

Super-Resolution Based Magnification of Endothelium Cells from Biomicroscope Videos of the Cornea

Dario Comanducci^a, Fabio Bellavia^b, Carlo Colombo^{b,*}

^aEB Neuro S.p.A., Firenze, Italy

^bComputational Vision Group (CVG), Dept. of Information Engineering, University of Florence, Firenze, Italy

Abstract. We present a practical, robust and effective pipeline to compute a high-resolution image of the corneal endothelium starting from a low-resolution video sequence obtained with a general purpose slit lamp biomicroscope. An image quality typical of dedicated and more expensive confocal microscopes is achieved via software magnification by exploiting information redundancy in the video sequence. In particular, the high-resolution image is generated from the best low-resolution frames, obtained by identifying the most suitable endothelium video subsequence using a SVM-based learning approach, followed by a robust graph-based frame registration.

Results on long, real sequences show that the proposed approach is fast and produces better quality images than both classical multi-frame super-resolution approaches and commercial state-of-the-art mosaicing software. Only low-cost equipment is required, that makes the proposed method a valid diagnostic tool and an affordable resource for medical practice in both developed and developing countries.

Keywords: super-resolution, corneal endothelium, slit lamp biomicroscope image enhancement, machine learning, image mosaicing.

*Carlo Colombo, carlo.colombo@unifi.it

1 Introduction

The corneal endothelium is a monolayer of specialized, flattened cells delimiting the posterior surface of the cornea and facing the anterior chamber of the eye. The density and shape of endothelium cells change with age and are related with the health state of the cornea. In-vivo analysis of endothelium is nowadays a routine diagnostic test, performed either in manual or automatic way, throughout dedicated *confocal microscopes* capable to obtain high quality images. Nevertheless, the endothelium is also visible, at lower resolution, with more inexpensive instrument such as the slit lamp biomicroscope.

To observe the endothelium with a slit lamp biomicroscope, the specular reflection technique is employed (see Fig. 1). The goal of this work is to generate high-resolution (HR) images of the endothelium cells with confocal microscope quality from a low-resolution (LR) video sequence

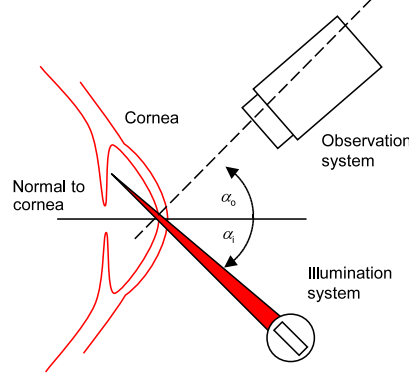


Fig 1 The operation principle of the specular reflection technique used in the slit lamp biomicroscope. This can only be achieved when the angle α_i of incident illumination to the normal of the corneal surface equals the angle α_o of the observation system.

of about 640×480 pixels of the endothelium obtained with a slit lamp biomicroscope. Additionally, low-resolution image acquisition grants real-time frame rates (30 fps or more), avoiding the presence of significant motion blur in the frames, due to the involuntary micro-movements of the patient's eye. This results in apparent very rapid Brownian camera movements over the corneal endothelium (see e.g. 7(c) later in the paper), causing blurring effects that would strongly affect the in-vivo analysis.

Typical low-resolution images obtained with a slit lamp biomicroscope are shown in Fig. 2. Images were acquired by an expert technician with an ad hoc setup, so as to get the best quality in term of visibility and clarity of the relevant content. Only a little portion of the frame contains a visible part of the endothelium area (green rectangle in Fig. 2(a)), that for some lamp/camera configurations can be barely visible (see Fig. 2(c)), making the frame unusable. Notice also that a relevant portion of the images can be occupied by the vertical illumination band produced by the lamp direct illumination onto the cornea (see Figs. 2(a),(c)). A single and even well-taken LR image from a slit-lamp biomicroscope cannot get a detail level comparable to a HD image of the corneal endothelium obtained with a confocal microscope (see Fig. 2(d)), due to the higher reso-

lution of the latter instrument. Nevertheless, we can apply super-resolution and photo-mosaicing techniques on several LR images, obtaining comparable results with cheaper equipment, making the proposed approach an affordable and valid diagnostic tool for both developed and developing countries.

The rest of the paper, anticipated in part in a previous work,¹ is organized as follows. Section 2 describes related work, while the details of the proposed approach can be found in Sec. 3. Finally, Sec. 4 reports the experimental results, while conclusions and future work are addressed in Sec. 5.

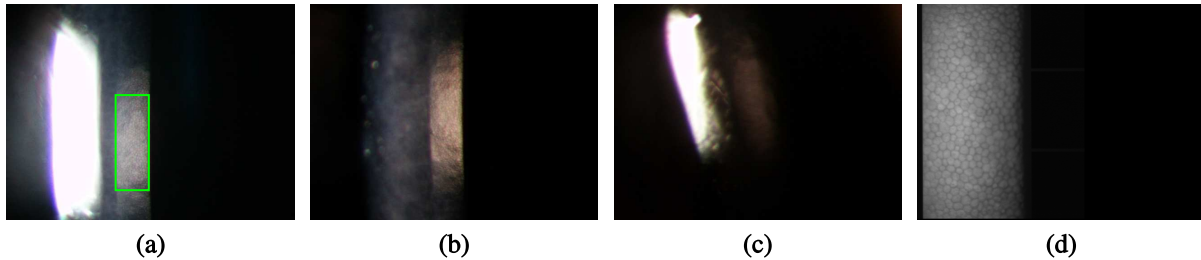


Fig 2 A typical endothelium image obtained with the biomicroscope: (a) Both the lamp corneal reflection and the endothelium are visible. (b) Only the endothelium is visible. (c) The endothelium is not visible. (d) An example of confocal image of the endothelium.

2 Related Work

The goal of super-resolution (SR) methods is to recover a HR image from one or more LR input images. SR is a ill-posed problem, since many HR images can produce the same LR image when downsampled as the magnification factor increases.² However this issue can be circumvented by imposing additional constraints to stabilize the solution, that depend on the data and the model. In this sense, SR methods rely upon the nature of the acquisition process they are developed for. SR has been successfully employed in medical imaging,³ such as for X-ray computer tomography (CT), magnetic resonance imaging (MRI) and positron emission tomography (PET). As reported in recent surveys,^{4,5} the majority of current SR methods on optical images works on the spatial

domain, and can be roughly divided in two classes according to the employment of a single image or multiple images as input, respectively.

In the former case, non-linear interpolation method⁶ or learning algorithms, going from the Bayesian learning approach⁷ to the more recent Deep Neural Networks,⁸ are generally used to try to hallucinate (i.e., to estimate from a general a-priori data knowledge plus the image content) the missing information of the super-resolved images,⁹ exploiting in some cases the relationship between LR and HR images from a training database of image patches,^{10,11} that can also be made up by the input LR image only.¹²

Several multi-image SR methods exist. Such methods generally assume that the input LR images have some relative geometric and photometric displacements that are exploited to reconstruct the original HR image given an image formation model, usually of the form

$$I_k(\mathbf{x}) = D(b(\mathbf{x}) * J(w_k(\mathbf{x}))) + n_k(\mathbf{x}) \quad (1)$$

Here, the LR images I_k are obtained by downsampling with the operator $D(\cdot)$ the blurred and warped HR image J with an added noise $n_k(\mathbf{x})$ for each pixel $\mathbf{x}$. Pixel warping is given by the function $w_k(\mathbf{x})$ and blur is obtained from the convolution $*$ with a kernel image b . Under this assumption, SR becomes an inverse problem that can be handled by using several approaches according to the model specifications.

Image registration assumes a critical role in the development of multi-image SR algorithms. While in the case of small displacements dense optical flow methods can be applied,¹³ robust approaches based on keypoint matching are needed where more severe geometric transformations are present.¹⁴

Direct multi-image SR methods simply upscale the registered LR images and then merge them into the output HR image. Different methods can be used to get the pixel value in the HR image starting from the corresponding pixel values in the LR images. Local mean or median filters,¹⁵ also on a weighted window,¹⁶ and AdaBoost¹⁷ are examples of such methods. Deblurring algorithms can be applied before the image fusion¹⁸ or on top of the HR output image.¹⁹ Starting from an initial solution provided by a direct method, the HR output can be further refined using iterative back-propagation,^{20,21} by repeatedly minimizing the error between the input LR images and those simulated from the current HR image according to the model.

A similar error minimization scheme can be formulated in terms of Maximum Likelihood Estimation and refined into a Maximum a Posteriori problem,²² when a regularization factor that models the a priori knowledge on the image is added. In particular, Markov Random Fields²³ have been used to model image neighborhood, further specialized through the Huber function to preserve sharp edges.²⁴

Similarly to single image SR methods, example-based SR techniques using a training database on multiple images exists as well.²⁵ The effectiveness of these techniques highly depends on the supporting image database, especially as magnification increases.²⁶ It is worth noting that, unlike the previous multiple image SR approaches, the hallucinated HR details obtained by these techniques are not guaranteed to correspond to the true unknown details. This represents a serious limitation for medical imaging, where high-resolution details are often used for diagnostic purposes.

3 Proposed Method

Our SR pipeline, can be divided into three main computational steps:

- The selection of the best endothelium sequence and of the proper ROI inside the image frame;
- A coarse-to-fine image registration and alignment of the visible endothelium areas through a robust graph-based approach and global error minimization;
- The generation of the HR image from the LR aligned images.

3.1 Endothelial Image Selection

3.1.1 ROI Selection

The goal of this step is a coarse detection of the region of interest (ROI) within each video frame of the input sequence that includes the endothelium cells. As shown in Fig. 3, it is assumed that the slit lamp is almost parallel to the image frame y-axis, that is the standard operative scenario for this specific application. Under this setup, the column intensity profile H_y is computed by cumulating the 8-bit intensity values inside each image column x . Assuming as requirement that the lamp is visible if half of the pixels inside the brightest intensity column are strongly illuminated (with values greater than 230, the remaining pixels in the worst case have zero value), this means that it must hold that $H_y(m) > t$, where m is the position of the maximum peak of H_y , $t = 230 * \frac{h}{2}$ and h is the frame height. In this case, one proceeds to compute the position l of the lowest peak of the derivative H'_y of H_y to the left of m and the position r of the highest peak of H'_y to the right of m , that define the column range of the pixel directly illuminated by the lamp. Otherwise, the lamp is assumed to be not visible and $r = 0$. The ROI column range $[x_l, x_r]$ can be found by identifying again the leftmost lowest peak x_l and the rightmost highest peak x_r of H'_y such that $x_l, x_r > r$. Restricting to the column range $[x_l, x_r]$ only, the row intensity profile H_x is computed

and a Gaussian distribution with mean μ and standard deviation σ is fitted into the histogram. The ROI row range $[y_-, y_+]$ is then defined as $y_- = \mu - c\sigma$ and $y_+ = \mu + c\sigma$, where $c = 0.8$ was found experimentally.

Figure 4 shows an example of the automatic ROI selection on a typical input video sequence. Notice the presence of the right red band, an extra noisy area not occurring in the standard cases depicted in Fig. 2(a), probably due to spurious reflections from the iris. Notice also that ROI selection is robust enough to properly segment all but just frames 42 and 43, while the absence of endothelium region in frames 12 and 45 will be handled in the next step of the pipeline.

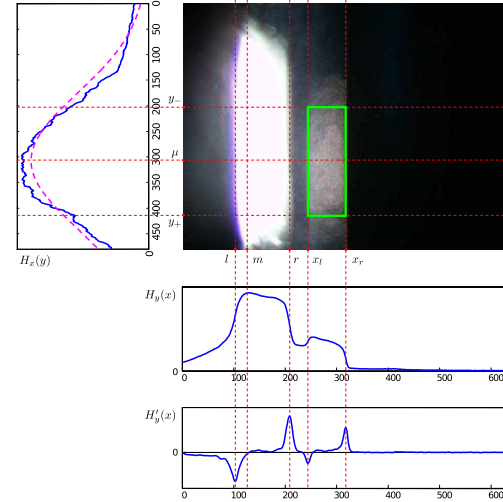


Fig 3 The intensity profile H_y and H_x , together with the derivative H'_y , for the image frame shown in Fig. 2(a). Gaussian fit of H_x is represented by a dashed curve (best viewed in color).

3.1.2 Best Subsequence Extraction

In this step a buffer of good consecutive frames is selected as input to the SR method, since the final quality of the HR image strictly depends on the input LR images. The main factor that can affect the quality of the endothelium image is blur, either due to fast eye/lamp motion or to out of focus configurations.



Fig 4 Automatic ROI selection on frames of a typical input video sequence. The dashed green rectangle shows the endothelium ROI, frame resolution is 640×480 pixels (best viewed in color).

Our choice is based on the combination of two different quality criteria. Each image ROI is processed by a Radial Basis Function (RBF) SVM trained on MPEG-7 descriptors and by the average image Laplacian. The synergy between SVM-based classification and Laplacian-based ranking is exploited due to their complementary strengths and weaknesses. On one hand, SVM can discriminate the presence of endothelial content in the image frame, but cannot provide a measure of the blur present in the image. On the other hand, the Laplacian filter is a powerful sharpness indicator employed in digital photography as a pixel-based autofocus heuristic,^{27,28} but it is unreliable when applied to images without endothelium (in that case the image shows a dark background with spurious spot lights, providing a high Laplacian score). For the SVM classification we used the LibSVM implementation with class probability output instead of binary labels.²⁹ Different MPEG-7 color and texture descriptors,³⁰ including Color Layout (CLD), Color Structure (CSD), Scalable Color (SCD), Edge Histogram (EHD), Homogenous Texture (HTD) were tested. As shown in Sec. 4, the best SVM results (about 92.6% of correct classification) is achieved by concatenating the CLD, CSD and HTD descriptors.

Within the input sequence, the best continuous subsequence of 60 frames was considered, corresponding to 2 seconds at the standard acquisition rate (30 fps). This is a reasonable choice, since the superposition of the endothelium segments can be lost as the acquisition time increases, due to the fast Brownian eye/lamp jumps occurring in the images. Using a sliding window of 60 frames running along the whole sequence, the average SVM probability and the mean Laplacian energy are computed for frames inside the window positively classified by the SVM (i.e. with a probability greater than 0.5). Finally, among the subsequences that correspond to a local maximum for the average SVM probability, the one that gives the best mean Laplacian score is selected.

Figure 5 shows an example on an typical input video sequence of 470 frames (i.e. about 15

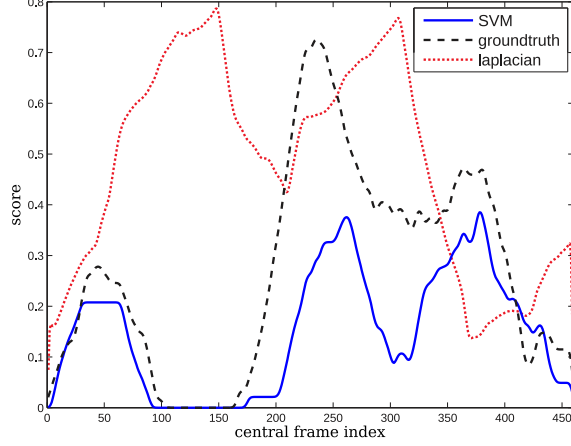


Fig 5 Best subsequence selection criterion. Average SVM and ground-truth probabilities and mean Laplacian energy are respectively plotted in blue, black and red (best viewed in color).

seconds). A manually labeling was done to provide a SVM classification ground-truth. Positive (negative) labeled frames correspond respectively to a 1 (0) probability. Notice that subsequences with frame index less than 200 have a high Laplacian score but no visible endothelium regions. Although the SVM score is often lower than the ground truth score, local maxima (good candidates) and minima (bad choices) are located at approximately the same frames of the sequence. As a result, the best subsequence is around frame index 250 both the SVM and ground-truth. The frames of the final selected sequence are those reported in Fig. 4.

3.2 Image Registration

3.2.1 Raw Image Alignment

Registration techniques taken for image mosaicing are used here, as for our purpose the cornea can be regarded as locally planar.

Referring to Fig. 4, the image ROI of frame I_k is registered with respect to the successive frame I_{k+1} using RANSAC matching between SIFT keypoints to estimate the affine transformation mapping the two frames³¹ (see Fig. 6(a)). Registration between two consecutive frames can fail

if the number of correct matches detected is below a given threshold, due to very blurred or not visible endothelium ROI. As a results, several chains of linked frames are produced, depicted as gray rectangles in Fig. 6(b). Finally, if a successful registration between frames belonging to different chains is possible, the corresponding chains are merged so that the initial image graph of Fig. 6(b) is created, where dashed edges merge the distinct chains.

Multiple trees arise when it is not possible to merge all the chains into a unique connected graph. In this case we choose between the two trees with the highest number of nodes the one with the highest mean average Laplacian energy. The reason for this strategy is that the value of the pixels of the HR image will be estimated on as many samples as possible, so it is convenient that the samples come from LR images of good quality. Figure 7(a)-(c) shows the mosaic obtained from the single tree of Fig. 6(b), together with the frame-by-frame apparent motion of the slit lamp with respect to the mosaic, while in Fig. 7(d) a case with multiple trees is presented.

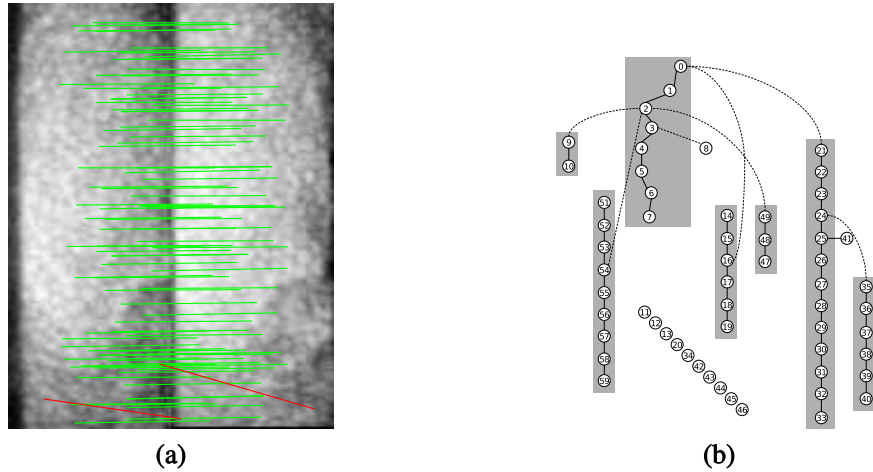


Fig 6 (a) SIFT correspondences between frames 0 and 1 of Fig. 4. RANSAC inliers and outliers are respectively shown in green, red (best viewed in color). (b) The initial graph built for the video sequence of Fig. 4. Nodes inside the gray rectangles represent the initial chains, while dashed edges correspond to the successive merging steps. Singleton nodes represent to frames not linked to any other frame.

3.2.2 Refinement

The initial alignment is further refined to improve the result. A node I_r is selected among the frames of the tree chosen as root and is used as reference for registering the other frames according to a projective transformation. Defining the ROI center of a frame as the intersection of the two diagonals of the quadrilateral in which a ROI is warped (see again Fig. 7(b)), I_r corresponds to the frame whose ROI center is the closest to the mosaic center, the latter being defined as the mean of the ROI centers of the frame composing the tree, weighted according to the average Laplacian energy of its frames.

New edges are then added incrementally, by linking frames with significant overlap as described in.³² The overlap measure between nodes I_i and I_j is defined as

$$\delta_{ij} = \frac{\max(0, |\mathbf{c}_i - \mathbf{c}_j| - |d_i - d_j|/2)}{\min(d_i, d_j)} \quad (2)$$

where $\mathbf{c}_i$, $\mathbf{c}_j$ and d_i , d_j are respectively the ROI centers and the longest diagonal lengths of the ROI quadrilaterals corresponding to nodes I_i and I_j after being projected onto the mosaic. A value $\delta_{ij} > 1$ corresponds to no frame overlap, and the diameter is the main diagonal of the warped image. A graph G corresponding to the tree is built, where the edge weights are the overlap values δ_{ij} . Using a greedy strategy, in turn a new link is added, corresponding to the minimum of

$$\gamma_{ij} = \frac{\delta_{ij}}{\Delta_{ij}} \quad (3)$$

among all the possible nodes pairs (i, j) , where Δ_{ij} is the current shortest path between nodes I_i and I_j on G . The range of γ_{ij} is $[0, 1]$, so that γ_{ij} approaches 0 as the edge (i, j) is a significant

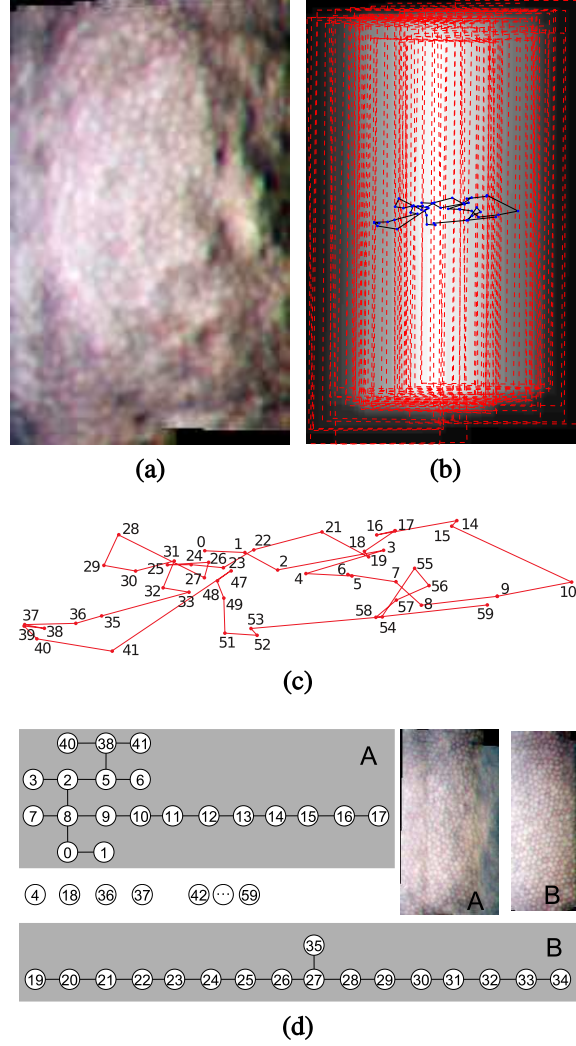


Fig 7 (a) The mosaic after the first image alignment. (b) The overlap of all ROIs, and the path followed by the slit lamp over the mosaic. Brighter regions correspond to a higher number of overlapping images, red parallelograms correspond to the warped ROIs onto the mosaic, and the points of the path are the frame ROI centers. (c) A better view of the path performed by the slit lamp onto the endothelium area, numbered according to the corresponding frames. Notice the sudden jumps from node 53 to 54 arisen during the scanning session (best viewed in color). (d) A case with multiple trees. The frame ROIs composing subgraph A have an average quality lower than the that of subgraph B. The latter is selected, even if the former has more nodes. Note that the mosaic of subgraph A do not overlap the mosaic of subgraph B. Frame 18, the missing node that would have connected subgraph A with B is very motion-blurred, because of the fast movement of the lamp jumping from a portion of the endothelium onto another one.

shortcut between I_i and I_j . The shortest path Δ_{ij} values are then updated according to the new link added to G , and the process is repeated until the δ_{ij} and Δ_{ij} corresponding to the candidate link to add are below given thresholds.

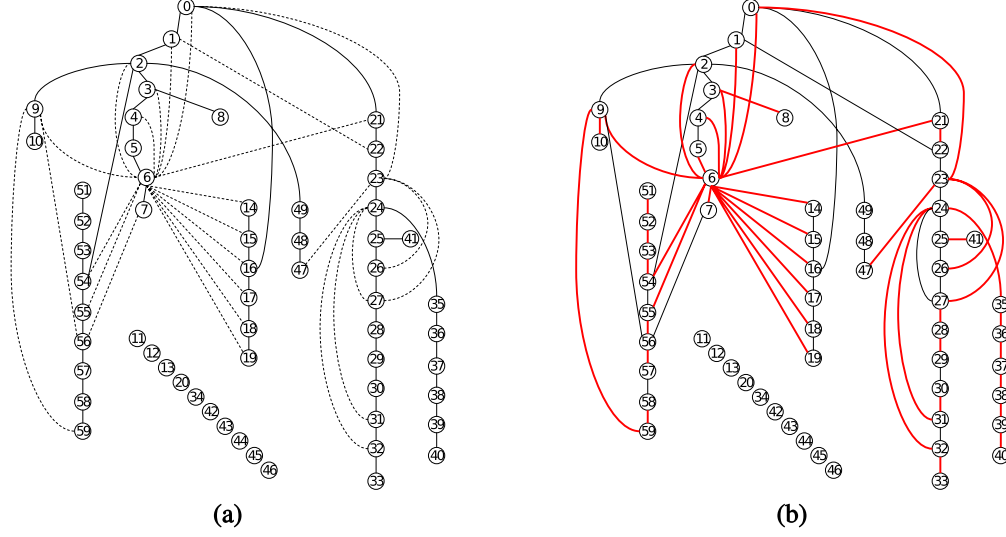


Fig 8 The extended graph. (a) Dashed edges are the new added edges. (b) The shortest path tree (bold red edges) having I_r as root, built on the extended graph. The weights are the average symmetric transfer errors of the affine maps registering the pairwise connected images (best viewed in color).

After these shortcuts are added (see Fig. 8(a)), each edge (I_i, I_j) of the expanded graph is weighted with the average symmetric transfer error³³

$$\frac{1}{|N_{ij}|} \sum_k \| \mathbf{m}_{ki} - \mathbf{H}_{ij} \mathbf{m}_{kj} \| + \| \mathbf{m}_{kj} - \mathbf{H}_{ij}^{-1} \mathbf{m}_{ki} \| \quad (4)$$

for the affine transformation $\mathbf{H}_{ij}$ between frames i and j , where k spans over the set N_{ij} of RANSAC SIFT keypoint correspondences $\mathbf{m}_{ki}$ and $\mathbf{m}_{kj}$ between the two considered frames. The transformation $\mathbf{H}_i$ registering each node I_i with respect to I_r is then computed by concatenating all the affine maps along the weighted shortest path connecting nodes I_i and I_r in the extended graph of Fig. 8(b). The final global registration is refined using Levenberg-Marquardt, by minimizing the cost³⁴

$$\min_{\mathbf{H}_i} \sum_{(i,j) \in L} \sum_k \mathcal{E}_{kij}^2 \quad (5)$$

where L is the set of edges in the tree made by chains (see again Fig. 6(b)) and

$$\mathcal{E}_{kij} = ||\mathbf{m}_{ki} - \mathbf{H}_i^{-1}\mathbf{H}_j\mathbf{m}_{kj}|| \quad (6)$$

Notice also that, in order to obtain a finer registration required for a good quality image magnification, the initial affine transformations $\mathbf{H}_i$ (6 degrees of freedom) are replaced by general homographies (8 degrees of freedom) in the minimization, to compensate and adjust towards the piecewise-planar approximation of the shape of the cornea.

3.3 HR Image Generation

The HR image J is obtained by merging the partial images $\tilde{I}_k$, resulting by warping the ROIs I_k , into a smooth preliminary estimate $\tilde{J}$ and then applying an unsharp filter as follows. For a pixel $\mathbf{p} \in J$

$$\tilde{J}(\mathbf{p}) = \sum_k \lambda_{k,\mathbf{p}} \tilde{I}_k(\mathbf{p}) \quad (7)$$

where the weight λ_k is the mean Laplacian energy of I_k , normalized so that $\sum_k \lambda_{k,\mathbf{p}} = 1$ and $\lambda_{k,\mathbf{p}} = 0$ if $\mathbf{p} \notin \tilde{J}$. Defining by $W_k(\mathbf{p})$ the map from a pixel $\mathbf{p} \in J$ to the corresponding aligned image ROI I_k through matrix $\text{diag}(\rho^{-1}, \rho^{-1}, 1)\mathbf{H}_k^{-1}$ in homogeneous coordinates, where ρ is the magnification factor, $\tilde{I}_k$ is given by

$$\tilde{I}_k(\mathbf{p}) = \sum_{\mathbf{p}' \in I_k} M_{\mathbf{p},\mathbf{p}'} I_k(\mathbf{p}') \quad (8)$$

Specifically, a pixel in $\tilde{I}_k$ is obtained by a weighted sum of the contributions from the corresponding neighborhood in the LR image ROI I_k , where

$$M_{\mathbf{p}, \mathbf{p}'} = \begin{cases} \frac{1}{z_{\mathbf{p}}^k \|W_k(\mathbf{p}) - \mathbf{p}'\|} & \text{if } \|W_k(\mathbf{p}) - \mathbf{p}'\|_{\infty} \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

weighs the neighborhood of $W_k(\mathbf{p})$ on I_k by the inverse distance to it, being $\|\cdot\|_{\infty}$ the Maximum norm and $z_{\mathbf{p}}^k = \sum_{\substack{\mathbf{p}' \in I_k, \\ \|W_k(\mathbf{p}) - \mathbf{p}'\|_{\infty} \leq 1}} \frac{1}{\|W_k(\mathbf{p}) - \mathbf{p}'\|}$ a normalization factor. An example of $\tilde{J}$ is shown in Fig. 9(b). Thanks to the contributions from all the pixels, $\tilde{J}$ is a slightly blurred, low noise image. Since image noise is low, the high-frequency blurred details can be enhanced by unsharp filtering³⁵ without significant noise magnification, so that the final HR image J is obtained by

$$J = \tilde{J} + \mu_s \left(\tilde{J} - (b_{\sigma_s} * \tilde{J}) \right) \quad (10)$$

where b_{σ_s} is a Gaussian blur kernel with standard deviation σ_s , $*$ indicates the convolution operator and μ_s is a parameter affecting the amount of sharpening. Notice that although the magnification can be applied to the whole registered mosaic, for the best result only the ROI of the reference frame I_r is taken into account, since this frame generally provides the best overlap with the other frames as shown in Fig. 7(b), thus maximizing the detail improvements due to the contribution from the other frames.

The effects of the unsharp filter are visible in the final HR image of Fig. 9(b). For the sake of comparison, Fig. 9(e) shows the effect of unsharp filtering on the image in Fig. 9(d), obtained by standard bicubic interpolation of the reference frame I_r . Note that the results of bicubic magnification are much inferior than those obtained with our method. Figure 9(d) contains the same

information of the LR I_r frame in Fig. 9(a) without including finer details so that, unlike the case of unsharpening Fig. 9(c), Fig. 9(e) only adds artifacts due to high frequency noise. Conversely, the high-frequency details obtained with our method are more realistic, as they summarize the information coming from the whole LR image sequence.

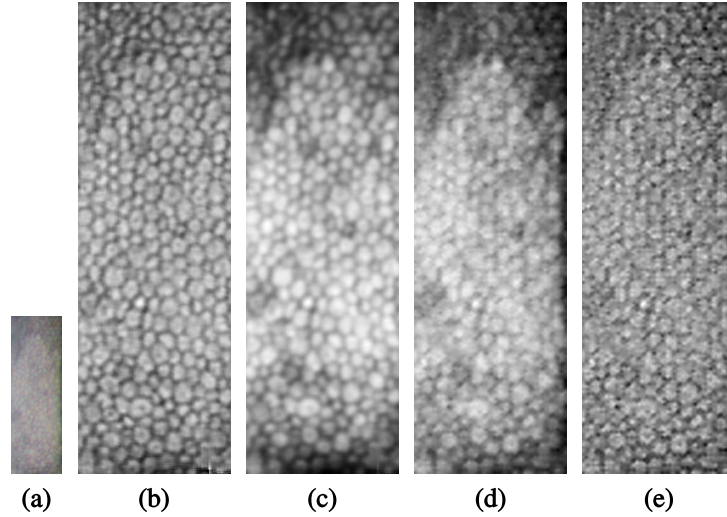


Fig 9 (a) The original endothelium LR image of size 60×208 pixels. (b) The proposed HR image J for a $3 \times$ magnification factor. (c) The blurred image $\tilde{J}$ before unsharpening as output of the standard mosaicing SR. (d) The bicubic interpolation of the corresponding LR frame for a $3 \times$ magnification factor. (e) Image (d) after unsharpening. Notice how in this last case the noise is amplified too.

4 Experimental Results

4.1 Endothelium Quality

In this section, results of several experiments performed on manually annotated datasets are reported. Results corroborate our choice of combining SVM classification and Laplacian ranking for the purpose of selecting the best endothelial images to be used in the super-resolution framework.

4.1.1 SVM-based Classification

The first experiment concerns the performance of the selected MPEG-7 descriptors (CLD, CSD, SCD, EHD, HTD) to train an SVM²⁹ in order to evaluate the visual quality of endothelium images. All the possible combinations of the descriptors (stand alone and in groups of two, three, four, and five) are considered.

The examples come from 20 video sequences acquired in different conditions and with different subjects. The sequences were segmented with the approach of Sec. 3.1.1 providing 2310 ROIs, manually classified as positive (1082) and negative (1228) examples. The example images were uniformly split into the training and validation datasets. The SVM output of other 2010 segmented frames coming from different, manually annotated, video sequences was checked as well so as to simulate the SVM classification at runtime. A multiclass 4-level labeling was also tested with classes going from 0 to 3 (from the worst to the best cases), with respectively 603, 625, 507 and 575 examples. Figure 10 show some representative endothelium ROI for each label.

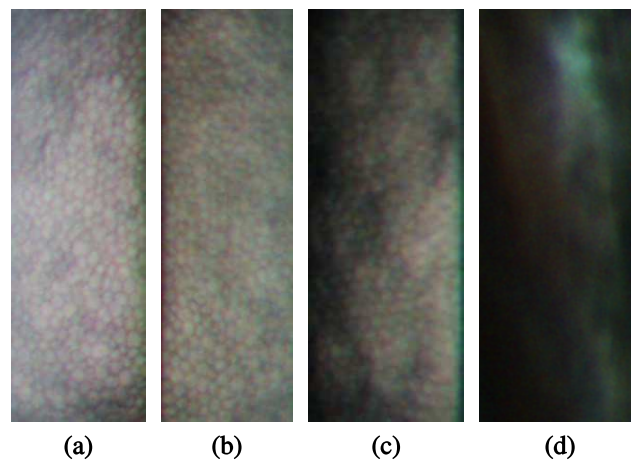


Fig 10 Examples of the four quality labels with decreasing level from left to right (best viewed in color).

As reported in Table 1, evaluation accuracy tends to increase with the number of descriptors, since there is more information in the features passed to the SVM. Anyway, the performance of

the SVM for the multiclass classification is not good enough for our purposes. On the other hand, in the case of just two labels, there are several cases with rates higher than 90%, both in the validation and runtime sets. Among them, we selected the combination CLD+CSD+HTD (bold), providing in the case of two classes the highest accuracy (91.86%) and precision (76.29%) in the runtime set and the second higher accuracy (92.58%) in the validation set, since for the application purpose discarding bad frames is more critical. The finer multiclass classification allows us to better understand the behavior of the SVM trained for just two classes: just 0.7% of images with the lowest quality (0 label) are classified as useful images (2-3 labels), and similarly good images (label 3) are considered as not useful (0-1 labels) only in 3.9% of the cases. Although the misclassification ratio is higher in the latter case, this is not a major drawback, since it is preferable to miss some good frames instead than selecting bad images. Crossover errors in the intermediate classes, with sufficient or low quality, are likely unavoidable.

Table 1 SVM classification based on MPEG-7 descriptors. Descriptor combinations are listed in the first column, the other columns provide evaluation results for the validation and runtime rate both for two or four quality labels

MPEG-7 Descriptors	Accuracy Validation (2 labels)	Accuracy Runtime (2 labels)	Accuracy Validation (4 labels)	Accuracy Runtime (4 labels)	Precision Runtime (2 labels)	Recall Runtime (2 labels)
CLD	76.08%	70.86%	60.53%	47.55%	21.85%	45.98
CSD	92.11%	82.61%	77.51%	60.60%	67.02%	55.92
SCD	80.86%	83.30%	67.22%	56.10%	35.41%	38.34
EHD	87.80%	82.89%	74.40%	54.05%	38.54%	89.17
HTD	89.47%	88.87%	74.88%	65.31%	61.12%	67.89
CLD+CSD	89.71%	89.16%	76.56%	64.20%	67.61%	60.89
CLD+SCD	80.62%	80.32%	65.79%	55.58%	30.48%	32.73
CLD+EHD	88.28%	81.36%	76.32%	55.54%	39.77%	93.88
CLD+HTD	90.19%	84.83%	73.92%	60.51%	42.75%	74.77
CSD+SCD	90.19%	80.51%	76.32%	65.56%	70.97%	60.12
CSD+EHD	89.47%	88.12%	76.79%	62.22%	52.80%	87.64
CSD+HTD	91.39%	89.91%	78.71%	66.45%	71.74%	77.32
SCD+EHD	90.19%	88.01%	79.19%	62.13%	50.75%	90.31
SCD+HTD	90.67%	88.20%	76.32%	65.54%	60.35%	64.20
EHD+HTD	91.63%	84.23%	78.71%	59.32%	46.45%	91.84
CLD+CSD+SCD	89.00%	88.25%	75.12%	67.42%	66.50%	53.88
CLD+CSD+EHD	90.67%	86.56%	77.75%	62.41%	50.58%	88.15
CLD+CSD+HTD	92.58%	91.86%	77.99%	70.30%	76.29%	69.29%
CSD+EHD+HTD	91.39%	88.59%	78.95%	63.97%	57.36%	91.33
CSD+SCD+HTD	92.82%	90.58%	78.95%	72.20%	73.53%	64.07
SCD+EHD+HTD	91.63%	87.58%	80.62%	63.94%	54.52%	91.33
CLD+CSD+SCD+EHD	90.67%	88.16%	78.95%	65.63%	56.75%	85.60
CLD+CSD+SCD+HTD	92.34%	91.16%	77.51%	67.66%	75.30%	55.92
CLD+CSD+SCD+EHD+HTD	91.63%	88.51%	80.14%	66.69%	60.36%	83.82

4.1.2 Laplacian-Based Ranking

In order to evaluate the quality of each frame of an endothelial sequence, four pixel-based auto-focusing techniques^{27,28} were tested. In particular, manually annotated epithelium content divided into four increasing labeled classes (“very bad”, “low quality”, “medium quality” and “very good”) of 70 video sequences of 60 frames, acquired in different conditions, was related through the Spearman Rank correlation coefficient with SMD, variance, Laplacian and Tenengrad operators considered as auto-focusing measures,²⁷ obtaining respectively a value of 0.58, -0.13, 0.63 and 0.5, yielding the best results for the Laplacian operator. The manual annotation was limited to images with endothelium only, since SVM classification provides a good frame selection, thus leaving the auto-focusing operator behavior to be evaluated only in presence of endothelium content. The plot of the Laplacian operator values for a subset of the considered frame sequences is shown in Fig. 11. The figure shows that the Laplacian almost invariably gives good results if considered as a relative measure within the set of frames of each sequence. Therefore, although it cannot provide an absolute quality measure because it is sensitive to the different acquisition conditions (i.e. illumination, shape and density of the endothelium cells), it can be successfully applied to the frames of each single sequence, because in that case the acquisition conditions can be regarded as uniform in time.

4.2 Comparative Results

We firstly compared our ad-hoc registration algorithm against a commercial state-of-the-art software (Photoshop). As shown in Fig. 12, both algorithms are able to successfully register almost all of 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 merg-

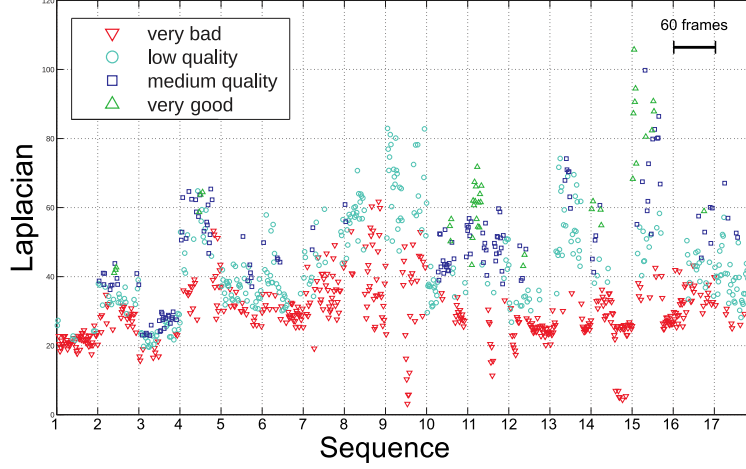


Fig 11 Frame quality evaluation for the Laplacian operator, as compared with manual ground truth. Results are shown for 17 sequences of 60 frames each. The sequences are separated by vertical grid lines. Ground truth frame quality is both shape- and color-coded, is classified into four classes: “very bad” (red ∇), “low quality” (cyan $\bigcirc$), “medium quality” (blue $\square$) and “very good” (green $\triangle$) (best viewed in color).

ing step: in fact, the result provided by Photoshop (Fig. 12(a)) has a wide region (enclosed by the yellow 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. For the special class of corneal endothelium images, this shows that using an information selection criterion based on the relative frame quality is undoubtedly beneficial.

A further evaluation compares our pipeline against a classical super-resolution method generating an HR image using the standard MAP super-resolution framework exploiting the Huber function as prior. This method requires to setup two parameters, i.e. the standard deviation of the Gaussian point spread function modeling the algorithm and the weight associated with the Huber prior. The code was run for different values of the parameters and the best result was retained. The best MAP result, shown in Fig. 13(b), is clearly of less quality with respect to ours, shown in Fig. 13(a), as it suffers of a posterization effect that flattens the appearance of the endothelium

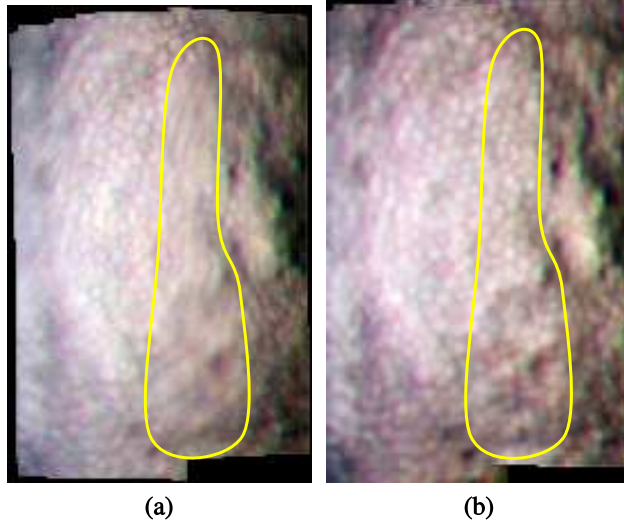


Fig 12 Comparison of mosaic images. (a) Photoshop. (b) Our result. The encircled area shows the mosaic portion where Photoshop fails to attain good image quality (see text).

cells, and makes cell boundaries much less definite. It is likely that the Huber prior, although providing excellent results in other contexts,³⁶ is less suitable for endothelial images. For the sake of completeness, Fig. 13(c) shows the best MLE solution (with the same Gaussian standard deviation used for MAP). As expected from the theory, a lot of high-frequency artifacts are generated in this case, making this solution totally useless for diagnostic purposes.

Finally, Fig. 14 reports an evaluation of the proposed method with respect to the bicubic interpolation of the reference frame I_r and the standard mosaic approach without Laplacian weighting and unsharpening using bilinear or bicubic interpolation, respectively, in terms of the BRISQUE index³⁷ according to varying magnification factors. The BRISQUE index is a quantitative image quality measure that does not require a reference ground-truth image, not available for the proposed application. Results are averaged over seventeen video sequences. The dotted line is referred to the average BRISQUE value for the reference frame I_r . As it can be noted, the proposed method achieves the best results. In particular, $3\times$ magnification represents the best trade-off between quality and scale, as evident from the fact that at this value the BRISQUE index related to our

method begins to rise significantly.

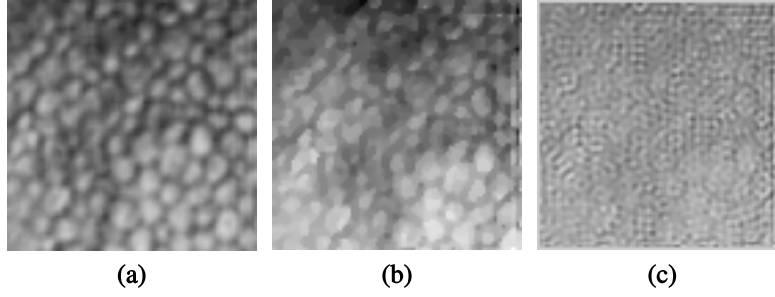


Fig 13 (a) Our result for the HR image. (b) The best MAP result employing the Huber function as prior. (c) The best MLE result.

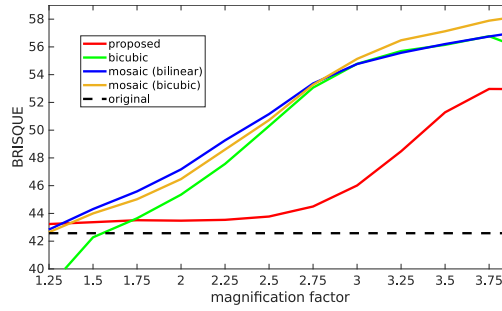


Fig 14 Evaluation of super-resolution methods according to the BRISQUE index for various magnification factors. Lower values are better (best viewed in color).

4.3 Image Gallery

Figure 15 shows twelve examples of $3\times$ magnification of 60-frame LR sequences with different subjects and acquisition conditions. For each sequence, the following images are shown from left to right: the original LR endothelium reference, the result obtained with our super-resolution approach, the result obtained with bicubic interpolation on the reference frame, and the results obtained respectively with the bilinear and bicubic classical mosaicing approaches. The average execution time of the proposed pipeline is about 7 second on a Intel[®] Core i3 CPU M330@2.13GHz notebook. In all cases, the HR image obtained with our method looks very detailed with respect to the original image, and of much better quality with respect to the interpolated image.

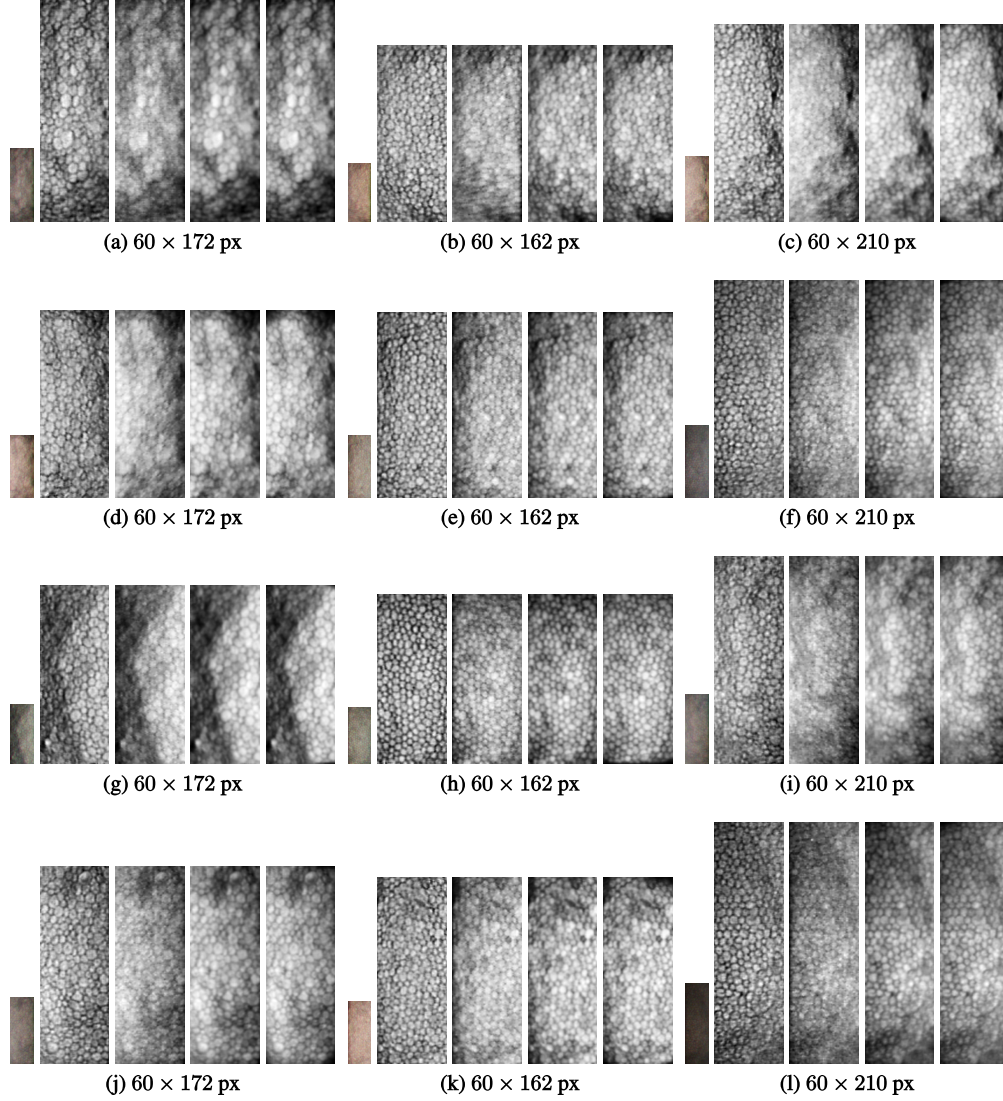


Fig 15 Several examples with magnification factor $3\times$. For each group of images, from left to right: Original reference endothelium frame, HR image obtained with our method, bicubic-interpolated image magnification, standard mosaicing approach with bilinear and bicubic interpolations. Images are of variable size, due to different illumination conditions. The original resolution of the reference frame is indicated in the corresponding caption.

5 Conclusions and Future Work

In this work, a fast and efficient method for obtaining good quality HR 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. Additionally, the proposed pipeline is fast and requires only low-cost equipment, making it a valid and affordable resource for both developed and developing countries.

Future work will address the development of a more general framework suitable for different applications, especially for medical imaging. Improvements to speed up the registration process will be addressed as well.

Acknowledgments

This work was partially funded by the *VISIA Imaging S.r.l.* company of S. Giovanni Valdarno (AR), Italy ¹. The authors would like to thank in particular Alessandro Foggi, Filippo Nencini and Chiara Menicucci of VISIA Imaging, both for their expert advice and for providing the annotated endothelial sequences obtained with slit lamp biomicroscopes.

References

- 1 D. Comanducci and C. Colombo, “Vision-based magnification of corneal endothelium frames,” in *9th International Conference on Computer Vision Systems ICVS*, 52–61 (2013).
- 2 S. Baker and T. Kanade, “Limits on super-resolution and how to break them,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**(1), 1167–1183 (2002).

¹<http://www.visiaimaging.com/>

- 3 H. Greenspan, “Super-resolution in medical imaging,” *The Computer Journal* **52**(1), 43–63 (2009).
- 4 K. Nasrollahi and T. Moeslund, “Super-resolution: a comprehensive survey,” *Machine Vision and Application* **25**(6), 1423–1468 (2014).
- 5 S. Park, M. Park, and M. Kang, “Super-resolution image reconstruction: a technical overview,” *IEEE Signal Processing Magazine* **20**(3), 21–36 (2003).
- 6 S. Battiato, G. Gallo, and F. Stanco, “A locally adaptive zooming algorithm for digital images,” *Image and Vision Computing* **20**(11), 805–812 (2002).
- 7 G. Pan, S. Han, Z. Wu, *et al.*, “Super-resolution of 3d face,” in *9th European Conference on Computer Vision (ECCV)*, 389–401 (2006).
- 8 C. Dong, C. Loy, K. He, *et al.*, “Image super-resolution using deep convolutional networks,” *IEEE Transaction on Pattern Analysis and Machine Intelligence* **38**(2), 295–307 (2016).
- 9 H. Liu and T. Zhu, “A correlative two-step approach to hallucinating faces,” *International Journal of Pattern Recognition and Artificial Intelligence* **28**(08), 1454004 (2014).
- 10 X. Lu, D. Wang, W. Shi, *et al.*, “Group-based single image super-resolution with online dictionary learning,” *EURASIP Journal on Advances in Signal Processing* **2016**(1), 84 (2016).
- 11 X. Li, J. Chen, Z. Cui, *et al.*, “Single image super-resolution based on sparse representation with adaptive dictionary selection,” *International Journal of Pattern Recognition and Artificial Intelligence* **30**(07), 1654006 (2016).
- 12 D. Glasner, S. Bagon, and M. Irani, “Super-resolution from a single image,” in *Proc. IEEE International Conference on Computer Vision*, (2009).

- 13 M. Irani and S. Peleg, “Image sequence enhancement using multiple motions analysis,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 216–221 (1992).
- 14 M. Tanaka, Y. Yaguchi, and M. Okutomi, “Robust and accurate estimation of multiple motions for whole-image super-resolution,” in *Proceedings of the International Conference on Image Processing (ICIP)*, 649–652 (2008).
- 15 A. Nasonov and A. Krylov, “Fast super-resolution using weighted median filtering,” in *20th International Conference on Pattern Recognition (ICPR)*, 2230–2233 (2010).
- 16 R. Hardie, “A fast image super-resolution algorithm using an adaptive wiener filter,” *IEEE Transactions on Image Processing* **16**(12), 2953–2964 (2007).
- 17 K. Simonyan, S. Grishin, D. Vatolin, *et al.*, “Fast video super-resolution via classification,” in *IEEE International Conference on Image Processing (ICIP)*, 349–352, IEEE (2008).
- 18 U. Mudenagudi, R. Singla, P. Kalra, *et al.*, *Super Resolution Using Graph-Cut*, 385–394 (2006).
- 19 M. Chiang and T. Boult, “Efficient super-resolution via image warping,” *Image Vision Computing* **18**(10), 761–771 (2000).
- 20 M. Irani and S. Peleg, “Super-resolution from image sequences,” in *IEEE International Conference on Pattern Recognition (ICPR)*, 115–120, IEEE (1990).
- 21 C. Chen, Y. Kuo, and C. Fuh, “Image reconstruction with improved super-resolution algorithm,” *International Journal of Pattern Recognition and Artificial Intelligence* **18**(08), 1513–1527 (2004).

- 22 D. Capel and A. Zisserman, “Automated mosaicing with super-resolution zoom,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 885–891 (1998).
- 23 P. .Cheeseman, B. Kanefsky, R. Kraft, *et al.*, *Super-Resolved Surface Reconstruction from Multiple Images*, 293–308. Springer Netherlands (1996).
- 24 R. Schultz and R. Stevenson, “Extraction of high-resolution frames from video sequences,” *IEEE Transactions on Image Processing* **5**, 996–1011 (1996).
- 25 D. Capel and A. Zisserman, “Super-resolution from multiple views using learnt image models,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2001)*, 627–634 (2001).
- 26 Z. Lin, J. He, X. Tang, *et al.*, “Limits of learning-based superresolution algorithms,” *International Journal on Computer Vision* **80**(3), 406–420 (2008).
- 27 K.-C. Ng, A. N. Poo, and M. H. Ang, “Practical issues in pixel-based autofocusing for machine vision,” in *Proc. IEEE International Conference on Robotics and Automation*, (2001).
- 28 M. Subbarao, T. Choi, and A. Nikzad, “Focusing techniques,” *Journal of Optical Engineering* **32**, 2824–2836 (1993).
- 29 C.-C. Chang and C.-J. Lin, “LIBSVM: A library for support vector machines,” *ACM Transactions on Intelligent Systems and Technology* **2**, 27:1–27:27 (2011). <https://www.csie.ntu.edu.tw/~cjlin/libsvm/>.
- 30 B. Manjunath, P. Salembier, and T. Sikora, *Introduction to MPEG-7: Multimedia Content Description Interface*, John Wiley & Sons, Inc. (2002).

- 31 M. Brown and D. Lowe, “Automatic panoramic image stitching using invariant features,” *International Journal on Computer Vision* **74**(1), 59–73 (2007).
- 32 R. Marzotto, A. Fusiello, and V. Murino, “High resolution video mosaicing with global alignment,” in *Proc. Conf. Computer Vision and Pattern Recognition*, (2004).
- 33 R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press (2004).
- 34 H. Y. Shum and R. Szeliski, “Construction of panoramic image mosaics with global and local alignment,” *International Journal on Computer Vision* **36**(2), 101–130 (2000).
- 35 A. Jain, *Fundamentals of Digital Image Processing*, Prentice-Hall (1989).
- 36 D. Capel, *Image Mosaicing and Super-Resolution (Cphc/Bcs Distinguished Dissertations.)*, SpringerVerlag (2004).
- 37 A. Mittal, A. K. Moorthy, and A. C. Bovik, “No-reference image quality assessment in the spatial domain,” *IEEE Transactions on Image Processing* **21**(12), 4695–4708 (2012).

List of Figures

- 1 The operation principle of the specular reflection technique used in the slit lamp biomicroscope. This can only be achieved when the angle α_i of incident illumination to the normal of the corneal surface equals the angle α_o of the observation system.

- 2 A typical endothelium image obtained with the biomicroscope: (a) Both the lamp corneal reflection and the endothelium are visible. (b) Only the endothelium is visible. (c) The endothelium is not visible. (d) An example of confocal image of the endothelium.
- 3 The intensity profile H_y and H_x , together with the derivative H'_y , for the image frame shown in Fig. 2(a). Gaussian fit of H_x is represented by a dashed curve (best viewed in color).
- 4 Automatic ROI selection on frames of a typical input video sequence. The dashed green rectangle shows the endothelium ROI, frame resolution is 640×480 pixels (best viewed in color).
- 5 Best subsequence selection criterion. Average SVM and ground-truth probabilities and mean Laplacian energy are respectively plotted in blue, black and red (best viewed in color).
- 6 (a) SIFT correspondences between frames 0 and 1 of Fig. 4. RANSAC inliers and outliers are respectively shown in green, red (best viewed in color). (b) The initial graph built for the video sequence of Fig. 4. Nodes inside the gray rectangles represent the initial chains, while dashed edges correspond to the successive merging steps. Singleton nodes represent to frames not linked to any other frame.

- 7 (a) The mosaic after the first image alignment. (b) The overlap of all ROIs, and the path followed by the slit lamp over the mosaic. Brighter regions correspond to a higher number of overlapping images, red parallelograms correspond to the warped ROIs onto the mosaic, and the points of the path are the frame ROI centers. (c) A better view of the path performed by the slit lamp onto the endothelium area, numbered according to the corresponding frames. Notice the sudden jumps from node 53 to 54 arisen during the scanning session (best viewed in color). (d) A case with multiple trees. The frame ROIs composing subgraph A have an average quality lower than the that of subgraph B. The latter is selected, even if the former has more nodes. Note that the mosaic of subgraph A do not overlap the mosaic of subgraph B. Frame 18, the missing node that would have connected subgraph A with B is very motion-blurred, because of the fast movement of the lamp jumping from a portion of the endothelium onto another one.
- 8 The extended graph. (a) Dashed edges are the new added edges. (b) The shortest path tree (bold red edges) having I_r as root, built on the extended graph. The weights are the average symmetric transfer errors of the affine maps registering the pairwise connected images (best viewed in color).
- 9 (a) The original endothelium LR image of size 60×208 pixels. (b) The proposed HR image J for a $3\times$ magnification factor. (c) The blurred image $\tilde{J}$ before unsharpening as output of the standard mosaicing SR. (d) The bicubic interpolation of the corresponding LR frame for a $3\times$ magnification factor. (e) Image (d) after unsharpening. Notice how in this last case the noise is amplified too.

- 10 Examples of the four quality labels with decreasing level from left to right (best viewed in color).
- 11 Frame quality evaluation for the Laplacian operator, as compared with manual ground truth. Results are shown for 17 sequences of 60 frames each. The sequences are separated by vertical grid lines. Ground truth frame quality is both shape- and color-coded, is classified into four classes: “very bad” (red ∇), “low quality” (cyan $\bigcirc$), “medium quality” (blue $\square$) and “very good” (green $\triangle$) (best viewed in color).
- 12 Comparison of mosaic images. (a) Photoshop. (b) Our result. The encircled area shows the mosaic portion where Photoshop fails to attain good image quality (see text).
- 13 (a) Our result for the HR image. (b) The best MAP result employing the Huber function as prior. (c) The best MLE result.
- 14 Evaluation of super-resolution methods according to the BRISQUE index for various magnification factors. Lower values are better (best viewed in color).
- 15 Several examples with magnification factor $3\times$. For each group of images, from left to right: Original reference endothelium frame, HR image obtained with our method, bicubic-interpolated image magnification, standard mosaicing approach with bilinear and bicubic interpolations. Images are of variable size, due to different illumination conditions. The original resolution of the reference frame is indicated in the corresponding caption.

List of Tables

- 1 SVM classification based on MPEG-7 descriptors. Descriptor combinations are listed in the first column, the other columns provide evaluation results for the validation and runtime rate both for two or four quality labels