A fast and reliable planar registration method with applications to document stitching

Maurizio Pilu
Hewlett-Packard Laboratories
Bristol BS34 8QZ, UK
mp@hpl.hp.com

Francesco Isgro'
Heriott-Watt University
Edinburgh EH14 4AS, UK
fisgro@cee.hw.ac.uk

Abstract

This paper presents a fast and extremely robust feature-based method for planar registration of partly overlapping images that uses a two-stage robust fitting approach comprising a fast estimation of a transformation hypothesis (that we show is highly likely to be correct) followed by a confirmation and refinement stage. The method is particularly suited for automatic stitching of oversize documents scanned in two or more parts. We show simulations, also supported by practical experiments, that prove both the robustness and computational efficiency of the approach.

1 Introduction and motivation

Image mosaicing, which consists of assembing multiple overlapping images of the same scene into a single, larger image, is an old topic in image processing and photogrammetry. Earlier research was for scientific or military applications such as stitching together different aerial images, sea-floor modeling, etc. More recently, and thanks to the success and large diffusion of digital photography, image mosaicing reached the mainstream consumer market with algorithms and tools for the creation of panoramic views from a set of overlapping shots (see e.g. [7]). Some software companies are also producing low-cost PC software for generating panoramic views such as Quick TimeVR, PhotoVista and PanaVue.

A particular practical problem where mosaicing techniques can be applied is for creating a digital copy of a large document. This can be done by mosaicing through cameras images [9, 10], or more commonly by using a flatbed scanner of smaller size, for instance when we have an A3 document and an A4 scanner.

Although a number of algorithms and commercial packages exist that address this problem, an investigation into many of them revealed serious drawbacks that prevented their practical use.

Firstly and above all there is the problem of robustness. While most of these packages might work well with photos, they have proven unreliable with documents, where the periodicity of text lines and structures such as words tend to

BMVC 2002 doi:10.5244/C.16.67

cause local minima and thus erroneous registration. Methods that track displacements over a large number of sub-images such as [10], where a document was mosaiced from an overhead video camera, might not suffer from this problem. However, in the case of document stitching[1] for flatbed scanners, image samples are necessarily minimal and typically with an unknown large displacement and possibly rotation.

Secondly, documents generally need to be at high resolution, causing most methods to fall short of performance, taking in the order of minutes to register document portions.

Thirdly, we wanted the method to be fully automatic, that is without requiring any manual intervention beyond knowing the relative position of the document portions with respect to each other. This requirement further stresses most registration methods since the "search area" becomes bigger and both the computational load and the likelihood of mis-registration are further increased.

This paper presents a fast and extremely robust planar registration method that overcomes all the problems outlined above. The method is basically a two-stage feature-based robust estimation method that quickly searches for a small number of *consistent* matches defining a transform that is highly likely to be the correct one, and then cheaply gathers further evidence for both validating and refining the transform estimate. The consistency test is based on the fact that if matched features are independently found and they happen to be defining the same Euclidean rigid transformation, it can be shown (Section 5) that with high likelihood the transformation is the correct one, a principle similar to that used by Viola [8] , for instance.

2 Overview of the registration method

The algorithm we have developed is outlined in the self-explaining pseudo-code of Figure 1. We find points of interest in the first image I' (Section 3). We then start picking random triplets of points $\mathbf{p}'_1$, $\mathbf{p}'_2$ and $\mathbf{p}'_3$ in the first image and search for their best matches $\mathbf{p}''_1$, $\mathbf{p}''_2$ and $\mathbf{p}''_3$in the second image I'' via a correlation procedure (Section 4). We stop sampling when the triplet is *consistent* with a single Euclidean rigid transform E_{12}, a situation that we show in Section 5 almost ensures that the transform is the correct one. Then we search for N other supporting matches in smaller search windows and apply a Least Median of Squares fitting to refine the estimate (Section 6).

Section 7 will overview some more specific implementation issues in the context of using the method for document stitching from flatbed scanners.

3 Feature extraction

The first step is to extract a set of interest points from the first image I'. We decided to use intensity corners as interest points; such features have been already used for mosaicing in a number of works, e.g. [11]. In particular we use the corner

[1] We shall henceforth indicate with *stitching* the mosaicing problem applied to documents scanned or captured in multiple parts with arbitrary overlap.

```
• Extract interest points in $I'$
• Find three consistent matches
    – While not found
        * Select three random points $\mathbf{p}'_1$, $\mathbf{p}'_2$ and $\mathbf{p}'_3$
        * Correlate each $\mathbf{p}'_i$ in large search window $\Gamma_i$ in $I''$
        * Identify best matched points in $I''$ as $\mathbf{p}''_1$, $\mathbf{p}''_2$ and $\mathbf{p}''_3$
        * Verify Euclidean consistency of matching triplet
        * If consistent then found=TRUE
    – Compute $\mathbf{E_0}$ with the matches $(\mathbf{p}'_1, \mathbf{p}''_1)$, $(\mathbf{p}'_2, \mathbf{p}''_2)$ and $(\mathbf{p}'_3, \mathbf{p}''_3)$
• Seek additional $N$ matches
    – While $i < 3 + N$
        * select random interest point $\mathbf{p}'_i$ in $I'$
        * search for match in small window centered around $\mathbf{E}_0\mathbf{p}'_i$
        * if correlation score is good then $i = i + 1$
• Robust fit to N+3 matches
```

Figure 1: Overview of the planar image registration algorithm proposed in this paper.

extractor described in [6] called SUSAN (Smallest Univalue Segment Assimilating Nucleus). This algorithm is fast since it does not use any derivative on the image to perform the corner enhancement. The number of points extracted depends of course on the amount of information in the image, but also on the threshold value set for the feature strength. For our purposes it is important that a large ($>$ 100) number of points distributed all over the image are identified in the first image in order to be sure that we can determine enough correspondences to accurately estimate the transformation. If the number of points extracted is not large enough we keep reducing the threshold value until a reasonable quantity of points is extracted.

4 Feature matching

In order to find the correspondent point $\mathbf{p}''_i$ in the second image $I''$of a point $\mathbf{p}'_i$ in the first image I' we use an intensity-based normalized cross-correlation technique. The technique works by defining a rectangular search window Γ in the second image and a rectangular correlation mask Λ. For each pixel $\mathbf{p}''$ in Γ the mask Λ defines two correlation windows in the two images, one centered in $\mathbf{p}''$ and the other one centered in $\mathbf{p}'_i$. By taking into account also a small rotation between the images, the correspondent candidate is the point $\mathbf{p}''_i$ maximizing the correlation coefficient.

In order to speed up the computation, in our implementation we rotate by an opportune angle the correlation window centered in $\mathbf{p}'_i$, and use correlation as per the translation case only. Moreover *the correlation is performed only with those points in the search window having mean value in a small neighborhood almost equal to the mean value of an equally sized neighborhood of the point* $\mathbf{p}''$.

In general the range for rotation angle need not be from 0° and 359° but in our document stitching problem we allow a skew between the two images no higher than ±10°.

The value correlation coefficient is a real number between -1 and 1 and this helps to control the quality of the correspondence found, which is done by setting a threshold value. A way for validating the correspondence ($\mathbf{p}'_i$,$\mathbf{p}''_i$) is to track back $\mathbf{p}''_i$ in the first image and accept the correspondence only if this second search by correlation returns $\mathbf{p}'_i$ as correspondent point [5]. This approach is computationally too expensive and so we preferred to use large correlation windows that, as observed in [3], increases the probability of a good match, especially in presence of a large amount of texture.

It will be apparent that the method presented in this paper is not limited to using cross-correlation and could employ other local similarity methods and techniques [1].

5 Matches consistency

As said earlier, the method presented in this paper uses an efficient strategy whereby we perform full-search correspondence in large search windows Γ_i until we find three point correspondences $(\mathbf{p}'_1, \mathbf{p}''_1)$, $(\mathbf{p}'_2, \mathbf{p}''_2)$, $(\mathbf{p}'_3, \mathbf{p}''_3)$ that are consistent with respect to a threshold value δ, where the relation of consistency is explained by the following definition:

Definition 1 *Let* $\mathbf{E}_{12}$ *be the Euclidean rigid transformation determined by the two point correspondences* $(\mathbf{p}'_1, \mathbf{p}''_1)$, $(\mathbf{p}'_2, \mathbf{p}''_2)$ *such that* $\mathbf{E}_{12}\mathbf{p}'_1 = \mathbf{p}''_1$. *We say the three point correspondences are consistent with respect to* $\mathbf{E}_{12}$ *if and only if the following two relations hold*

$$||\mathbf{E}_{12}\mathbf{p}'_2 - \mathbf{p}''_2|| < \delta \tag{1}$$

$$||\mathbf{E}_{12}\mathbf{p}'_3 - \mathbf{p}''_3|| < \delta \tag{2}$$

At this point we need to asses whether Definition 1 is enough to guarantee that the transformation $\mathbf{E}_{12}$ consistent with the three point correspendences is the correct one. This will of course be the case if the three point correspondences are correct (inliers) but since it is of course possible to have three mismatches which are consistent[2], it is desirable that this situation occur with a very low probability, for which we are going to derive an estimate in the next section.

5.1 Estimation of false positive probability

We need to estimate the probability that when three matches are consistent under an Euclidean rigid transformation, the transformation itself is the correct one. The probability however is more easily estimated by reversing the argument, that is we need to find the probability P_{fp} of the matches to contain outliers while still defining a consistent, yet incorrect, transformation (a false positive).

[2] E.g. think of three random points in a planar region that are aligned to form a line.

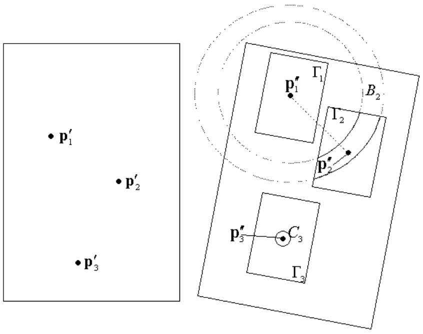

Figure 2: Graphical construction used for the estimation of the probability P_{fp} that outlier matches may be consistent with a single Euclidean rigid transform.

This probability is given by $P_{\text{fp}} = P_1 P_2 P_3$ where P_1 is the probability of the point correspondences to be outliers, P_2 is the probability of two matches to identify an Euclidean rigid transformation, and finally P_3 is the probability of the *third match* to comply to the same transformation determined by the other two.

In order to approximately derive these three probabilities, let us assume that ε is the probability of a search to fail and hence generate an outlier match $(\mathbf{p}'_i, \mathbf{p}''_i)$ and that $\mathbf{p}''_i$ can occur with uniform probability anywhere in Γ_i. Let us also indicate by $A(\Gamma_i)$ the area of the search region Γ_i, by $L(\Gamma_i)$ a typical dimension[3].

It is evident that an incorrectly consistent matched triplet could arise only from either two outliers and a single inlier match or from three outlier matches, since the Euclidean rigid transformation generated by two inliers is by definition the correct one and the single outlier could not be consistent and verify Eqns. 1 and 2. Hence we have

$$P_1 = \varepsilon\varepsilon\varepsilon + \varepsilon\varepsilon(1-\varepsilon). \tag{3}$$

The probability P_2 is determined through geometrical considerations with the aid of Figure 2. If we determine a planar transformation with two matches $(\mathbf{p}'_\mathbf{1}, \mathbf{p}''_\mathbf{1})$ and $(\mathbf{p}'_\mathbf{2}, \mathbf{p}''_\mathbf{2})$ independently shought in the search region Γ_1 and Γ_2, there is no guarantee that this transform is rigid. In order for the transformation to be rigid the second matched point $\mathbf{p}''_2$ should fall anywhere in a circular tollerance band $\mathcal{B}_2$ (bounded by the two dashed circles in Figure 2) distant $||\mathbf{p}'_2 - \mathbf{p}'_1||$ from $\mathbf{p}''_1$ and 2δ wide . Given that we searched for the match within Γ_2 and that an outlier match can occur anywhere and with the same probability in Γ_2, P_2 is approximately given by the ratio between the area of $\mathcal{B}_2 \cap \Gamma_2$ and the area of Γ_2, or

$$P_2 \cong \frac{A(\mathcal{B}_2 \cap \Gamma_2)}{A(\Gamma_2)} \cong \frac{2\delta \cdot L(\Gamma_2)}{A(\Gamma_2)} \tag{4}$$

Let now $\mathcal{C}_i$ be the region used to test the consistency of a transformation of a match $\mathbf{p}''_i$ to $\mathbf{p}'_i$ and let $A(\mathcal{C}_i) \cong 4\pi\delta^2$ be its area. P_3 is the probability that a third

[3] E.g. for a square region, one of its sides, or for a circular region its diameter

outlier match is consistent with the Euclidean rigid transformation determined by the first two matches, that is $\mathbf{E_{12}p'_3} \in \mathcal{C}_\mathbf{3}$). Using a similar geometric argument as for the estimation of P_2, the uniform distribution of the occurrence of an outlier match in Γ_3 implies that the probability P_2 is the ratio between the area of the tolerance region $A(\mathcal{C}_3)$ and the area of the whole search region $A(\Gamma_3)$, or

$$P_3 \cong \frac{A(\mathcal{C}_3)}{A(\Gamma_3)} \cong \frac{4\pi\delta^2}{A(\Gamma_3)} \tag{5}$$

In actual terms, let us immagine that $\varepsilon = 0.2$, $\delta = 5$ and let Γ_i, $i = 1 \ldots 3$ be a square region of 100x100 pixels. Then $A(\Gamma_i) = 10000$, $L(\Gamma_i) = 100$, $A(\mathcal{C}_3) = 63$ With these values we have $P_{\text{fp}} = 2.5 \times 10^{-5}$, which is sufficiently low to assume that in the overwhelmingly large majority of cases whenever three matches are consistent as of Definition 1 the transformation thus determined is the correct one.

5.2 Simulation results

In order to check the robustness of the method we performed some experiments on sets of synthetic points corrupted by Gaussian noise (with standard deviation $0 \leq \sigma \leq 3$), and with a fraction of outliers up to 0.5. We then randomly selected a large number triplets and computed: a) the ratio between the number of triplets including only inliers and the number of such triplets passing the consistency test; b) the ratio between the number of triplets including at least two outliers and the number of such triplets passing the consistency tests (false positives).

In Tables 1 and 2 we show the results of the simulations for different values of the consistency test threshold δ, expressed as function of the Gaussian noise. We can see that up to a threshold of $\delta = 6\sigma$ the fraction of false positives is negligible(Table 1). From Table 2 we see that, as expected, the smallest the threshold the more likely is a good triplet to fail the test, which might cause the algorithm to take more trials to find a consistent triplet. A good compromise is given by the threshold of 6σ.

5.3 Notes on the consistency test

Different kinds of consistency test on an Euclidean rigid transform could be worked out for two matches only, in which case the probability of a false positive would be $P_{\text{fp}} = P_1P_2$, or with four matches, in which case we would have $P_{\text{fp}} = P_1P_2P_3P_3$. We have however found by experimentation that the best trade off between robustness and number of expensive searches by correlation is by using three matches. Typically, in fact, four to six full-size searches for correspondence in I'' are necessary to find a consistent matching triplet that, by virtue of the low false-positive probability shown above, will be highly likely to be the correct one.

Compared to classic robust estimation methods where a high number of matches are sought and then robustly fit to a model, the present method can be seen as a two-stage robust estimation using a "hypothesize and test" paradigm, with the notable exception that although the hypothesis we make is obtained relatively cheaply, it tends to be correct.

threshold	3σ	4σ	6σ	8σ
min	0.0000571	0.0000560	0.0000561	0.5969463
mean	0.0002750	0.0004848	0.0007705	0.6391297
max	0.0033333	0.0088	0.0111661	0.6876274

Table 1: Ratio between the number of triplets including outliers and false positives. We show, for different thresholds, the minimum, mean value and maximum of the consistency test false positive probability over the values computed with $0 \leq \sigma \leq 3$ and fraction of outliers between 0 and 0.5.

threshold	3σ	4σ	6σ	8σ
min	0.1165533	0.2218625	0.4367468	0.5885839
mean	0.1421841	0.2572192	0.4785129	0.6391914
max	0.179039	0.3029571	0.5286849	0.6876274

Table 2: Ratio between number of triplets including only inliers and postives.. We show, for different thresholds, the minimum, mean value and maximum of the consistency test false positive probability over the values computed with $0 \leq \sigma \leq 3$ and fraction of outliers between 0 and 0.5.

6 Robust fitting via LMS

Once the initial transformation $\mathbf{E}_{12}$ is established by the process outlined in the previous section, we use the three point matches $(\mathbf{p}'_1, \mathbf{p}''_1)$, $(\mathbf{p}'_2, \mathbf{p}''_2)$ and $(\mathbf{p}'_3, \mathbf{p}''_3)$ to find a more accurate, new initial transformation $\mathbf{E}_0$. Next , we seek support for it by picking random interest points $\mathbf{p}'_i$ in the first image and searching if there are highly correlated pixels in a small neighbourhoods of $\mathbf{E}_0\mathbf{p}'_i$. However this does not imply that all the point correspondences established are correct. Therefore a robust estimation process is necessary to accurately determine the of the final tranform $\mathbf{E}$ after the feature matching procedure is complete. Although we do not expect a high fraction of outliers we preferred to make our estimation of $\mathbf{E}$ by means of a Least Median of Squares [2], which is able to cope with a fraction of outliers up to 0.5. The reasons for this choice are twofold: a) we make the stitching robust even in cases of of documents where repetitive patterns occur (this can create outliers in the matching process); b) given the small number of subsamples needed (~ 17) by the LMS, the speed of the algorithm is not affected by this step.

We preferred LMS to other sobust methods such as RANSAC, which is usually faster, because LMS does not need any *a-priori knowledge* such as error thresholds to be set.

Briefly LMS randomly selects a subset of p observations and uses only them to estimate the model. For each estimated model, the median of the square of the residuals is computed, and the temporary best model is the one which minimizes the objective function $\text{med}_i r_i^2$, where $r_i = ||\mathbf{E}\mathbf{p}' - \mathbf{p}''||$. Ideally the trials should be run on *each* subset of p observations, but this is usually computationally impossible. Thus the number m of trials is chosen in such a way to have a probability Υ that a subsample without any outlier has been included in our selection. The estimated model allows us to detect the outliers, if any, and to refine the estimation

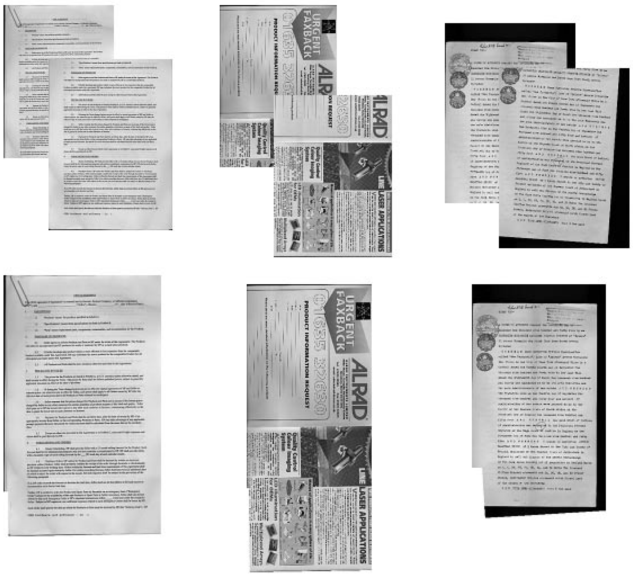

Figure 3: Examples of stitching various documents scanned in two parts.

using a standard Least Squares technique.

7 Application to document stitching

As anticipated in the introduction the main motivator for developing this registration algorithm was for performing fully automatic document stitching with applications to flatbed scanners.

We have used the algorithm in a flatbed stitching application that was able to stitch together two overlapping parts of an oversize 300dpi document, brochure, etc, scanned separately on an ordinary A4 flatbed scanner [4].

As is often the case, a number of additions to the basic algorithms were necessary.

First we impose and order on the scanning of the document portions, e.g. top first and then bottom. This was not strictly necessary but it was incidentally found to be also a natural way for users to scan in multiple document parts.

Moreover it is possible to set what type of document we are scanning. For instance a US Legal ($14'' \times 8.5''$) document scanned over an A4 scanner ($11.69'' \times 8.27''$) is best scanned in two parts and the displacement between top and bottom is typically $3''$, which we use to reduce the size of search windows Γ_i.

Second, once the transformation has been found, we need to rotate a high resolution image and paste it into the reference frame of another one. Since the overlap can be substantial, we are rotating and pasting only the non-overlapping part of the second image plus a small portion of the overlapping region that is

Number of Matches N	30
Y_{OFF}	$\frac{h}{4}$
Height of the search window Γ	$\frac{h}{2}$
Width of the search window Γ	$\frac{w}{5}$
Height of the correlation window	$\frac{h}{32}$
Width of the correlation window	$\frac{w}{40}$
Correlation threshold	0.8
Consistency test threshold δ	3.0
Range of angle θ	$-10 \leq \theta \leq 10$

Table 3: Parameters for the case of stitching document of unknown length. w indicates the number or columns in the input images and h the number of rows.

used only for blending the two images.

In order to minimize possible artifacts we determine the boundary of the overlapping region by seeking a path of low texture, which for text documents (see e.g. Figure 3) would typically be the spacing between the lines. In our implementation the boundary is simply the straight line cutting across the document that has the lowest pixel intensity variance.

The rotation of the second image is carried out using fast bicubic interpolation with a Keys kernel, as we need to preserve the quality of the scanned image for, e.g., printing.

A blending between the documents is done in a conventional way by gradually mixing in one image into the other as we move more and more into the overlapping region.

Figure 3 shows some results of stitching flatbed-scanned images, including a US Legal-sized document, a brochure with mixed text and graphics and another one with a rotated second portion.

Table 3 gives the parameters uses by the algorithm for stitching oversized document of unknown length, where h and w are the pixel height and the width, respectively, of the input image and Y_{OFF} is an average offset arising from the scanning process (e.g. $3''$, see above). Note the large size of the search window that is necessary for stitching without manual initialization.

On a mid-range Pentium-based PC the computation time was on average about 1s for the registration plus a couple of seconds for generating the output, which, as we said, involves the expensive rotation of part of the high-resolution second image into the first image.

8 Conclusions

In this paper we have presented a novel practical planar registration method that is fast and extremely robust and therefore particularly suitable for applications such as automatic document stitching for flatbed scanners.

The method, which could in principle be extended to more complex transformations, is based on a two-stage robust estimation process where an initial, yet highly probable, hypothesis about the image transform is followed by a more classic robust estimation process based on Least Median of Squares. The hypothesis is made by independently seeking three matches that are consistent with a single Euclidean rigid transform that, as we showed in Section 5, turns out to be, with high likelihood, the correct one. This approach crucially minimizes the amount of expensive initial searches for matches, which directly lead to a computation time vastly inferior to alternative registration methods, without compromising robustness. We have shown simulations and actual results on real documents. Other results and short videos showing the working of the method are available on http://www.hpl.hp.co.uk/people/mp/research/stitching/

Acknowledgments. The authors wish to thank Stephen Pollard and Adele Lorusso for their help and Iris Fermin for her advice on Section 5.1.

References

[1] L.G. Brown. A survey of image registration techniques. *ACM Computing Surveys*, 24(4):325–376, December 1992.

[2] P. Meer, D. Mintz, A. Rosenfeld, and D. Y. Kim. Robust regression methods for computer vision: a review. *International Journal of Computer Vision*, 6(1):59–70, 1991.

[3] H. K. Nishihara. PRISM, a practical real-time imaging stereo matcher. Technical Report A.I. Memo 780, MIT, Cambridge, MA, 1984.

[4] M. Pilu, S. Pollard, and F. Isgro'. Method and apparatus for scanning oversized documents. *European Patent EP1091560.*

[5] L.S. Shapiro, H. Wang, and J.M. Brady. A matching and tracking strategy for independently moving objects. In *Proceedings of the British Machine Vision Conference*, pages 306–315. BMVA Press, 1992.

[6] S.M. Smith and J.M. Brady. Susan: A new approach to low-level image-processing. *International Journal of Computer Vision*, 23(1):45–78, May 1997.

[7] R. Szeliski. Image mosaicing for tele-reality applications. In *Proceedings of IEEE Workshop on Applications of Computer Vision*, pages 44–53, 1994.

[8] P. Viola and W.M. Wells. Alignment by maximization of mutual information. In *Fifth International Conference on Computer Vision*, pages 16–23, Cambridge, 1995.

[9] A. Whichello and H. Yan. Document image mosaicing. In *International Conference on Pattern Recognition*, page SAP1, 1998.

[10] A. Zappala, A. Gee, and M. Taylor. Document mosaicing. *Image and Vision Computing Journal*, 17(8):589–595, June 1999.

[11] I. Zoghlami, O. Faugeras, and R. Deriche. Using geometric corners to build a 2D mosaic from a set of images. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 420–425, June 1997.