SurgIPC: a Convex Image Perspective Correction Method to Boost Surgical Keypoint Matching

Rasoul Sharifian[1,2,3*] and Adrien Bartoli[1,2,3]

[1*]Institut Pascal, UMR6602 CNRS/UCA, Clermont-Ferrand, France.
[2]Department of Clinical Research and Innovation, Clermont-Ferrand University Hospital, France.
[3] SURGAR, Surgical Augmented Reality, Clermont-Ferrand, France.

*Corresponding author(s). E-mail(s): rasoul.sharifian.cs@gmail.com;
Contributing authors: adrien.bartoli@gmail.com;

Abstract

Purpose. Keypoint detection and matching is a fundamental step in surgical image analysis. However, existing methods are not perspective invariant and thus degrade with increasing surgical camera motion amplitude. One approach to address this problem is by warping the image before keypoint detection. However, existing warping methods are inapplicable to surgical images, as they make unrealistic assumptions such as scene planarity.
Methods. We propose Surgical Image Perspective Correction (SurgIPC), a convex method, specifically a linear least-squares (LLS) one, overcoming the above limitations. Using a depthmap, SurgIPC warps the image to deal with the perspective effect. The warp exploits the theory of conformal flattening: it attempts to preserve the angles measured on the depthmap and after warping, whilst mitigating the effects of image resampling.
Results. We evaluate SurgIPC under controlled conditions using a liver phantom with ground-truth camera poses and with real surgical images. The results demonstrate a significant improvement in the number of correct correspondences when SurgIPC is applied. Furthermore, experiments on downstream tasks, including keyframe matching and 3D reconstruction using Structure-from-Motion (SfM), highlight significant performance gains.
Conclusion. SurgIPC improves keypoint matching. The use of LLS ensures efficient and reliable computations. SurgIPC can thus be easily included in existing computer-aided surgery systems.

Keywords: image perspective correction, image matching, conformal flattening

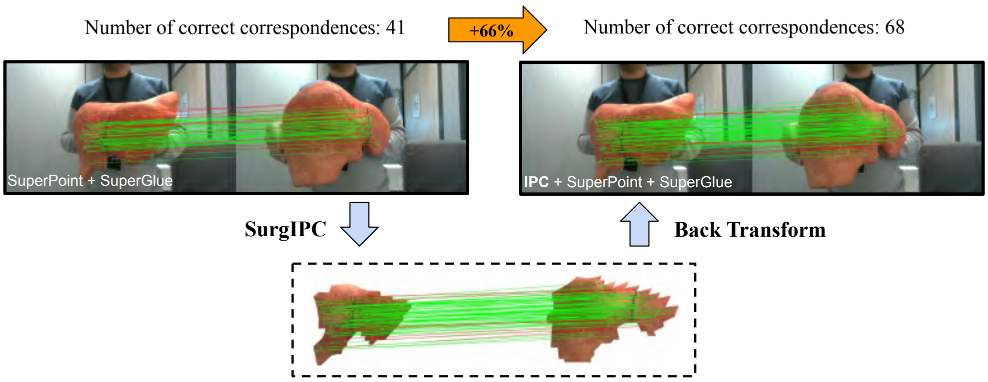
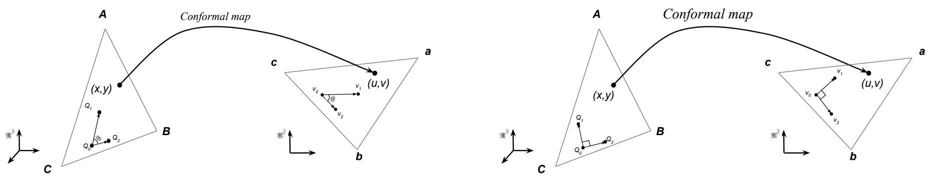


Fig. 1: SurgIPC cancels the effect of perspective and boosts the number of correct correspondences. In this example, SurgIPC is added to SuperPoint-SuperGlue and boosts matching by 66% (correspondences validated using the camera ground-truth).

1 Introduction

Keypoint detection and matching is a fundamental step in surgical vision, in particular in computer-assisted navigation, requiring multi-image organ reconstruction and organ-based camera pose estimation. The descriptors generated for the keypoints must be *invariant* against geometric and photometric changes, including illumination and blur [1]. In particular, detecting keypoints at multiple scales gains scale invariance [2] and rotating the patch to a dominant direction gains rotation invariance. However, dealing with the perspective effect, occurring when the camera viewpoint changes, is more challenging [1, 3–6]. Perspective distortion leads to significant keypoint appearance changes and causes image matching to fail. Perspective invariance has been attempted with two main strategies. The first strategy tries to develop perspective invariant descriptors; this so far has only been successfully achieved for affine transformations, a first-order approximation of perspectivities [3, 4, 7]. The second strategy is to warp the images with the purpose of mitigating the perspective effect prior to keypoint detection and description [5, 8–10]. Most methods in this category use homographies and are thus specific to planar scenes [5]. However, these methods are insufficient for the surgical setting. For non-planar scenes, existing methods require prior knowledge about the object geometry. In particular, [10] proposes a solution for developable surfaces, which again makes it impractical for the surgical setting. Consequently, image matching in the presence of viewpoint changes remains a major unsolved problem. The recent method [11] uses monocular depth estimation to compute the local surface curvature for each detected keypoint, which is then used as descriptive information in matching. This method needs the depth predictor and keypoint matcher to be jointly fine-tuned.

We introduce an image warping method named Surgical Image Perspective Correction (SurgIPC). It takes an image and a depthmap as inputs and generates a perspective-corrected image. SurgIPC exploits the concept of conformal flattening as



(a) General case for arbitrary bases (b) Special case for orthonormal bases

Fig. 2: Conformal mapping between two triangles with angle preservation.

a key component in warp computation. Conformal flattening is mostly used in computer graphics for 3D mesh texture mapping. It transforms a 3D mesh to a 2D flat domain while preserving angles [12], and thus limits the local geometric distortions to a 2D similarity (translation, rotation, and scaling), mitigating the perspective distortion [8, 9]. However, warping the image may negatively impact keypoint detection and matching, because of pixel resampling. This means that the image warp leads to the creation (over-sampling) and destruction (under-sampling) of pixels [13], harming the original image signal. We address this issue in the proposed warp using a second key concept: image resampling minimisation. In summary, SurgIPC warps the images while compromising between perspective cancellation and resampling minimisation. It involves a convex cost with two terms, leading to a globally optimal solution that can be reliably and effectively computed through simple linear least-squares minimisation.

2 Method

In SurgIPC, perspective cancellation is achieved by conformal flattening of the observed surface. For that, there exists a linear least-squares formulation, which is widely used in 3D geometry processing tools such as Blender and CGAL. The method is called Least Squares Conformal Mapping (LSCM) [12]. It approximates the Cauchy-Riemann equations in a least-squares manner. We propose a reformulation, the *3D conformal cost*, still convex, and with two main advantages over LSCM. Using LSCM or the 3D conformal cost on their own may cause resampling distortion, due to an excessive emphasis put on angle preservation. Additionally, LSCM requires one to fix the position of two vertices to avoid degenerate spurious solutions. The positions of these vertices may drastically change the overall solution. In SurgIPC, resampling minimisation uses a cost measuring the extent of over-sampling or under-sampling. We propose a convex *image displacement cost*. We next present the two cost terms and the method pipeline.

2.1 3D Conformal Cost

According to Riemann's theorem, any surface homeomorphism to a disk can be represented by a planar conformal parameterisation [14]. We use triangulated meshes constructed from the image's depthmap and the parametrisation is between 3D and 2D triangles. For a triangle ABC and its corresponding conformally flattened version

abc, shown in figure 2, the mapping between is affine and conformal if and only if the angles are preserved. We choose three points Q_0, Q_1, Q_2 in ABC forming two orthogonal vectors, hence:

$$Rot_{90}(\overrightarrow{Q_1Q_0}) = \overrightarrow{Q_2Q_0}. \tag{1}$$

Generalisation to non-orthogonal vectors is straightforward by choosing a custom angle. Under a conformal map, the following condition must thus hold in abc:

$$Rot_{90}(\overrightarrow{v_1v_0}) = \overrightarrow{v_2v_0}, \tag{2}$$

where v_i, $i \in [0,1,2]$ represent the positions of the mapped points within the 2D domain. In other words, the flat triangle must be a positively scaled version of its 3D counterpart. Let $[\alpha_j, \beta_j, \gamma_j]$, $j \in [0,1,2]$ be the barycentric coordinates of ABC; we thus have for the desired unknown vertices of abc:

$$v_i = \begin{bmatrix} \alpha_j & 0 & \beta_j & 0 & \gamma_j & 0 \\ 0 & \alpha_j & 0 & \beta_j & 0 & \gamma_j \end{bmatrix} \begin{bmatrix} x_a & y_a & x_b & y_b & x_c & y_c \end{bmatrix}^\top, \tag{3}$$

where $x_{a,b,c}$ and $y_{a,b,c}$ are the x and y coordinates of the abc vertices. Rewriting equation (2) in barycentric coordinates, we establish a linear system for the conformal transformation between the two triangles:

$$M_t p_t = 0, \tag{4}$$

where p_t holds the location of the t-th triangle vertices mapped to 2D and M_t simply relies on its geometry in 3D as:

$$M_t = \begin{bmatrix} \alpha_2 - \alpha_0 & \alpha_1 - \alpha_0 & \beta_2 - \beta_0 & \beta_1 - \beta_0 & \gamma_2 - \gamma_0 & \gamma_1 - \gamma_0 \\ \alpha_1 - \alpha_0 & \alpha_2 - \alpha_0 & \beta_1 - \beta_0 & \beta_2 - \beta_0 & \gamma_1 - \gamma_0 & \gamma_2 - \gamma_0 \end{bmatrix}. \tag{5}$$

The non-trivial solution of equation (4) gives the desired coordinates of the conformally flattened abc. To establish conformality for the entire 3D mesh, we minimise equation (4) in the least-squares sense for the entire 3D mesh triangle set T_{3D}, leading to the conformal cost as a function of the flattened vertices p:

$$C_{conf}(p) = \Sigma_{t \in T_{3D}} \|M_t p_t\|^2. \tag{6}$$

This proposed formulation of the conformal constraints is convex and has two main advantages over LSCM [12]. The first advantage is generality. The original LSCM formulation uses complex functions and their derivatives to derive the LSCM cost for a triangle, which preserves the angle between two arbitrarily selected orthogonal vectors on the triangle. In contrast, our formulation preserves the angle between any two custom vectors on the triangle. It can thus reproduce the LSCM formulation as a special case but can also use any other single or multiple vector pairs to express the conformal constraint. The second advantage is the need for the original LSCM method to fix the position of two vertices in the parameterisation domain to prevent trivial spurious solutions. In the application case at hand, which is image matching, this would incur an unreasonable image resampling. In contrast, our method does not require prescribing these two vertices: the proposed SurgIPC cost prevents trivial spurious solutions from occurring thanks to its image displacement cost.

2.2 Image Displacement Cost

We formulate a cost function which characterises the extent of image resampling, by measuring the displacement of pixels placed on a 2D grid within the image, as:

$$C_{disp}(p) = \|P_{init} - p\|^2, \tag{7}$$

where P_{init} represents the initial position of the grid and p its desired but unknown position. We initialize P_{init} by pruning a regular grid to approximately cover the region of interest. This linear least-squares cost imposes minimal pixel displacement image to reduce the resampling effect while texture-mapping.

2.3 Method Pipeline

SurgIPC takes an image and depth map as its initial input. It begins by generating the corresponding point cloud of the scene, using depth estimation. It then creates two triangulated meshes. The first mesh represents the 3D mesh of the target object within the scene. The second mesh is a 2D image grid corresponding to the same object. For each input image, application-specific object masks are applied to the image and its corresponding 3D mesh to segment the region of interest. This was done to ensure a fair evaluation, focusing solely on the target objects and also to prevent any discontinuity in the depth map. The warp is driven by the mesh vertices, which are computed by cost minimisation. The image is finally warped.

The total SurgIPC cost combines the 3D conformal and image displacement costs to achieve the simultaneous preservation of 3D angles and the prevention of resampling:

$$C_\lambda(p) = \lambda \mu_{conf} C_{conf}(p) + (1-\lambda)\mu_{disp} C_{disp}(p), \tag{8}$$

where μ_{conf} and μ_{disp} are fixed parameters used to normalise the range of the two terms to a consistent scale. The hyperparameter $\lambda \in [0, 1]$ is chosen to balance the two terms. The SurgIPC cost is convex, specifically linear least-squares, and thus can be efficiently and reliably solved in real time.

3 Experimental Results

We report experiments on several models.

3.1 Evaluation with a Liver Phantom

We quantitatively evaluate the use of SurgIPC with a 3D-printed and painted liver phantom. It was constructed by first reconstructing a 3D liver model from a patient CT obtained from our hospital within an IRB-approved protocol. Subsequently, randomly-spaced carved markers with known 3D locations were added to the 3D liver model

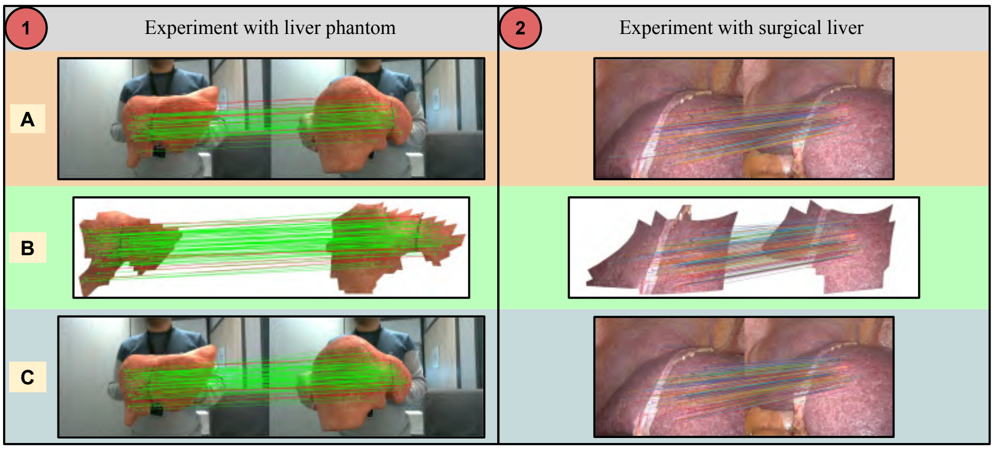


Fig. 3: 1) Quantitative experiment with 3D-printed and painted liver phantom. We positioned eight markers, represented as small black circles on the phantom surface, to facilitate stable pose estimation and ground-truth assessment. The green lines indicate correct correspondences, while the red lines are incorrect correspondences. In this experiment, SurgIPC improved the number of correct correspondences from 41 to 68. 2) Experiment conducted with surgical liver images. In this experiment, SurgIPC improved the number of correspondences from 174 to 217. In both experiments, A) represents the result of keypoint matching for the original images, B) represents the keypoint matching in flattened domain, and C) represents the result using SurgIPC. For both experiments presented in this figure, we have used SuperPoint and SuperGlue.

surface before 3D printing. These markers facilitate camera pose ground-truth estimation. This is particularly important in our experiments, as it enables the transfer of pixels between images and the automatic and reliable assessment of correspondence accuracy. A correspondence is considered valid if its transfer error is lower than a threshold set to 3 pixels. The transfer error is computed as the distance between the keypoint in the second image and the keypoint in the first image transferred to the second one, averaged with the distance computed by reversing the two images to ensure bidirectionality. Lastly, after printing, we painted the phantom to obtain the typical repetitive liver texture, enhancing realism and challenging image matching algorithms. We have captured images of this phantom using an intel realsense D405 RGB-D camera. Recall that the motivation of SurgIPC is that most keypoint detection methods cannot cope with perspective. This means that SurgIPC is an optional step usable in combination with any existing keypoint detection method. In other words, the evaluation should be done for representative keypoint detection and matching methods, comparing them without and with the use of SurgIPC. We use SuperPoint keypoint detector [7] combined with the SuperGlue matcher [15] as well as LoFTR [16]. These methods form the state-of-the-art in learning-based matching. We have used the main GitHub repository for SuperPoint and SuperGlue, and the Kornia library implementation [17] for LoFTR. We use SIFT and ORB, as popular classical keypoint methods. With these methods, we have the representative methods for both the classical and

	Phantom								Surgical
	#Crsp. ↑		#C. Crsp. ↑		Precision ↑		Recall ↑		#Crsp. ↑
	easy	*hard*	*easy*	*hard*	*easy*	*hard*	*easy*	*hard*	
SIFT+UBCMatcher	49	12	45	2	0.92	0.17	**0.80**	0.06	174
SurgIPC+SIFT+UBCMatcher	**50**	**24**	**47**	**12**	**0.94**	**0.74**	0.70	**0.32**	**217**
ORB+NN	**145**	0	94	0	0.65	0.00	**0.45**	0.00	25
SurgIPC+ORB+NN	142	**5**	**96**	**2**	**0.68**	**0.40**	0.36	**0.08**	**41**
SuperPoint+SuperGlue	**137**	53	**135**	41	**0.99**	0.52	**0.90**	41	245
SurgIPC+SuperPoint+SuperGlue	136	**76**	133	**68**	0.98	**0.65**	0.88	**68**	**284**
LoFTR	318	206	295	161	0.93	**0.81**	0.64	0.38	352
SurgIPC+LoFTR	**324**	**343**	**301**	**264**	0.93	0.77	**0.65**	**0.70**	**398**

Table 1: Comparison between methods with and without using SurgIPC. The phantom results are obtained using the liver phantom shown in Figure 1. The column '#Crsp.' gives the number of correspondences obtained by the matching algorithm. The '#C. Crsp.' column gives the number of correspondences validated by using the phantom's ground-truth pose. For all metrics, the higher, the better. Surgical column shows the results for surgical liver images. In this case, we only report the #Crsp., as no camera ground-truth was available.

learning-based approaches. Using this setup, we evaluated SurgIPC under two conditions: easy and hard. As a preliminary step, we performed a 'sanity check' of SurgIPC under easy conditions, characterised by rich textures and minimal viewpoint changes. For this evaluation, we selected two consecutive frames of the liver phantom with negligible perspective variation, referred to as the easy frames. Conversely, the hard frames involved significant perspective distortion. The results, shown in table 1, demonstrate that SurgIPC brings a significant performance boost in hard conditions across all compared methods, as validated by the ground-truth poses. For easy frames, there is no observable significant degradation in the number of correct correspondences when SurgIPC is used.

3.2 Evaluation with Surgical Liver Images

We evaluate the SurgIPC with real laparoscopic images of a patient's liver captured in our hospital. We selected frames where the surgeon has rotated the camera around the organ and induced the perspective effect. We used [18] monocular depth estimation network and created the 3D mesh of the scene using the camera intrinsic parameters. Note that as in SurgIPC formulation the scale of the 2D flattened mesh is constrained by the image displacement, only the shape of the 3D object is sufficient for the method. Therefore, depth estimation methods such as monocular depth estimation networks which estimate the depth map up to scale are compatible with SurgIPC. Figure 7 shows the output flattened images as well as the result of keypoint matching. Concretely, without SurgIPC we obtained 174 correspondences, whereas with SurgIPC and monocular depth, we observed an increase to 217 correspondences.

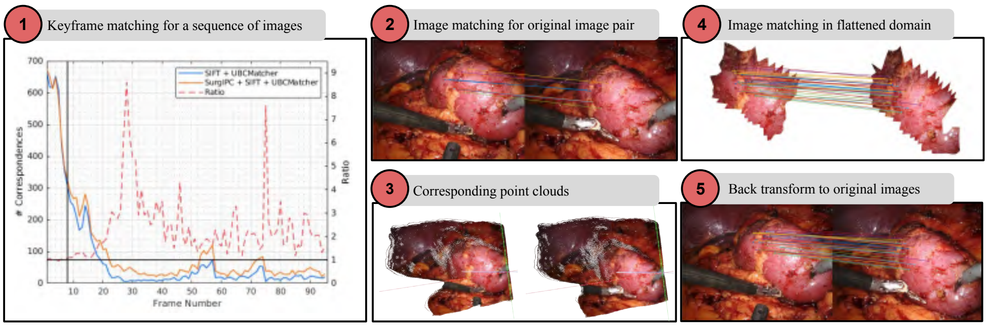


Fig. 4: Keyframe matching in robot-assisted partial nephrectomy. (1) Keyframe matching results for individual frames of a sequence. The original frame rate of 60 fps was reduced to 5 fps to ease visualisation. The left channel of the stereo camera was used. Frame #1 is chosen as the keyframe and matched to the subsequent frames. The graph shows the number of correspondences of the baseline method without and with SurgIPC, and their ratio. Up to frame #8 (indicated by a vertical black line), the frame difference is minimal and using SurgIPC does not make a difference against the baseline (the horizontal black line represents a ratio of one). Beyond frame #8, as perspective distortions intensify, SurgIPC demonstrates its distortion correction capability and significantly outperforms the baseline. (2) shows the keyframe on the left and frame #35 on the right, overlaid with the baseline matching result. (3) shows the monocular point clouds inferred from EndoDAC's depth map and the laparoscope's intrinsic parameters. (4) shows the matching result in the flattened domain. (5) shows the result of SurgIPC, as the matched keypoints back-transformed to the original images. Visual inspection did not reveal mismatches, showing that the matches are correct to an excellent extent.

3.3 Keyframe Matching in Partial Nephrectomy

We have conducted an organ tracking experiment involving a partial nephrectomy robot-assisted surgery sequence. In this sequence, the surgeon mobilises the kidney with a side push to expose it sideways, a very common type of gesture during the organ's initial inspection phase. This introduces substantial perspective changes from the beginning of the sequence. We established the initial frame of this sequence as the keyframe and evaluated SurgIPC's performance by matching all subsequent frames to this keyframe. We used EndoDAC [19] for depth estimation and the camera's intrinsic parameters to then reconstruct the point clouds. The results, as shown in figure 4, demonstrate a tremendous increase in the number of correspondences when using SurgIPC. Concerning the data used in this experiment, the patient gave written informed consent in accordance with the UroCCR project (French network of research on kidney cancer, NCT03293563).

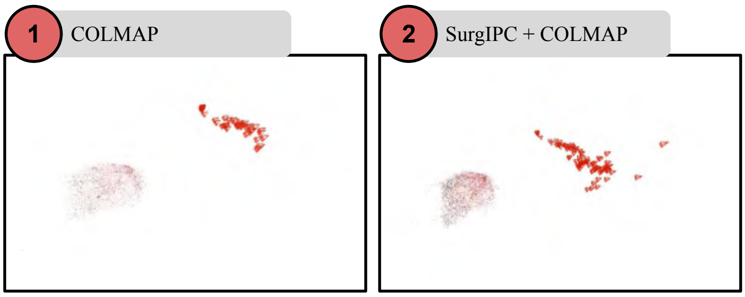


Fig. 5: Visual comparison of COLMAP reconstruction output. 1) COLMAP used as baseline, without using SurgIPC, and 2) COLMAP used with SurgIPC's keypoints.

3.4 3D Reconstruction

We have conducted a 3D reconstruction experiment by combining SurgIPC with COLMAP [20, 21]. The objective is to evaluate the integration of SurgIPC with standard Structure-from-Motion (SfM). We fed both the keypoints and descriptors obtained through SurgIPC, as well as the regular ones obtained without using SurgIPC for comparison purposes, into the COLMAP pipeline. For this experiment, we used the partial nephrectomy video sequence used in the experiment in 3.3. The reconstructions are illustrated by figure 5 and statistical results given in table 2. We observe that SurgIPC significantly boosts the number of matched images, from 37 images without SurgIPC to 88 images with SurgIPC, and achieved a significantly denser final reconstruction, from 3,303 points without SurgIPC to 4,976 points with SurgIPC. Additionally, the mean track length is extended from 8.43 to 11.23 frames, and the computation time is reduced from 3.28 minutes to 1.55 minutes. The mean reprojection error slightly increased from 0.87 to 1.02 pixels, an insignificant difference of only 0.15 pixels, compared to the boost obtained in the number of reconstructed cameras and structure density, indicating that SurgIPC significantly enhances the robustness and completeness of the 3D models generated by COLMAP. This experiment indicates the practical utility of SurgIPC in enhancing SfM methods.

Method	#Cameras	#Registered Imgs.	#Points	Reproj. Error (px)	Time (min)
COLMAP	37	37/95	3303	**0.87**	3.28
SurgIPC + COLMAP	**88**	**88/95**	**4976**	1.02	**1.55**

Table 2: SfM reconstructions obtained by COLMAP without and with SurgIPC.

3.5 Additional Data

We present three sets of additional data and results.

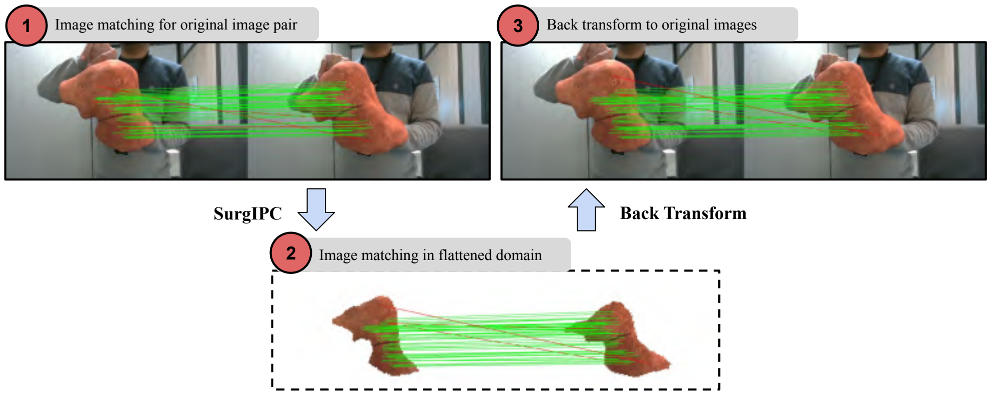


Fig. 6: Qualitative results for the experiment on 'easy' frames of the liver phantom with minimal viewpoint change.

3.5.1 Additional Qualitative Results for the Liver Phantom

We give additional qualitative results for the experiment conducted on the easy frames from section 3.1. Recall that these frames represent an easy case designed to verify the absence of performance decrease with the use of SurgIPC. These frames were chosen as consecutive frames of a liver phantom video, hence with minimal perspective variation. Figure 6 shows the frames and the matching results obtained with and without using SurgIPC. The results were obtained with SIFT combined with UBCMatcher, and are extensively described in table 1.

3.5.2 Results for an Ex-vivo Sheep Liver

We selected a sheep liver for its anatomical similarity to the human liver, featuring two lobes, as its human counterpart. This example features a case with limited texture and high resemblance between the point appearances. We acquired images by rotating the camera around the organ to induce perspective changes. Manual validation was performed to determine the correct correspondences for this dataset. The results shown in figure 7 demonstrate a boost in terms of the number of correspondences when SurgIPC is used, in combination to SuperPoint and SuperGlue. Without SurgIPC, we obtained 5 correspondences (all correct), whereas with SurgIPC, we observed an increase to 32 correspondences, among which 21 are correct.

3.5.3 Results with a Uterus Phantom

We used a surgical female pelvic trainer phantom. This phantom includes the uterus and other anatomical landmarks with standard shapes and textures. We used an Intel RealSense camera D415, which is a short-range stereo camera and provides sub-millimeter depth accuracy. To compare with this depth data, we used MiDaS for monocular depth estimation [18]. Using the estimated depth we created the 3D mesh of the scene using the camera intrinsic parameters. We used the proposed SurgIPC pipeline for both depth maps. Figure 8 shows the output flattened images as well as the

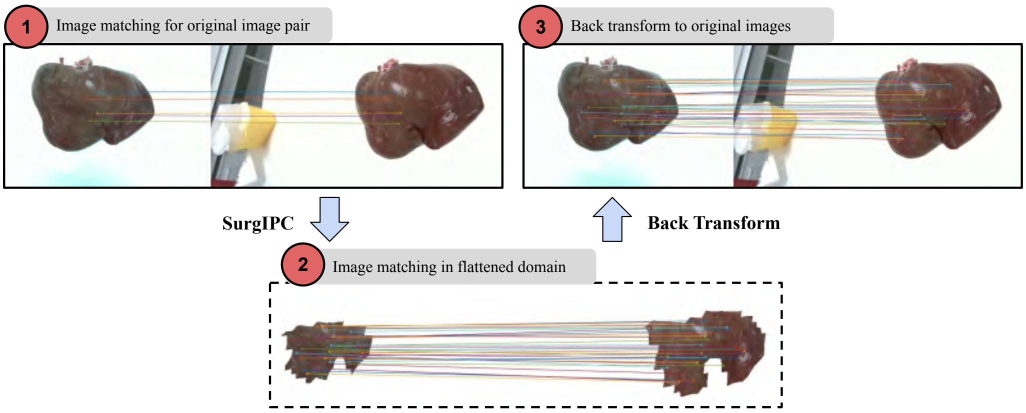


Fig. 7: Qualitative experiment with sheep liver. 1) Represents the original images and the result of keypoint matching, 2) represents the warped images using SurgIPC and the result of keypoint matching in flattened domain, 3) represents the result of transforming the keypoints back to original image-view. We have used SuperPoint as keypoint detector and SuperGlue as matcher. The results show a significant improvement in the number of correspondences.

result of keypoint matching. Concretely, without SurgIPC, we obtained 10 correspondences, of which 6 were correct, whereas with SurgIPC and deep learning monocular depth, we observed an increase to 16 correspondences, all of which were correct. Additionally, with SurgIPC and the depth sensor, we have 13 correspondences, with 12 of them being correct. For this experiment, we used SIFT combined with UBCMatcher. In both cases with different depth inputs, the warped images are fronto-parallel and the perspective effect is cancelled. As a consequence, the images are more similar, and thus the number of correspondences increases.

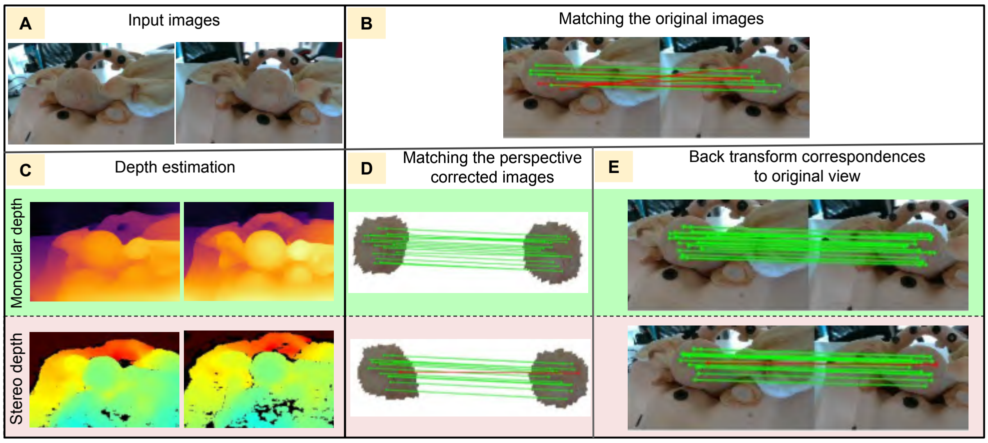


Fig. 8: Uterus phantom image matching with and without using SurgIPC, using different depth sources. A, Input images. B, Image matching using original images. C, Estimated depths via the monocular neural network and stereo. D, Image matching in perspective-free view, (Up: The input 3D mesh for SurgIPC is generated from monocular depth. Down: The input 3D mesh for SurgIPC is generated by the active stereo depth sensor). E, Back-transforming the correspondences to the original view.

4 Ablation Studies

We report an ablation study for the effect of resampling and for the depth accuracy.

4.1 Effect of Resampling on Image Matching

In this experiment, we have varied the value of λ from zero to one and measured the performance of the methods by quantifying the number of correspondences and the number of correct correspondences. The results are shown in figure 10 and example warped images for 5 different λ values are shown in figure 9. Recall that λ is a hyper-parameter of the cost function chosen within the [0,1] interval which allows one to trade-off the conformal cost and the resampling. When λ goes to one, the flattening tends to be highly conformal but with high resampling. The spurious trivial solution occurs when λ strictly equals one and we thus use an upper-bound of 0.995. This is shown by a yellow point in figure 10. Conversely, when λ goes to zero, the flattening is the least conformal but with the least resampling. The image simply does not change when λ strictly equals zero. This is shown by a red point in figure 10.

Several observations can be made from this comparison. First, the graphs clearly show the incompatibility of LSCM with the task at hand of image matching. Recall that LSCM preserves the angles and thus has the potential for resolving perspective distortion. However, its poor performance on image matching was expected: as LSCM is primarily designed to address texture mapping, it is not constrained on the right image scale, and thus significantly incur image resampling. In contrast, the proposed SurgIPC shows the desired performance as it is purposefully designed to address both perspective and resampling. Second, the performance of LSCM on image matching is even worse than matching the original images. This highlights the impact of the proposed SurgIPC formulation which effectively addresses the task of image matching by limiting the resampling effect of LSCM. Third, it is worth mentioning that we have used a fixed $\lambda = 0.95$ for all experiments, including phantom and ex-vivo data. While this indicates stable SurgIPC performance, there still remains future work for careful hyperparameter tuning and finding λ automatically.

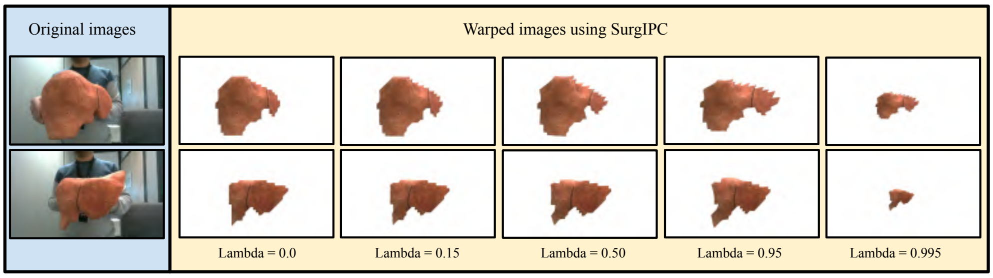


Fig. 9: Warped images obtained by SurgIPC for five sample values of λ. The original image pair is shown left and the warped images right. The warped images visually demonstrate the effect of λ selection and the trade-off between resampling and perspective correction. For λ equals zero, we obtain the original image without resampling, but one can clearly distinguish the perspective distortion effect upon careful observation. As λ approaches one, the perspective distortion is being corrected and the warped image looks fronto-parallel. However, the image starts to shrink, thus resampling becomes higher. For λ values in between, such as λ equals 0.95, we have a trade-off between perspective correction and image resampling: the warped image is almost equal to the original image in size and the perspective effect is effectively corrected.

4.2 Effect of Depth Accuracy on SurgIPC's Performance

In order to examine the quality of depth estimation network on SurgIPC, we have compared the impact of several depth estimation networks on the practical use-case of intraoperative 3D reconstruction by SfM, introduced in our experiments in section 3.4 . We have chosen depth estimation networks which were reported to improve in accuracy after fine-tuning on surgical datasets, namely AF-SfM Learner [22], Endo-DAC [19], and TRSMDV [23]. For this ablation study, we use the partial nephrectomy sequence used in experiments conducted in sections 3.3 and 3.4, and report the statistics of the final reconstruction metrics for each depth estimation method in table 3. We observe that the accuracy of the depth estimation methods has an important impact on the performance of SurgIPC. For instance, the AF-SfM Learner [22], which has the lowest reported accuracy among the tested methods [19, 23], resulted in the sparsest reconstruction, with 3,108 points. This is even fewer than the number of points generated without using SurgIPC.

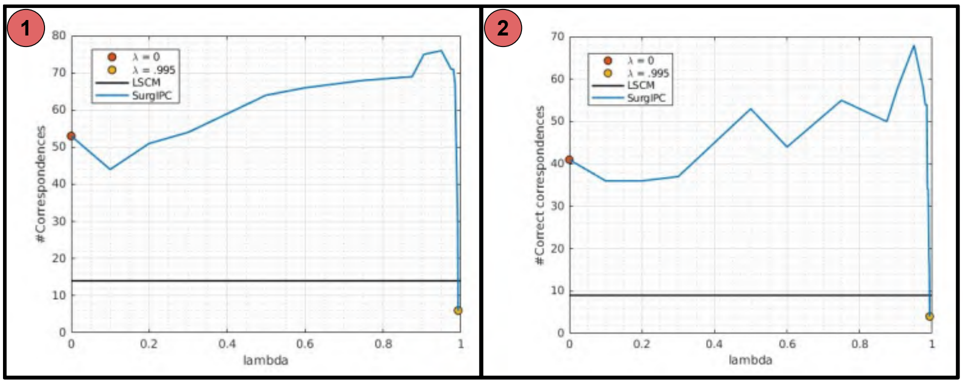


Fig. 10: Impact of λ on SurgIPC performance. 1) shows performance as number of correspondences and 2) shows performance as number of correct correspondences validated by ground-truth. For both graphs, the results were obtained by SuperPoint followed by SuperGlue. The blue line illustrates the number of correspondences achieved with SurgIPC against λ. Three special points are highlighted on the graphs. 1) The red circle at λ equals zero represents the zero resampling and the highest perspective distortion; it corresponds to the result obtained when matching the original images. 2) The yellow circle when λ goes to one represents the most conformal solution with the highest degree of resampling. 3) The graph maximum which is the optimal trade-off between resampling and perspective correction. The black line shows the performance of LSCM, which is independent of λ and clearly below par.

To further analyse the impact of the input depth quality, we qualitatively evaluate the point clouds generated using AF-SfM learner and EndoDAC, and the corresponding image warps produced by SurgIPC in figure 11. The point clouds are visualised from the same viewpoint, allowing for a direct comparison. Upon close examination, it is evident that the 3D shape of the kidney is inferred differently by both methods. These variations in 3D shape lead to differing SurgIPC image warps, resulting in notably different perspective corrections. For example, the output image warp for a close-up shown in figure D varies significantly between the methods.

Method	**#Cameras**	**#Registered Imgs.**	**#Points**	**Reproj. Error (px)**	**Time (min)**
COLMAP	37	37/95	3303	**0.87**	3.28
EndoDAC + SurgIPC + COLMAP	**88**	**88/95**	**4976**	1.02	1.55
AF-SfM + SurgIPC + COLMAP	28	28/95	3108	1.05	3.40
TRSMDV + SurgIPC + COLMAP	84	84/95	4251	0.97	**1.20**

Table 3: Impact of the depth estimation network on the final 3D reconstruction. The 3D reconstruction is obtained by SfM via COLMAP without and with SurgIPC.

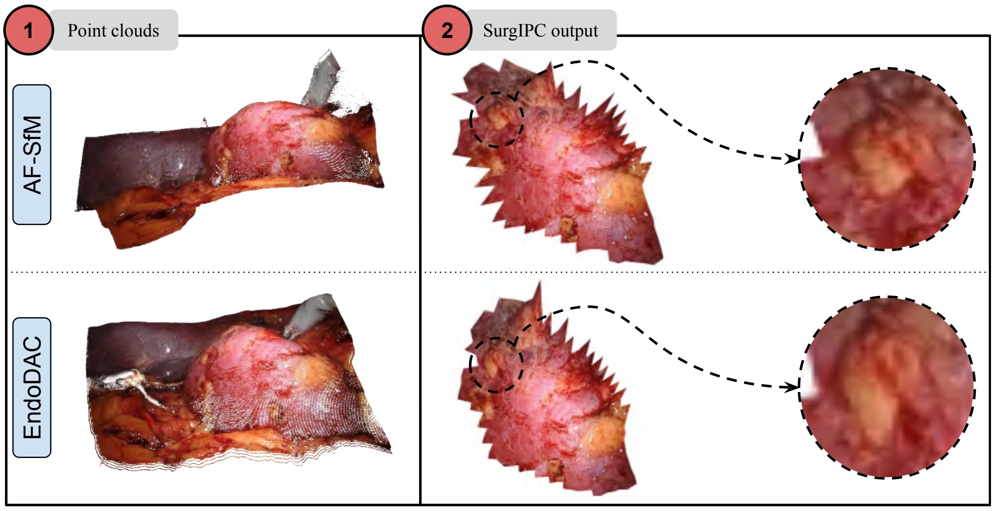


Fig. 11: Qualitative comparison of the impact of the depth estimator quality on SurgIPC's image warp. Both 3D visualisations on the left are rendered from the same viewpoint, showing how strongly different the predicted depths are, explaining why AF-SfM + SurgIPC underperforms, AF-SfM being having the least performance in depth estimation [19, 23], while EndoDAC + SurgIPC is consistently beneficial to the downstream tasks, such as SfM. We thus recommend using SurgIPC systematically in point matching tasks with EndoDAC as base depth estimator.

5 Discussion

We discuss three points. First, although improving matching would directly improve stereo reconstruction, SfM and visual SLAM, however, the number of benefitting downstream tasks is probably much more important. It would improve organ tracking, pose estimation and deformable reconstruction methods, to name but a few. Indeed, in contrast to rigid methods such as SfM, where the use of RANSAC brings a strong tolerance to spurious point correspondences, deformable reconstruction methods such as Shape-from-Template (SfT) and Non-Rigid Structure-from-Motion (NRSfM) do not benefit from a global geometric model and cannot use RANSAC. They are thus extremely sensitive to spurious correspondences, which the proposed SurgIPC considerably reduces, as shown in the experiments. Therefore, progress in keypoint matching drives one towards the use of deformable reconstruction which would strongly benefit navigation in soft organs.

Second, to assess the benefit in runtime, we have reimplemented the area-based resampling cost from [13] in our framework. This cost term is non-convex and requires nonlinear optimisation. On average, this previous term requires 12 seconds of iterative optimisation per image. This is clearly incompatible with real-time processing by several orders of magnitude. In contrast, the proposed SurgIPC framework uses a convex cost function and operates at 120 milliseconds per frame. Regarding the time needed for depth estimation, we evaluated AF-SfM learner [22], EndoDAC [19], and

TRMDSV [23] models in ablation studies presented in section 3.3.2, measuring inference times of 14 milliseconds, 22 milliseconds, and 20 milliseconds, respectively, on an RTX 2080 GPU. Including the depth estimation inference time with SurgIPC's, the complete overhead computation time amounts to 145 milliseconds per frame, which aligns with the real-time processing standards.

Third, a potential limitation of this work lies in the impact of depth estimation quality on the performance of SurgIPC, as highlighted in the ablation studies. In addition to the inherent inaccuracies in monocular depth estimation methods, challenging surgical conditions such as non-uniform lighting and image perturbations caused by smoke and bleeding can further impact the accuracy of depth estimation, potentially altering the inferred 3D shape. We leave a detailed investigation of these factors for future work.

6 Conclusion

SurgIPC corrects perspective distortions by preserving 3D angles and minimising image resampling. The results are very convincing: the method boosts the performance of existing keypoint matching in the presence of extreme perspective distortions. Importantly, SurgIPC is adaptable to non-planar scenes, is convex, and can be easily integrated into existing computer-aided surgery systems.

Acknowledgments. We thanks UroCCR for the anonymised surgical videos and ANR for support through the France 2030 program (ANR-21-RHU5-0015).

Conflict of Interest. Rasoul Sharifian declares that there are no conflicts of interest regarding the publication of this paper. Adrien Bartoli declares that there are no conflicts of interest regarding the publication of this paper.

References

[1] Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., Gool, L.V.: A comparison of affine region detectors. IJCV (2005)

[2] Lowe, D.G.: Distinctive image features from scale-invariant keypoints. IJCV (2004)

[3] Mikolajczyk, K., CORDELIA, S.: Scale and affine invariant interest point detectors. IJCV (2004)

[4] Morel, J.-M., Yu, G.: Asift: A new framework for fully affine invariant image comparison. SIAM Journal on Imaging Sciences (2009)

[5] Toft, C., Turmukhambetov, D., Sattler, T., Kahl, F., Brostow, G.J.: Single-image depth prediction makes feature matching easier. In: ECCV (2020)

[6] Rodriguez, M., Facciolo, G., Gioi, R.G., Muse, P., Delon, J.: Robust estimation of local affine maps and its applications to image matching. In: WACV (2020)

[7] DeTone, D., Malisiewicz, T., Rabinovich, A.: Superpoint: Self-supervised interest point detection and description. In: CVPRW (2018)

[8] Chandelon, K., Sharifian, R., Marchand, S., Khaddad, A., Bourdel, N., Mottet, N., Bernhard, J.-C., Bartoli, A.: Kidney tracking for live augmented reality in stereoscopic mini-invasive partial nephrectomy. Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization (2023)

[9] Kim, J.-H., Bartoli, A., Collins, T., Hartley, R.: Tracking by detection for interactive image augmentation in laparoscopy. WBIR (2012)

[10] Zeisl, B., Köser, K., Pollefeys, M.: Viewpoint invariant matching via developable surfaces. ECCV. Workshops and Demonstrations (2012)

[11] Wang, S., kannala, J., Pollefeys, M., Barath, D.: Guiding local feature matching with surface curvature. In: ICCV (2023)

[12] Lévy, B., Petitjean, S., Ray, N., Maillot, J.: Least squares conformal maps for automatic texture atlas generation. ACM Transactions on Graphics (2002)

[13] Gluckman, J., Nayar, S.K.: Rectifying transformations that minimize resampling effects. In: CVPR, Kauai, Hawaii, USA (2001)

[14] Riemann, B.: Grundlagen Für Eine Allgemeine Theorie der Functionen Einer Veränderlichen Complexen Grösse, Göttingen, Germany (1851). PhD thesis

[15] Sarlin, P.-E., DeTone, D., Malisiewicz, T., Rabinovich, A.: Superglue: Learning feature matching with graph neural networks. In: CVPR, Virtual (2020)

[16] Sun, J., Shen, Z., Wang, Y., Bao, H., Zhou, X.: Loftr: Detector-free local feature matching with transformers. In: CVPR, Virtual (2021)

[17] Riba, E., Mishkin, D., Ponsa, D., Rublee, E., Bradski, G.: Kornia: An open source differentiable computer vision library for pytorch. In: WACV (2020)

[18] Ranftl, R., Lasinger, K., Hafner, D., Schindler, K., Koltun, V.: Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer. TPAMI (2022)

[19] Cui, B., Islam, M., Bai, L., Wang, A., Ren, H.: Endodac: Efficient adapting foundation model for self-supervised depth estimation from any endoscopic camera. In: MICCAI (2024)

[20] Schonberger, J.L., Frahm, J.-M.: Structure-from-motion revisited. In: CVPR (2016)

[21] Schönberger, J.L., Zheng, E., Frahm, J.-M., Pollefeys, M.: Pixelwise view selection for unstructured multi-view stereo. In: ECCV (2016)

[22] Shao, S., Pei, Z., Chen, W., Zhu, W., Wu, X., Sun, D., Zhang, B.: Self-supervised monocular depth and ego-motion estimation in endoscopy: Appearance flow to the rescue. Medical image analysis **77**, 102338 (2022)

[23] Budd, C., Vercauteren, T.: Transferring relative monocular depth to surgical vision with temporal consistency. In: MICCAI (2024)