

Diss. ETH No. 23464

Wearable barcode scanning
**Advancements in visual code localization, motion blur
compensation, and gesture control**

A thesis submitted to attain the degree of
DOCTOR OF SCIENCES of ETH ZURICH
(Dr. Sc. ETH Zurich)

presented by
GÁBOR SÖRÖS

M.Sc. in Electrical Engineering
Budapest University of Technology and Economics
born on 02.06.1986, citizen of Hungary

accepted on the recommendation of
Prof. Dr. Friedemann Mattern, examiner
Prof. Dr. Otmar Hilliges, co-examiner
Prof. Dr. Dieter Schmalstieg, co-examiner

2016

Abstract

Visual codes like barcodes and quick response (QR) codes are the most prevalent linking elements between physical objects and digital information. They are found in numerous consumer applications such as shopping, electronic payments, ticketing, or marketing campaigns, but they are also found in logistics and enterprise asset tracking and provide employees access to detailed service records. Reading visual codes is also often the first step for pairing and interaction with physical appliances in various research projects in the area of pervasive computing. While these codes are found almost everywhere, the reading of codes usually requires expensive scanner devices which hinders the original goal of easy access to information about every physical object.

Technological advancements in wearable computing and mobile computer vision may radically expand the adoption of visual codes because smartphones, smartwatches, smartglasses, and other wearables enable instant barcode scanning on the go. The wide accessibility, the computing performance, the intuitive user interface, and the relatively low price make personal wearable computers strong competitors for traditional scanners while they also enable new use cases for visual codes. As these devices are primarily designed for other purposes, there are also a few shortcomings.

In this dissertation, we describe methods to overcome these shortcomings and to add advanced features that can make wearable barcode scanning an attractive alternative to traditional barcode scanning even outside the consumer domain. We present fast and robust solutions to the following problems on computationally restricted unmodified wearable devices:

(i) Fast and robust localization of visual codes: Current smartphone-based barcode scanning solutions require the user to hold and align the tagged object close to the camera. This is especially problematic with smartglasses and leads to lower user acceptance. On the other hand, today's wearable cameras have a high enough resolution to scan visual codes that are further away, once they are segmented in a preprocessing step. We propose a fast algorithm for joint 1D and 2D visual code localization in large digital images. The proposed method outperforms other solutions in terms of accuracy while it is invariant to scale, orientation, code symbology, and is more robust to blur than previous approaches.

We further optimize for speed by exploiting the parallel processing capabilities of mobile graphics hardware for image processing. Fast segmentation allows for scanning multiple codes at the same time and thus helps in application scenarios where interaction with multiple objects is necessary.

(ii) Fast and robust compensation of motion blur: When the camera or the code undergoes slight motion during scanning, in contrast to the performance of laser scanners, wearable cameras often suffer from motion blur that renders the codes unreadable. We build upon existing work in photograph deblurring and develop a fast algorithm for scanning motion-blurred QR codes on mobile devices. We exploit the fact that QR codes do not need to be visually pleasing for decoding and propose a fast restoration-recognition loop that exploits the special structure of QR codes. In our optimization scheme, we interweave blind blur estimation from the edges of the code and image restoration regularized with typical properties of QR code images. Our proposed restoration algorithm is on par with the state of the art in quality while it is about a magnitude faster. We also propose to combine blur estimation from image edges with blur estimation from built-in inertial sensors to make the restoration even faster. Fast blur compensation means that no precise code alignment is required but the user can simply swipe the camera in front of the code.

(iii) Fast and robust recognition of in-air hand gestures: Wearable devices have limited input capabilities; the way of interaction is usually limited to a few buttons or slim touchpads. We add natural hand gesture input to our wearable computers, but indirectly also to other smart appliances in our smart environment that can be automatically recognized through tiny visual codes, and that can “outsource” their own user interface to our wearable devices. We present a machine learning technique to recognize hand gestures with only a single monocular camera as can be found on off-the-shelf mobile devices. The algorithm robustly recognizes a wide range of in-air gestures and runs in real time on unmodified wearable devices. We further show that, with little modification, our method can not only classify the gesture but can also regress the distance of the hand from the camera. 3D in-air gesture control allows hands-free scanning with smartglasses which brings many advantages in enterprise scenarios. Furthermore, through user interface outsourcing, it also enables expressive vision-based gesture control even to those appliances that do not possess a camera by their own.

Along with the dissertation, we created several showcase scenarios and demonstrators of our contributions. All proposed algorithms have been designed and implemented to be compatible with various platforms and device families (e.g., PCs, tablets, smartphones, smartwatches, smartglasses), with their resource constraints in mind. The solutions presented in this dissertation are pushing forward the state of the art in terms of accuracy, speed, and robustness, and thus help to make wearable barcode scanning a promising alternative to traditional barcode scanning.

Kurzfassung

Visuelle Codes, wie beispielsweise Strichcodes oder QR-Codes, werden im Rahmen zahlreicher mobiler Anwendungsszenarien genutzt, sie stellen optisch erkennbare Marker als Verbindungselemente zwischen Objekten der physischen Welt und digitaler Information dar. Vom intelligenten Einkaufsassistenten über elektronische Zahlungen, Tickets, Marketingkampagnen bis hin zu Anwendungen in Logistik und Güteverkehr unterstützen visuelle Codes eine große Zahl an Einsatzmöglichkeiten. Auch bei vielen aktuellen Forschungsvorhaben im Bereich des *pervasive computing* ist das Erfassen visueller Codes der erste Schritt zur Interaktion mit physischen Objekten und Geräten in der näheren Umgebung. Doch obwohl visuelle Codes mittlerweile auf praktisch allen Produkten und vielen anderen Dingen zu finden sind, benötigt das optische Erkennen der Codes und das Auslesen der darin hinterlegten Daten (“scannen”) bisher noch dedizierte und teure, meist laserbasierte Lesegeräte. Der intendierte Zweck eines unmittelbaren und ubiquitären Zugangs zur damit verknüpften Information wird dadurch infrage gestellt.

Das zunehmende Interesse an visuellen Codes wird in der jüngeren Zeit auch durch rasante technische Entwicklungen im Bereich des *wearable computing* sowie des mobilen maschinellen Sehens induziert. Smartphones, Smartwatches, Smartglasses und andere tragbare Geräte bringen mit ihren eingebauten kleinen Kameras sowie ihren leistungsfähigen CPUs alle Voraussetzungen für eine effiziente Codeerkennung mit. Aufgrund ihrer breiten Verfügbarkeit, hohe Rechenleistung, intuitiven Mensch-Maschine-Schnittstelle sowie ihres günstigen Preises konkurrieren sie mit den traditionellen Lesegeräten für visuelle Codes und ermöglichen gleichzeitig neue Anwendungsszenarien. Da diese neueren mobilen Geräte aber vornehmlich für andere Verwendungszwecke entwickelt wurden, weisen sie in Hinblick auf die Codeerkennung als spezielle Eigenschaft allerdings einige markante Unzulänglichkeiten auf.

In der vorliegenden Dissertation beschreiben wir Methoden, die diese Defizite nicht nur überwinden, sondern darüber hinaus auch einige neue und interessante Funktionen ermöglichen. Damit wird der traditionellen Strichcodeerkennung eine attraktive Alternative gegenübergestellt, welche auch in vielen Bereichen außerhalb des Konsumentensektors, vor allem im Industrieinsatz, zusätzlichen Nutzen stiften sollte. Im Wesentlichen stellen wir schnelle und robuste Lösungen für die folgenden Problembereiche bei der Verwendung leistungsbeschränkter (und generell ressourcenlimitierter) *wearable devices* vor:

(i) Schnelle und robuste Lokalisierung von visuellen Codes: Gängige Smartphone-basierte Scannerlösungen verlangen, dass der Nutzer das Produkt dicht vor die Kamera hält und sorgfältig positioniert. Dies ist insbesondere bei Smartglasses problematisch und führt zu einer geringeren Akzeptanz. Andererseits verfügen die heutigen digitalen Kameras über eine genügend hohe Bildauflösung, um visuelle Codes auch aus großer Entfernung zu lesen, sofern die Codes durch einen vorverarbeitenden Schritt segmentiert werden. Um dieses Potential zu nutzen, entwickeln wir einen schnellen Algorithmus für die gleichzeitige visuelle Lokalisierung mehrerer ein- und zweidimensionalen Codes in großen digitalen Bildern. Die realisierte Methode übertrifft andere Lösungen an Genauigkeit, während sie bezüglich Skalierung, Orientierung und Symbologie der Codes invariant ist. Zusätzlich ist sie robuster gegen Unschärfe (*blur*) als bisherige Ansätze. Unter Ausnutzung der Parallelisierungsfähigkeiten moderner mobiler Graphikhardware zeigen wir, dass die Geschwindigkeit weiter optimiert werden kann. Die schnelle Codesegmentierung erlaubt das gleichzeitige Lesen mehrerer Codes und unterstützt Situationen, bei denen eine Interaktion mit mehreren Objekten erfolgen kann.

(ii) Schnelle und robuste Kompensation von Bewegungsunschärfe: Wenn Kamera oder Codemarker während des Scannens bewegt werden, verursacht dies im Unterschied zu den traditionellen laserbasierten Scannergeräten oft ein durch Bewegungsunschärfe verzerrtes Bild, das die codierte Information unleserlich macht. Um diesem Problem zu begegnen, bauen wir auf existierenden Ansätzen aus der Bildbearbeitung im Bereich *deblurring* auf und entwickeln einen schnellen Algorithmus, um unscharfe QR-Codes direkt auf mobilen Geräten lesen zu können. Wir profitieren dabei von der Tatsache, dass QR-Codes für ein erfolgreiches Decodieren nicht visuell ansprechend sein müssen und schlagen eine schnelle Restaurierungs- und Erkennungsschleife vor, welche die spezielle Struktur von QR-Codes ausnutzt. In unserem Optimierungsschema verbinden wir eine mittels dominanter Kanten erzielte Unschärfeschätzung des Bildes mit der Bildrestaurierung unter Ausnutzung typischer Eigenschaften von QR-Codes. Die von uns entwickelte Restaurierungsmethode entspricht qualitativ dem aktuellen Stand der Technik, ist aber um eine Größenordnung schneller. Darüber hinaus schlagen wir vor, die kantenbasierte Unschärferechnung mit einer Unschärfeschätzung mittels der in mobilen Geräten oft verbauten Trägheitssensoren zu kombinieren, um die Restaurierung weiter zu beschleunigen. Die von uns entwickelte Unschärfekompensation bedeutet, dass keine präzise Positionierung der Codemarkers mehr erforderlich ist und der Nutzer die Kamera einfach über den Code ziehen kann.

(iii) Schnelle und robuste Erkennung von Handgesten: Kleine Geräte haben oft limitierte Ein- und Ausgabemöglichkeiten, die Interaktion ist dabei auf wenige Tasten oder schmale Touchpads beschränkt. Aus diesem Grund haben wir eine Handgesteneingabe für mobile Geräte entwickelt, welche indirekt sogar von Objekten ohne Kamera und explizi-

te Bedienschnittstellen genutzt werden kann. Automatisch erkennbare kleinste visuelle Codes erlauben den beteiligten Objekten, ihre eigenen “virtuellen” Bedienelemente an andere mit Nutzerschnittstelle ausgestatteten mobilen Geräte auszulagern. Wir stellen ein maschinelles Lernverfahren vor, welches Handgesten mit einer einfachen monokularen Kamera, wie sie in den meisten mobilen Geräten verbaut ist, erkennen kann. Der Algorithmus läuft in Echtzeit auf physisch unmodifizierten Geräten und kann eine breite Auswahl von Handgesten zuverlässig erkennen. Des Weiteren zeigen wir, dass unsere Methode mit kleinen Änderungen nicht nur die Gesten klassifizieren, sondern auch die Distanz der Hand zur Kamera schätzen kann. Die Gestenerkennung erlaubt somit das Lesen von Strichcodes ohne Gebrauch der Hände; ein Vorteil, der besonders im kommerziellen Bereich zur Geltung kommt. Durch Bedienungsauslagerung wird ferner auch Geräten, welche selbst keine Kamera aufweisen, eine ausdrucksvolle Gestensteuerung ermöglicht.

Im Rahmen der Dissertation wurden mehrere Demonstratoren für diverse Anwendungsszenarien erstellt. Von allen beschriebenen Algorithmen liegen Implementierungen für unterschiedliche Plattformen und Geräte (z.B. PCs, Tablets, Smartphones, Smartwatches, Smartglasses) vor, welche die jeweiligen technischen Einschränkungen beachten. Die in der vorliegenden Dissertation präsentierten Verfahren zur mobilen Codeerkennung gehen hinsichtlich Genauigkeit, Geschwindigkeit und Robustheit deutlich über den bisherigen Stand der Technik und Wissenschaft hinaus. Damit stellt die kamerabasierte Codeerkennung mit alltäglichen mobilen Geräten nunmehr eine vielversprechende Alternative zu den traditionell verwendeten spezielleren Lesesystemen dar.

Kivonat

Valós tárgyak és digitális információ között mai világunkban a legelterjedtebb összekötő elemek a vizuális kódok, mint például a vonalkódok és a quick response (QR) kódok. A vizuális kódok számos alkalmazásával találkozhatunk a bevásárlástól az elektronikus fizetéseken és jegyvásárláson át a reklámkampányokig, de mindenütt megtalálhatóak a logisztika és az áruforgalom területén is. A *pervasive computing* területén sok kutatási projektben a fizikai eszközök konfigurálásában és az azokkal való interakcióban is vizuális kódok leolvasása az első lépés. Noha a vizuális kódok majdnem mindenhol jelen vannak, leolvasásuk általában drága olvasóeszközöket (*scannereket*) igényel, és ez meg-
hiúsítja a kódok használatának eredeti célját, hogy bármely tárgyról egyszerűen és gyorsan információhoz juthassunk.

A vizuális kódok további elterjedését jelentősen elősegíthetik a viselhető számítástechnika (*wearable computing*) és a mobil gépi látás (*mobile computer vision*) technikai vívmányai. Az okostelefonok, okosórák, okoszemüvegek és más viselhető eszközök lehetővé teszik, hogy bárhol és bármikor leolvassunk vizuális kódokat. A széleskörű elérhetőségnek, a fejlett számítási teljesítménynek, az intuitív felhasználói felületnek és a viszonylag alacsony árnak köszönhetően a viselhető számítástechnikai eszközök komoly konkurenciát jelenthetnek a hagyományos leolvasóknak, egyúttal utat nyithatnak a kódok számos új alkalmazásának is. Mivel azonban ezeket az eszközöket eredetileg más felhasználási célokra tervezték, kódolvasóként használva őket néhány területen hiányosságokat is mutatnak.

Jelen disszertációban olyan módszereket mutatunk be, amelyek ezeket a hiányosságokat áthidalják, és ezen felül egyéb új funkciókat is lehetővé tesznek. Így a viselhető vonalkódolvasás (*wearable barcode scanning*) új alternatívát kínál a hagyományos leolvasókkal szemben mind a fogyasztók körében, mind pedig ipari alkalmazásokban. A disszertációban gyors és hatékony módszereket mutatunk be a következő problémákra korlátozott számítási teljesítményű, módosíthatlan viselhető eszközökön:

(i) Gyors és hatékony kódlokálizáció: A mai okostelefon-alapú leolvasó megoldások megkövetelik, hogy a felhasználó a kódot stabilan, közel a kamera előtt tartsa. Ez gondot okozhat okoszemüvegek esetében, mert a felhasználók körében nemtetszést válthat ki. Másrészt a viselhető kamerák elég magas képfelbontással rendelkeznek ahhoz, hogy a

kamerától távolabb található vizuális kódokat is képesek legyenek leolvasni, ha azokat egy előzetes feldolgozó lépésben szegmántáljuk. Egy olyan gyors és hatékony algoritmust mutatunk be, amely nagy digitális képeken tud egyszerre egy-, és kétdimenziós vizuális kódokat lokalizálni. Az új módszer felülmúlja a korábbiakat pontosságában, miközben a kód méretét, orientációját, és a kód típusát tekintve nem támaszkodik előfeltételekre, valamint kevésbé érzékeny az elmosódottsággal szemben. A modern, mobil grafikus hardverek párhuzamos számítási képességeit kihasználva az algoritmus sebességét tovább lehet növelni. A gyors kódszegmentáció lehetővé teszi egyszerre több kód leolvasását is, amely olyan esetekben különösen hasznos, amikor egyszerre több tárggyal való interakció szükséges.

(ii) Gyors és hatékony elmosódás-helyreállítás: Ha leolvasás közben a kamera vagy a kód kis mértékben elmozdul, a lézerolvasókkal ellentétben a viselhető kamerák gyakran olvashatatlan, elmosódott képet rögzítenek. Ennek a problémának a kiküszöbölésére egy olyan gyors és hatékony algoritmust fejlesztettünk ki, amely képes rendkívül elmosódott QR kódokat is leolvasni. Az algoritmus a fényképretranszformálásban is alkalmazott módszereken alapul, amelyek minden lépését a QR kódok egyedi tulajdonságaihoz igazítottuk. Azt a tényt használjuk ki, hogy a QR kódoknak nem szükséges vizuálisan esztétikusnak lenniük ahhoz, hogy dekódolhatóak legyenek. A módszerünk alapja egy gyors javító-dekódoló ciklus, amely kihasználja a QR kódok speciális struktúráját. Az optimalizációs ciklusban felváltva becsüljük a kép elmosódottságát a kód éleiből és javítjuk a kép minőségét QR-típusú regularizációt alkalmazva. Ez a rekonstrukciós módszer minőségben a jelenlegi technikákkal azonos szintet képvisel, ám egy nagyságrenddel gyorsabb azoknál. Az elmosódottságot nemcsak az élekből, hanem a viselhető eszközökbe beépített mozgás-szenzorokból is becsülhetjük, hogy a helyreállítási folyamatot még gyorsabbá tegyük. A gyors elmosódás-kompenzáció lehetősége azt jelenti, hogy a leolvasás során nem szükséges többé a kód pontos pozícionálása, a felhasználó egyszerűen elhúzhatja a kamerát a kód előtt.

(iii) Gyors és hatékony gesztusfelismerés: A viselhető eszközök korlátozott be- és kiviteli lehetőségekkel rendelkeznek, az interakció többnyire pár gombra vagy vékony érintőképernyőkre korlátozódik. Ezért egy, a természetes kommunikációhoz sokkal közelebb álló, gesztikuláción alapuló beviteli módszert fejlesztettünk ki, amely indirekt módon akár más, az “okos” környezetünkben található eszközökhöz is használható. Az automatikusan felismerhető apró kódok lehetővé teszik a kommunikációban résztvevő eszközök számára, hogy saját kiszolgáló elemeiket a felhasználó viselhető eszközeire “ruházzák át”. Egy olyan tanuló algoritmust mutatunk be, amely a gesztusokat egy egyszerű, monokuláris kamera segítségével is felismeri, amellyel a legtöbb viselhető eszköz rendelkezik. Az algoritmus a kézi gesztusok széleskörű felismerésére képes, és valós időben fut módosíthatatlan viselhető eszközökön. Azt is bemutatjuk, hogy módszerünk kis változtatással nemcsak a

gesztusokat tudja felismerni, hanem a kéz és a kamera távolságát is meg tudja becsülni, és ezáltal háromdimenziós gesztusbevitelt tesz lehetővé. Módszerünk lehetővé teszi a vonalkódok leolvasását gombok és érintőképernyő használata nélkül, amely főként ipari alkalmazásokban jelenthet nagy előnyt, valamint a kezelőfelület viselhető eszközökre való átruházásával lehetővé teszi a gesztikuláción alapuló vezérlést olyan eszközökön is, amelyek maguk nem rendelkeznek kamerával.

A disszertációhoz szemléltető példákat és bemutatókat is készítettünk. Az algoritmusokat úgy terveztük és alkottuk meg, hogy különböző platformokon és eszközökön is működjenek, például személyi számítógépeken, okostelefonokon, okosórákon és okoszeművegeken – azok minden technikai korlátjára gondolva. Az eredmények, amelyeket itt bemutatunk mind pontosság, gyorsaság és megbízhatóság tekintetében jelentős előrelépést jelentenek. Segítségükkel a viselhető vonalkódozolás ígéretes alternatívát kínálhat a hagyományos vonalkódozásokkal szemben.