

3D-Grafik

WAS IST RAYTRACING?

Spätestens seit Nvidias Vorstellung der neuen RTX-2000-Grafikkarten ist das Schlagwort Raytracing in aller Munde. Aber was genau ist das überhaupt? Was haben Spieler davon? Und was kostet es? Von Nils Raettig

In der GameStar-Ausgabe 07/2004 haben wir uns zum ersten Mal mit dem Thema Raytracing auseinandergesetzt. Etwa 14 Jahre danach spielt die Render-Technik in Spielen immer noch so gut wie keine Rolle. Im Zuge der GDC 2018 samt Microsofts Vorstellung einer passenden Lösung für die Grafikschnittstelle DirectX 12 und vor allem der Präsentation der neuen GeForce-RTX-Grafikkarten von Nvidia auf der Gamescom hat sie aber wieder viel Aufmerksamkeit erlangt. Einerseits, weil Raytracing absolut im Fokus

von Nvidias Gamescom-Präsentation stand. Andererseits, weil die RTX-2000-Modelle besonders gut für das Echtzeit-Raytracing geeignet sein sollen. Außerdem wurden bereits kommende Spiele mit der Technik wie Shadow of the Tomb Raider oder Battlefield 5 in Aktion gezeigt. Was sich hinter Raytracing verbirgt, erklären wir in diesem Artikel.

Ein alter, neuer Render-Ansatz

Raytracing dient im Kern dazu, per Computer Bilder zu erstellen, die für die Darstellung auf einer zweidimensionalen Fläche wie einem PC-Bildschirm oder einer Kinoleinwand gedacht sind. Solche Flächen bestehen aus einzelnen Bildpunkten oder Pixeln, die sich zu einer bestimmten Auflösung zusammensetzen (etwa 1.920 Pixel in der Breite und 1.080 Pixel in der Höhe, auch als »Full HD« bekannt). Welches Bild wir auf einem Monitor oder einer Leinwand zu sehen bekom-

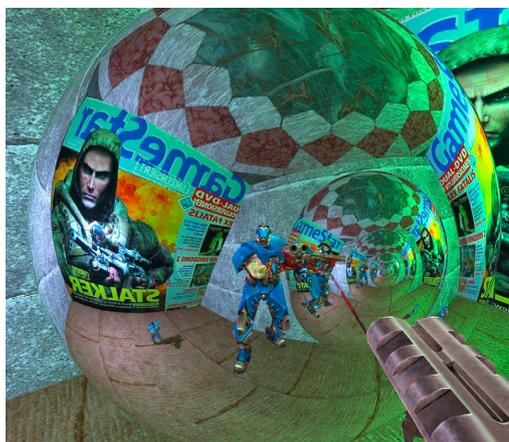
men, hängt von der Farbe ab, die die einzelnen Pixel jeweils annehmen. Damit diese Farbgebung zu einem möglichst realistisch aussehenden Bild führt, gibt es unterschiedliche Ansätze wie eben Raytracing oder die aktuell bei PC-Spielen meist genutzte »Rasterisierung« (englisch »Rasterization«), auf die wir gleich noch näher eingehen.

Raytracing ist dabei wie eingangs erwähnt kein grundsätzlich neuer Render-Ansatz, ganz im Gegenteil. Im Kinobereich und in der Werbung wird es schon seit vielen Jahren erfolgreich eingesetzt. Dank »CGI« (steht für »Computer Generated Imagery«) werden hier künstliche Bilder geschaffen, die täuschend echt wirken. Die Berechnung eines Bildes mittels Raytracing kann allerdings sehr lange dauern. Während das für Filme und die Werbung kein grundsätzliches Problem darstellt, sieht es in Computerspielen anders aus. Hier beeinflussen wir schließlich selbst, was wir sehen und was passiert, weshalb die passenden Bilder nicht vorgeordnet werden können und stattdessen in Echtzeit berechnet werden müssen.

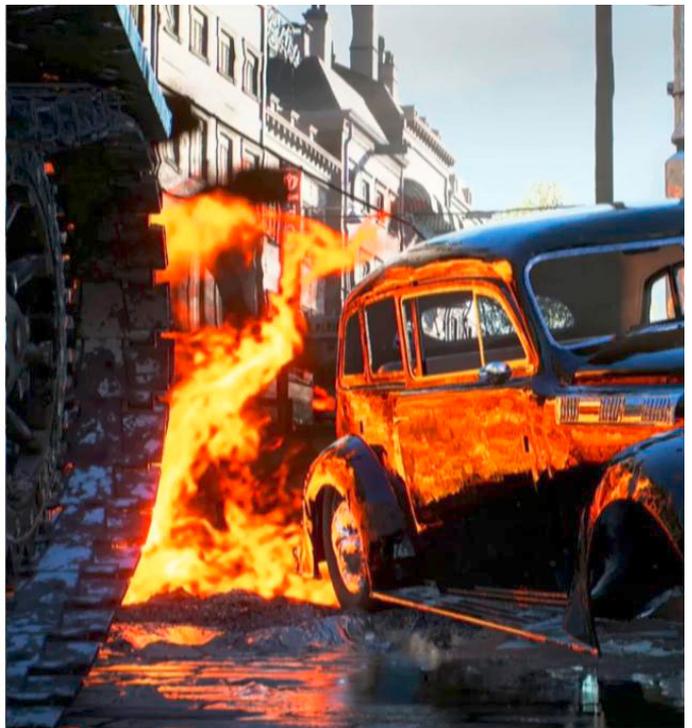
Für große Kino- und Werbeproduktionen steht außerdem meist immense Rechenleistung in Form von Serverfarmen zur Verfügung, während Computerspiele auf normaler PC-Hardware lauffähig sein müssen. Diese beiden Faktoren sind der Hauptgrund dafür, dass Echtzeit-Raytracing in Spielen (noch) keine Rolle spielt. Aber was genau ist Raytracing überhaupt? Und wie unterscheidet es sich von der Rasterisierung, die momentan meist in Computerspielen genutzt wird?

Die Rasterisierung

Um in Computerspielen ein passendes Bild der (3D-)Welt für die Darstellung auf dem (2D-)Monitor rendern zu können, ist zu-



Für die GameStar 07/2004 hat der Student Daniel Pohl (im kleinen Bild rechts zwischen seinem Professor Slussallek und unserem damaligen Hardware-Redakteur Daniel Visarius zu sehen) diesen Screenshot mit Raytracing in Quake 3 erstellt. Das Rendern erfolgte per CPU auf einem Linux-Rechnerverbund mit 48 Athlon-MP-Prozessoren und 12,0 GByte RAM.



Battlefield 5 wird wohl eines der ersten Spiele mit Echtzeit-Raytracing sein. Da sich das Feuer zu großen Teilen nicht im Blickfeld befindet, wird es links ohne Echtzeit-Raytracing nicht von der spiegelnden Oberfläche des Autos reflektiert, rechts mit Raytracing dagegen schon.

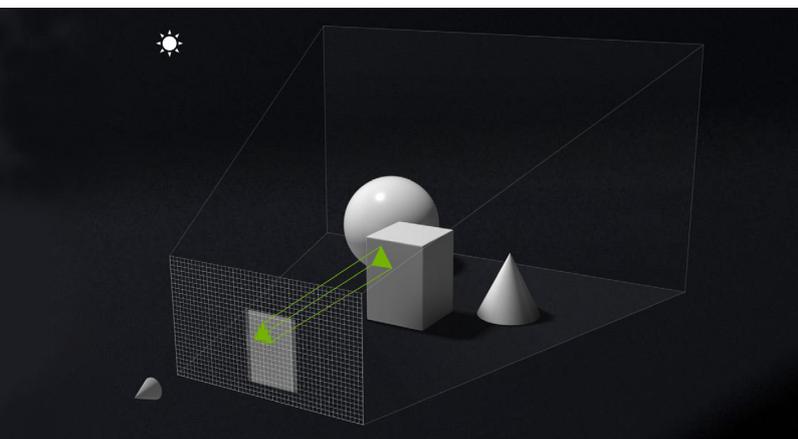
nächst ein Ausgangspunkt nötig. Dabei handelt es sich in der Regel um die Sicht, die wir als Spieler darauf einnehmen, also quasi um das virtuelle Auge, das auf die Spielwelt schaut. Die Objekte in dieser Spielwelt sind meist dreidimensional, zu der horizontalen x-Achse und der vertikalen y-Achse kommt also noch eine z-Achse für die Tiefe hinzu. Dabei bestehen die Objekte aus einzelnen Polygonen (»Vieleck«, genauer gesagt kommen in der Regel Dreiecke zum Einsatz), aus denen sich wiederum die bekannten Drahtgitter-Modelle zusammensetzen.

Bei der Rasterisierung wird vereinfacht ausgedrückt ein Abbild der dreidimensionalen Spielwelt auf ein zweidimensionales (Pixel-)Raster projiziert (siehe auch das Bild unten). Was das Abbild genau zeigt, hängt vom aktuellen Blickwinkel auf die Spielwelt ab. Die im Abbild enthaltenen Polygone liegen dabei auf bestimmten Pixeln und haben dadurch Einfluss auf ihre Farbgebung. Befin-

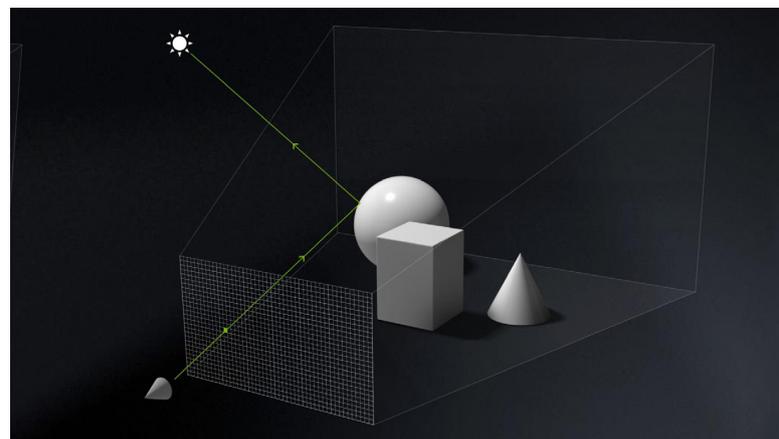
det sich ein Objekt außerhalb des Blickfeldes (oder hinter einem anderen Objekt), wird es dementsprechend nicht auf dem Raster abgebildet. Einerseits spart dieser geometriebasierte Ansatz Rechenleistung. Andererseits bedeutet er, dass die ausgelassenen Objekte nur auf Umwegen in die Farbgebung einberechnet werden können.

Dieser Umstand führt vor allem in Bezug auf die Beleuchtung und Reflexionen zu Problemen. Schließlich kann eine Lichtquelle auch dann Einfluss auf unsere Wahrnehmung einer Szene haben, wenn wir sie selbst nicht sehen. Und spiegelnde Oberflächen können Dinge zeigen, die sich nicht in unserem Blickfeld befinden. Um trotzdem ein möglichst realistisches Bild erzeugen zu können, bedienen sich aktuelle 3D-Engines allerlei Tricks wie zusätzlicher Kameraperspektiven. Insgesamt kann das die Rasterisierung zu einer durchaus leistungshungrigen, aber optisch auch überzeugenden

Angelegenheit machen. Ein großer Pluspunkt dieses Render-Ansatzes besteht dabei darin, dass sich die grundsätzlich sehr ähnlich ablaufenden Berechnungen für die Farbgebung jedes einzelnen Pixels sehr gut parallelisieren lassen. Während eine CPU wie etwa der Core i7 8700K oder der Ryzen 7 2700X in der Regel eher wenige gleichzeitig ablaufende, dafür aber komplexere Berechnungen für sehr unterschiedliche Zwecke übernimmt, ist im Gegensatz dazu die Parallelisierung von sehr vielen, größtenteils ähnlichen und weniger komplexen Berechnungen die Königsdisziplin von Grafikkarten. Deshalb können aktuelle Prozessoren auch nur eine vergleichsweise geringe Zahl von Aufgaben (beziehungsweise Threads) gleichzeitig bearbeiten, während schnelle Grafikkarten über eine große Zahl an Recheneinheiten (beziehungsweise Shader-Units) für parallel ablaufende Berechnungen ähnlicher Art verfügen.



Die Rasterisierung wird seit langer Zeit dafür genutzt, die dreidimensionale, virtuelle Welt von PC-Spielen für die Darstellung auf zweidimensionalen Flächen wie PC-Monitoren vorzubereiten.



Beim Raytracing werden Strahlen vom aktuellen Blickwinkel aus in die Spielwelt geschickt und verfolgt. Auf welche Objekte sie treffen und wie sie weitergeleitet werden, bestimmt die zweidimensionale Darstellung.

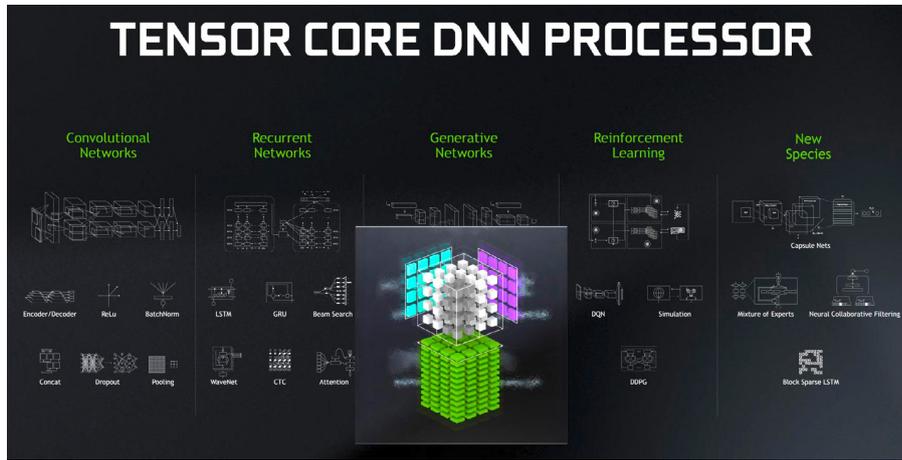


Auch (oder gerade) transparente Elemente wie Glas und spiegelnde Oberflächen lassen sich mittels Raytracing besonders realistisch darstellen.

Strahlen statt Raster

Der große Vorteil von Raytracing besteht darin, dass Tricks für eine realistische Darstellung damit nicht mehr (oder kaum noch) nötig sind. Der Grundansatz der Technik ist gleichzeitig relativ simpel: Im Prinzip bildet sie nach, wie wir die reale Umwelt über unsere Augen wahrnehmen. Hier treffen Lichtstrahlen (beispielsweise von der Sonne) auf Objekte und werden von ihnen reflektiert. Diese Reflexionen gelangen wiederum in unser Auge und vermitteln uns nach einer passenden Verarbeitung im Gehirn ein Bild von unserer Umgebung.

Raytracing dreht diesen Ansatz gewissermaßen um und schickt vom virtuellen Auge aus für jeden vorhandenen Bildpunkt Strahlen in die Spielwelt und verfolgt sie. Daher auch der Name der Technik: »Ray« bedeutet »Strahl«, »to trace« heißt übersetzt »etwas verfolgen«. Trifft einer dieser Strahlen auf ein Objekt (beziehungsweise auf ein Polygon), verrät das bei passender Programmierung nicht nur, dass es sich dort befindet, sondern auch, welche Eigenschaften (wie beispielsweise die Farbe oder die Oberflächenstruktur) es besitzt – und auf welche Art der Lichtstrahl davon beeinflusst wird. So kann er etwa je nach Einfallswinkel und Beschaffenheit des Objekts in eine bestimmte Richtung abgelenkt oder auch ge-



Die Tensor-Cores der Turing-Architektur sollen auch über das sogenannte DLSS per Deep-Learning-Algorithmus für eine Verbesserung der Bildqualität sorgen können. DLSS muss aber explizit von Spielen unterstützt werden.

brochen werden, beispielsweise in transparenten Materialien wie Glas.

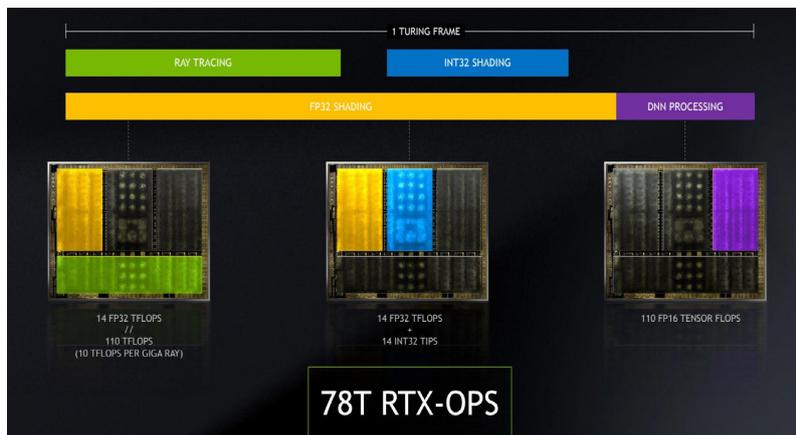
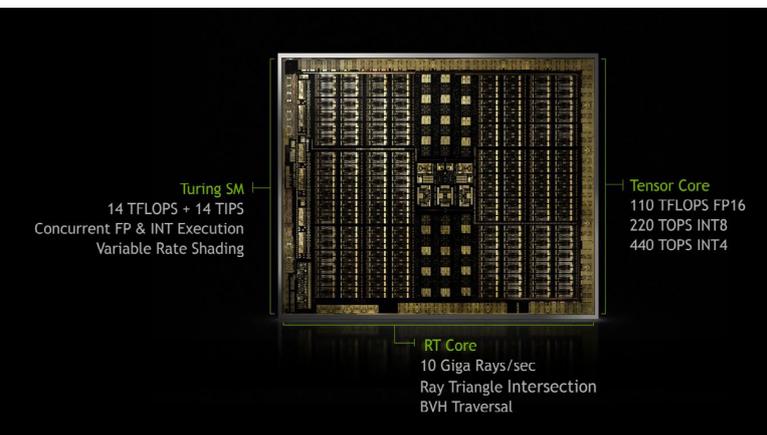
Die Strahlen lassen sich dabei auch in Bereiche verfolgen, die nicht im direkten Sichtfeld des virtuellen Auges liegen. Deshalb können auch Objekte und Lichtquellen außerhalb davon beim Raytracing die Darstellung der (Spiel-)Welt beeinflussen, was wir ja genauso von der realen Welt kennen. Aber auch wenn die Technik im Grunde relativ leicht umsetzbar ist, bedeutet es hohen Rechenaufwand, ausreichend viele Strahlen für das Rendern eines Bildes der Spielwelt zu verfolgen – der sich außerdem im Gegensatz zum geometriebasierten Vorgehen bei der Rasterisierung deutlich schwieriger parallelisieren lässt. Das ist auch der Grund, warum Echtzeit-Raytracing in Spielen bislang praktisch nicht zum Einsatz kommt.

Bei der Präsentation der neuen RTX-GPUs hat Nvidias CEO Jensen Huang davon gesprochen, dass Echtzeit-Raytracing eine Rechenleistung im Petaflop-Bereich erfordern würde – die aber erst in zehn Jahren erreichbar sei. Das entspricht einer Billiarde an Gleitkomma-Rechenoperationen pro Sekunde (10^{15}). Momentan kommen Grafikkarten »nur« auf Werte im Teraflop-Bereich, was einer Billion solcher Operationen entspricht (10^{12}). Aber wie will Nvidia dann jetzt schon Echtzeit-Raytracing umsetzen?

Raytracing per Hardware

Die neue Turing-Architektur von Nvidia ist speziell auf das Echtzeit-Raytracing ausgelegt. Eine ihrer wichtigen Neuerungen besteht dementsprechend in speziellen Recheneinheiten, die nur dazu da sind, das Raytracing zu beschleunigen. Nvidia nennt sie wenig überraschend »Raytracing-Cores« (beziehungsweise »RT-Cores«). Außerdem ein sehr wichtiger Faktor: Raytracing soll zunächst per »RTX Hybrid Rendering« in Spielen Einzug halten. Das bedeutet, dass die Berechnung des Bildes mit einer Kombination aus klassischer Rasterisierung und Raytracing erfolgt. So soll Raytracing beispielsweise in Shadow of the Tomb Raider zunächst nur im Hinblick auf die realistische (und damit unter anderem weichere) Darstellung von Schatten genutzt werden.

Bei der Gamescom-Präsentation sagte Jensen Huang, dass die Raytracing-Berechnung ohne die neuen RT-Cores zehnmal so lange dauern würde. Dazu passend liefert er auch einen neuen Wert für die Messung der Raytracing-Performance, die sogenannten »Giga Rays per Second«. Während die GTX 1080 Ti ohne spezielle RT-Einheiten laut Jensen auf 1,21 Giga Rays pro Sekunde kommt, sind mit der neuen Turing-Architektur 10 Giga Rays pro Sekunde möglich. Außerdem kann die Berechnung des Bildes per Deep



Nvidias neue Turing-Architektur bietet unter anderem spezielle Raytracing-Recheneinheiten, die in diesem Bereich für eine vergleichsweise hohe Leistung sorgen sollen.

Die Leistung aller Recheneinheiten kombiniert Nvidia zu »RTX-OPS«. Die neuen RTX-2000-Karten sollten aber auch unabhängig von Raytracing- und Tensor-Cores flotter unterwegs sein als die Vorgänger-GPUs.



Shadow of the Tomb Raider soll Raytracing für die Schattendarstellung nutzen, allerdings voraussichtlich erst per Patch nach dem Release.

“Turing improves raytracing up to 6X compared to Volta... Insane workloads still manageable.”
- Colin Barré-Brisebois and Henrik Halén — SIGGRAPH 2018

SEED // PICA PICA: Hardware Raytracing & Turing

Hybrid Rendering Pipeline

Deferred shading (raster)	Direct shadows (raytrace or raster)	Lighting (compute + raytrace)	Reflections (raytrace or compute)
Global Illumination (compute and raytrace)	Ambient occlusion (raytrace or compute)	Transparency & Translucency (raytrace and compute)	Post processing (compute)

Raytracing wird wohl noch längere Zeit in Kombination mit klassischen Render-Methoden eingesetzt, um Rechenleistung zu sparen.

Learning-Algorithmus ergänzt werden, wofür bei Turing die neuen »Tensor-Cores« vorhanden sind. Vereinfacht ausgedrückt soll der Algorithmus die korrekte Farbgebung von Pixeln ermöglichen, die noch nicht oder nicht korrekt im regulären Render-Vorgang erfasst wurden. Das bezeichnet Nvidia als »Deep Neural Network«. Darüber könnte beispielsweise das beim Raytracing häufig auftretende Problem des Bildrauschens behoben oder zumindest gelindert werden, das entsteht, wenn nicht genug Strahlen zur Erfassung von Bildinformationen ausgesendet werden. Auch in diesem Fall hat Jensen einen Leistungsvergleich zu der Vorgänger-Generation Pascal parat: Die Tensor-Cores von Turing sollen KI-Berechnungen ermöglichen, für die man zehn Geforce GTX 1080 Ti benötigen würde, die nicht über diese speziellen Recheneinheiten verfügen.

Ein weiterer Einsatzzweck der Tensor-Cores ist unabhängig von Raytracing gegeben: »DLSS« (»Deep Learning Super-Sampling«). Dabei wird ein KI-Algorithmus angewendet, der zuvor auf einem Supercomputer darauf trainiert wurde, die Bildqualität eines Frames anhand zuvor erlernter Beispiele automatisch zu erhöhen.

Was es bringt – und was es kostet

Im besten Fall führt Raytracing zu einer authentischeren Darstellung von Spielen, indem unter anderem alle vorhandenen Lichtquellen physikalisch korrekt berücksichtigt werden, Schattenwürfe weich und echt aussehen und Reflexionen Objekte so widerspiegeln, wie wir das von der realen Welt gewohnt sind. Die ersten Demos von der Gamescom lassen allerdings bereits vermuten, dass es noch so manchen Stolperstein auf dem Weg in die schöne, neue Raytracing-Welt gibt. Einer davon ist der erwartungsgemäß große Leistungshunger, der sich etwa am Beispiel von Shadow of the Tomb Raider andeutet. Ein Video zeigt, dass selbst das neue Flaggschiff Geforce RTX 2080 Ti das Spiel bei aktivierten »Raytraced Shadows« in vielen Szenen nur mit 30 bis 40fps darstellen konnte – und zwar in Full HD. Allerdings sind weder die gezeigten Spiele noch die RTX-2000-Karten respektive deren Treiber schon veröffentlicht. Etwas Zeit für Optimierungen ist also noch vorhanden. Außerdem wird man Raytracing voraussichtlich separat ein- und ausschalten können. Es sollte also möglich sein, für eine bessere Performance darauf zu verzichten.

Eine Möglichkeit, Ressourcen beim Raytracing zu schonen, dürfte darin bestehen, weniger Strahlen zu verfolgen – und sie weniger weit zu verfolgen. Das hat aber wiederum Einfluss auf die Darstellungsqualität. Solche Kompromisse werden aber vermutlich auch mit Nvidias neuen RTX-Karten noch länger unumgänglich sein, wenn die Zahl der Bilder pro Sekunde stimmen soll. Die Technik hat unabhängig davon durchaus das Potenzial, die Darstellung von Computerspielen deutlich zu verbessern. Eine wirklich neue Erkenntnis ist das aber nicht. Gleichzeitig wird sich wohl erst anhand von Testexemplaren der neuen RTX-Grafikkarten samt passender Raytracing-Spiele zeigen, wie groß der Fortschritt wirklich ist, den Nvidia in dieser Hinsicht mit Turing erzielt hat. ★



Ist Raytracing in dieser Szene aus Metro nicht aktiviert, sind auch Bereiche des Raums ausgeleuchtet, die das Licht aus dem Fenster nicht so stark erhellen dürfte (links). Mit Raytracing wirkt die Beleuchtung dagegen realistischer (rechts).



Nils Raettig
@nraettig

Es hat sich im Vorfeld der Gamescom-Präsentation schon abgezeichnet, dass Nvidia bei den neuen RTX-Grafikkarten einen starken Fokus auf das Raytracing legen würde. Dass er so stark ausfällt, hätte ich dann aber doch nicht gedacht. Ich persönlich stehe dem Ganzen ehrlich gesagt noch sehr skeptisch gegenüber. Das ist aber nicht deshalb der Fall, weil ich an der Idee hinter Raytracing oder den potenziellen Ergebnissen zweifle. Es liegt vielmehr daran, dass der Rechenaufwand auch für die neuen Turing-Karten noch zu viel sein könnte, während die optischen Verbesserungen beim Spielen selbst nicht immer im Verhältnis zum erhöhten Leistungshunger stehen dürften. Mehr Gewissheit darüber werden wir aber wohl erst in den kommenden Wochen und Monaten erlangen, wenn die ersten Geforce-RTX-Karten und vor allem die ersten Raytracing-Spiele (hoffentlich) zur Verfügung stehen.