالی ضروری به‌نظر می‌رسد، که در نتیجه ابرنقطه یکپارچه و باکیفیت بالا (دقت هندسی بهتر و نواحی پنهان کمتر) تولید می‌شود [7]، [8]، [9].

در بخش حاضر، مقدمه‌ای در باب اهمیت و کاربرد مدلسازی سه‌بعدی، تصویربرداری عمق به‌عنوان یک روش نوظهور در زمینه‌ی جمع‌آوری داده‌های سه‌بعدی و در ادامه با تشریح مشکلات پیش‌رو در فرآیند تولید مدل‌های پیوسته سه‌بعدی، به بیان مسأله و ضرورت این تحقیق اشاره شده است. بخش دوم از مقاله‌ی حاضر به معرفی اجمالی سنجنده‌ی Kinect و ساختار هندسی اخذ داده در آن اختصاص داشته است. در بخش سوم روش پیشنهادی برای تلفیق ابرنقاط تشریح شده و در بخش چهارم داده‌های مورد استفاده در این تحقیق معرفی شده و در ادامه، پیاده‌سازی و نتایج حاصل از روش پیشنهادی ارائه شده است. جمع‌بندی و نتیجه‌گیری نیز در بخش پنجم بیان شده است.

2- ساختار سنجنده‌ی Kinect و نحوه‌ی اخذ داده در آن

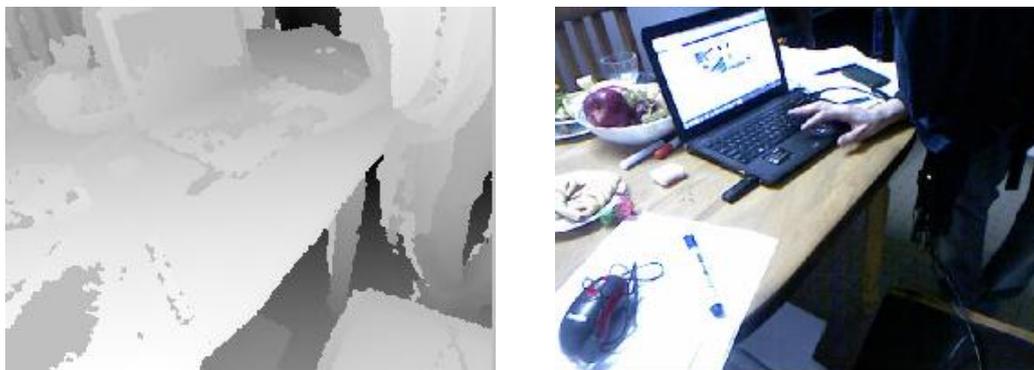
یکی از جدیدترین سیستم‌های تجاری در زمینه‌ی جمع‌آوری داده‌های سه‌بعدی، که جایگزین بسیار مناسبی برای اسکنرهای لیزری گران‌قیمت بوده، سنجنده‌ی Kinect است که توسط شرکت Microsoft در سال 2010 به بازار ارائه شده است. ابتدا Kinect جهت بازی‌های رایانه‌ای طراحی شد. به‌هرحال، مشخصه‌های اطلاعات گرفته شده توسط Kinect توجه محققان را در رشته‌ی نقشه‌برداری و مدلسازی سه‌بعدی، جلب نموده است. نمایش اخیر از پتانسیل‌های Kinect در مدلسازی سه‌بعدی در محیط‌های داخلی می‌تواند در کارهای هنری و همکارانش دیده شود [10]، [11]، [12]، [13]. این ابزار به جهت دارا بودن سیستم تشخیص سریع وضعیت انسان، بسیار مورد استقبال قرار گرفته و در صدر سنجنده‌های سه‌بعدی توسعه یافته است. ویژگی‌های منحصر‌بفرد سنجنده‌ی کینکت همانند: قیمت پایین سنجنده، نرخ بالا در نمونه‌برداری (30 Hz or 30 fps)⁴ و به‌دنبال آن افزونگی داده‌ی بسیار بالا، محاسبه‌ی خودکار تصویر عمق، قابلیت به‌کارگیری در شرایط

2- Line of Sight

3- Registration

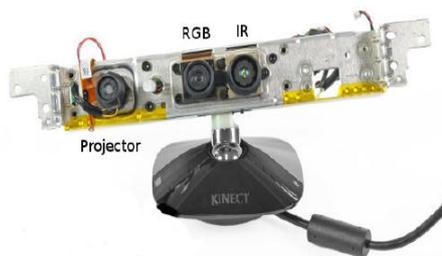
4- Frame Per Second

مختلف نوردهی و همچنین امکان استفاده در محیط‌های آزمایشگاهی و غیر آزمایشگاهی، انگیزه‌های استفاده از این ابزار نو و جذاب را در روند مدلسازی سه‌بعدی فراهم آورده است. Kinect یک سنجنده فعال است که خروجی لحظه‌ای این سنجنده، دو تصویر RGB و عمق در ابعاد 480×640 پیکسل می‌باشد. ادغام اطلاعات عمق و رنگ، منجر به ایجاد یک ابر نقطه رنگی سه‌بعدی می‌شود که حاوی حدود سیصد هزار نقطه در هر فریم است.



شکل 1. یک فریم از تصویر RGB (راست) و عمق (چپ) اخذ شده توسط Kinect

مناطق روشن در تصویر سمت راست از شکل بالا (ماکزیمم درجه خاکستری)، محل‌هایست که به دلایل مختلفی مثل Occlusion، فاصله زیاد، جنس مواد موجود در صحنه و زاویه سطح نسبت به پرتوهای تابیده شده با مشکلات تناظریابی همراه بوده‌اند؛ از طرفی درجات خاکستری در سایر مناطق موجود در صحنه، گویای دوری و نزدیکی اشیاء نسبت به سنجنده می‌باشند. این سنجنده در طراحی خود متشکل از سه بخش اصلی می‌باشد:



شکل 2: اجزاء سنجنده‌ی کینکت [14]

الف) پروژکتور در محدوده‌ی طیفی مادون قرمز نزدیک. این پروژکتور به‌طور پیوسته الگوی ثابتی از نقاط نورانی را به سمت صحنه‌ی روبروی خود می‌تاباند.

ب) سنجنده‌ی حساس به محدوده‌ی مادون قرمز نزدیک که به‌عنوان یک دوربین عمقی کار می‌کند و با هدف ثبت الگوی تابیده شده توسط پروژکتور طراحی شده است.

پ) دوربین مرئی و رنگی، جهت تشخیص محتوا و بافت تصویر که همزمان تصویر True Color از محیط اخذ می‌کند. در این پژوهش، از پتانسیل اخذ تصاویر RGB همزمان با تصاویر عمق در سنجنده‌ی Kinect به‌عنوان یک منبع اطلاعات رادیومتریکی از صحنه‌ی تصویربرداری بهره گرفته شده است.

3- روش پیشنهادی جهت تلفیق ابر نقاط

بدیهی است به هنگام ایجاد ابرنقطه از محیط‌های داخلی با استفاده از کینکت، اگر جابجایی زیادی در موقعیت سنسور وجود نداشته باشد، در یک بازه زمانی اندک، حجم زیاد نقاط کمکی به افزایش دقت نداشته و افزونگی داده، حافظه‌ی موردنیاز و زمان محاسبات را به شدت افزایش می‌دهد که این امر موجبات کندی سرعت پردازش را فراهم می‌آورد. از طرف دیگر، با افزایش حجم مشاهدات، هر یک از مشاهدات ابرنقطه، به تنهایی دارای دقت پایین و نواحی پنهان قابل توجهی است. برای جلوگیری از این موارد، بایستی ابرنقاط سه-بعدی فریم‌های متوالی با هم تلفیق شود. بدین منظور، به جای حجم عظیم داده‌ها، تعدادی از نقاط به صورت نمونه انتخاب و به عنوان نماینده سطح معرفی می‌شوند. بر همین اساس، میانگین‌گیری ملاک عمل قرار داده می‌شود. ذکر این نکته ضروریست که برخلاف لیزر اسکنر که دقت تمام نقاط برداشتی تقریباً یکسان می‌باشد، این فرض در مورد داده‌های کینکت ساده انگارانه است و اعمال این فرض موجب می‌شود که دقت واقعی نقاط در طی پروسه در نظر گرفته نشود. بدیهی است در هر گرید، دقت و چگالی نقاط سه‌بعدی بدست آمده از تصاویر عمق وابسته به میزان فاصله‌ی عوارض تا سنجنده بوده، و این عامل موجب شده که اگر نقطه‌ای از فاصله‌ی بسیار دور عکسبرداری شده باشد، چون به تبع دارای دقت پایینی می‌باشد، تأثیر کمتری نیز در برآورد نقطه‌ی نماینده خواهد داشت و برعکس نقاط نزدیکتر دارای دقت و اثرگذاری بالاتری در انتخاب نماینده خواهند بود. بر این اساس با توجه به رزولوشن موردنیاز کاربر، یک شبکه منظم سه‌بعدی مجازی روی ابرنقاط ایجاد شده و نقاط قرار گرفته در هر سلول متوسط‌گیری وزن‌دار می‌شوند، که وزن نقاط بر اساس کوواریانس دقت نقاط در نظر گرفته شده است. نکته حائز اهمیت دیگر توجه به این موضوع می‌باشد که دقت نقاط را نیز نمی‌توان صرفاً بر اساس فاصله در نظر گرفت. به عبارت دیگر دقت باید در سه مؤلفه x , y , z در نظر گرفته شود. باید توجه شود که نقطه‌ای ممکن است در فاصله‌ی خیلی دوری از سنجنده قرار داشته باشد اما طبق قانون انتشار خطا بیشترین خطای موجود در آن نقطه در یک مؤلفه نهفته باشد. بر این اساس، از آنجایی که مطابق معادله (1)، نقاط زمینی تابعی از مشاهدات عکسی و درجه خاکستری می‌باشند، یعنی:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = F(x, y, D_N) \quad (1)$$

به عبارت دیگر $X = F(x)$ ؛ با استفاده از رابطه‌ی زیر، دقت تک تک نقاط زمینی قابل محاسبه خواهد بود.

$$C_X = \begin{bmatrix} \frac{\partial F}{\partial x} & \frac{\partial F}{\partial y} & \frac{\partial F}{\partial D_N} \end{bmatrix} * C_x * \begin{bmatrix} \frac{\partial F}{\partial x} \\ \frac{\partial F}{\partial y} \\ \frac{\partial F}{\partial D_N} \end{bmatrix} \quad (2)$$

ماتریس واریانس کوواریانس نقاط زمینی: C_X

ماتریس واریانس کوواریانس مشاهدات عکسی: C_x

$$C_x = \begin{bmatrix} \delta_x & 0 & 0 \\ 0 & \delta_y & 0 \\ 0 & 0 & \delta_{DN} \end{bmatrix}$$

بگونه‌ای که δ_x و δ_y باندازه ابعاد پیکسلی و δ_{DN} باندازه یک واحد درجه خاکستری در نظر گرفته می‌شود.

بنابر رابطه‌ی (3-9)، واریانس تمامی نقاط در جهت X و Y و Z قابل محاسبه می‌باشد. در ادامه با گریدبندی فضای تحت پوشش نقاط، با میانگین‌گیری وزندار از تمامی نقاط داخل هر گرید یک نماینده انتخاب می‌کنیم.

$$X_c = \frac{\sum_{i=1}^n W_i . X_i}{\sum_{i=1}^n W_i}$$

$$Y_c = \frac{\sum_{i=1}^n W_i . Y_i}{\sum_{i=1}^n W_i} \quad (3)$$

$$Z_c = \frac{\sum_{i=1}^n W_i . Z_i}{\sum_{i=1}^n W_i}$$

باین ترتیب، دقت تمام نقاط موجود در گرید، در نقطه‌ای که به نمایندگی از تمام نقاط داخل گرید انتخاب می‌شود، لحاظ می‌شود. در ادامه، نتایج حاصل از تلفیق این نقاط هم‌مرجع شده در دو حالتی که دقت تمام نقاط موجود در گرید لحاظ شده باشد و یا حالتی که برای تک‌تک نقاط وزن یکسان در نظر گرفته شده باشد، ملاحظه می‌شود.

4- پیاده‌سازی و نتایج

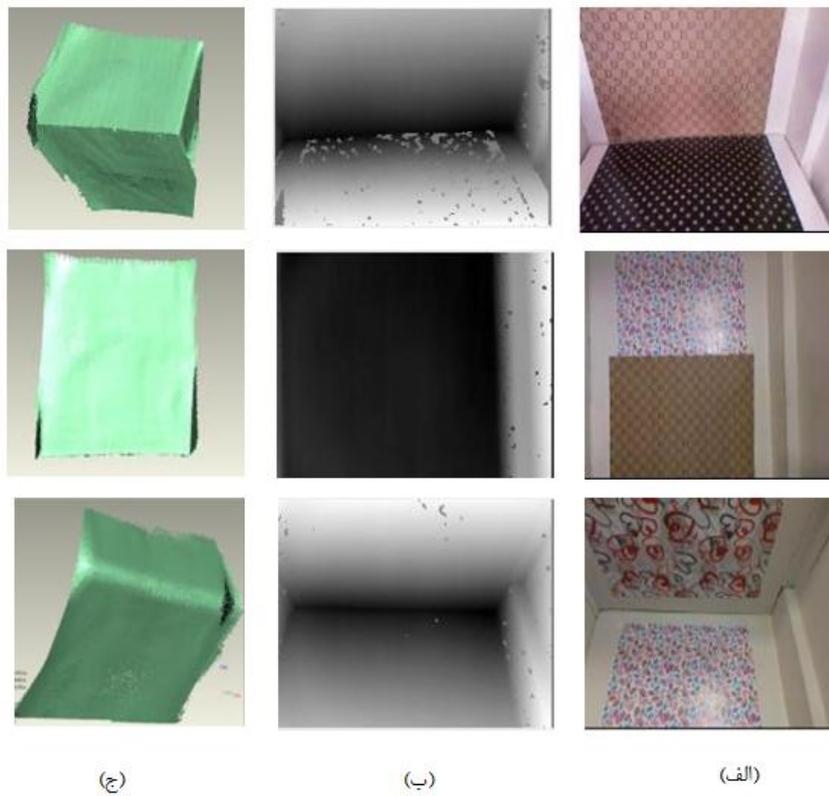
به‌منظور انجام آزمونهای ارزیابی روش‌های تلفیق این نقاط، از داده‌های کمد دیواری استفاده شد که در ادامه معرفی و تشریح می‌شود:

داده کمد دیواری: از سه سطح مسطح متعامد بافت‌دار (بالا- روبرو- پایین) در داخل کمد دیواری با ابعاد تقریبی کمتر از یک متر مکعب (شکل 3) تعداد دویست و پنجاه فریم متوالی از تصویر عمق و RGB سنجنده‌ی Kinect با حرکت توأم خطی و چرخشی متوالی اخذ شده است. از جمله موارد مورد توجه جهت به- کارگیری این داده، شرایط نوری محیط، زوایای واگرای فیلم‌برداری، عدم امکان مشاهده هم‌زمان سه صفحه در یک فریم، و فاصله‌ی دوربین تا محل فیلم‌برداری بوده که هر یک از این موارد به نوبه‌ی خود، تأثیر بسزایی در نتایج حاصل از آزمون خواهند داشت.



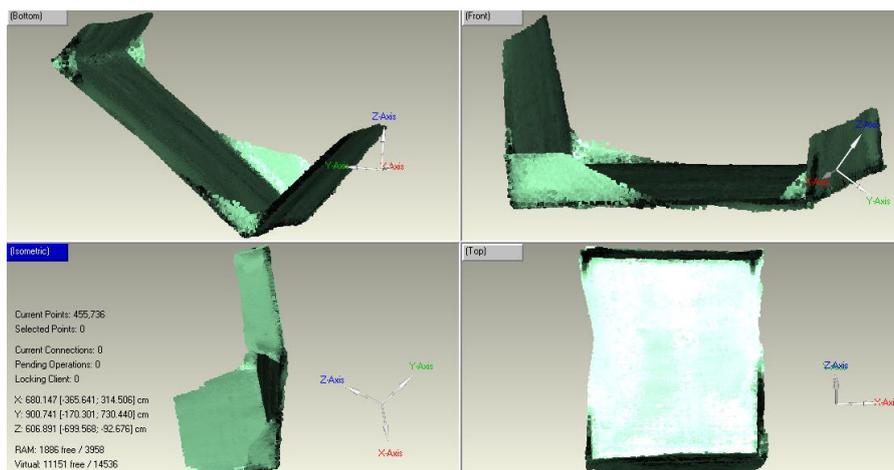
شکل 3: داده کمد دیواری

شکل‌های (4)، تصاویر رنگی و عمق میدان آزمون اخذ شده و ابرنقاط مربوط به فریم‌های متوالی مربوط به سه سطح مذکور را نشان می‌دهد.

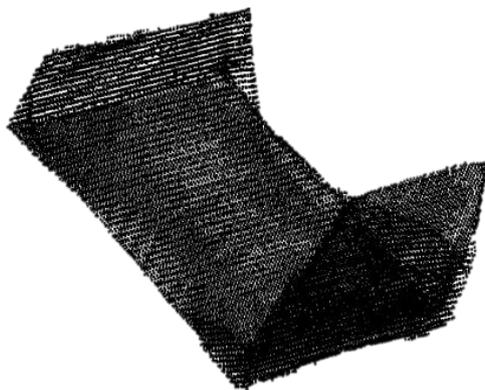


شکل 4: نمایشی از (الف) تصویر RGB، (ب) تصویر عمق، (ج) ابرنقاط در سه کنج

شکل‌های (5) و (6)، ابرنقاط هم‌مرجع شده را از زوایای متفاوت نمایش می‌دهد.

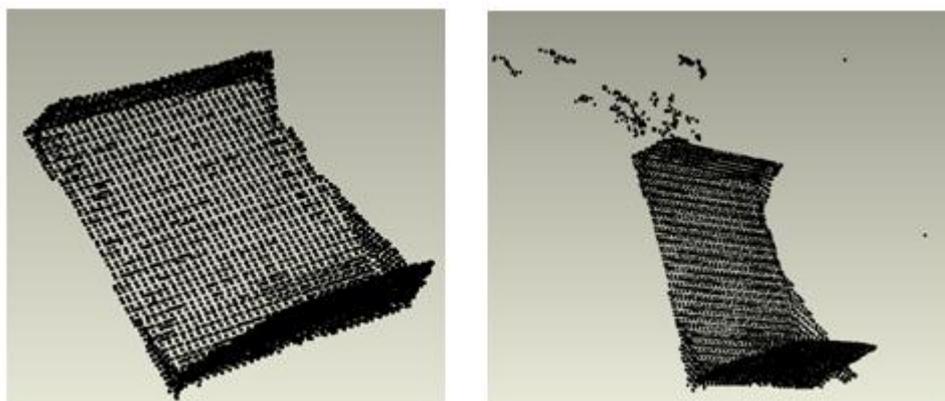


شکل 5: نمایشی از ابرنقاط حاصل از هم‌مرجع‌سازی فریم‌های متوالی مربوط به سه سطح مسطح



شکل 6: ابرنقاط تلفیق یافته از سه صفحه عمود برهم

همچنین دو روش متوسط‌گیری ساده (بدون وزن) و متوسط‌گیری وزندار مورد آزمون قرار گرفت. برای این-منظور ابرنقاط داده کمد دیواری هم‌مرجع‌سازی شد و یکبار با روش متوسط‌گیری ساده و یکبار با روش متوسط‌گیری وزندار ابرنقاط در هر مرحله افزایش فریم‌ها تلفیق گردیدند.



شکل 7: تلفیق وزندار و بدون وزن ابرنقاط هم‌مرجع‌شده

میزان خطای حاصل از برازش صفحه در روش تلفیقی وزندار 2.71 میلی‌متر است، در حالیکه این میزان خطا در روش تلفیقی بدون وزن 3.41 میلی‌متر می‌باشد و این نتایج حاکی از آن است که روش وزندار از عملکرد بهتری در تلفیق ابرنقاط برخوردار بوده است.

5- نتیجه‌گیری

توسعه‌ی ابزارهای کارآمد و ارزان قیمت در زمینه‌ی استخراج اطلاعات مکانی، لزوم توسعه روشهای جدید پردازش و الگوریتم‌های اجرایی جدید را می‌طلبد. سیستم‌هایی مثل Kinect قادر هستند حجم زیادی از موقعیت نقاط سه‌بعدی مقابل با میدان دید لحظه‌ای سنجنده را با نرخ بالایی اندازه‌گیری نمایند. ایجاد ابرنقطه از محیط‌های داخلی با استفاده از یک دوربین عمقی مانند کینکت، بدون اعمال هرگونه ساده‌سازی، به ایجاد نقاط اضافی زیادی منجر می‌شود که با افزایش حجم مشاهدات با گذشت زمان کوتاهی، حافظه‌ی موردنیاز و زمان محاسبات به شدت افزایش یافته که باعث کندی سرعت پردازش می‌شود. برای جلوگیری از این موارد، بایستی ابرنقاط سه‌بعدی در هر فریم با هم تلفیق شود. به عبارت دیگر؛ ارائه الگوریتمی کارآمد در جهت جلوگیری از افزایش تصاعدی داده‌های ابرنقطه از طریق تلفیق ابرنقاط هم‌مرجع شده فریم‌های متوالی ضروری

به نظر می‌رسد، که در نتیجه ابرنقطه یکپارچه و باکیفیت بالا (دقت هندسی بهتر و نواحی پنهان کمتر) تولید می‌شود از این رو در مقاله حاضر، یک راهکار جدید به منظور تلفیق ابرنقاط سه‌بعدی اخذ شده توسط دوربین‌های عمق ارائه شد و کارآمدی روش پیشنهادی در زمینه‌ی تلفیق با استناد به نتایج پیاده‌سازی تأیید گردید.

مراجع

1. Vincent Tao, Jonathan Li. (2007). “ Advances In Mobile Mapping Technology ” Taylor & Francis/Balkema, British Library.
2. Jan Smisek, Michal Jancosek and Tomas Pajdla. (2011). “3D with Kinect ”
3. J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, A. Blake. (2010). “Real-Time Human Pose Recognition in Parts from Single Depth Images ”, Microsoft Research Cambridge & Xbox Incubation.
4. Norbert Haala, Michael Petera, Jens Kremerb, Graham Hunterc, 2007, Mobile Lidar Mapping For 3D Point Cloud Collection In Urban Areas.
5. Gruen, A., Akca, D., 2004. Least squares 3D surface matching. ISPRS “ Panoramic photogrammetry Workshop ”, Dresden, Germany, February 19-22. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XXXIV, part 5/W16
6. Akca, D., (2007). “ Matching of 3D surfaces and their intensities ”, ISPRS Journal of Photogrammetry and Remote Sensing”, 62(2), 112_121
7. Tomi Kyostila, Daniel Herrera C., juho Kannala, and janne Heikkila, 2012. “Merging Overlapping depth maps into a nonredundant point cloud”
8. Merrell, p., Akbarzadeh, A., Wang, L., Mordohai, P., Frahm, j., Yang, R., Nister, D., Pollefeys, M.: Real-time visibility-based fusion of depth maps. In: IEEE 11th International Conference on Computer Vision (ICCV). (2007)
9. M. Wheeler, Y. Sato, and K. Ikeuchi. Consensus surfaces for modeling 3D objects from multiple range images. In IEEE International Conference on Computer Vision (ICCV), pages 917 – 924, 1998
10. Henry, P., Krainin, M., Herbst, E., Ren, X., Fox, D. (2010). “ RGB_D Mapping: Using Depth Cameras for Dense 3D Modeling of indoor Environment”, Proc. of International Symposium on Experimental Robotics (ISER).
11. K. Khoshelham. (2011). “Accuracy analysis of kinect depth data” In ISPRS Workshop Laser Scanning, volume XXXVIII.
12. Shahram Izadi, David Kim, O. Hilliges. (2011). “KinectFusion: Real-time 3D Reconstruction and Interaction Using a Moving Depth Camera”.

13. Peter Henry, Michael Krainin, Evan Herbst and Dieter Fox, RGB-D mapping: “Using Kinect-style Robotics indoor environments”, the International Journal of depth cameras for dense 3D modeling.
14. Weerasinghe, I. P. T. R. (2013). Automated Construction Worker Performance and Tool-time Measuring Model Using RGB Depth Camera and Audio Microphone Array System. University Of Calgary.