اولین کنفرانس مهندسی فناوری اطلاعات مکانی
دانشگاه صنعتی خواجه نصیرالدین طوسی
دانشکده مهندسی نقشه‌برداری
۲۹ و ۳۰ دی ماه ۱۳۹۴

The 1st National Conference on Geospatial Information Technology
K.N.Toosi University of Technology
Faculty of Geomatics Engineering
19 - 20 January 2016

انتخاب فریم های کلیدی در دنباله فریم های ویدئو به منظور بازسازی سه بعدی

حمیدرضا حسین پور ستوبادی[1*]، فرهاد صمدزادگان[2] ،فرزانه دادرس جوان [3]

۱- دانشجوی کارشناسی ارشد فتوگرامتری – دانشکده مهندسی نقشه برداری –پردیس دانشکده های فنی - دانشگاه تهران
۲- استاد دانشکده مهندسی نقشه برداری - پردیس دانشکده های فنی - دانشگاه تهران
۳- استادیار دانشکده مهندسی نقشه برداری - پردیس دانشکده های فنی - دانشگاه تهران

چکیده :

انتخاب فریم های کلیدی فرآیندی است که در آن فریم های حاوی اطلاعات مفید برای بازسازی سه بعدی از میان کل فریم های موجود در ویدئو انجام میگیرد. این کار با هدف بهبود روند بازسازی و کاهش حجم محاسبات انجام می شود. در این مقاله به معرفی روشی جهت انتخاب کردن فریم های کلیدی به منظور انتخاب بهترین فریم با کمترین خطای ریپروجکشن پرداخته شده است. روش پیشنهادی شامل حذف فریم های مات، فیلتر همپوشانی و طول بازه مناسب بین دو فریم و استفاده از معیار اطلاعات هندسی جهت جلوگیری از شرایط انحطاط است. نتایج حاصل از این تحقیق نشان دهنده کاهش خطای ریپروجکشن در مقایسه با روش های سنتی است. همچنین فرآیند بازسازی سه بعدی در صورت وجود شرایط انحطاط قابل پیاده سازی می باشد.

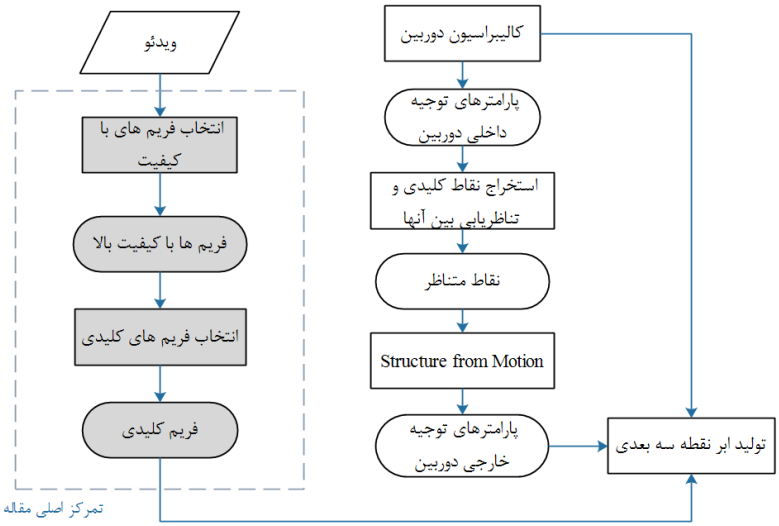
واژه‌های کلیدی : ویدئو، فریم کلیدی، معیار اطلاعات هندسی، خطای ریپروجکشن.

نویسنده مکاتبه کننده : حمیدرضا پورستوبادی
آدرس پستی : قزوین- خیابان شهید دستغیب – بین کوچه ۳ و ۵ پلاک ۱۸.
تلفن : ۰۹۱۲۷۸۵۵۱۴۴
آدرس پست الکترونیک : hosseinpoor@ut.ac.ir

۱- مقدمه

در سال های اخیر استفاده از روش های مبتنی بر بینایی کامپیوتر در بازسازی سه بعدی از فریم های ویدئویی (ویدئوگرامتری)، به عنوان روشی ارزان و قابل اعتماد نسبت به روش های سنتی، رشد چشمگیری داشته است[۱]. از دلایل مربوط به آن پیشرفت در تولید دوربین های با کیفیت بالا با توانایی اخذ هزاران پیکسل در هر فریم است. همچنین فریم های ویدئویی به صورت ترتیبی هستند. اطلاعات مربوط به هر فریم ویدئویی از فریم قبل خود قابل دستیابی است. از این نظر بازسازی سه بعدی سطح مورد نظر به صورت پیشرونده می باشد[۱، ۲ و ۳].

بازسازی سه بعدی از فریم های ویدوئی همواره با دو مشکل اساسی روبرو بوده است. اولین مشکل بروز پدیده مات شدگی فریم ها در طول فرآیند اخذ ویدئو است. وجود تصاویر مات در دنباله فریم های ویدئویی تاثیرات نامطلوب زیادی در فرآیند بازسازی دارد. مشکل دوم حجم بالای محاسبات در زمانی که از تمامی فریم های ویدئو استفاده گردد. به عنوان نمونه اگر دوربینی با نرخ ۳۰ فریم برثانیه اقدام به اخذ تصویر نماید در آن صورت در مدت زمان یک دقیقه حاوی ۱۸۰۰ فریم است، پردازش این حجم تصاویر از لحاظ محاسباتی زمان بر و ناکارآمد می باشد. بنابراین فرآیند ویدئوگرامتری نیازمند انتخاب فریم های کلیدی با کیفیت بالا و حاوی اطلاعات مفید از دنباله فریم های ویدئویی است. از آنجایی که فرایند بازسازی سه بعدی _ با تصاویر اخذ شده و یا فریم های انتخاب شده توسط کاربر _ به طور گسترده مورد مطالعه قرار گرفته است، ولی تحقیقات بسیار اندکی در زمینه معرفی الگوریتمی جهت انتخاب اتوماتیک فریم های کلیدی شده است[۴ و ۵]. با توجه به شکل (۱) پروسه تولید مدل سه بعدی از دنباله ویدئو نشان داده شده است. تمرکز اصلی این مقاله نشان داده شده است.



شکل ۱: پروسه ویدئوگرامتری جهت تولید مدل سه بعدی

در این مقاله به معرفی روشی برای انتخاب فریم های کلیدی در دنباله ویدئویی پرداخته شده است. فریم های کاندید پس از اندازه گیری متریک مات شدگی، با در نظر گرفتن طول مبنای مناسب بین فریم ها و جلوگیری از ایجاد شرایط انحطاط[1] در بازسازی سه بعدی انتخاب می گردند. درنهایت فریم های کلیدی با در نظر گرفتن بهترین توزیع نقاط استخراج شده در سراسر هر فریم انتخاب می شوند. در ادامه مروری بر تحقیقات صورت گرفته در این ضمینه پرداخته می شود، سپس مراحل مربوط به تعیین فریم های کلیدی به تفصیل بیان می شود. در نهایت نتایج حاصل از این تحقیق بر روی دو نمونه آزمایشی ویدئو پیاده سازی و ارزیابی می گردد.

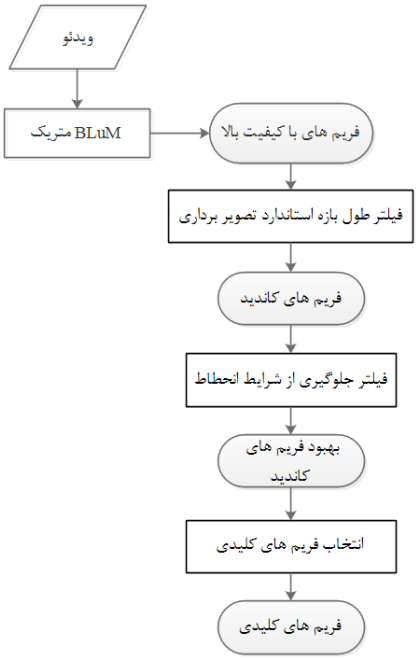
[1] Degeneracy

۲- مروری بر تحقیقات انجام گرفته

در سال های اخیر فعالیت های محدودی در زمینه استخراج فریم های کلیدی صورت گرفته است. [۴]. در این راستا [۳] روشی را جهت انتخاب فریم های کلیدی ارائه کرده است که بر مبنای اندازه گیری خطا در محاسبه ماتریس فاندامنتال و هموگرافی می باشد. در [۲، ۴، ۵ و ۹] از معیار اطلاعات هندسی درتعیین فریم های کلیدی استفاده کرده اند ، طبق این معیار از شرایط انحطاط در بازسازی سه بعدی توسط فریم ها جلوگیری به عمل می آید. در مقاله [۶] از اندازه گیری مقداری تیزی[۲] در فریم ها استفاده شده است. این مقدار در فریم ها نسبت مستقیمی با رزولوشن تصاویر دارد، بنابراین با اندازه گیری آن می توان فریم های غیر ضروری را حذف نمود. در مقاله [۷] از نسبت تناظر[۳] بین نقاط مشترک فریم ها جهت تعیین طول بهینه همپوشانی تصاویر استفاده کرده است. در مرحله بعد فریم کلیدی با کمترین خطای ریپروجکشن[۴] به منظور بازسازی تصویر انتخاب می گردد.

۳- روش پیشنهادی

همانطور که در بخش های پیشین اشاره گردید، تمرکز این مقاله بر تعیین فریم های کلیدی در دنباله فریم های ویدئویی می باشد. در همین راستا فلوچارت روش پیشنهادی در شکل (۲) آمده است.



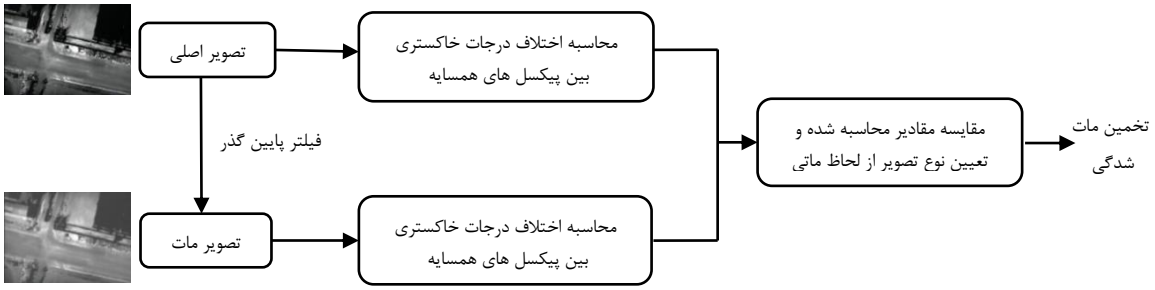
شکل ۲: فلوچارت پیشنهادی انتخاب فریم کلیدی

۳-۱- حذف فریم های مات شده

گام اول الگوریتم پیشنهادی حذف فریم های دارای اثر مات شدگی است. این مرحله با استفاده از روش ارائه شده در مقاله [۸] صورت می پذیرد. از آنجایی که پدیده مات شدگی در تصاویر به علت از دست رفتن مقادیر با فرکانس بالا می باشد، بنابراین می توان یک تصویر را با اعمال یک فیلتر پایین گذر مانند میانگین گیری مات نمود. با مطالعه طبیعت تصاویر مات تولید شده توسط فیلتر های پایین گذر متفاوت، می توان دریافت که تشخیص تفاوت بین یک تصویر تار و تصویر تار تولید شده از همان تصویر دشوار است. هر چقدر روند مات شدگی تصاویر ادامه پیدا کند، پیکسل

[2] Sharpness
[3] Correspondence Ratio
[4] Reprojection

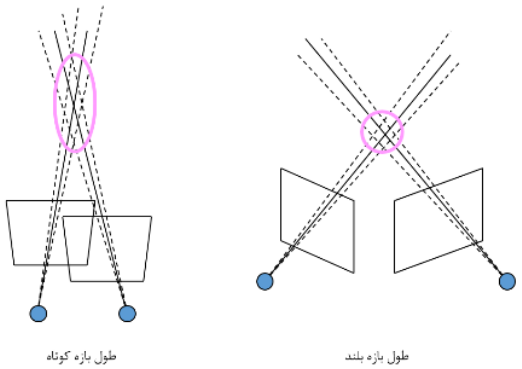
های همسایه به سمت درجه خاکستری یکسان تمایل پیدا می نمایند. اگر یک تصویر با کنتراست بالا مات گردد، مقادیر درجه خاکستری پیکسل های همسایگی با درجه بالایی از تغییرات دچار تغییر می گردد. حال اگر همین تصویر مات شده را با همان فیلتر پایین گذر دوباره مات نماییم مقادیر پیکسل های همسایه دچار تغییرات کمتری نسبت به حالت قبل می شوند. بنابراین برای شناسایی مات بودن یک تصویر می توان آن تصویر را مات نمود و مقادیر اختلاف بین پیکسل های همسایه را در دوحالت قبل و بعد از اعمال فیلتر پایین گذر مورد بررسی قرار داد. مرحله اول محاسبه مقایر اختلاف بین پیکسل های همسایه در تصویر اصلی می باشد. پس از اعمال فیلتر پایین گذر به این تصویر دوباره مقادیر اختلاف بین پیکسل های همسایگی محاسبه می گردد. مقایسه این دو مقدار اختلاف بین دوحالت بیان شده در تعیین کردن تصویر مات از غیر مات کمک می نماید. بنابراین اختلاف زیاد بین تصویر اصلی و مات شده نشان دهنده آن است که تصویر اصلی به عنوان یک تصویر شارپ بوده است. هر چقدر این اختلاف بین تصویر اصلی و تصویر مات شده زیاد تر باشد، در آن صورت تصویر اصلی به عنوات یک تصویر مات در نظر گرفته می شود. این مطلب را می توان به صورت خلاصه در شکل (۳) مشاهده می گردد.



شکل ۳: فلوچارت حذف فریم های مات شده

۳-۲- فیلتر کردن بهینه طول بازه و همپوشانی بین فریم ها

طول بازه بهینه و مقدار همپوشانی مورد نیاز بین دو فریم متوالی یکی از موارد مهم در بازسازی سه بعدی محسوب می گردد. همانطور که در شکل (۴) نشان داده شده است طول بازه کوتاه سبب افزایش مقدار خطای اندازه گیری در مقایسه با طول بازه بلند است[۷] .



شکل ۴: تاثیر طول بازه بین دو فریم در دقت بازسازی سه بعدی

برای بررسی طول بازه مناسب از نسبت تناظر بین دو فریم استفاده می گردد، که معادل نسبت تعداد نقاط مشترک بین دو تصویر به تعداد نقاط استخراج شده در تصویر اول می باشد. این نسبت از رابطه (۱) محاسبه می گردد:

$$R_c = \frac{T_c}{T_f} \quad (۱)$$

در رابطه فوق Rc به عنوان نسبت تناظر شناخته می شود، Tc به عنوان تعداد نقاط مشترک بین دو فریم و T_f تعداد کل نقاط استخراج شده در فریم انتخابی اول است. مقدار عددی Rc رابطه معکوسی با حرکت دوربین دارد. به طوری که در فریم های اولیه این مقدار نزدیک به یک می باشد و با حرکت دوربین این مقدار شروع به کاهش می نماید. در زمانی که دوربین حرکتی نداشته باشد این مقدار نزدیک به یک است. بنابراین این مقدار به عنوان اندازه گیری مناسب برای تخمین حرکت دوربین مورد استفاده قرار می گیرد.

۳-۳-فیلتر فریم های ویدئویی جهت جلوگیری از ایجاد شرایط انحطاط

ماتریس فاندامنتال برای بدست آوردن موقعیت یک دوربین نسبت به موقعیت دوربین دیگر مورد استفاده قرار می گیرد. با این وجود در حالت های انحطاط محاسبه موقعیت دوربین امکان پذیر نیست. از مهم ترین حالت های انحطاط می توان به انحطاط حرکت و انحطاط ساختاری اشاره نمود.

انحطاط حرکت: اگر دوربین حول محور خود دوران بدون انتقال داشته باشد. در این حالت هندسه اپی پولار قابل تعریف شدن نیست. با این حال با معلوم بودن نقاط کنترل در فضای سه بعدی می توان ماتریس پروژکتیو دوربین را محاسبه نمود.

انحطاط ساختاری: این حالت زمانی رخ می دهد که تمامی نقاط ساختار سه بعدی بر روی یک صفحه قرار گرفته باشد. هندسه اپی پولار قابل تعریف شدن نیست، و با استفاده از نقاط متناظر نمی توان ماتریس F را تخمین زد.

ارتباط بین نقاط متناظر و ساختار کلی دوربین در بین دو تصویر به طور مناسب با ماتریس فاندامنتال تعیین می گردد. در حالت های انحطاط جهت تطبیق زوج تصاویر به هم از مدل هموگرافی استفاده می شود. بنابراین برای مقایسه ماتریس فاندامنتال و مدل هموگرافی حالت های فوق مورد بررسی قرار میگیرند. برای این منظور از معیار انتخاب هندسی روباست (GRIC)[5] بین مدل هموگرافی و فاندامنتال استفاده می شود [۱۰]، این مدل بر مبنای اندازه گیری بهترین انطباق است. این مقدار از جمع دو مولفه حاصل می شود. استفاده کردن از تخمین کننده ماکزیمم احتمال برای تخمین مدل و ساختار های متفاوت جهت بررسی کردن با استفاده از GRIC بسیار مهم می باشد. در تعیین معیار GRIC، پارامتر n معروف به تعداد نقاط inlier، e_i بردار باقی مانده ها، σ انحراف معیار اندازه گیری نقاط، r ابعاد داده های اندازه گیری- مثلا این مقدار برای نقطه دو و برای خط چهار می باشد- ، مقدار عددی k برای پارامترهای حرکتی مدل و d ابعاد ساختار مدل مورد نیاز است. این معیار طبق رابطه (۲) معین می شود .

$$GRIC = \sum \rho(e_i^2) + (\lambda_3 nd + \lambda_3 k) \quad (۲)$$

$$\rho(e^2) = \min(\frac{e^2}{\sigma^2}, \lambda_3(r-d))$$

فریم کلیدی در صورتی انتخاب میگردد که مدل فاندامنتال در بین دو تصویر مقدار کمتری از لحاظ مقدار GRIC نسبت به ماتریس هموگرافی داشته باشد.

[5] Geometric Robust Information Criteria

۳-۴-انتخاب فریم های کلیدی

بعد از محاسبه ماتریس فاندامنتال و هموگرافی با استفاده از الگوریتم رنسک (RANSAC) [۱۱] ، در هر مدل درصد نقاط اینلیر (inlier) به تعداد کل نقاط محاسبه می گردد. مقدار عددی S برای تعیین فریم های کلیدی طبق رابطه (۳) محاسبه می گردد[۱۲].

$$S = (1 - \sigma)\frac{S_F - S_H}{S_F} \quad (۳)$$

در رابطه فوق S_H درصد نقاط اینلیر به تعداد کل نقاط برای محاسبه ماتریس هموگرافی، S_F درصد نقاط اینلیر به تعداد کل نقاط در محاسبه ماتریس فاندامنتال و انحراف معیار محاسبه شده از نحوه توزیع نقاط در سراسر فریم های کاندید می باشد. برای محاسبه فریم مورد نظر به چند زیر ناحیه تقسیم بندی می گردد. چگالی نقاط برای زیر نواحی و تمامی فریم ها به صورت مجزا محاسبه می گردد. در نهایت مقدار انحراف معیار با استفاده از رابطه (۴) محاسبه می شود.

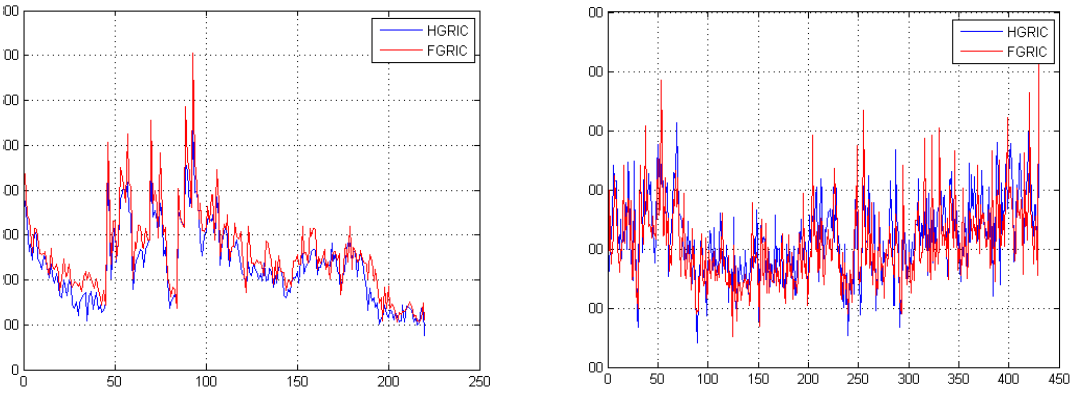
$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(N_i - \frac{N}{n})^2} \quad (۴)$$

Ni و N تعداد نقاط استخراج شده در هر ناحیه و در کل فریم می باشد. n معرف تعداد زیر نواحی است. فریم کلیدی در صورتی انتخاب می شود که مقدار عددی S در آن بیشترین مقدار ممکن باشد.

۴- پیاده سازی و ارزیابی نتایج

نتایج حاصل از این تحقیق در دو نمونه ویدئو اخذ شده مورد ارزیابی قرار گرفت. فریم های کلیدی به دوصورت انتخاب می شوند در حالت اول فریم ها به صورت دستی و با نرخ نمونه برداری ۱۰ فریم انتخاب می گردند. در حالت دیگر از روش پیشنهادی استفاده شده است. ابتدا فریم های مات شده از مجموعه فریم های ویدئویی حذف می گردند در این مرحله مقدار حد آستانه برای حذف فریم های مات در ویدئوی اول ۰.۳۶ و در ویدئوی دوم ۰.۴ انتخاب گردید. پس از حذف فریم های مات طول بازه همپوشانی جهت حذف فریم های با همپوشانی بالا طبق رابطه (۱) انجام می شود. در این تحقیق از دو حد آستانه بالا و پایین جهت این کار استفاده گردیده است. حد آستانه بالا نشان دهنده وجود درصد نقاط مشترک بین دو فریم است به گونه ای که اگر مقدار ضریب تناسب محاسبه شده از این مقدار بیشتر شود دو فریم دارای طول بازه بسیار کمی هستند که در کاربردهای بعدی مورد استفاده نمی باشد، حد آستانه پایین هم نشان دهنده آن است که اگر ضریب تناسب بین دو فریم از این مقدار کمتر گردد طول بازه مناسب بین آن ها زیاد می شود و این حالت هم مطلوب نمی باشد. در این مقاله این مقادیر برای حد آستانه بالا ۰.۸۵ و برای حد آستانه پایین ۰.۶۵ انتخاب شده است.

پس از فیلتر کردن فریم های دارای همپوشانی استاندارد، در صورت وجود شرایط انحطاط از معیار GRIC استفاده می شود. ابتدا ماتریس فاندامنتال و هموگرافی بین جفت فریم ها محاسبه می شود. سپس مقدار عددی GRIC هم برای ماتریس فاندامنتال و هم برای ماتریس هموگرافی طبق رابطه (۲) به دست می آید. در صورتی که مقدار عددی معیار GRIC برای ماتریس فندامنتال از ماتریس هموگرافی کمتر شود در آن صورت آن فریم را به عنوان فریم کاندید انتخاب می نماییم. نتایج حاصل از این مرحله برای دو نمونه ویدئوی به کار گرفته شده در این پژوهش در شکل(۵) نشان داده شده است.



شکل ۵: نمودار GRIC برای دو نمونه ویدئوی پردازش شده، محور افقی بیان کننده شماره فریم ها و محور عمودی نشان دهنده مقدار GRIC است. ویدئوی شماره یک (سمت چپ)، ویدئوی شماره دو (سمت راست).

در مرحله بعد ابتدا هر فریم به ۱۶ ناحیه تقسیم می شود مقدار انحراف معیار توزیع تعداد نقاط استخراج شده در هر فریم طبق رابطه (۴) تعیین می گردد. در نهایت فریم کلیدی با استفاده از رابطه (۳) در این مقاله انتخاب می گردد. مقدار خطای ریپروجکشن برای هر دو روش پیشنهاد شده محاسبه شده است. این نتایج را می توان در جدول (۱) مشاهده نمود.

جدول ۱: نتایج حاصل از پیاده سازی : روش پیشنهاد شده دارای کمترین مقدار خطای ریپروجکشن در دو نمونه ویدئوی پردازش شده می باشد.

ویدئو	تعداد کل فریم ها	روش انتخاب فریم کلیدی	تعداد فریم های کلیدی	خطای ریپروجکشن	
				میانگین	انحراف معیار
ویدئوی شماره یک	۲۷۰	روش دستی	۲۷	۰.۵۴۱۷	۰.۰۳۹۵
		روش پیشنهادی	۱۵	۰.۴۵۷۸	۰.۰۰۱۹
ویدئوی شماره دو	۵۰۰	روش دستی	۵۰	۰.۴۸۴۶	۰.۰۳۶۸
		روش پیشنهادی	۳۳	۰.۴۳۱۴	۰.۰۱۱۶

۵- نتیجه گیری

در این مقاله به ارائه روشی جهت انتخاب فریم های کلیدی از دنباله فریم های ویدئویی با ترکیب قید هایی مبتنی بر ضریب تناظر، مقدار GRIC برای ماتریس فاندامنتال و هموگرافی و روش انتخاب فریم کلیدی پرداخته شده است. قبل از اعمال روش پیشنهاد شده فریم های مات شده در هنگام تصویر برداری با اعمال متریک مات شدگ حذف می گردد. نتایج حاصل از این تحقیق نشان دهنده کاهش خطای ریپروجکشن در مقایسه با روش های سنتی می باشد. همچنین با محدود کردن پروسه تولید مدل سه بعدی به استفاده کردن از فریم های کلیدی علاوه بر این که حجم محاسبات به طرز چشمگیری کاهش پیدا می کند، از بروز پدیده انحطاط و یا وجود نویز در فریم ها در حین پروسه تولید مدل سه بعدی جلوگیری به عمل می آید.

مراجع

[1] R.Hartley, and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[2] J. Repko, and M. Pollefeys. "3D models from extended uncalibrated video sequences: Addressing key-frame selection and projective drift. 3-D Digital Imaging and Modeling", 3DIM 2005. Fifth International Conference on, IEEE, 2005.

[3] S. Gibson, J. Cook, T. Howard, R. Hubbold, and D. Oram. "Accurate camera calibration for off-line, video-based augmented reality". Proceedings of the 1st International Symposium on Mixed and Augmented Reality, IEEE Computer Society, 2002.

[4] A. Rashidi, F. Fei, I. Brilakis, and P. Vela. "Optimized selection of key frames for monocular videogrammetric surveying of civil infrastructure." Advanced Engineering Informatics 27(2): 270-282, 2013.

[5] M. T. Ahmed, N. Matthew, J. L. Landabaso, and N. Herrero . "Robust Key Frame Extraction for 3D Reconstruction from Video Streams." VISAPP, 2010.

[6] Nistér, D. "Frame decimation for structure and motion. 3D Structure from Images", SMILE 2000, Springer: 17-34, 2001.

[7] YH. Seo, SH. Kim, KS. Doo, and JS. Choi . "Optimal keyframe selection algorithm for three-dimensional reconstruction in uncalibrated multiple images." Optical Engineering 47(5): 053201-053201-053212, 2008.

[8] F. Crete, T. Dolmiere, P. Ladret, and M. Nicolas, "The blur effect: perception and estimation with a new no-reference perceptual blur metric.", Electronic Imaging 2007, International Society for Optics and Photonics, 2007.

[9] T. Thormählen, H. Broszio, and A. Weissenfeld, "Keyframe selection for camera motion and structure estimation from multiple views." Computer Vision-ECCV 2004, Springer: 523-535, 2004.

[10] P. Torr, A. Fitzgibbon, and A. Zisserman, "Maintaining multiple motion model hypotheses over many views to recover matching and structure.", Computer Vision. Sixth International Conference on, IEEE, 1998.

[11] Fischler, M. A. and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography." Communications of the ACM 24(6): 381-395, 1981.

[12] JK. Seo, SH. Kim, CW. Jho, and HK Hong, "3D estimation and key-frame selection for match move." ITC-CSCC: 2003: 1282-1285, 2003.

Keyframe Selection from Video Stream for 3D Reconstruction

Hossein Pour, H.R. *[1], Samazadegan, F. [2], Dadras Javan F

1-Ms.c student of photogrammetry in Department of Geomatics, College of Engineering, University of Tehran
2- Professor in Department of Geomatics, College of Engineering, University of Tehran
3- Assistant professor in Department of Geomatics, College of Engineering, University of Tehran

Abstract

Key frame selection is th process of selecting the essential frames in which the frames containing useful information for three-dimensional reconstruction from among many uncalibrated frames. It aims is to improve the process of restructuring and reducing the volume of calculations. This paper introduces a method for selecting key frames in order to select the best frames with minimal reprojrction error. The proposed method involves the removal of blurred frames, filters and during the period of overlap between the two frames and using geometric robust information criteria to avoid degeneration. The results of this research indicate reprojrction error reduction compared with traditional methods. The three-dimensional reconstruction process are implemented if the condition decline.

Keywords: video, keyframe, reprojection error, GRIC

Correspondence Address: Photogrammetry Group, Department of Geomatics, College of Engineering, University of Tehran, Tehran, Iran. **Tel: +98 9127855144.**

Email: hosseinpoor@ut.ac.ir