

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΣΤΗΝ
ΕΠΙΣΤΗΜΗ ΤΗΣ ΠΛΗΡΟΦΟΡΙΑΣ



Αναγνώριση Αντικειμένων για τα «Πράγματα» του Internet

Εργασία της φοιτήτριας του Β΄ εξαμήνου του ΠΜΣ

Αφροδίτη Φράγκου

στο πλαίσιο του μαθήματος:

Ηλεκτρονική Δημοσίευση

Υπεύθυνοι καθηγητές: κ. Σαράντος Καπιδάκης

κ. Μανόλης Γεργατσούλης

Αθήνα

Ιούνιος 2008

Περιεχόμενα

Περιεχόμενα.....	2
Περίληψη.....	3
Abstract.....	3
Εισαγωγή.....	4
Μεθοδολογία.....	7
Σχετικές εργασίες.....	7
Υπερσυνδεδεμένες Διαφάνειες: Διαδραστικές Αίθουσες	
Συνεδριάσεων.....	10
Λήψη διαφανειών και εξαγωγή χαρακτηριστικών	13
Σύστημα Αναγνώρισης Διαφανειών.....	14
Πειραματική διαδικασία.....	15
Υπερσυνδεδεμένα Κτίρια: Οδηγός πόλης για κινητό τηλέφωνο.....	18
Οπτικά Στοιχεία και Γεωγραφική Θέση.....	19
Σχεδιασμός Συστήματος.....	20
Μέθοδος Αναγνώρισης Αντικειμένων.....	22
Πειράματα.....	23
Συμπεράσματα και Προοπτικές.....	27
Βιβλιογραφία.....	30
Γλωσσάρι	34
Παράρτημα εννοιών που χρήζουν ανάλυσης.....	35
Beamer	35
Epipolar Geometry και epipolar lines	35
Θεμελιώδης πίνακας F.....	38
Homography.....	39
POST.....	41
RANSAC - RANdom SAMple Consensus.....	41
SIFT - Scale-Invariant Feature Transform	42
Stemming	42
SURF - Speeded Up Robust Features.....	42
Tablet PC	42
Timestamping / Timestamps.....	43
Ευρετήριο Εικόνων	44
Ευρετήριο Πινάκων	45

Περίληψη

Παρουσιάζεται ένα σύστημα που επιτρέπει την αναζήτηση πληροφοριών για φυσικά αντικείμενα μέσω της φωτογράφισής τους. Έτσι, με χρήση ενός κινητού τηλεφώνου με ενσωματωμένη φωτογραφική μηχανή, οι χρήστες μπορούν να αλληλεπιδράσουν με τα αντικείμενα ή «τα πράγματα» κατά τρόπο πολύ απλό. Ένα περαιτέρω πλεονέκτημα είναι ότι τα ίδια τα αντικείμενα δεν είναι απαραίτητο να χαρακτηριστούν με οποιοδήποτε είδος ετικετών (tags). Στον πυρήνα του συστήματος βρίσκεται μια μέθοδος αναγνώρισης αντικειμένων, η οποία προσδιορίζει ένα αντικείμενο με ένα ερώτημα εικόνας μέσω πολλών σταδίων αναγνώρισης, συμπεριλαμβανομένης της αναζήτησης τοπικών οπτικών χαρακτηριστικών, γεωμετρίας σφαιρών, και επίσης, προαιρετικά, μεταδεδομένων, όπως η θέση GPS. Παρουσιάζονται δύο εφαρμογές του συστήματος, μια εφαρμογή χαρακτηρισμού διαφανειών για τις οθόνες παρουσιάσεων στις έξυπνες αίθουσες συνεδριάσεων και μια εφαρμογή οδηγού πόλης για κινητά τηλέφωνα. Και τα δύο συστήματα είναι πλήρως λειτουργικά, συμπεριλαμβανομένης μιας εφαρμογής σε κινητό τηλέφωνο, το οποίο επιτρέπει αλληλεπίδραση με τα αντικείμενα με απλή φωτογράφιση του επιλεγμένου αντικειμένου (point-and shoot interaction). Πειράματα αξιολογούν την απόδοση της προσέγγισής μας και στα δύο σενάρια εφαρμογής και παρουσιάζουν καλά αποτελέσματα αναγνώρισης υπό τους προκλητικούς όρους.

Abstract

We present a system which allows requesting information on physical objects by taking a picture of them. This way, using a mobile phone with integrated camera, users can interact with objects or “things” in a very simple manner. A further advantage is that the objects themselves don’t have to be tagged with any kind of markers. At the core of our system lies an object recognition method, which identifies an object from a query image through multiple recognition stages, including local visual features, global geometry, and optionally also metadata such as GPS location. We present two applications for our system, namely a slide tagging application for presentation screens in smart meeting rooms and a city-guide on a mobile phone. Both systems are fully functional, including an application on the mobile phone, which allows simplest point-and shoot interaction

with objects. Experiments evaluate the performance of our approach in both application scenarios and show good recognition results under challenging conditions.

Εισαγωγή

Η επέκταση του Διαδικτύου στα φυσικά αντικείμενα - το Διαδίκτυο των Πραγμάτων – υπόσχεται στους ανθρώπους τη διαβίωση σε έναν έξυπνο, ιδιαίτερα δικτυωμένο κόσμο, ο οποίος επιτρέπει ένα ευρύ φάσμα αλληλεπιδράσεων με αυτό το περιβάλλον. Μια από τις καταλληλότερες αλληλεπιδράσεις είναι το αίτημα για πληροφορίες που αφορούν φυσικά αντικείμενα. Για αυτόν το λόγο εξετάζονται διάφορες μέθοδοι αυτήν την περίοδο. Οι περισσότερες από αυτές στηρίζονται σε κάποιο είδος μοναδικού δείκτη που ενσωματώνεται στο ή που συνδέεται με το αντικείμενο. Μερικοί από αυτούς τους δείκτες μπορούν να αναλυθούν χρησιμοποιώντας ασύρματη μικρού βεληνεκούς επικοινωνία (για παράδειγμα, RFID tags ή Bluetooth beacons ([Fuhmann and Harbaum, 2003](#)), άλλα είναι οπτικοί δείκτες και μπορούν να αναλυθούν με τη χρήση κάμερας, για παράδειγμα, τυπικούς μονοδιάστατους barcodes ([Adelmann, Langherinrich and Floerkemeier, 2006](#)) ή, με τους πιο σύγχρονους δισδιάστατους barcodes. ([Rohs and Gfeller, 2004](#)).

Μια δεύτερη ανάπτυξη αφορά τις συσκευές εισαγωγής για αλληλεπίδραση με φυσικά αντικείμενα. Τα τελευταία χρόνια τα κινητά τηλέφωνα έχουν γίνει προηγμένοι υπολογιστές πολυμέσων οι οποίοι μπορούν να χρησιμοποιούνται σαν σύγχρονες συσκευές αλληλεπίδρασης με το περιβάλλον του χρήστη. Πέρα από τις προφανείς δυνατότητες τηλεφώνου, οι σύγχρονες συσκευές προσφέρουν ενσωματωμένες κάμερες και ένα ευρύ φάσμα επιπλέον καναλιών επικοινωνίας όπως Bluetooth, WLAN ή πρόσβαση στο internet. Οι άνθρωποι είναι συνηθισμένοι να έχουν τη συσκευή κάθε μέρα μαζί τους. Ταυτόχρονα με τον αριθμό τηλεφώνου η συσκευή έχει ήδη αντιστοιχισθεί σε ένα συγκεκριμένο πρόσωπο. Συνεπώς είναι απολύτως φυσικό να χρησιμοποιούμε το κινητό τηλέφωνό σαν προσωπική συσκευή εισόδου για το internet των πραγμάτων.

Πράγματι κάποιες από τις τεχνολογίες που αναφέρονται παραπάνω έχουν ήδη ενσωματωθεί στα κινητά τηλέφωνα, όπως, για παράδειγμα, οι αναγνώστες Barcode ή οι αναγνώστες RFID. Το απόλυτο σύστημα παρ' όλ' αυτά δεν θα περιοριζόταν σε δείκτες για την αναγνώριση αντικειμένων αλλά περισσότερο θα αναγνώριζε την όψη τους, για παράδειγμα η χρήση οπτικής αναγνώρισης αντικειμένων από την εικόνα της κάμερας ενός

κινητού τηλεφώνου. Από τη στιγμή που μεγάλη πλειοψηφία κινητών τηλεφώνων περιέχει ενσωματωμένη κάμερα, για σημαντική βάση χρηστών μπορεί να κληθεί μονομιάς.

Ένα τέτοιο σύστημα φωτογράφισης ενός αντικειμένου θα ήταν ανεπαρκής για την αναζήτηση όλης της απαραίτητης σχετικής με αυτό πληροφορίας. Ενώ αυτό το όραμα είναι μακριά από την πραγματοποίησή του κάποιοι αφηρημένοι τύποι αντικειμένων με τις πρόσφατες εξελίξεις στο πεδίο των οπτικών υπολογιστών έχει οδηγήσει σε μεθόδους οι οποίες επιτρέπουν την αναγνώριση συγκεκριμένων τύπων αντικειμένων αρκετά αξιόπιστα και τους υπερσυνδέουν με ψηφιακή πληροφορία.

Η χρήση μεθόδων οπτικής αναγνώρισης για την υπερσύνδεση φυσικών αντικειμένων με τον ψηφιακό κόσμο φέρνει συγκεκριμένα πλεονεκτήματα. Για παράδειγμα ορισμένοι τύποι αντικειμένων δεν είναι ορθά συνδεδεμένοι με δείκτες. Αυτό συμπεριλαμβάνει τουριστικά αξιοθέατα τα οποία είναι συχνά μεγάλα κτίρια και ένας δείκτης μπορεί να επισυναφθεί σε ένα ή κάποια σημεία του κτιρίου, πείραμα το οποίο έχει πραγματοποιηθεί στο πλαίσιο του προγράμματος Semapedia¹. Επιπλέον ένας χρήστης μπορεί ν' αναζητήσει πληροφορία εξ αποστάσεως, για παράδειγμα το καμπαναριό μιας εκκλησίας το οποίο βρίσκεται αρκετά μέτρα μακριά. Ακόμα κι αν το αντικείμενο είναι κοντά, οι δείκτες μπορεί να είναι μερικοί. Ένα barcode ή RFID συνημμένο στην ταμπέλα ενός αντικειμένου που εκτίθεται σε Μουσείο θα ήταν δύσκολα προσβάσιμη σε ένα γεμάτο δωμάτιο. Αντίθετα, η λήψη φωτογραφίας του αντικειμένου μπορεί να γίνει από οποιοδήποτε σημείο από το οποίο είναι ορατό. Ακόμη συνεπής χαρακτηρισμός αντικειμένων συχνά είναι δύσκολο να επιτευχθεί. Ένα παράδειγμα είναι οι διαφημιστικές αφίσες εξωτερικού χώρου. Αν μια εταιρεία αφισών θέλει να υπερσυνδέσει όλες τα σημεία στα οποία έχει τοποθετήσει αφίσες θα πρέπει να εγκαταστήσει ένα αναγνωριστικό σήμα RFID ή bluetooth σε κάθε διαφημιστικό πίνακα ή να συνδέσει ένα barcode με καθεμιά από αυτές, η οποία απαιτεί ένα τυποποιημένο σύστημα και οδηγεί σε δαπάνες για εγκατάσταση και συντήρηση. Ένας άλλος τομέας της εφαρμογής είναι οι οθόνες προβολής σε «έξυπνες» αίθουσες συνεδριάσεων ή σε οθόνες πληροφοριών στους δημόσιους χώρους. Το περιεχόμενο που προβάλλεται στην οθόνη αλλάζει συνεχώς και θα ήταν ιδιαίτερα περίπλοκη διαδικασία να προστεθούν δείκτες σε όλο το προβαλλόμενο περιεχόμενο.

¹ <http://www.semapedia.org>

Η χρήση της αναγνώρισης αντικειμένων προκειμένου ο χρήστης να αλληλεπιδράσει με αυτά απαιτεί μόνο μια βάση δεδομένων με τις εικόνες τους. Αυτό συνεπάγεται δε και τ' ότι η αναγνώριση αντικειμένων δεν έχει περιορισμούς. Για παράδειγμα, αυτήν τη στιγμή (και ίσως πάντα) είναι αδύνατο να διακριθούν ιδιαίτερα παρόμοια αντικείμενα, όπως δύο ελαφρώς διαφορετικές εκδόσεις του ίδιου προϊόντος σε ένα κατάστημα. Επιπλέον, η αδυναμία επαρκούς ευρετηρίασης και αναζήτησης των οπτικών χαρακτηριστικών για εκατομμύρια ή δισεκατομμύρια αντικείμενα παραμένει μεγάλο άλυτο πρόβλημα.

Σε αυτή την εργασία παρουσιάζεται μια μέθοδος και ένα σύστημα που επιτρέπει στο «Διαδίκτυο των Πραγμάτων» να χρησιμοποιεί την αναγνώριση αντικειμένων για ορισμένους τύπους αντικειμένων ή «πραγμάτων». Στον πυρήνα του συστήματος του εξυπηρετητή (server), βρίσκεται μια μηχανή ανάκτησης που ευρετηριάζει αντικείμενα χρησιμοποιώντας μια κλίμακα οπτικών χαρακτηριστικών. Οι χρήστες μπορούν να πάρουν μια εικόνα ενός αντικειμένου του ενδιαφέροντός τους, η οποία αποστέλλεται στη μηχανή ανάκτησης. Το αντίστοιχο αντικείμενο αναγνωρίζεται και εκτελείται κάτι σχετικό, π.χ. ανοίγεται ένας ιστοχώρος σχετικός με το αντικείμενο. Το σύστημα ολοκληρώνεται με μια δεύτερη εφαρμογή στον πελάτη (client) που μπορεί να εγκατασταθεί σε ένα κινητό μικροτηλέφωνο και επιτρέπει την αληθινή αλληλεπίδραση ώστε με μία κίνηση να εντοπίζεται το ενδιαφέρον τμήμα, να φωτογραφίζεται και να συνδέεται με τις σχετικές πληροφορίες. Η δεύτερη εφαρμογή είναι ένας οδηγός πόλης για κινητό τηλέφωνο. Οι χρήστες έχουν τη δυνατότητα να φωτογραφίσουν ένα αξιοθέατο ή ένα σημείο της πόλης, να στείλουν τη φωτογραφία σε μια υπηρεσία αναγνώρισης, και να λάβουν το αντίστοιχο άρθρο της Wikipedia ως απάντηση. Για αυτήν την εφαρμογή, ο χώρος αναζήτησης είναι περιορισμένος μέσω της ενσωμάτωσης πληροφοριών για τη θέση, δηλαδή ταυτότητες κυψέλης (cell ids) ή GPS.

Και τα δύο συστήματα αξιολογούνται πειραματικά σε διαφορετικές διαστάσεις, συμπεριλαμβανομένων των διαφορετικών τηλεφωνικών προτύπων με τις διαφορετικές ιδιότητες φωτογραφικών μηχανών, χρησιμοποιώντας εναλλακτικά τα διαφορετικά είδη χωροταξικού περιορισμού της αναζήτησης (γεωγραφική θέση κ.λπ.), και με και χωρίς το στάδιο της επαλήθευσης με προβολική γεωμετρία.

Η υπόλοιπη εργασία διαρθρώνεται ως εξής: αρχίζει με μια επισκόπηση της βιβλιογραφίας και αναφορά σε παρεμφερείς, προς αυτές που παρουσιάζονται εδώ, εφαρμογές..

Το κύριο σώμα της εργασίας δομείται γύρω από τις δύο εφαρμογές που παρουσιάζονται, δηλαδή τις υπερσυνδέσεις από διαφάνειες για τις διαδραστικές αίθουσες συνεδριάσεων και τις υπερσυνδέσεις από κτίρια για ένα οδηγό πόλης. Για καθεμία από τις δύο παρουσιάζεται η μέθοδος και η υλοποίηση και ακολουθούνται από μια πειραματική αξιολόγηση του αντίστοιχου συστήματος. Τέλος, παρουσιάζονται τα συμπεράσματα και οι προοπτικές εξέλιξης.

Μεθοδολογία

Η μεθοδολογία που ακολουθήθηκε για τη διεξαγωγή της παρούσας εργασίας ήταν, αρχικά, η παρουσίαση τους κυρίως άρθρου των Quack, Bay and Gool, το οποίο αποτελεί και το βασικό αντικείμενό της. Στη συνέχεια αναζητήθηκαν άρθρα σχετικά με μεθόδους αναζήτησης και ανάκτησης ψηφιακής πληροφορίας για φυσικά αντικείμενα στο διαδίκτυο και πληροφορίες από αυτά ενσωματώθηκαν στην υπόλοιπη εργασία. Τέλος, διατυπώθηκαν κάποιοι προβληματισμοί που προέκυψαν από τη μελέτη της βιβλιογραφίας, ενώ παρατέθηκε ένα παράρτημα περιγραφής τεχνικών ζητημάτων που αναφέρονται στο κύριο άρθρο αλλά δεν αναλύονται σαφώς.

Σχετικές εργασίες

Η μέθοδος που παρουσιάζεται σ' αυτή την εργασία μπορεί να προσιδιάζει διάφορες πτυχές άλλων εργασιών. Μια πτυχή καλύπτει εργασίες σχετικές με την εφαρμογή των έξυπνων αιθουσών συνεδριάσεων, για παράδειγμα η χρήση κινητών τηλεφώνων με ενσωματωμένη φωτογραφική μηχανή ως συσκευή αλληλεπίδρασης για τις μεγάλες οθόνες. Σε αυτό, οι [Ballagas, Rohs και Sheridan](#) (2005) έχουν προτείνει ένα σύστημα που επιτρέπει στους χρήστες να επιλέξουν αντικείμενα σε μεγάλες επιδείξεις χρησιμοποιώντας το κινητό τους τηλέφωνο. Εντούτοις, η μεθοδός τους στηρίζεται στους πρόσθετους διαστάτους barcodes για να καθορίσει τη θέση της φωτογραφικής μηχανής και προορίζεται να χρησιμοποιήσει το κινητό τηλέφωνο όπως το ποντίκι του υπολογιστή προκειμένου να σύρει - και - ρίξει (drag – and – drop) τα στοιχεία στην οθόνη. Πολύ πρόσφατα, [[Boring \[et.al\]](#) (2007)] ένα σύστημα παρόμοιο με αυτό που παρουσιάζεται εδώ έχει προταθεί για την αναγνώριση εικόνων σε εκθετήρια. Ενώ οι οθόνες είναι εννοιολογικά παρόμοιες

με αυτές που χρησιμοποιούνται στις αίθουσες συνεδριάσεων, δεν είναι γνωστή άλλη εργασία που να έχει προτείνει τη χρήση κινητών τηλεφώνων με ενσωματωμένη φωτογραφική μηχανή για την προσθήκη ετικετών ή την ανάκτηση διαφανειών στις έξυπνες αίθουσες συνεδριάσεων. Οι πιο όμοιες εργασίες, από αυτή την άποψη, εξετάζουν την ανάκτηση διαφανειών από σταθερές συσκευές. Για παράδειγμα, οι [Vinciarelli και Odobez](#) (2006) έχουν προτείνει ένα σύστημα που εφαρμόζει την οπτική αναγνώριση χαρακτήρων (OCR) στις διαφάνειες που λαμβάνονται από το [beamer](#) της παρουσίασης. Η ανάκτηση και το ξεφύλλισμα γίνονται με το εξαχθέν κείμενο, δηλ. η μέθοδος δεν μπορεί να εξετάσει τις απεικονίσεις ή τις εικόνες στις διαφάνειες.

Το SlideFinder [[Niblack](#) (1999)] είναι ένα σύστημα που εξάγει δεδομένα εικόνας και κειμένου από τα αρχικά δεδομένα της διαφάνειας. Η ανάκτηση εικόνας είναι βασισμένη στα σφαιρικά ιστογράμματα χρωμάτων και έτσι περιορισμένη ως προς την αναγνώριση των γραφικών στοιχείων ή ως ένα ορισμένο βαθμό το σφαιρικό σχεδιάγραμμα της διαφάνειας. Η χρησιμοποίηση μόνο των αρχικά αποθηκευμένων αρχείων παρουσίασης αντί της χρησιμοποίησης των δεδομένων της ληφθείσας εικόνας δεν επιτρέπει το συγχρονισμό των διαφανειών με άλλα στοιχεία της συνεδρίασης, όπως ο καταγραφόμενος λόγος ή το βίντεο. Και τα δύο συστήματα προορίζονται μόνο για την ανάκτηση ερωτημάτων με λέξεις - κλειδιά από έναν προσωπικό υπολογιστή γραφείου. Ενώ το σύστημά που προτείνεται εδώ μπορεί επίσης να χρησιμοποιηθεί και για εκτός σύνδεσης ανάκτηση με ερώτημα με παραδείγματα, εστιάζει στην προσθήκη ετικετών (tagging) από κινητά τηλέφωνα. Αυτό απαιτεί τον προσδιορισμό της σωστής φωτογραφικής διαφάνειας αξιόπιστα από ποικίλες απόψεις, η οποία δεν θα ήταν δυνατή με τις προαναφερθείσες προσεγγίσεις.

Μια άλλη πτυχή σχετίζεται με την εργασία αυτή είναι οι εφαρμογές οδηγών για συσκευές κινητής τηλεφωνίας. Οι [Bay, Fasel, και Van Gool](#) (2006) έχουν προτείνει έναν οδηγό μουσείου σε [tablet PC](#). Το σύστημα εκτέλεσε με επιτυχία την αναγνώριση της τρισδιάστατης έκθεσης αντικειμένων χρησιμοποιώντας αμετάβλητης κλίμακας τοπικά χαρακτηριστικά. Εντούτοις, στο σύστημά τους, ολόκληρη η εν λόγω βάση δεδομένων εγκαθίσταται πλήρως στη συσκευή του client, πράγμα γενικά αδύνατον για μικρότερες συσκευές, όπως τα κινητά τηλέφωνα, και για μεγαλύτερες βάσεις δεδομένων. Ένα παρόμοιο σύστημα για κινητά τηλέφωνα, αλλά με κάπως απλούστερη αναγνώριση αντικει-

μένου είναι το προτεινόμενο από τους [Fockler \[κ.ά.\]](#) (2005). Η προτεινόμενη αναγνώριση στηρίζεται σε ιστογράμματα απλών χρωμάτων, η οποία αποδεικνύεται πολύ αναποτελεσματική στις αλλαγές φωτισμού στο εξωτερικό περιβάλλον. Η διάκριση των στιγμιότυπων των αντικειμένων στις εφαρμογές μας, δηλαδή των διαφανειών ή των υπαίθριων εικόνων των αξιοθέατων τουριστικού ενδιαφέροντος, είναι ακόμη λιγότερο αξιόπιστες με τα σφαιρικά ιστογράμματα χρωμάτων.

Η εργασία που προσομοιάζει περισσότερο με την εφαρμογή των οδηγών πόλης που σχεδιάστηκε από τους [Quack, Bay και Gool](#) (2008) είναι ίσως αυτή των [Paletta \[κ.ά.\]](#) (2006). Οι συντάκτες προτείνουν έναν οδηγό πόλης σε κινητό τηλέφωνο χρησιμοποιώντας τοπικά χαρακτηριστικά. Εντούτοις, εστιάζουν στη βελτίωση των δυνατοτήτων αναγνώρισης χρησιμοποιώντας τα πληροφοριακά και συμπαγή χαρακτηριστικά iSift αντί των [SIFT](#). Η παρούσα εργασία διαφέρει σημαντικά σε διάφορα σημεία: χρησιμοποιείται γεωμετρία πολλαπλών όψεων ώστε να βελτιωθεί η αναγνώριση, στηρίζεται σε χαρακτηριστικά [SURE](#), που είναι επίσης συμπαγέστερα και γρηγορότερα από τα [SIFT](#), σύμφωνα με τους [Bay, Tuytelaars και Van Gool](#) (2006) και ερευνά, επίσης, αριθμητικά τα αποτελέσματα του περιορισμού από το GPS ή των cell ids σε ποσοστά αναγνώρισης και στην ταχύτητα ταύτισης.

Δηλαδή αντί της βελτίωσης των ίδιων των χαρακτηριστικών η παρούσα εργασία προσθέτει ένα σφαιρικό γεωμετρικό φίλτρο ως τελικό στάδιο επαλήθευσης στο σύστημα αναγνώρισης. Τέλος, οι δοκιμαστικές βάσεις δεδομένων που προτείνονται περιέχουν εικόνες από πολύ περισσότερες οπτικές, με πολύ μεγαλύτερη ποικιλία από τις βάσεις που χρησιμοποιούνται από τους [Paletta \[et.al\]](#) (2006).

Έτσι, οι κύριες συνεισφορές αυτής της εργασίας είναι: ένα πλήρες σύστημα δικτύωσης αναγνώρισης αντικειμένων, συμπεριλαμβανομένου ενός δευτερεύοντος server αναγνώρισης και ενός δεύτερου λογισμικού για τον client για την αλληλεπίδραση με το περιβάλλον με ένα κλικ. Ένα πλήρες δίκτυο αναγνώρισης αντικειμένων για το «Διαδίκτυο των Πραγμάτων», το οποίο αρχίζει με αντιστοίχιση τοπικών χαρακτηριστικών, επαλήθευση με προβολική γεωμετρία, και αναζήτηση με χωροταξικούς περιορισμούς από τα πολύμορφα εμπόδια, όπως η θέση GPS· η αξιολόγηση της υλοποίησης για δύο δειγματικές εφαρμογές, δηλαδή την προσθήκη ετικετών σχολιασμού και σελιδοδεικτών σε διαφάνειες στις έξυπνες αίθουσες συνεδριάσεων, καθώς και μια εφαρμογή οδηγού πόλης για

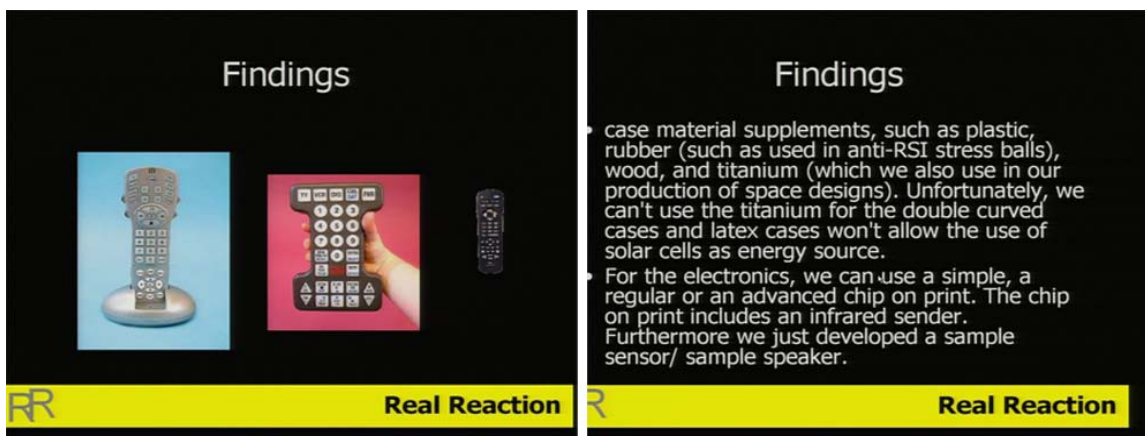
κινητό τηλέφωνο· τελευταία αλλά όχι ασήμαντη, είναι η αξιολόγηση και για τις δύο περιπτώσεις των προκλητικών συνόλων δεδομένων δοκιμής.

Υπερσυνδεδεμένες Διαφάνειες: Διαδραστικές Αίθουσες Συνεδριάσεων

Οι σύγχρονες αίθουσες συνεδριάσεων εξοπλίζονται με έναν αυξανόμενο αριθμό ηλεκτρονικών συσκευών λήψης, οι οποίες επιτρέπουν την καταγραφή των συνεδριάσεων σε διάφορες μορφές, όπως αναφέρεται στην εργασία του [Abowd](#) (1999) και των [Amir, Ashour, Srinivasan](#) (2001). Περιλαμβάνουν συχνά ηχογραφήσεις, καταγραφή βίντεο, λήψεις πινάκων και, τελευταίο αλλά εξίσου σημαντικό, λήψη πλαισίων από τον προβολέα φωτογραφικών διαφανειών. Αυτές οι εγκαταστάσεις αναπτύσσονται συνήθως για να διευκολύνουν δύο στόχους:

- να επιτρέπουν ανάκτηση εκτός σύνδεσης
- να επιτρέπουν ξεφύλλισμα στο σώμα των καταγεγραμμένων συναντήσεων και να μετατρέπουν τις αίθουσες συνεδριάσεων σε έξυπνα διαδραστικά περιβάλλοντα.

Η ανά χείρας εργασία εστιάζει στις ληφθείσες φωτογραφίες από διαφάνειες παρουσίασης που αποτελούν το κύριο μέρος της. Όπως φαίνεται στην [Εικόνα 1](#), οι φωτογραφικές διαφάνειες **περιέχουν** συνήθως γραπτώς τα κύρια σημεία που θίγει ένας ομιλητής, και συμπεριλαμβάνουν γραφήματα ή / και εικόνες, στοιχεία που διευκολύνουν την κατανόηση και την απομνημόνευση της παρουσίασης. Οι προβαλλόμενες διαφάνειες μπορούν να ειδωθούν ως ο «συνδεδετικός κρίκος» μεταξύ όλων των καταγεγραμμένων μορφών και κατά συνέπεια, αποτελούν ένα φυσικό σημείο εισόδου σε μια βάση δεδομένων των καταγεγραμμένων παρουσιάσεων.



Εικόνα 1. Χαρακτηριστικές φωτογραφικές διαφάνειες παρουσίασης από τη βάση δεδομένων AMI Corpus.
(Πηγή: [Carletta \[et.al\], 2006](#))

Ένα χαρακτηριστικό σενάριο χρήσης για το σύστημα αυτό είναι το ακόλουθο: Χρησιμοποιώντας την ενσωματωμένη φωτογραφική μηχανή του κινητού τηλεφώνου του, ένας συμμετέχων σε μια συνεδρίαση φωτογραφίζει μια διαφάνεια που τον ενδιαφέρει. Η εικόνα διαβιβάζεται σε έναν server αναγνώρισης μέσω της ασύρματης σύνδεσης του κινητού του στο Διαδίκτυο (UMTS, GPRS κ.λπ.). Στον server, εξάγονται από την εικόνα τα χαρακτηριστικά της και αντιστοιχίζονται στη βάση δεδομένων με τις ληφθείσες διαφάνειες. Η σωστή φωτογραφική διαφάνεια αναγνωρίζεται, προστίθεται στους προσωπικούς σελιδοδείκτες του χρήστη και λαμβάνει μια επιβεβαίωση στον φυλλομετρητή WAP στο κινητό του τηλέφωνό. Σημειώνεται ότι η αποστολή της εικόνας του ερωτήματος στο server μπορεί να πραγματοποιηθεί και με απλό MMS ή και με τη χρήση μια εφαρμογής στον client προγραμματισμένης σε C++ στην πλατφόρμα Symbian. Η [Εικόνα 2](#) παρουσιάζει στιγμιότυπα της οθόνης για ένα χαρακτηριστικό σενάριο χρήσης της εφαρμογής για κινητά τηλέφωνα.

Πίσω στον προσωπικό υπολογιστή του, ο χρήστης έχει ανά πάσα στιγμή πρόσβαση στο σύνολο των διαφανειών στις οποίες επέλεξε να προστεθούν σελιδοδείκτες. Από κάθε διαφάνεια στην οποία προσετέθη σελιδοδείκτης, ο χρήστης έχει τη δυνατότητα να ανοίξει τον φυλλομετρητή της συνάντησης και να αναπαράγει τις άλλες μορφές, όπως σχετικά βίντεο ή ηχογραφήσεις, ξεκινώντας από τη χρονική στιγμή της προβολής της διαφάνειας. Με τη φωτογράφιση μόνο ενός τμήματος μιας διαφάνειας, ο χρήστης έχει

επίσης τη δυνατότητα να δώσει έμφαση σε ορισμένα στοιχεία (κείμενο και αριθμούς) - με άλλα λόγια, το κινητό τηλέφωνο λειτουργεί ως ένα εργαλείο ψηφιακού δείκτη.



Εικόνα 2. Ο χρήστης προσθέτει ετικέτες στις προβαλλόμενες διαφάνειες χρησιμοποιώντας την εφαρμογή για κινητά βγάζοντας μια φωτογραφία (αριστερά) η οποία αυτόματα μεταβιβάζεται στον server όπου αναγνωρίζεται (μέση) και μια απάντηση δίνεται σε έναν φυλλομετρητή WAP που ανοίγει αυτόματα (δεξιά) (Πηγή: [Quack, Bay, Van Gool, 2008](#))

Ας σημειωθεί ότι κάποιος θα μπορούσε να υποθέσει ότι μια πολύ απλή μέθοδος επίσυναψης σελιδοδοκτών σε διαφάνειες θα μπορούσε να σχεδιαστεί, βασιζόμενη μόνο σε [timestamping](#). Σε αυτή την περίπτωση, το λογισμικό του client θα διεβίβαζε απλά τον τρέχοντα χρόνο, ο οποίος θα συγχρονιζόταν με τις διαφάνειες που έχουν ήδη [timestamps](#). Το σύστημα που προτείνεται εδώ, όμως, δεν επιτρέπει μόνο τις πιο ευέλικτες εφαρμογές (τις προαναφερθείσες «επισημάνσεις» - highlighting των στοιχείων των διαφανειών) αλλά είναι και αποτελεσματικό προς τον έγκαιρο εντοπισμό των λαθών συγχρονισμού. Στην πραγματικότητα, η χρησιμοποίηση ενός ελαστικού χρονικού περιορισμού που κυμαίνεται από μερικά λεπτά μέχρι ακόμη και αρκετές ώρες, θα καθιστούσε το σύστημά πιο εύκολα προσαρμόσιμο και θα ένωνε τα καλύτερα χαρακτηριστικά και των δύο μεθόδων.

Η βασική λειτουργία του προτεινόμενου συστήματος αναγνώρισης διαφανειών στον κεντρικό υπολογιστή είναι η ακόλουθη: για τα εισερχόμενα ερωτήματα, εξάγονται τοπικά χαρακτηριστικά που δεν μεταβάλλονται κλιμακωτά. Για κάθε χαρακτηριστικό εκτελείται μια αναζήτηση του κοντινότερου «γείτονα» στην αναφορική βάση δεδομένων των διαφανειών. Οι προκύπτουσες υποθετικές αντιστοιχίες ελέγχονται χρησιμοποιώντας όρια στην προβολική γεωμετρία. Οι επόμενες δύο υποενότητες περιγράφουν αυτά τα βήματα λεπτομερέστερα.

Λήψη διαφανειών και εξαγωγή χαρακτηριστικών

Αρχικά, θα αναλυθεί μια συλλογή διαφανειών παρουσίασης που αποθηκεύονται ως εικόνες.

Αυτό το αποτέλεσμα μπορεί να παραχθεί εύκολα χρησιμοποιώντας ένα μηχανισμό λήψης φωτογραφιών της οθόνης που συνδέεται με το [beamer](#) της παρουσίασης. Από τα αρχεία εικόνων, εξάγονται τα σταθερής κλίμακας χαρακτηριστικά γύρω από τα σημεία ενδιαφέροντος που έχουν εντοπιστεί. Τα τελευταία χρόνια, σημαντική πρόοδος έχει σημειωθεί στο πεδίο αυτό και έχει οδηγήσει σε ένα ποικιλόμορφο σύνολο μεθόδων εξαγωγής και περιγραφής χαρακτηριστικών, όπως παρουσιάζεται στο έργο των [Lowe](#) (2003), [Bay, Tuytelaars and Van Gool](#) (2006) και [Mikolajczyk, Schmid](#) (2005), που έχουν εφαρμοστεί επιτυχώς και σε τομείς όπως η ανάκτηση βίντεο από τους [Sivic και Zisserman](#) (2005)], αναγνώριση τάξεων αντικειμένων από τους [Leibe, Seemann, Shiele](#) (2005) κ.λπ. Προκύπτει ότι τέτοια τοπικά χαρακτηριστικά δεν μπορούν να χρησιμοποιηθούν μόνο για την περιγραφή και ταύτιση αντικειμένων και τοπίων, αλλά και για την εξίσου αξιόπιστη λειτουργία τους με κείμενο όπως οι πινακίδες αριθμού κυκλοφορίας [[Donoser, Bischof](#) (2006)]. Κατά συνέπεια, αυτή η τάξη χαρακτηριστικών είναι μια καλή επιλογή για περιγραφή του περιεχομένου των διαφανειών που περιέχει τόσο κείμενο όσο και οπτικά δεδομένα, όπως οι εικόνες και τα γραφήματα. Επιπλέον, σε αντιδιαστολή προς τα σφαιρικά χαρακτηριστικά που προτείνονται από τους [Niblack](#) (1999) και [Fockler, Zeidler, Bimber](#) (2005), επιτρέπουν, επίσης, στο χρήστη να φωτογραφίσει συγκεκριμένα τμήματα ή στοιχεία μιας διαφάνειας ως ερωτήματα στο συγκεκριμένο σύστημα. Σ' αυτή την υλοποίηση χρησιμοποιείται ο ευρέως διαδεδομένος [SURF](#) [[Bay, Tuytelaars and Van Gool](#) (2006)], που συνδυάζει ανίχνευση και περιγραφή. Το αποτέλεσμα του ανιχνευτή [SURF](#) αποτελεί-

ται από ένα πίνακα 64 θέσεων για καθένα από τα εντοπισθέντα σημεία ενδιαφέροντος σε μια εικόνα.

Σύστημα Αναγνώρισης Διαφανειών

Η προσέγγιση της αναγνώρισης διαφανειών αποτελείται από δύο στάδια:

- Ταύτιση χαρακτηριστικών,
- Σφαιρική γεωμετρική επαλήθευση.

Για την ταύτιση των χαρακτηριστικών συγκρίνεται το διάνυσμα του χαρακτηριστικού από την εικόνα στο ερώτημα προς αυτό της εικόνας στη βάση δεδομένων. Ακριβέστερα, για κάθε ερώτημα πίνακα 64 θέσεων, υπολογίζεται η Ευκλείδεια απόσταση προς τη βάση των διανυσμάτων. Μια ταύτιση δηλώνεται εάν η απόσταση προς τον κοντινότερο γείτονα είναι μικρότερη από $0,7^{(2)}$ φορές από την απόσταση προς το δεύτερο κοντινότερο γείτονα. Αυτή η στρατηγική αντιστοίχισης εφαρμόστηκε με επιτυχία από τους [Lowe](#) (2003), [Bay, Tuytelaars and Van Gool](#) (2006), [Bay, Fasel, Van Gool](#) (2006), [Miko-lajczyk, Schmid](#) (2005).

Η εύρεση του καλύτερου αποτελέσματος θα μπορούσε να επιτευχθεί απλά με την επιλογή ενός ζεύγους εικόνων ερωτήματος - βάσης δεδομένων, το οποίο θα συγκεντρώνει το μεγαλύτερο αριθμό αντιστοιχιών. Εντούτοις, χωρίς την επαλήθευση της γεωμετρικής διάταξης των αντιστοιχισμένων σημείων ενδιαφέροντος, υπάρχει περίπτωση ως αποτέλεσμα να επιστραφεί ένα λανθασμένο ζεύγος εικόνων ερωτήματος - βάσης δεδομένων. Αυτό ισχύει ιδιαίτερα στην περίπτωση του συστήματος που παρουσιάζεται εδώ, καθώς εφαρμόζεται σε μεγάλο αριθμό ταυτίσεων η μέθοδος του [stemming](#) σε τμήματα του κειμένου των διαφανειών. Αυτές οι αντιστοιχίες είναι καθ' όλα «σωστές» ως προς το επίπεδο της αντιστοίχισης χαρακτηριστικών, αλλά μόνο η συνεπής διάταξή τους με όλα τα γράμματα και τις λέξεις είναι σωστή σε σημασιολογικό επίπεδο.

Για τη λύση αυτού του προβλήματος, προσφεύγουμε στην προβολική γεωμετρία. Από τη στιγμή που τα αντικείμενα (οι διαφάνειες, εν προκειμένω) στη βάση δεδομένων είναι επίπεδα, μπορούμε να στηριχθούμε σε μια 2D ομόγραφη αντιστοίχιση [[Hartley, Zisserman](#) (2004)] από την εικόνα του ερωτήματος σε μια επιλεγμένη υποψήφια από τη

² Σύμφωνα με το λήμμα [SURF](#) της [Wikipedia](#) απορρίπτονται οι ταυτίσεις όπου η αναλογία της απόστασης του κοντινότερου είναι μεγαλύτερη από 0,8 φορές από αυτή του δεύτερου κοντινότερου γείτονα.

βάση δεδομένων προκειμένου να επικυρωθεί η επιλεγμένη ταύτιση. Αυτό σημαίνει ότι το σύνολο των σημείων αντιστοίχισης μεταξύ των ταυτισμένων σημείων ενδιαφέροντος από την εικόνα ερώτησης x_i^q του ερωτήματος και η εικόνα x_i^d της βάσης δεδομένων πρέπει να ικανοποιούν την ακόλουθη συνθήκη:

$$Hx_i^q = x_i^d \quad i \in 1 \dots 4$$

όπου [H ο 3x3 ομογραφικός πίνακας](#) του οποίου οι 8 βαθμοί ελευθερίας μπορούν να λυθούν με 4 σημεία αντιστοιχίας με το i να ανήκει από το 1 έως το 4.

Για ισχύ απέναντι στις προαναφερθείσες μετρήσεις που υπολογίζουμε το H εκτιμάται με τη χρήση του [RANSAC \[Fischler, Bolles \(1981\)\]](#). Η ποιότητα αρκετών από τα εκτιμώμενα μοντέλα υπολογίζεται βάσει του αριθμού των inliers, όπου ένα inlier ορίζεται από ένα κατώτατο όριο στο λάθος υπόλοιπο.

Το λάθος υπόλοιπο για το μοντέλο ορίζεται από την απόσταση των αληθινών σημείων από τα παραχθέντα σημεία όπως εκτιμήθηκαν από το H . Το αποτέλεσμα μιας τέτοιας γεωμετρικής επαλήθευσης με ομογραφία παρουσιάζεται στο [Εικόνα 3](#).

Πειραματική διαδικασία

Για τα πειράματά χρησιμοποιήθηκαν στοιχεία από την βάση δεδομένων AMI meeting room corpus [[Carletta \[et.al\] \(2006\)](#)]. Αυτό το σύνολο περιέχει τις εικόνες των διαφανειών που έχουν συγκεντρωθεί κατά τη διάρκεια μιας εκτεταμένης χρονικής περιόδου χρησιμοποιώντας μια κάρτα λήψης οθόνης σε ένα PC που συνδέεται με το [beamer](#) στην αίθουσα παρουσιάσεων. Οι διαφάνειες λαμβάνονται σε κανονικά χρονικά διαστήματα και αποθηκεύονται ως αρχεία JPEG. Για να είναι σε θέση να συγχρονίσει με τις άλλες μορφές (π.χ. ηχογραφήσεις και βίντεο), κάθε ληφθείσα διαφάνεια είναι [timestamped](#).

Για να δημιουργηθούν ακριβή, τεκμηριωμένα δεδομένα προβάλλονται διαφάνειες που αποκτήθηκαν από το AMI corpus σε αίθουσες συνεδριάσεων με ρυθμίσεις που όρισαν οι ερευνητές και πήραν φωτογραφίες με την ενσωματωμένη φωτογραφική μηχανή δύο διαφορετικών μοντέλων κινητών. Για την ακρίβεια, χρησιμοποιήθηκαν ένα Nokia N70, που είναι μοντέλο με φωτογραφική μηχανή υψηλών προδιαγραφών 2 megapixel, και τη Nokia 6230, που είναι παλαιότερο μοντέλο με χαμηλής ποιότητας φωτογραφική

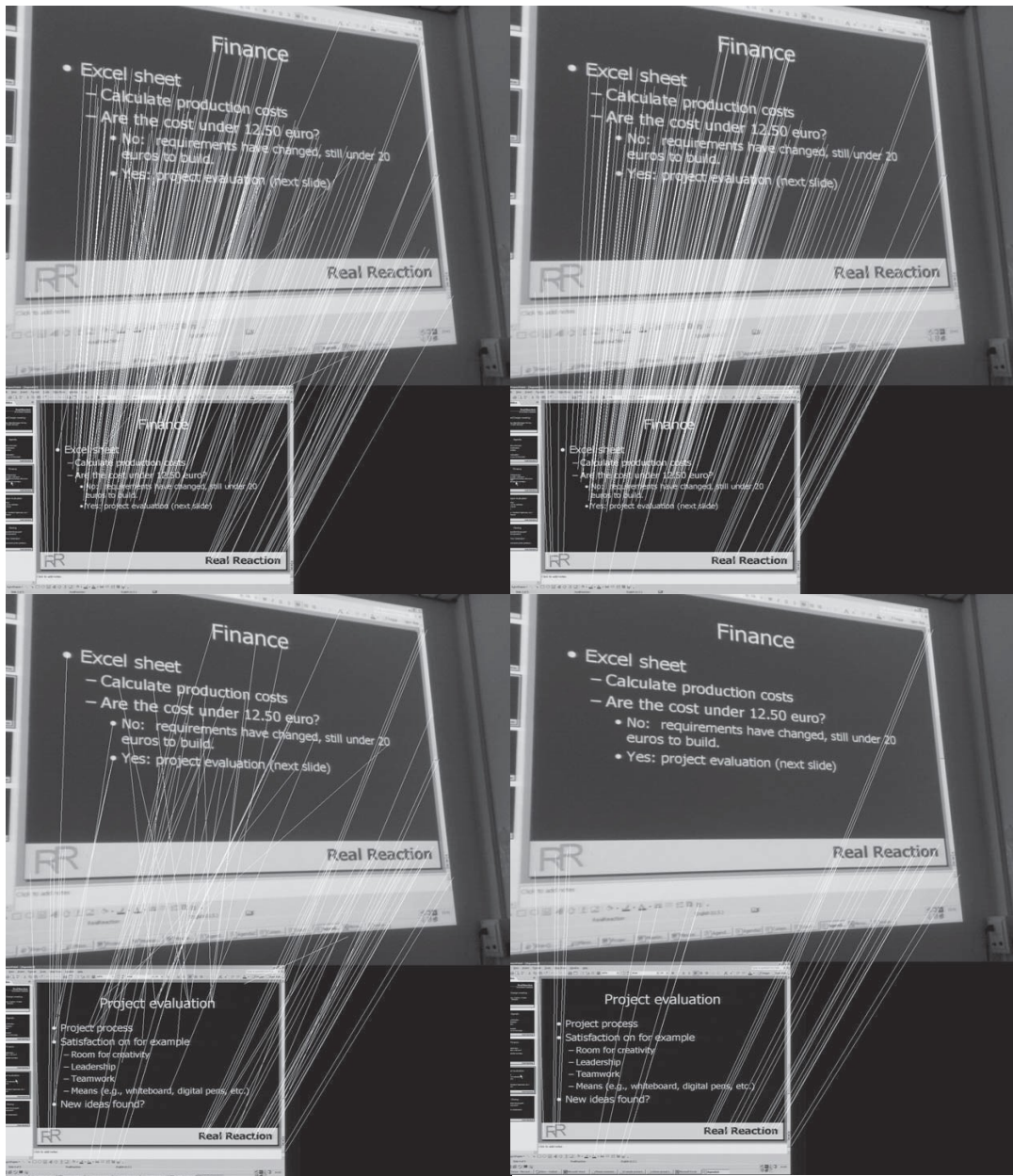
μηχανή VGA. Λήφθηκαν 61 εικόνες με το Nokia N70 και 44 εικόνες με το Nokia 6230³. Η [Εικόνα 4](#) παρουσιάζει μερικά παραδείγματα εικόνων ερωτημάτων. Η αναφορική βάση δεδομένων αποτελείται από το υποσύνολο AMI corpus για τις συνεδριάσεις του σεναρίου IDIAP, το οποίο περιέχει 1098 ληφθείσες εικόνες διαφανειών.

Εξήχθησαν τα χαρακτηριστικά [SURF](#) από τις διαφάνειες αναφοράς στη βάση δεδομένων σε 2 επίπεδα ανάλυσης, 800x600pixels και 640x480pixels. Για τις 1098 φωτογραφικές διαφάνειες αυτό οδήγησε σε $1,02 \cdot 10^6$ και $0,72 \cdot 10^6$ χαρακτηριστικά αντίστοιχα. Για την εξαγωγή χαρακτηριστικών [SURF](#) χρησιμοποιήθηκαν οι τυπικές ρυθμίσεις του ανιχνευτή που μεταφορτώθηκε από την ιστοσελίδα του δημιουργού του.

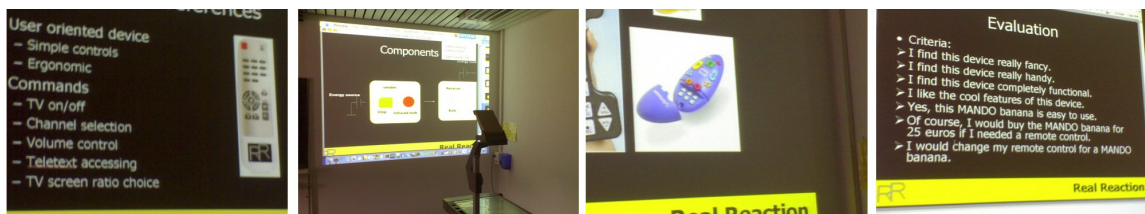
Οι αναλύσεις των εικόνων ερωτημάτων παρέμειναν χωρίς μεταβολή, όπως λήφθηκαν από τη φωτογραφική μηχανή των κινητών τηλεφώνων. Έγιναν πειράματα με και χωρίς ομογραφικό έλεγχο, και οι εικόνες των ερωτημάτων αντιστοιχίστηκαν με τις εικόνες της βάσης δεδομένων και στις δύο αναλύσεις.

Η ομογραφία υπολογίστηκε μόνο στα αποτελέσματα όπου υπήρχε ταύτιση τουλάχιστον 10 χαρακτηριστικών μεταξύ δύο διαφανειών. Εάν υπήρχαν λιγότερες αντιστοιχίες ή εάν κανένα συνεπές μοντέλο ομογραφίας δεν μπορούσε να βρεθεί με το [RANSAC](#), το ζεύγος κατατασσόταν ως «unmatched», δηλαδή, χωρίς αντιστοιχία. Εάν υπήρχαν περισσότερες από μία διαφάνειες που κάλυπταν τη συνθήκη του ελάχιστου των 10 ταυτόσιμων χαρακτηριστικών, τότε, ως αποτέλεσμα επιστρεφόταν αυτή που αξιολογήθηκε ως ακριβέστερη (που εμφάνιζε, δηλαδή, τα περισσότερα σημεία αντιστοίχισης). Δεδομένου ότι το corpus περιέχει μερικές διπλές διαφάνειες, μια πραγματική αντιστοιχία δηλωνόταν εάν αναγνωριζόταν τουλάχιστον ένα από τα αντίγραφα.

³ Οι εικόνες των ερωτημάτων με την τεκμηρίωση είναι διαθέσιμες για μεταφόρτωση από το <http://www.vision.ee.ethz.ch/datasets/> (Ανάκτηση 5/6/2008)



Εικόνα 3. Γεωμετρική επαλήθευση με μια ομογραφία. **Επάνω σειρά:** αντιστοιχίσεις για μια εικόνα ερωτήματος με τη σωστή εικόνα της βάσης δεδομένων. **Επάνω αριστερά:** πριν από το φίλτρο ομογραφίας. **Επάνω δεξιά:** μετά από το φίλτρο ομογραφίας. Δεδομένου ότι η αντιστοιχία μεταξύ των φωτογραφικών διαφανειών είναι σωστή οι περισσότερες από τις υποθετικές αντιστοιχίες χαρακτηριστικών «επιζούν» του φίλτρου της ομογραφίας. **Κάτω σειρά:** ίδια εικόνα αντιστοιχίζεται με μια λάθος εικόνα στη βάση δεδομένων. **Κάτω αριστερά:** πολλές λανθασμένες υποθετικές αντιστοιχίες θα προέκυπταν χωρίς γεωμετρική επαλήθευση, σε ακραίες περιπτώσεις η καταμέτρησή τους μπορεί να είναι παρόμοια με ή υψηλότερη απ' ό,τι για το σωστό ζεύγος εικόνων. **Κάτω δεξιά:** όλες οι λανθασμένες αντιστοιχίες αφαιρούνται, μόνο τα χαρακτηριστικά γνωρίσματα από το (ορθό) αντιστοιχισμένο πλαίσιο «επιζεί» και η διάκριση του σωστού ζεύγους αυξάνεται αισθητά. (Πηγή: [Quack, Bay, Van Gool, 2008](#))



Εικόνα 4. Παραδείγματα εικόνων ερωτημάτων, από αριστερά προς τα δεξιά: **α.** Συνθέσεις κειμένου και εικόνας, **β.** λήψη από τις ποικίλες οπτικές γωνίες, **γ.** διαφορετικά επίπεδα ζουμ ή μικτό περιεχόμενο, **δ.** παράδειγμα όπου επιλέγεται μια συγκεκριμένη περιοχή μιας διαφάνειας ή περιέχει μεγάλη ποσότητα κειμένου. (Πηγή: [Quack, Bay, Van Gool, 2008](#))

Ο [Πίνακας 1](#) παρουσιάζει τα ποσοστά αναγνώρισης, για τα διαφορετικά μοντέλα τηλεφώνου, διαφορετικές αναλύσεις και ομογραφία με και χωρίς φίλτρο. Στην ανάλυση 800x600, το φίλτρο ομογραφίας δίνει μια βελτίωση περίπου 2% ή 4% και για κάθε τύπο τηλεφώνου, αντίστοιχα. Το ποσοστό αναγνώρισης με ένα σύγχρονο τηλέφωνο, όπως το Nokia N70, που χρησιμοποιήθηκε εδώ, φθάνει το 100%, ενώ η φωτογραφική μηχανή χαμηλότερης ποιότητας στο παλαιότερο μοντέλο Nokia 6230 έχει ως αποτέλεσμα χαμηλότερα ποσοστά αναγνώρισης. Τα αποτελέσματα για την ανάλυση 640x480 επιβεβαιώνουν τα αποτελέσματα της περίπτωσης 800x600, αλλά γενικά επιτυγχάνουν χαμηλότερα αποτελέσματα αναγνώρισης. Αυτό οφείλεται στο γεγονός, ότι στη χαμηλότερη ανάλυση εξάγονται λιγότερα χαρακτηριστικά.

	Ακρίβεια με Γεωμετρικό Φίλτρο		Ακρίβεια χωρίς Γεωμετρικό Φίλτρο	
	800x600	640x480	800x600	640x480
Nokia N70	100%	98,30%	98,30%	96,70%
Nokia 6230	97,70%	93,20%	91%	86,30%

Πίνακας 1. Σύνοψη δεικτών επιτυχίας αναγνώρισης των διαφανειών της βάσης δεδομένων (Πηγή: [Quack, Bay, Van Gool, 2008](#))

Υπερσυνδεδεμένα Κτίρια: Οδηγός πόλης για κινητό τηλέφωνο

Η δεύτερη περίπτωση αξιοποίησης του συστήματος που παρουσιάζεται σε αυτή την εργασία ασχολείται μια διαφορετική κατηγορία "πραγμάτων". Υπερσυνδέει κτίρια (τουριστικού ενδιαφέροντος αξιοθέατα κ.τ.λ.) με ψηφιακό περιεχόμενο. Οι χρήστες μπορούν να αναζητήσουν πληροφορίες χρησιμοποιώντας μια εφαρμογή στο κινητό τους τηλέφωνο.

νο. Η διαδραστική διαδικασία, το λογισμικό και η διεπαφή για τους χρήστες είναι αρκετά όμοια με αυτή της εφαρμογής στις αίθουσες συνεδριάσεων. Όμως, αυτή τη φορά, ο αριθμός των αντικειμένων είναι σχεδόν απεριόριστος αν η εφαρμογή πρόκειται να αξιοποιηθεί σε παγκόσμια βάση. Για να ξεπεραστούν τα προβλήματα που προκύπτουν από τη δυνατότητα προοδευτικής αύξησης των επιδόσεων περιορίζεται ο χώρος αναζήτησης γεωγραφικά. Αυτό σημαίνει ότι περιορίζεται η οπτική αναζήτηση στη βάση δεδομένων που βρίσκεται στο γεωγραφικό πεδίο γύρω από τη θέση του χρήστη.

Ακολουθώς περιγράφεται αυτή η προσέγγιση με μεγαλύτερη λεπτομέρεια καθώς και η αξιολόγηση της εκτέλεσής της.

Οπτικά Στοιχεία και Γεωγραφική Θέση

Από την οπτική του χρήστη, η διαδραστική διαδικασία παραμένει η ίδια με αυτή στις αίθουσες συνεδριάσεων: Με το "κλικ" ενός κουμπιού στο κινητό τηλέφωνο γίνεται λήψη μιας φωτογραφίας και μεταδίδεται στον server. Σε αντίθεση όμως με την εφαρμογή των αιθουσών συνεδριάσεων η εφαρμογή του client για τον οδηγό προσθέτει στο αίτημα πληροφορίες για τη θέση του χρήστη και κατ' επέκταση του αντικειμένου το οποίο υποχρεωτικά θα βρίσκεται σε ένα ορισμένο πεδίο γύρω από τον χρήστη. Αυτές οι πληροφορίες για την τρέχουσα θέση ανάγνωσης αντλούνται από μία ενσωματωμένη ή εξωτερική συσκευή (bluetooth) GPS και από τον τρέχοντα πύργο κυψέλης (cell-tower), τον επονομαζόμενο CGI (Cell Global Identity).

Αυτός ο συνδυασμός δεδομένων εικόνας και θέσης διαμορφώνει ένα τέλειο ερώτημα για αναζήτηση πληροφοριών για στατικά, φυσικά αντικείμενα. Όπως αναφέρθηκε νωρίτερα, πληροφορίες αποκλειστικά για τη θέση εν γένει δε θα ήταν επαρκείς ώστε να δώσουν πρόσβαση σε σχετικές πληροφορίες: το αντικείμενο του ενδιαφέροντος θα μπορούσε να βρίσκεται αρκετές εκατοντάδες μέτρα μακριά (π.χ. το καμπαναριό μιας εκκλησίας ή θα μπορούσαν να υπάρχουν πολλά αντικείμενα ενδιαφέροντος στην ίδια περιοχή (π.χ. η πλατεία του Αγίου Μάρκου στη Βενετία περιβαλλόμενη από πολλά αντικείμενα ενδιαφέροντος). Επιπλέον σε αστικές περιοχές με ψηλά κτίρια και στενούς δρόμους τα δεδομένα του GPS είναι ανακριβή. Από την άλλη το να βασίζεται η αναζήτηση μόνο στην εικόνα δε θα ήταν ούτε αυτό εφικτό καθώς το μέγεθος της βάσης δεδομένων θα

έκανε τα ερωτήματα πραγματικού χρόνου και την ακρίβεια των αποτελεσμάτων πολύ δύσκολα επιτεύξιμα.

Μετά την επεξεργασία του ερωτήματος ο χρήστης λαμβάνει την αιτηθείσα πληροφορία απευθείας στην οθόνη του κινητού του. Στη δοκιμαστική εφαρμογή ανοίγει ένας φυλλομετρητής με τη σελίδα της Wikipedia που αντιστοιχεί στο αντικείμενο όπως φαίνεται στην [Εικόνα 5](#).

Σχεδιασμός Συστήματος

Το σύστημα του οδηγού πόλης αποτελείται από ένα λογισμικό για τον server και ένα λογισμικό για τον client.

Τα στοιχεία από πλευράς server αποτελούνται από μια σχεσιακή βάση δεδομένων για αποθήκευση των μεταδεδομένων της εικόνας (θέσεις GPS, πληροφορίες κυψέλης κλπ.) και πληροφορίες που αφορούν τα αποθηκευμένα αξιοθέατα. Για το σκοπό αυτό χρησιμοποιήθηκε το MySQL. Η αναγνώριση εικόνων υλοποιείται ως ένας server σε C++ που μπορεί να είναι προσβάσιμος μέσω HTTP.



Εικόνα 5. Λογισμικό πελατών για την εφαρμογή οδηγού πόλης: ο χρήστης σπάζει απότομα μια εικόνα, περιμένει μερικά δευτερόλεπτα, και επαναπροσανατολίζεται στην αντίστοιχη σελίδα Wikipedia. (Πηγή: [Quack, Bay, Van Gool, 2008](#))

Ερωτήματα από το λογισμικό του client μεταδίδονται στον server ως αιτήματα [HTTP POST](#). Ένα ενδιάμεσο πρόγραμμα (middleware) γραμμένο σε PHP και Ruby περιορίζει την αναζήτηση τοπικά αν είναι απαραίτητο και μεταβιβάζει αυτό το προ-επεξεργασμένο ερώτημα στον server αναγνώρισης. Το σχετικό περιεχόμενο της καλύτερης ταύτισης επιστρέφεται στο λογισμικό του client και προβάλλεται αυτόματα σε ένα φυλλομετρητή όπως φαίνεται στην [Εικόνα 5](#).

Τον λογισμικό του client στο κινητό τηλέφωνο έχει υλοποιηθεί και Symbian C++ και σε Java⁴. Σημειώνεται ότι η εξαγωγή χαρακτηριστικών του ερωτήματος γίνεται στον server δηλαδή το πλήρες ερώτημα εικόνας μεταδίδεται στον server. Επίσης είναι πιθανή η εξαγωγή χαρακτηριστικών [SURF](#) στο κινητό τηλέφωνο και μετάδοσής του στον server ως ερώτημα. Μια υλοποίηση της μεθόδου δείχνει ότι η εξαγωγή χαρακτηριστικών [SURF](#) στο κινητό τηλέφωνο, επί του παρόντος είναι πολύ αργή. Η μη βελτιστοποιημένη έκδοση στη Symbian C++ στο Nokia 6630 χρειάστηκε περίπου 10 sec για τον υπολογισμό των χαρακτηριστικών του ερωτήματος. Αντίθετα σε μια σύγχρονη εξαγωγή χαρακτηριστικών PC [SURF](#) χρειάζεται μερικές εκατοντάδες ms [[Bay, Tuytelaars, Van Gool](#) (2006)].

Από τη στιγμή που τα χαρακτηριστικά [SURF](#) δεν είναι πολύ πιο συμπαγή από την αρχική εικόνα (αρκετές εκατοντάδες χαρακτηριστικών πινάκων 64 θέσεων ανά εικόνα, με κύριο πλεονέκτημα της εξαγωγής χαρακτηριστικών στο τηλέφωνο να είναι η αυξημένη ιδιωτικότητα (μόνο χαρακτηριστικά μεταδίδονται αντί της εικόνας και η δυνατότητα να δώσει στο χρήστη στιγμιαία ανατροφοδότηση αν μια εικόνα ερωτήματος περιέχει πολύ λίγα χαρακτηριστικά, για παράδειγμα λόγω θολώματος, έλλειψη υφής ή χαμηλής αντίθεσης εξαιτίας φωτός πίσω από την εικόνα).

Εναλλακτικά το σύστημα μπορεί να είναι προσβάσιμο με χρήση της Υπηρεσίας Μηνυμάτων Πολυμέσων (MMS). Μια εικόνα μεταδίδεται στον server στέλνοντας την ως μήνυμα MMS σε μια διεύθυνση ηλεκτρονικού ταχυδρομείου. Η απάντηση (π.χ. το URL του αντίστοιχου λήμματος της Wikipedia) επιστρέφεται ως ένα μήνυμα SMS.

⁴ Δυστυχώς, μόνο η έκδοση της Symbian επιτρέπει πρόσβαση στις cell-tower ids

Μέθοδος Αναγνώρισης Αντικειμένων

Τα δεδομένα από την εφαρμογή του client μεταδίδονται στην αναγνώριση του server όπου ξεκινά μια οπτική αναζήτηση που περιορίζεται από τα δεδομένα της μεταδιδόμενης θέσης. Αν χρησιμοποιούνται δεδομένα του GPS, γίνεται αναζήτηση όλων των αντικειμένων της βάσης δεδομένων σε μια προεπιλεγμένη ακτίνα (στο πειραματικό τμήμα αυτής της εργασίας αξιολογούνται διαφορετικές ακτίνες). Αν χρησιμοποιούνται μόνο πληροφορίες του cell-tower ids, η αναζήτηση περιορίζεται στα αντικείμενα που επισημαίνονται με την ίδια σειρά δεδομένων CGI (Cell Global Identify).

Ως προσέγγιση, η αναγνώριση αντικειμένων είναι αρκετά όμοια με τη μέθοδο που εξετάστηκε για τις διαφάνειες των αιθουσών συνεδριάσεων. Δηλαδή οι φερόμενες ως ταυτίσεις μεταξύ ζευγών ερωτημάτων και εικόνων των βάσεων δεδομένων βρίσκονται με αναζήτηση του κοντινότερου γείτονα για τους περιγραφείς [SURF](#) τους [Bay, Tuytelaars, Van Gool](#) (2006). Αυτές οι αντιστοιχίσεις επικυρώνονται με ένα γεωμετρικό φίλτρο. Όμως, αφού πρόκειται για τρισδιάστατα αντικείμενα στην εφαρμογή του οδηγού πόλης, το ακριβές μοντέλο τώρα είναι ο [Θεμελιώδης 3x3 πίνακας F](#), αντί του [ομογραφικού πίνακα H](#) [[Hartley, Zisserman](#) (2004)]. Ο [F](#) αντιστοιχίζει τα σημεία σε μια απεικόνιση προς τις [epipolar lines](#) μιας διαφορετικής οπτικής. Σφάλματα στο υπόλοιπο για τα μοντέλα, ορίζονται γι' αυτό από την απόσταση των πραγματικών σημείων από τις [epipolar lines](#) που παράγονται από το εκτιμώμενο [F](#) [[Hartley, Zisserman](#) (2004)].

Από την πλευρά των πρακτικών για αντικείμενα, όπως κτίρια, που αποτελούνται από πολλές πλευρές (όψεις) μπορεί, εν τούτοις, κανείς να προσεγγίσει τα αποτελέσματα χρησιμοποιώντας ομογραφία, που απαιτεί λιγότερα σημεία αντιστοίχισης. Η εκτίμηση του μοντέλου από τα φερόμενα ως σημεία αντιστοίχισης μπορεί να γίνει με το RANSAC [[Fishler, Bolles](#) (1981)] και στις δύο περιπτώσεις.

Σημειώνεται ότι το μοντέλο είναι ιδιαίτερα σημαντικό να φιλτράρει τις λανθασμένα ορισθείσες ως σωστές αναγνώρισεις. Ειδικά για δομές κτιρίων, υπάρχουν πολλά επαναλαμβανόμενα μοτίβα που ταυτίζουν διαφορετικά μεταξύ τους κτίρια. Μόνο η σωστή διάταξή τους στο χώρο ή στην εικόνα, επιτρέπει αντίστοιχα μια ισχυρή απόφαση για το αν ένα αντικείμενο εντοπίστηκε στην πραγματικότητα. Το να τίθεται απλά ως αφετηρία ένας αριθμός αντιστοιχίσεων είναι εξαιρετικά επικίνδυνο καθώς η διάκριση μιας λανθα-

σμένης αναγνώρισης ως θετικής (π.χ. ένα ερώτημα εικόνας για ένα κτίριο το οποίο δεν βρίσκεται καν στη βάση δεδομένων) από ένα ερώτημα με κάποιες ταυτίσεις που προκύπτουν λόγω δυσχερών ή / και προκλητικών συνθηκών (π.χ. εικόνα που λαμβάνεται από απόσταση) είναι ανέφικτη.

Πειράματα

Για την αξιολόγηση της προτεινόμενης μεθόδου συνελέγησαν 147 φωτογραφίες που κάλυπταν 9 τουριστικά αξιοθέατα και τις τοποθεσίες τους. Οι 147 φωτογραφίες καλύπτουν τα 9 αντικείμενα από πολλές πλευρές, τουλάχιστον 3 ανά αντικείμενο. Οι εικόνες της βάσης δεδομένων λήφθηκαν με απλή φωτογραφική μηχανή (point-and shoot). Για να οριστεί η θέση GPS τους και οι CGIs (cell-tower ids) αναπτύχθηκε μια εφαρμογή εντοπισμού σε Symbian C++ η οποία τρέχει σε κινητά τηλέφωνα και αποθηκεύει τα τρέχοντα δεδομένα GPS (όπως αποκτήθηκαν από μία εξωτερική συσκευή bluetooth GPS) και πληροφορίες CGI κυψέλης σε τακτά χρονικά διαστήματα. Αυτό το αρχείο καταγραφής συγχρονίζεται με τα [timestamps](#) των φωτογραφιών της βάσης δεδομένων.

Συγκεντρώθηκαν αλλά 126 τεστ (ερωτήματα) εικόνων που λήφθηκαν από διαφορετικά μοντέλα κινητού τηλεφώνου (Nokia N70 και Nokia 6280, και τα δύο με camera 2 megapixel) σε διαφορετικές ημέρες και ώρες της ημέρας, από διαφορετικούς χρήστες και τυχαίες οπτικές γωνίες. Από τα 126 ερωτήματα εικόνων 91 περιέχουν αντικείμενα της βάσεις δεδομένων και 35 περιέχουν εικόνες άλλων κτιρίων ή του περιβάλλοντος χώρου (επίσης σχολιασμένα με GPS και cellid). Είναι σημαντικό το σύστημα να δοκιμαστεί με αρνητικά ερωτήματα, πείραμα το οποίο αμέλησαν να κάνουν πολλοί από τους άλλους ερευνητές. Σε σύγκριση με τη βάση δεδομένων MPG-20⁵ υπάρχουν λιγότερα αντικείμενα αλλά από περισσότερες πλευρές (συνολικά, περίπου, 30 μοναδικές αναπαραστάσεις), περισσότερο προκλητικές οπτικές γωνίες για κάθε πλευρά (σε απόσταση μέχρι 500 μέτρα), πλήρη επισήμανση με δεδομένα GPS και cell-tower ids και με περισσότερο από 4 φορές περισσότερες εικόνες ερωτημάτων. Η βάση δεδομένων με όλες τις επισημάνσεις (GPS, cellids, αντικείμενα από τις σελίδες της Wikipedia κ.τ.λ.) διατίθεται για μεταμόρφωση στο διαδίκτυο⁶. Και οι εικόνες της βάσης δεδομένων και αυτές των ερωτημάτων ανα-κλι-

⁵<http://dib.joanneum.at/cape/MPG-20/>

⁶<http://www.vision.ee.ethz.ch/datasets/>

μακώνονται σε 500x375 pixels. Εικόνες δείγματα από τη βάση δεδομένων φαίνονται στην [Εικόνα 6](#) και εξετάζονται λίγο παρακάτω.



Εικόνα 6. Εικόνες αποτελεσμάτων για την εφαρμογή του οδηγού πόλης, λεπτομέρειες για την ανάλυση εντός του κειμένου. (Πηγή: [Quack, Bay, Van Gool, 2008](#))

Σημειώνεται ότι το CGI (Cell Global Identity) εξαρτάται από το διαχειριστή του δικτύου, από τη στιγμή που κάθε διαχειριστής ορίζει το δικό του σύνολο cell ids. Αν ο διαχειριστής δεν αποκαλύψει τις τοποθεσίες των κυψελών, πράγμα που αποτελεί κοινή πρακτική σε πολλές χώρες για λόγους ιδιωτικότητας, πρέπει να βρεθεί μια αντιστοίχιση μεταξύ των cell ids διαφορετικών διαχειριστών. Μια τέτοια πειραματική αντιστοίχιση επετεύχθη με τη χρήση μιας εφαρμογής εντοπισμού: τα ίχνη αποκτήθηκαν μέσω των καρτών

SIM διαφορετικών διαχειριστών δικτύων κινητής τηλεφωνίας οι οποίες συγχρονίστηκαν με τις θέσεις GPS τους: αν τα σημεία GPS απέιχαν λιγότερο από 50 μέτρα δημιουργούνταν η αντιστοιχία μεταξύ των σχετικών cell-ids. Αυτή η αντιστοίχιση απέχει πολύ από την ολοκλήρωση της, προσομοιώνει, όμως, μία προσέγγιση η οποία ακολουθείται αυτή τη στιγμή από αρκετές αρχές του Διαδικτύου.

Παρουσιάζονται πειράματα για 3 εκδοχές: γραμμική αναζήτηση στο σύνολο της βάσης δεδομένων χωρίς τοπικό προσδιορισμό, περιορισμό κατά GPS με διαφορετική αναζήτηση ακτίνας και περιορισμό κατά cell-id. Σε κάθε περίπτωση συγκρίνονται οι διαφοροποιήσεις μεταξύ του χρόνου αναζήτησης και του ρυθμού αναγνώρισης. Ένα ζεύγος εικόνων θεωρείται ότι έχει ταυτιστεί εάν τουλάχιστον 20 χαρακτηριστικά ταυτίζονται. Από τις εικόνες που ικανοποιούν αυτό το κριτήριο αυτή με τις περισσότερες αντιστοιχίες θα επιστραφεί ως αποτέλεσμα. Ο [Πίνακας 2](#) συνοψίζει τα αποτελέσματα των επιδόσεων.

	Ακρίβεια με Γεωμετρικό Φίλτρο		Ακρίβεια χωρίς Γεωμετρικό Φίλτρο	
	Ρυθμ. Αναγνώρισης	Μέσος χρόνος ταύτισης	Ρυθμ. Αναγνώρισης	Μέσος χρόνος ταύτισης
	800x600	640x480	800x600	640x480
Nokia N70	100%	98,30%	98,30%	96,70%
Nokia 6230	97,70%	93,20%	91%	86,30%

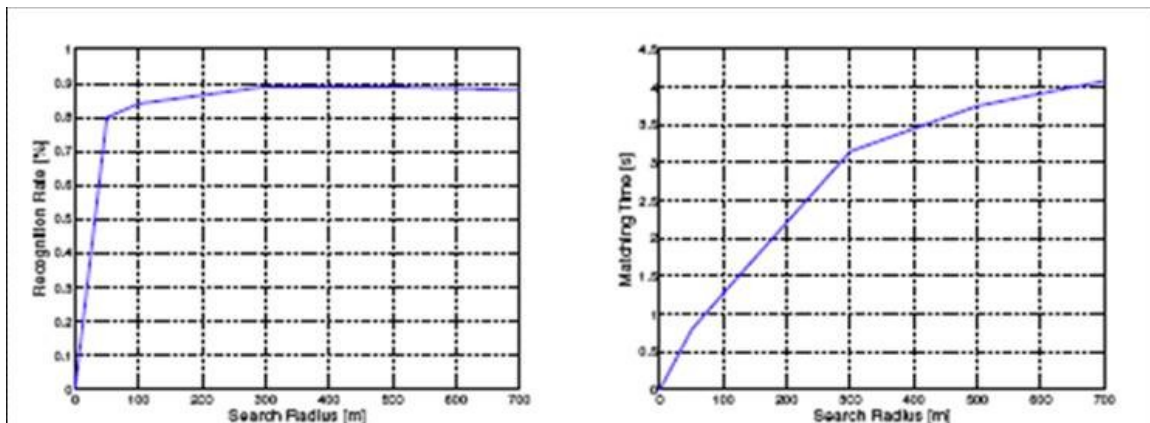
Πίνακας 2. Σύνοψη δεικτών επιτυχίας αναγνώρισης των διαφανειών της βάσης δεδομένων (Πηγή: [Quack, Bay, Van Gool, 2008](#))

Κατά βάση, γραμμική αναζήτηση, στο σύνολο της βάσης δεδομένων χωρίς γεωμετρικό φίλτρο επιτυγχάνει ρυθμό αναγνώρισης 67,4%. Αυτή η τιμή προσαυξάνεται κατά πάνω από 20% με την προσθήκη του γεωμετρικού φίλτρου, καταλήγοντας σε ρυθμό αναγνώρισης 88% . Αυτό οφείλεται στην απομάκρυνση των λανθασμένων θετικών ταυτίσεων. Η μεγάλη ακρίβεια κοστίζει όμως σε ταχύτητα.

Περιορισμός της αναζήτησης με βάση τη θέση GPS με μια ακτίνα 300 μέτρων είναι περίπου 40% γρηγορότερη ενώ αυξάνει ελαφρώς η ακρίβεια για την περίπτωση του γεωμετρικού φίλτρου και πιο ουσιαστικά για την περίπτωση χωρίς φίλτρο. Ο περιορισμός κατά CGI cell-tower είναι ελαφρώς γρηγορότερη άλλα σημαντικά χαμηλότερων επιδόσεων σε ακρίβεια αποτελεσμάτων. Αυτό φαίνεται να οφείλεται στο γεγονός ότι οι αντιστοι-

χίσεις CGI για διαφορετικούς διαχειριστές μπορεί να είναι ανολοκλήρωτες. Σε μια εφαρμογή στον πραγματικό κόσμο όπου ένας διαχειριστής δικτύου, ελπίζουμε, ότι θα συνεισέφερε τις πληροφορίες cell-id ή η ακτίνα αναζήτησης θα οριοθετούνταν από τις συντεταγμένες του GPS θα μπορούσαμε να αναμένουμε καλύτερα αποτελέσματα.

Συνολικά τα καλύτερα αποτελέσματα επιτυγχάνονται με GPS και μια σημαντικά μεγαλύτερη ακτίνα αρκετών εκατοντάδων μέτρων. Στην [Εικόνα 7](#) χαράσσεται η σχέση της ακρίβειας με το χρόνο για διαφορετικές ακτίνες. Στα 100 μέτρα τα περισσότερα αντικείμενα ανακτώνται σωστά αλλά μόνο μεταξύ 300 και 500 μέτρων επιτυγχάνεται ο ίδιος ρυθμός αναγνώρισης με τη γραμμική αναζήτηση αν και σε σημαντικά μεγαλύτερη ταχύτητα. Στη πραγματικότητα αυτή η επιτάχυνση έναντι της γραμμικής αναζήτησης προφανώς θα ήταν ακόμα μεγαλύτερη αν υπήρχαν περισσότερα αντικείμενα στη βάση δεδομένων. Οι χρόνοι αναγνώρισης θα μπορούσαν να επιταχυνθούν ακόμη περισσότερο με κατάλληλα ευρετήρια όπως στις εργασίες των [Indyk, Motwani](#) (1998) και [Nister, Stewenius](#) (2006). Έχουν συγκριθεί αρκετές μέθοδοι, όμως τα αποτελέσματα είναι προκαταρκτικά και εκτός του αντικειμένου αυτής της εργασίας.



Εικόνα 7. Ρυθμός αναγνώρισης (αριστερά) και ταχύτητα ταύτισης (δεξιά) βασισμένη σε ακτίνες γύρω από την θέση του ερωτήματος. (Πηγή: [Quack, Bay, Van Gool, 2008](#))

Οπτικά αποτέλεσμα εμφανίζονται στην [Εικόνα 6](#). Το πρώτο μέρος (α) δείχνει εικόνες ερωτημάτων στην αριστερή κολώνα και τις καλύτερες ταυτίσεις με τις εικόνες στη βάση δεδομένων για κάθε ερώτημα στη δεξιά. Παρατηρείται αντί η απόσταση στην εικόνα του ερωτήματος πως αυτή στην εικόνα της βάση δεδομένων στην πρώτη γραμμή και τις εστιάσεις και της χαμηλής εκκίνησης του ερωτήματος στη δεύτερη γραμμή. Το δεύτερο

μέρος (β) περιέχει ένα ερώτημα εικόνας επάνω και την καλύτερη ταύτιση της βάσης δεδομένων κάτω. Εκτός από την αλλαγή της οπτικής γωνίας και του διαφράγματος και άλλων ρυθμίσεων, της φωτογραφικής μηχανής με την οποία έγινε η λήψη, παρατηρείται ότι και η εικόνα του ερωτήματος και αυτή της βάσης δεδομένων έχει διαφορετικά σύννεφα και φωτισμό, αφού λήφθηκαν με διαφορά αρκετών εβδομάδων. Το τρίτο μέρος (γ) δείχνει ένα άλλο ζεύγος ερωτήματος της βάσης δεδομένων αυτή τη φορά για μία πρόσοψη με έντονη περικοπή και αλλαγή γωνίας. Η τελευταία εικόνα στο τέταρτο μέρος (δ), περιέχει ένα "αρνητικό" ερώτημα εικόνας, το οποίο δεν θα έπρεπε να επιστρέψει κανένα αποτέλεσμα αντιστοίχισης, όπως και πολύ σωστά συμβαίνει.

Τα αποτελέσματα αναδεικνύουν τα πλεονεκτήματα της προτεινόμενης προσέγγισης: το γεωμετρικό φίλτρο βελτιώνει δραστικά το ρυθμό αναγνώρισης. Ο περιορισμός της έρευνας σε μια γεωγραφική ακτίνα μερικών εκατοντάδων μέτρων αυξάνει σημαντικά την ταχύτητα ακόμη και στη δοκιμαστική βάση δεδομένων και θα είναι πρωταρχικής σημασίας για εφαρμογές μεγάλης κλίμακας στο πραγματικό κόσμο. Ταυτόχρονα τα αποτελέσματα δείχνουν ότι η επανάπαυση μόνο στις πληροφορίες των GPS (για αντικείμενα μέχρι μερικές δεκάδες μέτρα μακριά) δε θα ήταν δοκιμή για μια εφαρμογή οδηγού στον πραγματικό κόσμο. Έχοντας τη δυνατότητα να «επιλέξει» τα αντικείμενα με το κινητό του τηλέφωνο ο χρήστης απολαμβάνει σημαντικά πλεονεκτήματα ευχρηστίας.

Συμπεράσματα και Προοπτικές

Παρουσιάστηκε μια προσέγγιση για την αναγνώριση αντικειμένων για το Ίντερνετ των «Πραγμάτων», η οποία επιτρέπει στους χρήστες να αναζητήσουν πληροφορία για αντικείμενα φωτογραφίζοντάς τα. Έγινε εφαρμογή και επίδειξη ενός πλήρους συστήματος και αξιολόγηση των δυνατοτήτων του σε δύο προκλητικά σενάρια: tagging διαφανειών και προσθήκη σελιδοδεικτών από τις οθόνες έξυπνων αιθουσών συνεδριάσεων και ενός οδηγού πόλης σε κινητό τηλέφωνο. Και για τις δύο εφαρμογές ένα σύστημα αναγνώρισης αντικειμένων εκτελεί την ακόλουθη διαδικασία σε μια εφαρμογή στον server:

- Τοπικά χαρακτηριστικά εξάγονται από την εισερχόμενη εικόνα.
- Τα χαρακτηριστικά αντιστοιχίζονται στη βάση δεδομένων όπου ο χώρος αναζήτησης μπορεί προαιρετικά να περιοριστεί από μεταδεδομένα που λαμβάνονται με

το αίτημα, για παράδειγμα με τη γεωγραφική θέση από τις συντεταγμένες του GPS ή τις cell-tower ids.

- Τα υποψήφια αποτελέσματα αντιστοίχισης επικυρώνονται με ένα καθολικό γεωμετρικό φίλτρο.

Το σύστημα ολοκληρώνεται με ένα λογισμικό για τον client που μεταδίδει την εικόνα του ερωτήματος και τα μεταδεδομένα στον server με ένα μόνο κλικ.

Έγινε επίδειξη της ευελιξίας της προτεινόμενης προσέγγισης με μια πειραματική αξιολόγηση και για τις δύο δειγματικές εφαρμογές. Σε ό,τι αφορά αυτό, το σύστημα αξιολογήθηκε για δύο προκλητικά δοκιμαστικά σύνολα δεδομένων. Βασίζόμενοι σε τοπικά χαρακτηριστικά και ενδυναμώνοντας το ρυθμό αναγνώρισης με ένα γεωμετρικό φίλτρο επετεύχθησαν πολύ υψηλοί ρυθμοί αναγνώρισης. Αυτή η προσέγγιση λειτούργησε καλά τόσο για την ταύτιση διαφανειών με μεγάλες ποσότητες κειμένων και εικόνων όσο και για τα τουριστικά αξιοθέατα από σημαντικά ποικίλες οπτικές γωνίες, γεγονός που υπογραμμίζει την ευελιξία της προτεινόμενης προσέγγισης. Για την εξαιρετικά προκλητική εφαρμογή του οδηγού πόλης μπόρεσε να βρεθεί μια καλή ισορροπία μεταξύ εκτέλεσης και ρυθμού αναγνώρισης με περιορισμό του χώρου αναζήτησης χρησιμοποιώντας πληροφορίες θέσεις GPS.

Τα αποτελέσματα έδειξαν ότι το Ίντερνετ των Πραγμάτων με την αναγνώριση αντικειμένων μπορεί να γίνει πραγματικότητα ακόμη και σήμερα για συγκεκριμένους τύπους αντικειμένων. Στην πραγματικότητα το σύστημα μπορεί να ιδωθεί ως μια μηχανή οπτικής αναζήτησης για το Ίντερνετ των πραγμάτων. Βασιζόμενο στην απλή αποστολή εικόνων μέσω κινητού τηλεφώνου, το σύστημα μπορεί εύκολα να υιοθετηθεί και από τελικούς χρήστες και από προμηθευτές συστημάτων. με την εξέλιξη των οπτικών μεθόδων των υπολογιστών, αναμένεται πληθώρα πρόσθετων ιδιοτήτων στα επόμενα χρόνια.

Πέρα όμως από το Ίντερνετ των Πραγμάτων, μήπως η συγκεκριμένη εφαρμογή θα μπορούσε να αξιοποιηθεί και για άλλους σκοπούς; Για παράδειγμα κατά πόσον θα ήταν εφικτό να υπάρχει μια εφαρμογή όπου κάποιος χρήστης που τυχόν θα έπεφτε θύμα εγκλήματος να φωτογραφίζει το θύτη και να αποστέλλονται απευθείας δεδομένα σε κάποια βάση της αστυνομίας; Κάτι παρεμφερές συμβαίνει ήδη με τις κάμερες κυκλοφορίας. Και αυτό ακόμα θα μπορούσε να βελτιωθεί αυτοματοποιώντας τη διαδικασία και σχεδιάζοντας μία εφαρμογή όπου ο αριθμός κυκλοφορίας θα αναγνωρίζεται αυτοματο-

ποιημένα από το λογισμικό και θα αντιστοιχίζεται με τα στοιχεία του κατόχου στις βάσεις δεδομένων του Υπουργείου Συγκοινωνιών. Ανάλογες εφαρμογές μπορούν αξιοποιηθούν ακόμη και στο χώρο των μουσείων όπου κι εκεί τα αντικείμενα είναι τρισδιάστατα και σχετικά μετρήσιμα, ώστε ο χρήστης στα εκθέματα τα οποία επιτρέπεται να φωτογραφήσει να μπορεί αυτόματα να εντοπίζει τις σχετικές πληροφορίες και να δημιουργεί σελιδοδείκτες με αυτές στο κινητό του ώστε να μπορεί να έχει πρόσβαση και μετά την απομάκρυνσή του από αυτό. Τόσο οι τεχνικές όσο και οι εφαρμογές αυτές, αναπτύσσονται αυτή τη στιγμή και στην πορεία του χρόνου, σίγουρα θα βρεθούν πολύ περισσότεροι τρόποι για την αξιοποίησή τους.

Βιβλιογραφία

Abowd, G. Classroom, (1999). “An experiment with the instrumentation of a living educational environment”. In: *IBM Systems Journal*, Volume 38, Number 4. Ανάκτηση 31/5/2008 από <https://www.research.ibm.com/journal/sj/384/abowd.html>

Adelmann, R., Langheinrich, M., Floerkemeier, C. “Toolkit for Bar Code recognition and resolving on camera phones – jump starting the internet of things”. In: *Workshop at Informatik (2006): Mobile and Embedded Interactive Systems (MEIS 2006): Proceedings*. Ανάκτηση 31/5/2008 από <http://www.hcilab.org/events/meis06/meis-paper01.pdf>

Amir, A., Ashour, G., Srinivasan, S. “Towards automatic real time preparation of on-line video proceedings for conference talks and presentations”. In: *System Sciences, 2001. Proceedings of the 34th Annual Hawaii International Conference, on*: 3-6 Jan. 2001. Ανάκτηση 31/5/2008 από:

<http://ieeexplore.ieee.org/Xplore/login.jsp?url=/iel5/7255/20032/00926513.pdf?tp=&isnumber=&arnumber=926513>

Ballagas, R., Rohs, M., Sheridan, J., (2005), “Mobile Phones as Pointing Devices”. In: *PERMID Workshop, Pervasive 2005, Munich, May 2005*. Ανάκτηση 31/5/2008 από: <http://www.comp.lancs.ac.uk/~sheridaj/publications/2005/PERMIDPointingDevices.pdf>

Bay, H. Fasel, B. Van Gool, L. “Interactive Museum Guide: Fast and Robust Recognition of Museum Objects”. In: *Proceedings of the first international workshop on mobile vision, May 2006*. Ανάκτηση 31/5/2008 από: ftp://ftp.vision.ee.ethz.ch/publications/proceedings/eth_biwi_00394.pdf

Bay, H., Tuytelaars, T., Van Gool, L. “Surf: Speeded up robust features”. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) *ECCV 2006. LNCS, vol. 3951, pp. 404–417. Springer, Heidelberg (2006)*. Ανάκτηση 31/5/2008 από: <http://www.springerlink.com/content/e580h2k58434p02k/>

Boring, S., [et. al]. “Shoot & copy: Phonecam-based information transfer from public displays onto mobile phones”. In: *Proceedings of the International Conference on Mobile Technology, Applications and Systems (Mobility), Singapore, Sep. 2007*. Ανάκτηση 31/5/2008 από:

<http://www.medien.ifi.lmu.de/pubdb/publications/pub/boring2007ubicompdemo/boring2007ubicompdemo.pdf>

Carletta, J., et al. “The ami meeting corpus: A pre-announcement”. In: Renals, S., Bengio, S. (eds.) *MLMI 2005. LNCS, vol. 3869, pp. 28–39. Springer, Heidelberg (2006)*. Ανάκτηση 31/5/2008 από: <http://www.idiap.ch/~mccowan/publications/carletta-mlmi2005.pdf>

Donoser, M., Bischof, H. “Efficient maximally stable extremal region (MSER) tracking”. In: *Computer Vision and Pattern Recognition, 2006 IEEE Conference on*.

Fielding, R. [et. al], “9 Method Definitions”. In: *RFC 2616, Hypertext Transfer Protocol -HTTP/1.1*. Ανάκτηση 1/6/2008 από: <http://www.w3.org/Protocols/rfc2616/rfc2616-sec9.html>

Fischler, M.A., Bolles, R.C.: “Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography”. In: *Communications of the ACM* (1981). Ανάκτηση 13/6/2008 από:

<http://delivery.acm.org/10.1145/360000/358692/p381-fischler.pdf?key1=358692&key2=1506343121&coll=portal&dl=ACM&CFID=6279689&CF-TOKEN=93546774>

Fockler, P., Zeidler, T., Bimber, O. “Phone guide: Museum guidance supported by on-device object recognition on mobile phones”. In: *LNCS, vol. 3951, pp. 404–417. Springer, Heidelberg* (2006). Ανάκτηση 31/5/2008 από

<http://www.springerlink.com/content/vt27150741362620/>

Fuhrmann, T., Harbaum, T. “Using bluetooth for informationally enhanced environments”. In: *Proceedings of the IADIS International Conference e-Society 2003*.

<http://i30www.ira.uka.de/research/documents/p2p/2003/fuhrmann03informationally.pdf>

Hartley, R.I., Zisserman, A. “Multiple View Geometry in Computer Vision”, 2nd ed. Cambridge University Press, Cambridge (2004). Ανάκτηση 31/5/2008 από:

http://books.google.com/books?hl=el&lr=&id=si3R3Pfa98QC&oi=fnd&pg=PR11&dq=Hartley+Zisserman+%22Multiple+View+Geometry+in+Computer+Vision%22&ots=aPpZkvab5R&sig=XnoQXJwJL5DzoyiOO_YVJTIWd0#PPA26,M1

Indyk, P., Motwani, R. “Approximate nearest neighbors: towards removing the curse of dimensionality”. In *Proceedings of the Thirtieth Annual ACM Symposium on theory of Computing* (Dallas, Texas, United States, May 24 - 26, 1998). STOC '98. ACM, New York. Ανάκτηση 31/5/2008 από:

<http://portal.acm.org/citation.cfm?id=276876&dl=GUIDE.#>

Leibe, B., Seemann, E., Schiele, B. “Pedestrian detection in crowded scenes”. In: *IEEE Conference on Computer Vision and Pattern Recognition* (2005). Ανάκτηση 31/5/2008 από www.mis.informatik.tu-darmstadt.de/People/seemann/leibe-crowdedscenes-cv-pr05.pdf

Lowe, D. “Distinctive image features from scale-invariant keypoints”. In: [International Journal of Computer Vision](http://www.cs.ubc.ca/~lowe/papers/ijcv04.pdf), 60, 2, pp. 91-110, 2004. Ανάκτηση 31/5/2008 από <http://www.cs.ubc.ca/~lowe/papers/ijcv04.pdf>

Mikolajczyk, K., Schmid, C. “A performance evaluation of local descriptors”. In: *IEEE Transaction on Pattern Analysis and Machine Intelligence PAMI*, v. 27 no 10, pp. 1615–1630 (2005). Ανάκτηση 31/5/2008 από

<http://csdl2.computer.org/persagen/DLabsToc.jsp?resourcePath=/dl/trans/tp/&toc=comp/trans/tp/2005/10/ixtoc.xml&DOI=10.1109/TPAMI.2005.188>

Niblack, W., “Slidefinder: A tool for browsing presentation graphics using contentbased retrieval”. In: *Content-Based Access of Image and Video Libraries, 1999 IEEE Workshop*, pp. 114 - 118 Ανάκτηση 31/5/2008 από

<http://ieeexplore.ieee.org/Xplore/login.jsp?url=/iel5/6335/16935/00781134.pdf?temp=x>

Nister, D., Stewenius, H.: “Scalable recognition with a vocabulary tree”. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, v. 2, pp. 2161-2168, June 2006, Ανάκτηση 31/5/2008 από

http://www.vis.uky.edu/~stewe/publications/nister_stewenius_cvpr2006.pdf

Paletta, L. [et. al]. “A mobile vision service for multimedia tourist applications in urban environments”. In: *IEEE Intelligent Transportation Systems Conference (ITSC)*, (2006)
Ανάκτηση 31/5/2008 από

http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=1706801

Quack, T, Bay, H., Van Gool, L., “Object Recognition for the Internet Things” Quack, H. Bay and L. Van Gool, In: *LNCIS*, vol. 4952, pp. 230–246. Springer, Heidelberg (2008).
Ανάκτηση 13/6/2008 από

<http://www.springerlink.com/content/e80036wg23qu41w5/fulltext.pdf>

Rohs, M., Gfeller, B. “Using camera-equipped mobile phones for interacting with real-world objects”. In: Ferscha, A., Hoertner, H., Kotsis, G. (eds.) *Advances in Pervasive Computing, Austrian Computer Society (OCG)* (2004). Ανάκτηση 31/5/2008 από
<http://www.vs.inf.ethz.ch/res/papers/rohs-gfeller-visualcodes-2004.pdf>

Sivic, J., Zisserman, A. “Video Google: A text retrieval approach to object matching in videos”. In: *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on* (2003), vol. 2, pp. 1470-1477. Ανάκτηση 1/6/2008 από: <http://ieeexplore.ieee.org/Xplore/login.jsp?url=/iel5/8769/27772/01238663.pdf?tp=&isnum-ber=&arnumber=1238663>

Vinciarelli, A., Odobez, J.-M. “Application of information retrieval technologies topresentation slides”. In: *IEEE Transactions on Multimedia* (2006), vol. 8, no 5, pp. 981-995. Ανάκτηση 1/6/2008 από

http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=1703512

Want, R.: RFID - A key to automating everything. In: *Scientific American*, vol. 290 no. 1, pg. 46--55 (2004).

Wikipedia contributors. "Beamer". In: *Wikipedia, The Free Encyclopedia*. Ανάκτηση 14/6/2008 από <http://en.wikipedia.org/wiki/Beamer>

Wikipedia contributors. "Epipolar Geometry". In: *Wikipedia, The Free Encyclopedia*. Ανάκτηση 14/6/2008 από: http://en.wikipedia.org/wiki/Epipolar_geometry#Epipolar_line

Wikipedia contributors. "Fundamental matrix (computer vision)". In: *Wikipedia, The Free Encyclopedia*. Ανάκτηση 14/6/2008 από:
http://en.wikipedia.org/wiki/Fundamental_matrix_%28computer_vision%29

Wikipedia contributors. "Homography". In: *Wikipedia, The Free Encyclopedia*. Ανάκτηση 14/6/2008 από: <http://en.wikipedia.org/wiki/Homography>

Wikipedia contributors. "Hypertext Transfer Protocol". In: *Wikipedia, The Free Encyclopedia*. Ανάκτηση 14/6/2008 από:
http://en.wikipedia.org/wiki/Hypertext_Transfer_Protocol

Wikipedia contributors. “RANSAC”. In: *Wikipedia, The Free Encyclopedia*, Ανάκτηση 14/6/2008 από <http://en.wikipedia.org/wiki/RANSAC>

Wikipedia contributors. “Scale-invariant feature transform”. In: *Wikipedia, The Free Encyclopedia*, Ανάκτηση 14/6/2008 από http://en.wikipedia.org/wiki/Scale-invariant_feature_transform

Wikipedia contributors. "Stemming". In: *Wikipedia, The Free Encyclopedia*. Ανάκτηση 14/6/2008 από <http://en.wikipedia.org/wiki/Stemming>

Wikipedia contributors. "SURF". In: *Wikipedia, The Free Encyclopedia*. Ανάκτηση 14/6/2008 από <http://en.wikipedia.org/w/index.php?title=SURF&oldid=218411051>

Wikipedia contributors. "Tablet PC". In: *Wikipedia, The Free Encyclopedia*. Ανάκτηση 14/6/2008 από http://en.wikipedia.org/wiki/Tablet_pc

Wikipedia contributors. "Trusted time-stamping". In: *Wikipedia, The Free Encyclopedia*. Ανάκτηση 14/6/2008 από http://en.wikipedia.org/wiki/Trusted_timestamping

Γλωσσάρι

ΟΡΟΣ ΠΡΩΤΟΤΥΠΟΥ (Αγγλικά)	ΟΡΟΣ ΜΕΤΑΦΡΑΣΗΣ (Ελληνικά)
Annotation	Σχολιασμός
Barcode	Γραμμοκώδικας
Cell ids	Ταυτότητες κυψέλης
Celltower ids	Ταυτότητες πύργου κυψέλης
Client	Πελάτης
Global	Σφαιρικός
Global geometry	Γεωμετρία Σφαιρών
Highlight	Επισήμανση
Interaction	Αλληλεπίδραση
Interactive	Διαδραστικός
Mapping	Αντιστοίχιση
Marker	Δείκτης
Object	Αντικείμενο
OCR - Optical Character Recognition	Οπτική Αναγνώριση Χαρακτήρων
Pipelines	Δικτύωση / Δίκτυο
Presentation Screen	Οθόνη παρουσίασης
Projective geometry constraints	Όρια στην προβολική γεωμετρία
Query	Ερώτημα
Server	Εξυπηρετητής
Slide	Διαφάνεια
Tagging/Tag	Προσθήκη ετικετών / ετικέτες
Thing	«Πράγμα»

Με κόκκινο σημειώνεται ο όρος που προτιμήθηκε να χρησιμοποιηθεί στο κείμενο ανά περίπτωση

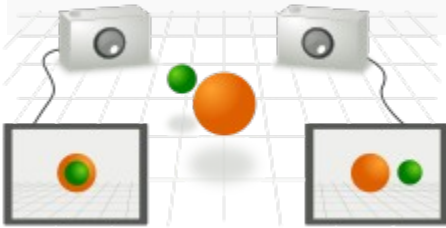
Παράρτημα εννοιών που χρήζουν ανάλυσης

Beamer

Ιδιωματισμός για τις έννοιες του προβολέα διαφανειών ή βίντεο-προβολέα.

Epipolar Geometry και epipolar lines⁷

Η Epipolar geometry αναφέρεται στη γεωμετρία της οπτικής στερεών. Όταν δύο φωτογραφικές μηχανές έχουν στο πεδίο τους ένα τρισδιάστατο αντικείμενο (σκηνικό, όψη, φυσικό αντικείμενο) από δύο διαφορετικές θέσεις, υπάρχει ένας αριθμός γεωμετρικών σχέσεων μεταξύ των τρισδιάστατων σημείων και της προβολής τους στις δισδιάστατες εικόνες που οδηγεί σε περιορισμούς μεταξύ των σημείων απεικόνισης. Αυτές οι σχέσεις εξάγονται βάσει της υπόθεσης ότι η φωτογραφική μηχανή μπορεί να προσεγγισθεί από ένα μοντέλο φωτογραφικής μηχανής οπής ([pinhole camera model](#)).



Εικόνα 8. Τυπική περίπτωση χρήσης της epipolar geometry. Δύο φωτογραφικές μηχανές φωτογραφίζουν το ίδιο αντικείμενο από διαφορετικές οπτικές γωνίες. Η epipolar geometry, στη συνέχεια, περιγράφει τη σχέση μεταξύ των δύο όψεων του αποτελέσματος. (Πηγή: Wikipedia Contributors)

Η [Εικόνα 9](#) απεικονίζει δύο φωτογραφικές μηχανές οπής που στοχεύουν το σημείο X. στις πραγματικές φωτογραφικές μηχανές, το επίπεδο της απεικόνισης βρίσκεται στην πραγματικότητα πίσω από σημείο εστίασης, παράγοντας μια ανεστραμμένη εικόνα. Εδώ, όμως, το πρόβλημα της προβολής απλοποιείται με την τοποθέτηση ενός εικονικού επιπέδου της εικόνας μπροστά από το σημείο εστίασης της κάθε φωτογραφικής μηχανής ώστε να παράγει μία μη ανεστραμμένη εικόνα. Τα OL και OR αναπαριστούν τα σημεία εστίασης των δύο φωτογραφικών μηχανών. Το X αναπαριστά το σημείο ενδιαφέροντος και στις 2 φωτογραφικές μηχανές. Τα σημεία xL και xR είναι οι προβολές του σημείου X στα επίπεδα των εικόνων.

⁷ Δεν βρέθηκε δόκιμος όρος για τη μετάφραση της ορολογίας.

Κάθε φωτογραφική μηχανή, απαθανατίζει μια δισδιάστατη εικόνα του τρισδιάστατου κόσμου. Αυτή η μετατροπή από τρισδιάστατο σε δισδιάστατο αναφέρεται ως μια οπτική προβολή και περιγράφεται από το μοντέλο της φωτογραφικής μηχανής οπής. Είναι κοινός τόπος η διαμόρφωση αυτής της λειτουργίας προβολής από ακτίνες που It is common to model this projection operation by rays that εκπέμπονται από την φωτογραφική μηχανή περνώντας από το σημείο εστίασής της. Σημειώνεται ότι κάθε εκπεμπόμενη ακτίνα αντιστοιχεί σε ένα μοναδικό σημείο της εικόνας.

Epipole

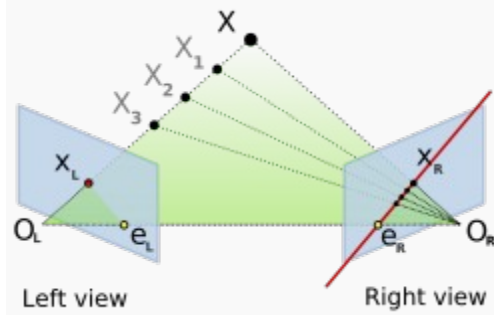
Από τη στιγμή που 2 σημεία εστίασης των φωτογραφικών μηχανών είναι διακριτά, κάθε σημείο εστίασης προβάλλεται σε ένα διακριτό σημείο στο επίπεδο της απεικόνισης της άλλης φωτογραφικής μηχανής. Αυτά τα δύο σημεία της εικόνας σημειώνονται στο σχήμα ως E_L and E_R και ονομάζονται *epipoles*. Και τα δύο epipoles E_L και E_R και τα σημεία εστίασης O_L και O_R ευθυγραμμίζονται (βρίσκονται πάνω σε μία μοναδική ευθεία).

Epipolar line

Η ευθεία O_L-X είναι ορατή από την αριστερή φωτογραφική μηχανή ως ένα σημείο γιατί είναι άμεσα ευθυγραμμισμένη με το σημείο εστίασής της. Όμως, η δεξιά φωτογραφική μηχανή βλέπει αυτή τη γραμμή σαν μια μηχανή στο επίπεδο της απεικόνισης. Αυτή η γραμμή (E_R-x_R) στη δεξιά φωτογραφική μηχανή ονομάζεται *epipolar line*. Συμμετρικά, η γραμμή O_R-X φαίνεται από την δεξιά φωτογραφική μηχανή όπως φαίνεται ένα σημείο στην epipolar line E_L-x_L από την αριστερή φωτογραφική μηχανή.

Epipolar plane

Ως εναλλακτική οπτικοποίηση, θεωρούμε ότι τα σημεία X , O_L & O_R σχηματίζουν ένα επίπεδο που καλείται *epipolar plane*. Το epipolar plane τέμνει το επίπεδο απεικόνισης της κάθε φωτογραφικής μηχανής στα σημεία που σχηματίζονται γραμμές, οι epipolar lines. Όλες οι epipolar lines τέμνουν το epipole ανεξάρτητα από το πού τοποθετείται το X .



Εικόνα 9. Epipolar geometry (Πηγή: Wikipedia Contributors)

Epipolar περιορισμός και τριγωνισμός

Αν η σχετική μετατόπιση (χωρίς αλλαγή της φοράς των αξόνων) και η περιστροφή των δύο φωτογραφικών μηχανών είναι γνωστή, η αντίστοιχη epipolar geometry οδηγεί σε δύο σημαντικές παρατηρήσεις:

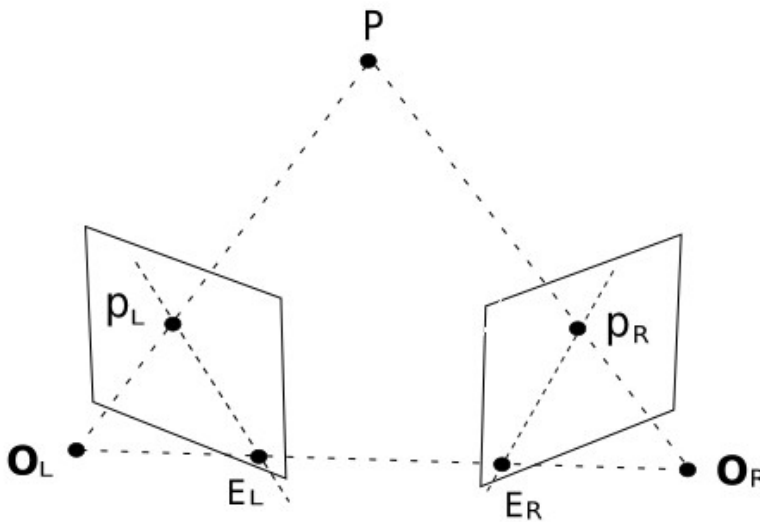
Αν το σημείο προβολής $\mathbf{x}_L$ είναι γνωστό, τότε η epipolar line $\mathbf{E}_R - \mathbf{x}_R$ είναι γνωστή και το σημείο $\mathbf{X}$ προβάλλεται στη σωστή απεικόνιση, σε ένα σημείο $\mathbf{x}_R$ που πρέπει να βρίσκεται πάνω στη συγκεκριμένη epipolar line. Αυτό σημαίνει ότι κάθε σημείο που παρατηρείται σε μια απεικόνιση το ίδιο σημείο πρέπει να παρατηρείται και στην άλλη απεικόνιση πάνω σε μια γνωστή epipolar line. Αυτό παρέχει ένα *epipolar constraint* (περιορισμό) για το ποια σημεία αντιστοίχισης της απεικόνισης θα πρέπει να μας ικανοποιούν και σημαίνει ότι είναι δυνατό να δοκιμαστεί το αν δύο σημεία στην πραγματικότητα αντιστοιχούν στο ίδιο τρισδιάστατο σημείο. Τα Epipolar constraints μπορούν επίσης να περιγραφούν από τον [essential matrix](#) ή τον [fundamental matrix](#) μεταξύ των δύο φωτογραφικών μηχανών.

Αν τα σημεία $\mathbf{x}_L$ και $\mathbf{x}_R$ είναι γνωστά, οι γραμμές προβολής είναι επίσης γνωστές. Αν τα δύο σημεία απεικόνισης αντιστοιχούν στο ίδιο τρισδιάστατο σημείο $\mathbf{X}$ οι γραμμές προβολής πρέπει να τέμνονται ακριβώς στο $\mathbf{X}$. Αυτό σημαίνει ότι το $\mathbf{X}$ μπορεί να υπολογιστεί από τις συντεταγμένες των δύο σημείων απεικόνισης, διαδικασία που ονομάζεται τριγωνισμός.

Απλοποιημένες περιπτώσεις

Η epipolar geometry απλοποιείται αν το επίπεδο των απεικονίσεων των δύο φωτογραφικών μηχανών είναι ίδιου μεγέθους. Σε αυτή την περίπτωση, ίδιου μεγέθους είναι

και οι epipolar lines ($\mathbf{E}_L - \mathbf{x}_L = \mathbf{E}_R - \mathbf{x}_R$). Επιπλέον, οι epipolar lines είναι παράλληλες προς τη γραμμή $\mathbf{O}_L - \mathbf{O}_R$ μεταξύ των δύο σημείων εστίασης και, πρακτικά, μπορούν να ευθυγραμμιστούν με τους οριζόντιους άξονες των δύο απεικονίσεων. Αυτό σημαίνει ότι για κάθε σημείο σε μία απεικόνιση, το αντίστοιχο σημείο του στην άλλη απεικόνιση μπορεί να βρεθεί απλά κοιτώντας κατά μήκος της οριζόντιας γραμμής. Αν οι δύο φωτογραφικές μηχανές δεν μπορούν να τοποθετηθούν κατ' αυτό τον τρόπο, οι συντεταγμένες της απεικόνισης από τις φωτογραφικές μηχανές, μπορεί να μετατραπεί για να την αποκαταστήσει έχοντας ένα κοινό επίπεδο απεικόνισης. Αυτή η λειτουργία ονομάζεται αποκατάσταση εικόνας.



Εικόνα 10. Παράδειγμα epipolar geometry. Δύο φωτογραφικές μηχανές με τα αντίστοιχα σημεία εστίασής του $\mathbf{O}_L$ and $\mathbf{O}_R$, παρατηρούν το σημείο $\mathbf{P}$. Η προβολή του $\mathbf{P}$ στο καθένα από τα επίπεδα της εικόνας σημειώνεται με τα $\mathbf{p}_L$ και $\mathbf{p}_R$. Τα σημεία $\mathbf{E}_L$ και $\mathbf{E}_R$ είναι οι epipoles. (Πηγή: Wikipedia Contributors)

Θεμελιώδης πίνακας $\mathbf{F}$

Ο θεμελιώδης πίνακας $\mathbf{F}$ είναι ένας πίνακας 3×3 της τάξης δύο που συσχετίζει τα αντίστοιχα σημεία σε απεικονίσεις στερεών. Στην [epipolar geometry](#), με συντεταγμένες ομογενούς απεικόνισης $\mathbf{y}_1$ και $\mathbf{y}_2$ των αντίστοιχων σημείων σε ένα ζεύγος στερεών εικόνων το $\mathbf{F}\mathbf{y}_1$ περιγράφει μια γραμμή (μια [epipolar line](#)) πάνω στην οποία πρέπει να βρίσκεται το αντίστοιχο σημείο $\mathbf{y}_2$ στην άλλη απεικόνιση. Αυτό σημαίνει ότι, για όλα τα ζεύγη των αντίστοιχων σημείων ισχύει ο τύπος:

$$\mathbf{y}_2^T \mathbf{F} \mathbf{y}_1 = 0.$$

Όντας της τάξης δύο και οριζόμενος μόνο πάνω σε κλίμακα, ο θεμελιώδης πίνακας μπορεί να υπολογιστεί με δεδομένα τουλάχιστον 7 σημεία αντιστοιχίας. Οι 7 παράμετροί του αναπαριστούν μόνο τη γεωμετρική πληροφορία για τις φωτογραφικές μηχανές που μπορούν να αποκτηθούν μόνο από σημεία αντιστοιχίας.

Homography

Η ομογραφία είναι μια έννοια της μαθηματικής επιστήμης της γεωμετρίας. Ορίζεται ως η σχέση μεταξύ 2 εικόνων, τέτοια που κάθε δεδομένο σημείο στη μία εικόνα αντιστοιχεί σε ένα μοναδικό σημείο της άλλης και αντιστρόφως.

Εφαρμογές οπτικής υπολογιστών

Στο πεδίο της οπτικής των υπολογιστών, μια ομογραφία σε ένα δισδιάστατο χώρο ορίζεται ως μια αντιστοίχιση μεταξύ ενός σημείου σε ένα επίπεδο όπως φαίνεται από μία φωτογραφική μηχανή, με το ίδιο σημείο στο επίπεδο όπως φαίνεται από τη δεύτερη φωτογραφική μηχανή. Αυτό έχει πολλές πρακτικές εφαρμογές, κυρίως παρέχει μέθοδο compositing δισδιάστατα ή τρισδιάστατα αντικείμενα σε μια εικόνα ή βίντεο με τη σωστή στάση. Ο ομογραφικός πίνακας είναι μερικές φορές γνωστός ως ομογραφία.

Εξίσωση τρισδιάστατου επιπέδου προς επίπεδο

Έχουμε δύο φωτογραφικές μηχανές a και b, που βλέπουν στα σημεία P_i ενός επιπέδου.

Περνώντας τις προβολές του P_i από το ${}^b p_i$ στο b σε ένα σημείο ${}^a p_i$ στο a:

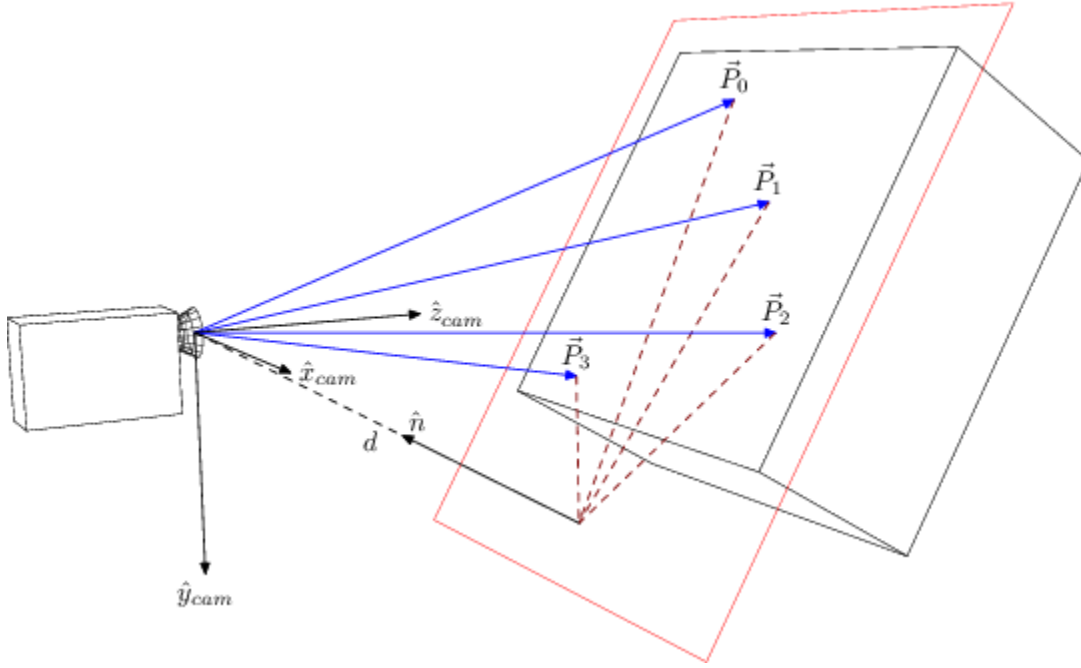
$${}^a p_i = K_a \cdot H_{ba} \cdot K_b^{-1} \cdot {}^b p_i$$

Όπου H_{ba} είναι:

$$H_{ba} = R - \frac{tn^T}{d}$$

Το R είναι ο πίνακας περιφοράς βάσει του οποίου το b περιστρέφεται σε σχέση με το a . t είναι η μεταφορά του διανύσματος από το a στο b ; n και d είναι το κανονικό διάνυσμα του επιπέδου και η απόσταση προς το αντίστοιχο επίπεδο.

K_a και K_b είναι οι πίνακες των εσωτερικών παραμέτρων των φωτογραφικών μηχανών.



Εικόνα 11. Το σχήμα δείχνει πώς η φωτογραφική μηχανή b βλέπει στο επίπεδο σε απόσταση d .

Μαθηματικός ορισμός

Δεδομένα

$$p_a = \begin{bmatrix} x_a \\ y_a \\ 1 \end{bmatrix}, p_b = \begin{bmatrix} x_b \\ y_b \\ 1 \end{bmatrix}, \mathbf{H}_{ab} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}$$

Ωστε

$$p_b = \mathbf{H}_{ab} p_a$$

and

$$p_a = \mathbf{H}_{ba} p_b$$

όπου

$$\mathbf{H}_{ba} = \mathbf{H}_{ab}^{-1}.$$

POST

Υποβάλλει δεδομένα για επεξεργασία (π.χ. από μια μορφή HTML) στον ταυτοποιημένο πόρο. Τα δεδομένα συμπεριλαμβάνονται στο σώμα του αιτήματος. Αυτό μπορεί να οδηγήσει στη δημιουργία ενός νέου πόρου ή ενημέρωσης / ανανέωσης του υπάρχοντος πόρου ή και τα δύο.

Το POST σχεδιάστηκε να επιτρέπει μια ενιαία μέθοδο κάλυψης των ακόλουθων λειτουργιών:

- Σχολιασμό υπαρχόντων πόρων,
- Ανάρτηση μηνύματος σε ένα δελτίο, newsgroup, mailing list, ή παρεμφερή ομάδα άρθρων,
- Παροχή ενός συνόλου δεδομένα, όπως το αποτέλεσμα της κατάθεσης μιας μορφής σε μια διαδικασία διαχείρισης δεδομένων,
- Επέκταση μιας βάσης δεδομένων μέσω μιας πρόσθετης λειτουργίας.

Η ακριβής λειτουργία της μεθόδου POST ορίζεται από τον server και συνήθως εξαρτάται από το Request-URI.

RANSAC - RANdom Sample Consensus

Επαναληπτική μέθοδος για την εκτίμηση παραμέτρων ενός μαθηματικού μοντέλου από ένα σύνολο εμπειρικών δεδομένων που περιέχει outliers. Ο αλγόριθμος δημοσιεύτηκε πρώτα από Fischler και Bolles το 1981. Μια βασική υπόθεση είναι ότι τα δεδομένα αποτελούνται από "inliers", δηλαδή, δεδομένα των οποίων η κατανομή μπορεί να ερμηνευθεί από ένα σύνολο παραμέτρων του μοντέλου και "outliers", που είναι δεδομένα που δεν αρμόζουν στο μοντέλο. Επιπλέον, αυτά τα δεδομένα μπορεί να γίνουν υποκείμενα θορύβου. Τα outliers προέρχονται, π.χ., από:

- ακραίες τιμές θορύβου
- εσφαλμένες μετρήσεις
- λανθασμένες υποθέσεις για την ερμηνεία των δεδομένων.

Το RANSAC, επίσης, υποθέτει ότι με δεδομένο ένα (μικρό, συνήθως) σύνολο inliers, υπάρχει μια διαδικασία που μπορεί να εκτιμήσει τις παραμέτρους ενός μοντέλου που ερμηνεύει ή προσαρμόζει βέλτιστα αυτά τα δεδομένα.

SIFT - Scale-Invariant Feature Transform

Είναι ένας αλγόριθμος στους οπτικούς υπολογιστές για την ανίχνευση και περιγραφή τοπικών χαρακτηριστικών γνωρισμάτων στις εικόνες. Ο αλγόριθμος δημοσιεύθηκε από το David Lowe το 1999

Stemming

Είναι η διαδικασία για την αναγωγή των κλιτών (ή μερικές φορές των παραγώγων) λέξεων στη ρίζα ή στο θέμα τους. Σε γενικές γραμμές πρόκειται για μια μορφή λέξης. Το θέμα δεν είναι απαραίτητο να είναι πανομοιότυπη με τη μορφολογική ρίζα της λέξης. Συνήθως αρκεί σχετικές λέξεις να αντιστοιχούν στο ίδιο θέμα, ακόμα κι αν το θέμα της λέξης αυτό καθαυτό, δεν είναι από μόνο του μια έγκυρη ρίζα. Ο αλγόριθμος του stemming αποτελεί ένα μακροχρόνιο πρόβλημα στην επιστήμη των υπολογιστών· το πρώτο σχετικό άρθρο δημοσιεύθηκε το [1968](#). Η διαδικασία του stemming συχνά καλείται και συγχώνευση, είναι χρήσιμο στις μηχανές αναζήτησης για επέκταση ενός ερωτήματος ή ευρετηρίαση και άλλα προβλήματα επεξεργασίας τη φυσικής γλώσσας.

SURF - Speeded Up Robust Features

Είναι ένας εξελιγμένος ανιχνευτής και περιγραφέας για σημεία ενδιαφέροντος αμετάβλητης κλίμακας και περιστροφής.

Tablet PC

Ένα Tablet PC είναι ένα notebook ή ένας επίπεδος φορητός υπολογιστής, που πρωτοεμφανίστηκε από την Pen Computing στις αρχές της δεκαετίας του '90 η οποία κυκλοφόρησε το PenGo Tablet Computer και διαδόθηκε από τη Microsoft. Η οθόνη επαφής ή οι πίνακες γραφικών / υβριδικής τεχνολογίας οθόνης επιτρέπει στον χρήστη να λειτουργεί τον υπολογιστή με μία ψηφιακή γραφίδα, ή με την άκρη του δακτύλου, αντί με τη χρήση πληκτρολογίου ή ενός ποντικιού.

Η μορφή του προσφέρει στο χρήστη ένα πιο ευέλικτο τρόπο αλληλεπίδρασης με τον υπολογιστή. Τα Tablet PCs χρησιμοποιούνται συχνά όπου τα κανονικά notebooks είναι δύσκολα ή δεν παρέχουν την αναγκαία λειτουργικότητα.

Timestamping / Timestamps

Ο όρος αναφέρεται σε ετικέτες με δεδομένα ημερομηνίας και ώρας που επισυνάπτονται στα αντικείμενα προκειμένου να επιβεβαιώνεται η χρονική στιγμή της δημιουργίας ενός ψηφιακού πόρου και κατ' επέκταση ο δημιουργός να διεκδικεί και να κατοχυρώνει τα δικαιώματά του.

Ευρετήριο Εικόνων

Εικόνα 1. Χαρακτηριστικές φωτογραφικές διαφάνειες παρουσίασης από τη βάση δεδομένων AMI Corpus. (Πηγή: Carletta [et.al], 2006).....	11
Εικόνα 2. Ο χρήστης προσθέτει ετικέτες στις προβαλλόμενες διαφάνειες χρησιμοποιώντας την εφαρμογή για κινητά βγάζοντας μια φωτογραφία (αριστερά) η οποία αυτόματα μεταβιβάζεται στον server όπου αναγνωρίζεται (μέση) και μια απάντηση δίνεται σε έναν φυλλομετρητή WAP που ανοίγει αυτόματα (δεξιά) (Πηγή: Quack, Bay, Van Gool, 2008).....	12
Εικόνα 3. Γεωμετρική επαλήθευση με μια ομογραφία. Επάνω σειρά: αντιστοιχίσεις για μια εικόνα ερωτήματος με τη σωστή εικόνα της βάσης δεδομένων. Επάνω αριστερά: πριν από το φίλτρο ομογραφίας. Επάνω δεξιά: μετά από το φίλτρο ομογραφίας. Δεδομένου ότι η αντιστοιχία μεταξύ των φωτογραφικών διαφανειών είναι σωστή οι περισσότερες από τις υποθετικές αντιστοιχίες χαρακτηριστικών «επιζούν» του φίλτρου της ομογραφίας. Κάτω σειρά: ίδια εικόνα αντιστοιχίζεται με μια λάθος εικόνα στη βάση δεδομένων. Κάτω αριστερά: πολλές λανθασμένες υποθετικές αντιστοιχίες θα προέκυπταν χωρίς γεωμετρική επαλήθευση, σε ακραίες περιπτώσεις η καταμέτρησή τους μπορεί να είναι παρόμοια με ή υψηλότερη απ' ότι για το σωστό ζεύγος εικόνων. Κάτω δεξιά: όλες οι λανθασμένες αντιστοιχίες αφαιρούνται, μόνο τα χαρακτηριστικά γνωρίσματα από το (ορθό) αντιστοιχισμένο πλαίσιο «επιζεί» και η διάκριση του σωστού ζεύγους αυξάνεται αισθητά. (Πηγή: Quack, Bay, Van Gool, 2008).....	17
Εικόνα 4. Παραδείγματα εικόνων ερωτημάτων, από αριστερά προς τα δεξιά: α. Συνθέσεις κειμένου και εικόνας, β. λήψη από τις ποικίλες οπτικές γωνίες, γ. διαφορετικά επίπεδα ζουμ ή μικτό περιεχόμενο, δ. παράδειγμα όπου επιλέγεται μια συγκεκριμένη περιοχή μιας διαφάνειας ή περιέχει μεγάλη ποσότητα κειμένου. (Πηγή: Quack, Bay, Van Gool, 2008).....	18
Εικόνα 5. Λογισμικό πελατών για την εφαρμογή οδηγού πόλης: ο χρήστης σπάζει απότομα μια εικόνα, περιμένει μερικά δευτερόλεπτα, και επαναπροσανατολίζεται στην αντίστοιχη σελίδα Wikipedia. (Πηγή: Quack, Bay, Van Gool, 2008).....	20
Εικόνα 6. Εικόνες αποτελεσμάτων για την εφαρμογή του οδηγού πόλης, λεπτομέρειες για την ανάλυση εντός του κειμένου. (Πηγή: Quack, Bay, Van Gool, 2008).....	24
Εικόνα 7. Ρυθμός αναγνώρισης (αριστερά) και ταχύτητα ταύτισης (δεξιά) βασισμένη σε ακτίνες γύρω από την θέση του ερωτήματος. (Πηγή: Quack, Bay, Van Gool, 2008).....	26
Εικόνα 8. Τυπική περίπτωση χρήσης της epipolar geometry. Δύο φωτογραφικές μηχανές φωτογραφίζουν το ίδιο αντικείμενο από διαφορετικές οπτικές γωνίες. Η epipolar geometry, στη συνέχεια, περιγράφει τη σχέση μεταξύ των δύο όψεων του αποτελέσματος. (Πηγή: Wikipedia Contributors).....	35
Εικόνα 9. Epipolar geometry (Πηγή: Wikipedia Contributors).....	37
Εικόνα 10. Παράδειγμα epipolar geometry. Δύο φωτογραφικές μηχανές με τα αντίστοιχα σημεία εστίασής του OL and OR, παρατηρούν το σημείο P. Η προβολή του P στο καθένα από τα επίπεδα της εικόνας σημειώνεται με τα pL και pR. Τα σημεία EL και ER είναι οι epipoles. (Πηγή: Wikipedia Contributors).....	38
Εικόνα 11. Το σχήμα δείχνει πώς η φωτογραφική μηχανή b βλέπει στο επίπεδο σε απόσταση d.....	40

Ευρετήριο Πινάκων

Πίνακας 1. Σύνοψη δεικτών επιτυχίας αναγνώρισης των διαφανειών της βάσης δεδομένων (Πηγή: Quack, Bay, Van Gool, 2008).....	18
Πίνακας 2. Σύνοψη δεικτών επιτυχίας αναγνώρισης των διαφανειών της βάσης δεδομένων (Πηγή: Quack, Bay, Van Gool, 2008).....	25