RESEARCH REPORT

VTT-R-04735-11

Image based linking

Editor: Olli Nurmi, VTT

Authors: Atte Kortekangas, VTT
Janne Laine, VTT
Janne Pajukanta, VTT
Jouko Hyväkkä, VTT
Tim Kuula, VTT
Markus Koskela,Aalto University
Xi Chen, Aalto University

Confidentiality: Public

Aalto University
School of Science

Report's title	
Image based linking	
Customer, contact person, address TEKES, Ubicom-programme	Order reference
Project name Image based linking	Project number/Short name 35434/Kuvalinkitys
Author(s) Olli Nurmi	Pages 29
Keywords Image linking, content based information retrieval, hybrid media	Report identification code VTT-R-04735-11

Summary

This project developed technology for linking digital information to the real world based on image recognition and image matching. This technology was applied for magazines because it was estimated to have potential for commercial services.

In the basic user case the subscriber could get additional information by taking a photo with his mobile camera phone of the magazine page. The client software sends the preprocessed photo for image recognition and it receives links to the additional information.

The test users regarded technology as easy to use and quite reliable. However the true value for the system depends on the content that the magazine publisher provides to the subscriber.

The challenge with image linking is to find uses that solve a real problem and enable something fundamentally new, useful or uniquely entertaining.

Confidentiality	Public

Espoo 22.6.2011

Written by	Accepted by
Olli Nurmi, Team Leader	Caj Södergård, Technology Manager

VTT's contact address
Vuorimiehentie 3, 02044 VTT, Finland

Distribution
TEKES, Aller Jukaisut Oy, StoraEnso Oy, Viestintäalan tutkimussäätiö, Mobicode Oy, Anygraaf Oy, VTT

Preface

The “Image linking”-project has been carried out during 2009 – 2011 in the TEKES Ubicom programme. The research partners in this project have been VTT Media Technologies and Aalto University School of Science ICS Department.

This report summarizes the main findings in the project and suggests how to develop the system further.

The steering group of the project consisted of the following members:

Juha Kuokka (Aller Jukaisut Oy)
Juha Maijala (StoraEnso Oy)
Manu Setälä (TEKES)
Erkki Oja (TKK)
Helene Juhola (VTS)
Olli Ikäheimo (Mobicode Oy)
Hanna Muukka (Anygraaf Oy)
Olli Nurmi (VTT)
Caj Södergård (VTT), Project manager

The practical work has been carried out in the project group consisting following persons:

Olli Nurmi, VTT
Atte Kortekangas, VTT
Janne Laine, VTT
Janne Pajukanta, VTT
Jouko Hyväkkä, VTT
Timo Kuula, VTT
Markus Koskela, Aalto University
Xi Chen, Aalto University

The project group wants to thank the steering group members for the support, quid lines and comments regarding the development work and TEKES for supporting the project financially.

The writers hope that the image linking technology may find many new application areas and helps creating services where digital information is linked with the physical world.

Espoo 22.6.2011

Contents

1 Executive summary

Mobile phones with integrated digital cameras provide new ways to access digital information and services. Recognizing the digital image of the physical world additional information can be provided to the user. In this report “image linking” means linking digital information with physical objects by recognizing the digital image of the physical object.

Several mobile application based on image linking have been introduced covering application areas like visual search, tourism, navigation and leisure time. The best known commercial solutions are Google Goggles, kooaba, Amazon Visual Search and Nokia Point and Find.

Recent examples of further development of this technology are methods where the knowledge of a phone’s position (through GPS) and direction (using an electronic heading sensor or internal compass) are used to allow users to point their phones at objects they are interested in to gain information about them.

Research on image retrieval algorithms is very active. Image matching algorithm should be robust against variations in illumination, viewpoint, and scale. Mobile applications should work with stringent bandwidth, memory and computational requirements. This requires the optimisation of the performance and the memory usage.

This project focused to scalable image recognition methods and integration of the image linking technology into magazine publishers’ processes. In this project a system based on distributed software architecture was demonstrated where the magazine readers can use image linking to get additional information and the reference database can be updated dynamically whenever a new magazine issue is published.

Image recognition is calculation intensive task and it needs powerful systems especially when there are several simultaneous users. The system performance was increased by applying distributed computing system.

This project provided new knowledge and technology background for starting activities and developing services related to image linking technology.

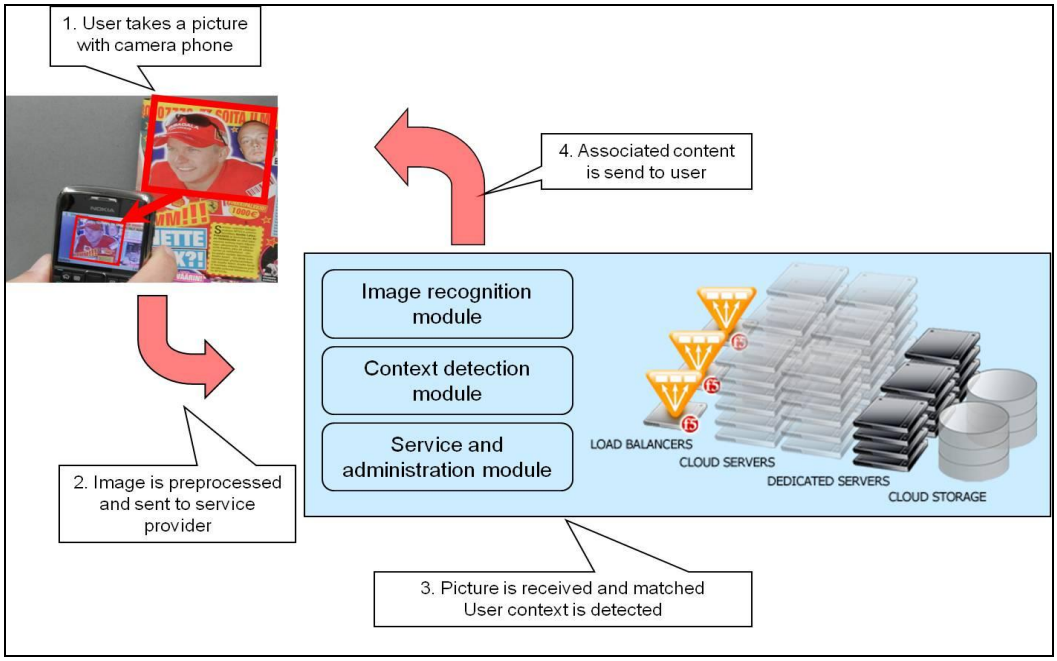

Figure 1. Basic user case for image linking. The magazine subscriber gets targeted information or services using digital image of the physical object and image recognition technology.

2 Introduction and goal of the research

There are several application areas for linking digital and physical world and the research and development work in the domain of linking technology is intense. Several means to carry out the linking are possible and each of them has its own characteristics. Optical codes and NFC-technology are some examples. The most advanced optical code generation is so called image based linking where the image itself acts as a link.

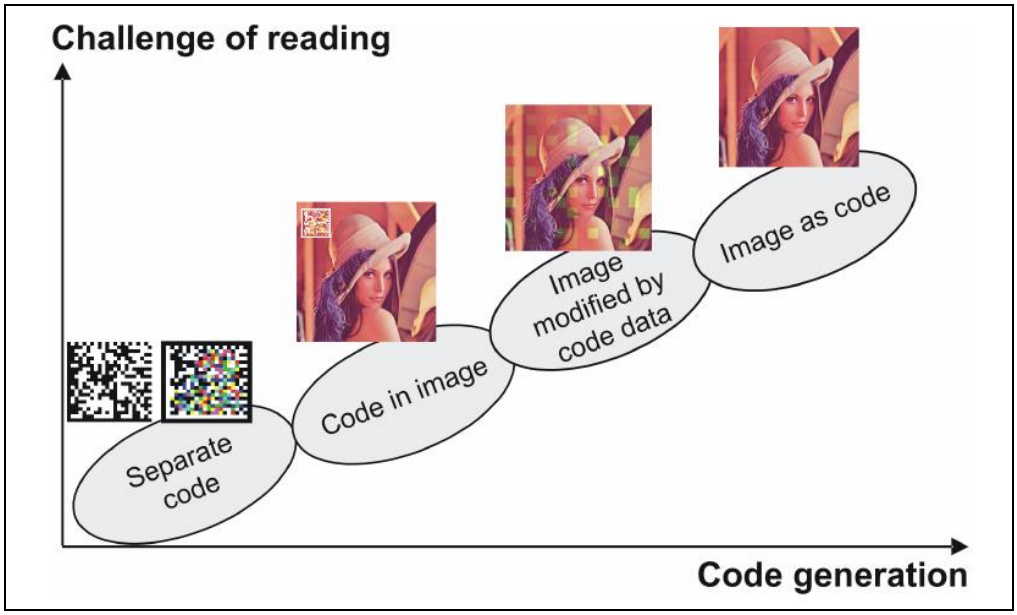

Figure 2.Different code generations in the domain of optical codes. [source: Aalto University]

The image based linking suits well to the situations where earlier optical code generations or NFC cannot be used because of problems in the product appearance, process technology or the long reading distance.

The image based linking is a generic enabler allowing linking of the physical world with the digital information. However this technology requires a-priori information stored in so called image feature database about the possible images matching to the captured image. This feature database can grow very large and the response time unpractical long if the possible matches or the size of the feature vector is not limited. One way to limit the search is to use GPS-information like in the Nokias' Point and Find-technology. This method fits well to the situations where the location is relevant and the number of possible matches in the same location is restricted.

There is active research to develop a robust feature calculation method and SIFT- and SURF-methods have gained popularity. SURF is further developed and simplified version of SIFT and it allows faster calculation speed. Another important development line is the compression of the image vectors allowing compression of the database. There are also several options to optimise the image search.

The key purpose of the project was to develop technology for the image based linking. Possible application areas cover outdoor advertising, magazine and newspaper advertising, tourist applications, shopping. Image based linking will not boom unless the applications and services with substantial added value for both the end-users and business can be developed.

Special effort was put on the further development of the feature vector calculation methods, the feature vector compression to decrease the size of the feature database, optimisation of the search from the database and methods to allow dynamic database updating. The target of the project was to demonstrate a system that is suitable for large feature databases.

This project plan didn't cover work that is required to integrate the system to the production processes of the media products such as magazines, newspapers or outdoor advertisements

3 Focus area of the project

It has been estimated in several reports that the amount of digital services will grow rapidly in the future. Image linking technology can be used to increase the access to these services and it may also provide new possibilities for new business models.

Several application areas were identified for the image recognition technologies covering visual search, media industry (magazines, newspapers and books), outdoor advertising, augmented reality applications, tourism, educational sector and ecommerce.

Hybrid media linking of the printed magazines was selected as the focus area of this project. In the magazine the visual appearance is very important and it is not

possible to use visual codes that may deteriorate the outlook. Also the attachment of NFC-tags is not possible during the high speed printing process.

It was estimated that the commercial application of the image linking technology will start from the media services and gradually find new application areas. This estimation was based on the following hypothesis:

- media has the means to inform the public about the possibilities of the image linking technology. For example the first pages of the magazines can be used to describe the image linking services.
- media companies have lot of digital content that cannot be used in printed magazines. Such as video material, extra images, music or social media services that cannot be printed.
- There is vital need for making the printed advertisements and editorial content interactive. Especially on the advertisement side, there are great economic benefits to be reached, if the advertiser can get feedback on his ads.

The use of the image linking technology was estimated to start from selected magazine brands and later applied in wider area. It may finally cover all of the publishers' magazines. On the other hand technology provided may offer image recognition service for many publishers.

Different user scenarios were illustrated and discussed. It was decided that this project will concentrate on the "reading" user scenarios with the following magazine-cases from Aller Publishing:

1. Katso- magazine: Image from TV-guide is used to link the user to the film trailer or social web services build around the TV-program. These services may include internet recording of the TV-program (for example Elisa Viihde-service).
2. 7-päivää magazine: Images from celebrity are used as a link to additional information/ pictures/social media services.
3. Miss Mix- magazine. Advertisements are used as a link to additional information regarding to the shopping.

4 Image detection approaches

The mobile visual searching system is divided into two parts: the mobile client and the server backend. The users take photos of interest using the client software, which then sends the image to the server for recognition. In this section, three image processing approaches intended for the server backend are briefly described.

4.1 Image matching using local features

The procedure for image matching in the server is based on matching local image features between the query image and the images in the database. The pairwise matching of local features is implemented using approximate nearest neighbor search, and the resulting correspondences are verified using a geometric

consistency check. Figure 3 shows an example matching between a query photograph and a magazine page in the database.

Figure 3 An example of successful matching. Left: the query photo, right: the matching magazine page.

In image matching based on local features, we need to detect reliable and repeatable local features and to describe the detected features by distinctive image descriptors. In our current system, as the feature detector we use a *fast multi-scale Hessian key point detector* and as the descriptor we apply *speeded-up robust features* (SURF) [1]. When the server receives a query image, it first detects the local features and extracts the corresponding descriptors, and then begins to search the most similar descriptors from the database.

In order to speed up the descriptor search from the database, we build the database index using *multiple randomized kd-trees* [2] which is a method for sub-linear indexing and which performs approximate nearest neighbor search. The candidate images are ranked according to the number of matched features. For a small number of the most promising candidate images, a validation step is used to check the spatial consistency of the matching features.

In our current application setup, new magazine are added to the database when they are published, and old ones are removed from the database when they become outdated. The standard kd-tree cannot, however, be modified after construction, and the time to build the tree is also relatively long when the dataset is large. Therefore, if new data is continuously added and old data is removed from the database, it is infeasible to constantly keep rebuilding the kd-trees. In order to handle the constant changes in a dynamic database, we use *multiple forests of randomized kd-trees* [3]. When a new magazine issue is published, it is added to the database as a new randomized kd-forest. Similarly, the outdated magazine issues are removed from the database by removing the corresponding forests from the database.

Each magazine page can generate a large number of descriptors, as there are large portions of text on many pages, which is both a common cause of wrong matches and increases the memory consumption of the image matching server. A common method for limiting the number of descriptors is to reduce the resolution of the

images. In this project, we cannot however reduce the resolution of the images too much as the system has to be able to recognize also small details from the magazine pages. For this reason, we have developed methods for reducing the number of descriptors without compromising the search accuracy [3].

4.1.1 Testing the image matching server

In the experiments reported in [3], we built a database with nine issues from three different magazines, each issue containing about 80–130 pages. The size of each page image was 771×1024 pixels, and a total of 6.5 million image descriptors were extracted. Three different descriptor pruning methods were applied before creating the image indices as described in the previous section.

For testing the recognition accuracy, we took a total of 300 query images from three issues, each from a different magazine. The images were taken by a Nokia E71 phone camera and resized to 640×480 pixels. The images were taken of such content that could be potentially interesting to the readers of the magazines and mostly contain only a small portion of whole page. Some of the query images are illustrated in Figure 4.

Figure 4 A random sample of the query images used in the experiment

The results of the experiments are shown in Figure 5, where we can observe that the matching accuracy of the clustering-based method is somewhat higher than the other two methods, especially when the size of the indices is only a small fraction of the whole database. In particular, the matching accuracy remains near 0.95 with only 18% of the whole database remaining. The average matching time is about 500 ms.

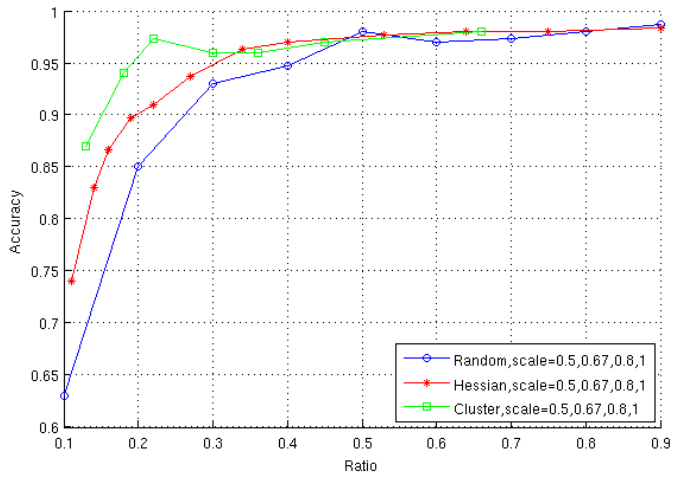

*Figure 5 A comparison of the recognition accuracies of the three descriptor pruning methods. "*Ratio*" on the x-axis refers to the fraction of the descriptors remaining on the database.*

We have also applied our image matching algorithm to other application areas. The matching algorithm was shown to perform well with a standard database of building images [3]. We have also experimented with images of CD covers, books, other products, outdoor landmarks, business cards, text documents, museum paintings, and video clips using the freely available Stanford Mobile Visual Search dataset [4]. The results show that our system can achieve the same level of accuracy that was reported in [4] and where exhaustive one-by-one matching was used.

4.2 Development of an elastic template matching method

4.2.1 Introduction

Elastic template matching was studied as a possible alternative approach for picture matching. Matching is done in terms of i) a *pixel-wise distance criterion*, which combines matching in *colour space* and *geometry*; ii) organisation of the reference image into a hierarchy of adjacent sub-images progressively merged at ascending hierarchy levels, iii) transformation of discrepancy of pixel match to discrepancy of subarea match while constructing the hierarchy and iv) algorithms for computing to the lowest level distance criterion and their transformed versions on iterated levels of the ascending hierarchy. We address this approach as *hierarchical template matching* below, referring to the structure of the reference template as well as the overall structural organisation of the matching procedure.

Matching itself is in the simplest case equivalent of computing the match criterion with the method for all reference candidates, selecting the closest one and assigning the output according to the result. The main methodological approach of the *Image Link Project* (“Kuvalinkitys”, in Finnish) is a variation of a matching algorithm based on SURF descriptors and stratified matching of robust spatial features computed for the target picture (to be identified) and matched with a collected database of similar features computed for the set of reference pictures.

The goal of the elastic matching approach was to develop and realise a method for recognising the target, a specific page of a magazine, in snapshots taken by a picture phone based on an approach related with traditional template matching, but avoiding the most well-known problem with the approach, i.e. sensitivity of recognition to perturbation of target. Traditional template matching is designed to compensate for two-dimensional positioning errors like sliding the co-planar target image with the horizontal and vertical degrees of freedom in otherwise strictly fixed 3-D alignment with the detector plane. Then, the images are compared at positions specified at the intersection of a line orthogonal the two parallel planes.

However, when a target picture is being photographed with a handheld camera and only this result is being compared with a reference image for deciding the likelihood of target picture matching the reference, such two-dimensional sliding mechanisms is not sufficient nor accurate except for an exceptional case of fully aligned target with the reference at a unity distance (so that target and reference images get mapped to the same scale). Even so, camera optics is not perfect so that e.g. aberrations will affect matching accuracy.

The hierarchical template matching procedure can be modelled as fitting a 6-D geometric mapping with the parameters (x,y,z,φ_X,φ_X,φ_X) for best possible match between hypothetic target content (image) and the a reference. In practice, the match should be done vice versa, i.e. transforming a reference to all possible realisations within the appropriate scope of the six basic parameters and their active scope. Figure 6 depicts the coordinate system being referred to in the following discussion.

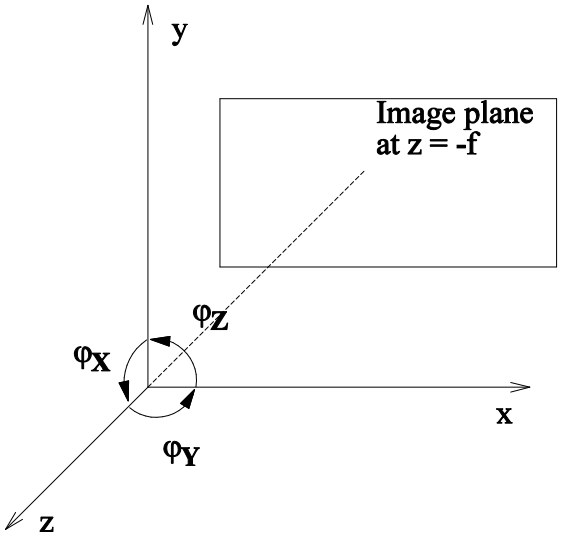

Figure 6. Coordinate system and denotation for imaging setup; target objects at negative z.

4.2.2 Algorithmic principle

Figure illustrates the principle of merging lower levels match cost to higher lever cost with hierarchical template matching. The template is subdivided by a systematic pyramidal division to a refined hierarchy of sub-regions. We have implemented a scheme, where single pixels make the bottom level of division and the division is constructed so that on the next higher level sub-regions at row index positions $(2i, 2i+1)$ and column index positions $(2j, 2j+1)$, i=0,.., j=0,… are merged for the next higher level. The associated costs are merged as well.

The idea is to select a weight function for match discrepancy such that the cost C_k of discrepancy for the merged level is expressed as a sum

$$C_k = C_{1,1,k-1} + C_{1,2,k-1} + C_{2,1,k-1} + C_{2,2,k-1} + C_{0,k}$$

of discrepancies $C_{i,j,k-1}$,where $i,j = 1,2$ are inherited from the lower level and match discrepancy on the merged level $C_{0,k}$ is not singularly minimised with $C_{0,k}\equiv 0$. Essentially, the weight function should be such that it is generally more cost effective to describe discrepancy in terms of discrepancy explained on the higher level and leave minimal amount on discrepancy to be explained on the lower level (when the higher level is unable to model such remaining discrepancy).

We have used versions of weighted Euclidean distances as such measure. Discrepancies can be imagined as elastic strings or rubber bands, whose tension is a measure the cost of discrepancy. Tension values are initialised from pixel difference in colour space. At the merging stage the strings are bound together with the pixel matrix resulting from the merger. The matrix is able to move at later stages of the algorithm, but only as a rigid aggregate and along and aligned with the grid positions of the original pixel array. The merged sub-area entity is tied with this aggregate with a string having lower string constant than the united string constants of the strings from the lower level. Of course, the alignment of the lower level strings would affect the magnitude relation bringing more complexity to selection and properties of the iterated criterion of discrepancy.

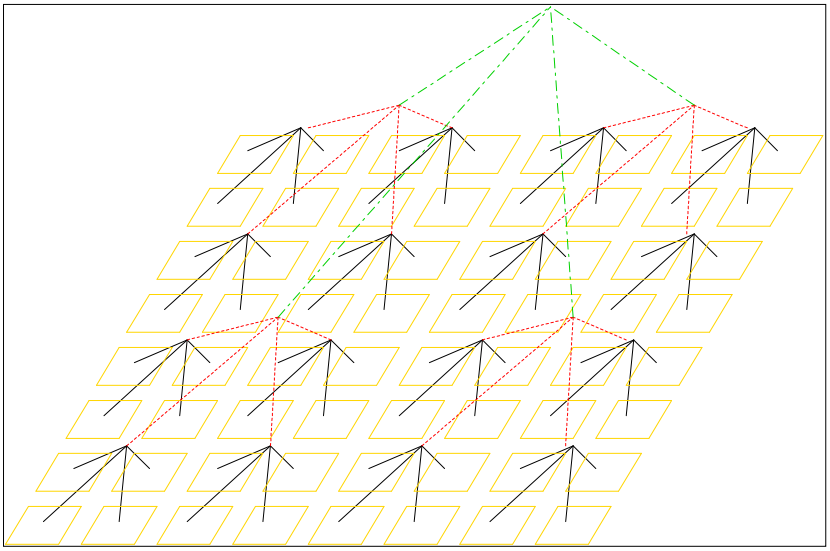

Figure 7. Hierarchical merging of match costs.

An essential part of the algorithm is the calculation of cost of perturbation of recently rigidified sub-region aggregates. This should be done primarily along the two degrees of freedom translation along the image plane; secondly, along rotation around a normal of the image plane and possibly along a few variations of scale, somewhat related with the distance of camera from target in space. Tilt angles have much less effect on matching and go as penalty to the match criterion, At this stage, only the two translational degrees of freedom are considered at this stage. The complexity of matching stage will be linear with respect to the number of pixel positions, if the criterion is a norm (especially, conforms to the triangle inequality and positive scalability).

4.2.3 Variations of the matching scheme tested

Most time is spent with the matching algorithm with manipulation of data-structures for each sub-region. We assume that the target area in the snapshot is a portion of the page. Then, it is most natural to require that the snapshot will match a sub-region of the original reference page image, i.e. each pixel of the snapshot should be transformed to some pixel of a reference page image and there is a cost dependent on the data and the way in is being modelled. This will lead to a principle, where the hierarchy is built on the page reference image, but the target

snapshot is used entirely as a flat arrays “needing a valid explanation” by the set of reference page images.

In principle, there has to be an aggregate storage vector for each pixel position of the target snapshot image on each level the template hierarchy. As a worst-case scenario this could lead to space complexity of the order of product of the pixels with the target and reference being matched, but with careful ordering of the hierarchy merging procedure the order can be dropped to the product of the number of the pixels in the target image times the number of hierarchy levels, a logarithm of the number of pixels with the page reference. However, the lower levels can be discarded as soon as the corresponding sub-regions are merged, but the total number of aggregate cells needed is still proportional to the number of pixels of the reference image.

4.2.4 Experiments carried out

Preliminary experiments were carried out on three development versions of the hierarchical matching scheme. The scheme was realised as a computer programme and tested with a set of snapshots taken by a camera phone (*Nokia N900*) from pages of a magazine. Figure contains examples of images being experimented with. Matching was done with reduced accuracy (image size 51x38 pixels, number of rows and columns reduced 2.5 per cent of original), but still the processing speed was not sufficient (only a few pairwise matches could be completed per one minute of CPU time).

Figure 8. Examples of image shots on page content being matched with original page data.

Initially, no normalisation of any of the images was done and only Euclidean cost of translation of merged sub-regions was done on each level. It was soon observed that the average density of images had considerable accuracy of matching besides finer detail in the image.

As the second algorithmic modification, both images (the target snapshot image and the synthesised page image) were normalised with respect to averages of the colour channels and their variances. This appeared to improve recognition accuracy, but a fraction of the images were still recognised incorrectly. The scheme also has obvious shortcoming, i.e. the reference page image will get normalised according to overall image contents, not according to the target image being recognized and decided to be the best match.

Therefore, a third version, where normalisation is carried along the process on building the template hierarchy specific to each individual position, local

deformation and level on construction of hierarchy. Considerable new complexity and new degrees of freedom and need for parameter choice were brought in with this line of development. Experimentation was done with some ad hoc selection of system parameters, but no obvious improvement was observed compared with the second stage of development. This is no indication that such improvement could not be achieved, but the new degrees of freedom make the algorithm so much more complex that considerable new research with them should be done both theoretically and numerically with controlled artificial data, before it makes sense to carry out further experimentation with real page image data. There have not been sufficient resources to do such extra work so far.

Figure illustrates the values of the match at the topmost level of the matching hierarchy for the case of good match. The vertical and horizontal pairs of dotted lines denote the boundary for the sub-region for result array, where the target snapshot fits entirely within the boundaries of the reference page image of the magazine. The dark area in the centre is associated with a low value of the match criterion, i.e. good match. Since the global criterion tolerates local translation of sub-regions with low penalty, if the regions can be combined at a higher level, the lower the higher the merging point can be brought, the minimum is smooth with respect to translation along the image plane. The same applies for slight rotation and change of scale (i.e. zooming or translation orthogonal to the image plane), but then the discrepancy is distributed over all levels of the hierarchy resulting in higher penalty, as well.

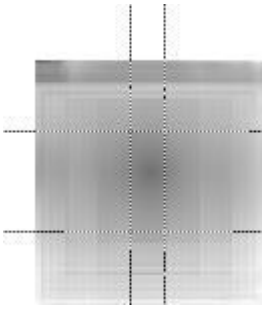

Figure 9. Example of location match image being matched with original page data (Dark means good match; implicit image crop limits denoted by lines and central rectangular area).

4.2.5 Discussion and concluding remarks

Speed has to been a problem for making the approach and method practical. Speed is likely to be improved by an optimised implementation, but memory allocation and the sub-region merging process remain as sources of heavy computational load. Even smaller images might possibly be used, but some accuracy is already lost at the current level of subsampling.

More efficient strategies could possibly be developed e.g. based on organising reference images into clusters and doing a stratified search progressing from larger groups of reference images to smaller and smaller ones, but we have not yet pursued such approaches. It is obvious that a rather large set of images is needed

before the possible benefits could be demonstrated apart from being studied experimentally.

Especially the approach, where image normalisation is aligned with the construction of the matching hierarchy need considerable more theoretical and numerical analysis with synthetically generated test material to justify the selection of most critical parameters like the mutual weighting of geometric and colour-photogrammetric distance criteria as well as the details of the weights of the region merging procedure at all levels in general and the mutual components of the total criterion in special.

4.3 Automatic blur detection

The use of a blur metric in the task of automatic detection of very blurry query images in image recognition applications was studied. The motivation for this work was to detect images that are so blurry (due to focusing errors, or motion blur resulting from camera movement during exposure, for instance) that they cannot be reliably recognized by the image recognition algorithm. This detection functionality enables an image recognition application to give appropriate feedback to the user about image quality problems.

The blur metric uses an algorithm based on the re-blurring of the image to calculate an index for the magnitude of blurriness in the image. The image analysis algorithm is based on blur metrics previously proposed in literature, modified to make it less susceptible to noise in the image and better suited to estimation of a wide range of blur magnitudes. In the application area being considered, the metric has the further desirable properties of being relatively insensitive to image content (the blur estimate does not change too much with changing image content) , not requiring a reference image for blur estimation, and having an approximately linear behaviour over a wide blur range and an overall monotonic characteristic. Experimental results showed that a meaningful threshold could be set for the blur metric value above which the images in the test set could not be reliably recognized, separating these images from the class of images that were generally correctly recognized.

5 System description

5.1 System architecture

From the user point of view the system architecture can be considered as a quite ordinary web service accessible with any kind of relatively modern mobile phone equipped with a camera and an internet connection [5]. The system architecture is outlined in Figure 10.

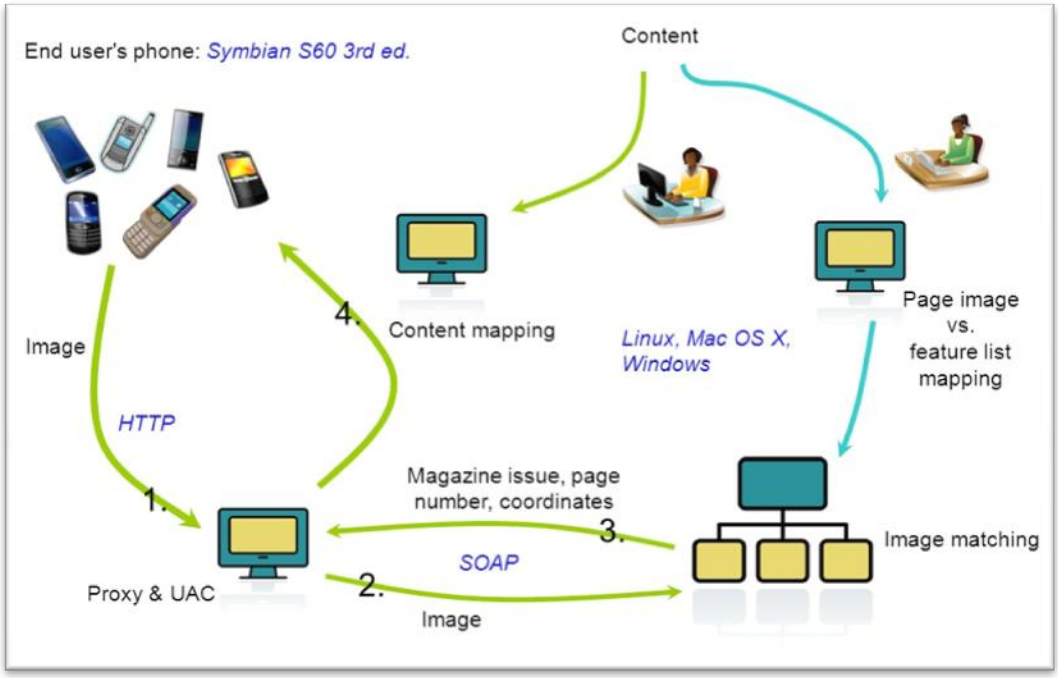

Figure 10. The system architecture.

The image recognition and the content retrieval processes start when the user takes a photo of a magazine page with a mobile phone application and sends it for instance as a multimedia message (MMS) to the recognition system. Preferably, the image transfer is handled with a dedicated application which can add more features to the service than just the phone's own camera application. Additional data such as the user preferences and contextual information can also be sent if the data can be utilized and if the application supports it. The user's privacy and legal issues must naturally be taken into consideration when collecting the data from the user.

The query is first processed in a proxy server which can handle both anonymous and authenticated users. The data from the query is filtered and the image is forwarded to an appropriate image matching service using the SOAP protocol. The data filtering removes the unnecessary data and adds all the known meta-data so that image matching service can narrow the search as much as possible.

The image matching block returns the matching results, consisting of the matching magazine issue, page number, x and y coordinates, and the transformation matrix. The results are mapped to the related content, such as additional information, news, and videos, provided by the publisher. Links to the related content are then returned to the user's mobile phone.

Before the image matching block can process the query images, it requires access to all the supported magazine pages from the magazine publisher. The image matching database has to be constantly updated, also by removing out-dated magazine pages, for example when a certain magazine campaign period is closed. Moreover, the additional content visible to the end users is even more temporary and volatile. The content provider therefore has to be able to add frequent updates to the content while the magazine page information located at the image matching service remains unchanged.

5.2 End user's mobile client

A dedicated mobile phone application was created in order to make it easier for the end user to use the image linking system. It should be noted that the whole mobile ecosystem went through large changes while the project was carried out. Some common features nowadays, such as image upload and sharing were available only in limited number of suitable mobile devices when the project started.

The main tasks for the client application are:

- function as user interface
- utilize phone's resources; camera, display, loudspeaker, keypad, network
- handle communication with the proxy server

The client was implemented in Symbian C++ language and the main target was set to Nokia E71 phone (S60 3rd edition, feature pack 1). One minor generation back and one forward were taken into consideration so the whole S60 3rd edition was covered, i.e. S60 3rd ed., 3rd edition FP1 and 3rd edition FP2. There are dozens of different S60 3rd edition mobile phones but nowadays they have disappeared or, at least they are disappearing rapidly from the market.

The functionality was successfully tested with the following Nokia phone models;

- N80, N92, N93 (S60 3rd edition)
- 6110 Navigator, E71, E90 Communicator, N95 (S60 3rd edition FP 1)
- E72, E75, N96 (S60 3rd edition Feature Pack 2)

The application can handle functionalities such as auto-focus and dynamic screen rotations if the phone provides them. The same binary installation file is used for all the supported phone models and the installation procedure is more or less typical Symbian installation procedure. For instance, if the installation file is sent to a phone using Bluetooth the installation starts automatically but it is not finished unless the user accepts the installation and grants privileges such as network access to the application.

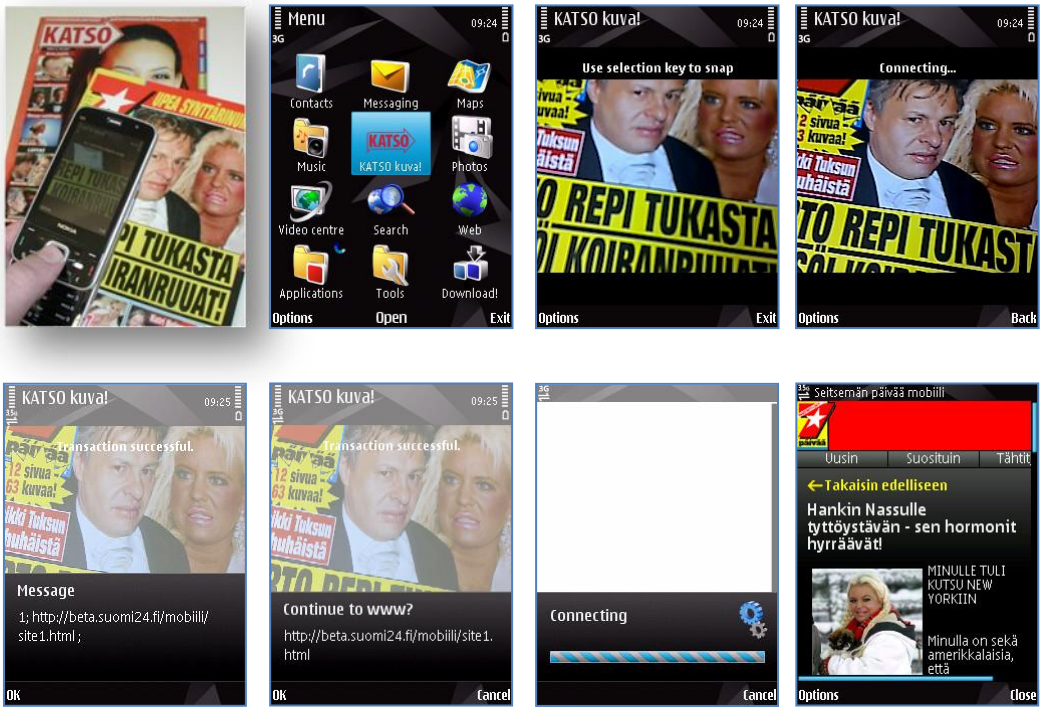

Figure 11. Usage example and some screenshots from the application running on Nokia N96.

5.3 Servers and load balancing

The system consists of frontend and calculation server and it runs on hardware containing 2 processors a'6 cores. The interface between the servers has been implemented with SOAP - Simple Object Access Protocol allowing the applications exchange information over HTTP over existing firewalls and proxies.

Distributed system is used for practical reasons. For example, it was more cost-efficient to obtain the desired level of performance by using a cluster of several low-end computers, in comparison with a single high-end computer. A distributed system can be more reliable than a non-distributed system, as there is no single point of failure. Moreover, a distributed system may be easier to expand and manage than a monolithic uniprocessor system.

In this system each processor has its own private memory (distributed memory). Information is exchanged by passing messages between the processors.

A simple load balancing method is applied to distribute workload across computer cluster to achieve optimal resource utilization, maximize throughput, minimize response time, and avoid overload. The method is to use for every CPU cores 12 instances of the software, instead of a single instance, which increases reliability through redundancy.

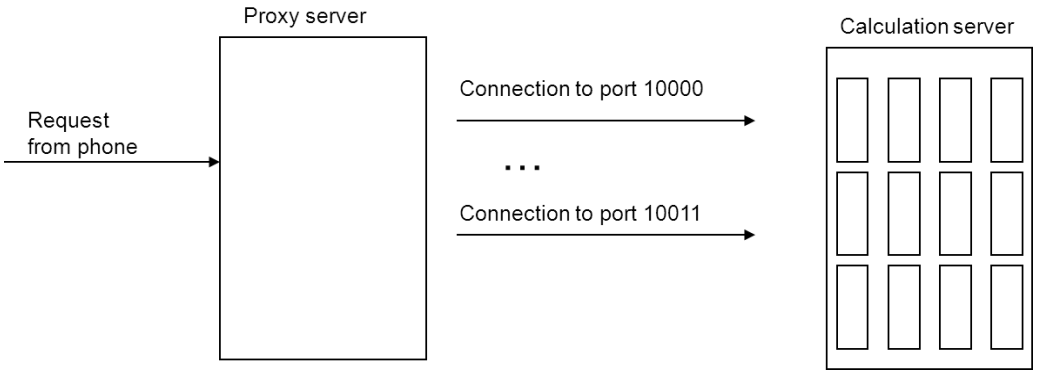

Figure 12.Load balancing method used to allow maximal usage of the 12 cores of the calculation server.

The load balancer is a software program that is listening on the port where external clients connect to access services. The load balancer forwards requests to one of the "backend" servers, which replies to the load balancer. This allows the load balancer to reply to the client without the client ever knowing about the internal separation of functions. It also prevents clients from contacting backend servers directly, which may have security benefits by hiding the structure of the internal network and preventing attacks on the kernel's network stack or unrelated services running on other ports.

This load balancing approach allows the maximal usage of current hardware. It has also some flexibility even if some of the instances jam. The drawback of this approach is the memory shortage if lots of magazine issues are active. On the other hand it is not easy to update the system.

More sophisticated load balancers may take into account additional factors, such as a server's reported load, recent response times, up/down status (determined by a monitoring poll of some kind), number of active connections, geographic location, capabilities, or how much traffic it has recently been assigned. High-performance systems may use multiple layers of load balancing.

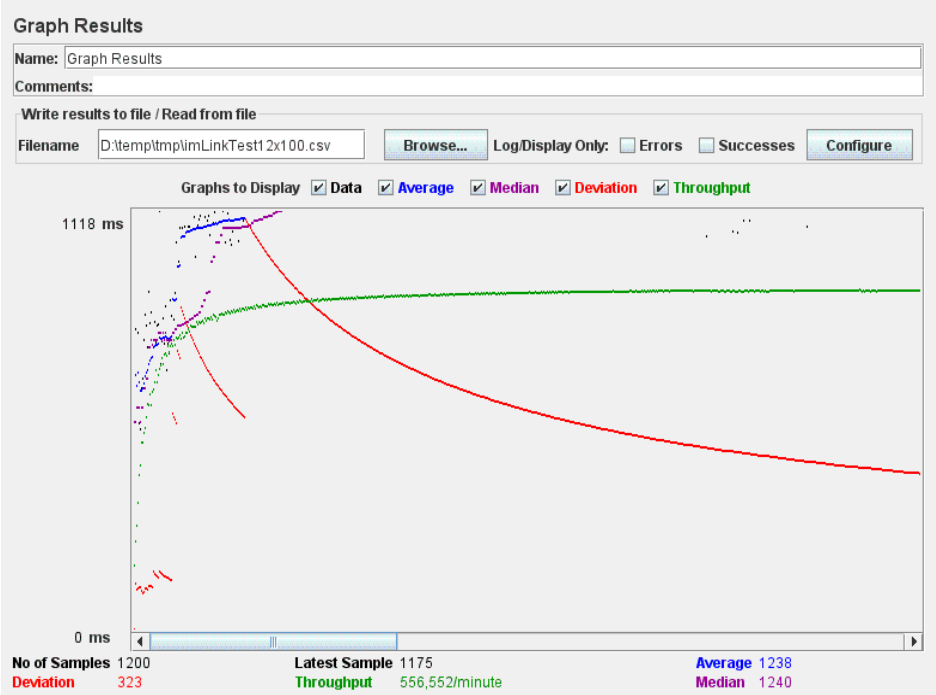

Figure 13. Result graph of system load test of 12 users.

A simple load test was carried out to find out system responsiveness. In the test to the system were sent in total 1200 requests by 12 test users at same time. Average response time were 1,2 seconds and throughput were 556 requests per minute.

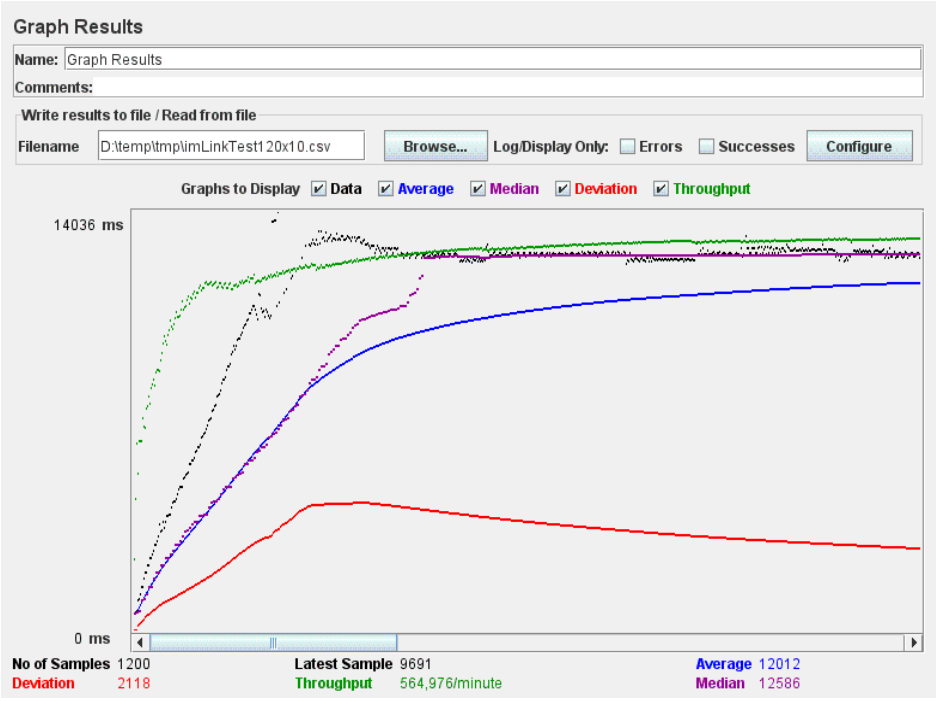

Figure 14. Result graph of system load test of 120 users.

When the system were loaded by 120 concurrent users, throughput were still 564 requests per minute, but average response time rose to 12 seconds.

5.4 Image linking process

Image linking process start with the publisher, where after the magazine PDF's and additional content are created the PDF pages are sent to VTT server. These PDF-files are converted to 300 dpi PNG images with Linux tools using automated scripts.

For simplicity the linking information is currently "embedded" in the filename of the page and which restricts the amount of links per page to one. In case of multiple links the user is supposed to select the interesting link.

Basically the system is capable to handle multiple links per page because the system recognizes also the coordinates of the search image on the page. These coordinates could be mapped to different links by smart mapping algorithm.

The next step in the process is key point extraction and removal of irrelevant key points with image recognition software. After this is the index creation and finally the feature database is updated so that the 3 last issues are active.

6 User testing

Two user tests were conducted in order to collect feedback on use and usability of the system. The chosen methods for user tests were group interview and usability questionnaire.

The target group was the subscribers of *7 päivää* magazine published by Aller Media Oy. Altogether seven persons participated in the tests. The both tests were held in the same day (25 May 2011) successively: three persons participated in the first test and four in the second test.

The actual test included three main phases:

1. Group interview: Background information; participants' use of magazines and use of smart phones
2. System test; filling in the usability questionnaire
3. Group interview: Evaluation and further ideas

The test started and ended with the group interview. The first interview was related to use of magazines (mainly *7 päivää*) and smart phones, and the latter included evaluation of the system and further ideas for development. During the second phase (System test), the idea of image based linking and use of the system were introduced to the participants. After the introduction all participants were given Nokia E71 phones with the client application and two different issues of *7 päivää* magazine. They were asked to use the system and explore the contents individually at their own pace. The "active" pages in the magazines were clearly marked and provided text, pictures and audio-visual materials (short video clips, music files) into the phones. Finally, the participants filled in the usability questionnaire.

6.1 The participants

Seven subscribers of *7 päivää* magazine participated in the tests. Five of the participants were female and two were male, and they were aged 24 – 38, five of them under 30. All lived in Greater Helsinki, six in Helsinki and one in Vantaa. Four of the participants had university education, and only one had no other education than comprehensive school. The overall picture of the participants was that they were young and educated people living in Helsinki.

6.1.1 Reading of *7 päivää* magazine

The participants read *7 päivää* magazine because of the short stories that are easy and fast to read. Reading it is a good way to separate work from leisure; at work one needs sometimes to read long and heavy texts, thus *7 päivää* is a good counterbalance to work-related and other more profound reading. The magazine includes stories about interesting and familiar persons, and provides topics for coffee break discussions. Hollywood stories and TV-guide are also among the favourite contents. Reading serves curiosity as well: one does not necessarily look for anything special from the magazine, reading it is just a nice way to spend time and get information about what is going on in the world. Some mentioned that the magazine is subscribed mainly because of children or spouse.

The interviewees read *7 päivää* whenever the situation and context is suitable, and when they have spare time. Some read it briefly right after it has been delivered and probably return to it later on. A good context for reading longer periods is when using public transport. Reading can also be tied to breaks in daily routines (morning coffee, lunch break, coffee break etc.). Sometimes it may replace watching of TV, especially when there isn't anything interesting on TV. For parents, reading *7 päivää* is a way to relax during their private time, after the kids have gone to bed.

6.1.2 Use of smart phones

Three of the participants did not have smart phones (with internet access) and used their phones mainly only for talking and sending text messages. Two of the participants with smart phones were currently active users of mobile services and applications, such as email, Facebook, Microsoft Office, map service and mobile newspapers. The rest two smart phone owners used their phones mainly for talking and SMS, and occasionally used applications and services. One of them mentioned that she would like to have a phone with touch screen in order to play mobile games (Angry Birds). With that phone she would use other applications and services more frequently as well (her current phone had poor usability).

6.2 Results from usability questionnaires

Since the number of participants was low (N=7), the results based on statistical analysis cannot be generalised and only represent the opinions of the test groups. The questionnaire used in the tests was based on QUIS (Questionnaire for User Interface Satisfaction) by Chin et al. [9]. QUIS measures user's subjective rating of the human-computer interface. The original QUIS was modified for this study.

As an overall reaction to the system, the result (average) was **7,29** with scale being 0 = 'terrible' to 9 = 'wonderful'. Variation in answers in this question was very low with six 7's and one 9. Furthermore, with scale being 0 = 'difficult' to 9 = 'easy', the overall reaction to the system resulted **7,86** with variation between 6-9. The lowest overall reaction score was **6,0** (variation 4-7) with scale 0 = 'frustrating' to 9 = 'satisfying'.

Three questions related to learning how to use the system gave the highest scores in the whole questionnaire with variation in answers between 7-9:

Scale: 0 = 'difficult'; 9 = 'easy'

Question	**Average**
Learning to operate the system	8,29
Exploring new features by trial and error	8,29
Remembering necessary operations	8,14

The questions with lowest scores were related to sound quality, help messages (system guidance) and system reliability as follows:

Scale: 0 = 'terrible'; 9 = 'wonderful'

Video sound quality	6,57
Music (audio clips) sound quality	6,67

Scale: 0 = 'unhelpful'; 9 = 'helpful'

Help messages on the screen	6,0

Note! In the question about help messages the variation in answers was very high, between 3-9.

Scale: 0 = 'unreliable'; 9 = 'reliable'

System reliability	6,57

When asked whether the participants would recommend the image based linking system to their friends, three (3) would definitely recommend, three (3) would recommend with doubts, and one (1) would not recommend.

The one person, who would not recommend the system for her friends, explained her opinion in the questionnaire as follows:

> *"It is enough for me to read the paper version of 7 päivää, so it's hard to say if anyone else would be interested in image based linking service. I wouldn't pay for it – even though I'm curious, I'm not that curious that I would bother using the application (despite the fact that it was easy to use)." (Woman, 26)*

6.3 Results from group interviews – evaluation and further ideas

6.3.1 Usability

Generally, the system was easy to use according to participants. It was also considered as interesting and handy way to get extra information directly into the

phone. However, using computer was still seen as better way to look for similar content, because computer provides faster internet connection. The phone could be used when computer is not at hand. It was also mentioned that reading *7 päivää* is purposefully done *without* the phone, since the idea is to "*leave all the thoughts behind*" and relax with the printed magazine. The usability and features of the phone could affect the motivation for use. The size of the phone screen (Nokia E71) was basically considered as large enough, but some thought it was too small for watching videos.

Downloading the client application was not included in the system test. It was mentioned that the one of biggest reasons for not using the system could be the lack of motivation for downloading it.

6.3.2 Content and pricing

Video clips were seen as a nice thing when combined with reading the magazine. The service with list of songs (audio files) was considered nice as well. It was emphasised, that image based linking must provide some *additional* content compared with the content in the magazine. The same content would not be interesting.

The content via image based linking could be consumed in boring situations with lots of spare time and nothing much to do. These could be for example long journeys in a car or when standing in line. On the other hand, the use could be fun in social situations as well: It could be used for "*laughing together with friends*", as stated by a participant.

It was also mentioned that the content would not necessarily be consumed in public places, since it could be embarrassing to watch videos about certain things and celebrities. More private places, such as home, would be more suitable for audio-visual contents thus. Additionally, one participant said that it would be important to know what kind of content is to be expected. Otherwise the content might even shock.

The participants were not eager to pay for the use of the system and extra content. It was mentioned that some two Euros additional price for the magazine would be too much. Still, when choosing from many similar magazines, image based linking service could affect the decision positively.

Handing over personal information to the advertiser in order to use the system seemed to be a very unwanted idea. Some mentioned that they would rather pay for the use than give information to the advertiser. However, some thought that many advertisers already have the information, so it wouldn't be a big deal.

6.3.3 Further ideas

"I think this is a very potential thing…it probably works, if it is reasonably priced and it offers content interesting enough." (Woman, 27)

It was generally thought that the service would be suitable for those who use a lot of applications for smart phones. Mainly young people and digital natives

were seen as the most potential groups. Further ideas for development and use of the service are listed below:

- The service could provide background information about the stories in the magazine
- The content should be exclusive in a way that it can *only* be accessed via image based linking
- Additional video clip could be available related to a celebrity doing something silly
- Movie trailers could be linked to movie reviews
- Nude pictures could be available at extra price
- For children, content related to pets and pop stars

6.4 Conclusion

Based on the user tests, the client application is generally easy to use. It is especially easy to learn how to operate the system, which might lower the threshold of implementation among users. The reliability of the system is considered as slightly lower when compared with ease of use.

The system has some novelty value among users and the overall reaction to the service is positive. However, it seems to be difficult for the user to think what the true added value of the service is when compared with printed magazine and magazine website. In many situations, computers provide a better access to internet and desirable contents. Digital natives are seen as the most potential users of the service.

The users are not eager to pay for the use of the service. In some situations, the availability of the service in a magazine might affect the decision to buy positively.

7 Summary

The work in image linking project resulted content-based image retrieval system where computer vision techniques are applied to the image searching and matching in large database containing magazines pages. As a result the mobile client gets additional information related to a certain magazine page.

This system has some specific features because it is adapted especially to printed magazines and the features in the database are extracted from the pdf-files of the magazine pages. The system is capable to provide additional information even though the search image contains only part of the page such as interesting article or advertisement. The other key feature of the system is the support for dynamic database, so the addition and removal of magazine issues is quite straightforward from the feature database. This is essential in magazine publishing process where new issues should replace the out-dated ones.

The image recognition software runs on scalable calculation server and it is based on SURF-method which allows search image to be partially covered and the result doesn’t change as function of scale, orientation, geometric distortions or

illumination. The amount of key points are reduced by clustering them in useful and non-useful and removing the non-useful from the database.

Figure 14. The search image and the page PDF. The blue lines illustrate matched key points.

The system is capable to provide several links on one magazine page, because the image recognition software provides the coordinates of the middle point of the search image in the magazine page. Using these coordinates allows concluding which part of the page is interested to the user and provide links to the specific information related to that part of the page.

Image analysis algorithms were developed to detect blurriness in query images, in order to enable feedback to the user about image quality problems in the cases where out-of-focus or motion blur of the query image makes reliable image recognition impossible.

One quite interesting use scenario is to use not only the mobile client to view the additional information but also some other media devices like internet TV. This could be useful especially in viewing additional video material related to the magazine page.

8 Further development areas

The work in the project could be continued to develop both the image recognition methods, calculation system and mobile client.

8.1 Image recognition methods

Currently the system adopts a full representation method where the local features are directly used for in matching in a pairwise manner and in indexing the databases. The accuracy of the retrieval is remarkably high, but with certain trade-

offs in scalability and memory consumption. Another alternative matching strategy uses the so-called bag-of-visual-words paradigm [6], in which the standard image matching techniques are based on inverted files, min-hashing, or local sensitive hashing (LSH) [7]. This paradigm is typically used with large image databases to retrieve multiple images with not as strict requirements for retrieval accuracy. For the image linking application, the method would enable the use of very large databases, but further development is required to preserve the current high accuracy in the matching results.

Another way to enlarge the scale of the database is to replace the current SURF descriptor with some other descriptor. SURF is very robust and performs well in matching, but the descriptor is a 64-dimensional vector, which implies that the nearest-neighbor distance calculations are relatively slow. There are novel descriptors proposed in the literature such as the binary robust independent elementary features (BRIEF) [8], which can alleviate the deficiencies of SURF in memory and computation costs.

The current method for performing geometrical validation assumes a homography or a projective transformation between the input and database images. This is sufficient for many applications, including hydrid media linking of printed magazines, but may be a limitation to some other use cases as it intrinsically assumes that the target object is planar. A more flexible geometrical validation check might be required for some applications.

On a technical level, the system will be converted to use the latest version of the OpenCV library, as it supports a wider selection of both detectors and descriptors of image key points. This enables the easy replacement of the SURF descriptors, which may have a positive effect on the performance on some applications, even though the SURF descriptors have turned out to be very robust and reliable.

8.2 System development and mobile client

System scalability and integration to cloud services should be developed further by adding the necessary administrative tools. Another important development area is multithreading which is a widespread programming and execution model that allows multiple threads to exist within the context of a single process. These threads share the process' resources but are able to execute independently.

However, perhaps the most interesting application of the technology is when it is applied to a single process to enable parallel execution on a multiprocessor system.

This advantage of a multithreaded program allows it to operate faster on computer systems that have multiple CPUs, CPUs with multiple cores, or across a cluster of machines — because the threads of the program naturally lend themselves to truly concurrent execution.

Another advantage of multithreading, even for single-CPU systems, is the ability for an application to remain responsive to input. In a single-threaded program, if the main execution thread blocks on a long running task, the entire application can appear to freeze. By moving such long running tasks to a worker thread that runs

concurrently with the main execution thread, it is possible for the application to remain responsive to user input while executing tasks in the background.

The collection of the use statistics and analysis should be based on commercial systems. The area that might be interesting for the publisher is the use and search profiles and the distribution of the time when the service is used.

A special are in the image linking of magazine pages is the administration of the multiple links in the same magazine pages. This area needs special tools that should be integrated to the PDF-file production. Possibly the information could be integrated into the PDF files to avoid errors.

Developing blur detection further it would be possible to alarm the user about too blurry search image before it is send to recognition server. Another development area is the support for other mobile operation systems. Android, iOs and Windows are currently very promising.

9 Conclusions

The technology developed in this project for image matching using local features has turned out to be robust, flexible, and scalable to relatively large databases while retaining a high matching accuracy.

The system maintains a dynamic image database, which is constantly begin updated by adding new issues and removing old ones, without the need for rebuilding the index due to the use of separate randomized kd-forests for each magazine. A second characteristic of the system is the high matching accuracy even with query images that are often of poor quality. Third, the pages of magazines often contain large numbers of useless descriptors mainly due to presence of text on the pages. For this reason, we have developed methods to reduce the number of descriptors, even up to 80% of them, while preserving the matching accuracy.

Experimental algorithms and software realisations were developed for elastic hierarchical template matching. The approach was intended as a supplementary or alternative approach for the main (SURF-related) approach with possible special cases needing further arsenal of processing tools. Preliminary experiments with real images indicated that appropriate image normalisation is essential for achieving accurate matching.

A unified approach integrating image normalisation with construction of the matching hierarchy was realised, but the selection of the more complex set process parameters proved to be critical and this case. Further analysis of the selection of these parameters is needed in order to assess the feasibility of the unified approach for the real matching problem.

Based on the user tests, the client application is generally easy to use. It is especially easy to learn how to operate the system, which might lower the threshold of implementation among users.

The system has some novelty value among users and the overall reaction to the service is positive. However, it seems to be difficult for the user to think what the true added value of the service is when compared with printed magazine and magazine website.

The challenge with image linking is to find uses that solve a real problem and enable something fundamentally new, useful or uniquely entertaining.

References

[1] H. Bay, T. Tuytelaars, and L.V. Gool. “SURF: Speeded up robust features”. In Proceedings of ECCV 2006. May 2006.
[2] C. Silpa-Anan and R. Hartley. “Optimized KD-trees for fast image descriptor matching”. In Proceedings of CVPR 2008.
[3] X. Chen and M. Koskela. “Mobile visual search from dynamic image databases”. In Proceedings of 17th Scandinavian Conference on Image Analysis (SCIA 2011). Ystad, Sweden. May 2011.
[4] V. Chandrasekhar, D. Chen, S. Tsai, N.-M. Cheung, H. Chen, G. Takacs, Y. Reznik, R. Vedantham, R. Grzeszczuk, J. Bach, and B. Girod. "The Stanford mobile visual search dataset". In Proceedings of ACM Multimedia Systems Conference. February 2011.
[5] X. Chen, M. Koskela, and Jouko Hyväkkä. "Image Based Information Access for Mobile Phones". In Proceedings of 8th International, Workshop on Content-Based Multimedia Indexing (CBMI 2010). Grenoble, France. June 2010.
[6] J. Sivic, B. C. Russell, A. A. Efros, A. Zisserman, and W. T. Freeman. Discovering object categories in image collections. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2005.
[7] Ondrej Chum and Michal Perd’och and Jiri Matas. Geometric min-Hashing: Finding a (Thick) Needle in a Haystack. CVPR, pages 17–24, 2009.
[8] Michael Calonder, Vincent Lepetit, Christoph Strecha, and Pascal Fua. BRIEF: Binary robust independent elementary features. In Kostas Daniilidis, Petros Maragos, and Nikos Paragios, editors, *Computer Vision - ECCV 2010*, volume 6314 of *Lecture Notes in Computer Science*, chapter 56, pages 778–792. Springer Berlin / Heidelberg, Berlin, Heidelberg, 2010.
[9] Chin, J.P., Diehl, V.A. & Norman, K.L. (1988). Development of an Instrument Measuring User Satisfaction of the Human-Computer Interface. Proceedings of the SIGCHI conference on Human factors in computing systems (CHI’88). Washington, DC, USA, May 15 - 19, 1988. pp. 213–218