

Computergestütztes Visual Storytelling

Seit mehreren Jahren gewinnt **VISUAL STORYTELLING** als Trend in der Verlagsbranche stetig an Bedeutung. Sei es online, auf mobilen Geräten oder in klassischen gedruckten Zeitungen – die Nachrichten werden nicht mehr ausschließlich mit Text erzählt, sondern großzügig mit Bildern, Infografiken und Illustrationen angereichert. Dem Visuellen wird so in zunehmendem Maße Platz eingeräumt.

Erfreulicherweise steht Journalisten und Bildredaktionen heute mehr hochwertiges Bildmaterial als je zuvor für ihre Medienberichterstattung zur Verfügung. Viele Fotoagenturen schicken heute fünfmal so viele Fotos in die Welt wie noch vor wenigen Jahren. Dementsprechend können sich größere Verlagshäuser über einen täglichen Bildzufluss von mehr als 20.000 Fotos freuen, die den Redakteuren für ihre Arbeit zur Verfügung stehen.

Diese Bilderflut bringt jedoch auch neue Probleme und Herausforderungen mit sich. Schon die schiere Menge der zur Verfügung stehenden Bilder macht eine Sichtung des kompletten Materials unmöglich. Die Potenzialausschöpfung dieser wichtigen Ressource wird jedoch auch durch wirtschaftliche Zwänge erschwert: Laut dem Pew Research Center ist die Anzahl der redaktionellen Mitarbeiter, einschließlich der Bildredakteure, seit 2000 um 35% gesunken. Statt professionelle, ausgebildete Archivare und Bildredakteure einzusetzen, müssen Journalisten und Redakteure immer häufiger selbst die Bildsuche in die Hand nehmen. Der Trend zum Stellenabbau sowie der Druck zur Prozessoptimierung entlang der gesamten Wertschöpfungskette bereiten auf der anderen Seite wiederum den Weg zur Einführung innovativer technischer Lösungen, die nicht nur zu Kosteneinsparun-

gen und einer Effizienzsteigerung führen, sondern auch die optische Anziehungskraft einer Veröffentlichung dramatisch steigern können.

Beim Visual Storytelling geht es aber nicht nur darum, einen Artikel mit beliebigen Fotos zu bebildern. Vielmehr müssen Redakteure die Motive finden, die bei den Konsumenten bereits auf den ersten Blick Emotionen erwecken, die Handlung der Story veranschaulichen, auf Aktionen hindeuten und dabei einen nachhaltigen Eindruck hinterlassen. Die Suche nach passenden Motiven, welche die oben genannten Voraussetzungen erfüllen, scheidet aber häufig an einer mangelhaften und uneinheitlichen Verschlagwortung des Bildmaterials sowie den Einschränkungen einer schlüsselwortbasierten Suche. Weil viele wichtige Informationen zum Bildinhalt, z.B. Datum, Ort und eine Beschreibung des aufgenommenen Ereignisses, meist vom Fotografen in den Bildmetadaten adäquat erfasst werden, müssen Redakteure dennoch nach dem ineffizienten Prinzip Versuch und Irrtum suchen und immer wieder neue Suchbegriffe probieren, um eine zielführende Kombination von Schlagwörtern zu finden.

Obwohl der menschliche Betrachter o.g. Konzepte innerhalb weniger Millisekunden

interpretieren kann, ist eine vollständige und umfangreiche Verschlagwortung mit allen notwendigen kontextbezogenen Daten in der gängigen Praxis nicht oder nur mit erheblichem Aufwand realisierbar. Schon aus wirtschaftlichen Gründen ist es also unmöglich sämtlichen Fotos mit allen notwendigen kontextbezogenen Daten zu verschlagworten, um eine Suche nach semantischen Konzepten zu ermöglichen. Aufgrund dessen ist es nicht überraschend, dass man bei der Suche nach einer bestimmten Gestik, genauen Personenzahl, zwischenmenschlichen Interaktionen oder einem semantischen Konzept schnell an die Grenzen der Volltextsuche stößt. Vor diesem Hintergrund ist es seit Jahren ein ambitioniertes Ziel, technische Systeme zu entwickeln, deren visuelle Leistungen denen des menschlichen Sehsystems, hinsichtlich der Erkennung von abstrakten semantischen Konzepten, gleichwertig sind.

Bislang galt die automatische Erschließung des Bildinhalts bzw. der Schließung der sogenannten „semantischen Lücke“ als äußerst komplexe Aufgabenstellung. Dementsprechend entwickelt PAMA Technologies maschinelle Lernverfahren, die bestimmte Aspekte menschlicher Wahrnehmung nachahmen können. Sie versetzen somit Maschinen in die Lage wesentliche Merkmale wie bspw.

die Körperhaltung, die Armposition oder ein Handzeichen zu extrahieren und ihre gegenseitigen Bezüge einzuordnen. Im Zuge dessen lernen Maschinen zwischen diversen Körperhaltungen und Handgesten zu unterscheiden, um durch eine Analyse des Gesamtbildes, eine semantische Bedeutung wie Jubeln, Winken oder Zeigen zu identifizieren.

Darüber hinaus ist es mittlerweile ebenfalls möglich, zwischenmenschliche Interaktionen, zum Beispiel einen Zweikampf im Bereich Sport oder das Händeschütteln zweier Personen auf einem Gipfeltreffen, im Bild zu erkennen. Mittels neuartiger Gesichtserkennungsverfahren können auch Prominente und Politiker vollautomatisch im Bild ausgemacht werden. Durch eine automatisierte Verschlagwortung mit ent-

sprechenden semantischen Metadaten ist es jetzt realisierbar, Fotos zu finden, bei denen sowohl die Person also auch deren Emotionen, Bewegungen und Handlungen im Vordergrund stehen. Eine automatische Logoerkennung ermöglicht überdies, z.B. im Falle eines Skandals oder einer Produkteinführung, die Bildsuche nach Personen in Verknüpfung mit entsprechenden Markenzeichen. In der Vergangenheit wären solche Suchanfragen nahezu unmöglich gewesen.

Neben den personenbezogenen Konzepten sind die lernfähigen Systeme von PAMA auch in der Lage das Setting eines Bildes durch eine inhaltsbasierte Analyse zu erkennen und somit zwischen Tages-, Nacht-, Außen- oder Innenaufnahme zu unterscheiden. In einer maßgeschneiderten Visual Search-Oberfläche können die

erkannten Bildinhalte mit diversen Visual Filtern entsprechend verknüpft werden, um sekundenschnell optisch und inhaltlich anziehende Fotos zu liefern. In Industrie und Forschung profitiert die Bildverarbeitung demnach längst vom Einsatz unterschiedlicher Technologien, welche versuchen, die intelligenten Fähigkeiten eines Menschen auf Maschinen zu übertragen. Steigende Performance von Hard- und Software, höhere Rechenleistungen und der Einsatz von Algorithmen, die die Organisations- und Verarbeitungsprinzipien des menschlichen Gehirns imitieren, werden die Lernfähigkeit und Robustheit automatisierter Bildverarbeitungssysteme in den kommenden Jahren noch einmal beträchtlich steigern.

Diese technologischen Fortschritte im Bereich des maschinellen Sehens tragen somit enorm zu Verkleinerung der semantischen Lücke bei. Vielmehr noch stellen sie schon heute im Bereich des Verlagswesens einen Paradigmenwechsel dar, welcher das gesellschaftliche Verlangen nach einer stetig zunehmenden Visualisierung jeglicher Lebensbereiche, allem voran dem Visual Storytelling, bedient. <



GETTY IMAGES · PAMA · THOMAS BARWICK · 530281095



GREG FAWSON
 Leiter Business
 Development
 PAMA
 Technologies GmbH
www.blindtext.de