

Ostbayerische Technische Hochschule Regensburg
Fakultät für Informatik und Mathematik

Wissenschaftliche Abhandlung

Möglichkeiten zur Bestimmung der visuellen Ähnlichkeit zwischen zwei Bildern

Vorgelegt von: Andreas Huber <andreas.huber@st.oth-regensburg.de>

Matrikelnummer: 3370380

Studiengang: Master Informatik (Schwp. Software Engineering)

Betreuer: Prof. Dr. techn., Dipl.-Ing. Markus Kucera

Abgabedatum: 5. Dezember 2022

Inhaltsverzeichnis

Abbildungsverzeichnis	II
1 Einleitung	1
2 Inhaltsbasierte Algorithmen zur Ähnlichkeitsbestimmung zweier Bilder	1
2.1 Mittlere quadratische Abweichung (Mean squared error)	2
2.2 Perceptual Hashing	3
2.3 Histogram Intersection	4
2.4 Scale-Invariant Feature Transform (SIFT)	5
2.5 Structural Similarity Index (SSIM)	6
3 Fazit und Ausblick	7
Quellenverzeichnis	8
Anhang	8

Abbildungsverzeichnis

- 1 Einleitung
- 2 Inhaltsbasierte Algorithmen zur Ähnlichkeitsbestimmung zweier Bilder

2.1 Mittlere quadratische Abweichung (Mean squared error)

2.2 Perceptual Hashing

Eine der einfachsten, aber auch anfälligsten Methoden zur Bestimmung der Ähnlichkeit von zwei Bildern ist das Hashing. Dabei gibt es viele verschiedene Ansätze und Vorgehensmodelle. Die gängigste Methode ist die Anwendung des pHashes, auch genannt Perceptual Hashing. (**hashing-apiumhub**)

Beim Perceptual Hashing werden sowohl für Bild A als auch für Bild B ein Hash berechnet. Die daraus resultierende Hamming-Distanz zwischen den Hashes ergibt die Ähnlichkeit der Bilder. Je geringer die Distanz, desto ähnlicher. Das Verfahren ist nicht normiert und noch ein offenes Forschungsthema. (**hashing-phash**)

Im Folgenden wird ein beispielhaftes pHash-Verfahren erläutert. Zuerst wird sowohl das Referenz- als auch das Suchbild in eine Graustufen-Grafik umgewandelt. Schließlich wird die Grafik auf 32x32 Pixel skaliert. Auf das entstandene Grauwertbild folgen zwei Diskrete Kosinus Transformationen (1. Pro Zeile, 2. Pro Spalte). Die hochfrequenten Abschnitte befinden sich nun links oben in einer 8x8 Matrix. Daraufhin wird der Median-Grauwert der 64 Pixel berechnet. Jeder Pixel, dessen Grauwert unter dem Durchschnitt liegt wird weiß eingefärbt, der Rest schwarz. Daraus ergibt sich ein 64-Bit langer Hashwert (schwarz: 0, weiß: 1). Zuletzt wird durch eine Subtraktion der beiden generierten Bitfolgen die Hamming-Distanz berechnet. Mittels einem vorher festgelegten Schwellenwert wird jetzt auf Basis des Hamming-Abstands bestimmt, ob die Bilder als ähnlich eingestuft werden. (**hashing-apiumhub**)

Vorteile

- Schnelle Performance, wenn die Hashes der Referenzbilder bereits in einer dafür geeigneten Datenstruktur (z.B. k-d-Bäume, VP Bäume oder Kugelbäume) vorliegen (**hashing-lvngd**)
- Einfach zu implementieren
- Keine Trainingsdaten nötig
- Robust gegen Wasserzeichen, Farbfilter, leichte Helligkeits- und Kontraständerungen, Gammakorrekturen, Skalierungen sowie Komprimierungen (**hashing-phash**)

Nachteile

- Nicht robust gegen Spiegelungen, Rotierungen und Verzerrungen
- Nicht robust gegen Zuschneidungen, neu eingefügten Elementen oder Änderungen des Blickwinkels

2.3 Histogram Intersection

Ein robusteres, aber auch rechenaufwendigeres Verfahren ist der Histogram Intersection Algorithmus von Swain und Ballard. Histogramme sind stabil gegenüber Translation und Rotation um die Betrachtungsachse. Außerdem ändern sie sich bei Änderungen des Blickwinkels, des Maßstabs und bei Verdeckung von Elementen nur langsam. Gegen Belichtungsänderungen können stark reduzierte Farbhistogramme jedoch Probleme bereiten.

Im Standardfall wird die Histogram Intersection auf Farbhistogramme angewandt. Zuerst muss die Menge an zu diskretisierenden Farben im Histogramm festgelegt werden - zum Beispiel 100. Folglich gibt es 100 mögliche „Farbeimer“ in die wir jeden Pixel der Bilder einsortieren. Mit den durch das Referenz- und das Suchbild berechneten Histogrammen, kann schließlich die Überschneidung festgestellt werden. Je stärker sich die Histogramme überschneiden, desto ähnlicher sind die Bilder. Zusätzlich wird wie beim Hashing ein gewisser Schwellenwert vorher definiert. Der Vorgang kann beispielsweise auch auf die einzelnen Farbkanäle aufgeteilt werden. Bei RGB würde das bedeuten, dass man jeweils den Rot-, Grün-, und Blau-Kanal einzeln betrachtet.

Weitere Untersuchungen haben ergeben, dass sowohl die Wahl des Farbraums als auch die Festlegung des Quantization Levels (Anzahl der Farbeimer) eine große Rolle bei der Erfolgsschance dieses Vorgehensmodells spielen. Hierbei sorgte wohl der CIELab-Farbraum im Allgemeinen für die besten Ergebnisse.

Eine zusätzliche Möglichkeit die Genauigkeit des Algorithmus zu verbessern ist das Hinzufügen eines sogenannten „Texture Direction“-Histogramms. In diesem Fall wird durch eine Eckenerkennung die Struktur und Richtungen des Bildes bestimmt. Diese werden wie bei den Farbhistogrammen auch in einem Histogramm mit vorher definierten Bins einsortiert. Abschließend kann, wie oben, beschrieben die Überschneidung und somit die Ähnlichkeit der Struktur zweier Bilder mithilfe des „Histogram Intersection“-Algorithmus von Swain und Ballard berechnet werden. Offensichtlich wird jedoch jeder weitere zusätzlicher Berechnungsschritt die allgemeine Performance des Systems beeinträchtigen.

2.4 Scale-Invariant Feature Transform (SIFT)

David G. Lowe hat mit der Einführung des SIFT-Algorithmus einen für die Bildverarbeitung wertvollen Beitrag geleistet. Der SIFT-Algorithmus findet lokale Merkmale in Bildern. Diese Merkmale sind invariant gegenüber Bildskalierung und -drehung und teilweise invariant gegenüber Änderungen der Beleuchtung sowie des Betrachtungswinkels. Die Wahrscheinlichkeit einer Störung durch Verdeckung einzelner Elemente oder auftretenden Bildrauschens wird, durch räumliche Streuung und einem vielseitigen Frequenzbereich, erreicht. Im Folgenden wird die grobe Vorgehensweise zur Findung der lokalen Merkmale erläutert.

1. Scale-space extrema detection

Zu Beginn wird aus einem Bild durch repetitives Unschärfen (mittels Gaußschen Unschärfefilter) und Halbieren der Größe eine Serie an Bildern erzeugt. Auf die Bilder der Bilderserie folgt die Anwendung der Gaußschen Differenzfunktion.

2. Keypoint localization

Aus den vorher differenzierten Grafiken werden jetzt mit dem Marr-Hildreth-Operator interessante Schlüsselpunkte bzw. Merkmale herausgefiltert. Eine dem Harris-Corner Detector ähnliche Prozedur wird angewandt, um Punkte auf Linien zwischen Ecken und Punkte mit wenig Kontrast (uninteressante Punkte) herauszusieben.

3. Orientation assignment

Mithilfe eines Histogramms kann den Schlüsselpunkten nun eine Orientierung zugewiesen werden. Dadurch werden die Punkte neben der Robustheit gegenüber Skalierung und Translation auch gegen Rotation invariant.

4. Keypoint descriptor

Zuletzt wird nach Anwendung einiger mathematischer Verfahren eine Art eindeutiger Fingerabdruck des Merkmals berechnet. Schließlich wird dadurch die Stabilität gegenüber begrenzter lokaler Formverzerrungen und Beleuchtungsänderungen erreicht.

Nach Anwendung dieses Algorithmus muss man bei einem 500x500 Pixel großen Bild mit in etwa 2000 robusten Merkmalen rechnen. Werden die gefundenen Merkmale in einer Datenbank gespeichert, können sie schließlich in nahezu Echtzeit auf die Schlüsselpunkte eines Suchbilds verglichen werden.

Ein ähnlicher patentierter Algorithmus zur Erkennung von Bildmerkmalen ist der „Speeded up robust features (SURF)“. Ferner gibt es viele ähnliche frei verfügbare Algorithmen der Gruppe „Bildmerkmalsfindung“.

2.5 Structural Similarity Index (SSIM)

Anders als die meisten anderen Vorgehensmodelle versucht die menschliche Wahrnehmung nicht Fehler zu Quantifizieren, sondern strukturelle Informationen aus einer Szene wiederzuerkennen. Um dieses Verhalten nachzuahmen, haben Forschende Mitglieder der IEEE den Structural Similarity Index entworfen.

Der Index extrahiert, vergleicht und kombiniert folgende drei Parameter:

- Leuchtdichte
- Kontrast
- Struktur

Im ersten Schritt wird die Leuchtdichte durch Mittelwertbildung über alle Pixelwerte gemessen. Anschließend wird der Kontrast durch Berechnung der Standardabweichung aller Werte ermittelt. Vereinfacht dargestellt wird die Struktur durch die Teilung des Eingangssignals durch seine Standardabweichung geteilt, sodass eine genormte Standardabweichung entsteht.

Sobald die drei Werte berechnet wurden, werden diese anhand, im Original-Paper, veröffentlichten mathematischen Funktionen verglichen und schließlich kombiniert.

Obendrein sei es nützlich den Index nicht global auf das Bild anzuwenden. Das Paper gibt an den Algorithmus vorzugsweise auf verschiedene lokale Stellen einzusetzen. Erstens seien für die Analyse interessante Objekte oft instationär und zweitens können Bildverzerrungen auch räumlich variabel sein. Um die Referenz zur Nachahmung der menschlichen Wahrnehmung zu ziehen, ist ebenfalls festzustellen, dass Menschen ebenfalls nur lokale Bereiche einer Grafik in hoher Auflösung wahrnehmen können.

Auch wenn der SSIM durch seine Einfachheit, gegenüber anderen Verfahren, eine überragende Performance aufweist, ist der Algorithmus sehr anfällig für Skalierungen, Translationen und Rotierungen.

Aufbauend auf den SSIM-Algorithmus gibt es weitere optimierende Forschungen, wie zum Beispiel der Complex Wavelet Structural Similarity Index. Der CW-SSIM adressiert die vorher genannten Probleme und erzeugt durch Verwendung von Wavelet-Transformationen eine Robustheit gegenüber kleinen Translationen und Rotierungen. (**ssim-complex-wavelet**)

3 Fazit und Ausblick

Anhang