Vision-Based Magnification of Corneal Endothelium Frames

Dario Comanducci and Carlo Colombo

Computational Vision Group
Dipartimento di Ingegneria dell'Informazione
Università di Firenze – Via S. Marta 3, 50139 Firenze, Italy
{comandu,colombo}@dsi.unifi.it

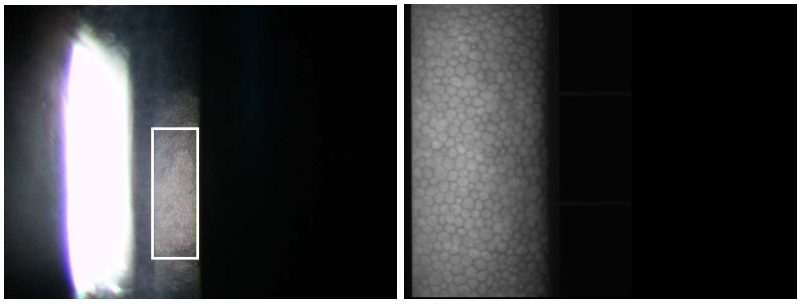
Abstract. We present a fast and effective method to compute a high-resolution image of the corneal endothelium starting from a low-resolution video sequence obtained with a general purpose biomicroscope. Our goal is to exploit information redundancy in the sequence so as to achieve via software a magnification power and an image quality typical of dedicated hardware, such as the confocal microscope. The method couples SVM training with graph-based registration, and explicitly takes into account the characteristics of the application domain. Results on long, real sequences and comparative tests against general-purpose super-resolution approaches are presented and discussed.

Keywords: Biomedical imaging, Machine learning, Super-resolution, Video mosaicing.

1 Introduction

The density and shape of endothelium cells change with age and are related with the health of the cornea; checking them is nowadays a routine diagnostic test, performed either in manual or automatic way. For this purpose, dedicated *confocal microscopes* capable to obtain a high quality image of the endothelium cells, are often used. Nevertheless, the endothelium is also visible, at lower resolution, with less powerful and expensive instruments such as the slit lamp biomicroscope. A typical low-resolution frame obtained with a slit lamp biomicroscope is shown in Fig. 1(a): Only a little portion of the frame contains a visible part of the endothelium area, enclosed by the rectangle in the picture. The white area on the left is due to the corneal reflection of the slit lamp light used to illuminate the cells. For the sake of comparison, Fig. 1(b) shows an image of the endothelium obtained with a confocal microscope. Here the endothelium is visible in better detail, thanks to the higher resolution of the instrument.

The goal of this work is to generate a high-resolution (HR) image (with confocal microscope quality, i.e., a zooming factor of at least $3\times$) of the endothelium cells, given a low-resolution (640×480 pixels) video sequence of the endothelium area obtained with a slit lamp biomicroscope. Coping with low resolution images is practically unavoidable for general *in vivo* analysis with a biomicroscope,



(a) (b)

Fig. 1. (a): A typical endothelium image obtained with the slit lamp biomicroscope. (b): A confocal image of the endothelium.

where a real-time frame rate (30 fps) is required to ensure that the frames are acquired without significant motion blur. As of course the required image quality cannot be achieved by mere bicubic interpolation of a single frame, in this work *all* frames are exploited within a super-resolution (SR) framework based on photo-mosaicing. As shown in the experimental tests section, our solution drastically reduces computations with respect to classical multi-frame SR approaches (see [1] for a general overview, and [2] for a survey on the use of SR methods in medical imaging). More details on theoretical and implementation aspects can be found, together with additional experimental results, in an extensive technical report devoted to this work [3].

2 The Magnification Method

The proposed method can be split into the following steps: (1) Automatic segmentation of the visible endothelium within each frame; (2) Selection of the best endothelium subsequence via a trained SVM; (3) Image alignment and mosaicing of the selected visible endothelium segments; (4) HR image generation by exploiting all the pixels from the aligned images.

Raw endothelium segmentation. Fig. 2 shows the result of endothelium region segmentation for Fig. 1. The endothelium is contained in the image rectangle $[x_l, x_r] \times [y_-, y_+]$, whose extremes are computed by analyzing the shape and slope of cumulative horizontal and vertical histograms of pixel intensity.

SVM-based selection of effective endothelial images. The quality of the HR image depends on the quality of the images contributing to it. Hence, a quality measurement is required in order to select the best images to be used. In order to identify endothelial images with good visual quality, color and texture descriptors compliant with the MPEG-7 standard are used to train an SVM with a radial basis function as kernel. Furthermore, SVM classification is combined

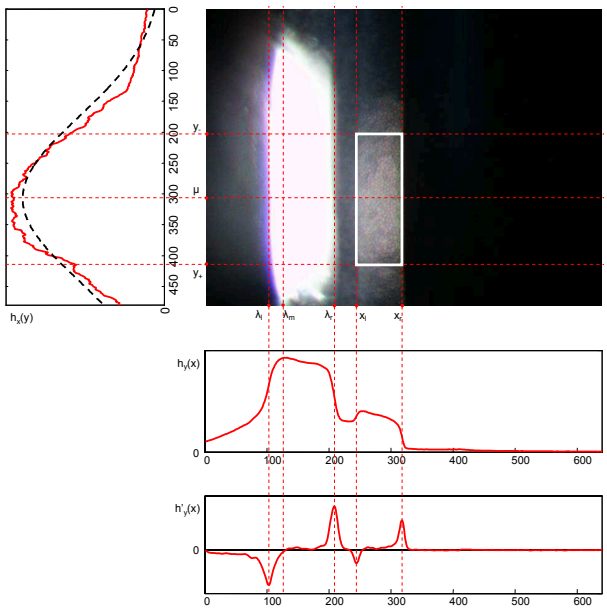


Fig. 2. The cumulative histograms $h_y(x)$ and $h_x(y)$, together with the derivative $h'_y(x)$, computed for Fig. 1(a). **Best viewed in color.**

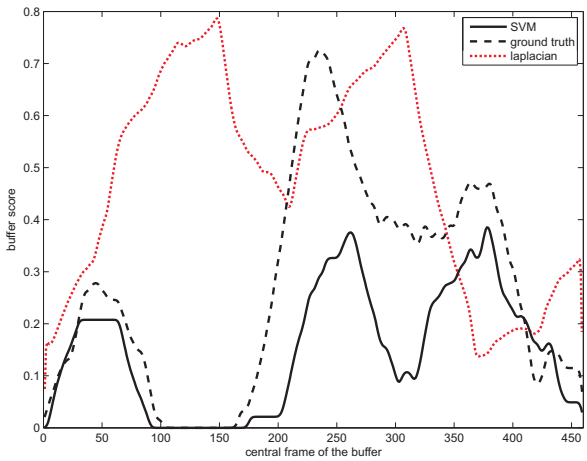


Fig. 3. Score of being a good visual quality subsequence. Bold line: Buffer score based on SVM average probability. Dotted line: Buffer score based only on average Laplacian energy. Dashed line: Buffer score based on ground truth.

with the average energy of *image Laplacian* of the segmented frames. Such an operator, usually employed in digital photography as a pixel-based autofocusing heuristic [4], is used here to identify the most focused images. Indeed, the main factor that can affect the quality of the endothelium image is blur, either due to fast eye/lamp motion or out of focus. The synergy between SVM-based classification and Laplacian-based ranking is exploited since they have complementary strengths and weaknesses. On the one hand, SVMs can discriminate between segments with endothelial content and images where the endothelium is absent, but cannot provide any quality ranking among the images within the same class. On the other hand, the Laplacian operator is a powerful sharpness indicator, but is unreliable when applied to images without endothelium.

SVM classification is employed here to select a buffer of good consecutive frames (i.e. a subsequence) as input to our SR method. In all the experiments presented in this paper the buffer is composed of 60 frames, corresponding (at 30 fps) to 2 seconds of acquisition time. This is a reasonable number of frames since, due to fast eye saccades/lamp jumps, the superposition of the endothelium segments is likely to be lost with larger buffers. Extensive tests show that the best SVM performance (92.58% correct classification in the validation set) is obtained by using just two labels, "useful" and "not useful", and concatenating the CLD (Color Layout), CSD (Color Structure) and HTD (Homogeneous Texture) MPEG-7 descriptors. Frame buffer selection is obtained by averaging the SVM probability of the "useful" frames inside a sliding window running along the segmented sequence. The closer such a buffer score to 1, the higher the chance that the buffer is composed of good images. Fig. 3 shows an example of the buffer score obtained with the SVM classification applied to a video sequence of 470 frames, compared with the same score using manual ground truth. The sequence contains several frames without visible endothelium regions. Although the SVM score is often lower than the ground truth score, local maxima (good buffers) and minima (bad buffers) are located at approximately the same frames of the sequence, thus allowing us to select the best subsequence. When local maxima are comparable, as in the case of Fig. 3, the final decision is taken by looking at the highest average energy of image Laplacian (red dotted curve) of the segmented frames in each subsequence.

Photo-mosaicing. Since the cornea can be regarded as locally planar, photo-mosaicing can be exploited as a way to obtain the HR image of the endothelium region. The resulting mosaic covers the endothelium area visible during the whole scanning session. The quality of the HR image depends on the quality of the images contributing to the mosaic. Again, a quality measurement is required in order to select the best images to use for the purpose of mosaic creation. Since the SVM classification has already found a good subsequence made by (almost) all useful images, the average modulus of the image Laplacian can be directly applied as quality measure. A robust, graph-based implementation of the mosaicing algorithm, akin to the one proposed in [5], is employed, so as to cope with possible outlier frames of class "not useful" that can occur in the subsequence.

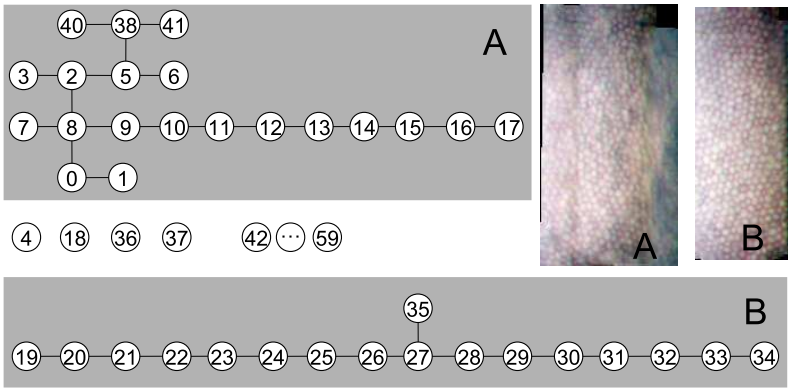
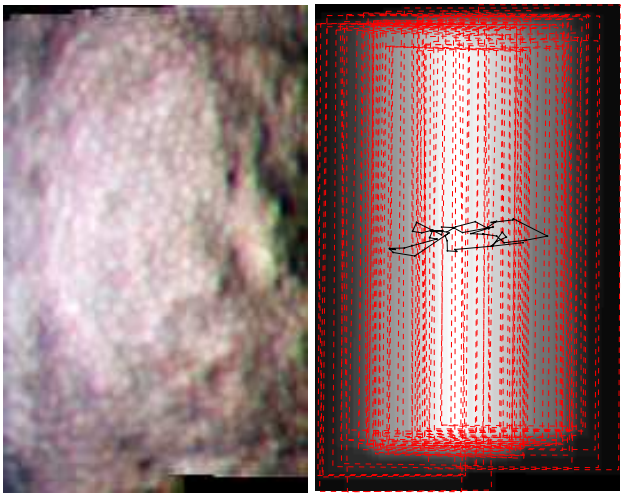


Fig. 4. A case of multiple trees for a sequence of 60 frames. The tiles composing subgraph A have an average quality lower than the quality of the tiles in subgraph B. Hence the latter is selected, even if the former has more nodes.

The algorithms is summarized as follows. First, each segmented image I_k is roughly aligned w.r.t. the next frame I_{k+1} by means of an affine transformation. Since in some frames the endothelium area can be very blurred or not visible at all, the registration of subsequent frames can fail, thus producing several distinct chains of linked images. Each chain is a tree composed by a set of subsequent image nodes. Chains are then merged together to build wider trees. Two chains are merged if an alignment transformation is found between any two nodes belonging to them. Multiple trees arise when it is not possible to merge all the chains in an unique connected graph. When this happens, each tree corresponds to a different mosaic, one of which is chosen to produce the HR image. The criterion for the tree selection is based on the size (i.e. the number of nodes) of the tree and average quality of its images. By default the selected tree is the largest, unless a second one has a better average quality (Fig. 4 shows an example of this situation). The reason for this strategy is that the value of every pixel of the HR image will be estimated on as many samples as possible, and it is convenient that the samples come from low resolution images of good quality. In Fig. 5, the mosaic obtained at the end of the raw alignment step is shown, together with the frame-by-frame apparent motion of the slit lamp with respect to the mosaic.

After the raw, affine registration, a finer image alignment takes place. For this purpose, a node I_r is selected, among the frames of the chosen tree, as root. This root acts as the reference frame of the finer mosaic and all the other images are registered w.r.t. it, according to a full projective warping transformation (2D homography).

Once all the low-resolution (LR) images have been registered w.r.t. the reference frame, the creation of the HR image can start. The HR image is a magnified version of the reference frame. Hence, the transformation $\mathtt{W}_k$ mapping the HR



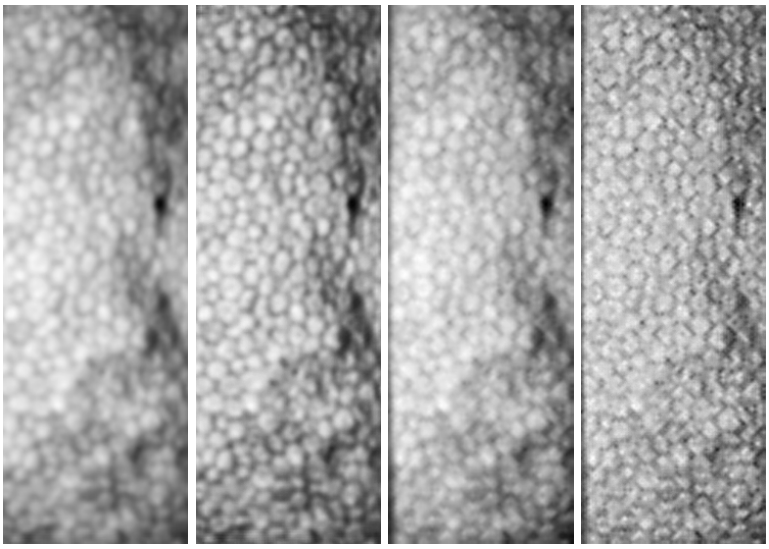
(a) (b)

Fig. 5. (a): The mosaic after raw image alignment. (b): The overlap of all the tiles, and the path made by the slit lamp over the mosaic. The whiter the pixel, the higher the number of overlapping images. Note the sudden jumps that can arise during the scanning session. **Best viewed in color.**

image onto each LR image I_k is obtained as $\mathtt{W}_k = \mathrm{diag}(\rho^{-1}, \rho^{-1}, 1)\mathtt{H}_k^{-1}$, where $\rho > 1$ is the magnification factor and $\mathtt{H}_k$ is the homography mapping I_k onto the reference frame. Our method recovers the HR image J in closed form as a linear combination of several pixels coming from the LR images. The solution can be written as

$$\mathbf{j} = \mathtt{U} \sum_k \lambda_k \mathtt{N}_k \mathbf{i}_k \quad , \tag{1}$$

where vectors $\mathbf{j}$ and $\mathbf{i}_k$ contain (in lexicographic order) the pixel values of the HR and I_k image, respectively. In Eq. 1 the λ_k's are Laplacian weights; each matrix $\mathtt{N}_k$ has size $n_j \times n_k$, while $\mathtt{U}$ is a square matrix of dimensions $n_j \times n_j$, being n_j and n_k the number of elements of $\mathbf{j}$ and $\mathbf{i}_k$ respectively. The matrices $\mathtt{N}_k$ and $\mathtt{U}$ take into account respectively the number of pixels that influence each HR pixel value, and the local amount of unsharp filtering. They can easily be computed as shown in [3]. The effects of matrix $\mathtt{U}$ applied to Fig. 6(a) are visible in Fig. 6(b). For the sake of comparison, Fig. 6(d) shows the effect of unsharp filtering on the image in Fig. 6(c), that was obtained by standard bicubic interpolation of the reference frame. Note that the results of bicubic magnification are much inferior than those obtained with our method. In fact, on the one hand, unlike Fig. 6(b), Fig. 6(d) contains artifacts due to high frequency noise. On the other hand, the bicubic-magnified image contains the same information than the LR frame it comes from, and therefore its fine details are not realistic. Conversely, the high-



(a) (b) (c) (d)

Fig. 6. (a): The first guess for the HR image. (b): The final HR image after sharpening. (c): The bicubic interpolation of the corresponding LR frame after image equalization. (d): image (c) after sharpening; note how the noise is enhanced too.

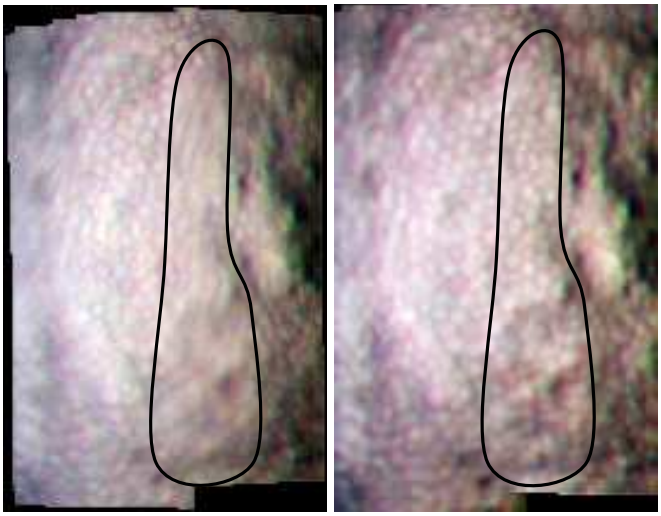
frequency details obtained with our method are realistic, as they summarize the information coming from the whole LR image sequence.

3 Experimental Results

In this section, results of comparative experiments are presented and discussed.

The first experiment compares our ad hoc mosaicing approach against a commercial state-of-the-art software (Photoshop^{TM} CS3). As shown in Fig. 7, both algorithms are able to successfully register almost all the frames containing the endothelium region (in the case of Photoshop, frames already cropped by our segmentation algorithm were provided to the software). The key difference is in the merging step: in fact, the result provided by Photoshop (Fig. 7(a)) has a wide region (enclosed by the black curve) that appears blurred, while in our solution the same region is of good quality. This is due to the fact that our solution merges the images by Laplacian averaging, while Photoshop is less selective. As a result, the quality of the Photoshop mosaic is highly affected even by very few blurred frames. Hence, for the special class of corneal endothelium images, using an information selection criterion based on the relative frame quality (as done in our approach) is undoubtedly beneficial.

A second comparison is made with a standard classical super-resolution method, according to which the HR image is generated by using a standard



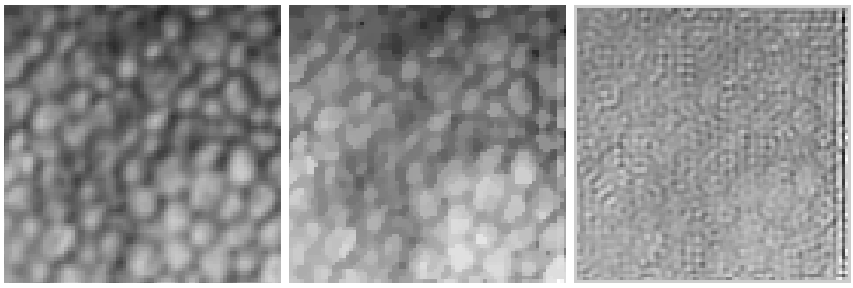
(a) (b)

Fig. 7. Comparison of mosaic images. (a): Photoshop CS3. (b): Our result. The encircled area shows the mosaic portion where Photoshop fails to attain good image quality (see text).

MAP super-resolution framework and exploiting the Huber function as prior [6]. The best MAP result, shown in Fig. 8(b), is clearly of less quality with respect to ours, as it suffers of a posterization effect, that flattens the appearance of the endothelium cells, and makes cell boundaries much less definite. It is likely that the Huber prior, although providing excellent results in other contexts (see again [6]), is less suitable for endothelial images. For the sake of completeness, Fig. 8(c) shows the best MLE solution to the super-resolution problem. As expected from the theory, being the MLE a simple least squares approach, a lot of high-frequency artifacts are generated in this case, thus making this solution totally unuseful for diagnostic purposes.

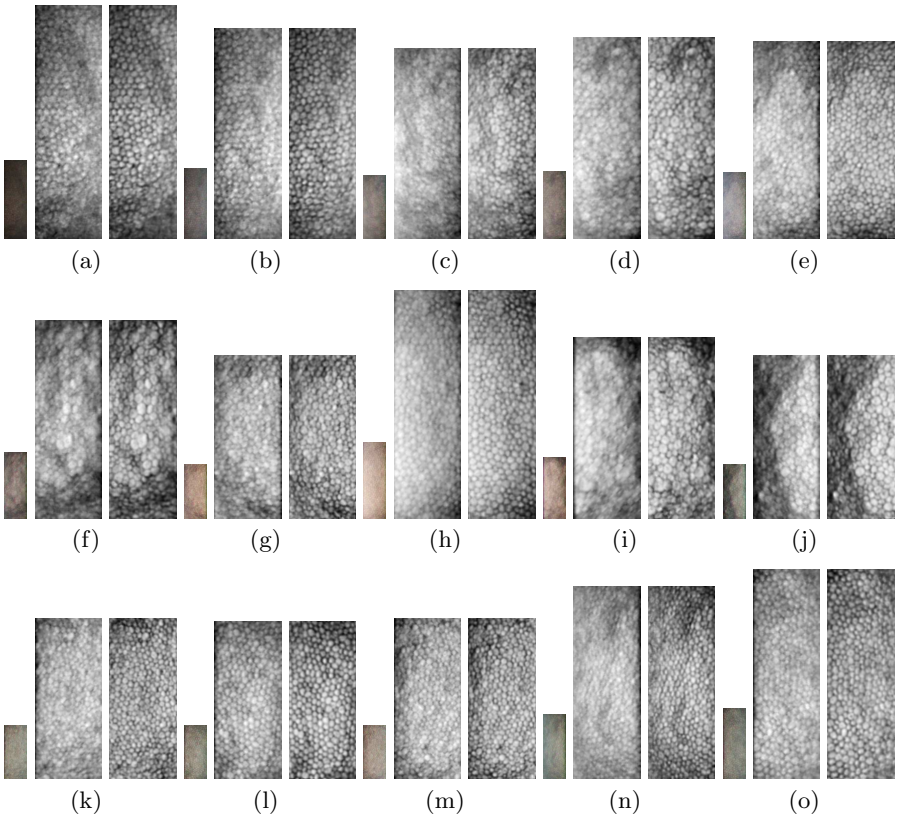
3.1 Gallery

Fig. 9 shows fifteen examples of 3× magnification of 60-frame LR sequences with different subjects and acquisition conditions. For each example are shown: one original LR endothelium image (left), the result of image magnification by bicubic interpolation (middle), and the result obtained with our super-resolution approach (right). For our approach, the average execution time for a single magnification process with an off-the-shelf notebook (processor Intel® core i3 CPU M330 at 2.13GHz) is about 7 seconds. In all cases, the HR image obtained with our method looks very detailed w.r.t. the original image, and of much better quality w.r.t. interpolated image.



(a) (b) (c)

Fig. 8. (a): Our result for the HR image. (b): The best MAP result employing the Huber function as prior. (c): The best MLE result.



(a) (b) (c) (d) (e)

(f) (g) (h) (i) (j)

(k) (l) (m) (n) (o)

Fig. 9. Several examples with magnification factor 3×. Images are of variable size, due to different illumination conditions. For each group of images, from left to right: original endothelium image; enhanced bicubic-interpolated image; HR image obtained with our method.

4 Conclusions and future work

In this work, a fast and efficient method for obtaining good quality magnified images from a low resolution slit lamp biomicroscope was proposed. Compared against classical super-resolution techniques based on multiple images of the same scene, our method produces images of higher quality, as it is specifically tailored to the endothelial image domain.

Future work will address the development of a more general framework suitable for different applications. Indeed most of the computational steps of the pipeline could be adapted to new (medical and not) image domains, after a proper training of SVMs. For the sake of generalization, the ad hoc endothelium segmentation procedure described in section 2 could be replaced by a cascade of classifiers line in [7] to perform endothelium detection. Such classifiers could be trained again to detect the region of interest in each frame if the super-resolution pipeline is applied to a different domain.

Improvements to speed up the registration process will be addressed as well.

References

1. J. van Ouwerkerk, Image super-resolution survey, Image and Vision Computing 24 (10) (2006) 1039–1052.
2. H. Greenspan, Super-resolution in medical imaging, The Computer Journal 52 (1) (2009) 43–63.
3. D. Comanducci, C. Colombo, A Super-Resolution Based Pipeline for the Magnification of Corneal Endothelium Images. Technical Report #CVG-10/2011, Università di Firenze, Dipartimento Sistemi e Informatica, October 2011. (Submitted for publication.)
4. K.-C. Ng, A. N. Poo, M. H. Ang, Practical issues in pixel-based autofocusing for machine vision, in: Proc. IEEE International Conference on Robotics and Automation, 2001.
5. R. Marzotto, A. Fusiello, V. Murino, High resolution video mosaicing with global alignment, in: Proc. Conf. Computer Vision and Pattern Recognition, 2004.
6. D. Capel, Image Mosaicing and Super-Resolution (Cphc/Bcs Distinguished Dissertations.), SpringerVerlag, 2004.
7. P. Viola, M. Jones, Robust real-time face detection, International Journal of Computer Vision 57 (2) (2004) 137–154.