

Cell Phones as Imaging Sensors

Nina Bhatti^a, Harlyn Baker^a, Joanna Marguier^b, Jérôme Berclaz^b, and Sabine Süssstrunk^b

^a Hewlett-Packard Laboratories, 1501 Page Mill Drive, Palo Alto, California 94304 USA;

^b School of Computer and Communication Sciences (IC), EPFL, Lausanne, Switzerland

Abstract

Camera phones are ubiquitous, and consumers have been adopting them faster than any other technology in modern history. When connected to a network, though, they are capable of more than just picture taking: Suddenly, they gain access to the power of the cloud. We exploit this capability by providing a series of image-based personal advisory services. These are designed to work with any handset over any cellular carrier using commonly available Multimedia Messaging Service (MMS) and Short Message Service (SMS) features. Targeted at the unsophisticated consumer, these applications must be quick and easy to use, not requiring download capabilities or preplanning. Thus, all application processing occurs in the back-end system (i.e., as a cloud service) and not on the handset itself. Presenting an image to an advisory service in the cloud, a user receives information that can be acted upon immediately. Two of our examples involve color assessment – selecting cosmetics and home décor paint palettes; the third provides the ability to extract text from a scene. In the case of the color imaging applications, we have shown that our service rivals the advice quality of experts. The result of this capability is a new paradigm for mobile interactions — image-based information services exploiting the ubiquity of camera phones.

Keywords: Image Processing and Computer Vision, Applications and Expert Systems, Recommendation systems, Mobile based image services, Cloud computing

1. INTRODUCTION

Over the past decade, mobile phones have evolved from being a luxury accessory to becoming an object indispensable to life. By 2008 there were more than 3 billion mobile subscribers worldwide, with some of the more mature markets having over 100% penetration [1]. With respect to camera phones, there were over 700 million shipped in 2007 and 1.3 billion expected for 2012 [2]. A 13% annual growth rate is anticipated. Camera phones have rapidly put a network-connected ubiquitous visual sensor in everyone's pocket. While technical capabilities of mobile phones have evolved greatly, consumer use has not kept pace. Kindberg *et al.* [3] report that most people still use their cell phone cameras as they would a standard digital camera—that is, for storing photographs—and do not take advantage of other offerings such as MMS. In this paper we report a first step in creating intelligent services that go beyond simple snapshotting, to provide picture interpretation via powerful *cloud services*, unconstrained by the software and handset platform. These services access massive computation in delivering unique capabilities of personal image assessment. In this context, there is a high potential for new multimedia-based services that make use of underexploited cell phone hardware.

Cameras to the Cloud

For many consumers, a cell phone is simply a device used for spoken communication. What makes it special, however, is that it is mobile and is evolving to contain a plethora of sensors such as camera, GPS, accelerometer, proximity and light detector, and microphone [4]. While equipped with increasing numbers of sensing modalities, we feel that none has quite the potential of the camera, because of the capability of computing advisory and other services linked to its images. Such image-based computation could be performed on the handset, but our view is that the most interesting opportunity comes in accessing the cloud. In short, we can take a picture, request expert analysis, and receive feedback and advice of immediate value. This *in situ* measurement capability is unparalleled. The convenient and easy-to-use nature of the device greatly facilitates its use by consumers. *Snap, Analyze, and Go* is a novel operational model.

An additional driving factor for using the cloud as the application server is that with the large number of cell phone brands, operating systems, technologies employed, and service providers out there, it would be impractical to develop a cell phone application supported across all these variants. Using the cloud lets us bypass this difficulty; we simply rely

on the mobile client being connected to cloud services, and let the network be the unifier. In this paper we describe three applications built around this notion of using any camera-equipped network-connected cell phone as the front end for accessing personalized image processing services.

A New Paradigm: Image-based Queries

In demonstrating these applications, we also argue that this presents a new paradigm for mobile applications; image-based queries. If a picture is worth a thousand words, we suggest that a picture may be worth a thousand key strokes. Applications would be much easier to use if the only requirement were taking and sending a picture — about the simplest possible interface. If we tried to rely on the usual speech or text for such an application, the process would be much more difficult. We have created specific-interpretation image-based systems for cosmetics selection, home décor, and text extraction, and carried out initial explorations in healthcare, language translation, and other areas.

Note that the use of cell phones as the imaging sensor gives immediate ubiquity. Cell phone figures on a world-wide basis make it clear that, for many populations, this will be their single communication and computation device. Cell phone penetration in the US is already at 89% and continuing up [5]. For many in the developing world it is their only computing platform. Consider as well that many of the engines of innovation revolve around commercial applications, bringing an accelerating effect to the injection of technology to this space.

The three examples we demonstrate have identical interfaces—up to the destination number keyed. MMS is used for the outgoing, and SMS brings back the response. In what follows, we discuss fundamental challenges in this approach, requirements of the individual applications, our means and methods in performance evaluation, and plans for future developments.

Color Assessment from Camera Phones

Despite the ease of taking a picture, objective color assessment remains difficult, particularly with the variable quality of devices generally found in cell phones. The same scene imaged with different cameras can result in significant aesthetic and quantitative differences, primarily due to uneven illuminant compensation, in-camera processing, and varying imager characteristics. It is impossible to accurately assign a color from an uncalibrated digital imager — both the camera and the scene present unknown influences [6]. Color properties of objects are fully characterized by their reflectance spectra, i.e. the percentage of light reflected by the object's surface at each wavelength and incident angle. In many applications, however, it is sufficient to retrieve only tristimulus values, and these can be acquired from an RGB camera. Several approaches using calibrated trichromatic imaging systems have been presented ([7,8]).

Our method requires just a simple calibrated reference target to be present in the scene, providing a cheap alternative to using an expensive calibrated device (that would only provide for the camera side of the unknowns, anyhow). The object of interest is imaged together with the reference target, its reference values observed, and a transform derived, allowing color correction of images independently of both the imaging device and the scene illuminant. The next three sections detail the applications and our solutions.

2. COSMETIC ADVISORY SERVICE

The beauty and cosmetics industry presents one of the highest areas of consumer spending — it is a \$160 billion market. This industry includes makeup, skin and hair care, fragrances, cosmetic surgery, etc. [9]. The US color cosmetics industry alone is a \$29 billion market [10]. Research shows that foundation makeup is the second most popular product in the color cosmetics category. Foundation is a tinted liquid, cream, or powder that is applied to the face before other color cosmetics. Its main purpose is to even the skin tone, hide flaws, and provide a base for further cosmetic application. For this product to work effectively it must be selected to match the skin tone of the user. This requires the accurate perception of the consumer's skin tone. Our user studies have confirmed what the cosmetics advertisements claim — it is difficult for women to select a foundation that best matches their skin tone, and this often results in purchase dissatisfaction.

We created a mobile imaging-based application in response to a commercial request to help women shop for cosmetics[11]. The goal was to increase women's satisfaction with cosmetics purchases by giving personalized product recommendations. We conducted a series of in-depth interviews with women and discovered that they were quite frustrated with their cosmetics shopping experiences even when purchasing at an attended counter. Users felt that they

wanted to be able to use an application at a time of their choosing, they were unlikely to plan the use of the application, and therefore were unlikely to download software in anticipation. The technology also needed to be convenient to use and likely to be present with the user when the shopping opportunity appeared. This led us to believe that the mobile phone in the consumer's pocket could be the ideal platform. Interestingly enough, statistics at project inception indicated that women took more digital photographs per month than men [12] and continued to lead men in camera phone usage. In 2006, 51% of women used their camera feature occasionally to frequently while only 40% of men did so [13]. We considered these conditions and expect that women would indeed be comfortable in a mobile imaging application.

Cosmetics selection requires visual acuity, color judgment, and knowledge about fashion and trends, which makes for a complex and possibly frustrating situation for the typical shopper. We describe here a system that provides these capabilities on any camera phone connected over any service provider. The camera phone will be used as a mobile imaging device with easy-to-use network transmission capability via MMS. Note that a mobile technology in the hands of the consumer may be ideal, as she can interact with an application in a store, at home, and at any time she needs assistance.

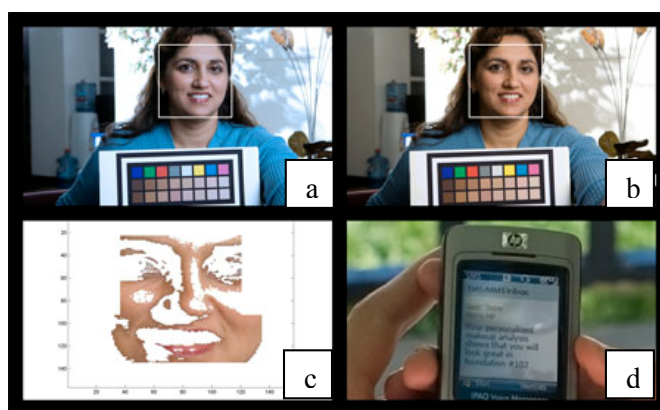


Figure 1. (a) Original image face detection (b) color correction, (c) sorted skin pixels, and (d) SMS recommendation.

To address the issue of makeup selection, we have devised a mobile phone based solution that allows a user to photograph herself using a cell-phone camera while holding a specially designed color chart. As mentioned, such a reference is critical, since color accuracy is generally impossible to guarantee with uncontrolled illuminant conditions and arbitrary cameras [6]. A captured image is sent by the consumer via MMS to an advisory service host at a backend server. The system uses image processing algorithms to locate the face (Fig. 1a), color science to correct the image color (Fig. 1b), and statistical classifiers to sample the image (Fig. 1c), determining the user's foundation makeup color with accuracy comparable to that of an expert. The photo analysis works regardless of image quality, camera specifications, and lighting conditions. Within seconds of sending the MMS, the consumer receives the SMS text containing the foundation shade recommendation judged best for her complexion (Fig. 1d). Figure 2 sketches the flow of information through the recommendation process.

Our studies have shown that our system gives a distribution of predictions comparable to that found between two cosmetic experts under different lighting conditions, different consumer imaging devices, and different cosmetic product lines [6]. The system can be configured to plug in any expert opinion database popular with consumers, a brand-oriented expert, a retailer-sponsored expert, etc., providing regional selectivity, product differentiation, and scalability.

The user experience is tied to the consumer's phone's MMS and SMS facilities. We do not require any software downloads, special camera software, or hardware camera quality. It is designed to work with the lowest quality consumer camera-phone and still give good results.

The color assessment is required for obtaining accurate color information about the users face. For the system to be most effective, the chart must be held near the face and roughly in its plane so that the face and the chart receive roughly the same illuminant. A chart covered with a shadow or a bright light from the side will give poor results. The next section

describes ways of detecting this. The user could receive a diagnostic SMS reply with errors such “avoid shadows on the chart, please, and try again,” “no chart found in image, try again”. We found that feedback was essential to help novice users with the system.

2.1. IMAGING PIPELINE

The cosmetic recommendation system is composed of image processing, chart detection, color correction, skin pixel extraction, and statistical classification of skin color to a database of expert opinions. In the next sections we will describe these aspects of the application.

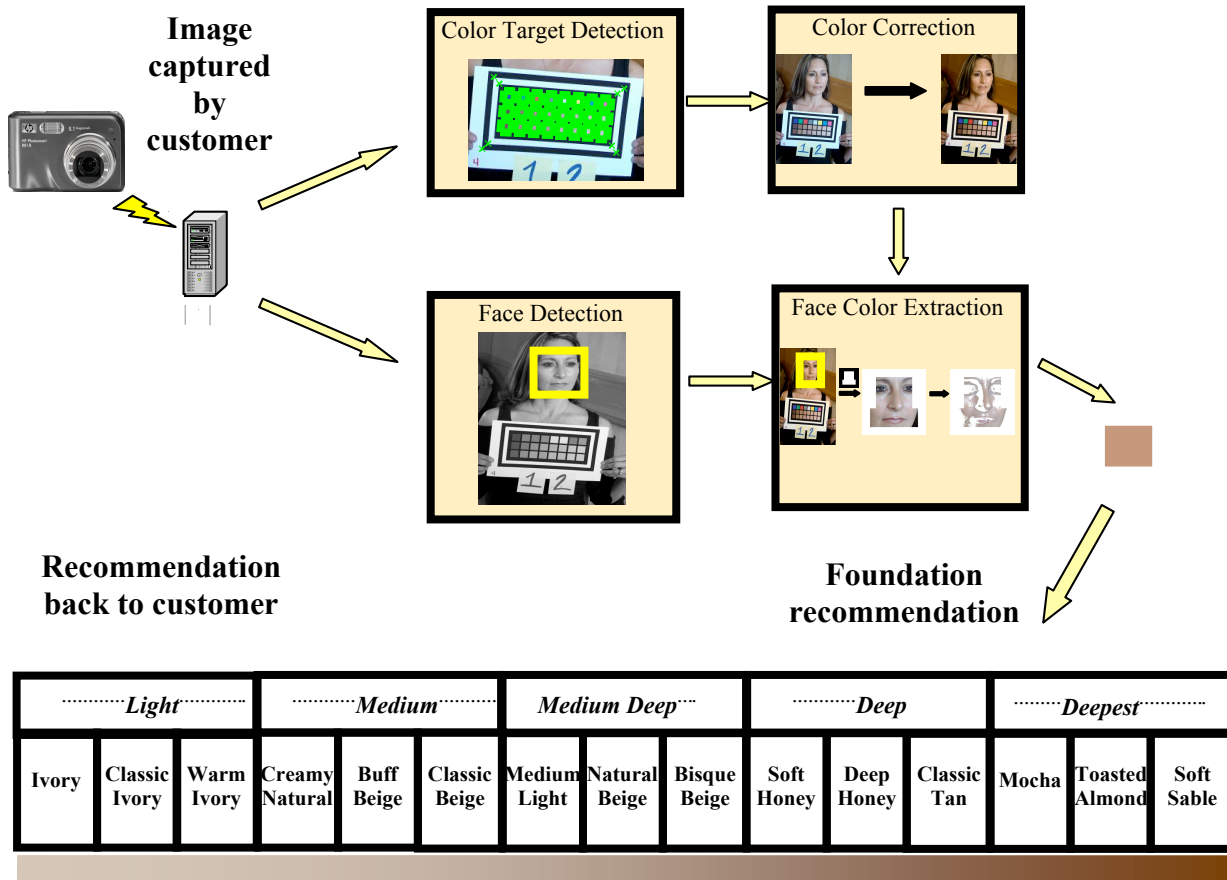


Figure 2. Overview of automated recommendation system including imaging pipeline (color target detection, color correction, face detection, face color extraction resulting in skin color representative color and finally the mapping by the recommendation system to the correct foundation shade).

Development of such a service is beset by many significant technical challenges. First, while accurate perception of a person’s coloration is the most critical factor in determining a proper shade of cosmetic, this perception is affected by the lighting conditions as well as the imaging device sensor’s characteristics and its processing. For our service to work in the uncontrolled lighting of stores and other environments — with any camera — we must find a way to compensate for these effects. In addition, we must devise a robust system to find the user’s face in the image, for a wide variety of poses, facial expressions, and distracting backgrounds. Further, once the face is found, we must judiciously sample it to determine the color of the skin and possibly other features, avoiding contamination by shadows, blemishes, sun-damaged areas, and specularities. In our view, successful automated cosmetic recommendation hinges on selection of appropriate facial appearance features and development of a recommendation system that ties these through machine learning to an expert-generated ground truth.

To make recommendations, our system processes an image, decides upon and delivers a response to the user. Since the imaging technology is computationally intensive, and the goal is a result within a few seconds on *any* cell phone, we perform our computation at the server-end and not on the mobile device.

Color Correction and Skin Color Assessment

In this section, we present consistent, fast techniques for measuring and classifying facial skin color from a single, casually posed digital camera image. The techniques are designed to allow the comparison of skin tones across a wide variety of cameras and environments, for all races of people. We also investigate the effects of varying image quality on our methods. Our aim is to develop methods that are practical for application in many computer vision and image processing tasks, beyond what prior art has enabled.

Reference Color Chart

Color calibration is achieved through the use of the reference color chart which has been specifically designed to calibrate for the color range of human skin tones. This allows the system to perform an approximation of a spectral analysis. Using robust computer vision algorithms the color chart is located in the image regardless of its orientation or placement. While no special positioning is required, results are best when the illuminant falling on the face and the chart are the same, so we recommend the chart be held in front at the user's chest area. The chart can take many forms. The simplest is gridded squares, as seen in Figure 1. But we have created several charts more in keeping with the esthetics of consumer brands (see Figure 3). We have found through our studies that we require a minimum of 5 skin-tone colored sampling areas, at least 1 inch in diameter [6]. The appearance of the color values in the chart is used to correct for both illuminant conditions and camera variations.



Figure 3. Example color chart designs: incorporated into a butterfly brand symbol, as a necklace.

Figure 3). We have found through our studies that we require a minimum of 5 skin-tone colored sampling areas, at least 1 inch in diameter [6]. The appearance of the color values in the chart is used to correct for both illuminant conditions and camera variations.

Validation of Color Correction

A basic question can be posed in addressing the validity of this effort: how effective can automated analysis of casually captured images be at providing the metric information needed for successful color-based product recommendation. Our studies have addressed this in several ways. With a desire to better understand the generality and metric character of our approach, we have submitted our methods to fairly rigorous analytic review. We pose the following three questions:

1. Can a cell-phone imager provide the color discrimination needed for a task when, before, this had been thought to require expensive colorimeters and spectral analysis?
2. Can a simple printed color pattern serve as a reliable reference palette for repainting an image to true color?
3. How broadly applicable is the technology in the space of inexpensive and varied-resolution cell-phone-type cameras?

We have addressed these issues in great detail in [6,13], here we present a summary of the service's ability to color correct facial skin from uncalibrated images.

Imagers as spectra-reading substitutes

We first examine the question of adequacy of cell phone imagers for a color correction task that was previously the domain of spectral analysis — noting 3-band RGB imagers only sample a small part of the scene's color, while spectral analysis samples the energy over 40-50 wavelengths through the visible range. Our subjects, holding a copy of our color calibration target, were captured with four different single-chip CCD cameras — an HP 850 (3.9 MPixel), a Nikon D1 (2.7MPixel), an HP iPAQ camera (1.3 MPixel), and a Nokia 6820 CIF-quality cell phone camera (352x288 pixels). Faces were detected and extracted using Viola-Jones face detection [13,14] with only those selected face-region pixels having lightness between 10% and 90% of the distribution being considered (this removes outliers due to hair, contour shadows, eyes, lips, and blemishes — see Fig. 1c).

The mean color-corrected values of qualifying pixels are then used as the skin color estimate. Contrasting with these values, we measured ground truth as the skin reflectance of each subject's face from a uniform area of the cheek using a

portable Microflash spectrometer (with $0^\circ/45^\circ$ measurement geometry). sRGB skin values computed for each of these using the observed skin reflectance spectra were then represented in normalized color coordinates. These values were compared with face colors extracted from the images, color corrected with the transform computed using all skin patches. We determined the correlation between imaged face color values and spectrally derived values to be high. The details of this analysis are published elsewhere [6].

Printed inks as true color reference

Previous work [6,11,13] has addressed our assessment of the color gamut and ability to use a linear transform on sampled values to adequately map in this space. The target itself being the only well-defined element in the image, we chose it to evaluate its own correction. The target is composed of color patches, primaries, gray and skin colors. Figure 4 shows the same image corrected with transforms computed using all 16 skin patches, eight skin patches covering the whole range of lightness, and the eight lighter patches only.

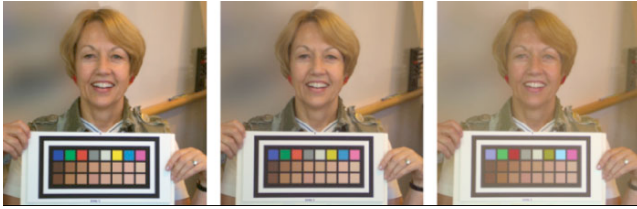


Figure 4. Correction with 16, covering-8, and lightest-8

The number and size of the patches influences the quality of color correction. Our color transform has twelve degrees of freedom, so we must have at least 12 measures — four 3-element RGB patches. We found the mean face color values obtained using four patches forming a convex hull in CIELAB space and the mean face color values obtained using the 16 skin colored patches to be very similar (Fig. 4 left and center) — differing about 1% in mean value and 5% in maximum value — showing sample quantity to be less significant than coverage.

Sensitivity to camera quality

Our test cameras covered the widest range of resolutions available — from megapixel through CIF. We had concerns about our ability to both detect the target in low quality images and to avoid sampling artifacts inherent in cell-phone transmission where providers often permit only much reduced bandwidth data upload. In these conditions, compression can be a major issue.

Larger color patches will be more reliably detected, especially in low-resolution cameras, and reducing the number of patches allows us to increase their size. In the images taken with the Nokia 6820 CIF cell phone, patches can be so small that JPEG artifacts begin to introduce significant errors in the estimation of patch color, and this presents the limiting case for our analysis. Notice the blocking and re-sampling artifacts in the roughly 15x15 CIF patches in Figure 5. While skin tones can be color corrected using only four patches, for robustness this is not recommended. A shadow projected on just one patch, for example, will cause the resulting color transform to be seriously degraded. More robust methods such as RANSAC [15] could ameliorate this, and would be advised for a production system. Figure 6 shows this and several other capture anomalies that could deter correct target recognition and/or color adjustment. As with the shadow handling, future additions to the system will accommodate to these challenges as well.



Figure 5. JPEG compression effects.

A detailed analysis using mapped lightness and normalized chromaticity showed that our color correction is coherent across resolutions, with ΔE values ranging from 1.8 to 5 [6]. The highest error came from the CIF imagery, indicating that, while discernible to a human, the difference in correction still provided good classification value over the tested recommendation classes throughout a broad range of camera characteristics. As with the overall color correction issue — where it was not essential that we make everything appear better, just the skin — here we only need performance that permits sufficiently accurate recommendation.

The high correlation across devices indicates that with a solid ground truth, the system can be correctly trained to give sufficiently consistent results for reliable use. While the assessment of color is not perfect, our method requires neither expensive calibrated imaging devices nor a controlled illuminant. It has been demonstrated to perform across a variety of consumer cameras. Our analyses have given us confidence in the system's ability to meet the demands of wide deployment.

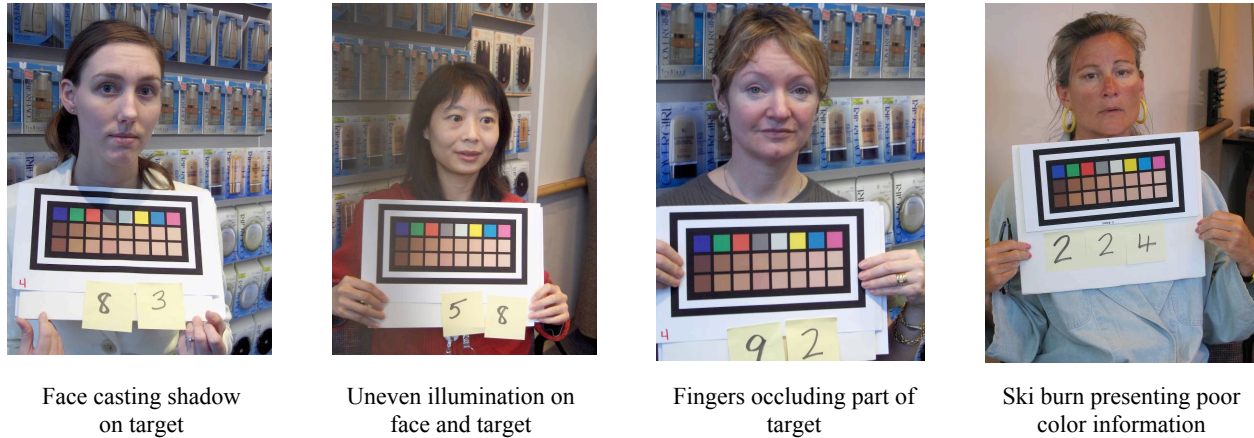


Figure 6. Acquisition anomalies that could deter accurate recognition and color adjustment.

2.2. CLASSIFICATION OF SKIN COLOR AND PRODUCT RECOMMENDATION

While the imaging service can compensate for lighting and camera system variations, it can not provide cosmetics selections. There is one final and critical step to formulating a cosmetics recommendation — a machine learning system is used to replicate expert opinion. We have created a statistical classifier that relies on the use of expert generated ground truth in establishing the professional opinion upon which our recommendations will be based.

Cosmetics is not “paint-by-numbers,” that is, detecting exact skin color is not sufficient and color matching to a foundation is not sufficient. To illustrate, if a user is sunburned she is unlikely to want to match this shade; instead the system should recommend a color of foundation that will compensate for this condition. Cosmetics experts know to correct too much red in the skin by applying a yellow toned foundation. We seek to present the same quality of advice. *The goal of our system is to be as good as a professional beauty consultant.* This means that we need to collect the functional beauty knowledge of products, and we accomplish this through systematically collected ground truth coupled to a statistical classifier to replicate cosmetic expert opinion.

Collection of Ground Truth

We have developed a protocol that ensures consistent data for statistical classification and captures the artistry as well as knowledge of cosmetics products. We have performed this protocol with 300 human subjects, several experts, and 6 international brands of cosmetics from low cost to luxury priced. The subjects were selected for ethnic diversity to ensure inclusion of all complexion colorings. Our cosmetics experts were instructed to take as much time as needed and that consistency was the critical factor in assigning cosmetic foundation. The data collection is done in the following steps:

1. The clean faced subject is photographed (with a color chart). A photometric, full spectral analysis reading of her skin is taken.
2. The cosmetic expert is presented with the human female subject with cleansed skin. She judges her complexion and selects the most appropriate foundation for the subject in given brand of cosmetics.
3. The cosmetic is applied and the “dry down” effects are observed. Corrections can be made to the selection at this time if the color blending of the cosmetic is not pleasing.

4. The foundation color is recorded for each brand in our study.

We analyze the expert’s cosmetic ground truth and locate any odd effects that would affect the consistency of recommendations. That is, we suppose that if the skin color of two subjects is very close (using spectral reflectance and photographic data) it implies that the cosmetic recommendation should be the same. If two subjects are found to have the same photometric coloring and yet the recommendations are quite different, the data is removed from the system. This happened quite rarely but this potential expert “mislabeling” could affect our results. Classifiers do not perform well with inconsistent data. We also check the data to ensure that each cosmetic shade has a minimum number of *exemplars*, subjects for which the foundation shade is correct. This ensures the classifiers ability to identify appropriate wearers of each shade.

Statistical Classification

Several machine learning approaches were evaluated to create classifiers that could predict foundation choices. All rely on high quality ground truth data. One of the more successful methods we use is a Kuiper Statistic method which compares distributions of color values. We compare skin color distribution of R,G,B values (after color correction is applied) to each exemplar from our ground truth set. The exemplar whose face has minimal difference in coloring distribution is used as the basis of product recommendation. The product selected is chosen from our exemplar indexed database, which contains the beauty expert’s techniques and judgments. We use the logic that if the subject looks most like an exemplar, then her best cosmetic will be the same as for the exemplar. In practice, this has led to good results. We present a blind test between two experts and show that our method results in the same differences as two experts.

Results

To evaluate our performance we invited two experts to classify the same group of women. Expert 1 was a professional makeup artist and educator who classified by observing the subject, selecting a foundation from the line, and testing it on the subjects skin. Expert 2 used a proprietary photometric based classification table. She measured the subject’s skin and indicated the foundation selection. Experts were not aware of the classification by the other expert so no cross influencing was present. To measure error from our foundation recommendation from another, we devised a consistent error metric D, Distance. Shades are not uniformly distributed, e.g. there are several nearly identical shades in the light shades and much greater gaps in the darker shades. The makeup expert assigned a 1 for distance for general one shade difference, .5 for very close shades, and greater than 1 for shades quite far apart.

In table 1 we list the error from Expert 1 to Expert 2 given in Distance. Note the two experts on average have a difference of 1 D. Corresponding to one shade off, this establishes the flexibility inherent in expert opinions and a bound on our prediction error.

Table 1 Camera based performance measures

| Camera Image Source | Resolution as tested | Average D (distance from Expert 2) |
|---------------------|----------------------|------------------------------------|
| HP 850 | 4 M / 4 | 0.764 |
| Canon S400 | 4 M / 4 | 0.867 |
| Nokia 6820 | 0.1 M | 0.896 |
| Nikon D1 | 2.7 M / 4 | 0.970 |
| HP R707 | 5 M / 4 | 0.943 |
| HP R817 | 5 M / 4 | 0.702 |
| Expert1 | n/a | 1.094 |

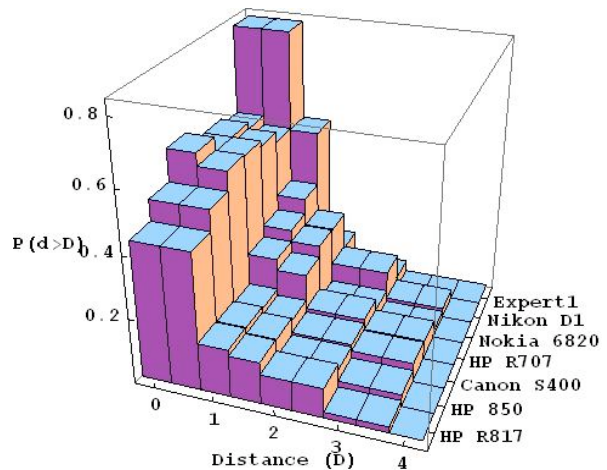


Figure 7. Comparisons of our system to Human Experts.

So how well does our system do compared to the human experts? We compare our application to the expert opinion. This study was performed on a data set of 73 subjects. The 73 subjects were the basis of the classification and we used a “leave one out” evaluation method on each subject. The subjects were classified by our system after being photographed with the color correction chart. The subjects were also classified by Expert 1 and Expert 2. Note that several photographic devices were tested to see if our results would be too sensitive to photographic devices. The resolutions used are given in table 1 and are typically equal to or less than MMS services typically support. In Figure 7 you can see that the probability of the color match system doing as well as or better than the expert is consistent across all imaging devices. We conclude that our system’s correctness rival’s the quality of expert opinion. This validates not just the classification but the entire application.

3. HOME DÉCOR ADVISORY SERVICE

We have demonstrated the feasibility of calibrating and measuring skin color from casually posted consumer images by placing a reference calibration target in the scene, but can this technique be applied to other domains also requiring color assessment? Another industry that is exceptionally dependent on color assessment and measurement is home décor. Home décor requires the selection of interior finish products that, when taken together, form an esthetically attractive combination. Home owners often rely on professionals to provide them with expert advice regarding paint, textiles, carpet, flooring, and hard finishes such as marble, granite, and tile. The selection is overwhelming and the novice often feels paralyzed by the multitude of products. Concepts such as “undertone”, “neutral”, “contract”, and “harmony” are fairly complex to master and the homeowner wants actionable advice delivered quickly. While there have long been paint matching solutions, what is lacking is the ability to determine a palette of colors — that, is a set of colors that form an esthetic group and are in keeping with current trends. Our goal is to determine palette selection that should coordinate well with collected image samples.

In the home décor application the calibration chart is composed of very different color values than were used in the skin classification problem example. The color values for skin color correction require a very small space in the entire color spectrum (see Figure 8). For home décor, the color range is much broader and we have devised a color chart system that gives complete coverage by dividing the color space in color ranges, each covered by a chart. The chart must be roughly in the same color space as the material sample to be measured. We have found good results with a selection of Munsell Colors covering a hue angle of roughly 60°. Operationally, the consumer can look at the sample and roughly approximate the color correction chart, e.g. blue green, reds, yellows, grey, etc. The details of the color chart design and effectiveness in color correction are given in [16].

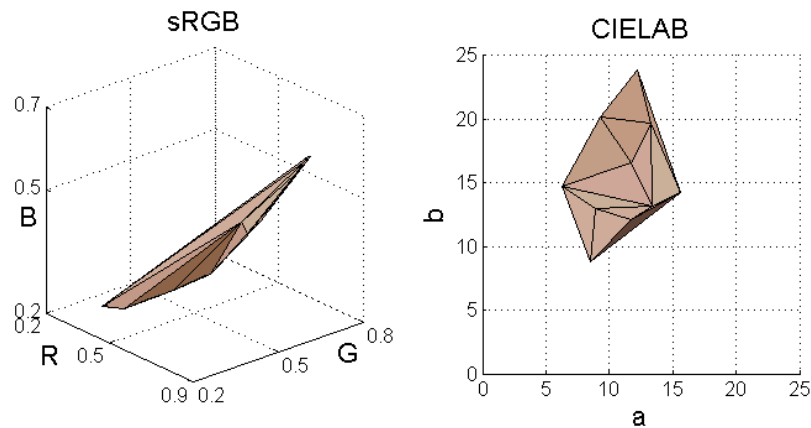


Figure 8. Skin tones only represent a limited color range and span a hue angle of about 20° in CIELAB. The surfaces represent the color values of the target used for skin color correction covering the range of human skin tones in sRGB and CIELAB.

The application is simple and follows these steps:

1. The user visually selects a color target whose hue is similar to that of the object, and captures an image of the object and target together. Note that a specific illuminant is not required; unknown but even illumination across the chart is sufficient for good results. As in the cosmetics application, the image can be transmitted using MMS, email, or other communication facility.
2. Using a computer vision algorithm, the chart is located in the image and the target patches are extracted. The color correction transform mapping the image target values onto reference values is computed and applied to the entire image.
3. The object pixel values are extracted, averaged, and converted to CIECAM02.
4. The object color is compared against a collection of coordinated paint palettes, each consisting of four harmonious colors. The best matching paint is selected based on the color distance of the object in CIECAM02.
5. The palette containing the best match is selected, along with the three paints complimenting its color – hence complimenting the object.
6. The resulting palette is transmitted to the user as an SMS, email, printed brochure etc. Figure 9 shows examples of home décor textiles with a color chart. The object is placed on an off-white background. The first image is uncalibrated and it shows clearly the poor color results. The second image results from application of the color, and the final panel shows the palette selected for the object.



Figure 9. Paint samples are grouped together in palettes of four colors. The application matches the object color to one of the paints and returned the corresponding palette. The figure shows, from left to right, the uncorrected image, the corrected image, and the corresponding paint palette.

We next wanted to evaluate how we had achieved our goal of giving expert quality advice using *in situ* color measurements. To do this we did an evaluation study with a professional home décor expert. Our application does not generate palettes; rather it is used to recommend palettes. We created a collection of 63 common décor elements

composed of textiles, wood flooring, ceramic tile, glass tiles, vinyl flooring, etc. We selected 9 décor palettes from a major US retailer. Each palette contained four colors resulting in a total of 36 colors.

The expert was shown each sample along with all 9 palettes of four colors each, chosen to be of similar color to the object. She was first asked to choose a paint from the 36 that she would pick as the best match for each of the 63 samples. The database of paints did not offer enough colors for her to systematically pick an optimal color. She was thus asked to rate her own choices. She graded the selection from 1 to 5 according to the following rating:

1. The sophistication of a high quality expert recommendation; perfect, given the palette selections
2. Competent work by an expert, given the palette selections
3. Close, but not perfect; typical of an untrained consumer
4. Poor selection; other selections are much better
5. Terrible; unacceptable, even for a consumer.

The expert was asked to give integer grades and keep uniform intervals between them. After she made her choice, we presented the automated analysis results and asked her to rate them as previously. The application's and expert's results were then compared and analyzed. We did not ask the expert to rate the quality of the color palettes chosen through automated analysis but, rather, to rate the quality of the match. Results are shown in figure 10. This figure shows the occurrence of each rating for the service (left) and expert (right) choices. The colors in the right plot correspond to the grades of the automated choices, i.e., it shows the proportion of samples whose matches were similarly rated. The expert gave the grade 1, meaning "perfect," to 29 of the 63 paints selected by our system, while she similarly rated 47 of her own choices. The application and expert top choices match for one third of the objects, but the remaining results are also good. Many different paint patches have similar CIECAM02 values, especially in beige tones. The expert graded most of the automated results as 1 (perfect) or 2 (competent). The average grades are 2.09 ($\sigma = 1.34$) and 1.57 ($\sigma = 1.21$) for the automated analysis and the expert, respectively. These results are satisfying. While the expert was only asked to rate the quality of the match of the object to a paint sample, she made several remarks regarding the applications selection of paint palettes that she had not chosen; on reconsideration, she found the provided choice to be quite good. This led her to remark that she, as an expert, would find this application quite helpful, since selecting palettes and paint matches is an exhausting task. The application had opened her eyes to palettes she would not have initially considered, so found that both expert and novice users could benefit from its use.

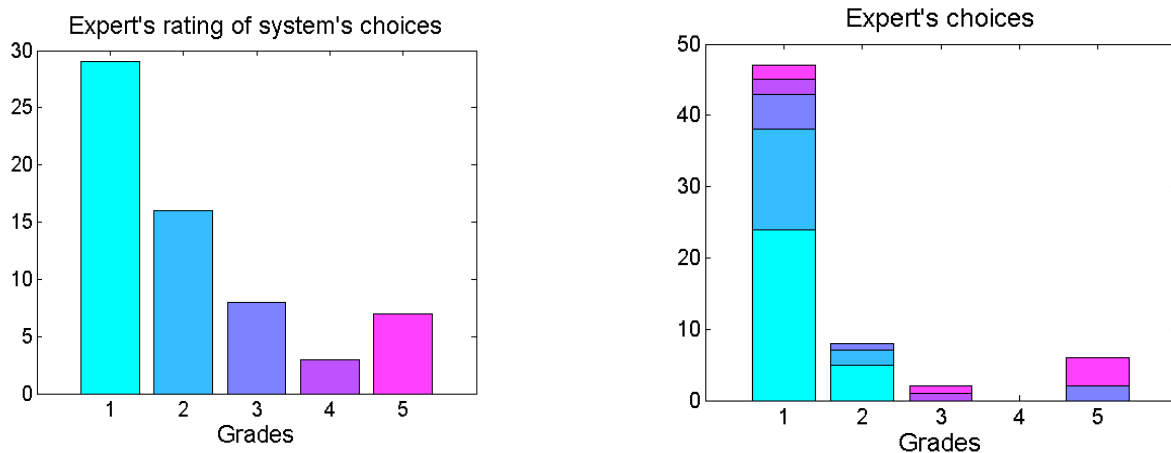


Figure 10. The occurrence of each rating, for the system (left) and the expert (right) choices. The colors in the right plot correspond to the grades of the system's choices, i.e. it shows the proportion of samples whose matches were similarly rated.

4. TEXT-BASED ADVISORY SERVICE

Imagine a traveler dining in a restaurant in a country on the other side of the world and struggling with the menu. “Is it Beef Steak or Lizard Steak?” could be one of the more pressing questions as the traveler reads through the menu. What if this traveler could use her camera phone to get a translation simply by taking a picture of the menu, sending an MMS and promptly receiving an SMS with the answer? In this section we describe an unsupervised text extraction service. Once the text is extracted from an image any advisory services taking text inputs can be used, such as Wikipedia, search, etc. For our example, we used a language translation service.

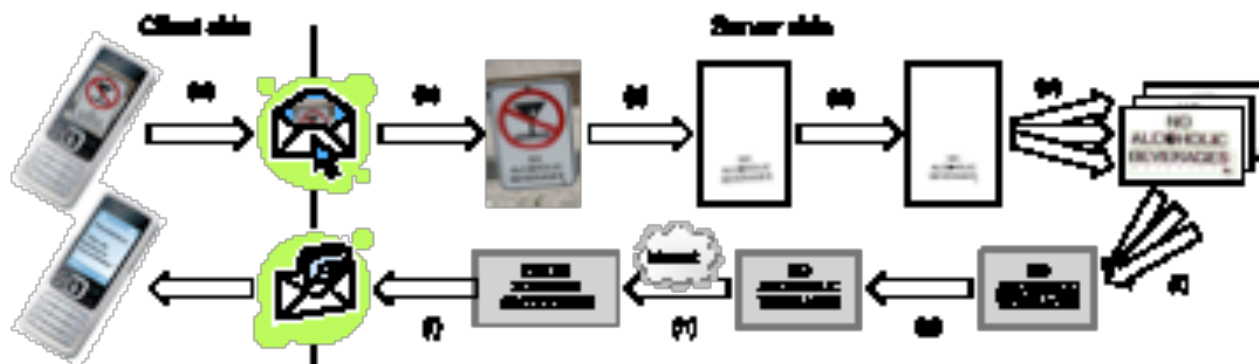


Figure 11. **System Overview.** A user takes a picture of text and sends it as an MMS (a). The server receives the MMS, extracts the image content (b) and performs text extraction to isolate the text region (c). Skew correction is applied if necessary (d). Then several OCR engines are invoked in parallel (e) and their output is merged (f). Spell checking is applied (g) to correct potential OCR mistakes. The resulting text is sent to an online translation engine (h), and the final translation sent back to the user in an SMS (i).

As illustrated on Fig. 11, the translation process is initiated by a user sending a picture by MMS. This MMS is received by the back-end server, where the image is passed to the text extraction module. This latter module attempts to isolate text elements from the rest of the image. It produces a binary image along with the coordinate information used for OCR. The text extracted by the OCR process is passed to web-services-based translation engines and, finally, the results are sent as SMS to the user.

While this may appear simple, there are several issues that complicate the process of extracting text from casual images. An over view is provided here and more details can be found in [17]. Currently available OCR algorithms have been developed to deal with printed documents. Typically, OCR engines process documents that have been previously scanned by a flatbed scanner, and expect them black and white, relatively clean, and well structured. The problem of text recognition from natural images is much less constrained. Not only is the image colored and cluttered with non-text objects — some of which might however present text-like textures — but the text itself can take very different shapes, colors, and fonts. Moreover, it is likely to be affected by perspective and lighting effects as well as compression artifacts. To make things worse, a standard text structure cannot be expected; thus finding text location within the image might represent a challenge by itself. In these conditions, the direct application of an OCR engine on a scene image containing text would result in an almost certain failure. Therefore, the first task in the server-side pipeline consists of processing the image to transform it in a way that current OCR engines are able to manage.

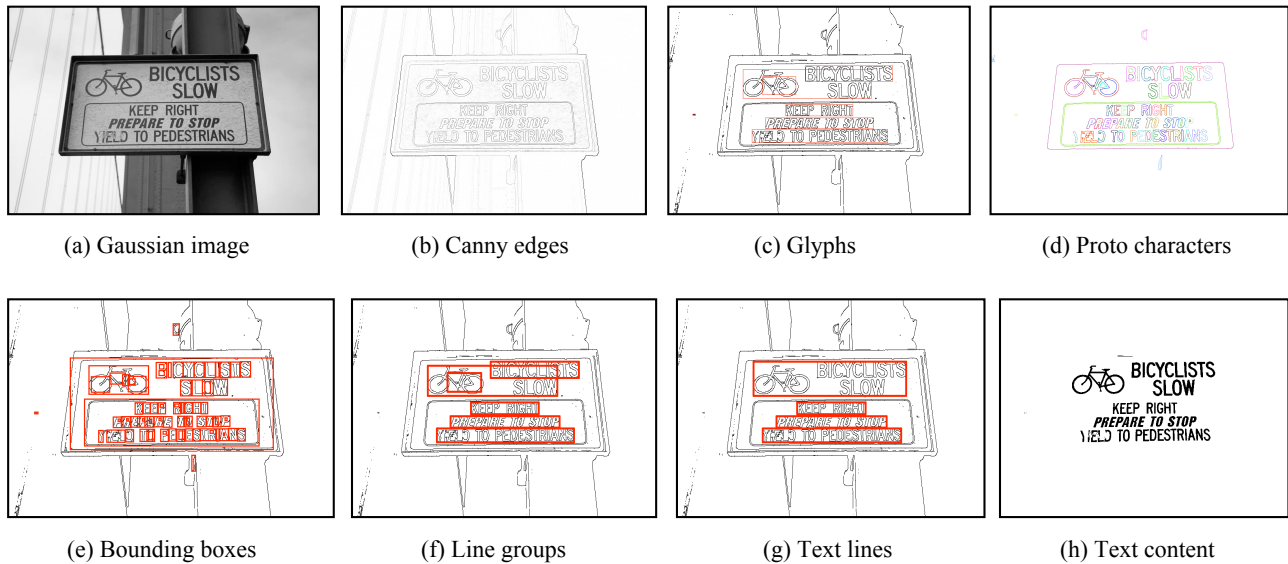


Figure 12. Steps of text extraction.

4.1 Scene Text Extraction

One of the fundamental elements in our pipeline is scene text extraction. Its goal is to identify and isolate text elements in a picture. Scene text extraction is designed to operate on raw images, and produce binary versions, in which all non-text elements have been removed. Our approach to scene text extraction is illustrated by Fig.12. After applying a convolution with a Gaussian kernel to the input image (a), in order to get rid of the possible noise and compression artifacts, we perform edge detection using Canny's method [18] (b). This result is then thresholded to yield a binary edge image, and the end points of open shapes are merged into closed shapes, based on the gradient of their linking path. At this stage, the binary edge image is segmented into individual glyphs, by performing connectedness analysis (c). The glyphs thus retrieved are analyzed one by one, and those less likely to represent characters are rejected, according to a set of heuristics based on size, aspect ratio, and topological constraints (d). For every glyph accepted, a bounding box is determined (e). The boxes are then clustered into words, and the words grouped into lines (f), discarding isolated boxes that do not fit into the structure. We end up with a set of bounding boxes covering the detected lines of text in the image (g). The final image is then generated by extracting the content of the original image in the areas covered by the line bounding boxes (h) and, for each of them, finding the appropriate support, in order to deal with text in inverse video.

4.2 Image Skew Correction

Even transformed by scene text extraction into binary versions with only text elements, the images may not yet be in an appropriate format for OCR processing. At this stage, images might still be plagued by rotation or perspective effects, to which most traditional OCR engines are very sensitive. Additional post-processing is thus needed to expect satisfying OCR results. In this work, we concentrate on rotation correction, and do not address the problem of perspective deformation. This shall be considered in future work in order to make the system more robust to any kind of deformation. Here, we base our reasoning on the fact that a large majority of texts written in western countries are aligned horizontally. Therefore, OCR engines expect characters aligned to the document in which they appear. However, user photographs, especially those taken with a cell-phone camera, are rarely perfectly level. Our goal is thus to determine whether the glyphs found by the text extractor are inclined with respect to the horizontal, and, if they do, to compute this angle α . We can later rotate the image by the inverse angle $\square\alpha$, in order to align the text horizontally. Note that we do not apply the rotation directly on the binary images, as this would alter their quality significantly. Instead, if the rotation is needed, we perform it on the original input image and re-apply text extraction. Assuming that, in most cases, characters are aligned horizontally into words, we compute the angle between any couple of glyphs identified by the text extractor and store the result in a histogram. Because of the horizontal alignment, the number of angles between couples of glyphs located on the same line should be significantly higher than other various angles between couples on

different lines. Thus, we expect to observe a well defined peak in the histogram bin corresponding to the roll angle. As shown by results of Fig. 4, this is the case in most images.

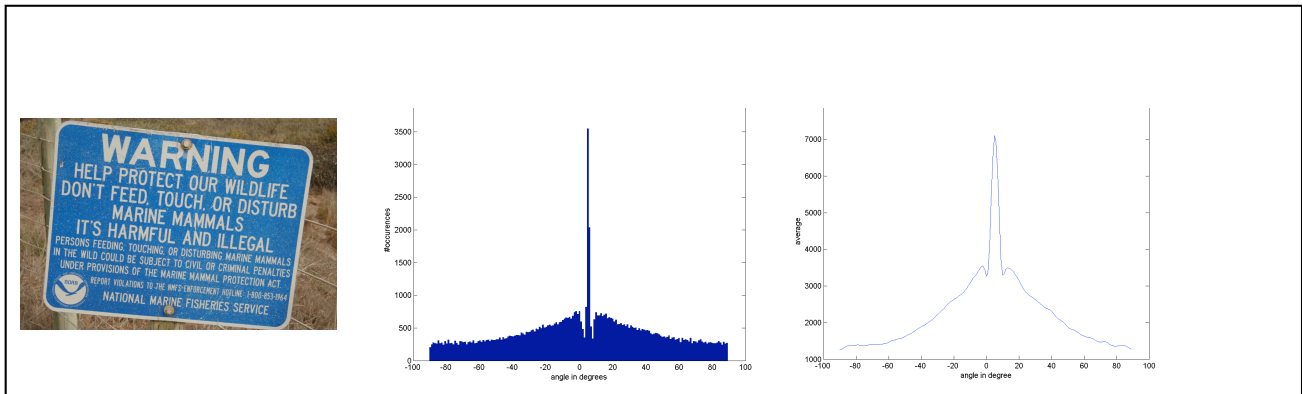


Figure 13. Text deskewing results. The large number of characters makes it fairly easy to determine the roll angle.

4.3 Optical Character Recognition

Optical Character Recognition has been a very active area of research for several decades and, as a result, many different analysis engines have been implemented, some of which are available in the public domain. For this reason, we chose to rely on an already-tested and solid implementation. Although OCR on scanned documents is generally considered as a solved problem, our input images, even corrected, are still not close to the ideal binary image coming from a flatbed scanner. Thus, we cannot expect as good OCR results on our images as they would be on scanned documents. In addition, the cloud services architecture lends itself easily to configuration of web based services. Thus, we chose to rely on multiple OCR services to generate a strengthened combined result. If new services appear that generate better results, these can quickly be configured into the system. These reasons motivated our design choice to rely on multiple OCR engines. For this system, we opted for the three following open source OCR engines: Tesseract [19] developed by HP in the 1980's, released as open source in 2005 and currently supported by Google; Ocrad [20] maintained by the Free Software Foundation; and GOCR [21]. All engines take a binary image as input and produce various output formats, ranging from complex XML with character locations and probabilities, to simple text-only output. Multiple OCR engines are integrated into our system in this way: We simply feed the corrected image processed by the text extraction and skew correction modules to every OCR engine and obtain as many outputs as there are engines. Ideally, all engines should have correctly recognized characters, yielding identical outputs. In reality, such a good coordination is rare, and we most likely have to post-process the OCR results and merge them into a unique answer before we can proceed with the translation.

Combining OCR results

Combining OCR results is only feasible if they present an equivalent output format. The first task we address is to unify the multiple output formats into a common XML one. Since we use open source software OCR engines, we were able to get access to their source code and modify these easily. After modification, every OCR engine produces an XML output containing a list of bounding boxes for every detected character, along with a character hypothesis and (on occasion) a second guess. A confidence score is also provided for each hypothesis. With output format unified, we proceed with the comparison. First, overlapping detections are merged. Then characters are clustered into words and words into lines. Isolated and low confidence character detections are dropped. At this stage, we integrate the spell checker method (to be described below) into the combiner in order to produce interpretable results. When character detections differ across engines, their confidence score and the spell-checker hints are taken into account to solve the conflict. The final merged result is produced in text-only format.

4.4 Spelling Correction

Despite using several OCR engines, the combined OCR result is likely to still contain mistakes, especially when the input images are of low quality. Typical errors are usually not a big challenge for humans, since they often consist of mixed visually similar symbols, such as 'o' and '0' (zero), or 'l' and '1' (one). Even when contaminated with these errors, a text is still easily readable by a person. However, the situation is very different when dealing with a translation engine. Those simple OCR mistakes have a strong impact on translation performance since misspelled words will not be matched in the translator's dictionary, and are often ignored. To address this problem, we propose the application of a spell correction program to the output of the OCR engines. Classical spell correction software is typically able to check if a word belongs to a dictionary, and otherwise suggest a list of possible replacements, sorted by occurrence probability. In this work, we use the open source software spell checker Aspell [22] which is available both as a stand-alone program and as a library. It is widely used in text editing software and has a strong reputation at suggesting meaningful replacements for misspelled words. We decided to merge the spell correction step directly into the OCR engine's combiner program. It makes much sense to do it at this stage because of the confidence score provided for each character by the OCR engines: Using both the OCR confidence information and the replacement probability provided by Aspell, we can take better decisions while merging OCR results, as explained above. The spell checker is also used during the merging process to split long words that have been merged mistakenly by OCR engines, or to merge adjacent words that were wrongly separated during character recognition. The use of software such as Aspell to correct spelling mistakes made by OCR is not without its problems. Most spell checkers have been primarily designed to correct human spelling errors, which turn out to be quite different from OCR errors. Human spelling mistakes tend to result from pressing the wrong keys of a keyboard, thus replacing a letter by another one located adjacent on the keyboard, or swapping a sequences of letters, for example writing 'teh' instead of 'the'. As discussed above, OCR tends to replace one character with another that is visually similar. Also OCR frequently glues two characters together – 'ot' could become 'd' – or misinterpret a small drawing for a character. Therefore, the replacement suggestions made by Aspell, are not always appropriate for these circumstances, and the distance score it provides with every replacement does not perfectly match our target application. Nevertheless, applying a generic spell checker such as Aspell still tends to improve performance. The use of a specific dictionary-based spell checker [23] designed for correcting OCR mistakes would definitely yield better results. We will address this in a future version of our system.

4.5 Translation

To this point, the efforts have been directed at extracting text from a cell phone camera image and parsing it to most resemble its original. The final step, before replying with an answer, is the translation itself. To perform this task, we again chose to rely on one of the numerous on-line translation tools. Those are attractive because of their free availability and constant revision for better performance. The drawback of this choice, as will be seen, is that they are usually very generic and designed to translate full texts. For our application, we selected Google Translation and Babel Fish Translation (Yahoo). They both provide free and reasonably good translation among English and the major other European languages. As can be deduced from the results of Fig. 6, we built our application translating English into French. Of course, as long as we are dealing with languages using western characters, our system can be easily modified to deal with any pairing of languages supported by the translation engines. Using online translation engines in an automatic process is not straightforward because they are built as websites and designed mainly for an interactive user experience. To be able to automate the process, we have to format our `http` requests to look as if they originate from a web browser. We perform this task thanks to the open source library `cURL` [24] which makes it possible to create well formatted `http` requests and manages to send them to a web server and collect their answer. Using such a web translation engine is a solution we chose out of convenience — it let us set up and deploy a working prototype quickly. In a commercial application, the use of a more specific type of translation tool would be advised. First, the format of the web pages can change without notice, requiring adaptation of the script that feeds the original language text and extracts the translation. Second, on-line translation solutions are very sensitive to misspelled words and incorrect or incomplete punctuation. When a word is not found in their dictionary, it is usually copied 'as is', and no further effort is made. Third, their dictionary is very generic and fails to translate some specific expressions. For an application like ours, where input text often comes from public signs, the use of a dictionary more adapted to the context could improve the general performance. Finally, translation engines are primarily targeted at translating whole and well-formed sentences. In practical application of camera phone translation, however, users are more likely to translate small texts originating from, for example, road signs or shop windows. These tend to consist of very simple sentences, with missing elements, such as verb, or subject.

5. RELATED WORK

Text Extraction

With the miniaturization of digital cameras and their widespread inclusion in mobile devices, text extraction from scene images or video has received a fair amount of recent attention. There are a number of published works that resemble our effort. The closest is that of Watanabe *et al.* [25], which presents an automatic translation tool targeted at translating Japanese text extracted from a photograph. That system runs on a PDA equipped with a digital camera. It is not fully automatic, requiring the user to specify the area of the image he wants to translate. Also, some strong assumptions are made: The background is supposed to be more or less uniform, and text fonts should not vary. The authors later implemented their system on a cell phone [26] using Java technology. In this implementation, only the text region selection runs on the cell phone. After this step, the cell phone connects to a server and sends it the image region to be translated.

A similar approach is developed by Haritaoglu [27] in which a scene text extraction and translation system for PDA is introduced. The user takes a picture of a scene containing some text in Chinese and selects the rectangle region he wants to translate. Due to memory and computing power limitations of the PDA, the computation is not performed onboard, but the handheld device connects to a back-end server via GSM modem and sends the picture. The server performs text translation to the user's native language and replaces the original text in the image.

Pilu and Pollard [28] developed a light-weight system performing text extraction from scene pictures. The focus is given on a fast implementation such that the algorithm can run entirely on a PDA. Demonstrated as a stand-alone application for text extraction with translation, it runs on a handheld device.

Automatic translation is used in a slightly different way in Yang [29,30]. The authors introduce an application capable of performing automatic sign translation from video. A first component identifies the text regions in the video and an interface lets the user choose the area he wants to translate. Another module then performs standard text extraction, recognition, and translation, and feeds the result back to the user. Chen [31] have refined the application such that user intervention is no longer needed.

Compared to the approaches described above, our method has a number of advantages. First, it is completely implemented on the server side, and relies on widely adopted communication channels. This has the combined benefit of making it readily deployable on virtually all mobile phones in use today, while requiring absolutely no software installation on the client side. Second, our method is entirely automatic, and requires no user intervention other than taking a picture and sending an MMS. Finally, our system is made of several interoperable modules, which makes it both easy to upgrade by changing one particular module, and simple to deploy in a cloud architecture to obtain high performance and scalability in response to strong user demand, should that occur.

Bar codes

Most past work in mobile imaging has centered on recognizing printed 2D bar codes and responding with a URL or related content [32,33,34]. These systems are very popular in Japan where many camera phones come preloaded with bar code recognition software. Neomedia [35] and Scanlife [36] offer this software for download on any phone.

RedLaser is an iPhone application for reading conventional barcodes. This application identifies the product from the barcode and provides pricing and retailer related information allowing the consumer to make instant price comparisons [37].

Image-Based services

A well established research area of camera phone applications is image-based localization. Ravi *et al.* [38] propose using cell phone cameras as indoor localization devices. A user walks inside a building with a cell phone worn around his or her neck; the phone takes pictures at regular intervals and sends them to a server. Localization is inferred by comparing the query picture to a database. Other image based services have been used for personal memories [39].

Image-based search query [40] is another possible application for mobile phone cameras. Arguing that entering a text query on a mobile phone is too cumbersome, we advocate the use of snapshots — supplemented with keystrokes as necessary — for querying search interaction.

Several companies are exploring image recognition for mobile marketing applications, Clic2C [41], Snapnow [42], and Snaptell [43], (now A9, part of Amazon). None of these use images in understanding the consumer's needs in creating a personalized response.

Another use of images is to track portion and food types consumed. Studies show that consumers prefer taking pictures of their meals than manual tracking. The imaging services determine portion and either use automatic food identification or manual classification [44,45].

6. CONCLUSIONS

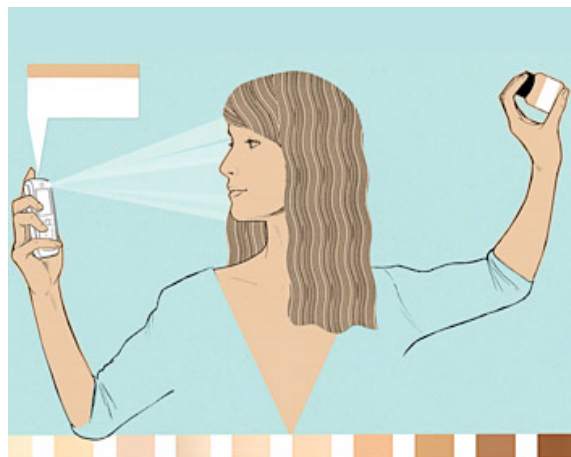


Figure 8: “O”, (Oprah Magazine) illustration

We have demonstrated several novel mobile applications targeting the unsophisticated user. Exploiting computation in the cloud rather than incurring complex operations and interactions on the user's side, these applications build on unique imaging and computer vision capabilities — and a use model — that allows them to work with any camera phone. We demonstrated and quantified a camera phone image-based system that provides high quality recommendations comparable to those of an expert who is viewing the situation. This reduces cost, increases user satisfaction, and extends the reach of expert advice to many who would not otherwise have the advantage of such information.

Recently our system received coverage in *O, The Oprah Magazine*, a leading women's lifestyle publication [46]. Here and elsewhere we have received interest from a press intrigued with the novelty of a mobile system that can offer advice as good as that of an expert. A predominant element of the feedback from

our user studies and cosmetics companies is that *people want rich services that are simple and easy to use* [47]. We feel that addressing a real user need with applications that are simple, intuitive, and not the least bit intimidating, is key to achieving ubiquitous mobile services. We anticipate extending this paradigm, and have begun study of other image-based query domains for mobile services, such as healthcare.

REFERENCES

- [1] P. Budde et al., [Global Mobile Communications – Statistics, Trends and Forecasts], BuddeComm, Feb 2008.
- [2] Infotrends, [Worldwide Camera Phone Forecast: 2007-2012] (July 2008).
- [3] T. Kindberg, M. Spasojevic, R. Fleck, and A. Sellen, “The ubiquitous camera: An in-depth study of camera phone use,” *IEEE Pervasive Computing* 4(2), 42–50, 2005.
- [4] <http://www.apple.com/iphone/iphone-3gs/high-technology.html>
- [5] Jupiter Research Mobile Marketing &Media. US Mobile Marketing Forecast 2007–2012. Jan 2008.
- [6] J. Marguier, N. Bhatti, H. Baker, M. Harville and S. Süssstrunk, “Assessing human skin color from uncalibrated images,” *International Journal of Imaging, Systems and Technology*, special issue on Applied Color Image Processing, Vol. 17, Nr. 3, 143–151, 2007.
- [7] J. Farrell, D. Sherman, and B. Wandell, “How to turn your scanner into a colorimeter,” *Proc. of IS&T 10th Int. Congress on Adv. in Non Impact Printing Technol.*, 579–581, 1994.
- [8] W. Wu, J.P. Allebach, and M. Analoui, “Imaging colorimetry using a digital camera,” *J. Imaging Sci. Technol.*, vol. 44, 267–279, 2000.
- [9] *The Beauty Business*, *The Economist*, May 2003.
- [10] M.L. Gavenas, *Color Stories: Behind the Scenes of America's Billion-Dollar Beauty Industry*, Simon & Schuster, 2002.
- [11] N. Bhatti, H. Baker, H. Chao, S. Clearwater, M. Harville, J. Jain, N. Lyons, J. Marguier, J. Schettino and S. Süssstrunk, “Mobile Cosmetics Advisor: An Imaging Based Mobile Service,” *Proc. of IS&T/SPIE Electronic Imaging: Multimedia on Mobile Devices 2010*, Vol. 7542, 2010.
- [12] Infotrends. 2004 Worldwide Camera Phone and Photo Messaging Forecast. March 2004.

- [13] M. Harville, H. Baker, N. Bhatti, and S. Süsstrunk, "Image-based measurement and classification of skin color" Proc IEEE Intl Conf Image Process 2, 374–377, 2005.
- [14] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," International Journal of Computer Vision, 511–518, 2001.
- [15] M. Fischler and R. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," Communications of the ACM 24, 6, 381–395, 1981.
- [16] J. Marguier, N. Bhatti, H. Baker and S. Süsstrunk, "A Home Décor expert in your camera", Proceedings of the IS&T/SID 17th Color Imaging Conference, Vol. 1, 85–90, 2009.
- [17] J. Berclaz, N. Bhatti, S. Simske, and J. Schettino, "Image-based mobile services: automatic extraction and translation," Proc. of IS&T/SPIE Electronic Imaging: Multimedia on Mobile Devices 2010, Vol. 7542, 2010.
- [18] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence* 8(6), 1986.
- [19] R. Smith, "Tesseract OCR," (1985-2008). <http://code.google.com/p/tesseract-ocr/>.
- [20] Free Software Foundation, "GNU Ocrad," (2003-2007). <http://www.gnu.org/software/ocrad>.
- [21] J. Schulenburg, "GOCR," (2000-2008). <http://jocr.sourceforge.net/>.
- [22] K. Atkinson, "GNU Aspell," (2004-2008). <http://aspell.net/>.
- [23] A. Dengel, R. Hoch, F. Hönes, T. Jäger, M. Malburg, and A. Weigel, [*Handbook of Character Recognition and Document Image Analysis*], ch. Techniques for improving OCR results, World Scientific, 1997.
- [24] D. Stenberg, "cURL." <http://curl.haxx.se/>.
- [25] Y. Watanabe, Y. Okada, Y. Kim, and T. Takeda, "Translation camera," International Conference on Pattern Recognition, 1, 613–617, August 1998.
- [26] Y. Watanabe, K. Sono, K. Yokomizo, and Y. Okada, "Translation camera on mobile phone," International Conference on Multimedia and Expo, 2, 177–80, July 2003.
- [27] I. Haritaoglu, "Scene text extraction and translation for handheld devices," Computer Vision and Pattern Recognition, 2, 408–413, 2001.
- [28] M. Pilu, and S. Pollard, "A light-weight text image processing method for handheld embedded cameras," British Machine Vision Conference, Sept 2002.
- [29] J. Yang, J. Gao, Y. Zhang, X. Chen, and A. Waibel, "An automatic sign recognition and translation system," Proceedings of the Workshop on Perceptive user interfaces], 1–8, 2001.
- [30] J. Yang, J. Gao, Y. Zhang, and A. Waibel, "Towards automatic sign translation," *Proceedings of the first international conference on Human language technology research*, 1–6, 2001.
- [31] X. Chen, J. Yang, J. Zhang, and A. Waibel, A., "Automatic detection and recognition of signs from natural scenes," [*IEEE Transactions on Image Processing*], **13**, 87–99, Jan 2004.
- [32] J. Rekimoto and Y. Ayatsuka, "Cybercode: Designing Augmented Reality Environments with Visual Tags", Proceedings of DARE 2000 on Designing augmented reality environments, 2000.
- [33] <http://www.mobilecodes.org/>
- [34] <http://www.activeprint.org/>
- [35] www.neom.com
- [36] www.scanlife.com
- [37] <http://redlaser.com/>
- [38] N. Ravi, P. Shankar, A. Frankel, A. Elgammal, and L. Iftode, "Indoor Localization Using Camera Phones," 7th IEEE Workshop on Mobile Computing, April 2006.
- [39] R. M. Jesus et al., "PhotoNav: A Retrieval Based Interface to Navigate in Physical Spaces Using Memories", WWW2007, May 2007.
- [40] X. Fan, X. Xie, Z. Li, M. Li, and W. Ma. "Photo-to-search: using multimodal queries to search the web from mobile devices," ACM SIGMM international workshop on Multimedia information retrieval, 143–150, Singapore, 2005.
- [41] www.clic2c.com
- [42] www.snapnow.com
- [43] www.snaptell.com
- [44] C. Boushey, D. Kerr, J. Wright, K. Lutes, D. Ebert, and E. Delp, "Use of technology in children's dietary assessment," European Journal of Clinical Nutrition, 2009.
- [45] I. Woo, O. Karl, S. Kim, D. Ebert, E. Delp, and C. Boushey. "Automatic portion estimation and visual refinement in mobile dietary assessment," Electrical Imaging science and technology, Computational Image VIII, San Jose, CA, Jan 2010.
- [46] O Magazine, Vol. 10, Nr. 2, Feb 2009.

- [47] Jain J., and Bhatti, N. "Imaging-Based Cosmetics Advisory Service," CHI '09 Extended Abstracts on Human Factors in Computing Systems, ACM, New York, NY, 2009.