



ID:

Passwort:

[L@g in](#)

## Der Plagiatsuchdienst Docoloc

**Dr. Jens Brandt**

76. Jahrestagung der DPG  
Berlin, 25. - 30. März 2012

# Plagiat



"Plagiat ist geistiger Diebstahl. Schuldig macht sich, wer sich als Autor eines fremden Textes ausgibt und fremde Gedankengänge oder Argumente als die eigenen verkauft."

(ZEIT Campus 01/2010)

# Verschiedene Plagiate



- ▶ Wörtliche Übernahme fremder Texte (Copy & Paste)
- ▶ Übersetzung fremder Texte
- ▶ Paraphrasierung fremder Texte
- ▶ Plagiieren eigener Texte (Eigenplagiat)
- ▶ Übernehmen fremder Ideen oder Argumente
- ▶ Übernehmen von Ergebnissen, Daten, Bildern

⇒ Jeweils ohne Kennzeichnung und Angabe der Quellen

# Textplagiate in Wissenschaft und Lehre



- ▶ arXiv.org: 67 Dokumente wurden 2007 aufgrund von Plagiaten entfernt (Nature, 06.09.2007)
- ▶ Die IEEE überprüft seit 2010 sämtliche Einreichungen zu 24 Magazinen und 30 Konferenzen. (IEEE, 05.02.2010)
- ▶ An der Universität Klagenfurt wurden zwei Doktortitel aufgrund von Plagiaten aberkannt. Sämtliche Abschlussarbeiten werden seit 2006 automatisiert auf Plagiate geprüft. (Kleine Zeitung, 16.02.2010)
- ▶ Die Universität Göttingen erkannte einem Juristen 2009 den Doktortitel aufgrund mehrerer Plagiate ab. (Lippische Landes-Zeitung, 17.02.2011)
- ▶ Verteidigungsminister Karl-Theodor zu Guttenberg plagierte in seiner Dissertation aus mehreren Quellen. (Zeit, SZ, FAZ, Februar 2011)
- ▶ ...

# Gründe für Textplagiate



- ▶ Digitalisierung erleichtert das Kopieren fremder Inhalte
- ▶ Überwindung sprachlicher Hürden bei Publikationen
- ▶ Kulturell anderer Umgang mit geistigem Eigentum
- ▶ Publikationsdruck (z.B. wegen Drittmitteleinwerbung)
- ▶ Recyceln eigener Arbeiten verlängert eigene Publikationsliste
- ▶ Zeitdruck, bspw. Studiengebühr und viele BA/MA Module
- ▶ Mangels besseren Wissens

# Vermeidung von Plagiaten



- ▶ Intensive Kontrollen und Begutachtungen
  - ⇒ Zeitintensiv
  - ⇒ Hoher Personaleinsatz
  - ⇒ Nur in Einzelfällen realisierbar
- ▶ Automatisierte Plagiatsuche erleichtert das Auffinden
- ▶ Aufklärung zu guter wissenschaftlicher Praxis
- ▶ Sensibilisierung für Plagiate

# Plagiatsuche im Internet



- ▶ Stetig wachsende Inhalte im Internet
- ▶ Qualitativ gute Inhalte
- ▶ Digitale Veröffentlichungen, Wikipedia, Blogs, etc.
- ▶ Internetsuchmaschinen helfen beim Auffinden von Inhalten
- ▶ Suchmaschinen sind jedoch nicht für die Plagiatsuche optimiert
- ▶ Suchfragmente dürfen nicht zu lang sein

⇒ Eine manuelle Plagiatsuche ist/bleibt zeitaufwändig

# Plagiatsuchdienst Docoloc



- ▶ Onlinedienst zur automatisierten Plagiatsuche
- ▶ Entstanden 2004 zur Überprüfung studentischer Arbeiten
- ▶ Institut für Betriebssysteme und Rechnerverbund, Technische Universität Braunschweig
- ▶ Nutzung zunächst von einzelnen Mitarbeitern des Instituts
- ▶ Ab 2006 Betrieb als kommerzieller Dienst für Bildungseinrichtungen
- ▶ 2010 Gründung der Docoloc UG (haftungsbeschränkt) & Co. KG
- ▶ Fokus auf wissenschaftliche und studentische Arbeiten
- ▶ Eingeschränkter Nutzerkreis, Freischaltung durch Lizenzschlüssel

# Plagiatsuchdienst Docoloc



- ▶ Textanalyse des zu untersuchenden Dokuments
- ▶ Suche nach Übereinstimmungen im Internet
- ▶ Nutzung verschiedener Internetsuchmaschinen
- ▶ Aufbereitung gefundener Inhalte im Originallayout des Dokuments
- ▶ Herkunftsreport kann offline genutzt und archiviert werden
- ⇒ Prinzipiell gleiches Vorgehen wie ein Mensch, nur automatisiert
- ⇒ Vereint Recherche, Aufbereitung, Auswertung und Dokumentation
- ⇒ Wesentliche Zeitersparnis gegenüber manueller Kontrolle

# Docoloc Web Interface



# Docoloc

ID:  Passwort:  [Login](#)

[Anleitung](#) [Login erstellen](#) [Prüfauftrag](#) [Ihr Konto](#)

Zu prüfende Datei:  [Durchsuchen...](#) [Eingabe als Web-Adresse](#)

demo  professionell [Plagiatsuche starten](#)

Ergebnisreport:  nur in Ihr Konto legen  
 nur per E-Mail senden  in Konto und per E-Mail

[Kontakt](#) - [Twitter](#) - [Preise und AGBs](#) - [Grundsätze](#) - [Datenschutz](#) - [Referenzen](#) - [Hilfe](#)

©2011 Docoloc KG - [IBR/ITM](#) Forschungspartner - Plagiatsuche in Milliarden Dokumenten  
[Deutsch](#) [English](#)

# Docoloc Account

[Anleitung](#)[Login ändern](#)[Prüfauftrag](#)[Ihr Konto](#)Abmelden: [Logout](#)7 Prüfdokumente im Konto. [Prüfauftrag hinzufügen](#)[neu laden](#)

31	<a href="#">Militär.pdf</a> (86.53 KB)	26.03.2012 12:01	14%	<a href="#">Herkunftsreport</a>	<a href="#">löschen!</a>
30	<a href="#">Bachelorarbeit_Lernen_und_Hirnforschung.pdf</a> (603.94 KB)	26.03.2012 11:55	0%	<a href="#">Herkunftsreport</a>	<a href="#">löschen!</a>
29	<a href="#">b0be66417b5870bbf0afa9e679776471F8.pdf</a> (1.52 MB)	26.03.2012 11:55	52%	<a href="#">Herkunftsreport</a>	<a href="#">löschen!</a>
28	<a href="#">pxc387875.pdf</a> (255.85 KB)	26.03.2012 11:52	32%	<a href="#">Herkunftsreport</a>	<a href="#">löschen!</a>
26	<a href="#">0_Complete_Work.doc</a> (4.36 MB)	26.03.2012 11:48	6%	<a href="#">Herkunftsreport</a>	<a href="#">löschen!</a>
24	<a href="#">plagiat.odt</a> (16.35 KB)	26.03.2012 11:38	75%	<a href="#">Herkunftsreport</a>	<a href="#">löschen!</a>
21	<a href="#">B5.pdf</a> (322.81 KB)	26.03.2012 11:33	3%	<a href="#">Herkunftsreport</a>	<a href="#">löschen!</a>

# Docoloc Report

## Report

Digital signed

Docoloc

Reviewed document: **testfragments.txt**

Processing date: **Tue, 15.6.2010 23:42:46 CEST**

A total of **51** fragments were analysed. As a result **22** fragments (43.1%) were found in other documents. In the document preview below the fragments are marked **yellow** and clickable. At most 6 found documents are shown with same text passages.

## Cross reference documents

Following list of found documents is grouped by document titles and ordered by found fragments. With a mousedown on "**x** fragments" the relevant fragments in the document are colored **orange** and the window scrolls to the first location. Click on "**x** fragments" again resets the special marks.

**5** fragments were found in a text with the title: "**KidsWingTsun und soziales Kompetenztraining für Kids im ...**", located on:  
[http://www.kidswingtsun-schule-koblenz.de/wtkids\\_gewalt.html](http://www.kidswingtsun-schule-koblenz.de/wtkids_gewalt.html)

**4** fragments were found in a text with the title: "**wingtsun.de | Gewaltprävention | Warum brauchen Kinder und ...**", located on:  
<http://www.wingtsun.de/gewaltpraevention/warum-gewaltpraevention.html>

**4** fragments were found in a text with the title: "**Skript: Qualitätsmanagement: Amazon.de: Rolf Mohr: Bücher**", located on:  
<http://www.amazon.de/Skript-Qualit%C3%A4tsmanagement-Rolf-Mohr/dp/3638703967>

25% with fuzzy search (1 fragment)

We employed a Microstrip-to-CPW-to-Microstrip transition and via holes to transfer the current from the top to the bottom substrate layer and vice versa. The presented phase shifter is operating in a wide bandwidth between 5.5 and 17.2 GHz, with low insertion loss and reflection coefficients. Because the input and output microstrip lines are on the same layer, the presented phase shifter is suitable for a modified class of feeding networks for phased antenna arrays.

1 match:

Various studies of throughput and channel utilization for wireless communication in asymmetric channels, multi-hop interference, high traffic demand and mechanisms for contention have been introduced to notify

[PIER\\_B Online - Wideband\\_180\\_Degree\\_Phase\\_Shifter\\_Using\\_Multihop\\_Interference\\_in\\_Asymmetric\\_Channels.pdf](http://ceta.mit.edu/PIER_B/PIER_B_Online - Wideband_180_Degree_Phase_Shifter_Using_Multihop_Interference_in_Asymmetric_Channels.pdf) Because the input and output microstrip lines are on the same layer, the presented phase shifter is suitable for a modified class of feeding networks ...

[http://ceta.mit.edu/PIER\\_B/PIER\\_B\\_Online - Wideband\\_180\\_Degree\\_Phase\\_Shifter\\_Using\\_Multihop\\_Interference\\_in\\_Asymmetric\\_Channels.pdf](http://ceta.mit.edu/PIER_B/PIER_B_Online - Wideband_180_Degree_Phase_Shifter_Using_Multihop_Interference_in_Asymmetric_Channels.pdf)

Given a mesh network, our goal is therefore to identify a set of maximal concurrent transmission sets whose union contains all the nodes in the network. By assigning one time slot to each of these sets, a node schedule can be created, maximizing throughput and ensuring that each node gets to transmit. This schedule is then repeated.

# Docoloc Report - Originallayout

## Efficient Pixel Substitution Using Match Batching for Real-Time Video Streaming

 Balabda, IEEE member  
ABC Laboratories, University of Somewhere

City, 12345 Some Land

**Abstract** In this paper we present a novel approach for 3D camera self-localization, that uses template matching in a particle filter framework. We propose a tracking scheme that explores the matching scenario, where image patches to extract the matching features of the filter are generated and updated by representing a template image with an ensemble of patches. The method is robust with respect to variations such as local appearance, variation, partial occlusion, and scale changes. Experimental results show that the proposed tracking scheme has shown that the proposed approach provides accurate tracking, especially for fast motion or long-term partial occlusions. Comparisons have been made with existing methods; results have shown that the proposed scheme has provided an improved tracking accuracy at the cost of more computations.

### INTRODUCTION

Real-time camera tracking is one of the most important tasks in Augmented-Mixed Reality (AM-MR) systems. One way to resolve this problem is to use GPS and compass sensors to estimate the camera pose (position and orientation) with respect to the environment [1][2]. However, these kinds of sensors are well adapted for large scale outdoor environments, but not for indoor environments or mobile applications. To simplify the problem for the indoor applications, it is common to use planar fiducial markers of known appearance. The tracking problem consists then to identify the markers from the images and computing the corresponding camera poses. Several marker-based tracking systems have been developed these last years, the most popular are ARToolKit [3], ARToolKitPlus [4] and ARKit [5]. The authors used these systems successfully to track both 2D and 3D patterns that can be easily detected. The four corners of the markers can then be assumed and the relative camera pose calculated. However, these methods still lack tracking robustness, as the fiducials must be visible at all times.

For real-time applications, most of tracking algorithms are based on Extended Kalman filtering. The idea is to estimate sequentially the state of the system using a noisy measurement. The first work that used this approach and for the first time to compute the 3D location by Wilson et al. [6]. Yoon et al. [7] presents a model-based object tracking to compute the 3D camera pose. Their algorithm uses an EKF to provide an incremental pose-update scheme in a prediction-verification framework. Recently, Abalos [8] developed a new robust camera pose estimation approach based on 3D lines features. The author used an EKF to incrementally update the camera pose in real-time. The principal contributions of this method include the expansion

of the RANSAC scheme in order to achieve a robust matching between 2D image edges and 3D model lines segments. However, the linearization process in EKF equations can cause instability of the filter. Also, the computation of the Jacobian matrices is generally complex and can be difficult to implement in the implementation of the EKF.

The particle filters [9][10][11], provide general solution to many problems where linearization and Gaussian approximations are intractable or would yield poor performance, particularly if the methods are to be extended to non-static scenes in which nonlinearity and multimodality are likely to be ...

**Efficient Pixel Substitution Using Match Batching for Real-Time Video Streaming**  
Balabda, IEEE member  
ABC Laboratories, University of Somewhere  
City, 12345 Some Land

<http://www.ijcai.org/Proceedings/11/IJCAI11-00537.pdf>

In this paper, we propose a new marker-based camera tracking approach that uses the particle filter framework. Our main contributions include: 1) improving the tracking robustness to occlusion by using an ensemble of patches instead of a single large template; 2) constructing an efficient likelihood function of the particle filter based on a template matching framework.

The rest of the paper is organized as follows: section III, describes the tracking formulation when combining the particle filtering framework within the square fiducials. In section III, we explain how we have constructed the observation model using the patch matching approach. Experimental results are presented in section IV. Finally, conclusions and future work are given in section V.

### CAMERA TRACKING USING PARTICLE FILTER

In this paper, the 3D camera pose is recursively estimated using a particle filter. The camera pose is represented by position and orientation of the camera with respect to a world coordinate system. The camera pose is represented by several different mathematical angles. Euler angles, quaternions, and axis-angle are proven very useful in robotics. The quaternion is a good choice to represent the camera pose, as it provides a quaternions to represent the rotation ...

**Efficient Pixel Substitution Using Match Batching for Real-Time Video Streaming**  
Balabda, IEEE member  
ABC Laboratories, University of Somewhere  
City, 12345 Some Land

<http://www.ijcai.org/Proceedings/11/IJCAI11-00537.pdf>

$$X = \begin{bmatrix} q_x & q_y & q_z & r_x & r_y & r_z \end{bmatrix} \quad (1)$$

# Features



- ▶ Unterstützte Dokumentenformate:  
PDF, DOCX, DOC, ODF, RTF, HTML, TXT
- ▶ Automatisierte Plagiatprüfung im Hintergrund
- ▶ Abgeschlossener, signierter Ergebnisreport
- ▶ Möglichkeit zum Versenden der Reports per E-Mail
- ▶ Webservice Schnittstelle zur Integration in bestehende Systeme
- ▶ Privater Suchbereich
- ▶ Einhaltung gängiger Datenschutz- und Urheberrechtsbestimmungen

# Verbreitung von Docoloc

- ▶ 73 Universitäten und Hochschulen in Deutschland
- ▶ Weitere 22 Hochschulen in Österreich und Schweiz
- ▶ 5 weitere Universitäten weltweit
- ▶ 85 berufs- und allgemeinbildende Schulen
- ▶ Schweizer Portal für Schulen: [www.copy-stop.ch](http://www.copy-stop.ch)
- ▶ Konferenz-Management System EDAS: [www.edas.info](http://www.edas.info)
- ▶ Jährlich ca. 250.000 geprüfte Dokumente mit 3,3 Mio Seiten

# Docoloc Webservice

- ▶ Großteil der Nutzer nutzt Docoloc über den Browser
- ▶ Weitere Automatisierung sinnvoll
- ▶ Docoloc Webservice:
  - ▶ Nutzung von Docoloc durch andere Computer
  - ▶ SOAP (Simple Object Access Protocol)
  - ▶ Entfernter Aufruf von Docoloc Methoden
  - ▶ Methoden zum Einreichen und Abrufen von Dokumenten
- ▶ Einbindung in bestehende Prozesse
- ▶ Integration in vorhandene Software
- ▶ Erstellen automatisierter Einreich-/Abgabeprozesse

# Private Suchbereiche



- ▶ Suchraum im Internet begrenzt
- ▶ Potentielle Quellen teilweise nicht veröffentlicht
- ▶ Aufbau eigener Suchbereiche
- ▶ Steht ausschließlich den Nutzern eines Kunden zur Verfügung
- ▶ Dokumente werden durch Docoloc Server indiziert
- ▶ Index wird bei Plagiatsuche mit einbezogen

# Datenschutz

- ▶ Keine dauerhafte Speicherung von Dokumenten
- ▶ Keine Verwertung übermittelter Dokumente
- ▶ Keine Speicherung personenbezogener Daten aus Dokumenten
- ▶ Absicherung der Server durch gängige Schutzmechanismen
- ▶ Serverstandorte ausschließlich in Deutschland
- ▶ Überprüft und Bestätigt durch externen Datenschutzbeauftragten

# Open Access Plagiatsuche



# Open Access und Plagiate



Frei zugängliche Inhalte erleichtern ein direktes Kopieren

- ▶ Schüler und Studenten kopieren Inhalte aus Wikipedia
- ▶ Autorin Helene Hegemann schreibt aus einem Blog ab
- ▶ Doktoranden kopieren Texte aus dem Internet

# Open Access und Plagiate



## Frei zugängliche Inhalte erleichtern ein direktes Kopieren

- ▶ Schüler und Studenten kopieren Inhalte aus Wikipedia
- ▶ Autorin Helene Hegemann schreibt aus einem Blog ab
- ▶ Doktoranden kopieren Texte aus dem Internet

## Frei zugängliche Texte erleichtern die Erkennung von Plagiaten

- ▶ Internet Suchmaschinen fördern die Quellen zu Tage
- ▶ Automatisierte Plagiatsuche möglich
- ▶ Aufwertung der Ergebnisse durch Metadaten
- ▶ Vermeidung von Eigenplagiaten

# Open Access Plagiatsuche



## Ziele

- ▶ Neuartiger OA Serviceprovider zur Plagiatsuche
- ▶ Vermeidung von Textplagiaten in OA Repositorien
- ▶ Stärkung der Qualität von OA Veröffentlichungen
- ▶ Sensibilisierung für Textplagiate in der Wissenschaft
- ▶ Unterstützung in Begutachtungsprozessen

## Umsetzung

- ▶ Aufbau eines Volltextindex aller verfügbaren OA Dokumente
- ▶ Volltext-Suchmaschine für OA Repositorien
- ▶ DFG Förderung: 2009 - 2011

# OAPS Projekt



funded by

**DFG** Deutsche  
Forschungsgemeinschaft

# Status



## OAPS Beta

- ▶ Gestartet Anfang Februar 2011
- ▶ Offene Beta-Phase
- ▶ Frei zugänglich nach erfolgter Registrierung
- ▶ Voller Funktionsumfang
- ▶ Begrenzter Umfang des Suchraums
- ▶ 10 Mio. Wikipedia-Artikel und 2 Mio. OA Dokumente

# Lessons Learned



- ▶ Nur wenige OA Repositorien bieten tatsächlich "open access"
  - ▶ Fehlende oder nicht erreichbare Volltexte
  - ▶ Fehlerhafte Daten
  - ▶ Einschränkungen durch Betreiber
- ▶ Open Access tatsächlich meist nur bei Metadaten
- ▶ Erreichter Suchindex nur bedingt für Plagiatsuche nutzbar

# Einschränkungen von Plagiatsuchdiensten

- ▶ Plagiatsuchdienste können nur so gut sein wie ihre Suchbasis
- ▶ Software kann lediglich Übereinstimmungen finden
- ▶ Entscheiden ob ein Plagiat vorliegt muss ein Mensch
- ▶ Fehlen von Übereinstimmungen garantiert keine Plagiatfreiheit
- ▶ Plagiatsuchdienste bieten lediglich eine Hilfestellung

# Zusammenfassung

## Open Access Plagiatsuche

- ▶ Plagiatsuche auf Basis von Open Access Dokumenten
- ▶ Aufbau eines Suchindex verfügbarer OA Dokumente



## Der Plagiatsuchdienst Docoloc

- ▶ Automatisierte Plagiatsuche
- ▶ Nutzung von Internetsuchmaschinen
- ▶ Integration durch Nutzung des Docoloc Webservice
- ▶ Private Suchbereiche

# Docoloc

<http://www.docoloc.de>

info@docoloc.de

Dr. Jens Brandt

brandt@docoloc.de

Testen Sie Docoloc:  
[www.docoloc.de/dpg2012.hhtml](http://www.docoloc.de/dpg2012.hhtml)

