Mosaicing for High Resolution Acquisition of Paintings

Massimiliano Corsini, Franco Bartolini, Vito Cappellini
Dipartimento di Elettronica e Telecomunicazioni, Università di Firenze
e-mail: barto@lci.det.unifi.it

Abstract

There is an every day increasing interest toward the applications of image processing techniques to the field of Cultural Heritage, the main objectives being high quality archival, image reproductions enhancement, virtual reconstruction, virtual restoration. This paper deals with the problem of setting up a simple and automatic procedure for obtaining a high resolution picture of a painting starting from a set of overlapped images showing different parts of the painting itself. With respect to other mosaicing techniques the proposed method allows to compose the whole picture even in the presence of strong colour distortions. This is possible thanks to the use of an ordinal measure as a similarity criterion for evaluating the correspondences between the images. Experimental results have demonstrated the good performances of the proposed method.

1. Introduction

In recent years many image processing and computer vision techniques have been used to develop various applications in the Cultural Heritage area. One application, for example, regards the assembly in a unique picture of images depicting parts of a painting and where the images can be of several types (e.g. infrared reflectograms, image acquired by X-rays, or in the visible band).

In many cases the size of the painting under examination is too large with respect to the acquisition device capacity, so it is not possible to obtain, in a single step, the desired spatial accuracy. In these cases it is necessary to repeat the acquisition process on several parts of the painting, and then to join the obtained sub-images in order to put together the data of the whole painting. For this purpose, several kinds of mosaicing techniques have been developed for a semiautomatic assembling of infrared reflectograms [1], for reconstruction of painting on curved surfaces [5], for colour images acquired under perspective distortion [2], for panoramic images [3], and so on.

The present work deals with colour images taken manually by a photographic machine and then scanned at high resolution. This acquisition process produces images characterized by strong colour distortions and small geometric distortions introduced when the operator translates the photo camera. In particular, during the camera translation, small rotations along all three-axis are inevitably introduced. The proposed method, in the hypothesis that translational motion of the camera is dominant with respect to tilt, slant and rotation, performs a fully automatic reconstruction of the acquired painting, based on a set of partially overlapped colour images.

Proceedings of the Seventh International Conference on Virtual Systems and Multimedia (VSMM'01)
0-7695-1402-2/01 $17.00 © 2001 IEEE

The proposed algorithm has two main features; the first is that it is fully automatic and requires a priori information only regarding the relative positions of the mosaic sub-images. The second feature regards the robustness with respect to colour changes between overlapped images: this latter feature is achieved thanks to the use of an ordinal measure [7] to estimate a set of image correspondences that are used to perform colour and geometric matching.

In the next section we give a general description of how our method joins a pair of sub-images (local mosaicing) and we analyze the most important steps of this process, in section 3 we extend the approach from local to global, to join all the sub-images together, and in section 4 we present some experimental results.

2. Local Mosaicing Algorithm

As briefly stated in the introduction the proposed method is based on a local mosaicing technique which joins a pair of images. Here this technique is described in detail. First of all, one of the two images is selected as a reference. Then, the non-reference sub-image is mapped on the reference one (from a geometric viewpoint) and its colours are modified in order to match the colours of the reference sub-image. We denote by I_R the reference sub-image, and by I_2 the sub-image that have to be mapped on I_R. The selection of I_R is done by a human operator, in this way the sub-image with the colours nearest to those of the painting can be selected, thus allowing to obtain the best final picture quality.

The first step of the local mosaicing technique consists of the estimation of the translational displacement components (t_x, t_y). In order to improve the performance of the algorithm, we estimate only an approximation of the exact value of (t_x, t_y). Nevertheless, this approximation is sufficient to extract the overlapped parts of the two sub-images (second step). In the following, we indicate the overlapped parts by I_R^{ov} and I_2^{ov}.

The third step is one of the most important and regards the evaluation of the pixel correspondences between I_R^{ov} and I_2^{ov}. The procedure for establishing correspondences will be detailed in subsection 2.2.

At this point we have obtained a set of correspondences (we denote it by C). So, we can use these data to compute a geometric transformation to map I_2 on I_R (4th step). Because the sub-images can be approximated as planar surfaces we can use homographies (mathematical relationships to model planar perspective transformations) for this mapping. The homography does not only eliminate the rotational components but also refines the estimation of (t_x, t_y). Subsection 2.3 treats the homography estimation.

Finally, colour adjustment is performed (as explained in subsection 2.4).

All the steps of the local mosaicing process are working on a gray-scale version of the two colour sub-images, but the last one which obviously uses the colour information.

2.1. Extraction of overlapping parts

In order to extract the overlapped parts of the sub-images we have to estimate the translational motion components (t_x, t_y) between I_R and I_2. To this aim we use the *phase correlation alignment algorithm* [6]. Let us suppose we have two images $I_1(x, y)$ and $I_2(x, y)$ related by

$$I_1(x, y) = I_2(x - t_x, y - t_y) \tag{1}$$

by computing the normalized correlation between the Fourier Transforms of I_1 and I_2 we get

$$C_\phi(u,v) = \frac{I_1(u,v)I_2^*(u,v)}{|I_1(u,v)I_2^*(u,v)|} = \exp(-j2\pi(ut_x + vt_y)) \quad (2)$$

where $I_1(u,v)$ and $I_2(u,v)$ are the DFT of $I_1(x,y)$ and $I_2(x,y)$ respectively. So, we can obtain (t_x, t_y) by :

$$(t_x, t_y) = \arg\max_{(x,y)} \mathcal{DFT}^{-1}\{C_\phi(u,v)\}(x,y) \quad (3)$$

where with $\mathcal{DFT}^{-1}$ we mean the inverse DFT.

In this way a first raw approximation of the translational displacement, which we assume dominant, between the two sub-images is obtained. To speed up this step we apply the phase correlation alignment algorithm on the 25% down-sampled versions of I_R and I_2 because, as we have just said, an approximation of the exact value of (t_x, t_y) is here sufficient to extract the overlapped parts of the two sub-images.

2.2. Establishing correspondences

To find correspondences between images with a common part is one of the main task in image processing and computer vision, and many different techniques have been developed during the years having this goal. It is possible classify these methods into two distinct categories: windows-based and feature-based. Here, a windows-based method is used. A first reason to use this kind of method instead of a feature-based one, resides in the fact that, sometimes, we have to mosaic images that do not present prominent features and, in these cases, a feature-based method could have some problems. In addition we will show that the windows-based method we used have some useful properties, making it attractive for our application. In general, the windows-based methods can be summarized as composed by the following steps: select a pixel on the reference image, extract a matrix of points around it (reference matrix), choose the correspondent pixel in the other image as that one for which the corresponding matrix of points have the maximum similarity with the reference matrix. The most used similarity criteria are based on correlation measures such as SSD (Sum of Squared Difference) and NCC (Normalize Correlation Coefficient), but for a fully automatic mosaicing algorithm having to deal with strong colour distortions (as it is our case) these measures are inappropriate because they are not enough robust. Better performances can be obtained with other similarity criteria like the ordinal measure that we used.

The ordinal measure is based on the relative order of the pixel values inside the matrix instead of on the matrix values themselves. An example of a 3×3 matrix and its corresponding relative order are shown below:

$$\begin{matrix} 120 & 118 & 100 \\ 145 & 144 & 99 \\ 143 & 146 & 87 \end{matrix} \quad -\rightarrow \quad \begin{matrix} 5 & 4 & 3 \\ 8 & 7 & 2 \\ 6 & 9 & 1 \end{matrix} \quad (4)$$

There are various kind of metrics to evaluate the difference between two sets of relative orders. The normalized distance $\kappa(W_1, W_2)$ (W_1 and W_2 are two generic windows) elaborated by Dinkar et al. [7] have a series of attractive properties like:

1. It is symmetrical, i.e. $\kappa(W_1, W_2) = \kappa(W_2, W_1)$. Hence, both W_1 and W_2 can be used as reference.

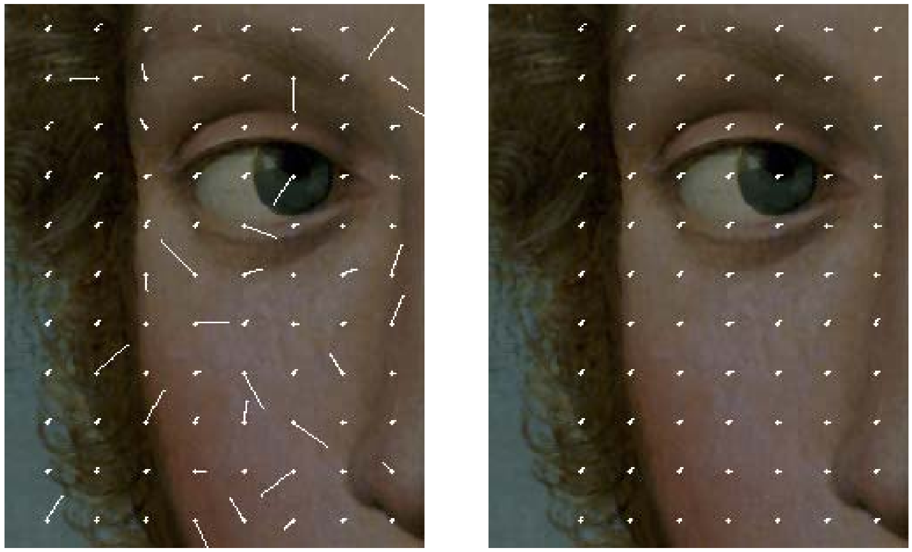

Figure 1. Correspondences found by the described search method (left). Regularized correspondences (right).

2. $\kappa(f(W_1), h(W_2)) = \kappa(W_1, W_2)$ where $f(.)$ and $h(.)$ are monotonically increasing functions. In particular this means that ordinal measure is invariant, for example, to transformations such as
$$I = gE^{\frac{1}{\gamma}} + m \tag{5}$$
where g is the camera gain, m is the reference bias factor, and γ accounts for the image contrast.

3. This measure captures the general relationship betw een data.

In particular, property 2 is very interesting because it allows us to deal with pair of images having strong colour distortions.

We are now going to detail the procedure to obtain the correspondences. First of all we define a grid of point on I_R^{ov} and I_2^{ov}. Then, for each point of the grid on I_R^{ov} (call the set of these grid points I_R^{grid}) w esearch its correspondence around its relative point position on the non-reference sub-image grid (I_2^{grid}).

After we hav e calculated the correspondences for every point of the grid we eliminate the outliers by performing a regularization of the obtained correspondence vector field. Taking into account that images we are working on can be approximated as rigid planar surfaces, w ecan perform the regularization b y applying a component-wise median filter. Figure 1 sho ws the result of the regularization process.

2.3. Homography estimation

The relationship betw een the coordinates of tw oimages related by a planar perspective transformation is:

$$x' = \frac{m_{11}x + m_{12}y + m_{13}}{m_{31}x + m_{32}y + 1}$$
$$y' = \frac{m_{21}x + m_{22}y + m_{23}}{m_{31}x + m_{32}y + 1} \tag{6}$$

where (x, y) represents the coordinates of a point on an image and (x', y') are the coordinates of the same point mapped through the homography on the other image. Rewriting (6) in homogeneous coordinates we get:

$$\begin{pmatrix} x' \\ y' \\ \lambda \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = M \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \tag{7}$$

where M (the homography) is a 3 by 3 matrix with 8 unknown parameters which we want to estimate.

Each correspondence gives a relation like (7). So, we can build a linear system considering N correspondences, and rewrite it by highlighting the unknown parameters:

$$\begin{bmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 & -x_1 x'_1 & -y_1 x'_1 \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -x_1 y'_1 & -y_1 y'_1 \\ x_2 & y_2 & 1 & 0 & 0 & 0 & -x_2 x'_2 & -y_2 x'_2 \\ 0 & 0 & 0 & x_2 & y_2 & 1 & -x_2 y'_2 & -y_2 y'_2 \\ & \vdots & & & \vdots & & \vdots & \vdots \\ x_N & y_N & 1 & 0 & 0 & 0 & -x_N x'_N & -y_N x'_N \end{bmatrix} \begin{bmatrix} m_{11} \\ m_{12} \\ m_{13} \\ m_{21} \\ m_{22} \\ m_{23} \\ m_{31} \\ m_{32} \end{bmatrix} = \begin{bmatrix} x'_1 \\ y'_1 \\ x'_2 \\ y'_2 \\ \vdots \\ x'_N \\ y'_N \end{bmatrix} \tag{8}$$

where (x_i, y_i) are the coordinates of several points on I_R^{grid} and (x'_i, y'_i) are the coordinates of the corresponding points on I_2^{ov}. The solution of this system provides an estimate of M.

To build the system (8) we have to choose a subset of N correspondences from the set C, call this subset C_N. Obviously, the solution of system (8) gives us an estimate $M(C_N)$. The goal is to find the $M(C_N)$ that best fit all the correspondences from an Euclidean distance point of view. To achieve this goal we iterate a procedure that we have called *pseudo-RANSAC scheme*. This name is derived from the fact that we are not using a truly RANSAC scheme but a similar algorithm with some of the RANSAC's basic characteristics. Let us now to describe this method. According to the RANSAC philosophy only the minimum number of data required for solving (8) are used, so we work on a subset of C_N with $N = 4$. In the first step the subset C_N is constructed by a random selection of correspondences from C. Then, we use this subset to compute $M(C_N)$. The next step consists of the evaluation of $M(C_N)^*$, i.e. the consensus set of $M(C_N)$. A correspondence lies in $M(C_N)^*$ if and only if it respects the condition:

$$\sqrt{(x'_p - x_2)^2 + (y'_p - y_2)^2} < \Theta \tag{9}$$

where Θ is a tolerance error, (x'_p, y'_p) is the mapping through $M(C_N)$ on I_2^{ov} of a generic grid point $(x_p, y_p) \in I_R^{grid}$ and (x_2, y_2) is the corresponding correspondence. Instead of using a stop condition (like in a standard RANSAC scheme) we iterate this processing a predefined number of times. At the end we take as a final estimate of M the $M(C_N)$ with the biggest consensus set (this is another difference with respect to the classical RANSAC scheme, where the final estimate of the parameters is performed over all data belonging to the biggest consensus set).

2.4. Colour adjustment

The main idea is to use a parametric transformation to model colour changes of correspondent pixels, estimate the parameters of the transformation and then use it to adjust all colours of the non reference sub-image. We have chosen a linear affine transformation:

$$\begin{pmatrix} R \\ G \\ B \end{pmatrix} = A \begin{pmatrix} R' \\ G' \\ B' \end{pmatrix} + B \tag{10}$$

where $A \in \mathbb{R}^{3\times 3}$, $B \in \mathbb{R}^{3\times 1}$, (R, G, B) represents the tristimulus value of one pixel on the reference sub-image and (R', G', B') represents the correspondent value on the other sub-image. T o simplify the notation we rewrite (10) in a more compact form:

$$\begin{pmatrix} R \\ G \\ B \end{pmatrix} = J \begin{pmatrix} R' \\ G' \\ B' \\ 1 \end{pmatrix} \tag{11}$$

where $J = [A\ B]$ is a 3×4 matrix of real values.

Our goal is to select those values of J that allow to obtain a good colour matching, so we have to estimate 12 parameters. The estimation of J is performed by an algorithm similar to the one used for the estimation of the homography. The main difference resides in the evaluation step; in this case we do not perform the computation of a consensus set but the capability of $J(C_N)$ to match colours is evaluated instead. In particular the quality test of $J(C_N)$ is based on the index:

$$\Delta E^*_{Lab}(P_1, P_2) = \sqrt{(L^*_1 - L^*_2)^2 + (a^*_1 - a^*_2)^2 + (b^*_1 - b^*_2)^2} \tag{12}$$

where (L^*_1, a^*_1, b^*_1) are the colour components of P_1 (a generic point on a image) in the standard, perceptually uniform, colour system CIELAB [8], and (L^*_2, a^*_2, b^*_2) are the corresponding colour components of P_2. The use of CIELAB colour space allows ΔE^*_{Lab} to give a measure of the colour difference between P_1 and P_2 from a perceptual point of view. F or example, a value of ΔE^*_{Lab} less than 3 denote a colour differences almost not perceptible by the eye. In order to give a numeric value to compare the quality of $J(C_N)$, the relative ΔE^*_{Lab} is computed for each correspondence and all values are summed up:

$$Q_{test} = \sum_C \Delta E^*_{Lab}(P(I^{ov}_R), P(I^{ov}_2, J(C_N))) \tag{13}$$

where $P(I^{ov}_2, J(C_N))$ represents the (L^*, a^*, b^*) components of the corresponding (R, G, B) values re-mapped through $J(C_N)$. We use Q_{test} as a quality measure for $J(C_N)$. As the final estimate of J we take the $J(C_N)$ with the minimum value of Q_{test}.

This algorithm provides a good colour adjustment for images with complex, even non-linear, colour distortions (figure 2 shows an example).

3. From Local to Global Mosaicing

Up to now the problem of mosaicing two sub-images has been solved, but when several sub-images have to be joined together to form the final composition, we have to consider

Figure 2. Two images before (right) and after (left) colour adjstment.

the issue related to extending the approach from a local to a global basis. This task is very difficult and usually expensive from a computational point of view [4]. The main problem regards the errors accumulation that appears in a concatenation of homographies. In particular, the accumulation of errors becomes evident when the sub-images are connected by a circular concatenation of homographies (as shown on left image in figure 3).

To avoid this problem we use an *incremental approach* for the global registration; starting from the sub-image selected as reference, one of its neighboring sub-images is registered and joined to it. Then, we repeat this step by considering the current mosaic as a new reference sub-image.

To work properly this approach requires some preliminary step. In particular we pre-calculate the relative (t_x, t_y) for all the pairs of sub-images; this means that we can obtain the global translation between all sub-images. Then, with this data, we can use the algorithm described in Section 2 (without the first step) to perform the registration of each sub-image in the incremental approach.

In order to automatically decide which sub-image has to be joined at each step, we have implemented a strategy that try several options during the mosaic construction to achieve the best visual result. For example, if at a certain moment, 3 sub-images can be added to the current mosaic, all of this 3 sub-images are registered and only the sub-image that gives the highest image quality is used. The best visual result is evaluated by the comparison of the ratio:

$$Q_{image} = \frac{\#M(C_N)^*}{\#C} \tag{14}$$

where the symbol # denote the cardinality operator. We consider values of Q_{image} nearest 1.0 as a perfect registration (from a geometric point of view) and an uncertain registration for small value (< 0.7). Although the Q_{image} index does not have a close relation to image registration (i.e, a good registration could have a small value of Q_{image}) it has shown to be a useful criterion to choose which sub-image can be joined to the current mosaic.

This incremental approach reduced the effects of the errors accumulation, but it works well only when few sub-images (about 20-30) have to be joined. In figure 3 (Right) an example of the obtained results is shown.

Figure 3. (Left) Circular concatenation of homographies. We can see at the top-left corner the inaccurate registration caused by the accumulation of errors. (Right) Global alignment with incremental approach.

4. Experimental Results

In this section we present the results obtained on a set of real sub-images taken from the Raffaello's painting *"Madonna con Liocorno"*. As we have just said the algorithm works in a fully automatic way; the user have only to select the starting sub-image.

In figure 4 the mosaic construction is shown; starting from a reference sub-image, the global mosaicing algorithm, at each step, adds a new sub-image to the current mosaic. The final result of this process is depicted in figure 5. As we can see the colour adjustment is good but not perfect. This problem is caused by two reasons. The first is that the relationship involved in the colour distortion is actually nonlinear and, the second is that sometimes the overlapping parts do not contains enough colours for correctly estimating J. A possible improvement to colour adjustment could be obtained by superimposing on each sub-image a reference colour table chart (i.e. AGFA IT8) during the acquisition process, and by calibrating the sub-image colours with respect to this reference. Despite this problem, the quality of the final image is similar to the one obtained by a human operator who composes the various sub-images manually with a photo-manipulation program.

In order to demonstrate that the algorithm can work well with images with strong colour distortion and to show the original colours of the sub-images before the correction process, we have repeated the procedure on the same set excluding the colour adjustment step. The obtained image is shown in figure 6.

5. Conclusions

In this paper an algorithm to mosaic pieces of a painting in order to obtain the whole picture without geometric and colour distortions has been presented. The most valuable characteristic of the developed algorithm is just the capability of working in presence of strong colour distortions among the sub-images that have to be joined through mosaicing.

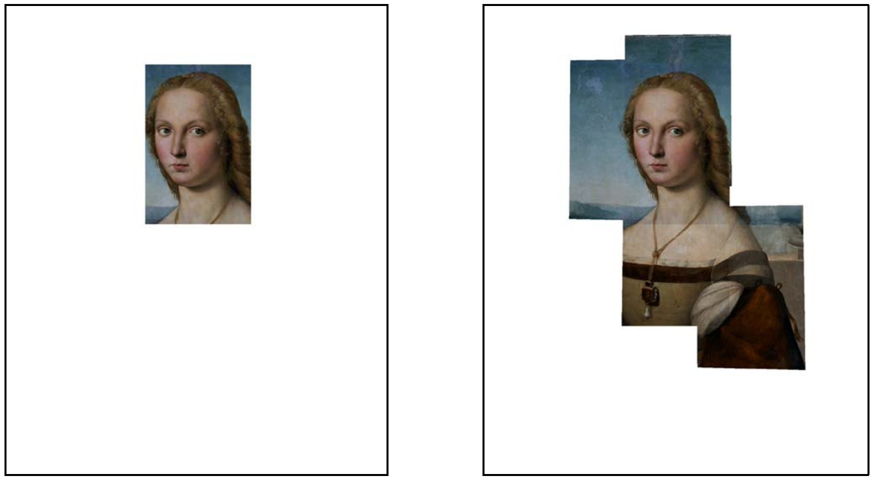

Figure 4. Mosaic under construction: on the left we show the starting subimage; on the right we show the mosaic produced by the algorithm after some step. Remind that at each step we join a new subimages to the current mosaic.

Figure 5. Final result of the mosaicing algorithm.

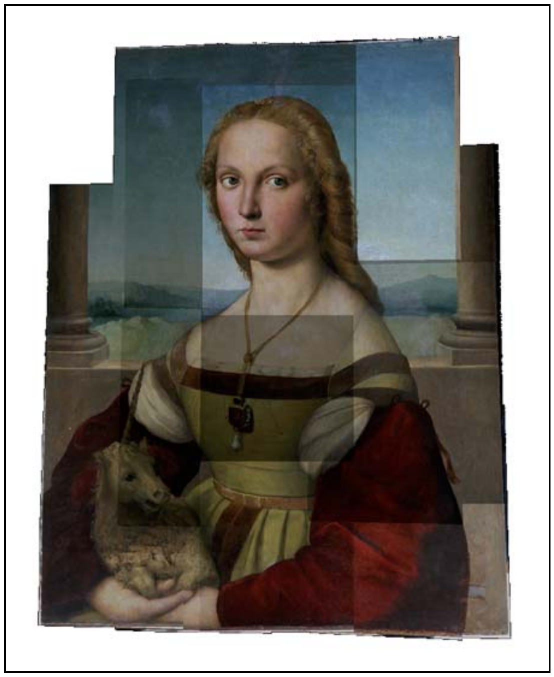

Figure 6. Final result of the mosaicing algorithm without colour adjustment.

6. Acknowledgment

The authors would like to thank Ing. Seracini of the Editech S.r.l. for having stimulated this work and provided the images.

References

[1] R. Billinge, J. Cupitt, N. Dessipiris, and D. Saunders, "A Note on an Improved Procedure for the Rapid Assembly of Infrared Reflectogram Mosaics.", *Studies in Conservation*, vol. 38, pp. 92-97, 1993.

[2] G. M. Cortelazzo, L. Lucchese , "A Technique for Matching Projective Views with Application to Color Image Mosaicking", *Proc. of ICIP '99* , Kōbe , Japan , 24-28 ottobre , volume IV , pp. 108–112.

[3] Heung-Yeung and Richard Szeliski, "Panoramic Image Mosaics", *Microsoft Research T echnial Report MSR-TR-97-23*, Redmond WA 98052, http://www.research.microsoft.com/.

[4] D. Capel and A. Zisserman "Automated Mosaicing with Super–Resolution Zoom", *Pr oc. of IEEE Int'l Conf. on Computer Vision and Pattern Recognition*, San ta Barbara, CA, 1998, pp. 442–447.

[5] W. Puech, A. G. Borş, and I. Pitas, "Mosaicking of Pain tings on Curved Surfaces", *Pr oc. of the* 3^{rd} *Workshop on applications of Computer Vision* , Sarasota, FL , 2-4 Dec. 1996 , pp. 44–49.

[6] C. D. Kuglin and D. C. Hines, "The Phase-Correlation Image Alignment method" , *Pr oc. IEEE 1975 Int. Conf. Cybern. So c.*, 1975, pp. 163–165.

[7] Dinkar N. Bhat and Shree K. Nayar, "Ordinal Measure for Visual Correspondence", *Columbia University Computer Science Technical Report, TR-CUCS-009-96*, 1996.

[8] Charles A. Poyn ton, "F requen tly Asked Question About Color", *A vailable on Internet*, 1997-03-02.

[9] Martin A. Fischler and Robert C. Bolles , "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography", *Communications of the ACM*, vol. 24:381-95, 1981.