

Joachim Griesbaum, Thomas Mandl,  
Christa Womser-Hacker (Hrsg.)

# Information und Wissen: global, sozial und frei?

Proceedings des 12. Internationalen Symposiums  
für Informationswissenschaft (ISI 2011)

Hildesheim, 9.–11. März 2011

**vwh**

Verlag Werner Hülsbusch  
Fachverlag für Medientechnik und -wirtschaft

# Ein erweiterbares Tool zur Annotation von Videos

*Marc Ritter, Maximilian Eibl*

Technische Universität Chemnitz – Professur Medieninformatik

09107 Chemnitz

{marc.ritter|eibl}@informatik.tu-chemnitz.de

## **Zusammenfassung**

Die Beschreibung audiovisueller Medien durch Metadaten bedarf aufgrund der Masse der Daten einer Unterstützung für die manuelle Annotation sowie eine möglichst weitgreifende automatische Annotation. Das vorgestellte Werkzeug unterstützt Markierung, interaktive Vorsegmentierung und Objekterkennung und -verfolgung. Mithilfe der integrierten *Shot Detection* kann ein Video strukturell zerlegt werden, um beispielsweise die halbautomatische Objektverfolgung über Szenengrenzen hinweg zu unterbinden. Die Anbindung des Forschungsframeworks AMOPA gestattet dem Nutzer beliebige Ketten von Operationen zur Bild- und Videoverarbeitung zu erstellen. Der Datenexport erfolgt über MPEG-7-DAVP.

## **Abstract**

We introduce an extensible annotation tool for videos. It allows marking, interactive segmentation and semi-automatic tracking of objects in video images as well as preprocessing steps like shot boundary detection. The combination with the research framework AMOPA enables the user to create arbitrary custom image processing chains depending on the underlying field of research. All data may be exported to MPEG-7-DAVP.

# 1 Einführung

Systeme zum Einsatz audiovisueller Medien, seien es Medienarchive, Mediatheken, Knowledge Management Systeme, E-Learningsysteme oder andere sehen sich in den letzten Jahren einer zunehmenden Flut von Daten gegenüber. Diese durch Beschreibungsdaten sinnvoll recherchierbar zu machen, entwickelt sich zu einer immer stärkeren Herausforderung an Kapazität und Zeit. Hier greifen Versuche, die Annotationsprozesse weitgehend zu automatisieren bzw. die Anwender in der Annotation zu unterstützen. So werden Algorithmen der Audio-, Bild- und Video-Verarbeitung speziell im Anwendungsgebiet des Maschinellen Lernens entwickelt und eingesetzt, indem Klassifikatoren zum Auffinden von Objekten trainiert werden. Deren Training erfordert jedoch oftmals zuvor aufbereitete, d.h. annotierte Trainingsdaten, die bereits Beispiele der zu klassifizierenden Objekte enthalten. Die Annotation derartiger Objekte in Bild- und Videodaten erfolgt zumeist manuell, ist mit wiederholten Arbeitsprozessen verbunden, zeitaufwendig und verlangt zudem eine hohe menschliche Aufmerksamkeit. Einfache, schnelle und zuverlässige Annotationsvorgänge können die Entwicklungszeit von Algorithmen positiv beeinflussen sowie die Aussagekraft über die Performanz von wissenschaftlichen Verfahren auf größeren Testdatensätzen erhöhen.

## 1.1 Related Work

In der Vergangenheit wurden zahlreiche Tools entwickelt, die Annotationen in Bildern und Videos erleichtern sollten. Das ViPER-Toolkit (Doermann & Mihalcik 2000) erlaubt die Markierung von Einzelobjekten in Videos und die Kreation von Deskriptoren mittels Schema-Editor und findet bei der *TREC Video Retrieval Evaluation* Anwendung (Smeaton 2006). Die Handlung von Bildinhalten lässt sich in *Caliph* mit einem gerichteten Graphen erfassen, jedoch nicht mit einem Bildobjekt vereinen. Die Komplementäranwendung *Emir* findet ähnliche Bilder mithilfe von MPEG-7 Low-Level-Deskriptoren (Lux 2009). Mit diesen durchsucht (Vezzani 2006) Videos nach zuvor trainierten Objekten. Deskriptoren finden auch beim *M-OntoMat-Annotizer* Anwendung, der es erlaubt, in Bildern Regionen zu segmentieren, um aus diesen dann Objekte und deren Eigenschaften zu bestimmen (Petridis 2006). Das *Multimedia Analysis and Retrieval System* (MARVEL) tauscht die übliche Annotationszeichenfläche gegen statistische Analysealgorithmen und ordnet

vorhandene Bilder in eine komplexe Taxonomie ein, um so die Fehlerrate einer zeitaufwendigen manuellen Annotation zu minimieren (Columbia 2004). Einen ähnlichen Weg verfolgt das *VideoAnnEx Annotation Tool*. Es hält vordefinierte Lexika zur Annotation von Schlüsselobjekten, Ereignissen und statischen Szenen bereit (Naphade 2002). In einer MATLAB-GUI verbindet (Kounoudes 2008) eine Szenenwechselerkennung mit der Möglichkeit Frames mit Kategorien erweiterbarer XML-Wörterbücher semantisch zu taggen. Eine semi-automatische Segmentierung von Objekten durch aktive Konturen und deren Verfolgung durch Interpolation verfolgt (Luo & Eleftheriadis 1999). (Goldmann 2008) modifiziert ein 2D-Trackingverfahren, um Bildobjekte zu transformieren und deren Position im Video zu ändern. Ursprünglich zur Annotation von Sprachdialogen entwickelt, gestattet *ANVIL* die Definition und Annotation diverser Kodierungsschemata auch für die Videoverarbeitung. Daten können XML-basiert abgespeichert und in der statistischen Toolbox SPSS weiterverarbeitet werden. (Kipp 2008) erweitert die Kodierungsschemata um eine räumlich-zeitliche Funktion zur Markierung von Objekten über den eigenen Frame hinaus. Das für Mac OS X verfügbare Tool *VCode and VData* ermöglicht die Verarbeitung multipler Videoquellen und verschiedener Sensorformate mittels Objektmarkierung auf einem Zeitstrahl. Als Besonderheit können Annotationssequenzen syntaktisch auf Konsistenz geprüft werden (Hagedorn 2008). Einen anderen Weg schlägt (Schroeter 2008) mit *FilmEd* ein, wo mehrere über heterogene Netzwerke verbundene Nutzer interaktiv gleichzeitig Annotationen durchführen können. Selbst mobile Geräte werden inzwischen mit allerdings sehr rudimentären Annotationsfunktionen ausgestattet (Concejero 2008).

## 1.2 Motivation

Zahlreiche Anstrengungen wurden unternommen, den mühsamen und weit verbreiteten Prozess der Frame-by-Frame-Annotation zu erleichtern. Einerseits ist es möglich, die Objektsegmentierung und -verfolgung zu automatisieren, was nur von wenigen Tools angeboten wird. Andererseits werden mehrere Bilder einer Sequenz mit einem semantischen Begriff versehen. Obgleich die Begriffswahl durch maschinelle Vorverarbeitung eingegrenzt werden kann, leidet doch oftmals die Exaktheit der Annotation aufgrund der Spezifität der zugrundeliegenden Wissenstransferfunktion. So versucht (Fan 2008) die semantische Lücke zwischen Low-Level-Merkmalen und High-Level-Konzepten zu schließen, indem saliente Objekte in Bildern er-

kannt und mit atomaren Bildkonzepten konzeptbasiert abgeglichen werden. In vereinfachter Form ist dies auch in der Videoannotation wünschenswert.

Zudem lassen einige Tools die Möglichkeit zum Datenexport zur anwendungsspezifischen Weiterverarbeitung vermissen. Das vorliegende Annotationstool versucht nun die Vorteile zu vereinen. Dazu wird es an das effiziente und flexible Lehr- und Forschungsframework AMOPA (Ritter & Eibl 2009) angebunden, dessen Verarbeitungsketten auf Geschwindigkeit optimiert sind und sich bei Bedarf skalierbar auf andere Rechner über *Java RMI* verteilen lassen. Somit können Videos nicht nur schnell strukturell zerlegt, sondern auch Objekte markiert, segmentiert und verfolgt werden. Da Videos im Gegensatz zu Bildern in der Regel Bewegungen und unterschiedliche Kameraperspektiven aufzeigen, lassen sich Variationen von Objekten leichter erfassen, erhöhen die Anzahl verfügbarer Trainingsdaten und präzisieren zugleich deren Qualität für die Weiterverarbeitung. Einhergehend verringert sich der Annotationsaufwand.

## 2 Systembeschreibung

Das Annotationstool besteht aus zwei Komponenten. Erstere bildet das Framework *Automated MOving Picture Annotator* (kurz: AMOPA), das (Ritter 2009) ausführlich beschreibt. Es wurde als Lehr- und Forschungsinstrument zum Entwurf und zur Implementierung von beliebigen prozessgesteuerten Workflow-Konzepten konzipiert. Der Zugriff auf Videodaten wird durch die offene C-Bibliothek FFMPEG und über das Interface von *Streambaby* in Java gekapselt (siehe Abb. 1).

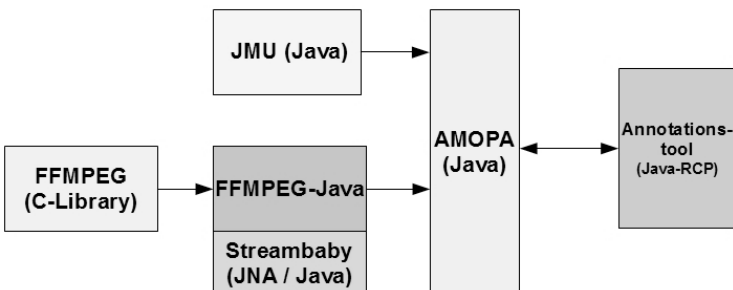


Abbildung 1: Architektur des Annotationstools (rechts). Dieses tauscht Daten mit AMOPA (Mitte) aus, welches die Funktionen von JMU und Streambaby aggregiert.

Neueste Erweiterungen des Prozesskonzeptes von *Java Media Utility* (kurz: JMU) erlauben die Verwirklichung von Prozessketten mit nicht-linearem Verlauf, womit Workflows nahezu beliebiger Topologie umgesetzt werden können.

Die Verschaltung der einzelnen Bildverarbeitungsschritte erfolgt in der Prozesskette wahlweise per Editor oder XML, wobei jeder Schritt als eigenständiger Thread gestartet wird, um Multi-Core-Rechnerarchitekturen besser zu unterstützen. Die zweite Komponente umfasst das eigentliche Annotationstool. Durch den Einsatz von Java-RCP als Entwicklungsplattform ist das Tool portierbar und unter Windows und Linux anwendbar. Eine schnelle Bilddarstellung wird durch *JFace* und *SWT* erreicht. In Abb. 2 wird die grafische Benutzungsoberfläche gezeigt.



Abbildung 2: Darstellung des Annotationstools mit einer bereits halbautomatisch vorsegmentierten Person in der Mitte (weiße Polygonlinie).

Sie enthält zunächst die Menüleiste mit Funktionen zum Laden, Speichern, Exportieren, Szenenwechselerkennung, Objektsegmentierung und -verfolgung. Unter der Menüleiste befinden sich Buttons zur manuellen Markie-

rung von Objekten. Hier stehen Rechteck, Kreis, offenes sowie geschlossenes Polygon zur Verfügung. Um die Sichtbarkeit der Markierung in jedem beliebigen Videomaterial zu gewährleisten, können Strichstärke und -farbe individuell gesetzt werden. Der aktuelle Annotationsframe ist in der Mitte dargestellt. Links reflektiert der *SceneView* in einer Baumstruktur die dazugehörige durch Annotationen erzeugte Objekthierarchie. Im unteren Bereich der GUI schließt sich das *ControlView* an, das über entsprechende Steuerelemente die Navigation in Videos ermöglicht. Der Zugriff auf Einzelbilder wird sequentiell in der oberen Zeile ermöglicht. Nach Ausführung der Szenenwechselerkennung kann zu den Einzelbildern einer kontinuierlichen Kameraeinstellung (engl. *shot*) über Klicken der repräsentativen Schlüsselbilder (unterste Zeile) navigiert werden. Zu annotierende Frames werden durch einen Doppelklick in die Annotationszeichenfläche transferiert.

## 3 Algorithmen

Die nachfolgenden Abschnitte geben einen Einblick in die im Annotations-tool verwendeten Algorithmen. Deren Umsetzung erfolgt in separaten Verarbeitungsketten ausschließlich durch AMOPA. Gesteuert werden sie über eine Schnittstelle gemeinsamer Objekte.

### 3.1 Navigation in Videos

In der Navigation kann unter allen Abspielfunktionen die Sprungfunktion an eine bestimmte Position des Videos unter Umständen problematisch erscheinen, da sie die sequentielle Abtastung (Streaming) eines Videos durchbricht. Das trifft in der aktuell angebundenen Version von FFMPEG besonders auf das MPEG-Format zu. Wird beim Ladevorgang ein solches Format festgestellt, werden alle Videoframes momentan serialisiert als Java-Objekte auf Festplatte ausgelagert. Im Gegensatz dazu funktioniert das Springen im Container-Format AVI reibungslos und benötigt lediglich die Verwendung eines internen Ringpuffers.

### 3.2 Szenenwechselerkennung

Die Detektion von Szenenübergängen erfolgt in Anlehnung an das Verfahren von (Liu 2006), nutzt jedoch nur wenige ausgewählte Merkmale zur Detektion von harten Schnitten (vgl. (Ritter & Eibl 2009)) und erreicht somit eine Verarbeitungsgeschwindigkeit von bis zu 2,2-facher Echtzeit (Testsystem: Dual Quad Core, 3 GHz mit Ø 65% Prozessor-Auslastung) unabhängig von der Auflösung des Videomaterials. Dazu wird das Bild zuerst auf halbe PAL-Auflösung skaliert und in  $48 \times 48$  große Blöcke zerlegt. Zu jedem Block werden die individuellen Bewegungsvektoren bestimmt. Der Fehler zwischen dem eigentlichen Block und seinem bewegungskompensierten Pendant ergibt sich über die minimale absolute Distanz aller Pixel. Der Quotient über das Verhältnis aus kumulativer Summe der Fehler aller Blöcke und einer additiven Glättungsfunktion über die Fehlersumme vergangener Bilder löst letztlich die Detektion eines Szenenübergangs aus. Gegenüber der Evaluation von (Zwicklbauer 2010) wurde die Genauigkeit des vorliegenden Verfahrens nochmals verbessert, indem MPEG-7-Kantenhistogramme innerhalb einer  $\varepsilon$ -Umgebung eines wahrscheinlichen Szenenwechsels auf je fünf benachbarten Frames berechnet und auf Unähnlichkeit geprüft werden. Auf dem getesteten Realmaterial kooperierender Fernsehsender ( $> 100$  Stunden) erreicht dieses Verfahren in eigenen Testreihen Detektionsquoten um die 99% mit einer Falsch-Positiv-Rate von max. 1,5%.

### 3.3 Objektsegmentierung

Die Erfassung von genauen Konturen eines Objektes kann für dessen Merkmalsextraktion und Weiterverarbeitung von entscheidender Bedeutung sein. Die manuelle Annotation von beliebigen Objekten mittels Freiform- oder Polygon-Tools ist oftmals zeitintensiv. Das Annotationstool verwendet zur semi-automatischen Segmentierung die Implementierung des *GrabCut* nach (Talbot & Xu 2006). Zur Anwendung genügt es, die zu segmentierende Fläche mit einer Bounding-Box geschlossen zu umranden. Zwei Gaußsche Mischverteilungsmodelle (GMM) werden für Vorder- und Hintergrund modelliert. Die Pixel innerhalb der Markierung werden dem Vordergrund-GMM zugeführt. Das Farbreduktionsverfahren nach (Orchard & Bouman 1991) bestimmt die Verteilungen innerhalb der GMMs. Danach werden alle Vordergrundpixel der wahrscheinlichsten GMM zugeordnet. Aus der neuen Pi-



xelverteilung wird mit den beiden GMMs als Quelle und Senke ein Graph aufgebaut und der minimale Schnitt nach (Boykov & Kolmogorov 2004) berechnet. Dieser Prozess wird iteriert bis sich die Pixelzuordnungen zur jeweiligen GMM nicht mehr ändern. Die Ergebnisse des Verfahrens visualisieren Abb. 3 und Tab. 1. Die Methode erzielt akzeptable Resultate, wenn die Farbunterschiede innerhalb des Objektes nicht zu groß sind. Probleme bereiten ausgeprägte Muster, Schatten und inhomogener Hintergrund, der keine scharfkantige Abgrenzung des Objektes zulässt.

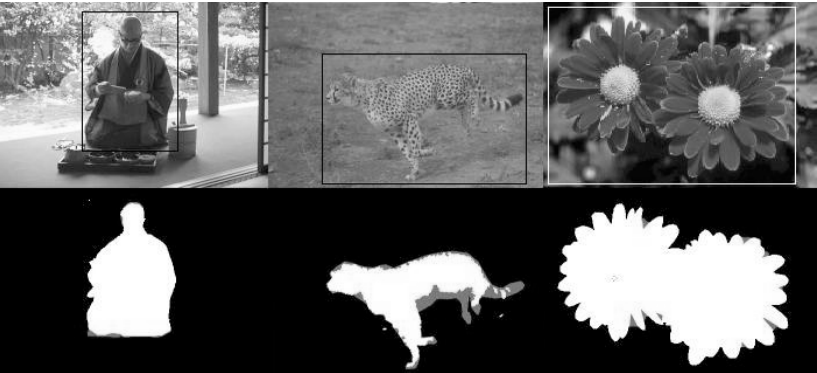


Abbildung 3: Evaluation der Genauigkeit des GrabCut-Algorithmus: Originalbilder mit rechteckiger Objektselektion (oben), Unterschiede zum manuell erstellten Ground Truth (graue Bereiche im unteren Bild). (modifiziert aus: Höhlig 2010)

Tabelle 1: Genauigkeitsanalyse GrabCut (Angaben in Pixel) (aus: Höhlig 2010)

Bild aus Abb. 3	Gesamtfläche	GrabCut-Segmentation	Manuelle Segmentation	Fehl-detektionen	Fehlerquote
Buddhist	151.526	24.826	24.501	575	2,35%
Leopard	150.416	19.973	24.510	4.669	19,05%
Blume	152.044	67.627	68.259	632	0,93%

### 3.4 Objektverfolgung und Datenexport

Das Annotationstool bietet die Möglichkeit, markierte Objekte wahlweise durch Angabe einer frei zu definierenden Anzahl von Bildern oder bis zum Ende des aktuellen Shots zu verfolgen. Dazu wird ein entsprechendes Objekt mit einer Bounding-Box markiert (Abb. 4 links). Das weiße Rechteck bildet den Rahmen des Suchfensters, in dem das zu verfolgende Objekt innerhalb

des nächsten Bildes aufgefunden werden sollte. In Anlehnung an (Beck 1999) wird ein einfaches Block-Matching-Verfahren mit  $n$ -Schrittsuche unter Nutzung minimaler absoluter Distanz angewendet. Der Export aller Annotationen erfolgt bisher wahlweise in das Format MPEG-7-DAVP gemäß (Bailer 2007) als *Plain Text* oder über frei konfigurierbares XML. Eine Konvertierungsfunktion in das gegenüber MPEG-7 hierarchisch flachere ViPER-Dateiformat (Doermann & Mihalcik 2000) ist angestrebt.



Abbildung 4: Tracking des markierten Objekts (schwarze Bounding-Box) im Initialbild (links). Resultat der automatischen Objektverfolgung 24 Frames später (rechts). Die Videosequenz stammt aus einer Überwachungssequenz aus TRECVID 2009 (Smeaton 2006). (entnommen aus: (Höhlig 2010))

## 4 Ausblick

Das beschriebene Annotationstool wird in aktuellen Forschungsarbeiten stetig weiter entwickelt. Dies umfasst sowohl die Bereitstellung einer zukünftigen Ontologie- und Thesauri-Funktion zur vereinheitlichten und konsistenten Kennzeichnung von markierten Objekten als auch Suche und Navigation nach annotierten Objekten mittels eines Objektbrowsers über Frame- und Shotgrenzen hinweg. Eine weniger speicherintensive Lösung zur Bearbeitung

von MPEG-Dateien ist über Ankopplung von externen Tools zur Indexierung der *IFrames* möglich. Die Einbindung von Algorithmen zur *Shot Composition*, die Dialoge und Nachrichtenszenen gleichen Inhalts strukturell zusammenfassen, ist ebenso angedacht wie eine Verbesserung der Segmentierungs- und Verfolgungstechniken (vgl. (Price 2009)). Verfahren zur Erkennung von Bildtexten, Sprecherwechsel und Sprache sowie zur Detektion von Gesichtern liegen in separaten Verarbeitungsketten in AMOPA vor und stehen zur Integration in das Annotationstool bereit, um zukünftig umfassendere Möglichkeiten zu Annotation und Analyse anzubieten.

## Danksagung

Diese Arbeit entstand partiell innerhalb des Projekts *sachsMedia –Cooperative Producing, Storage and Retrieval*, das im Rahmen des Programms Unternehmen Region vom BMBF gefördert wird.

## Literaturverzeichnis

- Bailer, Werner; Schallauer, Peter; Neuschmied, Helmut (2007). Description of the MPEG-7 Detailed Audiovisual Profile (DAVP), Techn. Bericht, Joanneum Research, Graz, Austria.
- Beck, Peter. (1999). Implementierung eines flexiblen Algorithmus zum Tracking von Objekten in Java. Studienarbeit, TU Darmstadt.
- Boykov, Yuri; Kolmogorov, Vladimir (2004). An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 26(9):1124–1137.
- Concejero, Pedro; Munuera, Jorge; Lorenz, Mirko (2008). The MESH mobile video annotation tool. In: Proc. of the 5<sup>th</sup> ACM Nordic conference on Human-computer interaction: building bridges, NordiCHI 2008. Schweden.
- Columbia University (2004). IBM T. J. Watson Research Center: MARVEL: Multimedia Analysis and Retrieval System.  
[http://domino.research.ibm.com/comm/research\\_people.nsf/pages/jsmith.projects.html](http://domino.research.ibm.com/comm/research_people.nsf/pages/jsmith.projects.html), 25.10.2010

- Doermann, David; Mihalcik, David (2000). Tools and techniques for video performance evaluation. In: Proceedings 15<sup>th</sup> International Conference on Pattern Recognition, vol.4, pp.167–170.
- Fan, Jianping; Gao, Yuli; Hangzai, Luo; Jain, Ramesh (2008). Mining Multilevel Image Semantics via Hierarchical Classification. In: IEEE Transactions on Multimedia, 10(2), pp. 167–187.
- Goldman, Dan B.; Gonterman, Chris; Curless, Brian; Salesin, David; Seitz, Steven M. (2008). Video object annotation, navigation, and composition. In: Proceedings of the 21<sup>st</sup> annual ACM symposium on User interface software and technology, Monterey, CA, USA.
- Hagedorn, Joey; Hailpern, Joshua; Karahalios, Karrie G. (2008). VCode and VData: Illustrating a new Framework for Supporting the Video Annotation Workflow. In: AVI 2008, pp. 317–321, Neapel, Italien.
- Höhlig, Sebastian (2010). Analyse und Implementierung eines Verfahrens zur interaktiven semi-automatischen Objektmarkierung und -verfolgung. Bachelorarbeit, Technische Universität Chemnitz.
- Kipp, Michael (2008). Spatiotemporal Coding in ANVIL. Proc. of the 6<sup>th</sup> International Conference on Language Resources and Evaluation (LREC).
- Kounoudes, Anastasis; Tsapatsoulis, Nicolas; Theodosiou, Zenonas; Milis, Marios (2008). A multi-level Video Annotatin Tool based on XML-dictionaries. In: Proceedings of the 10<sup>th</sup> WSEAS international conference on Mathematical methods, computational techniques and intelligent systems, Corfu, Griechenland.
- Liu, Zhu; Gibbon, David; Zavesky, Eric; Shahrray, Behzad; Haffner, Patrick (2006). AT&T RESEARCH AT TRECVID 2006. Workshop Contribution, AT&T Labs-Research.
- Luo, Huitao; Eleftheriadis, Alexandros (1999). Designing an interactive tool for video object segmentation and annotation. In: Proceedings of the 7<sup>th</sup> ACM international conference on Multimedia. Orlando, FL, USA, pp. 265–269.
- Lux, Mathias. (2009). Caliph & Emir: MPEG-7 photo annotation and retrieval. Proceedings of the 17<sup>th</sup> ACM international conference on Multimedia, pp. 925–926, Beijing, China.
- Naphade, Milind R.; Lin, Ching-Yung; Smith, John R.; Tseng, Belle; Basu, Sankar (2002). Learning to annotate video databases. In: Proc. SPIE Vol. 4676, pp. 264 bis 275, Storage and Retrieval for Media Databases.
- Orchard, Michael; Bouman, Charles (1991). Color quantization of images. IEEE Transactions on Signal Processing, 39(12), pp. 2677–2690.
- Petridis, Kosmas; Anastasopoulos, Dionysios; Saathoff, Carsten; Timmermann, Norman; Kompatsiaris, Yiannis; Staab, Steffen (2006). M-OntoMat-Annotizer: Image Annotation. In: Linking Ontologies and Multimedia Low-Level Features. En-

- gineered Applications of Semantic Web Session at the 10<sup>th</sup> International Conference on Knowledge-Based & Intelligent Information & Engineering Systems (KES 2006), Bournemouth, U.K.
- Price, Brian L.; Morse, Bryan S.; Cohen, Scott (2009). LIVEcut: Learning-based Interactive Video Segmentation by Evaluation of Multiple Propagated Cues. In: Proc. International Conference on Computer Vision (ICCV), Kyoto, Japan.
- Ritter, Marc (2009). Visualisierung von Prozessketten zur Shot Detection. In Workshop Audiovisuelle Medien: WAM 2009, Chemnitzer Informatik-Berichte, pp. 135–150. Technische Universität Chemnitz, Germany.
- Ritter, Marc; Eibl, Maximilian (2009). Visualizing steps for shot detection. In: LWA 2009: Lernen – Wissen – Adaption, Workshop Proceedings, pp. 98–100, Darmstadt, Germany.
- Schroeter, Ronald; Hunter, Jane; Kosovic, Douglas (2004). FilmEd: Collaborative Video Indexing, Annotation and Discussion Tools Over Broadband Networks. In: Proceedings of the 10<sup>th</sup> IEEE International Conference on Multimedia Modeling, pp. 346–353. Los Alamitos, California.
- Smeaton, Alan F.; Over, Paul; Kraaij, Wessel (2006). Evaluation campaigns and trecvid. In MIR '06: Proceedings of the 8<sup>th</sup> ACM International Workshop on Multimedia Information Retrieval, pages 321–330, New York, NY, USA.
- Talbot, Justin F.; Xu, Xiaoqian (2006). Implementing GrabCut. Brigham Young University, Provo, UT, USA. <http://www.justintalbot.com/course-work/>, 06.11.2010
- Vezzani, Roberto; Grana Costantino; Bulgarelli, Daniele; Cucchiara, Rita (2006). A Semi-Automatic Video Annotation tool with MPEG-7 Content Collections. In: Proceedings of the 8<sup>th</sup> IEEE International Symposium on Multimedia, San Diego, CA, USA.
- Zwicklbauer, Stefan (2010). Evaluierung und Implementierung von Shot-Boundary-Detection-Algorithmen zur automatischen Video-Annotation. Bachelorarbeit, Universität Passau, pp. 48–52.