



EINFACH
SICHER

Phishingerkennung mittels visuellem Ähnlichkeitsvergleich

Felix Hill

Ruhr-Universität Bochum

felix.hill@rub.de



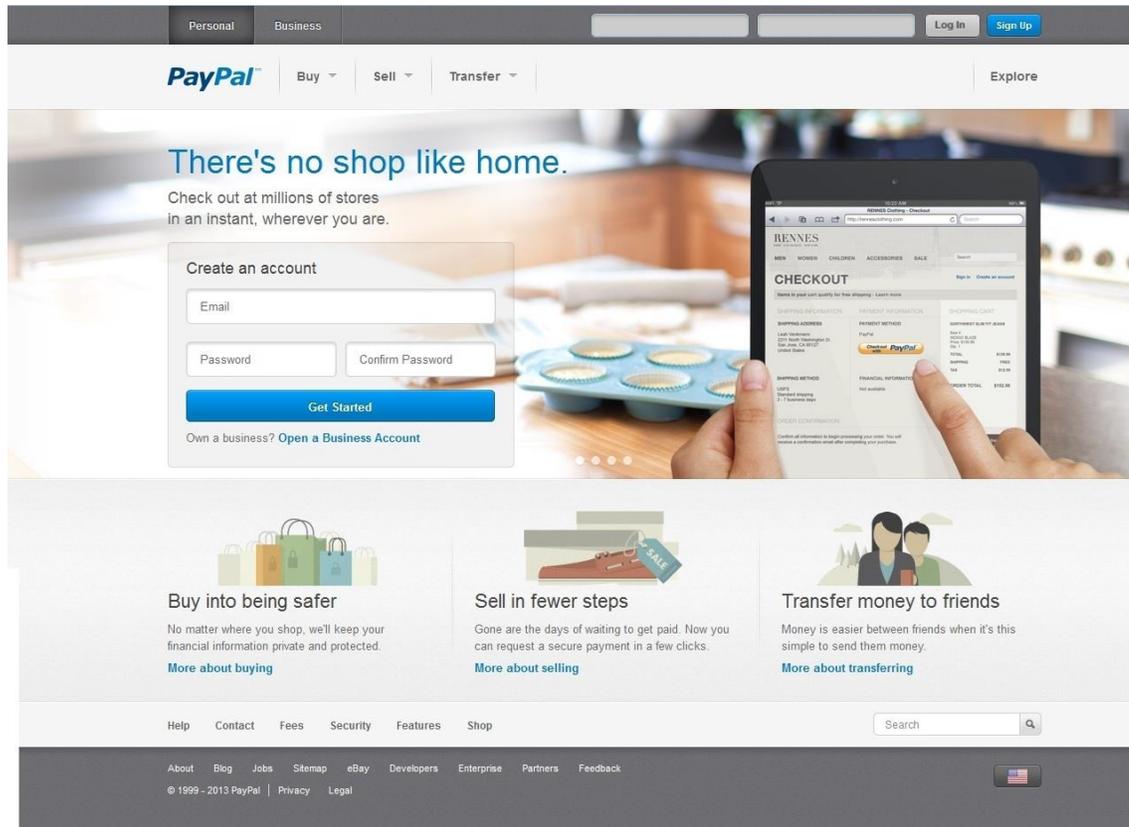
ÜBERSICHT

- Entwicklung im Bereich Phishing
- Ansatz
- Bilderkennung
- Evaluation

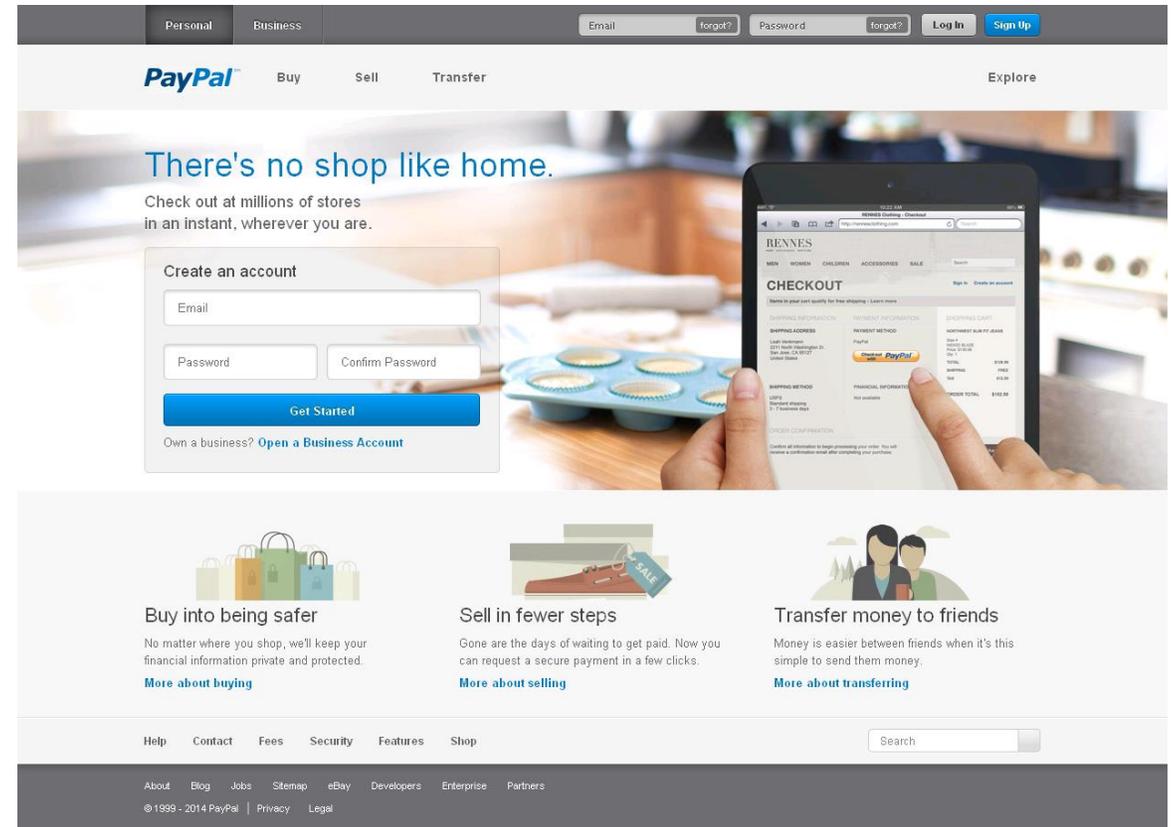
ENTWICKLUNG IM BEREICH PHISHING

- Phishing immer attraktiver
 - Onlinebanking
 - Onlineshopping
- Phishing schwerer für den Nutzer zu erkennen
 - Weniger Rechtschreibfehler
 - Original Bilder / Layout
- Phishing schwerer automatisiert zu erkennen
 - Quelltext-Obfuskiierung durch Javascript, Iframes, Flash...
 - Screenshots der originalen Webseite

ENTWICKLUNG IM BEREICH PHISHING



Original Webseite 14.01.2014



Phishing Webseite 30.03.2015

ANSATZ

- Ausnutzung der starken Ähnlichkeit der Phishing Webseiten zum Original
- Arbeit auf Daten, die der Nutzer sieht
- Screenshot Vergleich mit originalen Webseiten
 - 100% Vergleich
 - Inhaltsbasierte Bilderkennung
- Ziel: Automatisiertes Klassifikationsverfahren
 - Minimierung der falschen Zuordnungen

ANSATZ: INHALTSBASIERTE BILDERKENNUNG

- Algorithmen, die auf die Merkmale einer Webseite eingehen
- Aus Screenshots Merkmalvektoren extrahieren
- Vektoren miteinander vergleichen
- Pro
 - Unempfindlich gegenüber leichten Änderungen
- Contra
 - Rechenintensive Extraktion
 - Viele Vergleiche von Vektoren
 - Mögliche falsche Zuordnungen

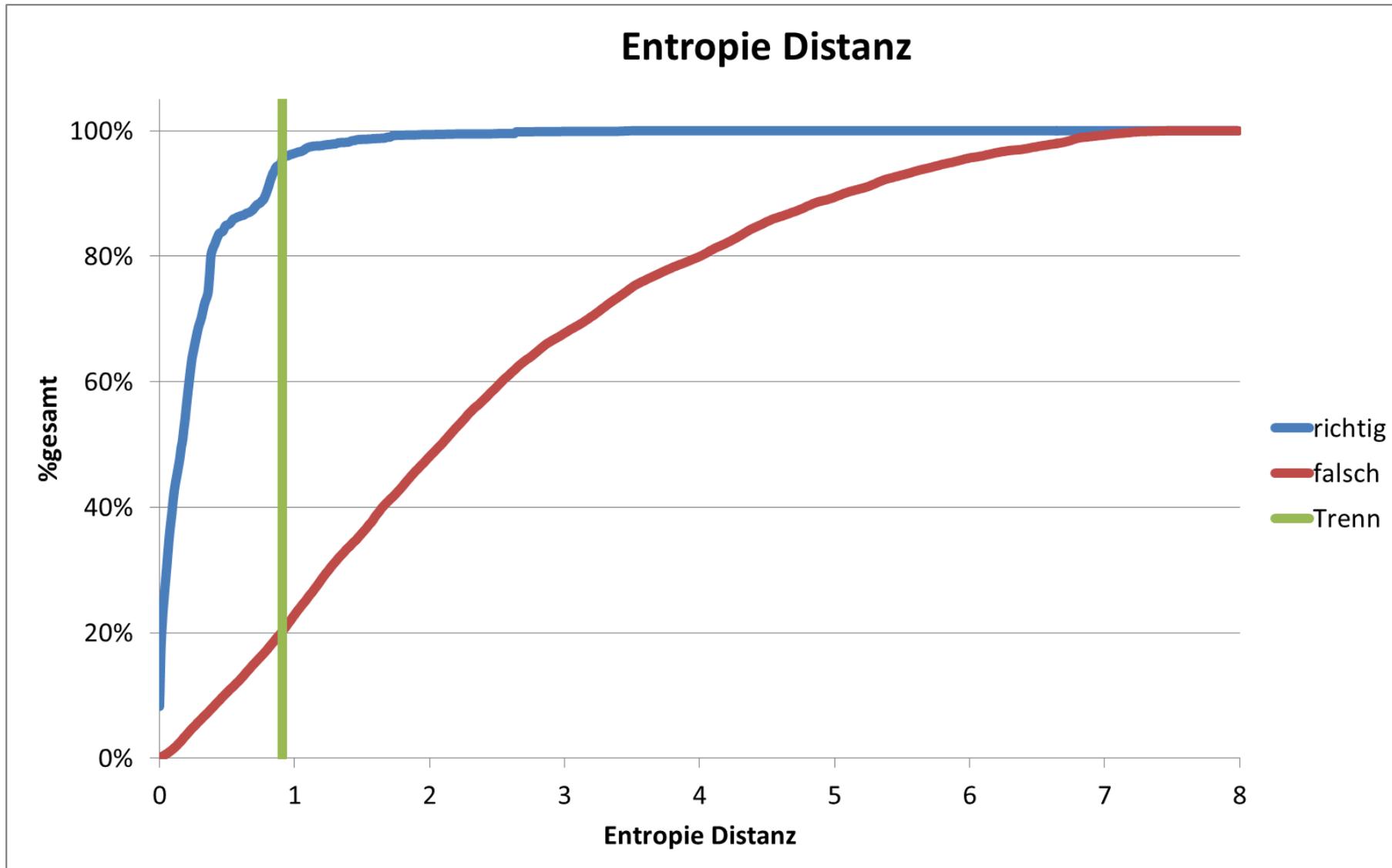
ANSATZ: INHALTSBASIERTE BILDERKENNUNG

- Algorithmen, die auf die Merkmale einer Webseite eingehen
 - Aus Screenshots Merkmalvektoren extrahieren
 - Vektoren miteinander vergleichen
 - Pro
 - Unempfindlich gegenüber leichten Änderungen
 - Contra
 - Rechenintensive Extraktion
 - ~~Viele Vergleiche von Vektoren~~
 - Mögliche falsche Zuordnungen
-  **Vorsortierung der Bilder**

VORSORTIERUNG

- Betrachte folgende Attribute:
 - Grau-Entropie
 - Farb-Entropie
 - Weißanteil
 - Anzahl unterscheidbarer Farben
 - Durchschnittshelligkeit
- These: Gleiche Screenshots haben gleiche Attribute
- Betrachte % und totale Abstände

VORSORTIERUNG



VORSORTIERUNG

- Trennung bei größter richtig / falsch Distanz
- Ausschluss aller Werte über Trennwert

- Ergebnisse nach Kombination von 6 Attributen
 - Alle Attribute müssen ähnlich sein
 - Behalte 75% der richtigen Zuordnungen (95% der Bilderkennungen)
 - Schließe 99,53% der möglichen falschen Zuordnungen aus

ALGORITHMEN

- 14 verschiedene Algorithmen
 - Inhaltsbasierte Bilderkennungsalgorithmen
 - Erstellung eines Merkmalvektors aus dem Bild
 - Berechnung eines Distanzwertes zwischen Vektoren

- Analyse der ursprünglichen Anwendungsgebiete

MERKMALE EINER WEBSEITE

The image shows the PayPal website interface. At the top, there are navigation tabs for 'Personal' and 'Business', a search bar, and 'Log In' and 'Sign Up' buttons. Below this is the PayPal logo and navigation links for 'Buy', 'Sell', and 'Transfer', along with an 'Explore' link. The main banner features the headline 'There's no shop like home.' and a sub-headline 'Check out at millions of stores in an instant, wherever you are.' To the left of the banner is a 'Create an account' form with fields for 'Email', 'Password', and 'Confirm Password', and a 'Get Started' button. To the right is a tablet displaying a 'CHECKOUT' page from 'RENNES Clothing' with a 'Check out PayPal' button. Below the banner are three feature highlights: 'Buy into being safer' (with shopping bags icon), 'Sell in fewer steps' (with a shoe icon), and 'Transfer money to friends' (with a couple icon). Each highlight includes a brief description and a 'More about...' link. The footer contains a navigation menu with links like 'Help', 'Contact', 'Fees', 'Security', 'Features', and 'Shop', a search bar, and copyright information: '© 1999 - 2013 PayPal | Privacy | Legal'.

ALGORITHMEN

- Unpassende Anwendungsgebiete
 - Iriserkennung
 - Fingerabdruckerkennung
 - Rotationserkennung

- Es bleiben 11 Algorithmen für weitere Tests

ALGORITHMEN

- Tests der verbleibenden Algorithmen auf Testset
- Betrachten der Bildpaare nach der Vorsortierung

- Finden eines Schwellwertes für jeden Algorithmus
 - Max. 0.1% falsche Zuordnungen
 - Keine falsche Zuordnung mit Abstand 0



ALGORITHMEN

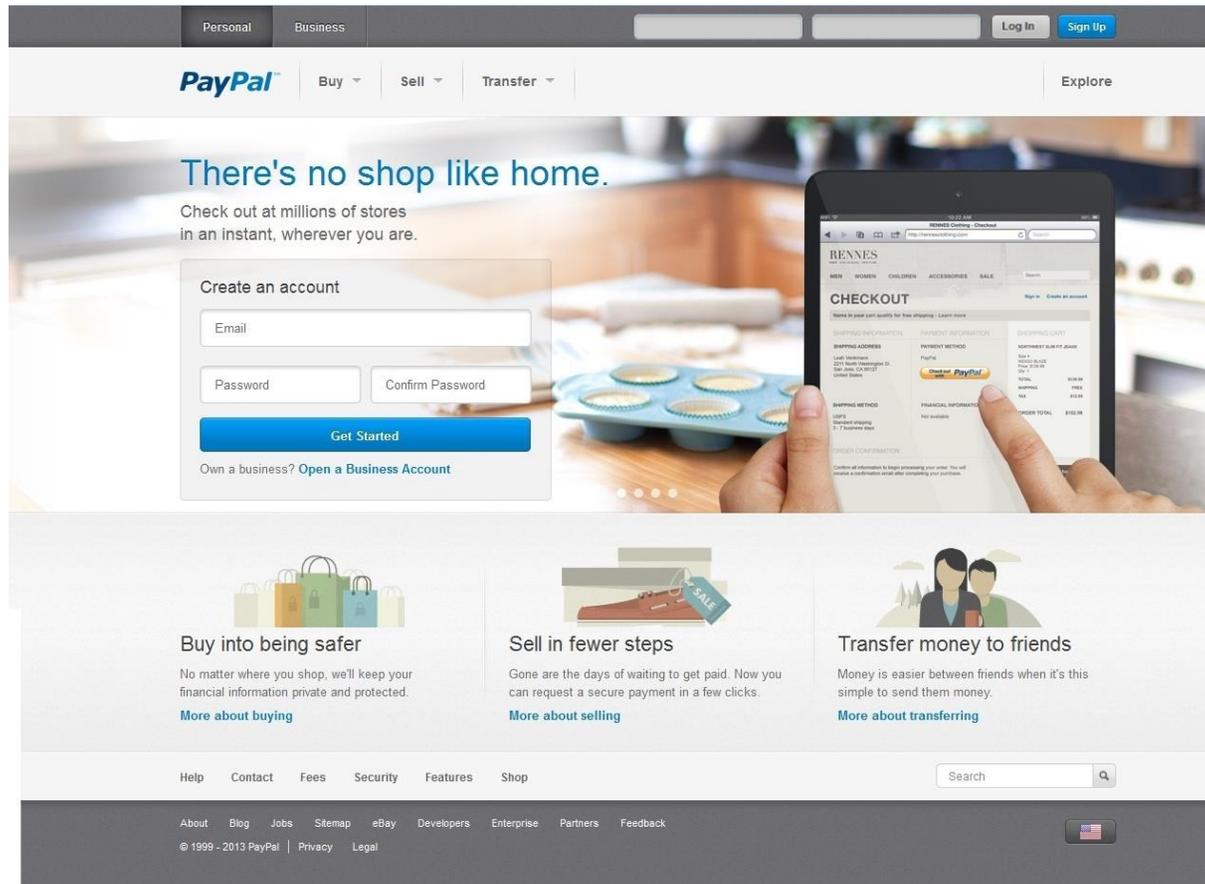
- Es verbleiben 3 Algorithmen:
 - Color Layout Descriptor (CLD)
 - Auto Color Correlogram (ACC)
 - Edge Histogram Descriptor (EHD)
- Hohe Erkennungsraten durch großes Signaturset
- Erkennungsraten:
 - 45-69% Erkennung
 - <0,1% Falschzuordnungen

ALGORITHMEN: CLD

- Entwickelt im MPEG-7 Standard
- Arbeitet auf 8x8 Bild
- Nutzt diskrete Kosinustransformation auf YCbCr Darstellung
- Speichert nur niedrige Frequenzen

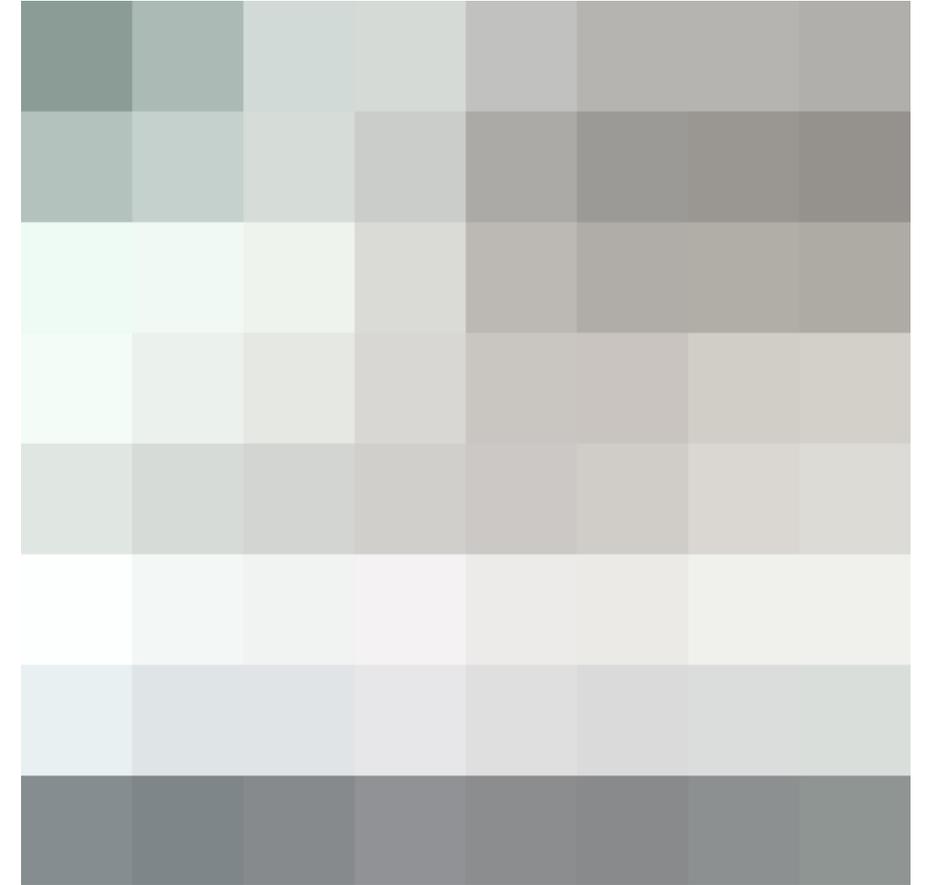
ALGORITHMEN: CLD

- Arbeitet auf 8x8 Bild

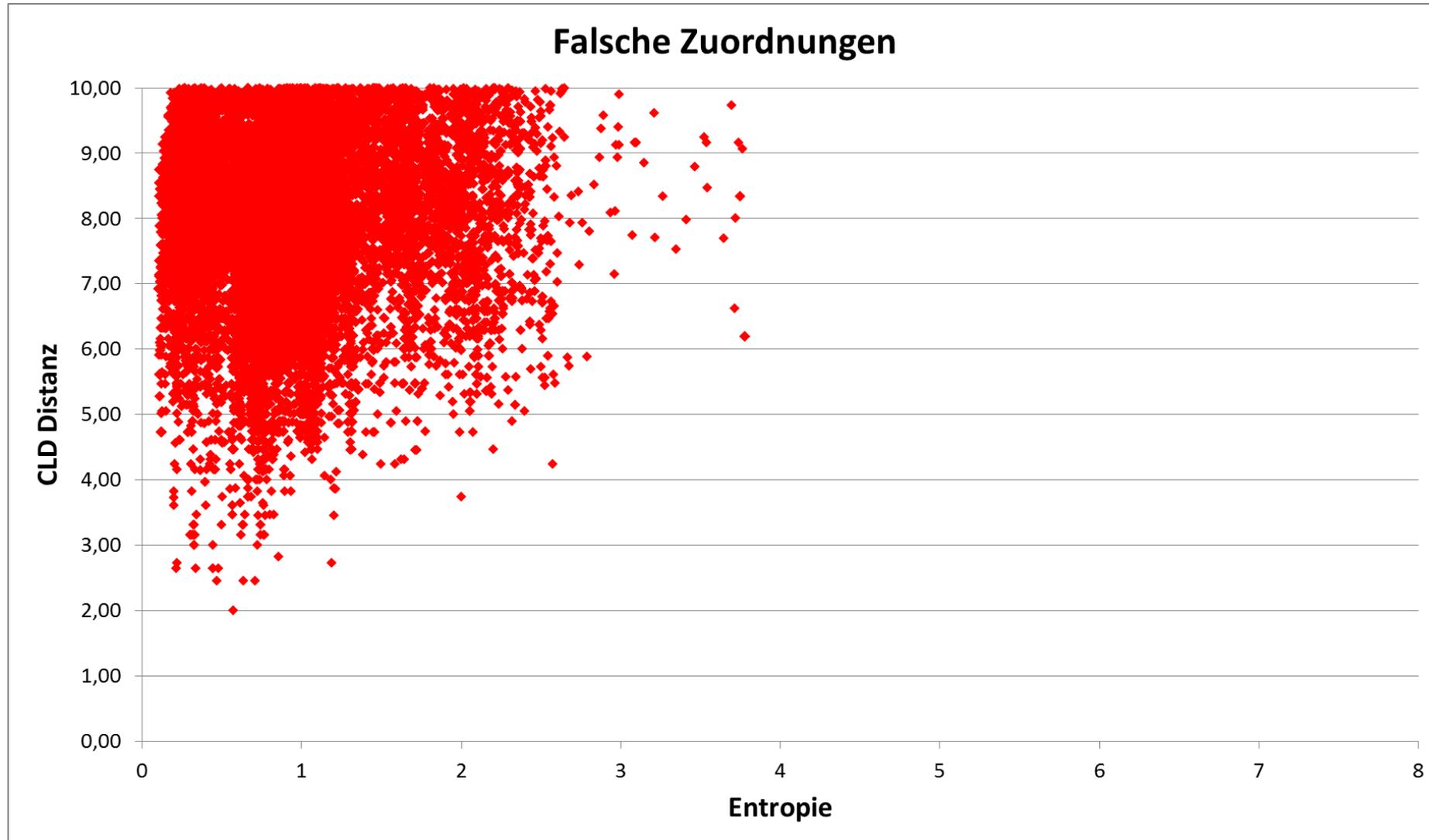


ALGORITHMEN: CLD

- Speichert nur niedrige Frequenzen

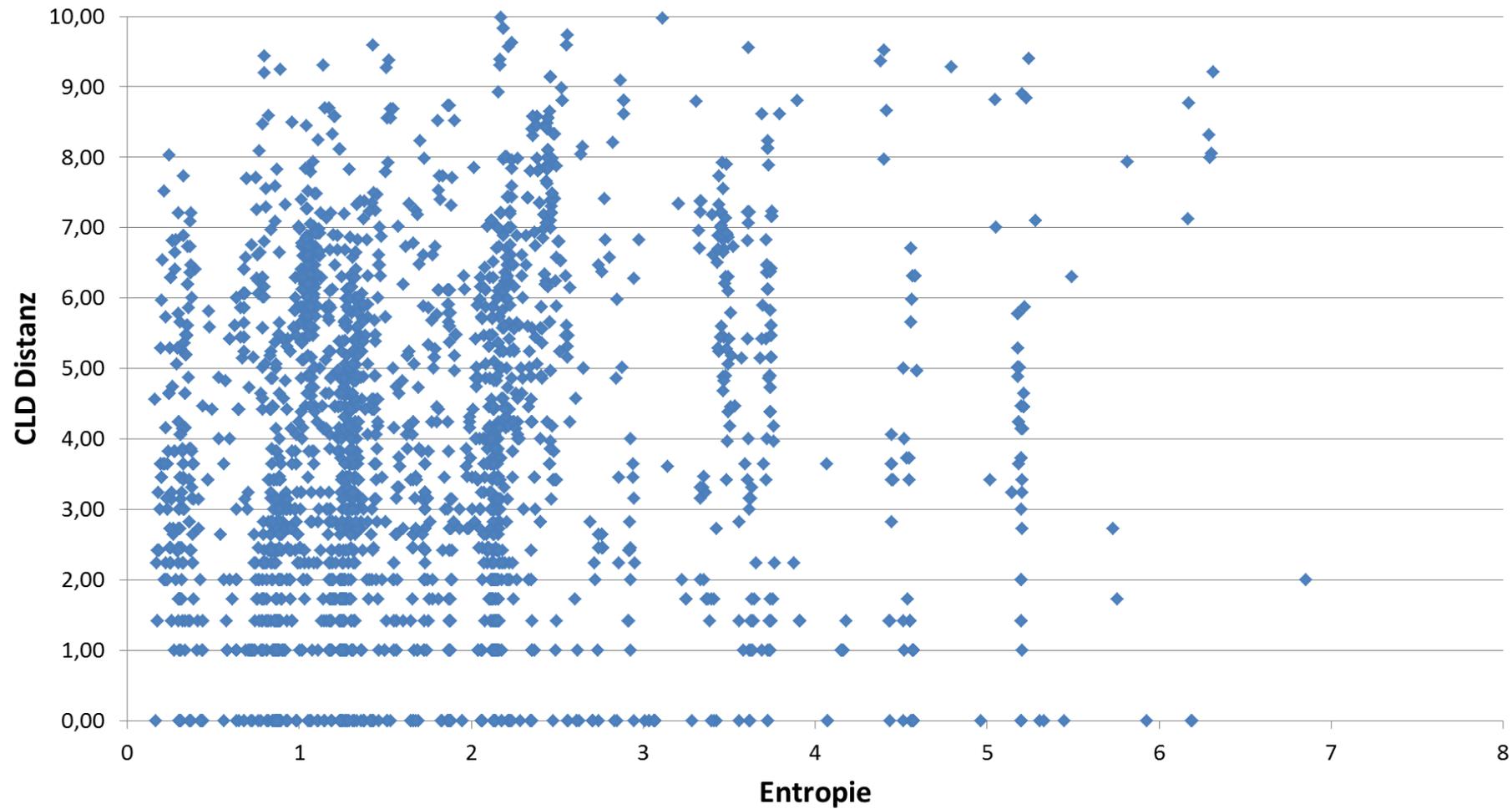


GRENZWERTBESTIMMUNG



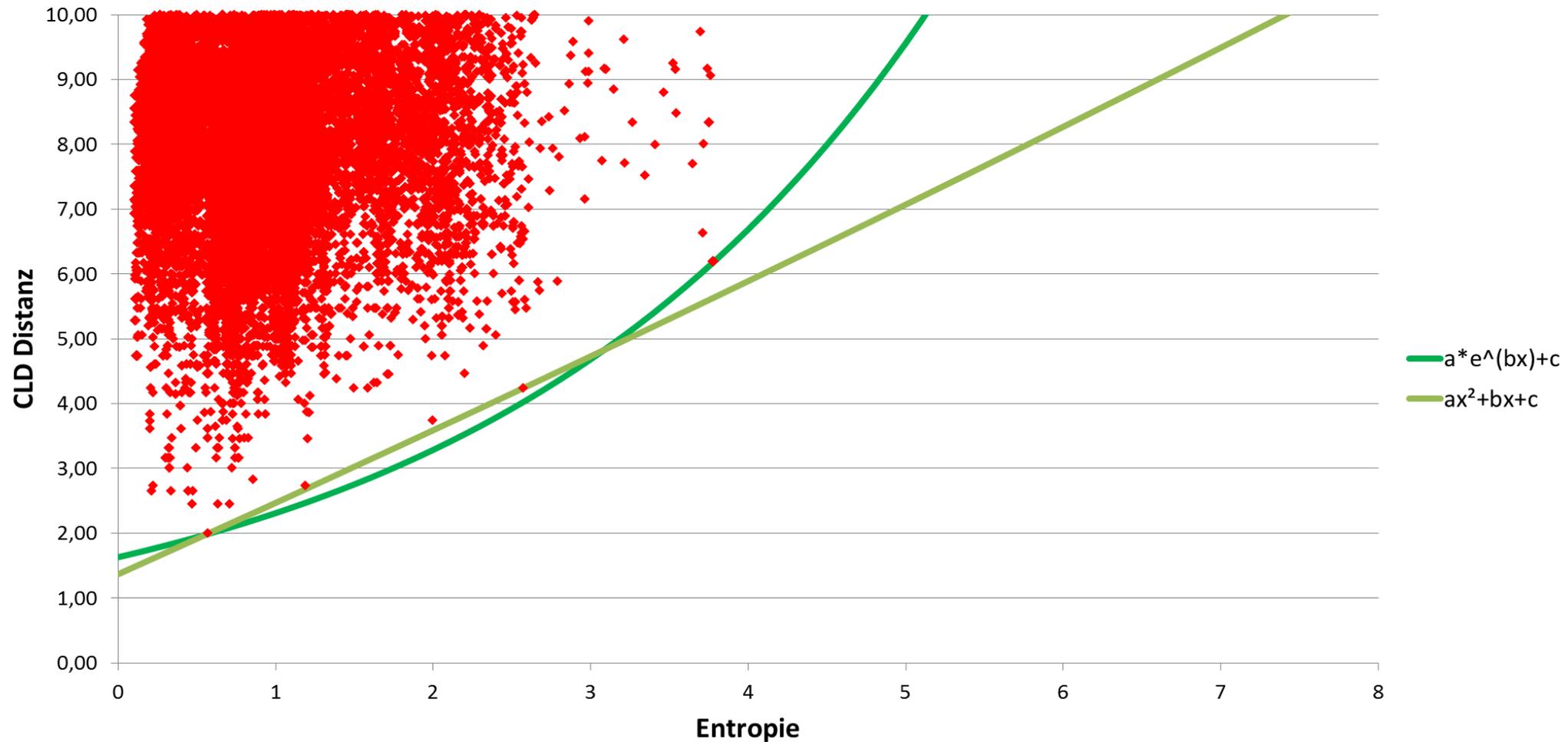
GRENZWERTBESTIMMUNG

Richtige Zuordnungen



GRENZWERTBESTIMMUNG

Falsche Zuordnungen



GRENZWERTBESTIMMUNG

- Grenzwerte abhängig von Attributen
 - 5 Grenzwert-Kurven
 - ODER Kombination

	Erkennung statisch	Erkennung dynamisch
CLD	58%	71%
EHD	45%	57%
ACC	69%	85%

GRENZWERTBESTIMMUNG

- Verbesserung der Algorithmen
 - Frequenzfilter verändern
 - Betrachtungsbereich verändern

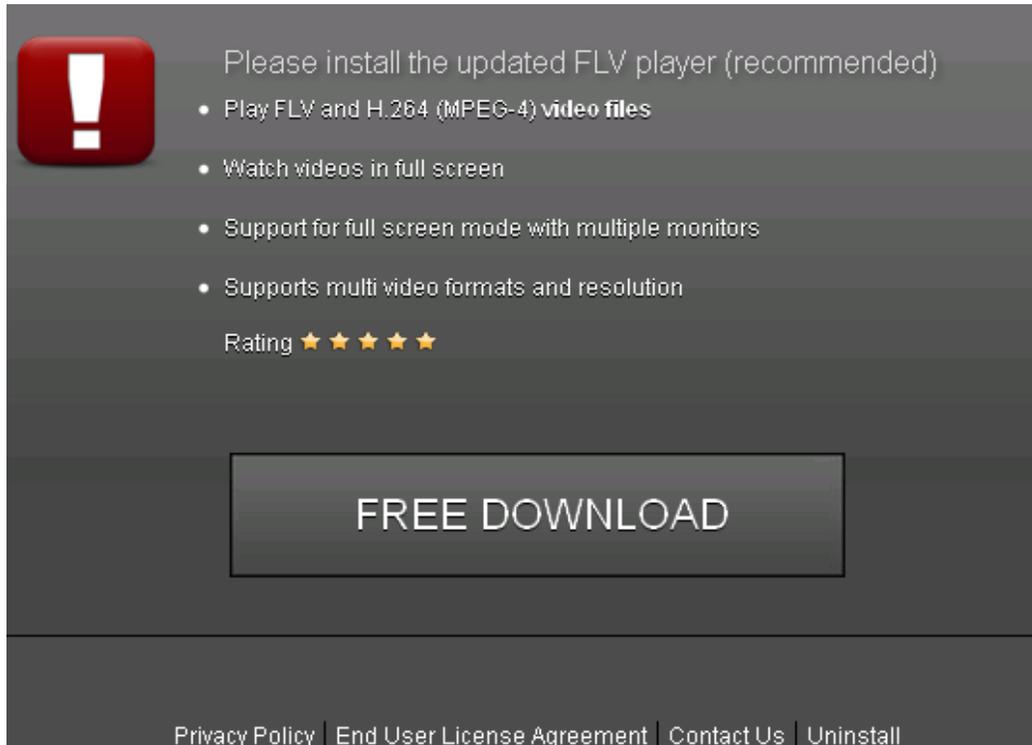
	Erkennung statisch	Erkennung dynamisch	Erkennung dynamisch 2
CLD	58%	71%	77%
EHD	45%	57%	-
ACC	69%	85%	85%

EVALUATION

- Set aus Livedaten
 - Alte Signaturbilder
 - 500.000 Screenshots
 - Nicht nur Webseiten
- EHD und ACC haben hohe FP-Raten
- CLD gute Ergebnisse

EVALUATION: CLD

- 33.000 Webseiten erkannt
- 7 falsche Zuordnungen



A screenshot of a website notification. On the left is a red square with a white exclamation mark. To its right, the text reads: "Please install the updated FLV player (recommended)". Below this are four bullet points: "Play FLV and H.264 (MPEG-4) video files", "Watch videos in full screen", "Support for full screen mode with multiple monitors", and "Supports multi video formats and resolution". Below the bullet points is a "Rating" section with five yellow stars. At the bottom of the notification is a large, dark grey button with the text "FREE DOWNLOAD". At the very bottom of the page, there are links for "Privacy Policy", "End User License Agreement", "Contact Us", and "Uninstall".

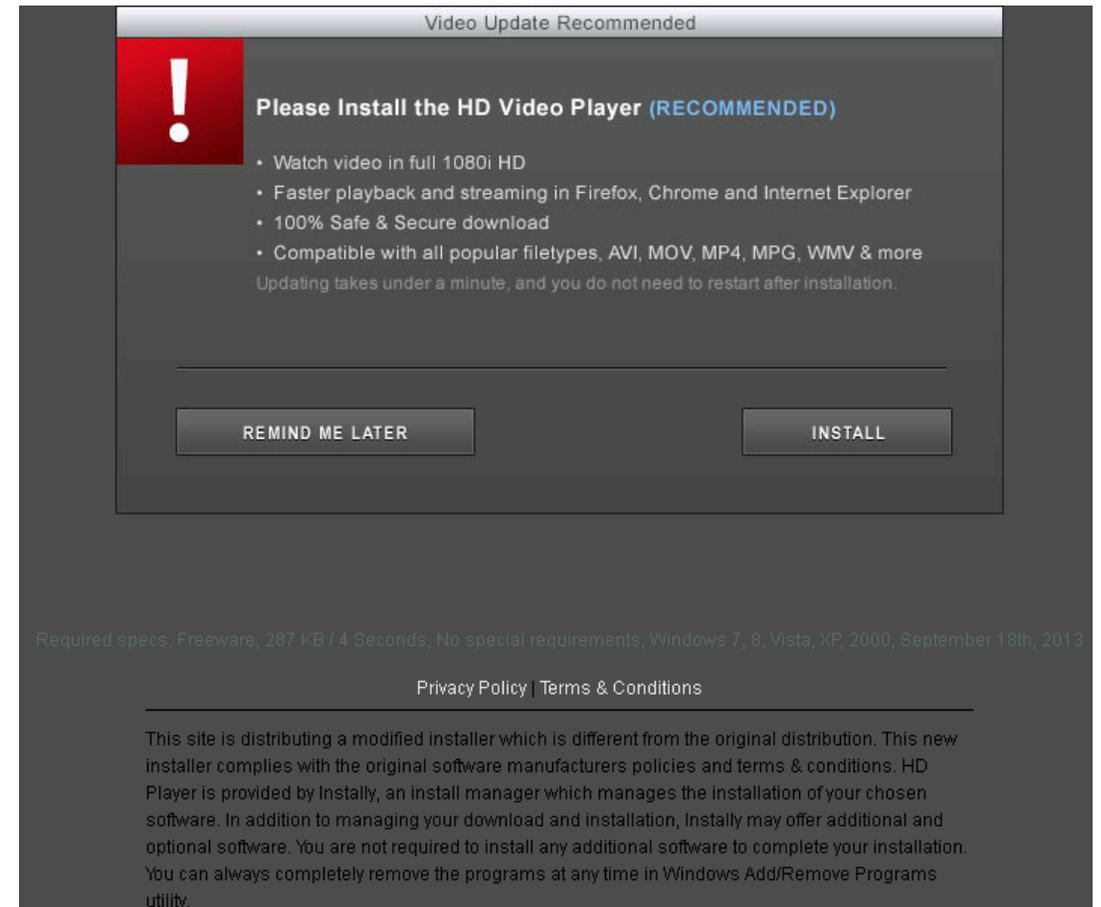
Please install the updated FLV player (recommended)

- Play FLV and H.264 (MPEG-4) video files
- Watch videos in full screen
- Support for full screen mode with multiple monitors
- Supports multi video formats and resolution

Rating ★★★★★

FREE DOWNLOAD

[Privacy Policy](#) | [End User License Agreement](#) | [Contact Us](#) | [Uninstall](#)



A screenshot of a "Video Update Recommended" dialog box. It features a red square with a white exclamation mark on the left. The main text says: "Please Install the HD Video Player (RECOMMENDED)". Below this are three bullet points: "Watch video in full 1080i HD", "Faster playback and streaming in Firefox, Chrome and Internet Explorer", and "100% Safe & Secure download". A fourth line of text states: "Compatible with all popular filetypes, AVI, MOV, MP4, MPG, WMV & more". A note at the bottom of the list says: "Updating takes under a minute, and you do not need to restart after installation." Below the list are two buttons: "REMIND ME LATER" and "INSTALL". At the bottom of the dialog, there is a line of small text: "Required specs: Freeware, 287 KB / 4 Seconds, No special requirements, Windows 7, 8, Vista, XP, 2000, September 18th, 2013". Below the dialog, there are links for "Privacy Policy" and "Terms & Conditions". At the bottom of the page, there is a paragraph of text: "This site is distributing a modified installer which is different from the original distribution. This new installer complies with the original software manufacturers policies and terms & conditions. HD Player is provided by Instally, an install manager which manages the installation of your chosen software. In addition to managing your download and installation, Instally may offer additional and optional software. You are not required to install any additional software to complete your installation. You can always completely remove the programs at any time in Windows Add/Remove Programs utility."

Video Update Recommended

Please Install the HD Video Player (RECOMMENDED)

- Watch video in full 1080i HD
- Faster playback and streaming in Firefox, Chrome and Internet Explorer
- 100% Safe & Secure download
- Compatible with all popular filetypes, AVI, MOV, MP4, MPG, WMV & more

Updating takes under a minute, and you do not need to restart after installation.

REMIND ME LATER **INSTALL**

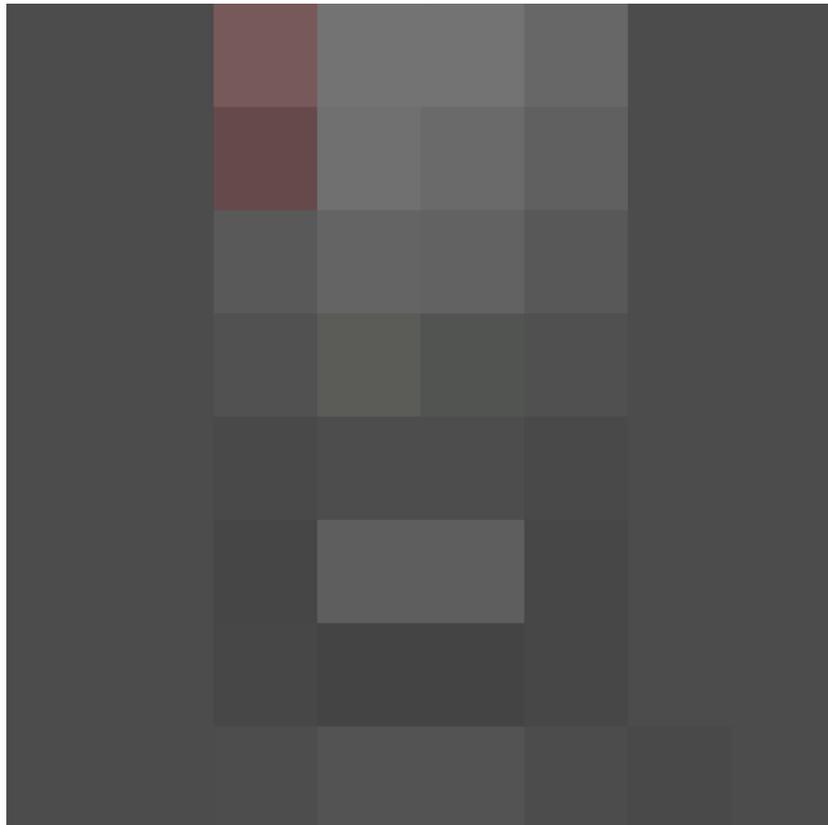
Required specs: Freeware, 287 KB / 4 Seconds, No special requirements, Windows 7, 8, Vista, XP, 2000, September 18th, 2013

[Privacy Policy](#) | [Terms & Conditions](#)

This site is distributing a modified installer which is different from the original distribution. This new installer complies with the original software manufacturers policies and terms & conditions. HD Player is provided by Instally, an install manager which manages the installation of your chosen software. In addition to managing your download and installation, Instally may offer additional and optional software. You are not required to install any additional software to complete your installation. You can always completely remove the programs at any time in Windows Add/Remove Programs utility.

EVALUATION: CLD

- 33.000 Webseiten erkannt
- 7 falsche Zuordnungen



...VIELEN DANK!

TESTSET

- Signaturset aus 1.500 Screenshots von 811 Webseiten
 - Unterschiedliche Werbung
 - Unterschiedliche Sprachen

- Testset aus 50.000 Screenshots
 - 74.990.258 mögliche falsche Zuordnungen
 - 9.742 mögliche richtige Zuordnungen

ALGORITHMEN: EHD

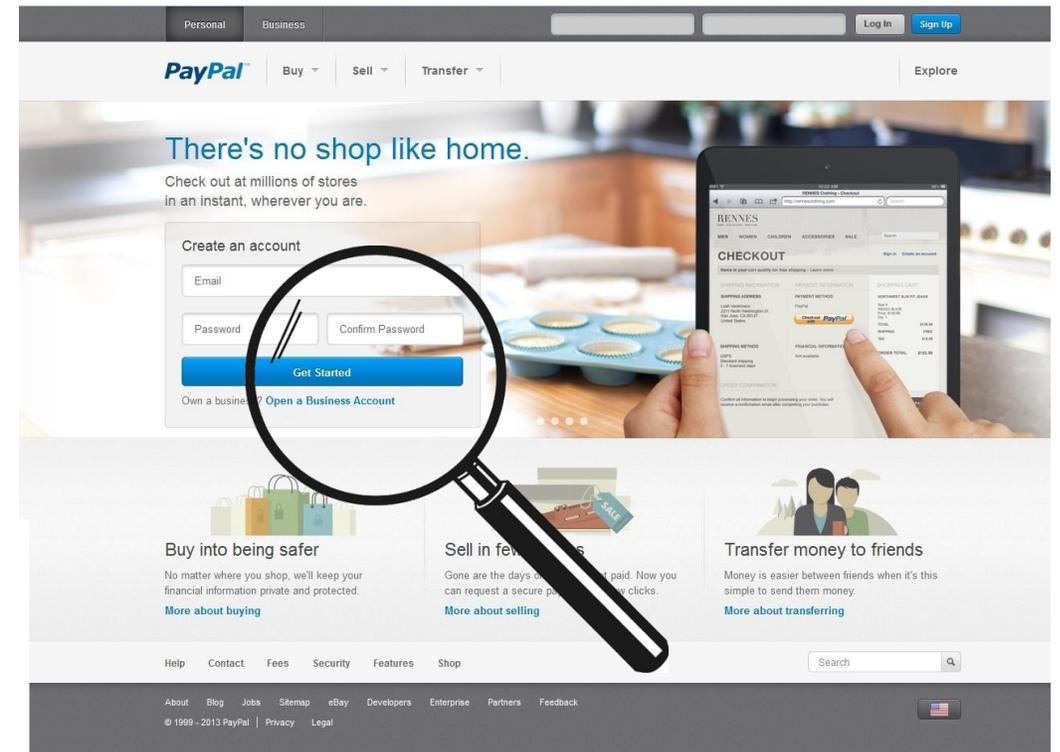
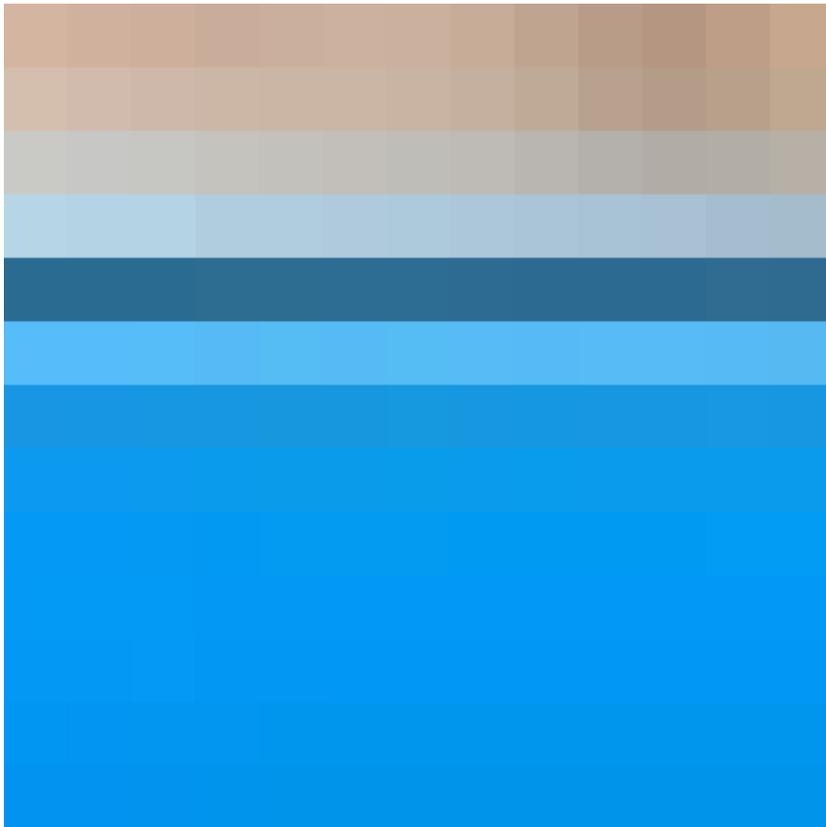
- Im MPEG-7 Standard entwickelt
- Betrachtet Kantenverteilung
 - Globale Kantenverteilung
 - Kanten in 16 Unterbildern
- 5 mögliche Kantenrichtungen
- Farben werden ignoriert
- 45% Erkennung im Test

ALGORITHMEN: ACC

- Betrachtet Farbnachbarschaft
 - Nachbarschaft eines Bildpunktes
 - Bis zu 3 Punkte entfernt
 - Zählt nur Punkte gleicher Farbe
- Positionsunabhängig
- 69% Erkennung im Test

ALGORITHMEN: ACC

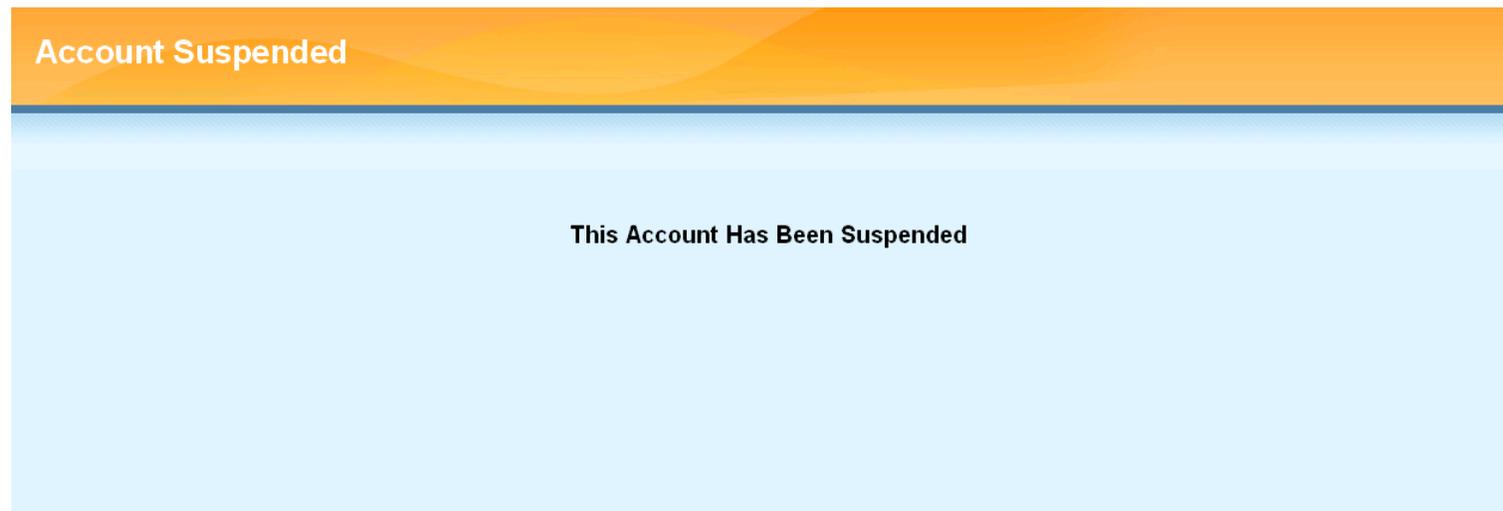
- Betrachtet Farbnachbarschaft



EVALUATION: EHD

- Grundannahme:
 - Kantenerkennung reicht, um eine Webseite zu identifizieren
 - Probleme bei gleich aufgebauten Webseiten
- Es kommt zu fehlerhaften Zuordnungen mit geringer Distanz
 - Algorithmus ist nicht automatisiert nutzbar

EVALUATION: EHD



EVALUATION: ACC

- Grundannahme:
 - Farbnachbarschaft genügt, um eine Webseite zu identifizieren
 - Probleme bei gleichfarbigen Seiten
- Es kommt zu fehlerhaften Zuordnungen mit geringer Distanz
 - Webseiten mit ausschließlich Textinhalten werden einander zugeordnet
 - Signaturset darf keine Webseiten ohne Farbinhalt enthalten

EVALUATION: ACC

Strict Standards: Non-static method JLoader::import() should not be called statically in `/home/octify/public_html/libraries/joomla/import.php` on line 29

Strict Standards: Non-static method JLoader::register() should not be called statically in `/home/octify/public_html/libraries/loader.php` on line 71

Strict Standards: Non-static method JLoader::import() should not be called statically in `/home/octify/public_html/libraries/joomla/import.php` on line 32

ATENÇÃO

O endereço IP de sua conexão encontra-se bloqueado neste servidor.
Leia atentamente as informações abaixo para evitar recorrência deste bloqueio.

Por que isto ocorre?

Os principais motivos de bloqueios são

1. Erros consecutivos de senha ou usuário no login nos serviços **cPanel, WebMail, WHM, FTP, E-mail (SMTP, POP e IMAP)** ou **diretórios protegidos pelo serviço HTTP (sites protegidos por senha)**.
2. Contas de email configuradas com senha ou login incorretos em clientes de email local tais como Outlook, Thunderbird ou similares.
3. Ocorrência de alguma ação suspeita ou acesso não autorizado a partir do seu IP.
4. Excesso de conexões ao serviço de email **POP3** (acima de 60 conexões por hora), para prevenir este tipo de bloqueio aumente o intervalo de checagem de novos e-mails em seu cliente de e-mail para ao menos uma vez a cada 5 minutos.
5. **Mais de 50 conexões simultâneas ao Serviço HTTP (acesso a páginas)** a partir de um mesmo endereço IP, comum em usuários de internet a rádio ou em empresas que compartilham a mesma conexão de internet para todos os usuários da rede.

O que devo fazer?

A maioria dos bloqueios são temporários e tem limite de aproximadamente 5 minutos, caso continue a ver esta mensagem mesmo após este período entre em contato com seu serviço de hospedagem.

Dados de seu acesso

IP bloqueado: **91.52.210.197**

Site acessado: **www.transpavimentacao.com.br**

Data: **Sexta-feira, 20 de Dezembro de 2013**

Caso seja necessário entrar em contato com o suporte informe os dados acima.

AUSBLICK

- Automatisierte Erkennung im Browser
- Erkennung von Phishingtrends / neuen Kampagnen
 - Clustering der Daten eines Eingangstroms

FORSCHUNGSSTAND

- Visuelle Erkennung möglich
- ABER: In der Praxis nicht eigenständig nutzbar
 - Zu viele falsche Erkennungen (11%)

FORSCHUNGSSTAND

- Visuelle Erkennung möglich



- ABER: In der Praxis nicht eigenständig nutzbar

- Zu viele falsche Erkennungen (11%)

