

Rekonstruktion von Stasi-Unterlagen per Computer

Bewährungsprobe für E-Jigsaw

Unmittelbar nach dem Fall der Berliner Mauer hat das Ministerium für Staatssicherheit der ehemaligen Deutschen Demokratischen Republik einen großen Teil seiner Akten vernichtet. Viele davon wurden zerrissen und lagern jetzt in der Behörde der Bundesbeauftragten für die Unterlagen des Staatssicherheitsdienstes (»Gauck-Behörde«) in etwa 17 000 Säcken à 20 000 Schnipsel. Seit Jahren bemüht man sich, diese Unterlagen von Hand wieder zusammzusetzen, kommt dabei aber nur schleppend voran - etwa ein Jahr braucht eine Person pro Sack. Die Arbeitsgruppe von Prof. Angelika Steger, Fakultät für Informatik der TUM in Garching, hat nun das Computerprogramm E-Jigsaw entwickelt, das die Machbarkeit der IT-gestützten Rekonstruktion zeigt.

Angelika Steger rief das Projekt im Herbst 2000 ins Leben, um motivierten Studierenden nach dem Vordiplom eine attraktive Aufgabe zu bieten, die eine echte Herausforderung sein und einen spannenden Hintergrund haben sollte. Seitdem sind daraus über 20 Studienarbeiten hervorgegangen. Die Diplom-Informatiker Martin Marciniszyn, Stephan Micklitz und Andreas Weißl haben sich in ihren Diplomarbeiten mit E-Jigsaw beschäftigt und promovieren heute bei Angelika Steger. Sie stellen das Projekt und seine Problematik vor:

Die Komplexität des Problems manifestiert sich in der Anzahl der Schnipselpaare, deren Zusammengehörigkeit das System zu bewerten hat. Obwohl dieser Prozess parallelisierbar ist, scheidet ein Brute-Force-Ansatz, der alle möglichen Kombinationen ausprobiert, aus - zu groß ist die Anzahl der Einzelteile. Selbst heutige High-Performance-Computer sind nicht in der Lage, alle

möglichen Anordnungen der Teile in vernünftiger Zeit zu bewerten. Nötig ist daher

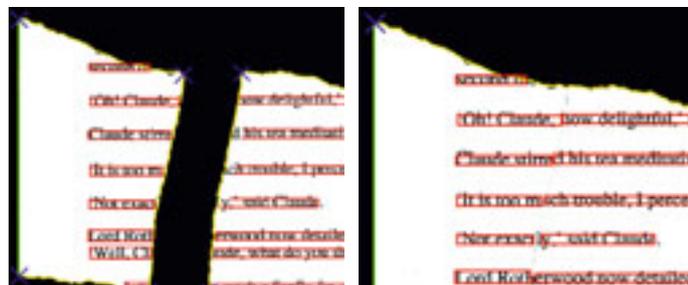


Das »Team E-Jigsaw« bei der Arbeit (v.l.): Andreas Weißl, Stephan Micklitz, Prof. Angelika Steger und Martin Marciniszyn.

Foto: Karlheinz Egginger

eine effektive Puzzlestrategie, um die Komplexität zu reduzieren. Das an der TUM entwickelte System E-Jigsaw setzt die Seiten nach dem Bottom-Up-Prinzip wieder zusammen, kehrt also den Ablauf des Zerreißvor-

gangs exakt um. Diejenigen Schnipsel, die beim Zerreißen als letzte entstanden sind, werden als erste aneinandergefügt. Somit lässt sich die



Zwei wieder zusammengefügte Schnipsel. Gelb: Konturen, Blau: Lokalisation der Eckpunkte, Rot: Lage und Ausdehnung der Textzeilen, Grün: Seitenränder.

Menge der möglichen Partner für einen Schnipsel stark einschränken und - bei Verwendung geeigneter Datenstrukturen - effizient ermitteln.

tion von Eckpunkten auf der Kontur, Orientierungswinkel sowie Lage und Ausdehnung von Textzeilen. Diese Informationen wer-

den in einer zentralen Datenbank abgelegt. Der rechenaufwendige Analysevorgang lässt sich einfach parallelisieren, indem man die Menge der zu analysierenden Bilder auf mehrere Computer verteilt. Die Rekonstruktion selbst besteht im Wesentlichen aus der iterativen Suche nach jeweils zwei zusammengehörenden Teilen. Für alle Paare, die gemäß effizient überprüfbarer Charakteristika zueinander passen könnten, stellt das System so genannte Hypothesen auf. Eine anschließende Hypothesenbewertung überprüft jede einzelne dieser Aussagen eingehend. Das Verfahren stellt die Wahrscheinlichkeit für die Richtigkeit einer Hypothese anhand geometrischer sowie auf Text basierender Übereinstimmungskriterien an der Risskante auf und lässt sich ebenfalls parallelisieren. Stuft es eine Hypothese als korrekt ein, vereinigt es die entsprechenden Schnipsel zu einem neuen, der wieder zu den zu rekonstruierenden Teilen gegeben wird. Dieser Prozess setzt sich solange fort, bis schließlich vollständige Seiten vorhanden sind.

Kritisch für die Effizienz des Systems ist die Toleranz bei der Charakterisierung der zu einem Schnipsel potentiell passenden Gegenstücke. Eine geringe Toleranz reduziert die Laufzeit, verhindert aber die Aufstellung gewisser eigentlich richtiger Hypothesen. Das passiert beispielsweise bei »Eselohren« an einem Schnipsel oder unterschiedlichen Zeilenstrukturen beidseits einer Risskante. Diese Problematik wurde durch Implementierung eines geeigneten Rundenkonzepts gelöst, bei dem die Toleranz mit der Verweildauer eines Schnipsels im System sukzessive erhöht wird. Der Prototyp wurde an einem Sack mit fiktiven Unterlagen aus etwa 1 700 zerrissenen DIN A4-Seiten einem Härte-test unterzogen. Nach einigen Anlaufschwierigkeiten bewies das System seine Tauglichkeit: Ein ganzer Sack ließ sich innerhalb einer Woche nahezu vollständig computergestützt rekonstruieren.

Weitere Informationen zu E-Jigsaw gibt es im Internet unter:
www.e-jigsaw.de

Ein Kernthema des Forschungsgebiets Telepräsenz, das die Deutsche Forschungsgemeinschaft im Rahmen des Sonderforschungsbereichs »Wirklichkeitsnahe Telepräsenz und Teleaktion« an der TUM fördert, ist die Fernsteuerung von Robotern durch Menschen. Dabei wird unter anderem versucht, dem Benutzer über seine natürlichen Sinneseindrücke den Eindruck zu vermitteln, er befände sich selbst an der Stelle des Roboters, ganz als ob beispielsweise seine Augen die Kameras des Roboters wären. Wenn der Bediener eines Wartungs-Roboters im Orbit den zu reparierenden Satelliten greifen und sehen kann, so wird er die Aufgabe bestmöglich erfüllen.

Die technische Umsetzung erfordert auf Seite des Bedieners, dessen Bewegungen zu erfassen, zum Roboter zu übertragen und dort auszuführen. Auf Seite des Roboters werden multimodale - etwa visuelle und haptische - Sinneseindrücke aufgezeichnet, zum Bediener geschickt und diesem vermittelt. Die Datenübertragung zwischen Bediener und Roboter ist jedoch mit teils erheblichen Zeitverzögerungen verbunden. Im Falle visueller Informationen, auf die sich das hier beschriebene Projekt beschränkt, nimmt der Bediener bereits Verzögerungen von 0,25 Sekunden zwischen Bewegung und dem daraus resultierenden Bild wahr - bei einer Sekunde ist kein intuitives Arbeiten mehr möglich. Der Bediener ver-

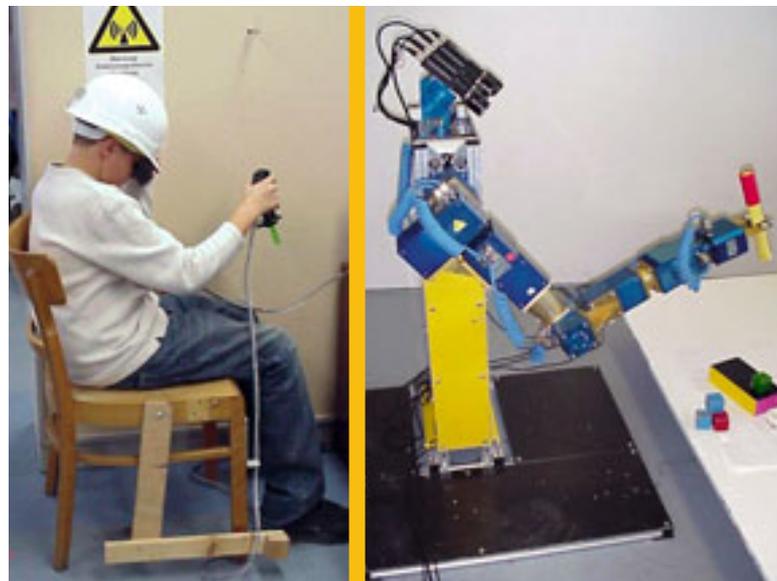
Telepräsenz: Fernsteuerung von Robotern

Bilder aus der Zukunft

Die Zukunft vorherzusagen ist Sache von Wahrsagern, Börsenspekulanten und Metereologen. Doch auch das spezielle Gebiet der Telepräsenz - ein Schwerpunkt am Lehrstuhl für Realzeit-Computersysteme der TUM (Prof. Georg Färber) - kann von Prädiktion profitieren, um Verzögerungen im Kommunikationskanal geschickt zu kompensieren.

fällt auf eine ganz und gar ineffiziente »move and wait«-Strategie. Der Beschleuniger

diener ein künstlich generiertes Bild der Umgebung des Roboters präsentiert,



Wissenschaft zum Anfassen: Dieser junge Besucher nutzte den »Tag der offenen TUM«, um am Lehrstuhl für Realzeit-Computersysteme den Roboter MinERVA mittels Telepräsenz fernzusteuern.

Foto: Jan Leupold

gung der Übertragung sind jedoch durch die Lichtgeschwindigkeit sehr harte Grenzen gesetzt.

Daher entwickelt der Lehrstuhl für Realzeit-Computersysteme der TUM ein »Prädiktives Display« zur Kompensation der Übertragungszeiten. Dabei wird dem Be-

das die zu erwartenden Veränderungen bereits zeigt, bevor sie in der Realität eintreffen. Bewegt der Bediener ein Objekt, so sieht er sofort das simulierte Ergebnis seiner Handlungen, obwohl der reale Roboter die Aktion erst später ausführt. Um eine derartige Bildprädiktion zu erreichen, ist ein