

Augmented vision: seeing beyond field of view and occlusions via uncalibrated visual transfer from multiple viewpoints

Antonella Nardi Dario Comanducci Carlo Colombo

Dipartimento di Sistemi e Informatica, Via S. Marta 3, 50139 Firenze, Italy

Email: {nardiantonella@gmail.com,comandu@dsi.unifi.it,colombo@dsi.unifi.it}

Abstract—We present an approach for merging into a single super-image a set of uncalibrated images of a general 3D scene taken from multiple viewpoints. To this aim, the content of either image is augmented with visual information taken from the others, while maintaining projective coherence. The approach extends the usual mosaicing techniques to image collections with 3D parallax, and operates like a virtual sensor provided with an enlarged field of view and the capability of seeing through visual occlusions in an “X-ray” fashion. Fundamental matrices are used to transfer visual information through the vertexes of an image graph. A dense stereo paradigm is employed to achieve photorealism by partitioning image pairs into corresponding regions. Results in oriented projective geometry are then exploited to both detect and handle occlusions by assessing the visibility properties of each transferred point.

I. INTRODUCTION

The availability of large image collections in the web has paved the way for new computer vision applications. For example, in [1] is described a vision-based approach for navigating inside a 3D scene constructed by several hundreds of images of touristic sites taken from the internet. The visual information in an image collection of a same 3D scene is usually highly redundant, since images have a large amount of overlapping content. In this work, visual redundancy is exploited so as to augment the visual content of one image of the collection, referred to as *reference*, using information from all the other images. The approach is uncalibrated, fully projective, and purely image-based (i.e., it does not require any 3D reconstruction). Moreover, no constraints are set on image acquisition conditions or scene structure. Visual augmentation has a two-fold meaning, namely (1) to enlarge the field of view of the original image, and (2) to see through the objects in the original image, thus visualizing occluded parts of the scene as it would occur with an “X-ray” view. The work extends to general 3D scenes with parallax and occlusions the usual mosaicing techniques based on homography registration [2], that cannot deal with viewpoint shifts if no constraints are set on scene structure. Related work has addressed the problem of synthesizing new views without explicit reconstruction of the 3D model of the scene. An early contribution on this topic is [3], where weakly calibrated cameras are employed. Novel views are rendered up to a projective deformation, that can be removed if cameras are fully calibrated. Other approaches to view synthesis employ the plane+parallax representation

[4]. For example, in [5] a virtual 3D camera path can be synthesized, provided that parallax information can be referred to the homography of the plane at infinity. The availability of three or more images has also been exploited for the purpose of novel view synthesis through trifocal tensors [6], [7]. Our approach propagates visual information through the basic tools of epipolar geometry and fundamental matrices. Building on a basic theoretical result on 4-node/5-edge image subgraphs established in [8], a graph representation of the image collection is constructed, allowing us to check visual transferability onto the reference image by means of an original graph traversing algorithm. In particular, we illustrate a method for transferring visual data between image pairs that do not have an overlapping visual content. This is another extension to the basic graph-based techniques used to construct global mosaics, where homographies are propagated through the graph [9]. An occlusion handling strategy based on results in oriented projective geometry is also introduced, with the goal of ensuring photoconsistency by assessing the visibility properties of each transferred point, and also to allow the visualization of occluded regions.

II. APPROACH

Given a collection of n uncalibrated images I_1, \dots, I_n of a 3D scene taken from different viewpoints, let us define an *image graph* $G = (V, E)$ with the following properties: (1) the graph has n vertexes, each representing a different image of the collection; (2) two distinct vertexes $i \in V$ and $j \in V$ are linked by an edge $\epsilon(i, j) \in E$ if the fundamental matrix F_{ij} between the corresponding images I_i and I_j is known, or may be computed. In such a case, vertexes i and j are said to be *adjacent* in the graph. Let us also denote as I_r the *reference image* onto which all the other images have to be registered. If i and j are adjacent vertexes, then any pair $({}^i\mathbf{x}, {}^j\mathbf{x})$ of corresponding points in I_i and I_j can be mapped onto I_r via *epipolar transfer* [3]:

$${}^r\mathbf{x} = (F_{ri}{}^i\mathbf{x}) \times (F_{rj}{}^j\mathbf{x}) \quad , \quad (1)$$

provided that vertex r is adjacent to both i and j . In this case, (i, j) is said to be an *r-connected pair* of vertexes. Similarly, k is referred to as an *r-connected vertex* if it belongs to at least an *r-connected pair*.

A. Conditions for point transferability

The simple point transfer mechanism of eq. 1 can only be employed for vertexes that are r -connected. However, *under certain conditions*, a non- r -connected vertex k can be transformed into an r -connected one, even if direct computation of F_{rk} is impossible—i.e., when I_k and I_r do not have an overlapping visual content. In such a case, k is said to be an r -connectable vertex, F_{rk} can be indirectly evaluated by exploiting information from other parts of the graph and, hence, points from image I_k can be transferred onto I_r .

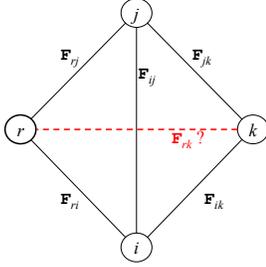


Fig. 1. Connectivity with respect to a reference vertex r . (i, j) is an r -connected pair, while k is a non- r -connected, yet r -connectable vertex (see text).

The problem of vertex connectability was addressed from a theoretical viewpoint in [8], where several graph topologies were discussed. Here we concentrate on the smallest fully connectable graph of Fig. 1, and provide for it a convenient and practical way (not requiring camera matrices and/or other 3D entities as postulated in [8]) to infer the unknown fundamental matrix from the known ones.

In the following, we provide a direct geometric construction for the unknown fundamental matrix ${}^rF_{rk}$ of Fig. 1—to follow the construction, please refer to Fig. 2. The matrix can be written as $F_{rk} = [{}^r\mathbf{e}_k] \times {}^rH_\pi$, where ${}^r\mathbf{e}_k$ is the epipole of view k in view r , and ${}^rH_\pi$ is any planar homography compatible with view r . Now, from the 3-view compatibility equations [10] ${}^r\mathbf{e}_k^\top F_{ri} {}^i\mathbf{e}_k = {}^r\mathbf{e}_k^\top F_{rj} {}^j\mathbf{e}_k = 0$, the epipole can be evaluated (epipolar transfer) as

$${}^r\mathbf{e}_k = (F_{ri} {}^i\mathbf{e}_k) \times (F_{rj} {}^j\mathbf{e}_k) . \quad (2)$$

Due to the symmetry of the configuration, the corresponding epipole in image I_k , ${}^k\mathbf{e}_r$, can be obtained in a similar way, by simply switching the indexes r and k .

Concerning the homography ${}^rH_\pi$, this can be obtained from the correspondence of the r and k views of four points in a scene plane Π . A suitable plane can be chosen as follows. Two of the four correspondences that define ${}^rH_\pi$ trivially derive from the images of the camera centers C_i and C_j : $({}^r\mathbf{e}_i, {}^k\mathbf{e}_i)$, $({}^r\mathbf{e}_j, {}^k\mathbf{e}_j)$. Similarly, a third correspondence is provided by the epipole pair $({}^r\mathbf{e}_k, {}^k\mathbf{e}_r)$. Finally, the missing fourth correspondence can be obtained as follows. Let us consider a generic point ${}^k\mathbf{x} \in I_k$. This gives rise to the two epipolar lines ${}^i\mathbf{l}_k = F_{ik} {}^k\mathbf{x} \in I_i$ and ${}^j\mathbf{l}_k = F_{jk} {}^k\mathbf{x} \in I_j$. Let us now arbitrarily choose a point ${}^j\hat{\mathbf{x}} \in {}^j\mathbf{l}_k$: this determines a

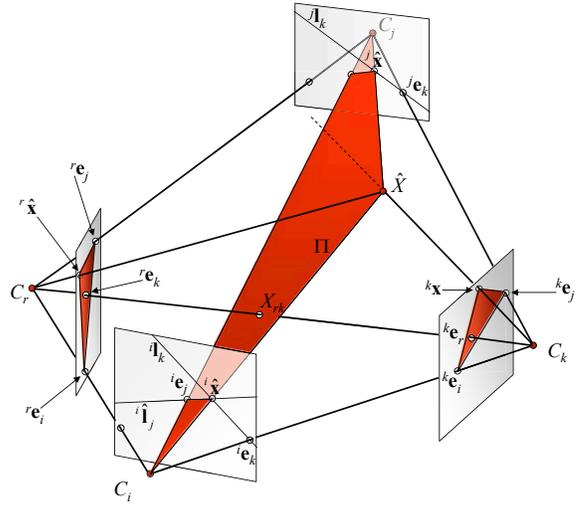


Fig. 2. The geometry of four views, with the definition of the plane Π used for the direct construction of the unknown fundamental matrix ${}^rF_{rk}$. (Best viewed in color.)

unique point \hat{X} in space as the pre-image of ${}^j\hat{\mathbf{x}}$ and ${}^k\mathbf{x}$. In the image I_i , the epipolar line ${}^i\hat{\mathbf{l}}_j = F_{ij} {}^j\hat{\mathbf{x}}$ intersects ${}^i\mathbf{l}_k$ in ${}^i\hat{\mathbf{x}}$. Π is defined as the plane through \hat{X} and the two camera centers C_i and C_j . The fourth correspondence can then be chosen as $({}^r\hat{\mathbf{x}}, {}^k\mathbf{x})$, where ${}^r\hat{\mathbf{x}}$ (i.e., the image of \hat{X} on I_r) is obtained as the intersection of the epipolar lines $F_{ri} {}^i\hat{\mathbf{x}}$ and $F_{rj} {}^j\hat{\mathbf{x}}$.

B. Graph traversal and epipolar propagation

By the technique above we can check, even without explicitly computing the associated fundamental matrices, which vertexes of the image graph are r -connectable, and are thus usable for transferring visual information onto the reference via epipolar propagation through the graph. After this analysis, which is based on an iterative graph traversal algorithm, the fundamental matrices corresponding to the newly added edges can be estimated and used as expounded in the next section.

Image graph traversal is carried out starting with an input graph $G_0 = (V, E_0)$ obtained by linking through an edge $\epsilon(i, j)$ all vertexes i and j for which F_{ij} can be estimated directly from image point correspondences using standard algorithms. The output graph is $G_f = (V, E_f)$, where the set E_f includes, beside all the edges in E_0 , all the new edges corresponding to the vertexes that were marked as r -connectable. The algorithm works as follows. The graph is traversed starting from the r vertex. For every visited vertex, r -connectability is tested. If the test is negative, the vertex is marked as re-visitabile, and the visit continues with a new vertex. If the vertex is found as r -connectable, it is added to the set Γ of r -connectable vertexes, initially set as empty. The exhaustive visit terminates when the set W_t of the re-visitabile vertexes at iteration time t is identical to the one at time $t-1$.

Fig. 3 shows a 5-vertex graph with two missing edges, corresponding to the two fundamental matrices relating the rightmost images to the reference image (the leftmost one in the figure). Such fundamental matrices cannot be computed

by point matching, due to the absence of overlapping visual content between each of the rightmost images and the reference. Nevertheless, the algorithm above allows us to assess the possibility of inferring the missing fundamental matrices, and hence the transferability onto the reference of overlapping visual content between the two rightmost images.

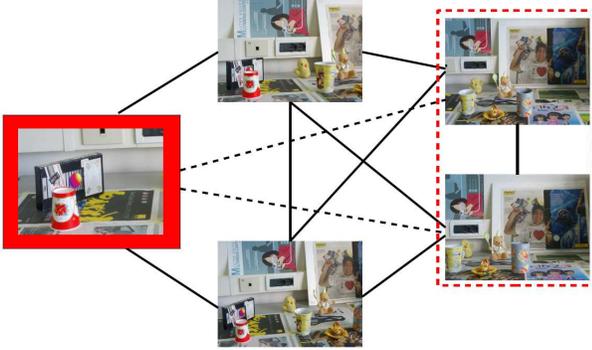


Fig. 3. A 5-vertex graph with two missing edges (image pairs have no overlapping visual content), shown as dashed lines. The corresponding unknown fundamental matrices can be inferred with the graph traversal algorithm expounded in the text, thus allowing image content transfer onto the reference image on the left. (Best viewed in color.)

III. IMPLEMENTATION

A. Estimation of fundamental matrices

1) *Direct estimation with visual overlapping*: Automatic initialization of the image graph relies on direct estimation of the fundamental matrix F_{ij} for all image pairs (I_i, I_j) with partially overlapping visual content. Visual correspondences $({}^i\mathbf{x}, {}^j\mathbf{x}) \in I_i \times I_j$ are extracted by SIFT detection and matching. The RANSAC algorithm is then run in order to get a robust estimate of F_{ij} using the epipolar constraint ${}^i\mathbf{x}^\top F_{ij} {}^j\mathbf{x} = 0$. Following the probabilistic approach used in [11] for homography validation, but using a more conservative threshold due to the more delicate nature of fundamental matrices, the RANSAC output is accepted as a valid F_{ij} if at least the 60% of all point correspondences are inliers.

2) *Indirect estimation by epipolar propagation*: An indirect estimate of the fundamental matrix between images I_r and I_k is obtained for all graph vertexes $k \in V$ marked as r -connectable. The method of construction employed in subsect. II-A to prove r -connectability is used here to get a raw estimate of F_{rk} as $[{}^r\mathbf{e}_k] \times {}^r\mathbf{H}_\pi$. Such a raw estimate is then used as the first guess solution in an iterative refinement scheme based on nonlinear optimization, working as follows. Let us first recall from subsect. II-A that, in order to be classified as r -connectable, vertex k must have been found adjacent to at least a pair (i, j) of r -connected vertexes. Hence, we have been able to estimate both F_{ik} and F_{jk} . Let $({}^k\mathbf{x}, {}^i\mathbf{x}) \in \mathcal{J}_{ik}$ be a pair of corresponding points chosen in the inlier set $\mathcal{J}_{ik} \subset I_k \times I_i$ obtained in the estimation of F_{ik} . We are now able to map, via epipolar transfer, the pair $({}^k\mathbf{x}, {}^i\mathbf{x})$ first onto I_r , then onto I_j , and finally back onto I_k :

$$\begin{aligned} {}^r\mathbf{x} &= (F_{rk} {}^k\mathbf{x}) \times (F_{ri} {}^i\mathbf{x}) \\ {}^j\mathbf{x} &= (F_{jr} {}^r\mathbf{x}) \times (F_{ji} {}^i\mathbf{x}) \\ {}^k\tilde{\mathbf{x}} &= (F_{kj} {}^j\mathbf{x}) \times (F_{ki} {}^i\mathbf{x}) \end{aligned} \quad (3)$$

After transferring in this way all pairs in the inlier sets \mathcal{J}_{ik} and \mathcal{J}_{jk} , F_{rk} can be refined by minimizing the following quadratic cost function:

$$\mathcal{E}_{k,(i,j)} = \sum_{\mathcal{J}_{ik}} \|{}^k\tilde{\mathbf{x}} - {}^k\mathbf{x}\|^2 + \sum_{\mathcal{J}_{jk}} \|{}^k\tilde{\mathbf{x}} - {}^k\mathbf{x}\|^2. \quad (4)$$

3) *Global refinement*: Once all computable F 's have been obtained, a further global refinement of all estimates is carried out. The adjustment mechanism is based on the cost function \mathcal{E}_k used in eq. 4. The global error to be minimized is

$$\begin{aligned} \mathcal{E} &= \sum_{i,j,k \in V} \mathcal{E}_{k,(i,j)} \\ \epsilon(k,i), \epsilon(k,j) &\in E \end{aligned} \quad (5)$$

Notice that while in the minimization scheme of eq. 4 only matrix F_{rk} is updated and the others remain fixed, in eq. 5 all the fundamental matrices are simultaneously updated at each step.

B. Dense stereo matching

We have described above how to transfer onto I_r a single pair of corresponding points of an r -connected pair (i, j) . We address here the problem of obtaining all possible correspondences from image pairs, in order to transfer as many points as possible onto the reference image. Points that can be put into correspondence in an image pair are said to be *matchable*. *First class* matchable points are all the SIFT points used to estimate the fundamental matrices. *Second class* matchable points can be recovered by guided matching along epipolar lines [10]. First and second class matchable points can be used as seeds for a dense matching algorithm [12], based on region growing, in order to set visually overlapping textured regions into pointwise correspondence. The output of this algorithm is a set of new, dense matchable points. Figs. 4(c) and (d) show (in blue) the matchable points extracted from the image point pair of Figs. 4(a) and (b).

Matchable points are not the only transferrable points. Indeed, even image points belonging to untextured image regions (for which no SIFT matching or other kinds of image search are possible) can be transferred onto the reference image, provided that they belong to image regions that are in correspondence to one another. Region-based correspondence is implemented as a variant of the pointwise dense matching algorithm mentioned above. In this variant, region growing is modified so as to take into account untextured image regions with uniform color. Figs. 4(e) and (f) show (in yellow) the points extracted with the above region-based matching. Transfer of points belonging to uniform regions can be accomplished by Delaunay triangulation of the regions (using matchable points on the region border as triangle vertexes) and affine mapping of all triangles. The process is illustrated

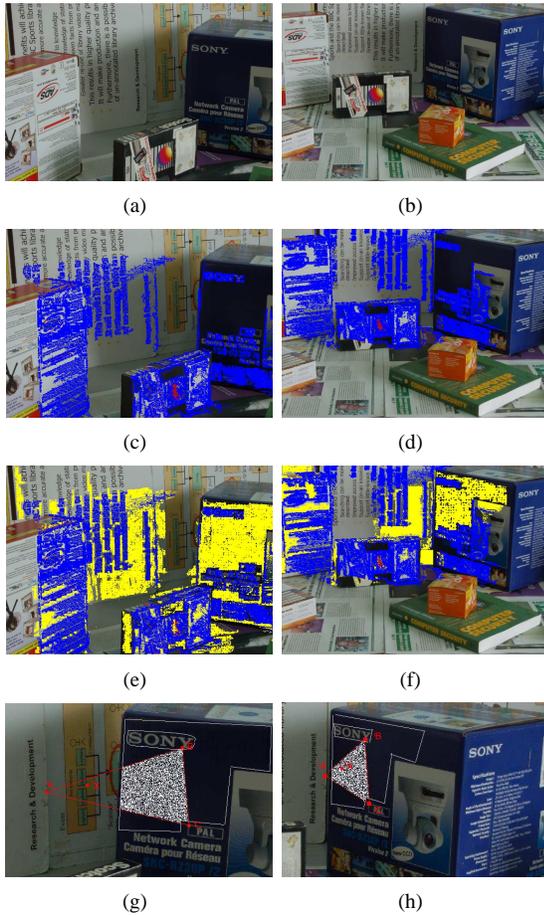


Fig. 4. Transferrable points recovered after classification. (a),(b): Input image pair. (c),(d): Matchable points (in blue). (e),(f): Points belonging to matchable regions (in yellow). (g),(h): Transfer of uniform image regions (dashed white line) is done by region triangulation followed by affine triangle mapping. (Best viewed in color.)

in Figs. 4(g) and (h). Note that only the points belonging to corresponding uniform regions are transferred.

C. Handling occlusions

The epipolar transfer mechanism of eq. 1 does not provide any information about the actual visibility of a transferred point. Hence, in order to preserve photoconsistency, a check for possible occlusions must be carried out for all points that are mapped onto the reference. This problem is dealt with using results from oriented projective geometry [13]. Let us consider the points ${}^i\mathbf{y} \in I_i$ and ${}^j\mathbf{y} \in I_j$, with the same 3D point Y as pre-image. Similarly, let us consider points ${}^i\mathbf{z} \in I_i$ and ${}^j\mathbf{z} \in I_j$, with pre-image Z . If Y and Z belong to the same optical ray from C_r , then they will both be projected onto the same point ${}^r\mathbf{x} \in I_r$. In order to find which, between Y and Z , is closer to C_r , and is then visible, we can observe the relative position assumed by their images ${}^i\mathbf{y}$ and ${}^i\mathbf{z}$, along the line ${}^i\mathbf{l}_r = F_{ir}{}^r\mathbf{x}$, with respect to the epipole ${}^i\mathbf{e}_r$ (see Fig. 5). If C_r is in front of the focal plane Φ_i , then the unoccluded point to be transferred is the one closest to ${}^i\mathbf{e}_r$. Conversely, if C_r is behind Φ_i , then the right point to transfer is the furthest one

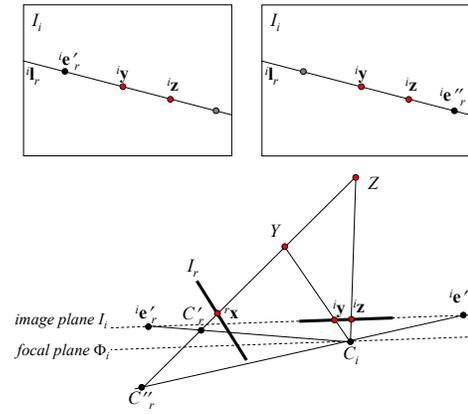


Fig. 5. Assessing visibility through point ordering in projective oriented geometry.

from ${}^i\mathbf{e}_r$. The position of the optical center C_r with respect to Φ_i can be estimated through the *cheirality* χ defined in [10]. This is a signed scalar quantity that can be computed after recovering the camera matrices $P_i \in P_r$ from F_{ri} . If $\chi > 0$ then C_r is in front of Φ_i , while C_r is behind Φ_i if $\chi < 0$.

The mechanism above can be used to choose, among several correspondences taken from a same image pair I_i and I_j that are mapped onto the same reference point, the one related to an unoccluded 3D point. Things are a bit more complicated in the case of conflicting image correspondences that belong to different image pairs. Let us consider for example the two pairs $({}^i\mathbf{y}, {}^j\mathbf{y}) \in I_i \times I_j$, and $({}^h\mathbf{z}, {}^k\mathbf{z}) \in I_h \times I_k$, both of which are transferred onto the same reference point ${}^r\mathbf{x} \in I_r$. To decide for visibility using the method above, it is required that at least another point, say ${}^i\mathbf{z}$, be computed. This is possible in the case that at least one between vertexes h and k are i -connectable.

IV. EXPERIMENTS

Results of experiments with two image collections are shown.

In a first set of experiments, a collection of six images of an indoor scene taken from distinct viewpoints was used (see Fig. 6). In order to match the augmented view with a ground-truth, the reference image was obtained by cropping a wider image containing all the scene (see Fig. 6(a)). Fig. 6(b) shows the augmented view of the scene obtained with our method. Notice that some regions in the augmented view are missing: This is due to the presence of textureless regions in the scene (for which the adopted dense stereo matching algorithm cannot work), and also to the presence of regions that were visible in at most only one image of the collection, and therefore could not be transferred (see, in particular, the region between the toy-bear and the mug on the right). A larger image collection would produce a more dense image rendering. Fig. 6(c) shows the standard mosaic obtained from the image collection. Notice that the mosaic is grossly incorrect, with evident ghosting effects and object replications, due to the presence of parallax in the image collection. Our approach prevents this problem

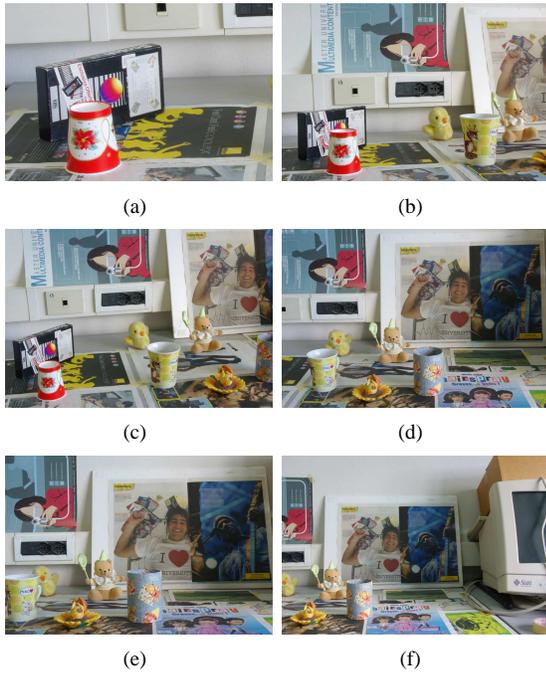


Fig. 6. The image collection for the indoor scene. The reference image is the first of the group.

to occur, and renders the super-image maintaining projective consistency. Fig. 8 shows the graph associated to the collection of Fig. 6: Solid lines are the original edges of set E_0 , for which direct computation of fundamental matrices was possible; edges $\epsilon(3, r)$, $\epsilon(4, r)$ and $\epsilon(5, r)$ (dashed lines) were added as the result of graph traversal. Three edges were inferred by graph traversal.

The ground truth image provides also a way to compute the transfer error for all matched SIFT points between the r -connected image pairs and the ground truth image. Tabs. I, II and III show the average μ and variance σ^2 of the transfer error for different image pairs and point transfer approaches.

In particular, Tab. I shows the performance when the visual content shared by I_1 and I_2 is carried onto I_r . Besides our approach (labeled as “F via r -connectability”), two other approaches based on the trifocal tensor T_{r12} (images I_1 and I_2 have a common overlap with the reference image I_r) were tested. T_{r12} allows us to obtain directly from a pair of corresponding points $(^1\mathbf{x}, ^2\mathbf{x})$ the value of $^r\mathbf{x}$ in I_r ; this approach is labeled as “points via T_{r12} ”. Furthermore, from the trifocal tensor T_{r12} all the three fundamental matrices F_{r1} , F_{r2} e F_{12} can be evaluated and exploited for epipolar transfer (“F via T_{r12} ” approach).

Tab. II shows the performance when the visual content shared by I_3 and I_4 is carried onto I_r . Also in this case F via r -connectability approach is compared w.r.t. two other techniques. The strategy labeled as “chaining F” maps the common points of I_3 and I_4 onto I_1 and I_2 ; the mapped points are then propagated onto I_r by the r -connectability approach. The other approach relies again on the trifocal tensor T_{r12} :



Fig. 7. (a): The ground-truth image. (b): Our augmented view. (c): Standard mosaic (note the ghosting effects due to parallax).

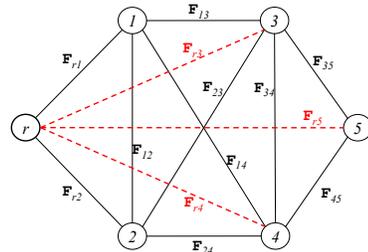


Fig. 8. The image graph for the collection of Fig. 6. Image (a) is the reference vertex r . The edges of the initial graph are indicated by solid lines. The red dashed lines are the edges added after graph traversal.

given two lines $(^1\mathbf{l}, ^2\mathbf{l})$ related to the same line L in space, the projection $(^r\mathbf{l})$ of L in I_r can be obtained by exploiting T_{r12} . In Our case $(^1\mathbf{l}, ^2\mathbf{l})$ are the epipolar lines on I_1 and I_2 related to a point \mathbf{x} in I_3 or I_4 . This approach is labeled as “epipolar lines via T_{r12} ”.

Tab. III shows the results related to the visual content shared by I_5 with I_3 and I_4 : in this case our approach has been compared only with the chaining F approach. Generally speaking, performances on y values are better than the results obtained for the x coordinate: This is due to the fact that all the epipolar lines involved are quite horizontal. The worst behavior of the approaches based on trifocal tensor are explained by the fact

that the tensor was computed on fewer point correspondences (triplets of points are required) w.r.t. the fundamental matrices. Our direct “F via r -connectability” approach outperforms also the “chaining F” strategy, since it avoids that the error propagates throughout the vertexes.

Method	μ_x	σ_x^2	μ_y	σ_y^2
F via r -connectability	3.19	6.87	0.98	0.49
F via T_{r12}	13.11	27.03	9.90	20.40
points via T_{r12}	9.09	138.36	6.20	15.50

TABLE I

ERROR PERFORMANCE FOR THE PROPAGATION, ONTO I_r , OF THE VISUAL CONTENT SHARED BY I_1 AND I_2 .

Method	μ_x	σ_x^2	μ_y	σ_y^2
F via r -connectability	5.30	36.04	3.06	10.64
chaining F	11.90	112.89	2.85	13.90
epipolar lines via T_{r12}	11.27	51.99	7.60	19.47

TABLE II

ERROR PERFORMANCE FOR THE PROPAGATION, ONTO I_r , OF THE VISUAL CONTENT SHARED BY I_3 AND I_4 .

Method	μ_x	σ_x^2	μ_y	σ_y^2
F via r -connectability	8.23	92.41	4.68	3.20
chaining F	13.90	125.80	2.85	14.90

TABLE III

ERROR PERFORMANCE FOR THE PROPAGATION, ONTO I_r , OF THE VISUAL CONTENT SHARED BY I_5 WITH I_3 AND I_4 .

An example of “X-ray” view is reported in Fig. 9, where the mug and the toy-bear are removed showing what is hidden behind them. Notice that hidden objects are rendered in the projective frame of the reference image.



Fig. 9. (top left): What’s behind the mug and the toy bear? (top right): Answer: Chick, poster detail. (bottom, left and right): “X-ray” view through the mug and toy-bear, and details of the two occluded regions. (Best viewed in color.)

Fig. 10 shows another image collection, an outdoor one, including touristic photos of Neptune’s fountain in Piazza della Signoria, Florence, Italy. In this case, the inference machinery was able to recover one missing fundamental matrix. Fig. 11

shows the augmented view and the standard mosaic obtained with the outdoor image collection. For the standard mosaic, the erroneous image alignment due to parallax is even more evident than in the previous case. Notice also that our augmented view, although providing a geometrically correct scene, is not illumination-compensated. This is particularly evident along the line of separation between the reference image (a) and image (e). A qualitative insight into our approach is gained by inspection of Figs. 12 and 13. The first figure illustrates photo-consistency performance in terms of straight line preservation for the overall augmented view. The second figure shows the characteristics of image-based rendering in the presence of occlusions. Notice how the occluded, and hence unrecovered, image region inside the square box reflects the shape of the occluding object, as if it was a shadow projected onto the fountain basement by a light source placed in the camera center of the left view.

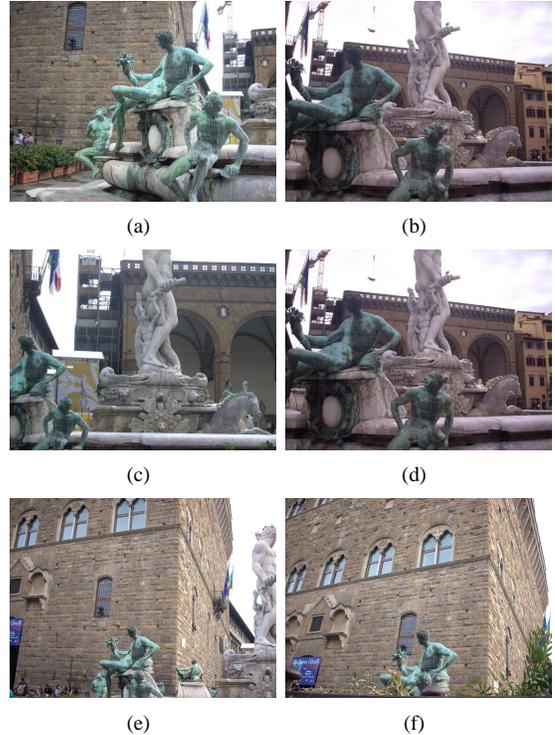


Fig. 10. Outdoor image collection. The reference image is the first of the group.

V. CONCLUSIONS AND FUTURE WORK

We have present an approach for constructing multi-view *epipolar networks*. These are graphs with images as vertexes and fundamental matrices as edges. After a first set of edges is obtained from the visual data at hand, an iterative procedure is employed so as to infer the missing edges from the already available ones. As a result, new vertex pairs are connected, and epipolar information is propagated throughout the network. The approach was demonstrated in an image-based rendering scenario, where a (small) set of uncalibrated images



(a)



(b)

Fig. 11. (a): Our augmented view for the outdoor scene. (b): Standard mosaic, with parallax-induced artifacts.

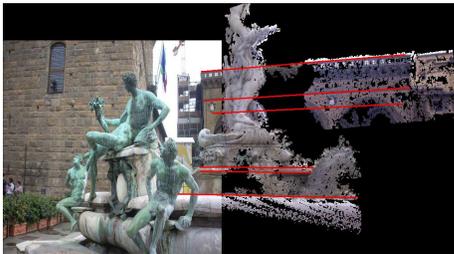


Fig. 12. Qualitative assessment of projective photoconsistency. Straight lines in the scene remain straight also in the augmented view.

of a general 3D scene taken from multiple viewpoints were merged into a single super-image while maintaining overall projective coherence. To this aim, results in oriented projective geometry were also discussed and exploited, to both detect and handle occlusions by assessing the visibility properties of each transferred point. Dense stereo techniques were also illustrated, that allow both pointwise and region-based visual transfer. Our epipolar network-based method extends the usual mosaicing techniques to image collections with 3D parallax, thus operating like a virtual sensor provided with an enlarged field of view and the capability of seeing through visual occlusions in an “X-ray” fashion.

Future work will address incorporating more basic image subgraphs from [8] inside our graph traversal inference machine. In addition, experiments with larger image collections than those presented here will be performed, so as to test the way our approach scales with respect to collection size in terms of photoconsistency accuracy, rendering density, and computation time. A further lesser refinement of the algorithm will be to incorporate in the approach an illumination compensation technique.



(a)

(b)



(c)

Fig. 13. (a)(b): Two r -connected images. The red frames mark image regions where an occlusion arises. (c) In the augmented view, the right part of the occluded area has the shape of the occluding arm of the statue. (Best viewed in color.)

REFERENCES

- [1] R. S. N. Snavely, S.M. Seitz, “Modeling the world from internet photo collections,” *International Journal of Computer Vision*, 80(2):189–210, 2008.
- [2] R. Szeliski, *Computer Vision: Algorithms and Applications*. Springer Verlag, 2010.
- [3] S. Laveau and O. Faugeras, “3-D scene representation as a collection of images and fundamental matrices,” INRIA, Tech. Rep., 1994.
- [4] M. Irani and P. Anandan, “Parallax geometry of pairs of points for 3d scene analysis,” in *Proc. ECCV*, 1996.
- [5] A. Fusiello, “Specifying virtual cameras in uncalibrated view synthesis,” *IEEE Trans. on Circuits and Systems for Video Technology*, 17(5):604–611, 2007.
- [6] R. Hartley, “Lines and points in three views and the trifocal tensor,” *International Journal of Computer Vision*, 22(2):125–140, 1997.
- [7] S. Avidan and A. Shashua, “Novel view synthesis by cascading trilinear tensors,” *IEEE Trans. on Visualization and Computer Graphics*, 4(4):293–306, 1998.
- [8] N. Levi and M. Werman, “The viewing graph,” in *Proc. IEEE CVPR*, 2003.
- [9] D. Capel, “Image mosaicing and super-resolution,” Ph.D. dissertation, University of Oxford, 2001.
- [10] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2004.
- [11] M. Brown and D. Lowe, “Automatic panoramic image stitching using invariant features,” *International Journal of Computer Vision*, 74(1):59–73, 2007.
- [12] M. Lhuillier and L. Quan, “Match propagation for image-based modeling and rendering,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(8):1140–1146, 2002.
- [13] S. Laveau and O. Faugeras, “Oriented projective geometry for computer vision,” in *Proc. ECCV*, 1996.