

A Fast Matching Method for Color Uncalibrated Images using Differential Invariants

V. Gouet[†], P. Montesinos[†], D. Pel[‡]

[†]EMA/LGI2P - Site EERIE

Parc scientifique Georges Besse

30035 Nmes Cedex 1 FRANCE

[gouet|montesin]@eerie.fr

[‡]France Tlcom CCETT/CNET/DIH

4 rue du Clos Courtel, BP 59

35512 Cesson Svign Cedex

Rennes FRANCE

pele@ccett.fr

Abstract

In this paper we present a new method for point matching in stereoscopic color images. Our approach consists first in characterizing points of interest using differential invariants. Then we define additional first order invariants using color information, which make sufficient the characterization till first order. In addition, we make our description robust to important image transformations like rotation, range of viewpoint and linear illumination variations. Second, we propose a new incremental technique for point matching using our characterization, which works robustly and rapidly whatever the number of points to be matched. Our stereo matching scheme is evaluated using stereo color images, with viewpoint and illumination variations. The very good results obtained clearly show the pertinence of our approach. Our color characterization produces a high rate of good matches, even though only first order derivatives are used. Results on images holding many points show that our matching process is robust and rapidly implemented even if the points to be matched are numerous. It is a great asset, when matching a high set of points is necessary for example to realize dense depth maps between images.¹

Key word: Color Images, Differential Invariants, Stereo Matching, Transfer Methods.

¹This work was supported by a CCETT grant 96-ME-24.

1 Introduction

This work addresses the problem of matching stereoscopic, uncalibrated color images, with the aim of doing transfer, i.e. computing synthetic images having intermediate points of view. In order to match point primitives, we revisit point characterization using differential invariants belonging to iconic methods in gray level images. Until now, these invariants were defined only for gray level images and used to be computed till order 3 to characterize efficiently points. We show that color information can robustly improve characterization, not only by multiplying the gray level features but also by providing new invariants, and this only using first order derivatives. In addition, we describe a method to make this characterization invariant to some usual transformations of the image. Then we implement a new fast method of matching using this characterization. We show that this approach is robustly and rapidly achieved, even if the points to be matched are numerous.

Our matching scheme is decomposed into three stages: In section 2, we define a new differential characterization of points of interest for color images, which is made invariant to image orthogonal transformations and to linear changes of intensities. In section 3, we define a new points matching process, working with our color invariants. To be effective, transfer methods impose matching a big set of points between the images. So we present in section 4 an incremental technique for matching rapidly and efficiently high sets of points. This technique integrates the simple matching process described in section 3, and enriches it with geometric constraints. In section 5, matching results firstly demonstrate the validity of our point characterization and our matching technique, on images differing from some usual image transformations. Second, results on big sets of points show the pertinence of our incremental matching technique.

2 Characterization with differential invariants

Some techniques of point matching called iconic methods use signal information directly to characterize the points to be matched. One of the most popular techniques is the correlation method [10] [13]. Other more recent methods, working on gray value images, consist in characterizing points by using differential invariants of the signal.

Among iconic methods, the correlation methods produce very good results but are very time consuming and very restrictive because of the small transformations they allow between images. On the contrary, methods based on differential invariants are more robust against rotations between images and are also faster to implement. But in gray level images, it is often necessary to consider invariants up to the third order [11] to obtain a good characterization of the primitives, and these are difficult to estimate in a stable way.

Our work consists in using differential invariants for point characterization. We revisit them in section 2.1 and we introduce new invariants of first order specific to color. In section 2.2, we present our characterization which overcomes the main drawbacks of the classical invariants, by considering only first order derivatives and color information. This new description is robust and invariant to orthogonal

transformations of image. We also make it invariant to linear illumination variations in section 2.3.

2.1 Invariant Image Attributes

As it has been shown by Hilbert [7], any invariant of finite order can be expressed as a polynomial function of a set of irreducible invariants. This set is well known for first and second order properties [4] [11] and is better expressed in a system of coordinates, no longer linked to the rotation, as the well known *Gauge coordinate* L , L_η , $L_{\eta\eta}$, $L_{\xi\eta}$, $L_{\xi\xi}$, where η is the unit vector given by $\eta = \frac{\nabla I}{|\nabla I|}$ and $\xi \perp \eta$. The 5 following invariants, which combine these attributes, perform quite well for matching two gray level images:

$$I \quad I_{\eta\eta} + I_{\xi\xi} \quad I_\eta \quad \frac{I_{\xi\xi}}{I_\eta} \quad \frac{I_{\xi\eta}}{I_\eta}$$

Considering a set of color image $\{R, G, B\}$ and the group of rotation (specified by just one parameter i.e the rotation angle), the set of invariants for first and second order presents 6 degrees of freedom by color channel, and so will include $6 \times 3 - 1 = 17$ invariants. These may include the fifth invariants for each channel, and two additional invariants that may be chosen from the following set:

$$\nabla R \cdot \nabla G \quad \nabla R \cdot \nabla B \quad \nabla G \cdot \nabla B$$

2.2 Characterization using first order differential invariants and color information

Our idea is to use Hilbert's invariants which involve only derivatives till first order. The characterization obtained with gray level images would be too poor, but we show here that it is widely compensated by color information. We obtain a vector of invariants against translation and rotation containing $2 \times 3 + 2 = 8$ components. We call it $\vec{v}_{col}$:

$$\vec{v}_{col}(\vec{x}, \sigma) = (R, |\nabla R|^2, G, |\nabla G|^2, B, |\nabla B|^2, \nabla R \cdot \nabla G, \nabla R \cdot \nabla B)^T \quad (1)$$

Using only first order invariants presents two main advantages: first it allows very robust characterization with regard to noise and second, the complexity of the method is small since only 14 images are computed: $R, R_x, R_y, |\nabla R|, G, G_x, G_y, |\nabla G|, B, B_x, B_y, |\nabla B|, \nabla R \cdot \nabla G, \nabla R \cdot \nabla B$. The reader can find further details and comparative results about this description in [9].

2.3 Color constancy

In order to obtain a robust characterization of our color image primitives, we want to make it invariant to changes of intensities. It requires first estimating a very precise model for these changes. According to the recent work of Finlayson on color constancy [3], we use its diagonal model ($\mathcal{D}_{(3 \times 3)}$) with an additional vector

of translation ($\mathcal{T}_{(3)}$) to get linear invariance to intensities. We obtain a model with six degrees of freedom :

$$\vec{x}' = \mathcal{D}\vec{x} + \mathcal{T} \quad (2)$$

where $\vec{x} = (r, g, b)^T$ is the pixel color and $\vec{x}' = (r', g', v')^T$ its linear transformation. Other more complex models exist but this one provides the best quality/complexity ratio. Then, to make the characterization invariant, we normalize the images in order to make them independent of the model, before computing $\vec{v}_{col}$. Our method is performed on each color channel {R,G,B}: we resize the gray value of each pixel in a given interval. This process makes the image independent of the parameters $\mathcal{D}$ and $\mathcal{T}$ of the model. We preserve the local properties of the pixels by implementing it locally in a window centered on the pixel to be normalized. For a complete description and comparative results about the image normalization described in this section, see [5].

3 Our matching method

When point characterization is achieved, matching method consists in comparing each feature vector of the first image with the ones of the second image, in order to find the points which looks the most similar. The most complete comparison method seems to be the Mahalanobis distance [11]. But this method involves the covariance matrix of the vector components, which is very complex and difficult to implement with exactness. So we present a more simple solution to compute the likeness of two vectors.

3.1 A method for comparing feature vectors

The components of $\vec{v}_{col}$ present their own range of values. So we resize each of them in a given interval, before using the Euclidean distance to make the comparison. This method is sub-optimal according to the Mahalanobis distance's one, but it looks sufficient, insofar as the percentage of correct matches obtained is very high.

3.2 Matching

Using these distances, the most natural process to realize the matching is to select the pairs (m_{1i}, m_{2j}) whose distance is lower than the one of all the possible pairs $(m_{1i}, m_{2j'})$ and $(m_{1i'}, m_{2j})$. But this technique may eliminate a lot of matches which could have been good matches. So we keep only matches associated to small distances and we eliminate the possible remaining ambiguities by using a relaxation technique which works with semi-local constraints [13] [11].

The reader can see in [9] matching results using the characterization developed in section 2.2 and the matching technique described in this section: the rate of correct matches overshoots 90% the most of the time, though matched images are very different (important rotations, different points of view, linear illumination variations). In addition, the implementation is much faster than the ones based on correlation. In next section, we show that the matching method can be notably improved, in particular for big sets of points matching.

4 An incremental algorithm for constrained matching

The main drawbacks of the comparison method described in section 3 is first its complexity. If we do not have any information about the disparity between the two images (for example when the cameras are not calibrated), the complexity of the comparison method is equal to $O(m \times n)$ for m given points in the first image and n in the second one. Consequently, the second inconvenient is the rate of ambiguities which arises with the number of points to be matched. It makes the relaxation process more time consuming and finally may generate more bad matches. To sum up, the matching process is efficient until 200 or 300 points but becomes unworkable with more points. It is for example necessary to match many points in order to realize dense depth maps between two images.

So our idea is to make the search area of the corresponding point in the other image as smallest as possible. If the disparity map between the two images is unknown, we must localize this area ourselves. Let us suppose that we have at our disposal a set $\mathcal{M}$ of good matches between the images. We show in the next section how these data can give us interesting geometric informations about the search area.

4.1 The available geometric information

We consider in this paper the model of the cameras as a Stenope model, which modelizes the projection as a pure perspective projection.

4.1.1 Using the epipolar geometry

If $\mathcal{M}$ contains at least 8 matches, the epipolar geometry of the two cameras system can be estimated. It is characterized by a fundamental matrix $F_{(3 \times 3)}$ which satisfies $m_2^T F m_1 = 0$ for two matched points m_1 and m_2 . This equation expresses that the corresponding point m_2 in the second image of the point m_1 of the first image lies on the line $F m_1$, and that the corresponding point of m_2 in the first image lies on $F^T m_2$. It represents all the information that can be estimated from two images when the cameras are not calibrated. When F is estimated, it is very easy to see that the complexity of the matching process is very reduced seeing that the search area becomes a line. The fundamental matrix F computed from the matches $\mathcal{M}$ will be called $F_{\mathcal{M}}$ in the next sections.

4.1.2 Using a Delaunay triangulation

Let us consider a 3D point P belonging to a triangle T , (p_1, p_2) its projections on the two images and (t_1, t_2) the projections of the triangle. It can be easily demonstrated that a triangle is transformed in a triangle by projective transformation, so t_1 and t_2 are triangles too. The point p_1 is necessarily located in t_1 and has necessarily its corresponding point p_2 in t_2 . If P does not belong to T , the position of p_2 related to t_2 is function of the disparity. In practice, we show that it is enough to consider only the closest neighboring triangles of t_2 . So we are able to define an area in which the corresponding point of a given point can

be found. We compute a triangulation on the matched points of the first image and we estimate the triangulation of the second image from it : a triangle of the second image has its 3 vertices matched to points of the same triangle in the first image. We have chosen for the first image triangulation a Delaunay triangulation because it produces triangles the most equiangular possible [2]. The triangulation of the second image obtained from it is not necessarily a Delaunay triangulation. We call in the next sections $\mathcal{T}_{\mathcal{M}}$ the two triangulations based on the matches $\mathcal{M}$.

By combining these two constraints, the search of the corresponding point is reduced to a segment, if the studied point belongs to a triangle of the triangulation, as shown in figure 1. The search is reduced to some segments or to the entire epipolar line if the studied point does not belong to any triangle.

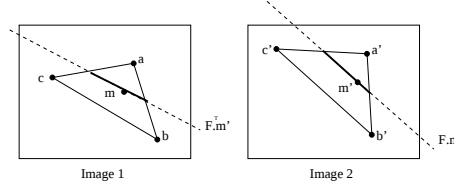


Figure 1: (a,a') , (b,b') and (c,c') are correct matches. A point m of Image1 located in the triangle (a,b,c) has its corresponding point in Image2 on the epipolar line $F.m$ and in the triangle (a',b',c') . The point m' has its corresponding point in Image1 on $F^T.m'$ and in (a,b,c) .

In this section we have introduced geometric constraints which allow us to reduce the search area of the point to be matched, even if we do not have any information about the disparity between the images and whatever the images are. These constraints are going to be integrated in the incremental algorithm described in the next section.

4.2 Our incremental algorithm

The constrained matching method presented above supposes that we have a set of matched points at our disposal to initiate the matching using the geometric constraints. So we can define an incremental algorithm of matching which computes at iteration i the set of matches $\mathcal{M}^i$ from the geometric constraints associated to the matches $\mathcal{M}^{i-1}$ of the iteration $i - 1$. This method proceeds in 6 steps :

1. Extraction (adding) of points of interest in the images : sets $\mathcal{P}_1^i$ and $\mathcal{P}_2^i$.
2. Characterization of each point using the feature vector $\vec{v}_{col}$ (see section 2.2 for the characterization method).
3. For each point $p_{1k} \in \mathcal{P}_1^i$ and $p_{2l} \in \mathcal{P}_2^i$, estimation of the search area in the other image, using $F_{\mathcal{M}_{i-1}}$ and $\mathcal{T}_{\mathcal{M}_{i-1}}$ if exist : $\mathcal{A}_{1k}$ and $\mathcal{A}_{2l}$.
4. Comparison of the feature vectors of p_{1k} and p_{2l} which satisfy : $p_{1k} \in \mathcal{A}_{2l}$ and $p_{2l} \in \mathcal{A}_{1k}$. A set of matches $\mathcal{M}_*^i$ with possible reduced ambiguities is obtained. The method of vector comparison has been described at section 3.1.

5. Relaxation method on $\mathcal{M}_*^i$ (see section 3.2): a set of matches $\mathcal{M}^i$ without ambiguities is obtained.
6. Computation of the geometric constraints: the triangulation $\mathcal{T}_{\mathcal{M}^i}$ and the fundamental matrix $F_{\mathcal{M}^i}$ associated to the matches $\mathcal{M}^i$.
7. Back to 1 while not enough matches.

Seeing that the matches $\mathcal{M}^i$ are computed using the ones of the iteration $i - 1$, it is very important to obtain a very high rate of good matches at each iteration. This condition is most of the time satisfied since the geometric constraints allow to eliminate efficiently many ambiguous solutions. It is interesting to notice that the average width of the segments decreases as the number of iterations increases; so the matching process is done faster and faster with less and less ambiguities. The main difficulty remains to obtain a very good estimation of the first set of matches $\mathcal{M}^0$ for which geometric constraints are not available. Our solution is to keep at the end of the relaxation process only a few percentage of the matches obtained (in practice approximatively 40% of the matches): the ones which have got the best matching scores. The experience shows that they represent a very good basis to compute matches of iteration 1.

5 Matching results

The points of interest to be matched $\mathcal{P}_1^i$ and $\mathcal{P}_2^i$ are computed using a generalization of the Harris detector to color data [1] [6] [9], which needs only first order derivatives. We characterize them using the feature vector $\vec{v}_{col}$ of equation 1.



Figure 2: *Two images Room differing from viewpoint and illumination: 155 matches were found using the basic matching technique. The epipolar line in the second image corresponds to the point 54 of the first image. The line of the first image corresponds to the point 112 of the second image.*

The images are locally normalized (see section 2.3). So the characterization is made invariant to rotation and change of intensities. The matching is realized using

the incremental method described in section 4. The fundamental matrices $F_{\mathcal{M}^i}$ can be estimated by using a Ransac algorithm [12]. But we estimate them using a robust linear method with the help of the Least Median of Squares regression (LMedS) [13] [8]. It produces a fast and reliable estimation of the $F_{\mathcal{M}^i}$, since the important thresholding of the first iteration and the geometric constraints applied in the other ones produce few outliers. The triangulation of the first image used to compute $\mathcal{T}_{\mathcal{M}^i}$ is a semi-dynamical Delaunay triangulation [2]: it is implemented in an incremental way in order to be the most efficient within our incremental matching process ; at iteration i , the triangulation is computed by inserting the points $\mathcal{P}_1^i$, belonging to $\mathcal{M}^i$, in the triangulation $\mathcal{T}_{\mathcal{M}^{i-1}}$ computed at the previous iteration.

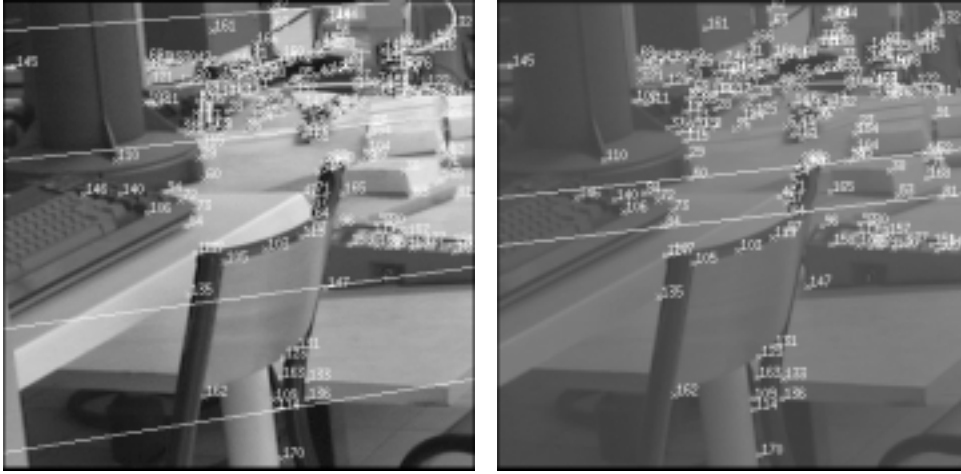


Figure 3: *The two same images Room : 170 correct matches were found using the incremental matching algorithm and the geometric constraints. The epipolar lines in the first image correspond to points $\{82,23,147,114\}$ in the second image ; the ones of the second image correspond to points $\{60,81\}$ in the first image.*

Figures 2 and 3 show the results of the matching process on (220×231) points of interest. Figure 2 brings to the fore matches (called $\mathcal{M}$) computed only with the basic matching method described in section 3 whereas figure 3 shows matches (called $\mathcal{M}_{inc}$) using the incremental method. These two results are very easy to compare:

- Time consuming: half an hour was necessary to obtain the set $\mathcal{M}$. Only 3 iterations and a few minutes were necessary to obtain $\mathcal{M}_{inc}$. It is due to the fact that the complexity in the feature vector comparison algorithm is smallest, since the search areas are limited. The other reason is that the rate of ambiguous matches is much less important at the beginning of the relaxation algorithm, which is consequently realized faster.
- Results: we obtain better results with the incremental method (170 matches all correct) than with the basic one (155 matches with few bad matches).

Some correct epipolar lines estimated using $\mathcal{M}_{inc}$ have been superposed on the images: using the epipolar geometry as a matching constraint (figure 3), we succeed in matching correctly the match 82 for example, whereas the same points in figure 2 have been incorrectly matched (matches 112 and 115). This comment shows the pertinence of our incremental approach. The reader can notice that the epipolar geometry obtained is very precise.

Figure 4 shows extracts of images for which more than 600 points have been matched in approximatively a quarter of hour, using our incremental matching method. The rate of correct matches is above 95% and the final epipolar geometry is very precise. For comparison, the basic algorithm has taken roughly 5 hours to obtain a solution having a higher rate of false matches ! This result shows clearly that our approach is the most adapted when there are many points to match.

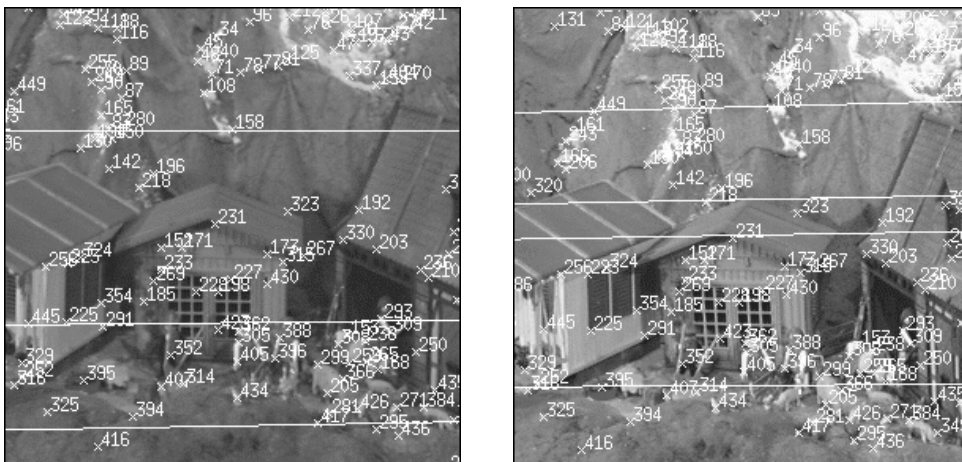


Figure 4: *Extracts of two images Toys differing from viewpoint and illumination. The epipolar lines in the first image correspond to points $\{158, 293, 417\}$ in the second image ; the ones of the second image correspond to points $\{449, 218, 231, 395\}$ in the first image.*

At the end of the matching process, we have obtained a set of matched points $\mathcal{M}$. In addition we have got the epipolar geometry of the cameras system characterized by $F_{\mathcal{M}}$ and a triangulation $\mathcal{T}_{\mathcal{M}}$ of the matched points. These geometric constraints have been very useful to realize an efficient matching and it is interesting to notice that their utility is not achieved insofar as they are necessary to implement methods of dense matching, images transfer or 3D scene reconstruction.

6 Conclusion

In this paper, we have defined a new method for matching stereoscopic, uncalibrated color images. We have first presented new differential invariants specific to color information. We have shown that adding them to Hilbert's invariants computed for each color channel gives a sufficient information to consider only deriva-

tives till first order. In addition, this description can be easily made invariant to the orthogonal images transformations and affine transformations of intensities. Second, we have implemented an incremental technique of point matching, based on this differential characterization, which works robustly and rapidly. Indeed, results of the last section show the pertinence of our approach; the percentage of correct matches is very high (above 95%) and the recovered epipolar geometry very accurate, in spite of the differences between the images (different points of view and illumination variations) and the high number of points to be matched. Using this matching scheme, we are able to match rapidly a high number of points, and so now we are able to implement efficiently methods of dense depth maps computation. Besides, we are already working on matching images involving changes of scale, as when the camera comes near the scene from one image to another. A method derived from the one we have described is producing promising results.

References

- [1] C. Bauckhage and C. Schmid. Evaluation of keypoint detectors. Technical report, INRIA, 1996.
- [2] J.D. Boissonnat and M. Teillaud. A hierarchical representation of objects: the delaunay tree. *Second ACM Symposium on Computational Geometry in YorkTown Heights*, 1986.
- [3] G.D. Finlayson, M.S. Drew and B. Funt. Color constancy: Generalized diagonal transforms suffice. *Journal of the Optical Society of America A*, 11(11):3011–3019, november 1994.
- [4] L.M.J. Florack, B.M ter Haar Romeny, J.J. Koenderink and M.A. Viergever. General intensity transformations and differential invariants. *Journal of Mathematical Imaging and Vision*, 4:171–187, 1994.
- [5] V. Gouet, P. Montesinos and D. Pelé. Stereo matching of color images using differential invariants. In *Proceedings of the International Conference on Image Processing, Chicago*, 1998.
- [6] C. Harris and M. Stephens. A combined corner and edge detector. *Proceedings of the 4th Alvey Vision Conference*, pages 147–151, 1988.
- [7] D. Hilbert. *Theory of Algebraic Invariants*. Cambridge Mathematica Library, Cambridge University Press, 1890.
- [8] S. Laveau. *Géométrie d'un système de N caméras. Théorie, estimation et applications*. PhD thesis, Ecole Polytechnique, France, 1996.
- [9] P. Montesinos, V. Gouet and R. Deriche. Differential Invariants for Color Images. *Proceedings of 14th International Conference on Pattern Recognition*, 1998.
- [10] K. Nishihara. Prism: a practical real-time imaging stereo matcher. *Proceedings of 3st International Conference on Robot Vision and Sensory Controls*, 1993.
- [11] C. Schmid. *Appariement d'images par invariants locaux de niveaux de gris - Application à l'indexation d'une base d'objets*. PhD thesis, INPG, France, 1996.
- [12] P.H.S. Torr and D.W. Murray. The development and Comparison of Robust Methods for Estimating the Fundamental Matrix. *International Journal of Computer Vision*, 24(3):271–300, 1997.
- [13] Z. Zhang, R. Deriche, O. Faugeras and Q.T. Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. Technical Report RR-2273, INRIA Sophia-Antipolis, France, 1994.