

„Intelligent Rooms“

Michael Würsch

01-701-754, wuersch@ifi.unizh.ch

Roman Haag

03-707-817, cac@kapsler.ch

12. Juni 2006

Inhaltsverzeichnis

EINFÜHRUNG	3
ANWENDUNGSSZENARIEN FÜR INTELLIGENTE RÄUME	5
<i>Szenario 1: Eine Kommandozentrale des Hurrikan Katastrophenschutzes in der Karibik.</i>	5
<i>Szenario 2: Eine Applikation im Kontext des Smart Homes, welche den Benutzer bei einer Vielzahl von Aktivitäten unterstützt.</i>	6
ENTWURFSPRINZIPIEN FÜR INTELLIGENTE RÄUME	7
DAS INTELLIGENT ROOM PROJECT VOM MIT ARTIFICIAL INTELLIGENCE LAB	10
DAS PERSONEN-TRACKING SYSTEM	11
DAS SYSTEM ZUR GESTENERKENNUNG	12
DAS SPRACHERKENNUNGSSYSTEM	13
THE SELF-ORGANIZING DESK	16
AUFBAU DES SYSTEMS.....	16
<i>Überwachungsfunktion (camera control)</i>	18
<i>Segmentierungsmodul</i>	18
<i>Informationsgewinnung</i>	20
<i>Informationsabfrage</i>	21
<i>Visualisierung</i>	22
PROBLEME DES SYSTEMS	23
ZUKUNFTSAUSSICHTEN UND SCHLUSSFOLGERUNG	23
FAZIT	24
LITERATURVERZEICHNIS	26
ANHANG	26

Einführung

Die Rolle zeitgemässer Arbeits-, wie auch Wohnräume wandelt sich zunehmend von simplen Aufenthaltsräumen hin zu hochtechnisierten multimedialen und multimodalen Einrichtungen. Diese "Intelligent Rooms" oder "Intelligenten Räume" bieten ihren Benutzern mannigfaltige Unterstützung indem sie Routinearbeiten automatisiert im Hintergrund durchführen, eine wahre Flut von Informationen bereitstellen, diese dem Anwender in adäquater und übersichtlicher Art und Weise zugänglich machen und ihn somit bei der Entscheidungsfindung unterstützen.

Da eine solche Computer- und Informationstechnologie im Alltag zunehmend an Bedeutung gewinnt und somit auch einen stärkeren Einfluss auf den Menschen ausübt, ist ein Umdenken beim Entwurf und der Realisierung der entsprechenden Lösungen angezeigt: Bestehende Systeme versuchen dem Benutzer mittels Metaphern (z.B. Schreibtischmetapher) eine Schnittstelle zum Verstehen und Bedienen der Maschine zu bieten und so ihre Handhabung intuitiv zu gestalten. Nichtsdestotrotz verlangen sie vom Benutzer stets eine mentale Umstellung: Marktübliche graphische Benutzeroberflächen verlangen in der Regel, dass ein Anwender einen Kontextwechsel vollzieht, sprich eine andere Tätigkeit unterbricht, sich an einen Bildschirm setzt und mittels Maus und Tastatur auf einem virtuellen Desktop virtuelle Repräsentationen von Daten in Form von Ordnern und Dateien manipuliert. Während diese auf Abstraktion basierende Art der Interaktion für informatik-versierte Personen heute selbstverständlich ist, bringt sie für unbedarfte Anwender einen enormen Lernaufwand mit sich (Bsp.: Die graphische Abbildung einer Bildlaufleiste birgt an und für sich keine Hinweise darauf, dass diese zum Blättern durch ein digitales Dokument verwendet werden kann. - Dieser Umstand muss vom Anwender erst erlernt werden.).

Im Zentrum der nachfolgenden Betrachtungen steht nun eine komplett neue Sichtweise, die Coen [1] sinngemäss wie folgt auf den Punkt bringt: *Künftiges Anwendungsdesign sollte weniger den Entwurf von Benutzerschnittstellen für Computer, als Computerschnittstellen für Benutzer, anstreben, wobei man sich unter diesen Anwendungen weniger Tabellenkalkulationen und Textverarbeitung, als Intelligente Räume und persönliche Assistenten, vorzustellen hat.*

Dieser human-zentrierte Ansatz bedeutet konkret, dass Anwender mit IT-gestützten Systemen auf dieselbe natürliche Art und Weise, wie mit anderen Menschen interagieren. Die neue Sichtweise soll einerseits dazu beitragen, die Benutzerfreundlichkeit künftiger Systeme zu maximieren und andererseits die zu Grunde liegende Technik unsichtbar werden lassen. Natürliche Instrumente und Informationsquellen wie Gesten, Sprache, Affekt und Kontext sollen Tastaturen, Computermäuse, Menüs und Fenster ablösen.

Im Folgenden sollen nun einige Szenarien konstruiert werden, welche durch die obenbeschriebenen Technologien realisiert werden könnten. Anschliessend werden wir einige Entwurfsprinzipien für Intelligente Räume näher beleuchten, bevor wir schliesslich anhand von Forschungsprototypen -- dem Intelligent Room Project von der Artificial Intelligence Group am MIT [4] und den Self-Organizing Desk [3] -- konkrete Realisierungen vorstellen.

Anwendungsszenarien für Intelligente Räume

Das Konzept der Intelligenten Räume lässt sich auf viele Problemstellungen anwenden, welche umfangreiche Routinearbeiten und einen Bedarf an der Bereitstellung, Beurteilung und Visualisierung von umfangreichen Informationen haben. Prominentes Beispiel ist der sog. "Command Post of the Future" [2], also ein militärischer Kommandoposten, welcher die kollaborative strategische Planung erleichtern soll, indem Informationen über Freund- und Feindbewegungen gesammelt und aufbereitet, sowie Werkzeuge zu virtuellen Strategie- und Einsatzbesprechungen bereitgestellt werden.

Dieses Grundkonzept lässt sich allerdings auch auf zivile Belange übertragen. Coen führt folgende Beispiele an:

Szenario 1: Eine Kommandozentrale des Hurrikan Katastrophenschutzes in der Karibik.

Katastrophenschutzbeauftragte sind im Falle eines schweren Tropensturm auf umfangreiche Echtzeitdaten über die Wetterlage, Verletzte, Position und Ausrüstung der Hilfskräfte und vieles mehr angewiesen. Eine beispielhafte Interaktion mit einem intelligenten Kommandoraum, könnte wie folgt aussehen:

Benutzer: "Computer, <pause> Bereitschaftsmodus".

[Das System schaltet in einen speziellen Zustand und akzeptiert fortan Kommandos, ohne dass der Benutzer diese mit dem Wort "Computer" einleiten muss.]

Benutzer: "Zeig mir die Virgin Islands!"

[Der Raum bestätigt das Kommando mit einem kurzen Zirpen und zeigt die gewünschten Information auf dem Bildschirm an, welcher sich am nächsten beim Benutzer befindet.]

Benutzer: [zeigt auf St. Thomas] "Vergrössern. Wie weit entfernt befindet sich der Hurrikan Marilyn?"

Raum: "Die Distanz zwischen dem Zentrum des Hurrikans Marilyn und der Stadt Charlotte Amalie auf St. Thomas beträgt 145 Meilen".

Benutzer: "Wo befindet sich das nächstgelegene Büro des Katastrophenschutzes?"

Raum: "Das Büro des Katastrophenschutzes auf St. Thomas befindet sich eine Meile ausserhalb von Charlotte Amalie. Michael, soeben ist ein neue neue Wetterprognose eingetroffen. Möchtest Du sie sehen?"

Benutzer: "Ja, zeig mir das Satellitenbild!"

Szenario 2: Eine Applikation im Kontext des Smart Homes, welche den Benutzer bei einer Vielzahl von Aktivitäten unterstützt.

Coen stellt sich eine Applikation ähnlich dem Hal-System im Film 2001: A space odyssey vor:

"Ich betrete Hal, schliesse die Tür und lege mich auf das Sofa. Hal bemerkt dies, reduziert die elektrische Raumbelichtung, schliesst die Rollläden und spielt leise Mozart im Hin-

tergrund. Schliesslich fragt er mich: "Michael, um welche Uhrzeit möchtest Du geweckt werden?"

Die Implementierung der oben beschriebenen Szenarien mittels Forschungsprototypen verfolgt mehrere Ziele: Zum einen soll auf eine explorative Art und Weise definiert werden, was denn nun einen Intelligenten Raum ausmacht und welche sensorischen Fähigkeiten dieser benötigt, um seine Funktion zu erfüllen. Zum anderen soll untersucht werden, inwiefern sich diese Technologie auf unser alltägliches Leben auswirkt. Aus diesen Erkenntnissen lassen sich dann für diese Systeme unter Anderem allgemeingültige Entwurfsprinzipien ableiten, auf welche wir im Folgenden kurz eingehen möchten.

Entwurfsprinzipien für Intelligente Räume

Intelligente Umgebungen sollen uns in unseren alltäglichen Tätigkeiten unterstützen. Dies impliziert, dass diese hochinteraktiven Technologien nahtlos in unser Leben eingebettet werden müssen, damit sie benutzerfreundlich aber dennoch unaufdringlich und im Idealfall unsichtbar für uns sind.

Um dies bewerkstelligen zu können, müssen diese Systeme multimodal, also mit umfangreicher Sensorik, zum Beispiel in Form von Kameras und Audioaufnahmegeräten, sowie mit Bild- und Spracherkennungssoftware ausgestattet werden. Auf diese Weise lassen sich Kontextinformationen nutzen, die bislang von der computerzentrierten Warte aus irrelevant waren. So tendieren Menschen beispielsweise dazu, auf jene Dinge zu zeigen, über die sie sprechen. Diese Geste kann nun vom System erkannt und z.B. dazu verwendet werden, um das entsprechende Element hervorzuheben oder Zusatzinformation über die-

ses bereitzustellen. Während bislang Tastatureingaben und Mausklicks die zentralen Mechanismen zur Bedienung von Computersystemen waren, können nun auch Ereignisse, wie das Betreten eines Raumes, welche ausserhalb der traditionellen Menge von Mensch-Maschine-Interaktionen liegen, für sinnvolle Zwecke verarbeitet werden.

Im Gegensatz zum Ubiquitous Computing bedeutet dies allerdings nicht, dass jeder Gegenstand des Alltags, seien es zum Beispiel Möbel mit Drucksensoren oder Kleider mit Sendern zur Positionsbestimmung, mit zusätzlicher Elektronik aufgewertet werden soll. Im Speziellen soll der Mensch nicht dazu gezwungen werden elektronische Geräte wie PDA's oder Rucksäcke voller technischer Spielereien mit sich herumzuschleppen, sondern - wie bereits mehrfach erwähnt - auf biologisch natürliche Art und Weise agieren.

Erkenntnisse aus der modernen Forschung auf dem Gebiet der künstlichen Intelligenz können hier wertvolle Dienste leisten: Bilddaten, aufgenommen durch eine Kamera, liefern zum Beispiel weit mehr Informationen als einfachere Sensorsysteme. Die Daten können zum Beispiel zur Positions-, Aktivitäts- oder Identitätsbestimmung oder gar zur Analyse des Gemütszustandes des Anwenders verwendet werden. Eine geschickte Auswahl der modalen Komponenten erlaubt es, dass sich imperfekte Modalitäten gegenseitig ergänzen und Synergien bilden [Cohen]. Indem die grundlegenden Entwurfsprinzipien der New A.I. bei der Entwicklung der Systeme befolgt werden, lassen sich einfache und preiswerte, aber dennoch leistungsfähige Lösungen realisieren.

Zusammenfassend lassen sich also bereits heute die folgenden Entwurfsprinzipien auflisten:

- Die Interaktion mit alltäglichen Gegenständen soll weiterhin unbelastet und frei von technischen Erweiterungen bleiben.
- Im Speziellen soll der Benutzer davor bewahrt werden, spezielle elektronische Geräte am Körper tragen zu müssen.
- Im Sinne der New A.I. sollen clevere Lösungen entwickelt und Synergien zwischen einzelnen Sensoren genutzt werden.

Das Intelligent Room Project vom MIT Artificial Intelligence Lab

Die Forschungsgruppe am MIT hat sich in ihrem Intelligent Room Project zum Ziel gesetzt, Lösungen für die oben beschriebenen Szenarien zu entwickeln, um herauszufinden, welche A.I.-Technologien hier bereits eine Anwendung finden könnten und auf welchen Gebieten weitere Forschung von Nöten ist.

Hierzu wurde ein Prototyp eingerichtet, welcher räumlich in zwei Bereiche unterteilt ist. Die eine Hälfte des Raumes ist wie ein gewöhnlicher Konferenzraum eingerichtet, in dessen Zentrum ein grosser Tisch mit Stühlen steht. Zwei helle LCD-Projektoren und mehrere kleinere Bildschirme, sowie Kamera- und Videoinfrastruktur stehen den Anwendern zur Benutzung zur Verfügung. Zusätzlich sind an verschiedenen Stellen insgesamt 12 Videokameras montiert, welche dem Raum sozusagen als Augen dienen.

Durch kleinere bauliche Massnahmen abgetrennt von der Konferenzfläche, befindet sich ein Bereich mit Workstations, welche die Kernsoftware des Raumes beherbergen und die anfallenden Berechnungen durchführen. Dadurch, dass dieser Bereich unmittelbar an den eigentlichen Intelligenten Raum angrenzt, werden Entwicklung, Fehlersuche und die Verkabelung vereinfacht; In einer marktreifen Lösung wäre dieser gesonderte Bereich nicht mehr nötig, bzw. in einem Nebenraum zu realisieren.

In den folgenden Abschnitten sollen nun drei zentrale Systeme des Raumes, das Personen-Trackingsystem, das Gesten-, sowie das Spracherkennungssystem näher betrachtet werden.

Das Personen-Tracking System

Der Intelligente Raum am MIT kann zur Zeit die Position von bis zu vier sich bewegenden Personen überwachen. Diese Fähigkeit ist zentral für viele Funktionen, die der Raum anbietet: Der Raum kann die Anzahl anwesender Person, so wie ihr Eintreten oder Verlassen detektieren und Informationen über die Position einer Person im Raum können zum Beispiel dazu genutzt werden, um Informationen auf jenem Bildschirm darzustellen, in dessen unmittelbarer Nähe sich die Person aufhält. Des Weiteren kann aus den gewonnenen Kontextinformationen abgeleitet werden, welche Tätigkeit eine bestimmte Person gerade ausübt (z.B. dass die Person gerade eine Präsentation durchführt). Dies führt unter anderem zu interessanten Synergien in Verbindung mit dem Gesten- und Spracherkennungssystem, welche aber in einem der folgenden Abschnitte angesprochen werden sollen. Zusätzlich besitzt der Raum Kameras, die beweglich montiert sind und vom Trackingsystem kontrolliert werden. Diese können dazu verwendet werden, um optimale Filmaufnahmen der anwesenden Personen zu erstellen (z.B. um eine Präsentation oder virtuelle Konferenz ins Internet zu speisen).

Realisiert wird das System mittels zwei stationären Kameras, deren Bilddaten synchronisiert und in ein künstliches neurales Netz gespiesen werden, welches dann ein dreidimensionales Modell des Raumes und seiner Einrichtung und den anwesenden Personen erstellt. Zwei Kameras sind nötig, um ein stereoskopisches Bild zu erzeugen, sprich die räumlichen Koordinaten von Gegenständen und Personen zu ermitteln. Die Implementierung mittels eines künstlichen neuronalen Netzes macht sich den Umstand zu Nutze, dass die Einrichtung des Raumes im Vergleich zu den sich bewegenden Anwesenden mehr oder weniger statisch ist und die Kameras stationär sind. So kann das Trackingsystem ein Modell des Raumes berechnen und dann mit den in Echtzeit ermittelten Bildinformationen vergleichen. Wenn eine Person den Raum betritt, wird ein Farbhistogramm ihrer Kleidung

ermittelt und gespeichert. Auf diese Art und Weise kann die Person später von anderen unterschieden werden.

Obschon das System in der vorliegenden Form bereits sehr leistungsfähig ist, sind an dieser Stelle doch einige Einschränkungen zu nennen: Durch die Verwendung eines künstlichen neuronalen Netzwerkes ist das System sehr anfällig gegenüber Änderungen an der Kamerainfrastruktur. Wenn beispielsweise eine der beiden Kameras um nur einen Zentimeter verschoben wird, so muss das künstliche neurale Netzwerk komplett neu trainiert werden, bevor das Trackingsystem wieder funktioniert. Dieser Trainingsvorgang bedingt einiges Fachwissen, was die Installation aufwendig und teuer und somit für technisch nicht versierte Anwender unpraktikabel macht. Das System ist ausserdem sehr anfällig auf Änderungen der Umgebungsbeleuchtung. So kann beispielsweise zusätzliches Licht, das durch eine geöffnete Tür oder ein Fenster einfällt das ganze System massiv stören oder Veränderungen im Schattenwurf können z.B. die Farbhistogramme der Personen, die sich im Raum aufhalten, in intolerablem Ausmasse verändern. Mittels geschickter Ausleuchtung des Raumes und dem Einsatz spezieller Materialien (z.B. ein dunkler Teppich, der Schattenwurf eliminiert), sowie aufwendiger Farbkorrektur-Algorithmen können diese Auswirkungen jedoch bereits heute ein Stück weit kompensiert werden.

Das System zur Gestenerkennung

Gesten sind wichtige Informationsquellen zur Kontextbestimmung. Menschen tendieren zum Beispiel dazu, stets auf jene Dinge zu zeigen, über die sie gerade sprechen. Der Intelligente Raum am MIT trägt diesem Umstand Rechnung, indem er diese Zeigebewegung erkennen und verwerten kann. Der Raum besitzt zwei LCD-Videoprojektoren, welche Bilder nebeneinander an die Wände des Raumes projizieren. Die projizierten Bilder erlauben nun sowohl das Zeigen mit blosser Finger, wie auch mittels Laserzeigern, wobei letztere

res eine höhere Auflösung und somit feinere Gesten erlaubt. So ist es beispielsweise möglich, dargestellte Objekte auszuwählen und mittels Drag'n'Drop zu verschieben. Desweiteren können die Projektionen als virtuelle Whiteboards genutzt werden. Markierungen und Zeichnungen können von Hand angebracht, gespeichert und später wieder abgerufen werden. In einem späteren Schritt soll die Auflösung verbessert und dann ebenfalls Handschriftenerkennung eingesetzt werden.

Das Spracherkennungssystem

Wie eingangs erwähnt, soll der Mensch im Zentrum der Interaktion stehen, und nicht der Computer. Dies bedeutet, dass die Interaktion auf eine Art und Weise vonstatten gehen soll, die für Anwender selbstverständlich ist und die ihn nicht an einen Computerbildschirm fesselt. Was würde sich hier besser eignen, als die natürliche Sprache?

Die schlichte Verwendung von Sprachbefehlen zur Steuerung des Systems führt aber noch nicht automatisch zu einer intuitiven Bedienung. So war es von vornherein erklärtes Ziel der Forscher am MIT nicht lediglich die Tastatur und Computermaus durch eine Sprachsteuerung zu ersetzen. Es sollten nicht einfach nur Befehle, die zuvor z.B. mit einem Mausklick ausgewählt wurden, an das System gerichtet werden, sondern eine intuitivere Art der Verständigung ermöglicht werden. Das System selber sollte so zurückhaltend wie möglich verbale Kommunikation einsetzen, um die Aufmerksamkeit des Benutzers nicht unnötig in Anspruch zu nehmen. Dies bedeutet konkret, dass das System den Empfang von Kommandos in der Regel lediglich mit einem kurzen unaufdringlichen Geräusch quittiert und lediglich dann Sprachausgabe verwendet, wenn eine explizite Kommunikation erforderlich ist. Wenn der Benutzer beispielsweise das Kommando "Computer, zeig mir meine Emails!" an das System richtet, so genügt es, wenn ein kurzes Klickgeräusch als Bestätigung ausgegeben wird und die Emails ansonsten kommentarlos angezeigt werden.

Es ist nicht nötig oder gar sinnvoll, dass das System mit "In Ordnung Michael, ich zeige Dir deine Email." antwortet.

Personen im Intelligenten Raum tragen kabellose Mikrophone um eine maximale Bewegungsfreiheit zu gewährleisten. Das System selber verwendet zwei verschiedene Spracherkennungsalgorithmen, wovon ein jeder spezifische Stärken und Schwächen besitzt. Eine Kombination zweier verschiedener Spracherkennungsalgorithmen erlaubt es, eben diese Schwächen auszugleichen und eine robustere Spracherkennung anzubieten. Während das eine System für kontinuierliche Sprache eine gute Erkennungsrate aufweist, liegen die Stärken des zweiten Systems beim Erfassen von einzeln diktierten Worten, gefolgt von einer Pause.

Zentrale Anforderungen an eine Spracherkennungsoftware sind eine hohe Erkennungsrate und ein grosser Sprachumfang - schliesslich soll sich ein Benutzer möglichst frei ausdrücken können und sich nicht spezielle Befehle antrainieren müssen. Hierfür sind entsprechend grosse Grammatiken nötig, welche aber enorme Leistungsfresser sind und Echtzeitverarbeitung schwierig machen. Hier machen sich nun die Forscher am MIT unter anderem Synergien zwischen Spracherkennung und Personentracking zu nutze: Je nach Position und Tätigkeit der Sprechenden Person, laden sie dynamisch entsprechende (Unter-)Grammatiken nach. So ist es beispielsweise nicht plausibel, dass ein Benutzer Kommandos zur Steuerung eines Videorecorders ("zurückspulen", "stop", etc.) gibt, wenn die entsprechende Person gar kein Video abspielt. Dementsprechend muss die zugehörige Grammatik auch nicht aktiv sein. Sobald eine Videokassette eingelegt und abgespielt wird, können die Grammatiken zur Videosteuerung geladen werden. Die Frage "Wie ist dort das Wetter?" macht auch nur Sinn, wenn der Raum zuvor eine Äusserung bzgl. eines geographischen Gebietes erkennen konnte. Wenn die Person allerdings neben einem Bildschirm steht, der gerade eine Landkarte der Stadt Zürich darstellt, so kann sich der

Raum die Information über die Position des Benutzers zu nutze machen und die gewünschte Daten über das lokale Wetter bereitstellen.

The Self-Organizing Desk

Jeder kennt sicher das Problem von sich stapelndem Papier auf dem Bürotisch. Entweder man sortiert alles sorgfältig, versucht nur noch elektronische Dokumente zu verwalten oder man wirft extensiv fort. Ein Mittelweg oder eine Hilfe kann der selbstorganisierende Bürotisch [3] darstellen.

Der selbstorganisierende Bürotisch kann grob als Unterstützungssystem für die Organisation eines Bürotischs angesehen werden. Genauer betrachtet erweitert das System einen herkömmlichen Bürotisch um elektronische unterstützende Informationen zu den darauf abgelegten Blättern in Papierform. Der selbstorganisierende Bürotisch hilft die Informationen aus den Dokumenten wiederzufinden, abzulegen, untereinander zu verlinken oder Planungsaufgaben zu vollziehen.

Ziel des Systems ist es, dass ein Benutzer Anfragen in der Form von "Wo ist der Brief von Hans Meier" stellen kann und ihm dann die Dokumente von Hans Meier einerseits digital gezeigt und andererseits deren Position auf dem Schreibtisch genannt wird. Auch beim Verschieben von Blättern oder ganzen Stapeln soll das System stets in der Lage sein die aktuelle Position zu nennen und so dem Benutzer ein schnelles Auffinden zu erleichtern.

Aufbau des Systems

Das System besteht aus Hardware in Form einer Kamera und verschiedenen Softwaremodulen.

Der selbstorganisierende Bürotisch wird von einer möglichst hochauflösenden Videokamera überwacht. Diese besitzt elektronisch steuerbare Bewegungsmotoren für die X- als auch Y-Richtung, sowie eine ansteuerbare Zoom-Funktion.

Die Software ist in drei Module aufgeteilt:

- intelligente Informationsgewinnung (information extraction)
- ein Darstellungsmodul für verschiedene Ansichten
- eine Benutzeroberfläche (GUI), welches die Interaktion mit dem System ermöglicht

Beim Ablegen eines neuen Papiers auf dem Bürotisch wird das Modul zur Informationsgewinnung aktiv und legt den aus dem Dokument extrahierten Inhalt in einer Datenbank ab. Zusätzlich wird die Position des Blattes, darüber und darunterliegende Blätter, die Ablagezeit und eine Historie über alle Geschehnisse mit dem Dokument auch in der Datenbank gespeichert. Die verfügbaren Operationen sind das Hinzufügen eines Dokumentes, das Entfernen, das Verschieben sowie das Abfragen des Systems via GUI.

Die folgende Abbildung zeigt die vorhandenen Module und deren Verknüpfung untereinander. Sie werden anschliessend einzeln beschrieben.

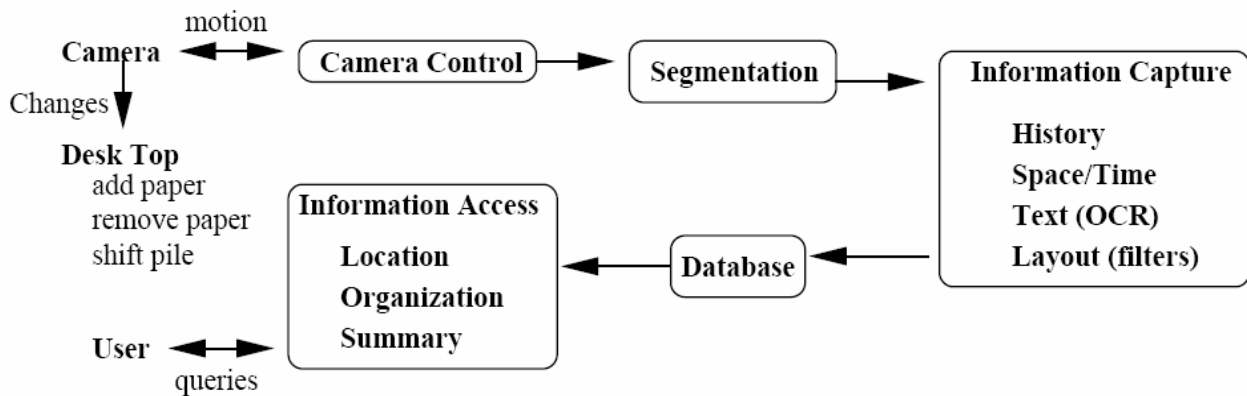


Abbildung 1: Ablaufdiagram für alle Operationen

Überwachungsfunktion (camera control)

Bei der Kamera handelt es sich um eine Canon VC-C1 Kamera, welche mit einem Silicon Graphics Indy verbunden ist. Dieser wiederum ist für die Steuerung der Kamera verantwortlich.

Es wird angenommen, dass sich auf dem Tisch nur Blätter in A4-Grösse befinden und sich immer nur ein Blatt oder ein Stapel zur gleichen Zeit ändert. Die Kamera wandert im Normalzustand über den zu überwachenden Bereich und vergleicht die aufgenommenen Bilder mit denjenigen von einer vorherigen Aufnahme. Stellt die Software, welche das Kamerabild überwacht nun eine Veränderung fest, so werden die Koordinaten der Änderung berechnet. Die Kamera stellt nun auf diese Position scharf und versucht mit geeignetem Zoom ein Bild mit maximaler Auflösung zu schießen. Die Analyse und der Vergleich der Kamerabilder geschehen im Segmentierungsmodul.

Segmentierungsmodul

Das hochauflösende Bild gelangt nun zum Segmentierungsmodul, welches versucht den Inhalt zu extrahieren. Es hat sich herausgestellt, dass das Auslesen von Informationen aus einem an und für sich dreidimensionalem Bild eine schwierige Aufgabe ist. Probleme für

das Erkennen des Bildinhaltes stellen insbesondere Rauschen oder die Verdeckung dar. Rauschen kann durch Lichteinfall in den Raum entstehen oder auch nur schon, wenn der Benutzer ein Blatt berührt oder nur ganz wenig verschiebt. Der Algorithmus sollte ausserdem in der Lage sein ein erkanntes Dokument auch dann noch für dasselbe zu halten, wenn es verdeckt wird. Dies ist wichtig, weil die selbstorganisierende Arbeitsfläche nicht eine real-time Überwachung der Dokumente an sich vornimmt - zwar durch die Überwachung aktiviert wird - dann aber die sich auf dem Tisch befindlichen Dokumente statisch zu erfassen versucht.

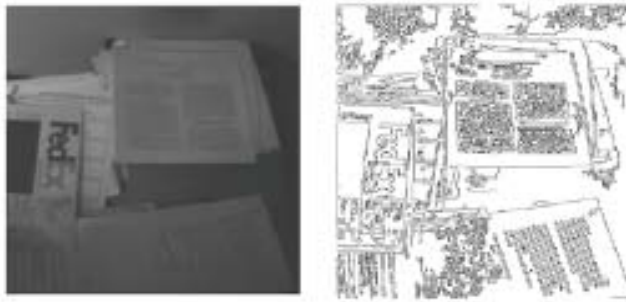


Abbildung 2: Bild vor und nach der Aufbereitung durch das Segmentierungsmodul

Als erstes muss natürlich der genaue Bereich festgelegt werden. Ist nur ein Papier neu auf das Pult gelegt worden, so dürfte diese Aufgabe relativ einfach sein. Lagen aber zum Beispiel bereits 2 Blätter dort und es wird ein drittes so hinzugelegt, dass sich die Blätter teilweise überlappen, so muss erst erkannt werden, wo die Grenzen der einzelnen Blätter sind.

Hierfür wird ein statistisches Programm verwendet, welches nach Kanten im Bild sucht. Aufgrund verdeckter Ecken, den weissen Seitenrändern kann so genau festgestellt werden, innerhalb welcher Koordinaten ein Dokument liegt und welcher Teil des Bildes dazu gehört und welcher nicht. Nun kann innerhalb des festgelegten Teiles das Texterfassen versucht werden.

Mit Erfassen ist gemeint, dass ein OCR-Modul das Bild durchläuft und den Text zu extrahieren versucht. Die dadurch erreichten Resultate sind gemäss dem von mir gelesenen Paper jedoch zu wenig genau, weswegen andere Techniken zu Hilfe gezogen werden mussten.

Um dennoch zu einem brauchbaren Resultat zu gelangen, wird zuerst das erhaltene Bild in viele Teilbilder zerlegt. Diese Teilbilder werden linear interpoliert und dann vektorisiert, so dass vor allem Ecken und Kanten übrigbleiben. Auch der Text lässt sich, ausser er ist handschriftlich geschrieben, relativ gut rastern. Die Teilbilder werden wieder zusammengefügt und das Gesamtbild nun ins OCR gespiesen zur Extraktion des Textinhaltes. Durch dieses Verfahren konnte die Erkennungsgenauigkeit von 60% auf ganze 95% gesteigert werden. Diejenigen Teile des Dokumentes, welche nicht als Text erkannt werden konnten, werden als Grafiken abgelegt. Zudem wird versucht gewisse Layoutmerkmale gesondert zu erkennen und zu speichern, dazu aber später noch mehr.

Informationsgewinnung

Nachdem durch das obig beschriebene Verfahren der Text hoffentlich möglichst vollständig erkannt werden konnte, soll dieser natürlich auch abgelegt werden. Hierzu wird eine Datenbank verwendet, die auch später für die Abfragen wieder ihren Nutzen findet.

Natürlich will ein Benutzer nicht nur die Dokumente an sich finden können, sondern auch Zugriff auf diverse Meta-Informationen erhalten. So werden in der Datenbank auch Merkmale wie die Ablagezeit, Layoutinformationen, Farbinformationen und Links zu anderen potentiell ähnlichen Dokumenten abgelegt. Jedes Dokument erhält ein eigenes Farbhistogramm, wofür das Dokument in ein 3x3 Raster zerlegt wird. Gespeichert wird dann an welcher Stelle im Dokument welche Farbe auftritt. Ausserdem werden Tabellen und Grafi-

ken als solche erkannt und auch über diese Objekte wird in der Datenbank Buch geführt. Somit werden später Abfragen in der Form "Wo ist das Dokument mit einer orangen Grafik in der rechten oberen Ecke und einer Tabelle in der Mitte?" möglich. Bezüglich der mit dem Dokument gespeicherten Zeitinformation ist zu sagen, dass nicht nur die Zeit der Ablage auf dem Pult, sondern auch jegliche Änderungen, Verschiebungen oder das Entfernen protokolliert werden und in einer Historie zum Dokument angezeigt werden können. Die zur Gewinnung des hochauflösenden Bildes berechnete Position wird ebenso gespeichert. Wichtig ist auch die Information, welches Dokument unterhalb und ev. später oberhalb liegen. Nur so ist es möglich beim Verschieben eines ganzen Stapels von Blättern die Positionen aller Dokumente aktuell zu halten. In diesem Fall würde die Datenbank automatisch für alle betroffenen Dokumente aktualisiert und die Historie zu jedem der Dokumente entsprechend ergänzt.

Informationsabfrage

Mit all den gesammelten Informationen zu einem Dokument kann man natürlich so einiges anstellen. Der selbstorganisierende Bürotisch soll nicht nur zeigen können, wo sich ein Dokument befindet, denn eventuell wurde dieses in der Zwischenzeit ja auch bereits entfernt, sondern er soll es auch ermöglichen weitergehende Informationen zu erhalten. Wie im Abschnitt Informationsgewinnung beschrieben, ist es möglich nach all den gespeicherten Merkmalen zu suchen und sich die Resultate entsprechend anzeigen zu lassen.

Das System ermöglicht es Volltext-Suchanfragen zu stellen, bei welchen alle relevanten Dokumente ähnlich einer Suchmaschine aufgelistet werden. Speziell am selbstorganisierenden Bürotisch-System ist jedoch, dass von jedem Dokument auch auf ähnliche Dokumente verwiesen wird, ganz ähnlich einem Threading in einem Forum, wo alle Nachrichten zu einem Thema geordnet erscheinen. Die Verlinkungstechnik basiert auf der Annahme,

dass ähnliche Dokumente auch ähnliche Wörter verwenden. Die Texte der Datenbank und diejenigen der Suchanfrage werden als gewichtete Vektoren abgebildet, dann verglichen und letzten Endes die Suchresultate daraus generiert. Mittels einem "star algorithm" genannten Verfahren wird versucht die Dokumente nach Themen zu gliedern. Somit wird es auch möglich Subthemen zu erkennen, was wiederum hilft Links auf ähnliche Dokumente zu erstellen.

Nebst Volltext lässt sich auch nach Dokumenteigenschaften suchen, welche sein können: Farbe, Layout, oder auch die Position auf dem Schreibtisch. Natürlich lassen sich auch alle Suchargumente mit booleschen Operatoren verbinden, wodurch komplexe Suchanfragen möglich werden. Für die Behandlung der Suchanfragen wurde eine abgewandelte Version von Smart verwendet. Smart ermöglicht es auch mit den durch das OCR-Modul teils unvollständig erkannten Texten umzugehen und dennoch gute Suchresultate zu liefern. Die verwendeten Verfahren, um die Ähnlichkeit zwischen Dokumenten zu erkennen, ermöglichen es dem System auch eine nach Themengebiet gegliederte Zusammenfassung über die Dokumente zu erstellen.

Visualisierung

Um die erhaltenen Suchresultate besser darstellen zu können wurden drei mögliche Visualisierungsoptionen implementiert. Einerseits kann die Position des Dokumentes auf dem Schreibtisch in einer dreidimensionalen Grafik veranschaulicht werden, was das rasche auffinden ermöglicht. Andererseits wird ein Graph erzeugt, welcher die Ähnlichkeit zwischen den Dokumenten visualisiert. Die dritte Option zeigt die Themengliederung auf dem Schreibtisch optisch an. Alle diese Ansichten sind in der Interaktionsmaske (dem GUI) für den User ersichtlich und sind untereinander verknüpft. Das Anwählen eines Dokumentes in einer Ansicht markiert dieses auch gleich in den anderen beiden Ansichten.

Probleme des Systems

Der selbstorganisierende Schreibtisch wurde in der vorhandenen Spezifikation mehrfach getestet, dabei kamen verschiedene Probleme zu Tage. Einerseits ist das System relativ träge und langsam, was das Erkennen einer Veränderung auf dem Schreibtisch angeht. Andererseits braucht es sehr lange um die erkannte Änderung zu bearbeiten, genannt werden 15 Minuten Rechenzeit. Diese allerdings dürfte heute um einiges tiefer liegen. Das Update oder das Abfragen der Datenbank hingegen ist einer Sekunde oder weniger möglich. Weitere Störfaktoren für ein zuverlässiges Arbeiten mit dem selbstorganisierenden Schreibtisch sind die zu schlechte Auflösung der Kamera als auch die ungenaue Texterkennung mittels OCR. Auch hier denke ich, dürfte man heute um einiges weiter fortgeschritten sein, gerade was die Auflösung der Kamera betrifft. Ebenfalls besteht ein Problem darin, dass sich überlappende Blätter, bei sehr kleiner Überlappung nicht mehr unterscheiden lassen. Genannt wird ein Grenzwert von 1 Zentimeter, was vor allem mit den weissen , resp. leeren Seitenrändern zu tun haben dürfte.

Zukunftsansichten und Schlussfolgerung

Im Paper, das aus dem Jahre 1997 stammen dürfte (eine genaue Jahreszahl fehlt leider) wird aufgrund der Restriktionen der Kamera und des OCR empfohlen, dass die Dokumente für das System per Scanner anstatt mittels Kamera eingescannt werden. So empfehlen sie das System für eine automatisierte Archivierung von Dokumenten, also eher ein intelligenter Aktenschrank denn ein selbstorganisierender Schreibtisch. Dies mag für damalige Zeiten womöglich zutreffen, ich denke, dass mit heute vorhandener und besserer Hard-

ware das System sicherlich realisierbar wäre oder zumindest viel bessere Trefferquoten und Erkennungsraten erzielt werden könnten.

Im Rahmen des "Intelligent Room"- Projektes [4] wurde eine interessante Erweiterungsmöglichkeit vorgeschlagen. Eine Idee wäre auf dem Pult aufzuklebende Post-It - Kleber mit gewissen Funktionalitäten zu belegen. Diese wären beliebig einstellbar, z.B. wären die Operationen Licht verdunkeln, Radio einschalten, Instant Messenger auf abwesend stellen alle auf ein Post-It "belegbar", so dass bei der Berührung desjenigen die Operationen ausgeführt werden. Diese Funktionalität eines intelligenten Raumes würde sich sicherlich sehr gut mit dem selbstorganisierenden Bürotisch vereinen lassen, da dieser ja bereits per Kamera überwacht wird.

Ich könnte mir auch vorstellen, dass das System so ausgebaut wird, dass physische Dokumente als auch virtuelle vereint durchsucht werden könnten. Z.B. eine "Google Desktop Search", die sich nicht auf den virtuellen PC-Schreibtisch beschränkt, sondern die Suchmaske des selbstorganisierenden Bürotischs gleich integriert und so eine zentrale für das Verwalten von allen persönlichen Dokumenten schafft.

Fazit

Was wir als wesentliche Vorteile gegenüber anderen System im Bereich "Context Aware Computing" sehen ist, dass beide Systeme "Intelligent Room" als auch der "selbstorganisierende Bürotisch" keine Voreinstellungen benötigen, keine Geräte am Körper getragen werden müssen, sondern die reale Welt plötzlich durch intelligente Systeme unterstützt wird, ohne dass sich der Benutzer gross umgewöhnen muss. Die Sensoren sind Kameras, welche aktiv überwachen und die Systeme aktivieren. Die Steuerung könnte sicher auch

beim unterstützenden Schreibtisch noch so ausgebaut werden, dass sie wie beim intelligenten Raum per natürlicher Sprache ermöglicht wird.

Wie ich am Tag der offenen Türe bei den IBM Forschungslaboratorien in Rüschlikon gesehen habe, gibt es auch schon Arbeitsplätze die genau das realisieren, was unter einem intelligenten Raum verstanden werden kann. Ein Bürotisch mit darauf projizierter Arbeitsfläche, Bildschirme und Wall-Screens die je nach Kontext verschiedene Informationen anzeigen und sich per Sprache oder mittels Berührung von virtuellen Schaltflächen steuern lassen.

Die Zukunft wird sicher vermehrt solch unterstützende Systeme hervorbringen, denn das Arbeiten mit Maus und Tastatur wird sicherlich nicht für Ewigkeiten Standard bleiben.

Literaturverzeichnis

- [1.] Design Principles for Intelligent Environments, Michael H. Coen, MIT Artificial Intelligence Lab, Cambridge, 1998,
<http://people.csail.mit.edu/mhcoen/Papers/aaai98.pdf>
- [2.] "Command Post of the Future", <http://www.defense-update.com/products/c/cpof.htm>
- [3.] The Self-Organizing Desk, Daniela Rus, Peter de Santis, Department of Computer Science, Dartmouth College, 1997, <ftp://ftp.cs.dartmouth.edu/TR/TR97-305.pdf>
- [4.] The Intelligent Room Project, Rodney A. Brooks, MIT Artificial Intelligence Lab, Cambridge, <http://people.csail.mit.edu/brooks/papers/aizu.pdf>

Anhang

Der erste Teil zum Thema "Intelligent Room" wurde von Michael Würsch erarbeitet und beschrieben. Der zweite Teil zum Thema "selbstorganisierender Bürotisch" wurde von Roman Haag erarbeitet und beschrieben. Die Folien für die Präsentation hat M. Würsch erstellt, das Layouting des Berichts hat R. Haag übernommen.