Still life stereo

Tony Kasvand

▶ **To cite this version:**

Tony Kasvand. Still life stereo. [Research Report] RR-1372, INRIA. 1991. inria-00075189

HAL Id: inria-00075189
https://inria.hal.science/inria-00075189

Submitted on 24 May 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNITÉ DE RECHERCHE
INRIA-ROCQUENCOURT

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
B.P.105
78153 Le Chesnay Cedex
France
Tél.:(1) 39 63 55 11

Rapports de Recherche

N° 1372

Programme 4
Robotique, Image et Vision

STILL LIFE STEREO

Tony KASVAND

Janvier 1991

RR-1372

STILL-LIFE STEREO

A summary of studies carried out during Sept. 1989 to July 1990 as "Professor Invite" at Project Syntim.

INRIA
Domaine de Voluceau, Roquencourt
78153 Le Chesnay cedex, France

T. Kasvand

July 1990

STEREOVISION INTEMPORELLE

STILL LIFE STEREO

Tony KASVAND
Professeur Invité
Concordia University
Montreal, CANADA

RESUME :

Reconsidérant le problème de la stéréovision dans sa totalité, l'auteur résume ici les études qu'il a menées à cette occasion, lors de son séjour d'1 an à l'INRIA comme professeur invité.
Par l'application judicieuse de techniques existantes de traitement et d'analyse d'images, on espère avoir démontré que le "Problème de correspondance" peut être résolu de diverses manières. Plus particulièrement, la mise en correspondance ne nécéssite pas de "connaissances extrinsèques" du type "Modèles d'Objets" ou "Représentation du Monde", et aucune n'a été utilisée pour cette étude.
L'approche présentée est purement expérimentale, dans la mesure où les techniques utilisées sont bien connues. Le but est plutôt de reformuler le(s) problème(s) et les contraintes qui y sont attachées, puis d'analyser le résultat des expérimentations et faire partager le bénéfice des conclusions tirées.

Mots-clés: *analyse de scènes, analyse d'images, approximation analytique, contraintes de similarité, mise en correspondance, reconstruction 3D, relaxation, reconnaissance de formes, segmentation, stéréovision, traitement d'images, vision par ordinateur*

SUMMARY :

The report is a summary of studies carried out while rethinking the entire 3D Stereovision problem, during a one year long stay as Professeur Invité.
By judicious applications of existing image processing and analysis techniques, it is hopefully, partially demonstrated that the so called "Correspondance Problem" can be solved in several ways; especially that,contrary to presently popular opinions, no "external knowledge" such as "Object or World Models" is required to establish correspondance, and none are used in the report.
The present approach is purely experimental, since the techniques are well known. Its aim is rather to reformulate some problem(s) and the related requirements, then analyse are share results of carried out experiments.

Keywords: *analytical approximation, computer vision,image processing, image analysis, matching, pattern recognition, relaxation, segmentation, scene analysis, stereovision, similarity constraints, 3D reconstruction*

ABSTRACT

Measurement of distances to objects in a scene is one of the many practical applications of artificial or computer vision. Distances to objects enable a mobile robot to establish its position, it is necessary for avoiding collisions, and it allows the robot to carry out productive tasks. In other words, vision helps the robot to "see" where it is and what others are doing.

A method of measuring distance is by using stereo vision, exactly like we ourselves do. Two images are required which are taken from slightly different positions (the stereo pair of images), for example, the images from each of our two eyes, from two TV cameras, or by using two photographs or digitized images ("still-life" or "nature morte" stereo). The technical name for this field is (close-range) photogrammetry. Photogrammetry has a long history, for example, in the 1860's Colonel Aime Laussedat used photogrammetry to survey Paris, France. The principle, summarized in one sentence, is: "Given a point in the left image of the stereo pair and a point in the right image, such that these two points correspond to the same point in the scene, then the distance to the point in the scene can be computed from the measurements of the positions of the two points in the images". To find these two points in the images is called the "correspondence problem". Automated machinery exists for establishing correspondence but in close-range photogrammetry the correspondence is mainly determined by a human operator. The automatic methods used are insufficient (cross-correlation and/or Fourier analysis).

By judicious application of existing image processing and analysis techniques it is, hopefully, partially demonstrated in this report that the "correspondence problem" can be solved in several ways. It is mainly a question of our willingness to carry out the necessary work and to design and program suitable computers to make the methods practical. Contrary to presently popular opinions, no "scene understanding", no "object recognition", and no "object or world models" ("external knowledge") are required to establish correspondence and none are used in the present report. The availability of "models" greatly simplifies the problem but they introduce unrealistic constraints if the models are obtained from "external sources". The production of "models" as an integral part of the image analysis procedures is beyond present experimentation and largely beyond contemporary thinking.

The present work is purely experimental since the techniques are well known. Detailed results are shown at the different stages of processing to establish correspondence. The form of the report is: Problem, solution, results; new problem, solution, results; ... . The results are shown in image form but most of the images may carry meaning only for those versed in various aspects of image analysis. Numerous one character per pixel

prints, gray level prints, and stereo pairs are used to illustrate the results. Unfortunately, the gray level prints, specially after copying, look "dismal". For those interested, references are given to computer files from which the images may be displayed or printed. The entire report should only be considered as "introductory reading for dispelling a mystery" for students wanting to study image analysis applied to 3D vision. Research, after all, is a learning experience: Make experiments, see what happens, learn from them in order not to repeat the same mistake twice, and inform others "where not to step".

Chapter 1 briefly discusses some perfectly obvious aspects of stereo images which appear to have been ignored by many or considered as being too trivial. A "similarity principle" is defined, which simply states that "the images of a stereo pair are very similar". The rest of the chapter is devoted to what should be computed from the images and how the correspondence (matching) is to be accomplished. A large number of possibilities are pointed out.

The work starts in Chapter 2 with image processing requirements, followed by grouping of "pixels" to simplify the detection of similarities between the left and right images of the stereo pair. Classical techniques are used, consisting of pixel feature computations, classification of features to create homogeneous regions in the image space, and analytic approximation of the regions followed by "relaxation". (The processes and decision spaces required for images of thin line structures and textures are not included.) Great care is taken to only use features that preserve "stereo fidelity", i.e., images of the features can be seen in stereo with a pair of stereo glasses. It is highly doubtful that features that "ruin the stereo effect" will be of much use in subsequent analysis.

The similar regions found are put into approximate correspondence ("raw matches") in Chapter 3. Two techniques were studied in some depth, called "matched classification" and "direct and-ing" of the regions. Several other methods are mentioned but could not be studied in detail. Since the stereo pair of images used was not of high quality, an inordinate amount of effort was spent on correcting the dissimilarities between the images. This became more of an intellectual challenge than practical necessity since we can see the stereo effect despite the dissimilarities. In practice one would simply readjust the iris on the camera and take another picture. The approximate correspondence is then "adjusted" for better agreement with the information in the two images. In the preliminary experiments a second order polynomial is used for analytic approximation of the gray levels of the matched regions. This is followed by "relaxation". However, better analytic functions and more controlled relaxation processes are required for recovering the "stereo effect" to greater accuracy.

The report ends rather abruptly at this stage with a review and critique. Several reasons existed:

1. It became apparent where the "weak links" in the chain of reasoning were but there is no time to correct these at present.

2. The results "couple to and simplify" the presently popular "edge-based" stereo techniques since the combinatorial "explosion" problem is removed.

3. The procedure is iterative (hierarchical) where the matched facets obtained are simply viewed as "smaller images" to which the same techniques can be applied again, resulting in finer and finer correspondence between the details of the images.

A brief "illustrated summary" is given below. The correctness of the results is verified stereoscopically, i.e., stereo glasses are required to see the 3D effects in the images:

The original image pair is shown in Figures 1a and 1b. Even though it looks like a "room scene" to us, such "understanding" is not available to the machine, nor is it needed. The same stereo pair on a smaller scale is shown in Figure 1ab.

The scene is split into regions of pixels with homogeneous characteristics, see Figures 2a, 2b, and 2ab. (Seven gray levels or "colours" were used to "paint" the regions such that adjacent regions have different "colours".)

The homogeneous regions are matched and as many additional "matched" regions are found as is feasible, see Figures 3a, 3b, and 3ab. These are the "raw matched facets" ("painted" in six "colours"). The "raw" matches, when viewed stereoscopically, create a scene that appears to be partially "exploded". One set of (matched) facets are approximately in their "proper positions" in the scene, while others are "hanging in free space" in front of the scene". Different matching methods tended to produce different visual effects. However, the visually most disturbing effect is caused by minor differences in adjacency relations between the matched regions, i.e., if there is a small space between two facets in one image but not in the other, this space "cannot be matched" by our vision.

The analytic approximation and relaxation steps are an attempt to correct the raw matches. Some preliminary results are shown in Figures 4a, 4b, and 4ab (painted in six colours). As can be seen, the analytic approximation and relaxation techniques described correct the facets only to a limited degree but leave much "fine-grain" noise and may create occasional "gross errors". The small dissimilarities at the boundaries of matching facets produce "visual disturbances" when the results are viewed with

stereo glasses, since our vision is extremely sensitive to such small incompatibilities. In the image space most of these errors only involve a few pixels at a boundary and could be "written off" as "spatial quantization effects".

However, since such small differences are vital to our stereo vision, they are also important for any automated stereo vision system. The differences can be found if the scene is recreated from the analytic approximation. For example, the pixels that could not be assigned to any region are shown as "black zigzags" in Figures 5a, 5b, and 5ab. The pixels at which additional corrections are required are found by comparing the gray levels of the pixels in the original image and in the analytic approximation. For example, if the pixels in the analytic image are replaced by those in the original when the error is greater than 15 units or where the approximation does not exist, the "stereo effect" reappears, see Figures 6a, 6b, and 6ab. Thus, it is known where future corrections are to be made to achieve "full stereo fidelity".

Figure titles:

Figures 1a, 1b, 1ab: The stereo pair of images used for analysis. The left image (1a) and the right image (1b) form a stereo pair but no equipment existed for viewing such images. Hence, a smaller version of the same pair is shown in (1ab).

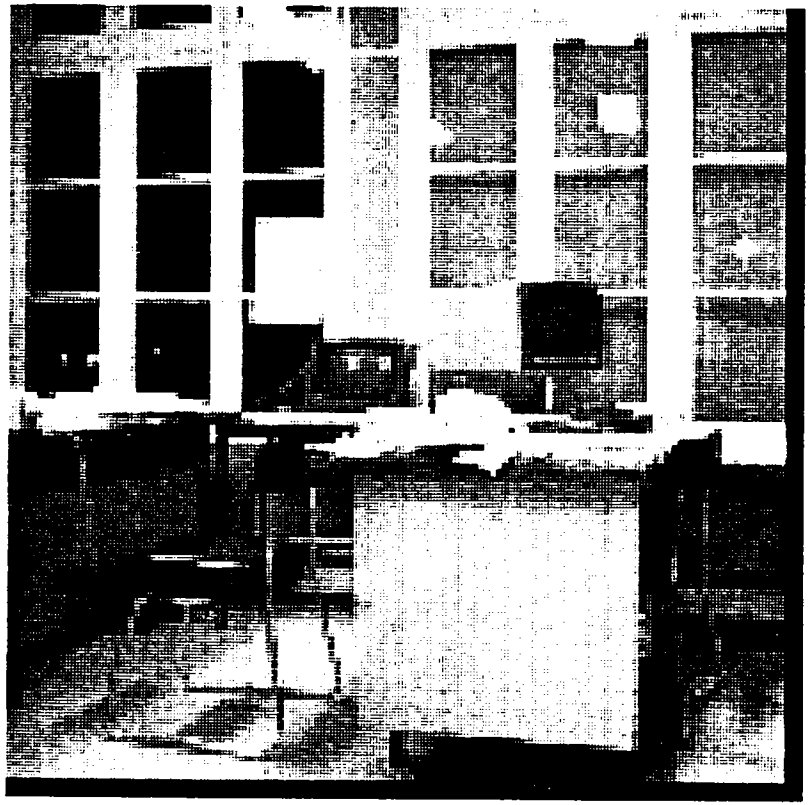
Figures 2a, 2b, 2ab: Regions of pixels with homogeneous characteristics "painted" in seven "colours" such that adjacent regions have different "colours". (2a) is the left image, (2b) the right image, and (2ab) shows a smaller version of the same pair.

Figures 3a, 3b, 3ab: The "raw matched facets" ("painted" in six "colours"). (3a) and (3b) are the left and right images, and (3ab) shows a smaller version of the same pair.

Figures 4a, 4b, 4ab: The matched facets after analytic approximation and relaxation ("painted" in six "colours"). (4a) and (4b) are the left and right images, respectively, and (4ab) is a smaller version of the same pair.

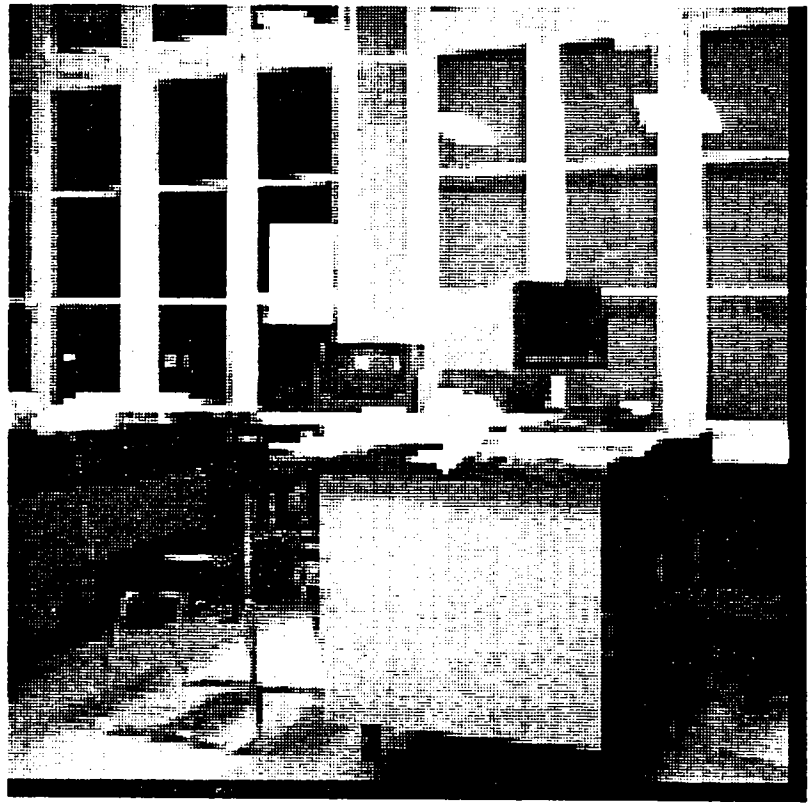
Figures 5a, 5b, 5ab: The scene in Figure 1 recreated from the analytic approximation after relaxation. Pixels where the approximation does not exist are shown as black "zigzags".

Figures 6a, 6b, 6ab: The scene in Figure 1 recreated from the analytic approximation after relaxation. Pixels where the approximation does not exist or where the error between the approximation and the original image is "too large" (>15 units) have been taken from the original image.



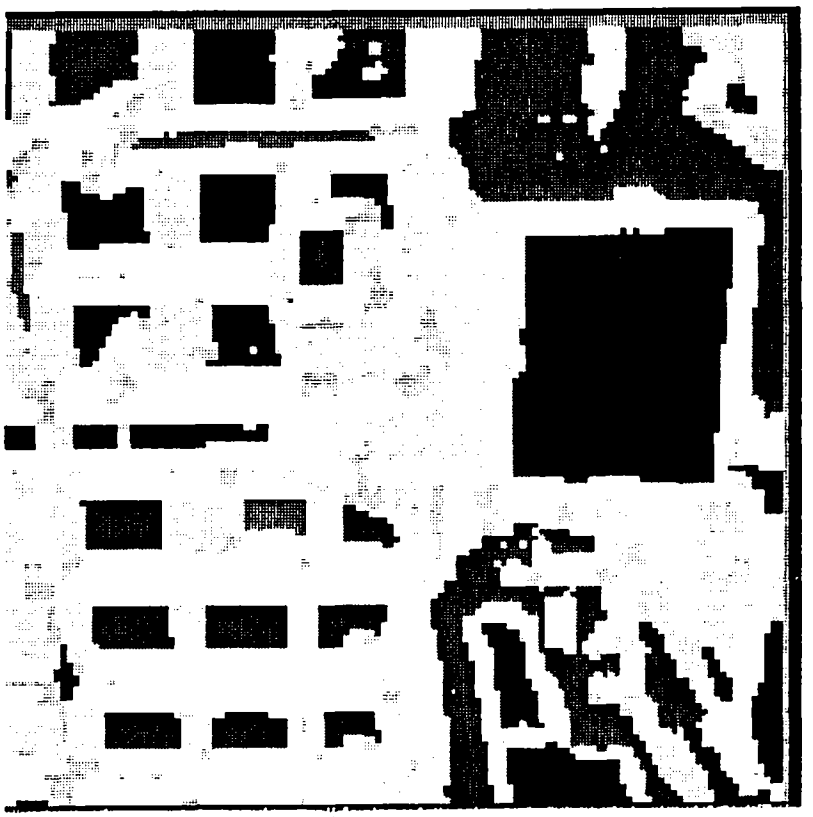
if8gray.lsh

Fig.: 1a.



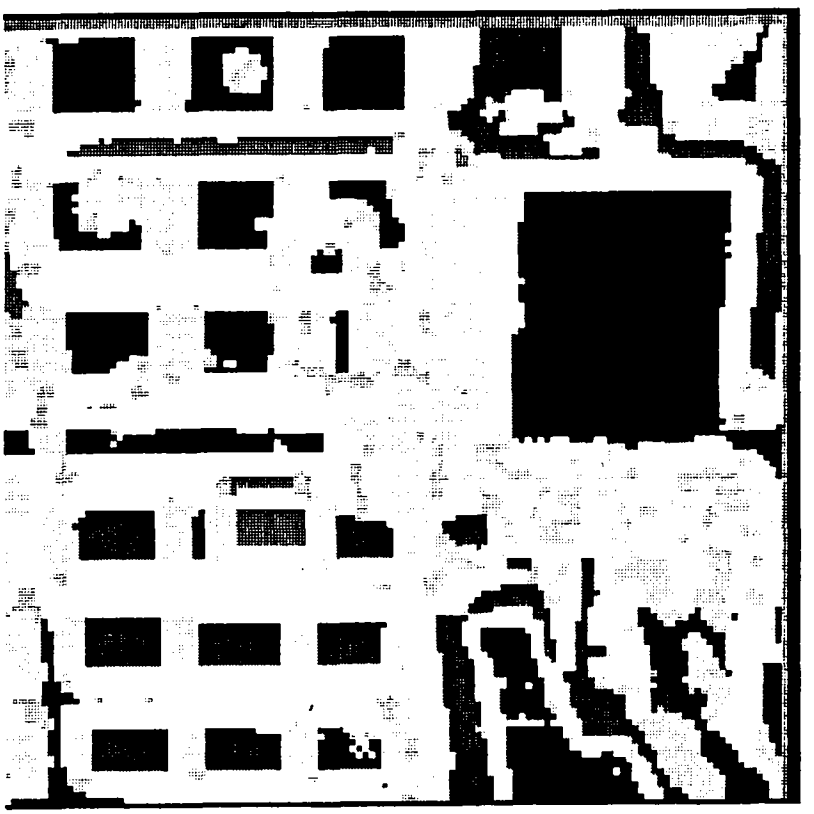
if8gray.rsh

Fig.: 1b.



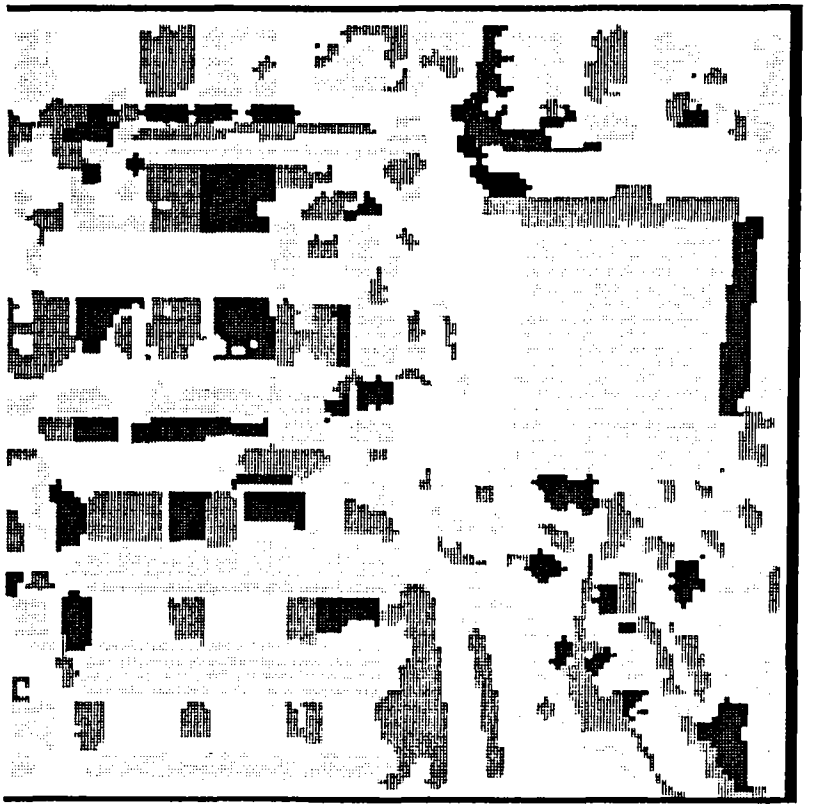
ifshkpp2.lgr

Fig.: 2a.



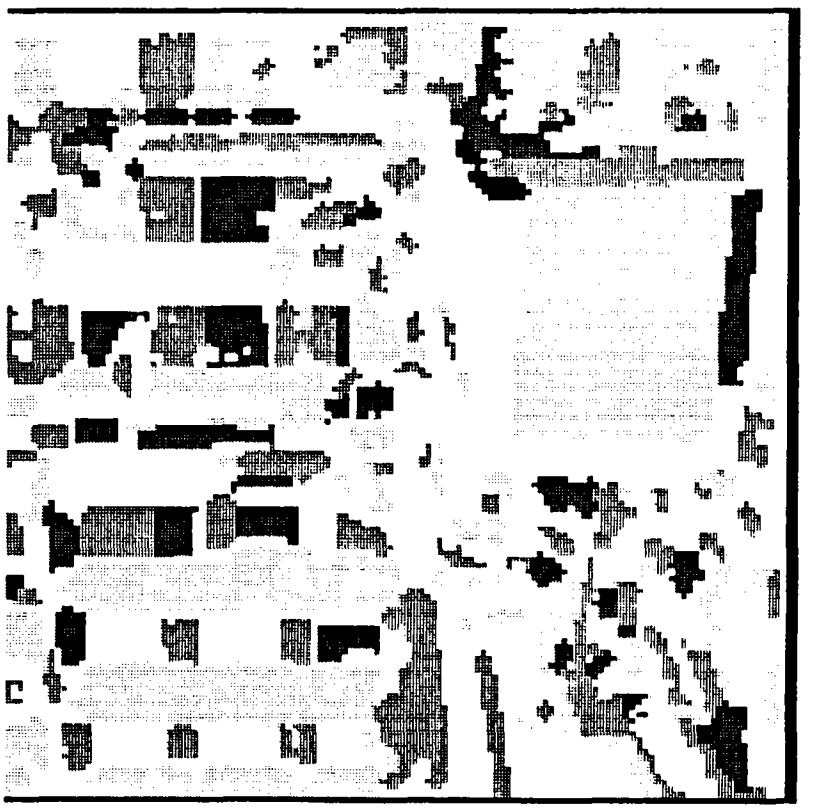
ifshkpp2.rgr

Fig.: 2b.



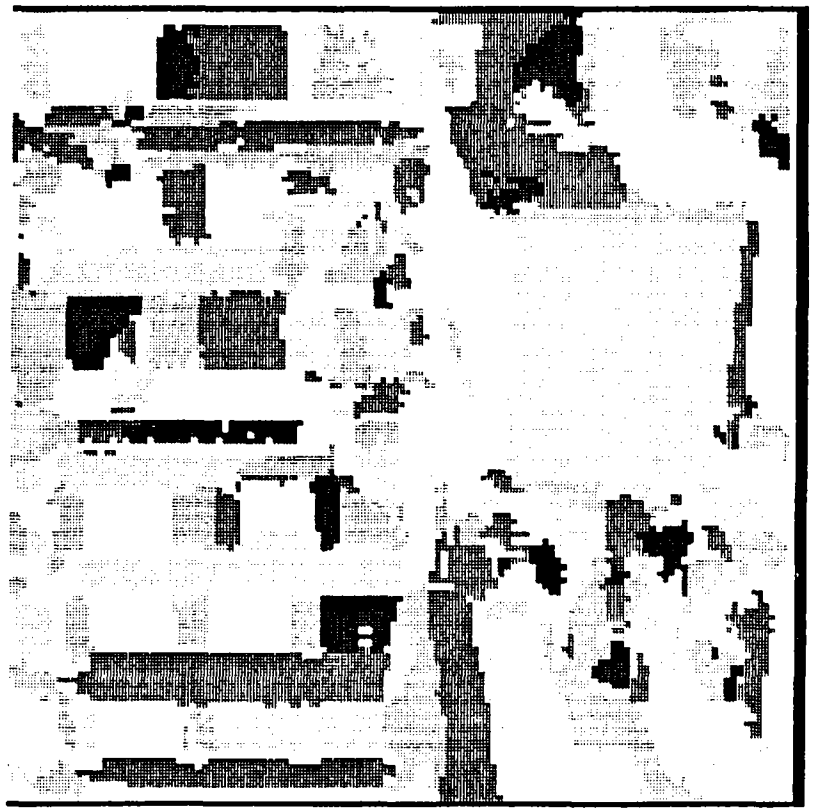
if8flkan3.rgr

Fig.: 3b.



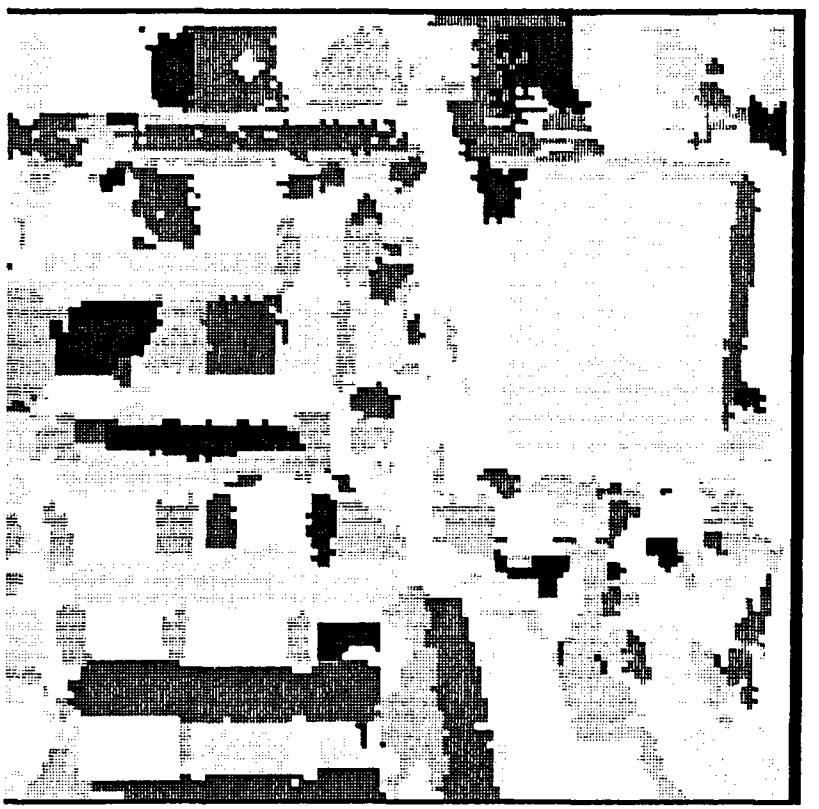
if8flkan3.lgr

Fig.: 3a.



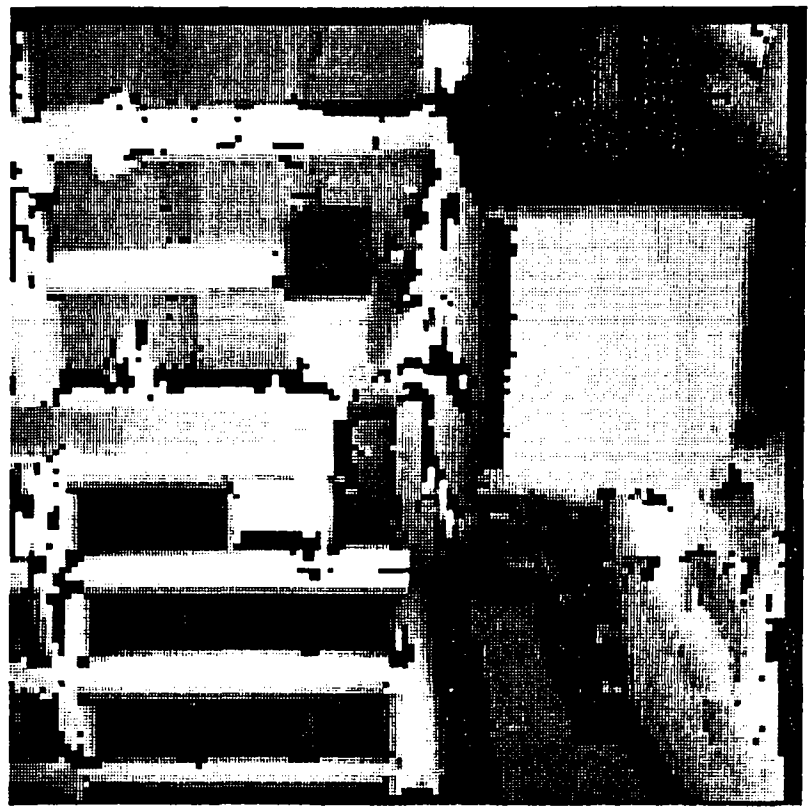
if8flkan4.rgr

Fig.: 4b.



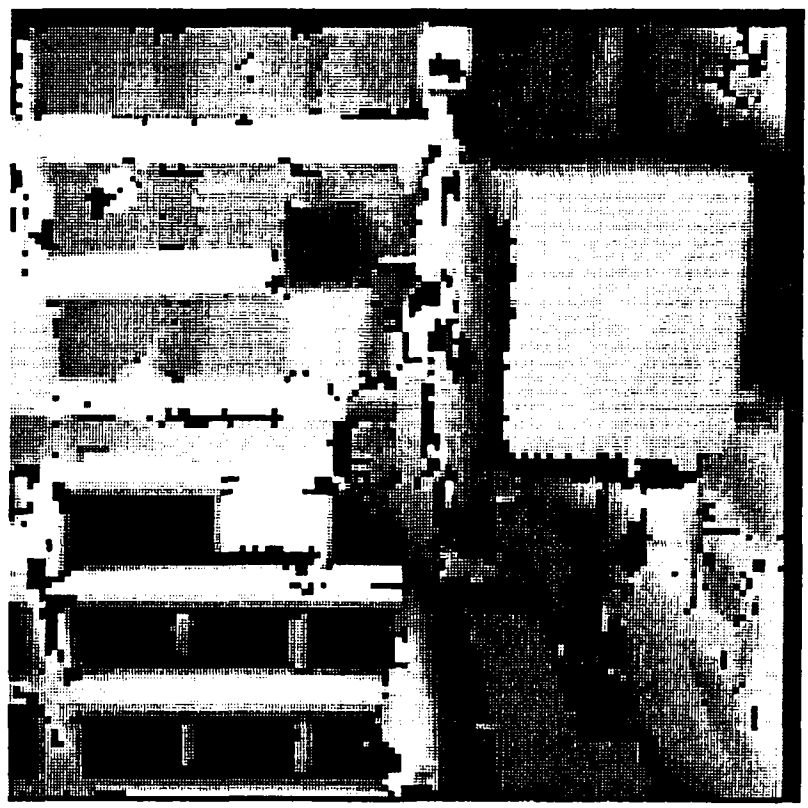
if8flkan4.lgr

Fig.: 4a.



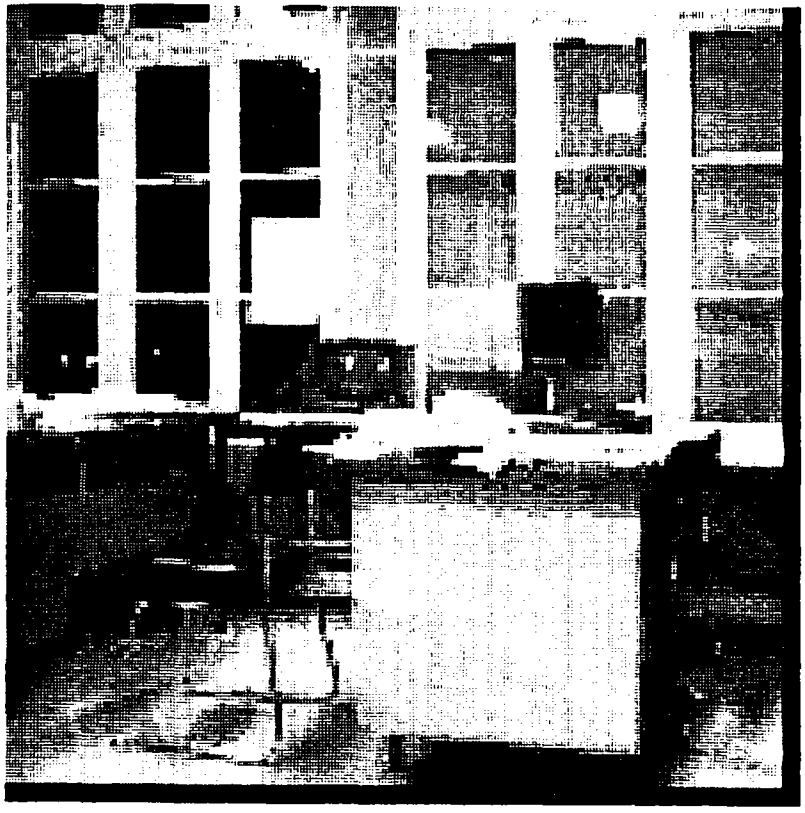
if8gacan9.rgr

Fig.: 5b.



if8gacan9.lgr

Fig.: 5a.



if8gacana.lgr

Fig.: 6a.

if8gacana.rgr

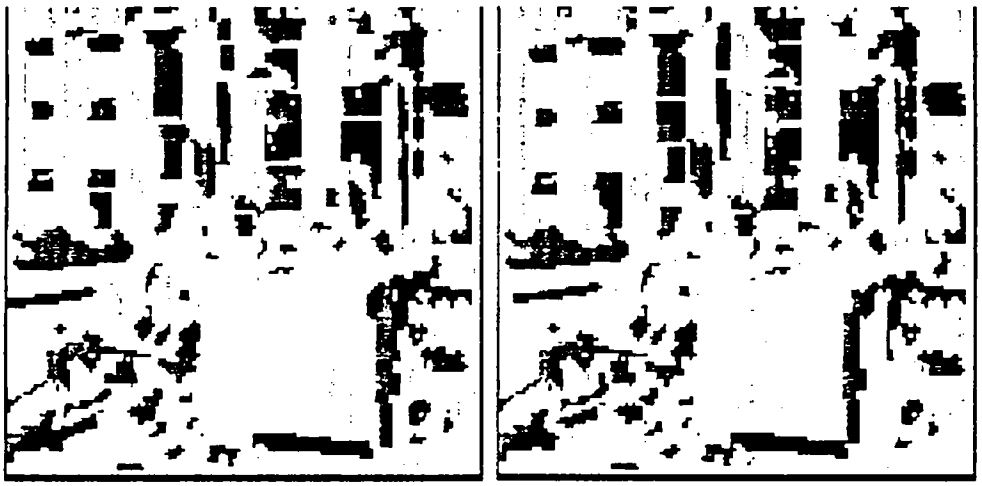
Fig.: 6b.

if8gray.lsh if8gray.rsh

Fig.: 1ab.

if8hkpp2.lgr if8hkpp2.rgr

Fig.: 2ab.

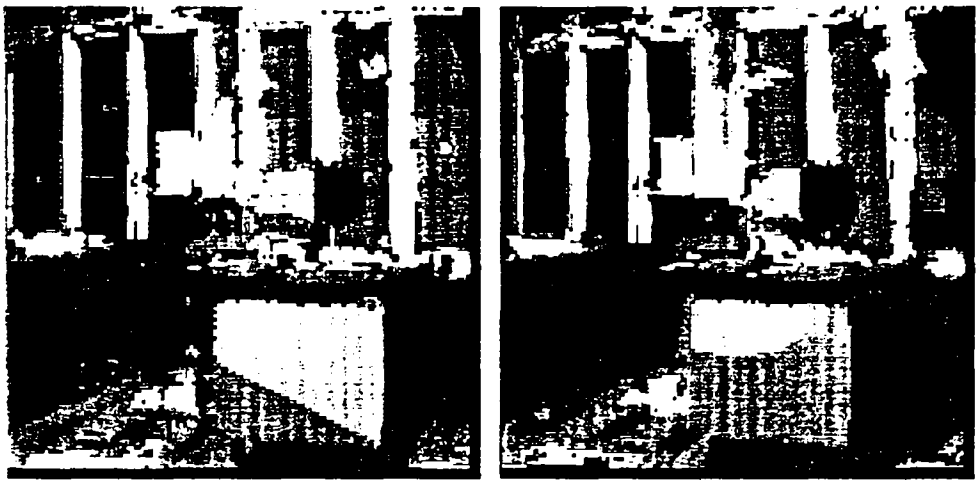


if8flkan3.lgr if8flkan3.rgr

Fig.: 3ab.

if8flkan4.lgr if8flkan4.rgr

Fig.: 4ab.



if8gacan9.lgr if8gacan9.rgr

Fig.: 5ab.

if8gacana.lgr if8gacana.rgr

Fig.: 6ab.

INDEX

Abstract
Index
Acknowledgements

Chapter 1: THE OBVIOUS page
1.1 Introduction 1
1.2 Seeing distance 4
1.3 The similarity principle 7
1.4 The hierarchical approach 8
1.4.1 The quantities "x" and "X" 10
1.4.2 Matching "X" 11
1.4.3 Comments 13
1.5 Validity requirements 14
1.5.1 The ideal stereo camera 14
1.5.2 The irreducibles 22
1.6 Conclusions 24

Chapter 2: PRELIMINARY COMPUTATIONS
2.1 Introduction 31
2.2 The image pair 39
2.3 Computable and computed features 41
2.4 Hierarchical feature classification 50
2.5 Analytic approximation and relaxation 73
2.6 Super-regions versus scale-space 97
2.7 Conclusions 99

Chapter 3: FACET MATCHING
3.1 Introduction 103
3.2 Matching methods 108
3.2.1 Matched classification 109
3.2.2 Direct "and-ing" 119
3.2.3 Correlation matching 130
3.2.4 Facet recognition matching 132
3.2.5 Pixel feature matching 135
3.2.6 Other methods 136
3.2.7 Comments 136
3.3 Modifying the various matches 139
3.3.1 The Lh, L&Rh, and Rh triplet 140
3.3.2 The raw Lm and Rm images 151
3.3.3 Misclassifications 155
3.3.4 Cut 159
3.3.5 Transplant 162
3.3.6 Clip 165
3.3.7 The gaps and the un-matchabels 173
3.3.8 Comments 174
3.4 Analytic methods 177
3.4.1 The coupling 178
3.4.2 Constraints on relaxation 181
3.5 Comments 206

REVIEW AND CRITIQUE 207
References 211
Appendix: IMAGE FILES 212

Acknowledgements

The invitation to INRIA, project Syntim, has been highly appreciated for the rare opportunity it offered to rethink the entire 3D stereo problem. Previously, the author had "stayed away" from the stereo problem while "observing" the efforts of others. During this one year the 3D stereo problem was restarted form beginning but in a different direction from the presently popular methods. This is only possible in a place where "liberte, egalite, et fraternite" rules.

Special thanks are due to Dr. Andre Gagalowicz for the invitation, to Mr. Jean-Paul Chieze for general assistance with the "Bora-Cumulus-Archille-Unix-C" environment and for several programs (delret, kasv2im, etc.) without which the work would have been impossible, to Mr. Laurent Vinet for more programs (im_laser, etc.), and to all the other colleagues who have provided valuable advice and assistance, and last but not least, to Mlle. Laurence Bourcier for taking care of the "administrative aspects".

Chapter 1: THE OBVIOUS

1.1 Introduction

Since the history of photogrammetry is rather long, the meaning of the word "photogrammetry" has been redefined several times. According to the Manual of Photogrammetry (1.1), the original meaning was "the science or art of obtaining reliable measurements by means of photographs", later "interpretation of photographs" was included, and by now the definition includes "remote sensing" and more. Possibly automatic image processing, analysis, and "pattern recognition" techniques will be included in the future.

The theoretical as well as the practical aspects of photogrammetry have been well "worked out" and will neither be discussed nor summarized here. Perusal of the more than 1000 pages of the "Manual of Photogrammetry" can give the interested reader a fair overview and numerous references.

As stated in the Abstract, if two corresponding points have been found, one in the "left" and the other in the "right" image of the stereo pair, then the distance to this point in the scene (the 3D world) may be computed from the measurements of point positions in the images. The computational methods include corrections for the way the two photographs are "tilted and rotated" with respect to each other, camera displacements and optics, lens distortions, and so on. The correspondence between the two points is established by an operator (using appropriate equipment), after which the calculations are performed by computers. Methods and equipment exist for establishing correspondence automatically. The methods are base on cross-correlation or (equivalently) on Fourier analysis of small regions in the two images. When the two regions "agree" the (phase) shift corresponds to the height or distance that is to be found. Automatic comparisons give reliable results for (contrasting) "far-away" or rather "flat" scenes for which the two (stereo) images are, essentially, "shifted copies with local variations" of each other.

If the stereo pair is taken at "close range", then the variations between the images are considerably larger and one "eye" may see "things" that the other does not. Of course, automatic comparisons still work, but only in selected parts of the images. Consequently, a human operator is needed to guide and/or assist the machine. Note that the fundamental requirement is to "establish unique correspondence between points in the two images of the stereo pair". When these techniques are applied to, for example, the vision system of a mobile robot then the human operator cannot be involved (all the time). Each image pair has to be analyzed purely automatically in order to establish cor-

respondence between points in the images and to compute the distances.

Since images of real scenes have very complex content, the two standard approaches that have been used are:

1. Ignore the scenes altogether and put some point light sources onto the objects in the scene that are of interest.

2. Ignore the scene and put some corner reflectors onto the objects of interest.

In both cases the images of the scene consist of "uniform black" with some "bright points". The bright points are located with (binary) image analysis methods, put into correspondence, and the distances are computed. The essential difference between the two methods is that in the first (1) the lights can be turned "on and off" thereby making the correspondence problem trivial, while in the second (2) reflectors are placed "far enough apart" to again make the correspondence problem trivial, or the machine is first assisted by an operator who "puts a box around each mark" and establishes correspondence, after which the machine "follows the points in the images" while the object moves, thereby keeping the points in correspondence. Formally, three non-collinear points with known spacing are sufficient to compute the six "degrees of freedom" of an object (marks). Commercial hardware exists (or existed) for both methods.

Numerous variations are possible and have or are being used, such as:

3. Mark the scene but keep the background around the mark "clean" such that the mark can be located with elementary image processing techniques. The marks may be of different shapes and/or in different colours.

4. Use "prominent feature points" in the images of the scene.

5. Use "edges" in the images of the scene.

The last two methods (4,5) are current research topics in many laboratories (1.2). The basic difficulty with these approaches is that they create a "combinatorial explosion" while trying to establish correspondence.

In industrial applications many "special effects" are used to establish distance or position of some "object of interest". In stereo vision the accuracy of the range measurement depends on the distance over which the measurement is made, on the displacement of the two cameras, and on the focal length of the lenses versus the resolution of the image sensor from which the stereo parallax is obtained. Clearly, with closely placed cameras and far-away objects the parallax (or displacement "d") in the two images is practically zero. The accuracy of the distance measures

is proportional to 1/d*d, i.e., the parallax "d" has to be "significant" for obtaining any accuracy at all with stereo based methods.

However, in addition to all this, there is the entire field of "active vision" for detecting depth. (Unfortunately, the word "active" has now become confused and has two meanings. The "old" meaning used to be "active = emit energy (light)" versus "passive = do not emit energy". The "new" additional meaning is "active = moving or mobile"). The principal active (= emit energy) methods of measuring distance are:

6. The laser range scanner which uses triangulation to measure distance. The scanner gives the distance readings directly and to very high accuracy.

7. The "light strip" method where a flat plane of (laser) light is projected onto a scene. In the images the strip appears as a "crooked" line. The shape of the line is related to the shape of the object and the position of the line is related to the position of the object in the scene.

8. Moiré patterns of interference fringes have been used.

9. Radar, lidar ("light radar"), and sonar ("acoustic radar") measure distance as a function of travel time.

Another categorization of "depth vision systems" is that they are:

a) Passive, such as stereo, where the observer can be immobile and "emits no radiation". Thus, except for the two "eyes", the observer can remain concealed (hidden) since only the ambient light is used.

b) Active, such as the laser range finder, which emits a laser beam to triangulate for distance. In the active systems the observer "advertises" his presence by emitting energy in order to "see".

It is hoped that this brief survey places the stereo depth (range) vision system in its proper "perspective", i.e., it is one of many methods. In a practical application all the possibilities should be considered before making a committment.

1.2 Seeing distance

It may appear from the brief survey in the previous section and from much recent literature that "all aspects of seeing distance (depth, range)" have been covered, but this is not the case at all since the need for seeing the distance has not been defined. The need for seeing distance defines the required accuracy of the distance measurements.

In practice the required accuracy of the distance measurement depends on the application. For close objects distance is a very important "variable", but frequently distance is not important for "far-away" objects. A "relative" measure of distance is often far more important, i.e., is the object "in front of" or "behind" something else. The knowledge that the object is far away is sufficient in itself and it is far more important to know the speed of the object relative to the observer. Even though, mathematically, velocity is but a derivative of distance, the important aspects of "velocity" can be observed without measuring distance. To "see" this (superficial) contradiction between mathematics and reality, consider the driver of a car. Does the driver really measure distances between say 5 to 100 meters to an accuracy of centi- or deci-meters in order to compute the relative velocity vectors of all the other cars to the required accuracy? The human vision cannot measure a distance to the accuracy of a centimetre within a range of even one meter! Clearly, much of the vital information required to "operate" in a 3D world is both different and obtainable by other means.

The human visual system uses about a dozen methods at the same time for detecting depth or distance. Only one among these methods is based on stereo pairs of images and it is claimed that more than 10% of the population cannot see depth based on pure stereo vision, but they may not even be aware of this loss (1.3). There are many cues to depth and it appears that almost all of them are used. The following is a very brief description to indicate the diversity and finessing that one finds in biological information processing systems. For illustrations and better appreciation of the details the original sources should be studied (1.4, Gibson). The ordering of the phenomena used are according to Gibson. Some comments have been included to indicate the nature of the computer vision problem if the use of this effect is contemplated. Thus, directly quoted or slightly modified:

1) The texture perspective or the texture gradient. The sizes and shapes of the texture elements in the image vary as a function of displacement. The detection of texture gradients is a topic in computer vision.

2) Size perspective. Objects vary in size as a function of distance. This requires object recognition and the effect can be used to create visual illusions by producing conflicting situations between the foreground (objects) and the background. Such conflicts, however, seldom occur in nature.

3) Linear perspective or parallel lines "meet" at infinity. From the image processing side this requires line or contour detection, estimation of slopes and curvatures, and clustering of the results. Certain aspects of this effect are used when the scene is illuminated with structured light.

4) Binocular perspective or stereo vision. This uses both the displacement and skew of the image in one eye with respect to the image in the other eye. The necessary comparisons can only be carried out if the image contains details (contours, textures). In order to simplify the comparison problem, for example, the scene may be illuminated with randomly structured light.

5) Motion perspective. The changes that occur in the image due to motion are very complex but also highly informative. The scene expands, contracts, skews, or rotates around the fixation point, i.e., where we fix our gaze. Added complexity is introduced when the objects in the scene are also moving independently of the observer. These effects are easy to observe, for example, when watching the landscape from a moving train or car. The total "visual flow" gives a very compelling impression of one's state (position, orientation, and motion) with respect to the rest of the world. It is also well known that actions can be easily recognized by motion alone even when the objects themselves are invisible, for example, when the moving objects carry some bright spots while the objects themselves are in the dark. Illusions occur when the observer is not aware that he is being moved. Clearly, the analysis of motion is exceedingly well developed in the biological vision systems. Only some simple aspects of motion and visual flow have been studied in computer vision.

6) Aerial perspective or the haziness, blueness, and desaturation of colours as a function of distance. This varies with the illumination of the scene and the scattering properties of the intervening medium. Good artists know how to use this effect in their paintings. These effects have not been used in computer vision, even though the scientific aspects of light scattering are quite well developed and sometimes used to correct satellite or high altitude photographs.

7) The perspective of blur or the variation of the quality of blur as a function of displacement from the centre of clear vision. This effect is hard to observe since we automatically focus on what we observe.

8) Relative upward location in the visual field (on the retina or in the image) if the background is assumed to be a (horizontal) terrain. This effect is easy to measure once the objects in the image have been identified.

9) Sudden shift of texture density or linear spacing between texture elements (usually coincident with changes in brightness or colour). The sudden change in texture density generates a

contour where the more densely packed texture elements appear farther away. This is the texture contour problem in computer vision.

10) Shift in the amount of double imagery occurring at a contour. When a contour is observed with two eyes the images differ near the contour. This effect is one source of problems for automatic stereo mapping.

11) Shift in the rate of motion of the texture elements in the image on one side of a contour with respect to the other side of the contour when the observer's head moves. The shift of a contour in the image is directly related to the distance between the contour and the observer, measured with respect to a (stationary) background. Simpler aspects of this problem are called "change detection" in computer vision. Similar triangulation is used in the laser range finder but, of course, the physical realization is very different.

12) Completeness of an object or the continuity of its outline is an indicator of what is in front of what. Complete objects appear closer since only they can obscure what is behind of them. In computer vision the closest parallel is the analysis of "blocks worlds".

13) Transition between light and shade. Sudden shifts in the brightness of adjacent regions generate contours but also generate the appearance of depth. However, this problem appears to be unresolved. For smooth surfaces the brightness gradient has been used to infer shape in computer vision (1.5).

These thirteen effects are cues to depth or variables computed from the image by our visual system, as described in literature. The stationary one-eyed observer can see texture, size, and linear perspective (1,2,3), aerial and blur perspective (6,7), the relative positions of objects (8), as well as depth at a contour (9), continuity of outlines (12), and transitions in brightness (13). The one-eyed mobile observer can, in addition, also see a form stereo with time delay (4), many aspects of motion perspective (5), and the texture shifts occurring at contours (11). The stationary two-eyed observer can, of course, see everything that the stationary one-eyed observer sees, and also stereo (4) and the doubling of texture at contours (10). The tilt, orientation, and curvatures of a surface are also immediately visible but how these are computed in the human visual system is not certain.

The biological vision system is designed for survival. Consequently, in order not to "advertize" one's position or even existence, the vision system has to be passive. Motion is a very easy "feature" to detect and consequently very dangerous. Hence, the understanding of the visual scene must not require motion on the part of the observer. Stereo vision provides reasonably good distance information (lens focus is used by chameleons). Stereo

vision requires minimally two convergent eyes. In principle, three or more eyes could provide "better coverage" but more than two eyes occur only in very primitive organisms. The biological vision system has to "understand" motion, both its own and that of other moving objects, in order to be able to navigate in the three-dimensional environment. Consequently, motion understanding is very well developed.

As may be observed, stereo vision only plays a minor role and when the ability is missing (in about 10 to 11 per cent of the population) the lack may not even be noticed! Motion based "stereo" is far more informative. However, the present work has been restricted to "still-life" (nature morte) stereo where all motion is excluded.

1.3 The similarity principle

The nature of the stereo problem may be "seen" by simply looking at a reasonably complex scene. Look at the scene with one and the other eye without moving your eyes or head and ask: What did one eye see and what did the other eye see? The answer is, of course, perfectly obvious: We see practically the same scene with one eye or the other! Alternatively, we may look at a pair of stereo photographs. Usually it is hard to see any difference at all between the two photographs without careful inspection of object relationships. After one is convinced that this is true, the rest is just a matter of "working out" the technical details.

The "secret" of the stereo vision problem is thus nothing more than knowing that the two images of the stereo pair are highly similar but, of course, not identical. This is the basis of the "similarity principle" upon which the solution is based. The similarity principle is very powerful since it does not require object or scene "recognition", nor any "scene modelling", no "scene understanding", and the "meaning" of the scene can remain unknown even after stereo reconstruction. In other words, none of the "high level methods" are needed, which are currently very popular. In this sense the present approach is "orthogonal" to the current thinking, since all that the present approach requires is a high degree of similarity between the two images of the stereo pair. The technical requirements for similarity and the differences from photogrammetry are discussed in section 1.5.

Of course, the similarity principle has been used implicitly by all stereo based methods. Automatic cross-correlation or Fourier transformation of small areas in the images finds the "shift" between the similar regions in the two images. Feature point based methods "look for" similar features points, and edge based methods search for matching edges in the two images. However, why stop here when there is a practically unlimited number of additional "similarities" to be found between the two images!

1.4 The hierarchical approach

The "similarity principle" only requires the left and the right images of the stereo pair to be "similar", allows unknown images to be matched (put into correspondence), and allows the 3D scene to be reconstructed without the need to know what the scene represents. Consequently, "similarity" requires no "image understanding", object "recognition", or "models", etc., but if this information is available, it can be easily incorporated.

"Similarity" can be used on many "levels" and, consequently, there are may different approaches to solving the stereo problem. The final goal, however, is the same in all cases, namely, each of the pixels in the two images have to be put into a unique correspondence, except where the two "eyes" see different things. The match cannot be one-to-one on the pixel level since, for example, a "stick" in the scene may be represented by 10 pixels in one image and 12 in the other, but this is only a quantization and resampling problem. The phrase "pixel level match" is used for convenience but it should be understood to include resampling or treated as an approximation given the present quantization.

The basic problems to be solved are, thus:

A. How to extract some quantities "x" from the two images, and what are these x-s?

B. How to group these "x-s" into some larger units, groups, or "X-s", if necessary, and what are the conditions that these groups have to satisfy?

C. How to match the x-s or the X-s according to similarity in the two images?

D. How to compute the distances after the matching has been accomplished?

The technical requirements for similarity are discussed in section 1.5. Here it may suffice to say that: "The larger the region in the images that is represented by one "X", the simpler it is to match the "X-s" in the two images, the simpler it is to find a unique match, and an increasing number of techniques becomes available for matching".

An hierarchical strategy is proposed as sketched in Figure 1.4-1, out of which only the steps marked with a star (*) have actually been programmed and verified experimentally. These steps are described in detail in Chapters 2 and 3.

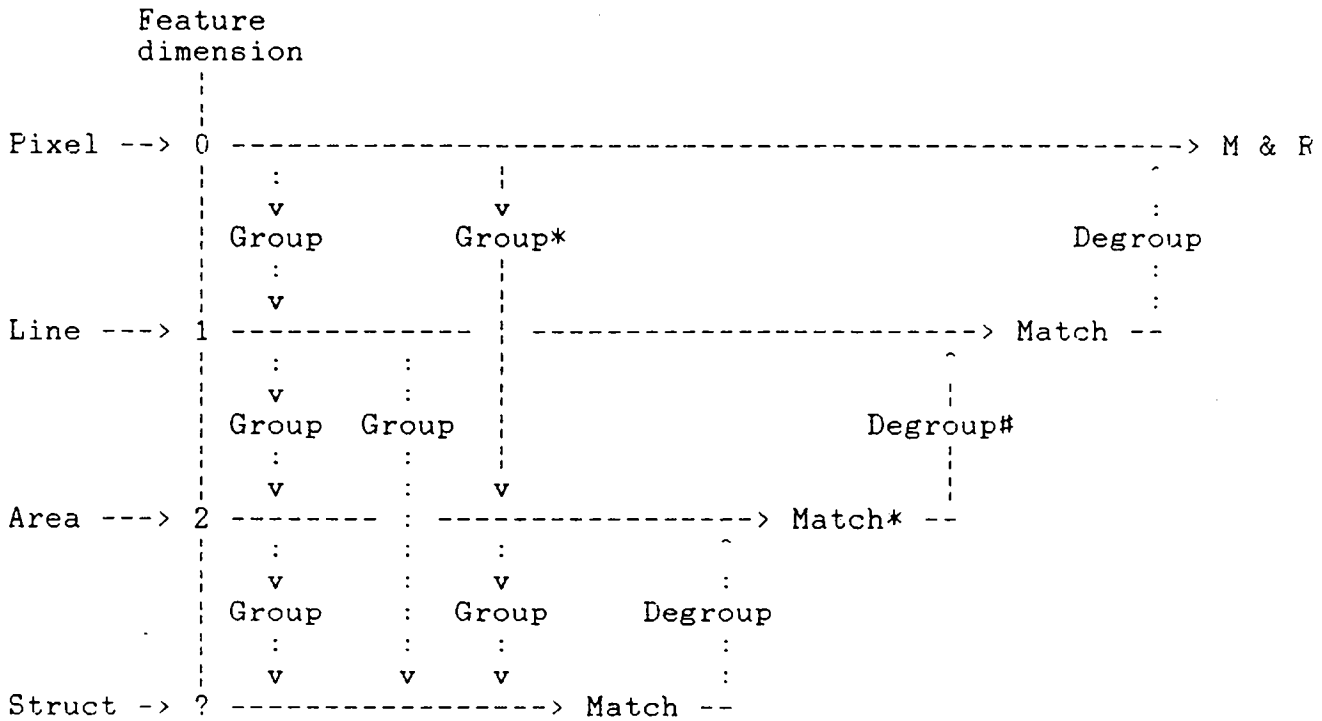


Figure 1.4-1: An hierarchical stereo matching method.
Struct = structure.
M & R = match and reconstruct.
* = Discussed in Chapters 2 and 3.
# = Edge matching.

1.4.1 The quantities "x" and "X"

The quantities indicated by "x" and "X" have many different names in image processing and analysis literature. In the present case these quantities are categorized according to their dimensionality in the image space:

1. Zero-dimensional. These are features, characteristics, or properties that "belong" to a pixel. In short, these are called the "pixel level" features.

2. One-dimensional. These are single or non-branching one-dimensional regions in the image space normally called "lines" or "edges". In the present case all the features associated with edges, i.e., slope, curvature, strength, etc., are included.

3. Two-dimensional. These are usually called segments, regions, facets, areas, etc.

4. Structures. These are various compositions of the above but in the present case only the composites of (3) are implied.

The zero-dimensional features are of two kinds:

i) The features measured at pixel level, i.e., the gray level intensity of the pixel or the intensity of its colour components if the image is in colour. Temporal effects (flicker) are also pixel level features if a series of registered images is available.

ii) The features computed from a local neighborhood around each pixel. The number of such features is practically unlimited.

For the one- and two-dimensional regions in the images and, of course, even more for structures, the numbers and types of features computable is high and difficult to enumerate briefly. From the practical point of view the number and variety of such features may be considered unlimited.

In the usual cases the one-dimensional features are "edges" and the process is called "edge detection", i.e., the pixels are grouped according to selected zero-dimensional features (or functions of such features) into one-dimensional configurations in the image space. Of course there are many methods among which the conventional "edge detection and edge linking" is but one. Clearly, it is highly desirable that the one-dimensional groups (edges) in the left and right images are "true" (very similar) such that when we look at such images, the stereo effect is immediately apparent. However, this may not be an absolute necessity if subsequent processing of the "raw" results will "chop", "cut", and "clip" these regions such that the stereo effect

reappears. This "avenue" was not explored in the present study, partially due to much prior work already completed on straight edges, and the simplifications that can result from using two-dimensional features.

The two-dimensional quantities "X" are pixels grouped according to selected zero-dimensional (but sometimes also one-dimensional) features (or functions of such features) such that these groups represent two-dimensional regions (areas) in the image space. These methods are normally known as "region segmentation" or "region growing". However, much better results are possible by using classical pattern recognition techniques. The classes or clusters obtained are simply mapped or projected back into the image space, for example, as is done in crop classification used in remote sensing. Actually, both the grouping and subsequent matching process is mainly an application of classical pattern recognition methodology, where both much theoretical as well as practical experience is available.

In the present study the grouping was based on the gray levels and the gradients of the gray levels. The resultant regions in the image spaces had poor "stereo appearance" since there was a marked difference in the two images of the stereo pair. However, the "stereo effect" reappeared after the regions were processed.

Structures represented by "X" can have infinite variety, their descriptions can become exceedingly complex but also unique such that the matching problem reduces to a triviality (but the construction of the descriptions that match in the left and right images is by no means trivial). Prior work in these areas is called "structural pattern recognition". In psychology certain types of these structures are known as "gestalt groupings" or simply "gestalt".

1.4.2 Matching "X"

Matching is just another name for uniquely pairing the quantity "X" in the left and right images of the stereo pair. As usual, there are many methods depending on what the quantity "X" represents. For example, cross-correlation and/or Fourier analysis used in existing photogrammetric instruments uses the gray levels (possibly enhanced), i.e., they depend on one zero-dimensional pixel feature distributed over a small area of the image. Other examples are the "feature point methods" which are also essentially based on zero-dimensional features computed from local neighborhoods, and the "edge-based" methods which use one-dimensional features.

Clearly, the matching method depends on the dimension of the quantity "X". In principle, matching "X" in the left and right images is only a variant of the classical pattern recognition

problem. Consequently, much theory and practical experience is already available. In addition there are other methods such as correlation and even straight-forward "and-ing" (a variant of the logical "and" operator).

If there are enough pixel level (zero-dimensional) features available such that each pixel in the left and right image can be uniquely "paired" then the pairing procedure is the same as the "minimum distance classifier". Clearly, a high number of pixel features is required and the classifier has to include the spatial position of the pixel as well as a condition for resampling to achieve "unique pairing". The last two conditions (resampling and unique pairing) are not part of the classical procedure but the required modifications are obvious. Since the size of the decision space will be hugh and exceeded available computational facilities, this method could not be experimented with. However, it is very likely that biological vision systems use some variant of this approach (1.6).

A possible modification of direct pixel feature matching is to combine adjacent pixels in pairs, in triplets, in quadruplets, etc., to create pixel feature combinations. Given enough pixels these n-tuplets will become unique and the matching problem is back in the classical domain. This may also be viewed as another method of "texture detection and matching". This approach was not tried.

Since direct pixel feature matching was not feasible and n-tuplets were not used, the approach was modified by grouping the pixel features into "larger units" or groups. The unique pairing will now be between the groups. Pixel feature grouping is standard practice in classical pattern recognition. The details may be found in Chapters 2 and 3. Grouping of the groups into structures was not tried due to lack of time.

1.4.3 Comments

An hierarchical procedure for matching the left and right images of a stereo pair was suggested mainly due to computational feasibility but also since the classical pattern recognition techniques are directly applicable. The computation of the distances (reconstruction) requires pixel level matching.

If direct pixel matches are available then one could proceed directly to 3D reconstruction and create a "dense" or complete distance (range) image. Solutions are directly available from photogrammetry or, if only a "pin-hole camera" model is used, then the derivation of the necessary equations is only a matter of elementary 3D geometry.

If the matching is only for one-dimensional features (edges), then the points on the edges (pixels) will still have to be matched. Only in case of straight line segments is it possible to interpolate after the line ends are matched, since straight lines in the scene project to straight lines in the images. However, except for the edges, the reconstructed image is "very sparse" i.e., a "vacuum" with only strings of pixels present.

If the matching is on a region level (two-dimensional feature matches) then it is necessary to first match the edges of the regions and then the edges on the pixel level. The interior of the region is simply "another smaller version" of an image and is processed as was done for the whole image initially (iterated hierarchical procedure).

1.5 Validity requirements

The "similarity principle" requires the two images of the stereo pair to be "similar". Of course, the images must differ since they represent the scene "seen" from two slightly different positions. This is a fact geometry and the difference (parallax) must exist for the "stereo effect" to be present. However, it may be easier to comprehend and to develop a "feeling" for the stereo problem if we restart from beginning.

1. Sketch the design of a stereo camera system that best suits image processing and subsequent analysis requirements. Elementary optics, some understanding of image processing, the fundamentals of control systems, knowledge of hardware capabilites, and so on, are sufficient basis for a conceptual design.

2. Assuming that the ideal camera system exists as designed in (1), what are the irreducible problems that still remain, being characteristics of the stereo problem rather than artifacts introduced by equipment or by the conditions under which the stereo image pair was obtained.

1.5.1 The ideal stereo camera

It may be easiest to see the whole problem at a "glance" if we "design" the camera system and its controls in order to satisfy the similarity principle rather than accept an existing design and try to adapt the principle to it. A simple "pinhole" camera model is assumed since there is no need to complicate the situation and this is the "usual" camera model assumed in many articles that deal with the stereo problem.

A "pinhole" camera model is shown in Figure 1.5.1-1. The camera is indicated by a sphere called "eye". It has a pinhole for a "lens" and the image is formed at the back of the sphere indicated by "image". A line that goes through the pinhole and the centre of the sphere will be called the "optic axis". Let the origin of a coordinate system be fixed at the pinhole. However, at the moment we are only concerned with the controls we expect to have in a properly designed system.

In order to have sufficient control over the image, we must be able to "swing" the camera horizontally, vertically, rotate it, zoom it, and adjust the "iris" of the pinhole, all under computer control. In terms of movie maker's or aeronautical engineering jargon, we need controls for yaw (pan), pitch (tilt), roll, zoom (focus), and image intensity (gray level). This allows us to point the camera anywhere in the scene and to focus or zoom in on it and to obtain a properly resolved light intensity image. Control of "exposure" is standard practice in cameras but it is better to control it via programs. Automatic focussing is avail-

able but it is only a feedback loop in the focus system that maximizes the sum of the gradient magnitudes over the image or the distance is measured acoustically. The place where to point the optic axis and the zoom (magnification of image) are defined when two cameras are adjusted to satisfy the "similarity principle".

A two-camera system is shown in Figure 1.5.1-2. One of the cameras ("eyes") will be called "left" or L and the other "eye" is called "right" or R. Both cameras have independent controls over yaw, pitch, roll, zoom, and iris. It is now simple to see how to control the cameras to satisfy the similarity principle and to simplify the image processing problems as much as possible.

1. Swing the cameras such that the optic axes intersect at a point called the "fixation point".

2. Zoom the cameras such that the images in the left and right "eyes" are of the same size (have the same scale).

3. Rotate the optic axes such that the images "line up" in the two "eyes" on or near the fixation point.

4. Control the irises such that the images have the same intensities.

If a camera system of this kind and its "drives" are properly "wired" to a computer system then the programming required to achieve the above control structure appears "straight-forward" and no "scene understanding" (and all that) is needed. However, in the absence of such a system to experiment with, these remain conjectures only. Note that the human eye "musculature" (control mechanism) has all these controls and they operate automatically. A simple way to "see" what happens to our stereo vision is to move one of the eye balls (push it slightly with a finger) while looking at a scene.

In photogrammetry a certain amount of control over the above variables (items 1, 2, 3, 4) and others is available but they "take the images as they are" (including lens defects, etc.) and apply appropriate corrections to compute the distances. They need the precision and, consequently, the formulas become rather complicated. Even the "famous" Pl = A*Pr + T formula makes no use of the similarities found in the stereo pair. In Pl = A*Pr + T, Pl and Pr are the two corresponding points, one in the left and the other in the right image, A is a general rotation matrix, and T is a translation vector.

These transformations become very simple if the coordinate systems are chosen as shown in Figures 1.5.1-3 and 1.5.1-4. Such a set of coordinates may be called "ego-centric" since they only refer to the observer. In essence, the x-axis goes through the two pin-hole lenses, the y-axis is halfway between the two pin-

holes, and the z-axis goes through the fixation point. The "right-hand" rule is preserved if the directions of the axes are chosen as shown. The eye or camera coordinates are chosen at the pin-hole lenses, see Figure 1.5.1-4. The yl and yr axes of the left and right eyes are set parallel to the ego-centric y-axis and, of course, the zl and zr axes intersect at the fixation point. If we also assume a set of yaw, pitch, and roll controls for the "head" that carries the two eyes (the "third imaginary eye controls" at the origin of the ego-centric system, which keep the "nose" always pointed in the direction of the fixation point), then the eyes only need to be rotated ("panned") by the same amount (in opposite directions) to "fixate" any point in the scene. The image plane coordinates are chosen as xl´, yl´, zl´ for the left, and xr´, yr´, zr´ for the right eye, where xl´||xl, yl´||yl, zl´||zl, and, xr´||xr, yr´||yr, zr´||zr, with || indicating "is parallel to".

Under the above assumptions, a point P in the scene and its corresponding image points Pl and Pr in the left and right images (which are found by solving the correspondence problem), give:

El = (xl, yl, zl) = (0, 0, -xo) = left pin-hole
Er = (xr, yr, zr) = (0, 0, +xo) = right pin-hole

Vl = (-xpl´, -ypl´, -d) = vector from Pl in left eye coordinates.
Vr = (-xpr´, -ypr´, -d) = vector from Pl in right eye coordinates.

In the ego-centric coordinates

Vle = Al*Vl + Bl = (xl, yl, zl)
Vre = Ar*Vr + Br = (xr, yr, zr)

$$Al = \begin{pmatrix} \cos(-T) & 0 & -\sin(-T) \\ 0 & 1 & 0 \\ -\sin(-T) & 0 & \cos(-T) \end{pmatrix}$$

$$Ar = \begin{pmatrix} \cos(+T) & 0 & -\sin(+T) \\ 0 & 1 & 0 \\ -\sin(+T) & 0 & \cos(+T) \end{pmatrix}$$

$$Bl = \begin{pmatrix} -xo \\ 0 \\ 0 \end{pmatrix}$$

$$Bl = \begin{pmatrix} +xo \\ 0 \\ 0 \end{pmatrix}$$

The unit vectors Vleu and Vreu (direction cosines) of Vle and Vre are

Vleu = Vle/|Vle| = (Vlx, Vly, Vlz)
Vreu = Vre/|Vre| = (Vrx, Vry, Vrz)

Approximately, the distances Dl and Dr from the left and right eyes form a triangle with a base of 2xo. Let the coordinates of the point P in the scene be (x,y,z) then

x = Vlx*Dl + xl = Vrx*Dr + xr
y = Vly*Dl + yl = Vry*Dr + yr
z = Vlz*Dl + zl = Vrz*Dr + zr

which can be solved for Dl and Dr, and (x,y,z), but the best numerical values should be selected during computation.

If these rays defined by Vleu and Vreu intersect (at the fixation point) then let (x implies cross product)

Nu = Vleu x Vreu = (Nux, Nuy, Nuz)

which is the normal of a plane in which the vector Vleu lies and which is parallel to the Vreu vector. The minimum distance Q between the rays Vleu and Vreu is

Q = Nux*xr + Nuy*yr + Nuz*zr + d'
0 = Nux*xl + Nuy*yl + Nuz*zl + d'

The "squint" Q is a line which has the same direction as Nu and may be used to compute a minor correction to the distance z.

The above brief derivation is only intended to show that with a suitable selection of coordinates, and if there is enough control over the cameras, then the computations can be rather simple (using the shortest distance between two lines in space). The above computations have not been checked experimentally for their numerical stability. If the eyes are independently mobile, i.e., the z-axis does not go through the fixation point, then the computations are slightly more complex but the similarity principle remains the same.

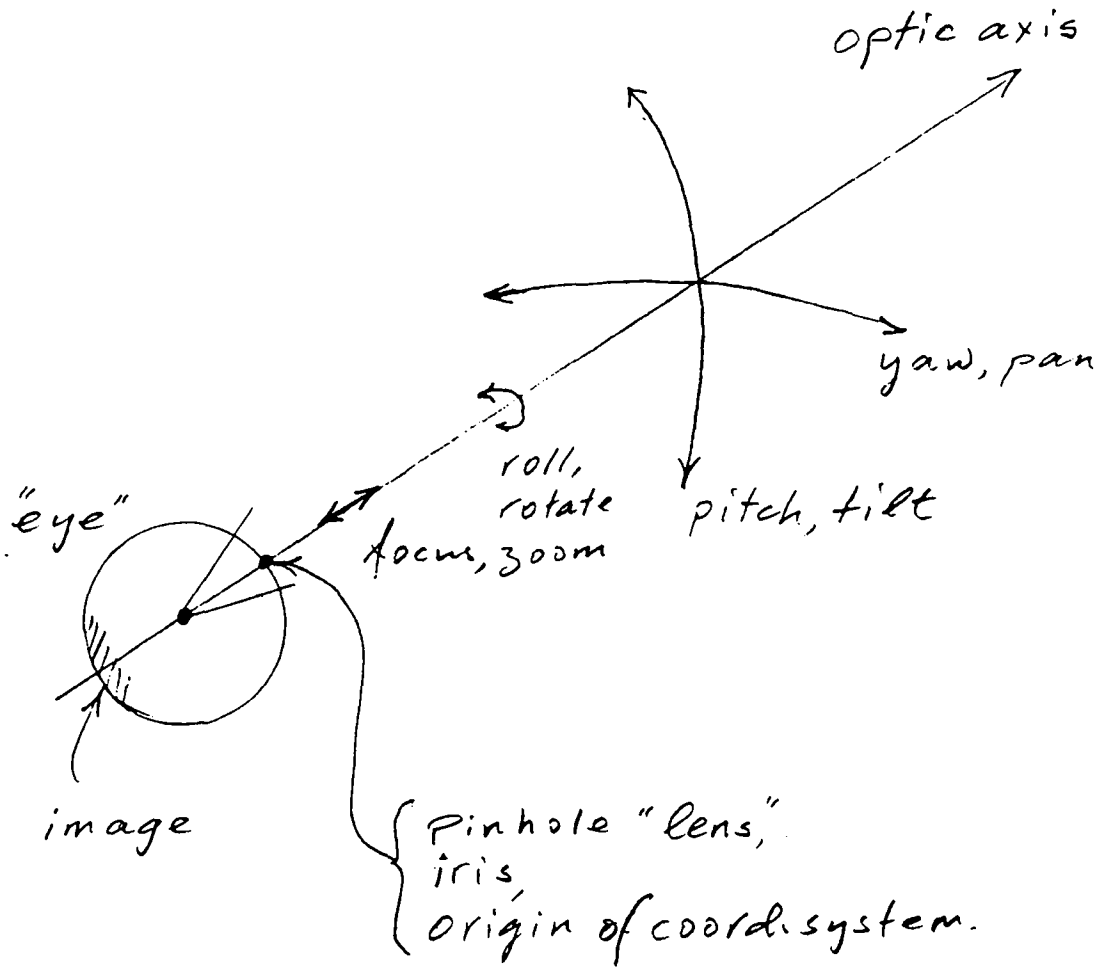


Figure 1.5.1-1: The pin-hole camera model.

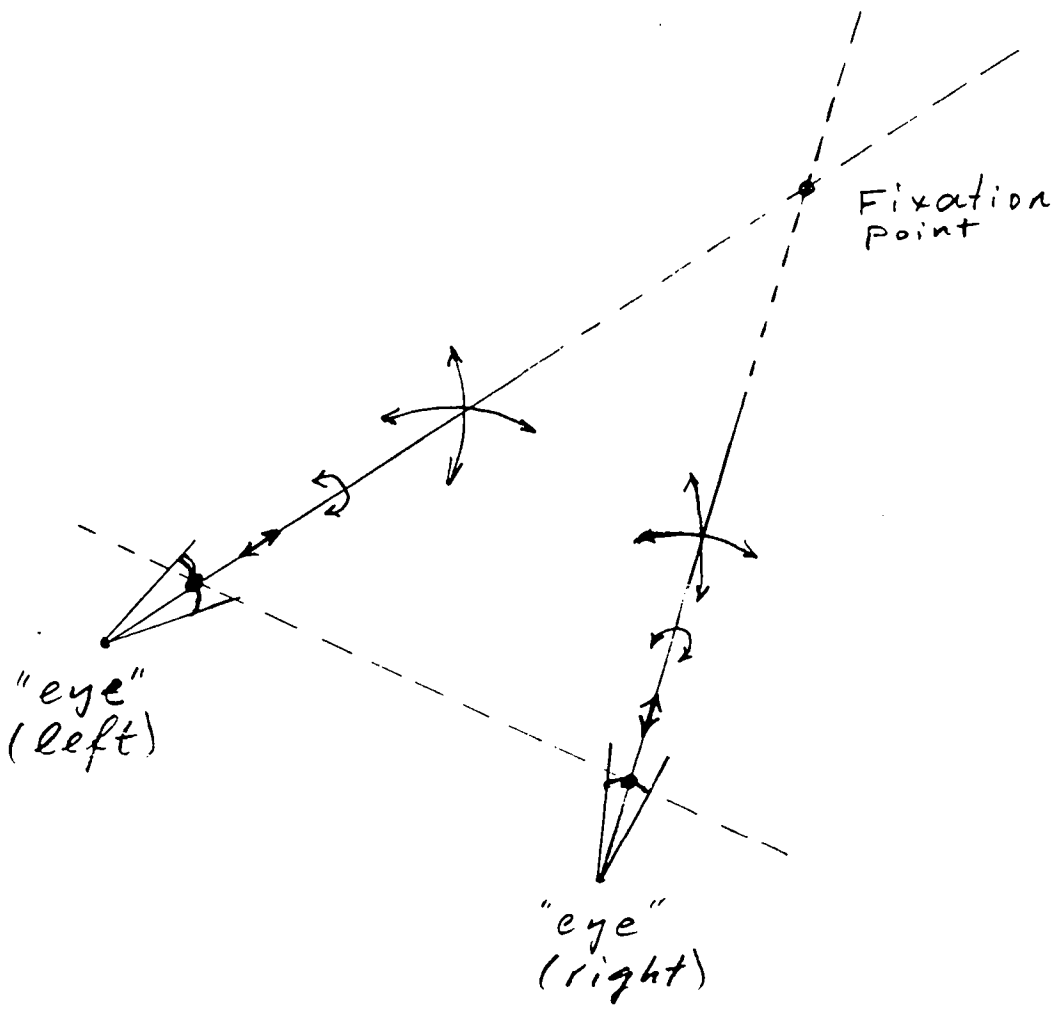
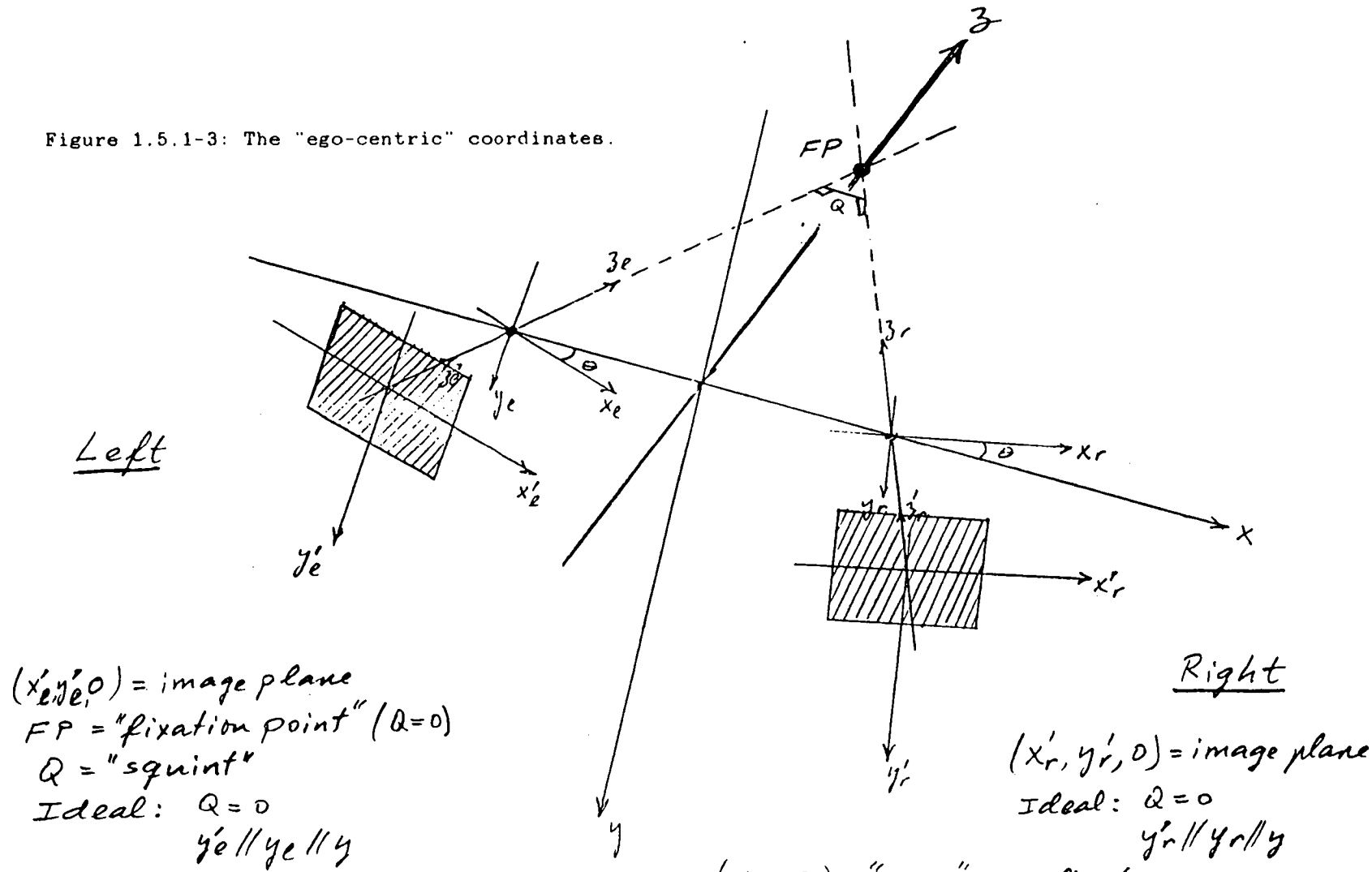


Figure 1.5.1-2: Two pin-hole cameras.

Figure 1.5.1-3: The "ego-centric" coordinates.



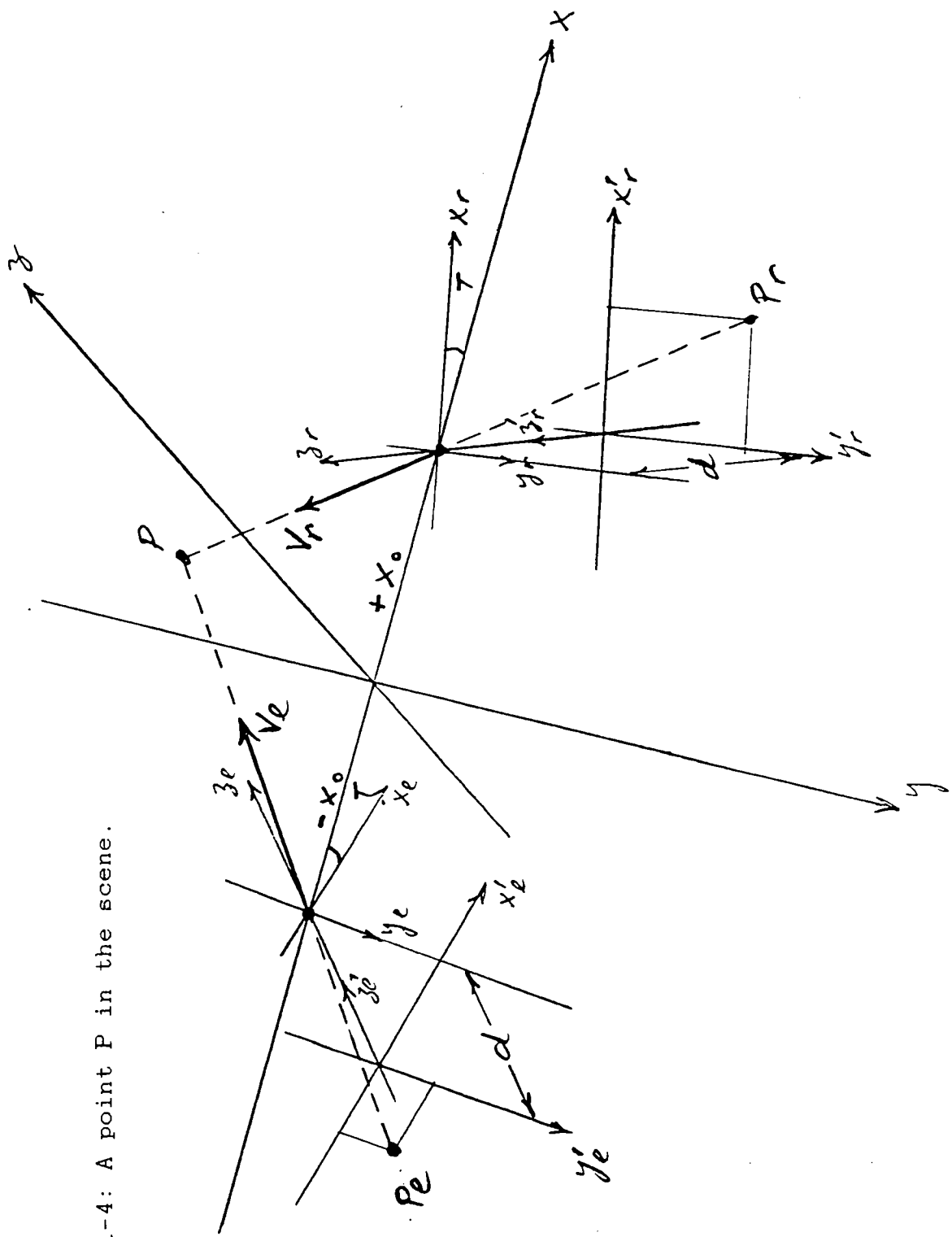


Figure 1.5.1-4: A point P in the scene.

1.5.2 The irreducibles

An object in the 3D space can be anywhere (in any position in the space). An object can have any shape, size, rotation, and orientation. The objects need not have any particular predefined shapes, they may move, rotate, split, merge, and change shape. Objects can be semi-transparent, can obscure each other, and so on. In addition to all this, the illumination adds another set of "variations". Thus, in general, a scene can be very complex, but all these variations do not seriously affect the similarity principle, provided that some elementary conditions are satisfied:

1. The stereo pair is taken at high enough speed and proper focus so that motion blur and defocussing are not noticeable in the images at the resolution required for analysis.

2. The two images have been taken simultaneously or with a single camera of a stationary scene.

3. The cameras are "sufficiently" close such that the disparity (parallax) is not excessive. It should be remembered that there is both a practical and a theoretical limit to the parallax, i.e., one can make the problem ridiculous - point one camera east and the other west and try to reconstruct the 3D scene!

4. The fixation point is on a surface with texture or on an edge and not in "free space".

5. The cameras are "properly aligned" such that:

i) The optic centers intersect at the so-called "fixation point".
ii) The (TV camera) scan lines are in the same plane.
iii) Both cameras have the same light sensitivity (iris setting).
iv) Both cameras have the same focus setting.

None of the above conditions are unrealistic. The additional constraint of "still-life" (nature morte) implies that only one pair of images is available, being one "frozen instant" from a live system. It may, however, be interesting to observe some differences between the "still-life" and a "live" stereo system:

a) In a "live" vision (active) system there is a difference in the analysis if both eyes and head move together or when eye motion and head motion are independent.

b) In "live" vision there are problems of "how to": Focus the eyes. Line up the images. Position the eyes. Rotate the eyes, etc. There is also the question of "where to look", how to integrate the various views, and so on.

c) Calibration either has to be carried out at every change of fixation point, or the calibration problem is an artefact dictated by the mathematical method used.

d) The fixation point can be placed anywhere in the scene. This is an exceedingly important aspect of stereo vision, at least from the processing point of view since, at the fixation point, the parallax is zero. Thus, a "live" system can put any point, edge, or region into correspondence by simply "fixating" on it.

e) "Still-life" excludes any possibilities of using motion, flicker, exploration (change of fixation point), etc., for helping to establish correspondence. In a "gray still-life" the images only contain gray levels (colours are excluded).

The "still-life" case thus avoids certain problems that a "live" would have to solve but does this complicate or simplify the stereo problem? If the additional computations that the "live" system has to carry out are not considered (they are processes that can be carried out in parallel) the stereo problem tends to "vanish" since one can "look at" any point in the scene that causes problems with stereo correspondence. The "still-life" system does not have this privilege thereby unnecessarily complicating the stereo problem.

However, given a "still-life" pair where the fixation point is in the middle of the images and on a surface, what types of variations or problems can one expect? Most of these are easily understood with the help of a few simple sketches:

1. The absurd case: Even in a "properly taken" stereo pair the two images can be totally different. Put a two-sided thin mirror edge-wise between the eyes and look, see Figure 1.5.2-1. This may also be a thin cardboard with different pictures on either side.

2. The absurd but normal case: Even in a "properly taken" stereo pair parts of the two images can be totally different. Two normal situations are sketched in Figures 1.5.2-2a and 1.5.2-2b. In the first case (Fig. 1.5.2-2a) one has focussed onto a hole or maybe a "fixation target" but the scenes seen through the hole can be different. In the other case (Fig. 1.5.2-2b), which is fairly usual, one eye can see much more than the other of a part of the scene but, of course, this part could also be totally different.

3. A point P in the scene which appears fixed in one image can be anywhere (on a line) in the other image, or not even visible in one, see Figure 1.5.2-3. The trajectory of P may be a "stick" seen "end on" in the other image. If the trajectory is a plane surface (thin) then it appears as a stick in one image and as a surface in the other.

4. The two pin-hole lenses for the left and right eyes are two points Pl and Pr in the ego-centric or a "world" coordinate system, see Figure 1.5.2-4a. Given any point Po in the

scene, these three points (Po,Pl,Pr) define a plane P in the space. The left and right image planes Il and Ir are also planar surfaces. Two plane surfaces that cross each other form a straight line S at the intersection. If the intersection of P and Il is called Sl and the intersection of P and Ir is called Sr, then the two lines Sl and Sr are called the epipolar lines. Only if Po coincides with the fixation point Fp then Sl and Sr are in the same plane, otherwise these two lines are "tilted" with respect to each other, see Figure 1.5.2-4b. Clearly, if a point Px is constrained to move in the plane P, its images are constrained to the lines Sl and Sr.

These simple sketches may be sufficient to indicate how to "see" many of the situations that can be encountered. For example, if there are two indistinguishable points Px and Px in each image on Sl and Sr, which of these are to be put into correspondence to obtain the correct reconstruction in the 3D space (cannot decide); do the centers of gravity of the facets in the image space correspond to the centre of gravity of the facet in the 3D scene (no), and so on.

1.6 Conclusions

The brief description in this chapter has, hopefully, put the 3D stereo problem into "perspective". It is also hoped that some of the "mystery" has been removed.

Adequately designed stereo camera systems are available from firms specializing in photogrammetry. However, the problems they address are different from those required in "live" robot vision systems. For the study of still-life (nature morte) stereo a pair of adequately digitized images is sufficient in principle. However, as already mentioned, the solutions to such problems are of more interest to photogrammetry than to the designers of computer vision systems for mobile robots.

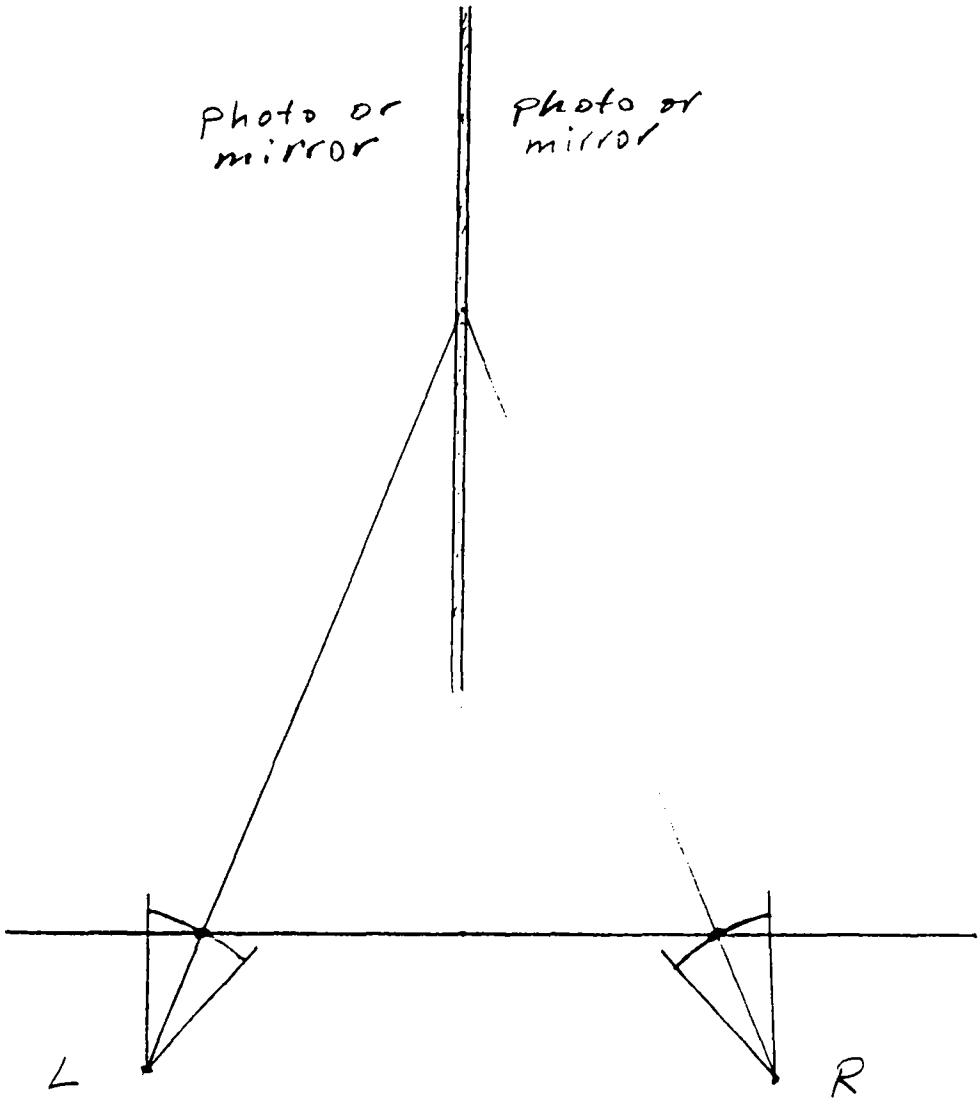


Figure 1.5.2-1: The absurd case.

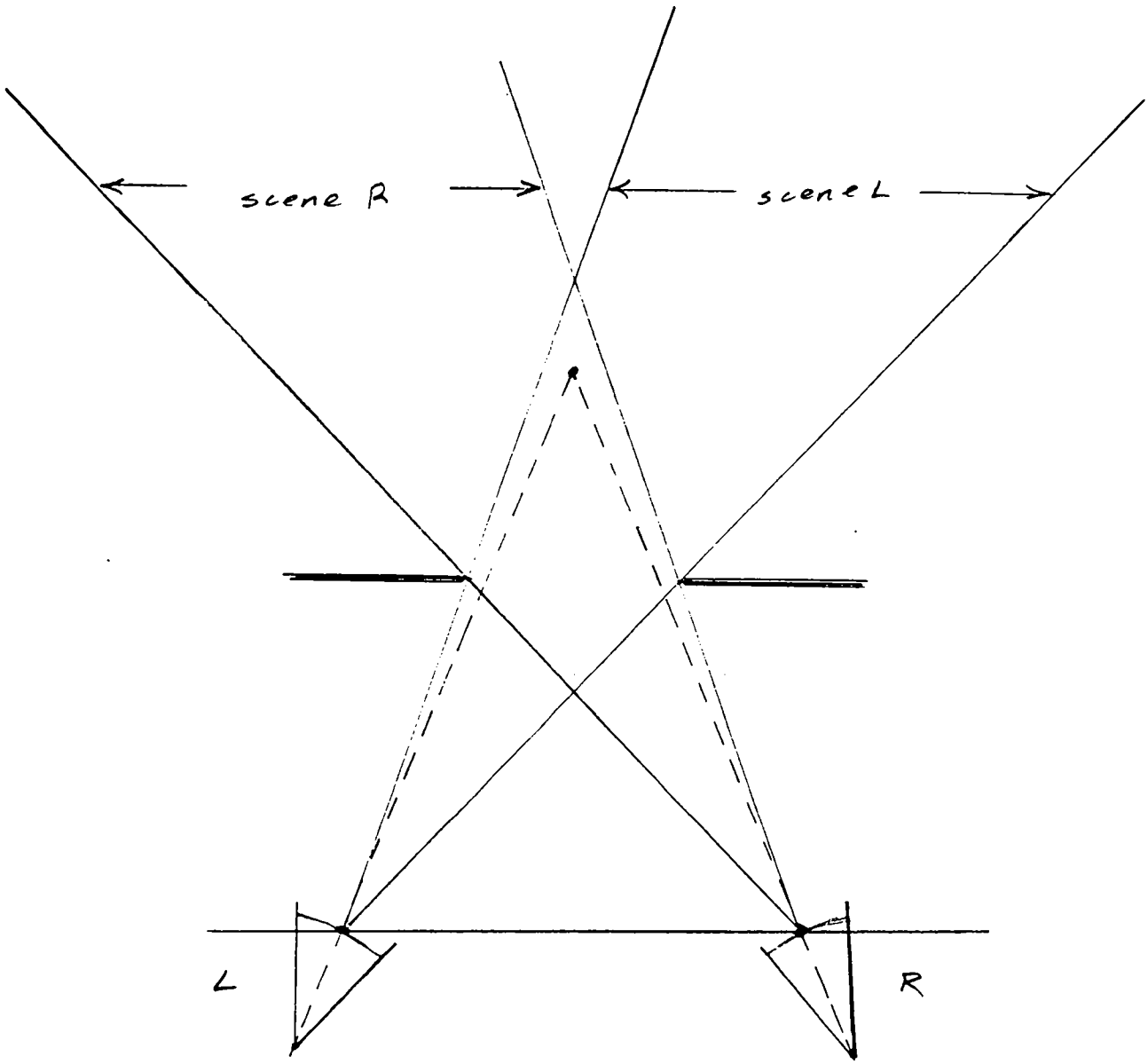


Figure 1.5.2-2a: The absurd but normal case.

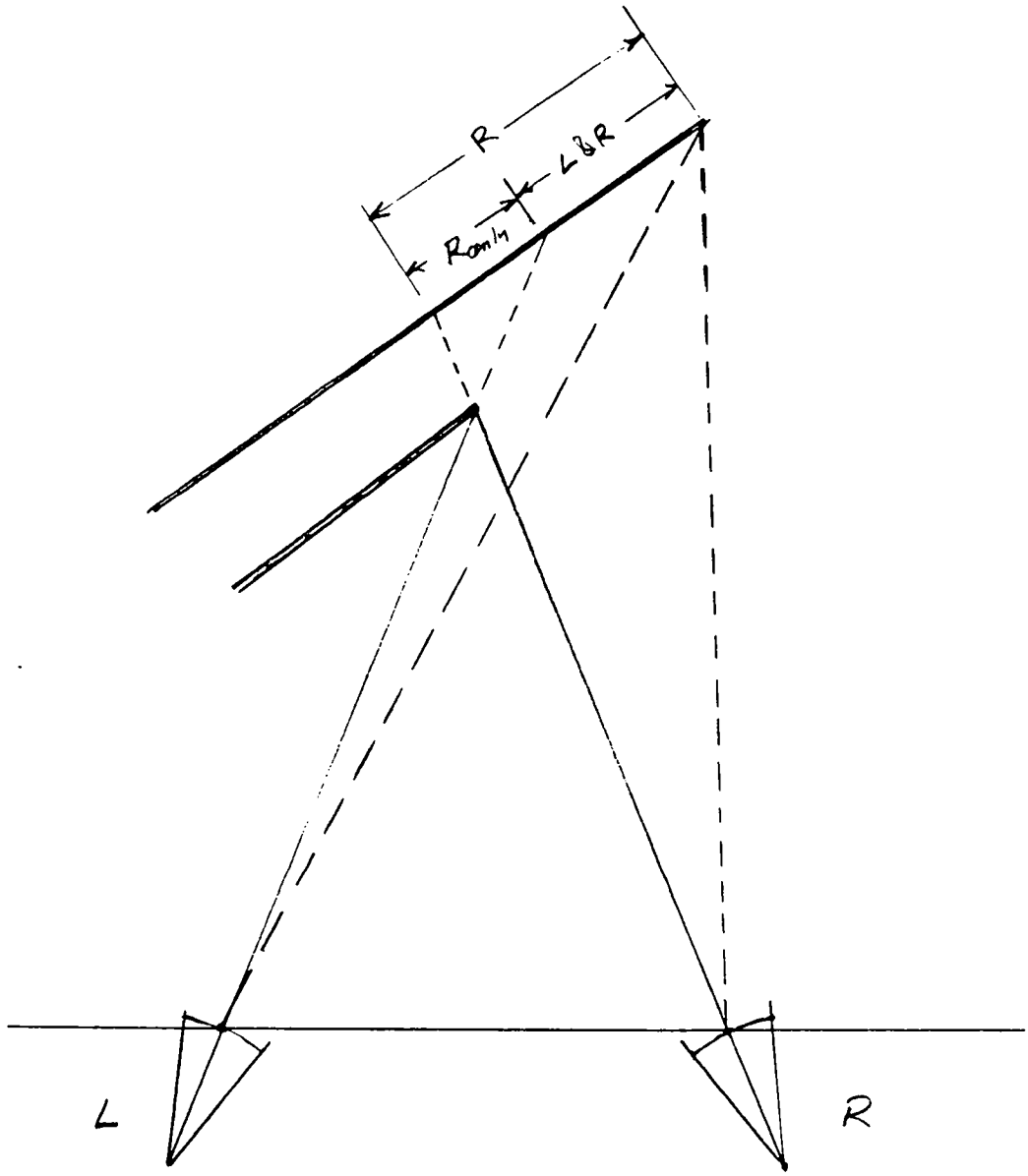


Figure 1.5.2-2b: The usual case.

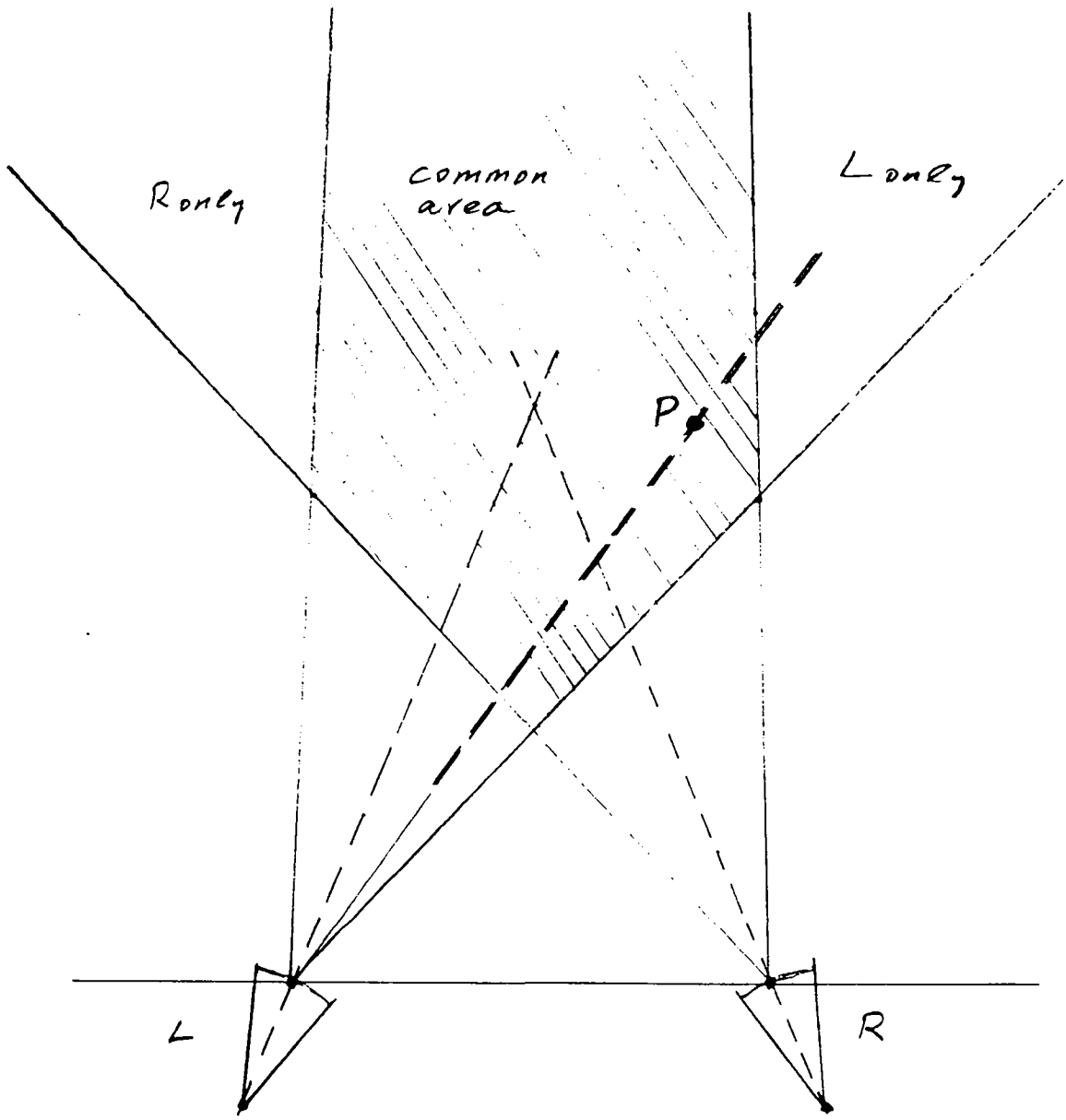


Figure 1.5.2-3: A point P may be anywhere on the epipolar line in one image while it is stationary in the other image.

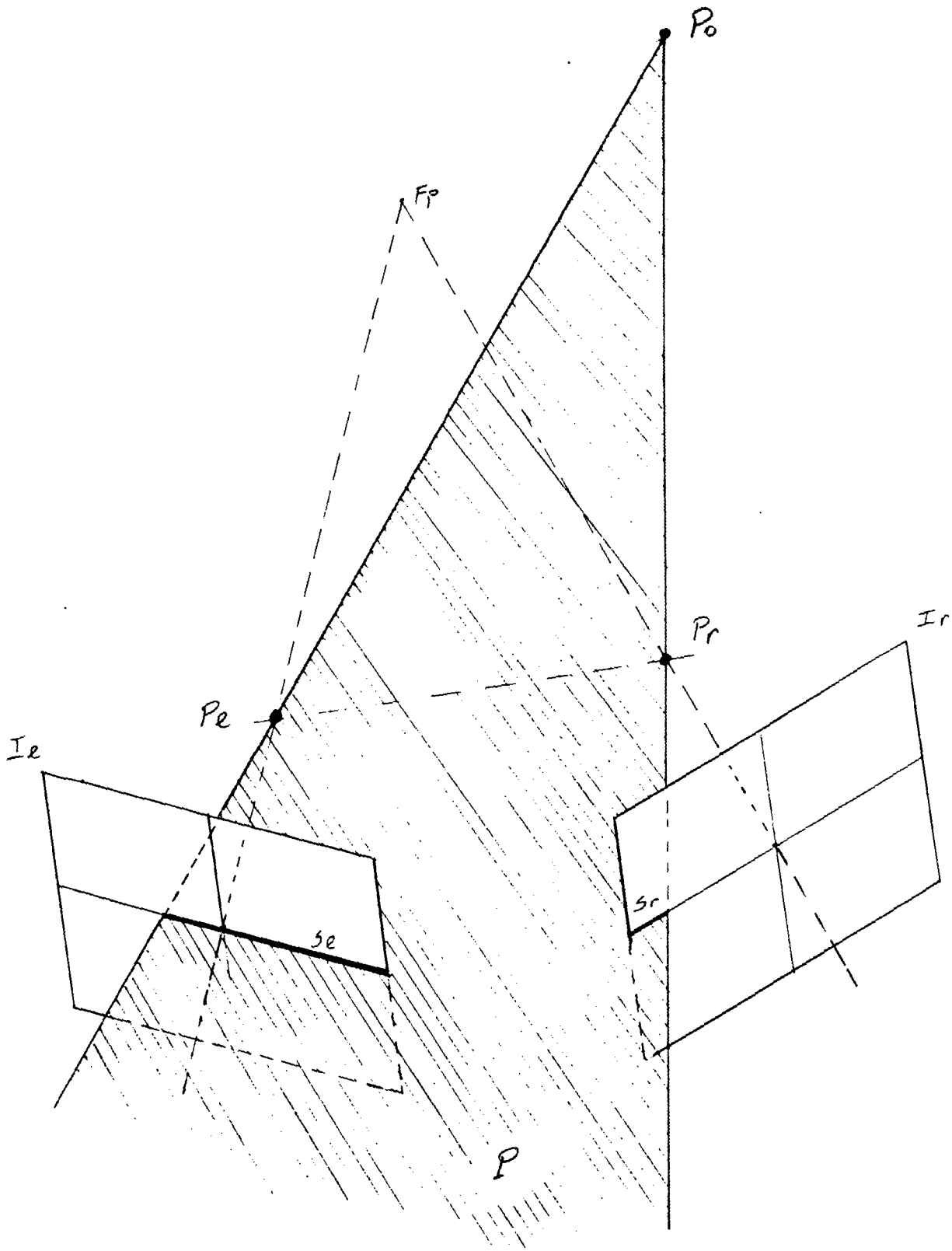


Figure 1.5.2-4a: The epipolar line.

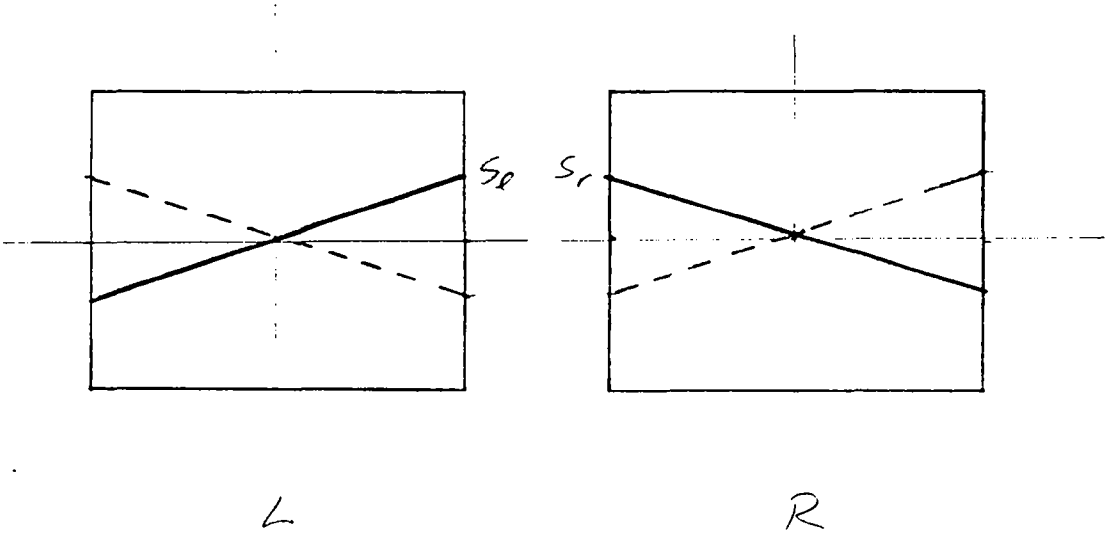


Figure 1.5.2-4b: The epipolar line.

Chapter 2: PRELIMINARY COMPUTATIONS

2.1 Introduction

In order to develop a "feeling" for the actual difficulties with a 3D stereo problem and to "interface" with the existing work at project SYNTIM, the "INRIA canonical image pair #1" (I001g_o and I001d_o) was chosen for experimental studies. However, due to the relatively short visit to INRIA, a rather large personal collection of image processing programs, and the differences in the programming languages and operating systems (Fortran, VMS, and DOS versus C, Unix, and X-windows), the experimental environment consisted of a compromise between private computer equipment brought from Canada and the facilities available at INRIA. In essence, the compromise consisted of the "computational structure" sketched in Figure 2.1-1.

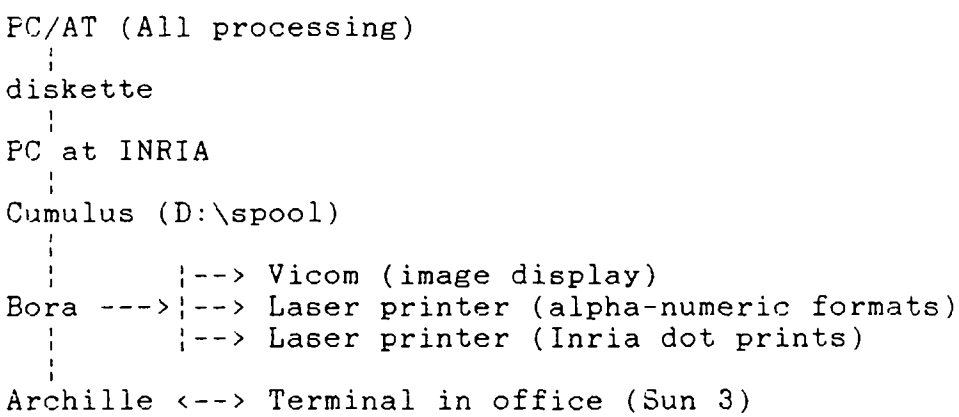


Fig. 2.1-1: The computational structure used for the studies.

The INRIA image pair is shown in Figure 2.1-2 (and called I001.log and I001rog, respectively). These two images appear to form a stereo pair and the depth is visible if the images are looked at through a stereo viewer. Consequently, at least in principle, this image pair should also be adequate for computer vision. The images are arranged side-by-side in Figure 2.1-3a for viewing (using a "Peak 2x viewer", (2.1)) to allow the reader to verify that visual integration of these images presents no difficulties for our vision. However, it was noticed during the experiments that one's stereo vision varies, whatever the rea-

sons. On "good days" the images generated clear impressions of depth, while on the "bad days" stereo integration proved problematic, specially on processed images and even if the originals were relatively far apart. The author is not one of the approximately 11% of the population who, apparently, cannot see "pure stereo" (Julez'es dot stereograms).

In brief, the computational environment was the following:

1. Private PC/AT computer using the DOS operating system. This rather small system consists of 512 Kbytes of memory, three hard disks of 32 Mbytes each, a 1.2 Mbyte "floppy", a printer, and a terminal. On this system approximately 1000 previously developed image processing programs (in Fortran) were available from earlier studies of various aspects of image analysis. The computer system was familiar and available "day or night". Consequently, it was preferred over the large systems at INRIA. Initially a telephone interconnection between these systems was considered but these efforts were abandoned due to a variety of reasons and the programming and experimentation was carried out mainly during evenings and week-ends. This compromise placed constraints on some aspects of the research but it represented the "path of least resistance" for quickest results.

2. All the image analysis was carried out on the PC/AT. In order to fit the rather small memory (512 Kbytes) which was accessible to the operating system, the images were reduced to 128x128 bytes by selecting every odd row and every odd pixel on a row. To accommodate the existing "defaults" in the PC, the last row was eliminated, leaving images of 128 by 127 pixels stored in four bytes per pixel. The "integer*4" format is used throughout the analysis. Gray level resolution is preserved by multiplying all applicable data by 100. This stereo pair is shown in Fig. 2.1-3b and a "blown up" version is in Fig. 2.1-4. (These are called "if8fort11.gry" and "if8fort12.gry" in the INRIA Unix system). An 8-level "alphabetic gray" image pair, which can be printed on any alpha-numeric device, is shown in Fig. 2.1-5. Image quality has suffered due to decimation but sufficient remains for processing. Furthermore, since the world is "fractile", it does not matter at what scale the image is digitized, there is always a loss of detail. It was initially intended to develop the programs on the PC/AT using the reduced (decimated) images and then transfer the programs to INRIA facilities to be run at "full resolution". This would have offered a check of the effect of scale on the processing strategies. Unfortunately, the transfer never took place.

3. The "communication" between the PC/AT and the INRIA systems occurred via diskettes using formatted data. In brief, the whole experimental "structure" was as shown in Fig. 2.1-1.

Some difficulties were experienced in obtaining good quality prints rapidly in order to verify the results visually as soon as they were obtained. The program used for printing images on the laser printer, "im_laser -T -s K -inv file_name", had size options from "K=1" to "K=6". Size 6 prints the entire image, while size 5 drops some lines from the image and size 3 drops more lines. The exact cause has not been investigated. A gray level chart was constructed to test the laser printer. The chart has 128, 64, 32, 16, 8, 7, 6, 5, 4, 3, gray levels. The number of distinguishable gray levels is only 4 to 5 on the originals and 4 or less on copies. No texture printing programs existed to bypass this problem. A "4-colour" map painting program was written to try to accommodate the printer but in many cases more than 4 "colours" were needed due to "unknown" regions in the image that required the same colour. (The problem is equivalent to trying to paint the map of the Eurasian continent in 4 colours under the constraint that the seas around it and the lakes within have to be in blue). The size "5" print of the chart is shown in Figure 2.1-6a and two of them fit on one page. Only the size "3" print shown in Figure 2.1-6b could be viewed with the available stereo equipment.

Many of the results are shown as one pixel per character alphabetic prints of images of facets labels (modulu 26). Initially these prints may appear to the reader as "alphabet soup". However, for those who like to see details, these prints are far more informative than gray level or even false colour prints. The texture effects created by groups of similar letters (regions or facets) may suffice for an overall impression. In the vicinity of line 56 these prints may show a "discontinuity" since they were assembled manually from two sheets (X11, xprint, small size).

Figure titles:

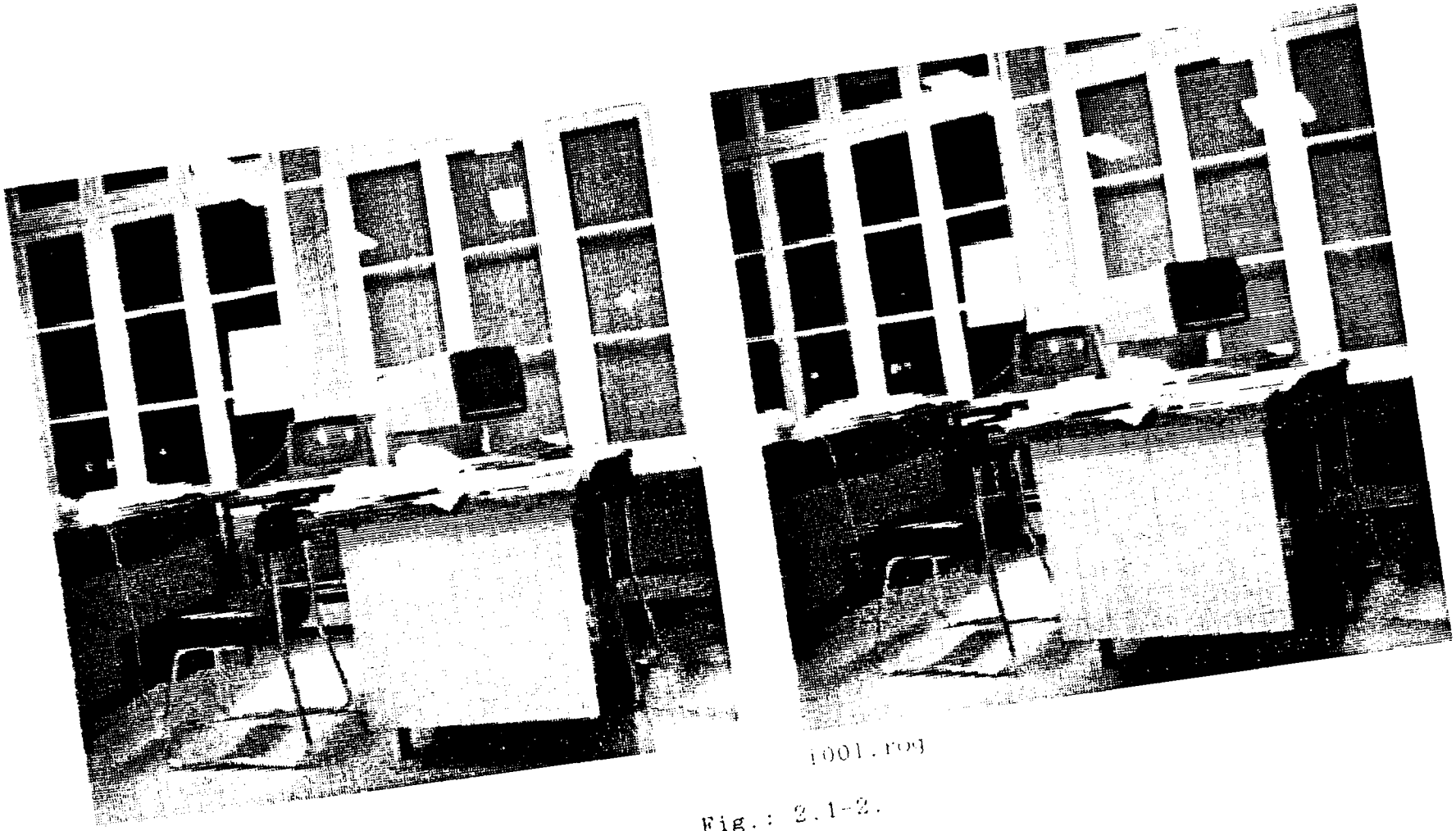
Fig. 2.1-2: The "INRIA canonical image pair #1" printed in relatively large size to show detail. I001g_o is the left image and I001d_o is the right image. The digital image matrix size is 256 by 256 bytes. The gray level range = 0 to 255.

Fig. 2.1-3a and -3b: The "INRIA canonical image pair #1" arranged for stereo viewing (using a "Peak 2x viewer"), and renamed as "i001.log" and "i001rog". (-3a) is the original image while (-3b) shows the decimated image used in computing.

Fig. 2.1-4: The reduced image pair shown in larger size for illustration of the loss of detail.

Fig. 2.1-5: The reduced image pair printed in eight gray levels using alphabetic symbols. (Files: Inr021.fig and Inr022.fig)

Fig. 2.1-6a and -6b: The gray level chart in size 5 (-6a) and size 3 (-6b).



i001.log

i001.rog

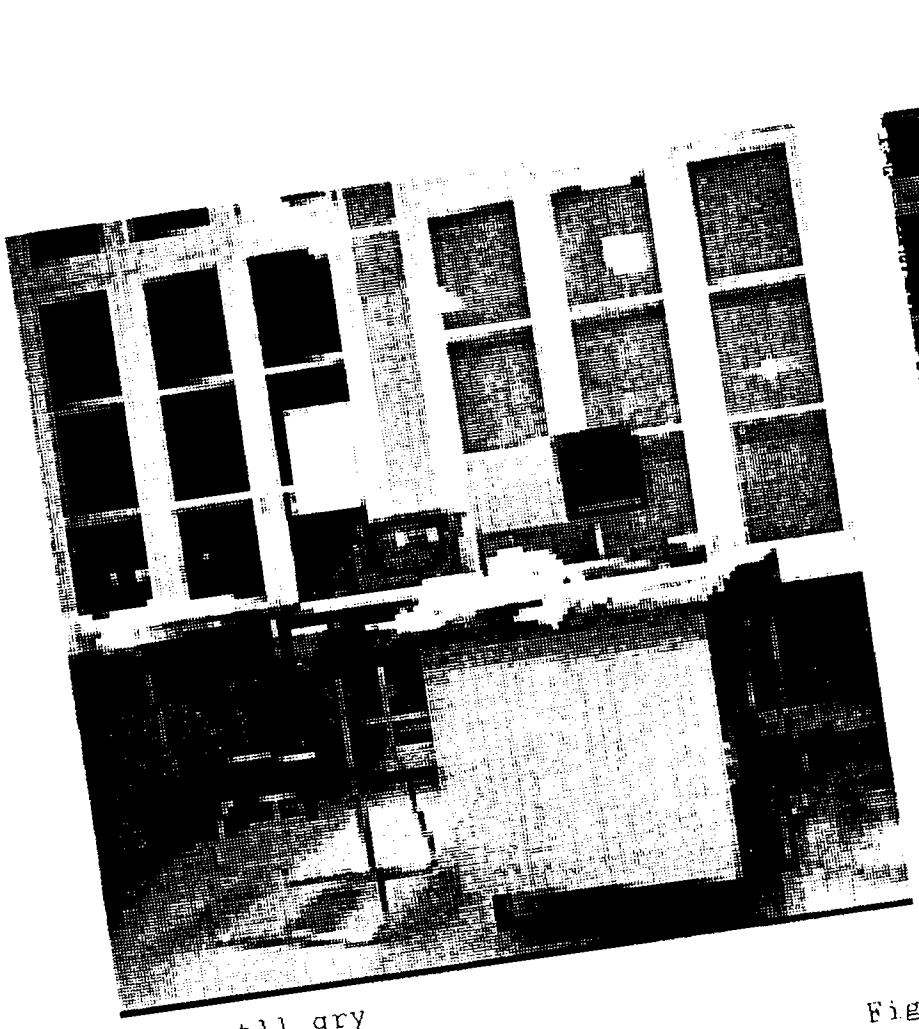
Fig.: 2.1-2.

i001.log i001.rog

Fig.: 2.1-3a.

if8fort11.gry if8fort12.gry

Fig.: 2.1-3b.



if8fort11.gry

if8fort12.gry

Fig.: 2.1-4.

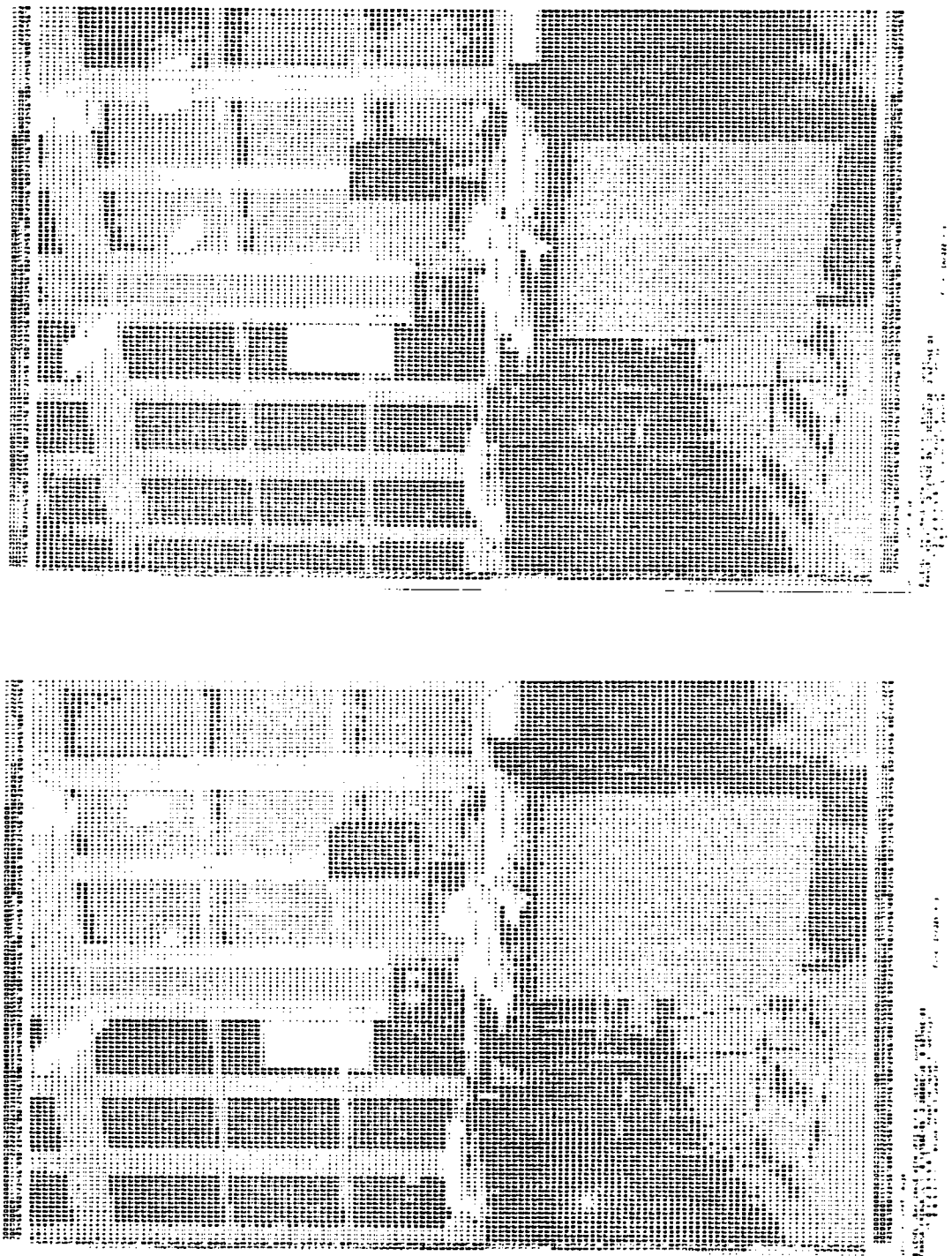
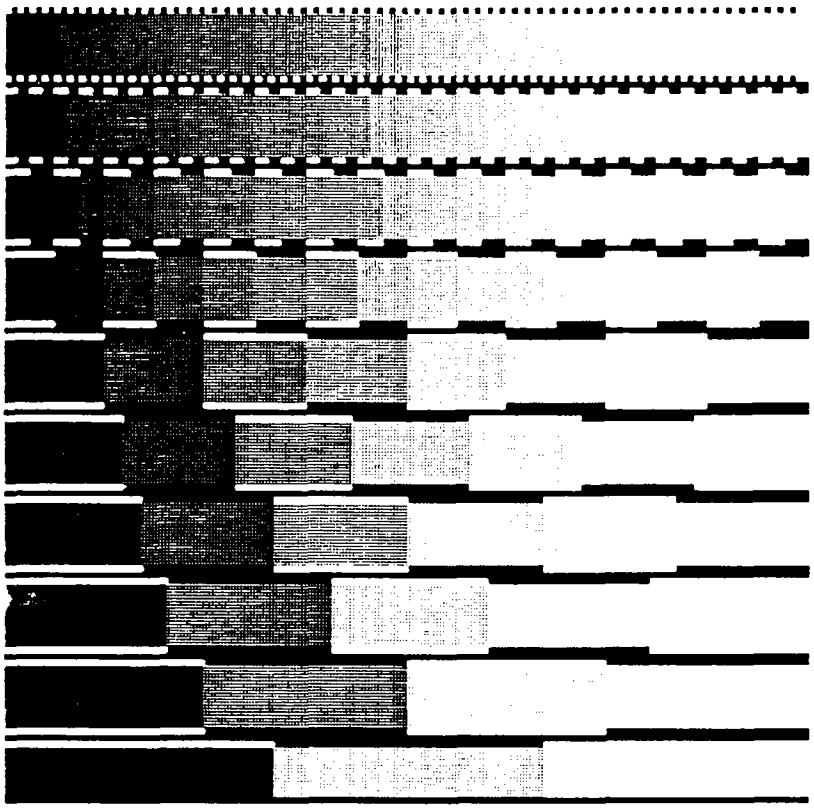
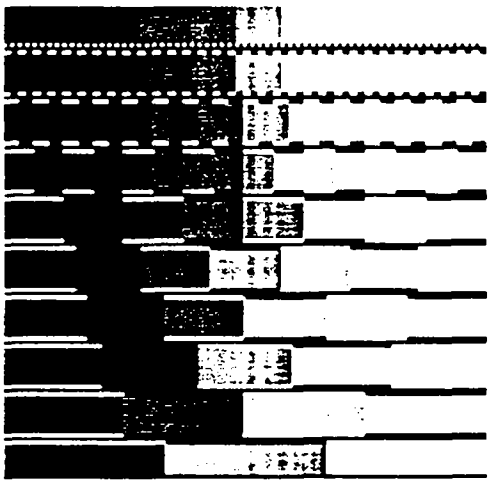


Fig.: 2.1-5.



if8gray.bar

Fig.: 2.1-6a.



if8gray.bar

Fig.: 2.1-6b.

2.2 The image pair

Even though this image pair (I001g&d_o) is adequate for human vision and the "room scene" appears in its correct shape when the images are viewed through stereo glasses, this pair contains numerous aberrations. In brief, the difficulties that became apparent during the analysis were of the following type, where I001g_o (or its decimated version) will be referred to as "left image", "left", or "L", and I001d_o (or its decimated version) will be referred to as "right image", "right", or "R":

1. The gray levels in the two images are rather different. Histograms of the gray levels normalized to a histogram peak of 100 units are shown in Fig. 2.2-1. Normalization to the peak in an histogram tends to "over-dramatize" the difference between the histograms. Non-parametric statistical tests (for ex., Kolmogorov-Smirnov) were considered unnecessary.

2. Closer inspection of the images reveals that, for example, the lower right corner of L contains details which are not visible in the lower right corner of R. A loss of information has occurred in one image with respect to the other. Some of these losses cannot be recovered by filtering.

3. These two 256x256 images appear to be "extracts" from a larger pair of images (>256x256 bytes). The shift between L and R produced in this extraction is unknown and the image centre cannot be relied upon as the optical centre of the two cameras.

4. There does not appear to exist a "fixation point" where the optical axes of the two cameras cross. Consequently, the fixation point, if any, is unknown.

5. The rotation of the cameras around their optical axes is unknown, but appears to be small.

6. The calibration (focal length and camera displacement for the images) is unknown.

7. Reflections of light sources occur in several places, which is normal in ordinary environments.

Initially, some experiments were carried out in order to attempt to "balance" the gray levels in the two images, but the results proved futile due to the near-total loss of detail in some regions of the images. After further thought, however, all the defects in the two images and the missing information about the cameras were viewed as challenges rather than drawbacks. The analysis programs must be able to handle such situations since they occur normally and present no difficulties for our visual system.

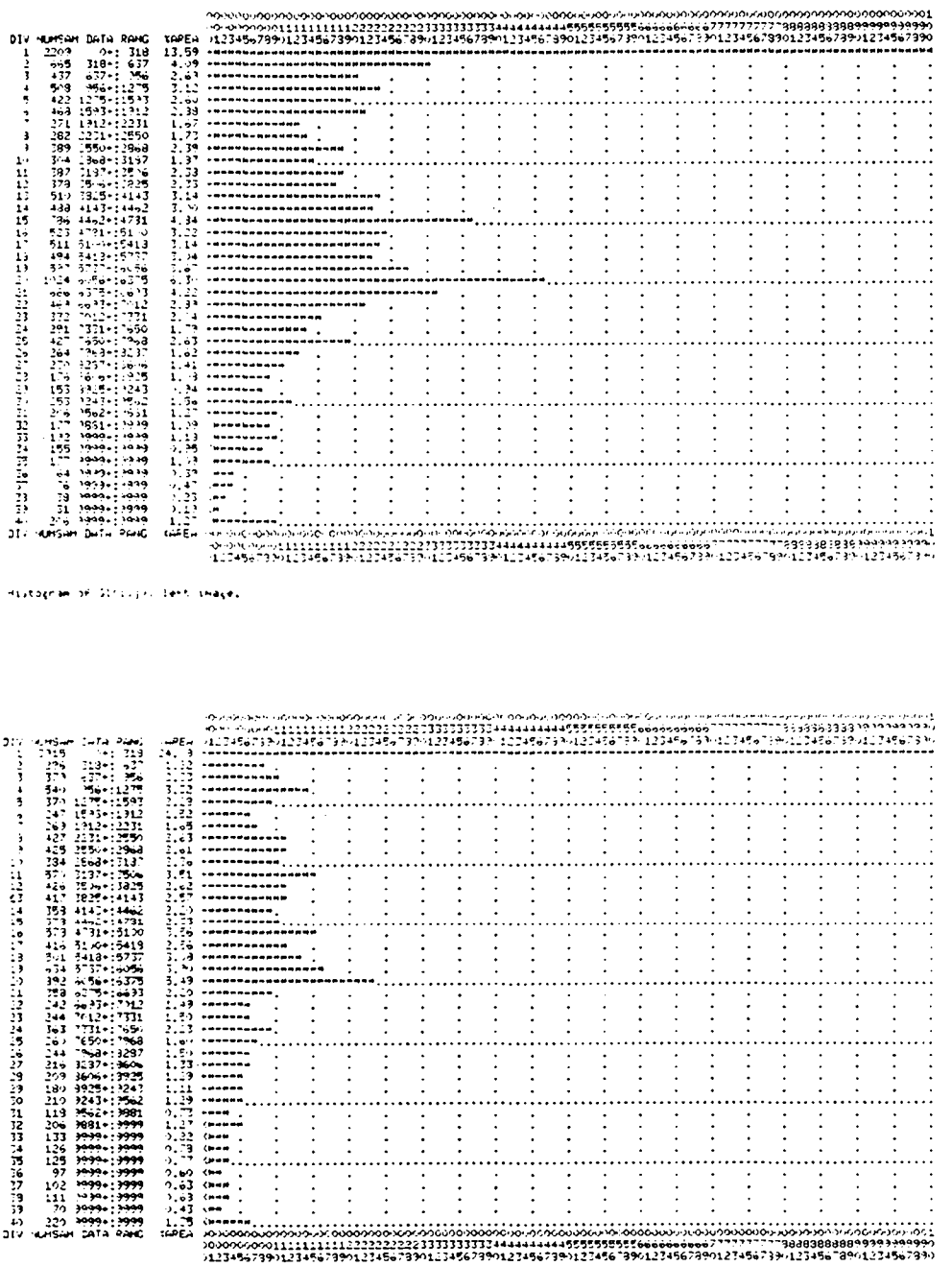


Fig. 2.2-1: Histograms of the reduced image pair. (File: Inr001.fig)

2.3 Computable and computed features

It was conjectured in chapter 1 that stereo matching on the pixel level is only feasible if an hierarchical image analysis and image matching structure is used. In three steps, the processing may take the following form:

A) Segmentation and matching on the "large" scale.

B) Segmentation and matching of the "subsegments" within each segment obtained in (A).

C) Matching of edge and pixel features within subsegments of (B).

In other words, hierarchical segmentation and matching is carried out on ever-decreasing regions until there are a sufficient number of pixel-level features to allow unique pixel matching within each sub-region. The only critical requirement is segment and feature similarity between the left and right images, not the meaning of the segments nor the meaning of the pixel features. Consequently, stereo matching or "depth vision" requires no "image understanding" and no "models of the scene" are necessary. This may contradict presently held popular opinions.

The digital image matrix is represented in the computer as a two-dimensional matrix of the form

```
        1  2  3  4  ... i  ... Idim --- The i-coordinate
    1   *  *  *  *  ... *  ... *
    2   *  *  *  *  ... *  ... *
    3   *  *  *  *  ... *  ... *
    .   .  .  .  .  ... *  ... .
    j   .  .  .  .  ... p  ... .
    .   .  .  .  .  ... *  ... .
    .   .  .  .  .  ... *  ... .
 Jdim   *  *  *  *  ... *  ... *
    ¦
    The j-coordinate
```

where a single pixel (*) will be referred to as (i,j) or "p". The "i" and "j" are image coordinates (the pixel address) of an image matrix Q(i,j). The "Q" may be any alphanumeric combination. For example, in G(i,j) the "G" refers to "gray level". G(p) is "short-hand" for G(i,j). If necessary, the left (gray level) image is represented as Gl(i,j) and the right image as Gr(i,j), where "l" and "r" imply "left" and "right".

In brief, the information available from the images may be classified as follows:

1. Features directly available at single pixel level from the input images. Normally these are intensity, colour, and "flicker", i.e., the gray level, spectral, and temporal

compositions. In the present case only the gray level G(i,j) is available (in intensity range 0 to 255).

2. Features computable from a relatively small local neighborhood N around each pixel p. There is no clearly definable limit to the number of such features. Furthermore, the results depend on the size and shape of N. The more common features are:

a) Average gray level Gav(i,j) in N.

b) The x and y gradients Gx(i,j) and Gy(i,j) in N as well as higher order derivatives.

c) "Edge" functions in N, i.e., various "edge detectors" which locate either the extremal value of the gradient in N or compute an "edge amplitude" in N.

d) Curvature or change of gradient direction in N.

e) Numerous linear or non-linear functions in N, for example, "busyness".

f) Co-occurrence matrices and other "structural descriptions" within N.

g) Statistical parameters in N, such as, variance, skew, (invariant) moments, etc.

Since there is no clearly definable upper bound to what could, in principle, be computed for a pixel p from the gray levels in N, a logical choice is to restrict the computations only to features that are:

1. Computable in the presence of existing noise.

2. Invariant to aspects in the image that are not wanted.

3. Statistically independent from each other.

4. Able to retain stereo integrity in L versus R.

Noise and improperly resolved details tend to be suppressed by using a relatively large N. Invariance depends on what is wanted and its importance depends on the subsequent processes. In Ref. 2.2 (table 1, page 86) it is shown that the gray level G(i,j) and its first (partial) derivatives Gx(i,j) and Gy(i,j) are independent, and that the first derivative is independent of the second, etc. Verification of "stereo integrity" is carried out by printing the results in image form (whenever feasible) and inspecting the results with stereo glasses. Clearly, if our vision cannot integrate the results into a "3D scene" then subsequent matching by computer is going to be difficult or impossible.

For the present case, the local neighborhood N is chosen relatively large and as isotropic as feasible, see Fig. 2.3-1. The computed features at each pixel p or (i,j) are:

i) The x and y gradient images Gx(i,j) and Gy(i,j), and the corresponding gradient magnitude Gpm(i,j) and gradient angle Gpa(i,j) images. Gradient angle is measured with respect to the x=i axis. The left and right gradient magnitude images Gpml(i,j) and Gpmr(i,j) are shown as a stereo pair in Fig. 2.3-2 for visual verification. Both the "black on white" and "white on black" versions are shown in this case (Figures 2.3-2a and -2b) in order to select the "easiest to see" version. The author preferred the "white on black" in most cases. Stereo integrity ("fidelity") is retained in a "ghostly" scene. The corresponding gradient angle images Gpal(i,j) and Gpar(i,j) are show in Fig. 2.3-3a and -3b. Only the four major directions are printed as - / ¦ \ (- = 0 +/- 22.5; / = 45 +/-22.5; ¦ = 90 +/-22.5; \ = 135 +/-22.5 degrees) for gradient magnitude values > 2.5 (in order to avoid the erratic directions at low gradient values).

ii) An "edge amplitude" function Ef(i,j) which indicates "the amount of edge" present at pixel (i,j).

$$Ef(i,j) = [\,|G'(i,j)| - W*|G''(i,j)|\,]^{+}$$

where, $|G'(i,j)| = Gpm(i,j)$ = gradient magnitude, $|G''(i,j)|$ = the magnitude of the second derivative in the direction of the gradient angle Gpa(i,j), W = a positive weight, and the $[.]^{+}$ indicates If(Ef(i,j).Gt.0) Then Ef(i,j)=[.] Else Ef(i,j)=0 (.Gt. means "greater than"). It should be noted that the weight W can be used to "sharpen" the edge function (until Ef becomes so narrow that it may "vanish" at but not between the pixels). Numerous more formally defined "edge functions" may be found in (2.3). The edge functions Efl(i,j) and Efr(i,j) are shown as a stereo pair in Fig. 2.3-4. Again, stereo fidelity remains in a "ghostly" manner.

iii) A curvature function C(i,j) at values where the edge function Ef(i,j) is nonzero. In essence, C(i,j) represents the change of the direction of the gradient vector within N.

iv) The "automatic volume control" type of contrast enhancement

$$G''(i,j) = [G(i,j) - Gav(i,j)] / [Gstd(i,j) + Const]$$

where Gav(i,j) is the average gray in N, Gstd(i,j) is the standard deviation of G(i,j) in N, and Const is a constant. A stereo pair of G''(i,j) is shown in Fig. 2.3-5 (Const = 1.0) and a "larger version" in Fig. 2.3-6 in order to show the "texture" better. (There is a "name" and considerable theoretical analysis of this type of filter, presumably in astronomy, but the author has forgotten the references.)

At the moment the Ef(i,j) and G˜(i,j) images have only been used to verify stereo consistence ("integrity", "fidelity") for these types of features. The Gpm(i,j) and Ef(i,j) images do present initially a "ghostly" appearance, presumably since we are not used to such images and they have been highly degraded by the low intensity amplitude ("gray") resolution in these prints. G˜(i,j) is interesting since it tends to confirm the hypothesis that hierarchical segmentation is feasible.

Since these processes are well known, the interest in these experiments centered on stereo fidelity. Clearly, at least some of the features computable from the local neighborhood N retain stereo integrity. As a question to think about: "If stereo integrity were not maintained, is it meaningful to proceed"?

Figure titles:

Fig. 2.3-1: The neighborhood N used in the present studies. (File: Inr002.fig)

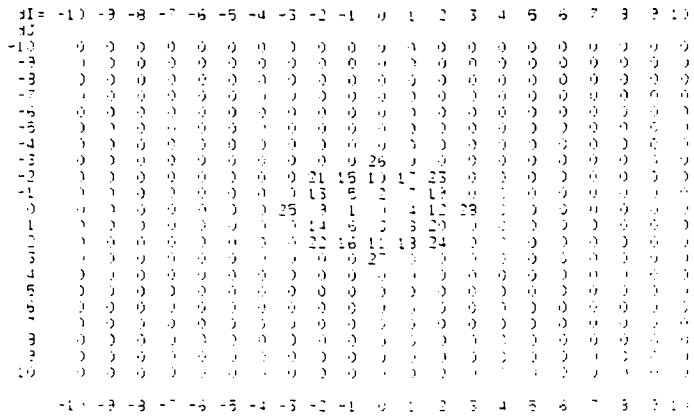
Fig. 2.3-2a and -2b: The left and right gradient magnitude images Gpml(i,j) and Gpmr(i,j) shown as a stereo pair. (a) Printed as black on white. (b) Printed as white on black.

Fig. 2.3-3a and -3b: The gradient angle images Gpal(i,j) and Gpar(i,j). Only the four major directions are printed as - / ¦ \ (- = 0 +/-22.5; / = 45 +/-22.5; ¦ = 90 +/-22.5; \ = 135 +/-22.5 degrees) for Gpm(i,j) values > 2.5 (in order to avoid the erratic directions at low gradient values). (Files: Inr041.fig and Inr042.fig)

Fig. 2.3-4: The edge functions Efl(i,j) and Efr(i,j) are shown as a stereo pair.

Fig. 2.3-5: The "automatic volume control" functions G˜l(i,j) and G˜r(i,j) shown as a stereo pair.

Fig. 2.3-6: The "automatic volume control" functions G˜l(i,j) and G˜r(i,j) shown as a "magnified" stereo pair.



K = Lookup table address.
dI = I-coordinate increment.
dJ = J-coordinate increment.
R = Radius from (0,0) to pixel at (dI,dJ).
A = Angle (degrees) from (0,0) to pixel at (dI,dJ).

K	dI II(K)	dJ JJ(K)	R RRK(K)	A AAK(K)
1	-1	0	0.1000000E+01	0.1800000E+03
2	0	-1	0.1000000E+01	0.9000000E+02
3	0	1	0.1000000E+01	0.2700000E+03
4	1	0	0.1000000E+01	0.0000000E+00
5	-1	-1	0.1414214E+01	0.1350000E+03
6	-1	1	0.1414214E+01	0.2250000E+03
7	1	-1	0.1414214E+01	0.4500000E+02
8	1	1	0.1414214E+01	0.3150000E+03
9	-2	0	0.2000000E+01	0.1800000E+03
10	0	-2	0.2000000E+01	0.9000000E+02
11	0	2	0.2000000E+01	0.2700000E+03
12	2	0	0.2000000E+01	0.0000000E+00
13	-2	-1	0.2236068E+01	0.1534350E+03
14	-2	1	0.2236068E+01	0.2065650E+03
15	-1	-2	0.2236068E+01	0.1165651E+03
16	-1	2	0.2236068E+01	0.2434349E+03
17	1	-2	0.2236068E+01	0.4343493E+02
18	1	2	0.2236068E+01	0.2965651E+03
19	2	-1	0.2236068E+01	0.2656505E+02
20	2	1	0.2236068E+01	0.3334349E+03
21	-2	-2	0.2828427E+01	0.1350000E+03
22	-2	2	0.2828427E+01	0.2250000E+03
23	2	-2	0.2828427E+01	0.4500003E+02
24	2	2	0.2828427E+01	0.3150000E+03
25	-3	0	0.3000000E+01	0.1800000E+03
26	0	-3	0.3000000E+01	0.8999998E+02
27	0	3	0.3000000E+01	0.2700000E+03
28	3	0	0.3000000E+01	0.0000000E+00

File: Inc002.fig

Fig.: 2.3-1.

if8gradm.lgm if8gradm.rgm

Fig.: 2.3-2a.

if8gradm.lgm if8gradm.rgm

Fig.: 2.3-2b.

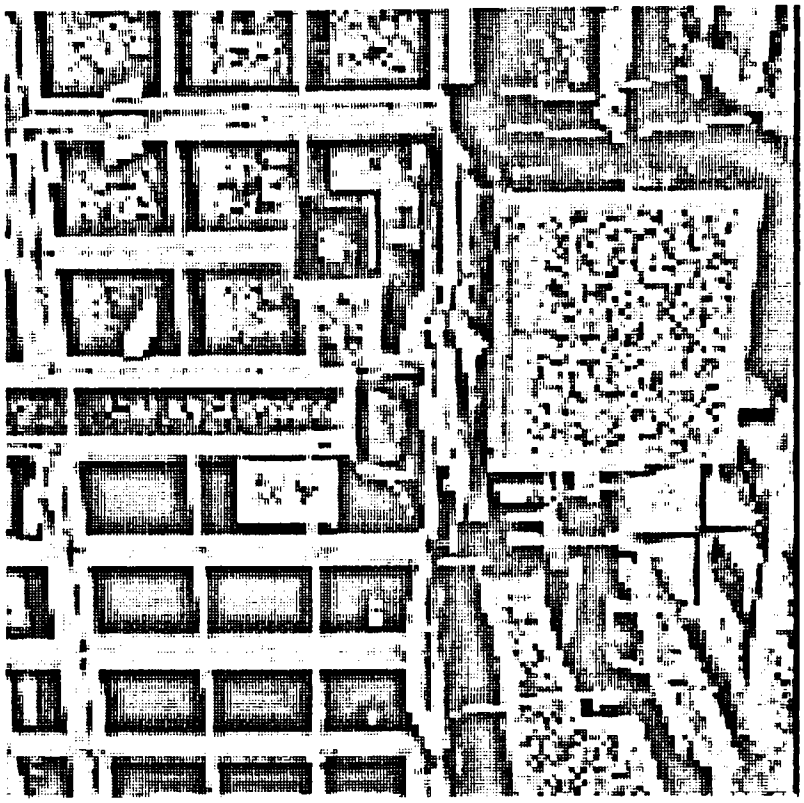
Figs.: 2.3-3a and -3b.

if8edgefl.lef if8edgefl.ref

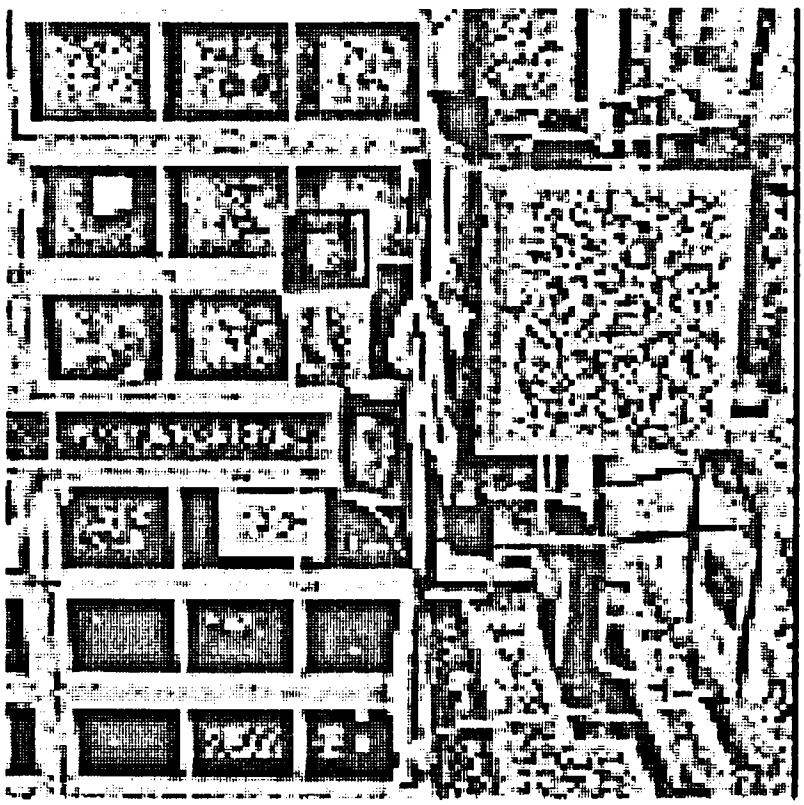
Fig.: 2.3-4.

if8volc2.lgr if8volc2.rgr

Fig.: 2.3-5.



if8volc2.rgr



if8volc2.lgr

Fig. : 2.3.6.

2.4 Hierarchical feature classification

The next stage in the processing of the left and right images is to detect larger "entities" in the images which should be easier to match than the individual pixel features. The word "entity" is used in the sense of "something larger than a pixel". At the moment the choice is between two kinds of "entities", namely, the one-dimensional entities called "edges", and the two-dimensional entities called "regions". The choice is thus between:

1. "Edge detection", followed by assembling the "edge pieces" into longer edges.

2. "Region detection", consisting of grouping selected pixel features into larger wholes.

In both cases a profusion of techniques is available and, at least in principle, edge and region detection operations are complementary since edges surround regions or regions are surrounded by edges. In practice, however, the contours (edges) around "regions" are automatically guaranteed to be continuous closed (digital) curves. If the "edges" are detected first, then the result is usually a profusion of "edge pieces" which are difficult or impossible to assemble into unique continuous closed curves, except in highly contrasting images. In region detection frequently many small "noise regions" are obtained which partially negate the advantages of a region-based approach. A "contour follower" is a compromise between the two methods but it has its own "problems".

The "region-based" approach was chosen for the following reasons:

a) Much work has already been done on the edge-based techniques. It is difficult to improve the results and meaningless to repeat these efforts. Furthermore, "edge matching" is difficult in the absence of "additional information".

b) Regions should be easier to match in the left and right images than edges and the edges are easily obtained after regions have been detected. The regions provide the "additional information" for edge matching without the need of any other sources of "knowledge".

c) The region-based approach coincides with the endeavors in the SYNTIM project.

Among the many methods available for "segmentation", a modified version of the "potential function" method was adapted (long ago) for low-dimensional decision spaces. The decision space is quantized and tabulated in the computer and then processed to become a look-up table for pixel feature based classification.

This leaves some apparently very important questions unanswered, namely:

1. Which features should be used for segmentation?

2. What is the criterion for judging segmentation results?

Both questions are fundamentally trivial since the left and right images are of the same scene but from slightly different "points of view". Consequently, if the same feature extraction algorithms are applied to the two images, the extracted features are expected to be rather similar (except in some restricted regions in the images where, for example, reflections occur and regions are visible in one but not in the other image). Thus, the only important criterion is that the features form clusters in the decision space and that these clusters represent relatively large connected (or detectable) regions in the image space. Neither the features nor the segments need to have any meaning in the human sense, as long as stereo integrity is maintained in the segmentation.

In the present experiments the pixels were classified according to their unfiltered gray levels G(i,j), the magnitude of the spatial gray level gradient Gpm(i,j), and the direction angle Gpa(i,j) of the gray level gradient with respect to the x=i axis, i.e.:

The left image is characterized by:

Gl(i,j) = Original gray levels in the reduced image of I001g_o.
Gpml(i,j) = Gradient magnitude computed over N from Gl(i,j).
Gpal(i,j) = Gradient angle computed over N from Gl(i,j).
Gxl(i,j) = Gradient in x-direction computed over N from Gl(i,j).
Gyl(i,j) = Gradient in y-direction computed over N from Gl(i,j).

And the right image is characterized by:

Gr(i,j) = Original gray levels in the reduced image of I001d_o.
Gpmr(i,j) = Gradient magnitude computed over N from Gr(i,j).
Gpar(i,j) = Gradient angle computed over N from Gr(i,j).
Gxr(i,j) = Gradient in x-direction computed over N from Gr(i,j).
Gyr(i,j) = Gradient in y-direction computed over N from Gr(i,j).

The features computed from any input image are stored in image form and always kept in registration with the input image. The Gx(i,j) and Gy(i,j) images are used later in analytic relaxation.

The original images are assumed to form a stereo pair but image content is assumed unknown. The number of feature classes in the decision space, the shape of the clusters (classes), and their extraction is to be "automatic", i.e., without operator intervention.

The processing consists of the following steps:

1. Individual histogramming of the gray level images Gl(i,j) and Gr(i,j), and gradient magnitude Gpml(i,j) and Gpmr(i,j) images to establish upper and lower bounds for significant data. The gradient angle Gpa(i,j) bounds are 0 and 360 degrees.

2. A two-dimensional decision space (2D histogram) H(Gpm,G) constructed from both the left and right images scaled to the bounds obtained from step 1. H(Gpm,G) is a two-dimensional histogram of Gpm(i,j) versus G(i,j) quantized to 73 by 73 intervals and tabulated in the computer. The left and right image features are amalgamated in H(Gpm,G). An one character per "pixel" alpha-numeric print of H(Gpm,G) is shown in Fig. 2.4-1. Each "pixel" value is represented by 0, 1, ..., 9, A, B, ..., Z, (36 levels, see the scale at the bottom on the figure). In order to show the "less populated" slot values in H(Gpm,G), the histogram peaks were clipped at 100.

3. Dynamic thresholding of H(Gpm,G) and conversion to a binary version Hb(Gpm,G).

4. Labelling of connected clusters in Hb(Gpm,G).

5. Label "spreading" since dynamic thresholding (in step 3) only "captures" the local peaks in H(Gpm,G).

6. Application of a constraint after label "spreading". In the present case the constraint consist of only retaining the clusters in H(Gpm,G) that contain low (and zero) values of the spatial gradient magnitude Gpm. After application of the constraint the remaining clusters are relabelled. The labelled decision space Hl(Gpm,G), which is now a look-up table, is shown in Fig. 2.4-2, where the characters represent cluster labels (C=2, D=3, etc.). Cluster labels are 2, 3, 4, etc.

Comments: The major variables that control the processes from steps 1 to 6 are:

a) The histogram bounds within which the data values are considered to be significant. Usually 1% of the upper and lower histogram areas of G(i,j) and Gpm(i,j) are rejected as "outliers".

b) The size of the tabulated decision space H(Gpm,G). The 73x73 quantization of the histogram is a "default" value from 128x127 range data classification. Clearly, cluster formation depends on the size of the decision space but this dependance is not critical and is also controllable by the parameters in step 3.

c) Dynamic thresholding consists of regularization of H(Gpm,G) with two different kernels, subtraction of the results, and thresholding. The parameters are the "sigmas" of the two Gaussian regularization kernels and the subsequent threshold level. Prior investigations (for range images) indicated that these are not critical and, consequently, the old "default" parameters were retained.

7. Classification of the left and right images based on Hl(Gpm,G) and the gray levels G(i,j) and gradient magnitudes Gpm(i,j). These classified images were only used as "masks" to select the unclassified pixels for the next level of classification (see step 8).

8. A two-dimensional decision space H(Gpm,Gpa) amalgamating the gradient magnitudes Gpm(i,j) and gradient angles Gpa(i,j) for the unclassified pixels in step 7 for both the left and right images. The alpha-numeric print of H(Gpm,Gpa) is shown in Fig. 2.4-3. The 73x73 quantization was a "default".

9. Dynamic thresholding of H(Gpm,Gpa) using the same parameters as in step 3. This gives Hb(Gpm,Gpa).

10. Constraints and labelling adapted to stereo image pairs and subsequent processes.

Comments, continued: Assume that the cameras are horizontally displaced. In order to distinguish between horizontal, vertical, and +/- 45 degree directions (with respect to the plane of the two cameras), artificial clusters were created in the thresholded (binary) decision space Hb(Gpm,Gpa). The horizontal and vertical "directional preferences" could be dictated by gravity, but the others may be purely "cultural effects". Other types of "major directions" are equally easy to impose. The clusters were labelled as in step 4 and the labels were "spread" as in step 5. Another constraint consisted in separating the region in Hb(Gpm,-Gpa) corresponding to low values of the gradient magnitude Gpm(i-,j) since the corresponding gradient directions Gpa(i,j) are rather erratic and cannot be classified correctly. This region was assigned a "don't care" or "take-me" label (-1). The resultant labelled Hl(Gpm,Gpa) is shown in Fig. 2.4-4, where the "take-me" label is indicated by @. The parameters that control steps 5 and 9 are the same as those for processing H(Gpm,G). The new constraints are:

d) The number of orientations into which the Hb(Gpm,Gpa) space is split. In the present case the 45 degree interval was chosen.

e) The subsequent processing was simplified by grouping the 45 degree directions into four major directions, i.e., horizontal, vertical, and +/- 45 degree tilts. For further details, see "Relaxation".

f) The size (width) of the "don't care" or "take-me" region in Hl(Gpm,Gpa). A slight improvement in classification was obtained by using the "take-me" labels, see "Post processing" for details.

11. Classification of the remaining pixels in the left and right images using Hl(Gpm,Gpa).

12. Combining the two classifications. Since the pixels classified by Hl(Gpm,G) and Hl(Gpm,Gpa) are mutually exclusive, the classified images may simply be merged (if the labelling is continued from Hl(Gpm,G) to Hl(Gpm,Gpa). The raw results of classification after absorbtion of the "don't care" pixels are shown in Figures 2.4-5a and -5b for the left and right images.

The processes described in steps 1 to 12 are summarized in block diagram form in Figures 2.4-6a to -6h.

Comments, continued: It should be noted that the clusters found by using H(Gpm,G) represent relatively uniform gray level regions in the left and right images, while the clusters from H(Gpm,Gpa) represent relatively narrow and elongated regions with "cylindrically distributed" gray level gradient directions. The constraint that unified the 0 and 180, 45 and 225, etc., degree directions was deliberate.

g) The mistakes in the classification are of the following types:

i) A rather sharp reflection of a point light source which creates an approximately spherical light distribution is split into four or more regions. Such a region has known class label distribution but no attempts have been made to collect these into a larger regions.

ii) If a narrow (and long) region is wider than the approximate diameter of the local operator N used in computing the gradients, then the region is split into three narrow strips. One strip represents the centre region (uniform gray with low gradient) and the remaining two are "edge" regions where the gradients are high. However, in "critical" cases the splitting depends on minor variations in the gray levels and is thus erratic.

iii) Very "busy" areas which contain a more or less random distribution of high gray level gradients. Such areas are split into many small regions.

iv) A portion of one image (L or R) contains detectable regions while the corresponding region in the other image (R or L) is saturated (approximately 0 or 255 valued gray levels). Consequently, the regions found will be rather different.

Examples of all of these defects may be found in the classified Lh(i,j) and Rh(i,h) images, see Figures 2.4-5a and -5b.

13. Post processing. In practice the classification stage is followed by a certain amount of "post processing" since the Lh(i,j) and Rh(i,j) images contain:

i) Unclassified pixels.

ii) Pixels where the classification is uncertain (indicated by the "take-me" labels).

iii) Wrongly classified pixels, presumably due to noise and quantization effects.

Clearly, the nature of the post processing will depend on the purpose of the classification, which indicates the nature of additional knowledge that could be introduced at this stage. In the present case the classification is intended to help in matching of the corresponding regions in the left and right images. Consequently, the regions are only made "slightly more uniform" by using "majority logic" in a 3 by 3 local neighborhood N3. The steps are essentially as follows:

1) Replace the "take-me" or "don't care" labels by the majority label in N3.

2) Replace the unclassified or 0-valued labels by the majority label in N3.

3) Replace the "unsupported" or "lonely" and "poorly supported" labels by the majority label in N3.

These processes may be iterated but it should be remembered these are "ad hoc" processes and should not be carried "too far". In the present case all the "take-me" labels were replaced if they could be replaced and the remaining ones were zeroed (since some of them may be isolated). This was followed by one iteration for replacing 0-valued labels and one iteration of replacing "lonely" and "poorly supported" labels. The results are shown in Figures 2.4-7a and -7b.

Several attempts were made to try to display the results for visual 3D (stereo) inspection and to represent the results as fairly as possible. The different class boundaries have to be shown but they can also be very disturbing (distracting) to our vision since points and sharp corners are highly noticeable. In the first attempt the gray levels for each class were replaced by their respective average values. Figure 2.4-8a shows the class average gray levels before post processing and Figure 2.4-8b shows them after post processing. The averages differ very little (except for class 1 which is the boundary around the image). The

stereo pair of average gray for each class is shown in Figure 2.4-9. It mostly shows the defects of the printer and no stereo effect is visible. However, on the Vicom film-loop display one could see (or possibly imagine) a stereo effect. The classes were "painted" in a minimal number of "colours". This is the 4-colour problem with the constraint that certain regions have to be assigned common colours. The adjacency matrix for the classes is shown in Figure 2.4-10a and the new assigned colours are in Figure 2.4-10b. Six colours were needed, which is somewhat beyond the capabilities of the laser printer (for printing a proper stereo pair) but the result is much "livelier". The stereo pair of class labelled images "painted" in the new colours is shown in Figure 2.4-11. There may be a hint of stereo in these images but it is very difficult to see. A blown-up version of the stereo pair is shown in Figure 2.4-12.

Comments, continued: The hierarchical classification of the pixel features (gray levels G(i,j), gray level gradient magnitudes Gpm(i,j), and gray level gradient directions Gpa(i,j)) produced two types of regions, namely:

a) Regions of relatively uniform gray values and low gray level gradient magnitudes.

b) Regions of highly varying gray levels with approximately (within +/-22.5 degree) uniformly directed gray level gradient angles grouped around 0--180, 45--225, 90--270, and 135-315 degree directions.

Clearly, the regions of type (a) may be considered "flat" (uniform) while those of type (b) have "cylindrically distributed" gray levels. The classification has also produce a form of segment recognition based on the illumination distribution over the segment, i.e., it is now known which segments are "flat" and which are "curved". The "curved" segments have been separated into four orientation directions. The post processing step may have introduced some pixels into the various regions which are of either type. The gray level gradients were computed over a relatively large local neighborhood N. Hence, the regions with highly varying gray levels tended to become magnified at the expense of the more uniform regions. This effect is unavoidable since a local operator always needs a local neighborhood N over which to "operate". In the present case this "magnification effect" was also enhanced intentionally (by choosing a large N) in order to assist in matching of the left and right image segments.

From this stage onwards several "avenues" have opened up, i.e., analytic representations of the illumination levels for regions, numerous "shape descriptions" for the segments, direct matching, and so on.

Figure titles:

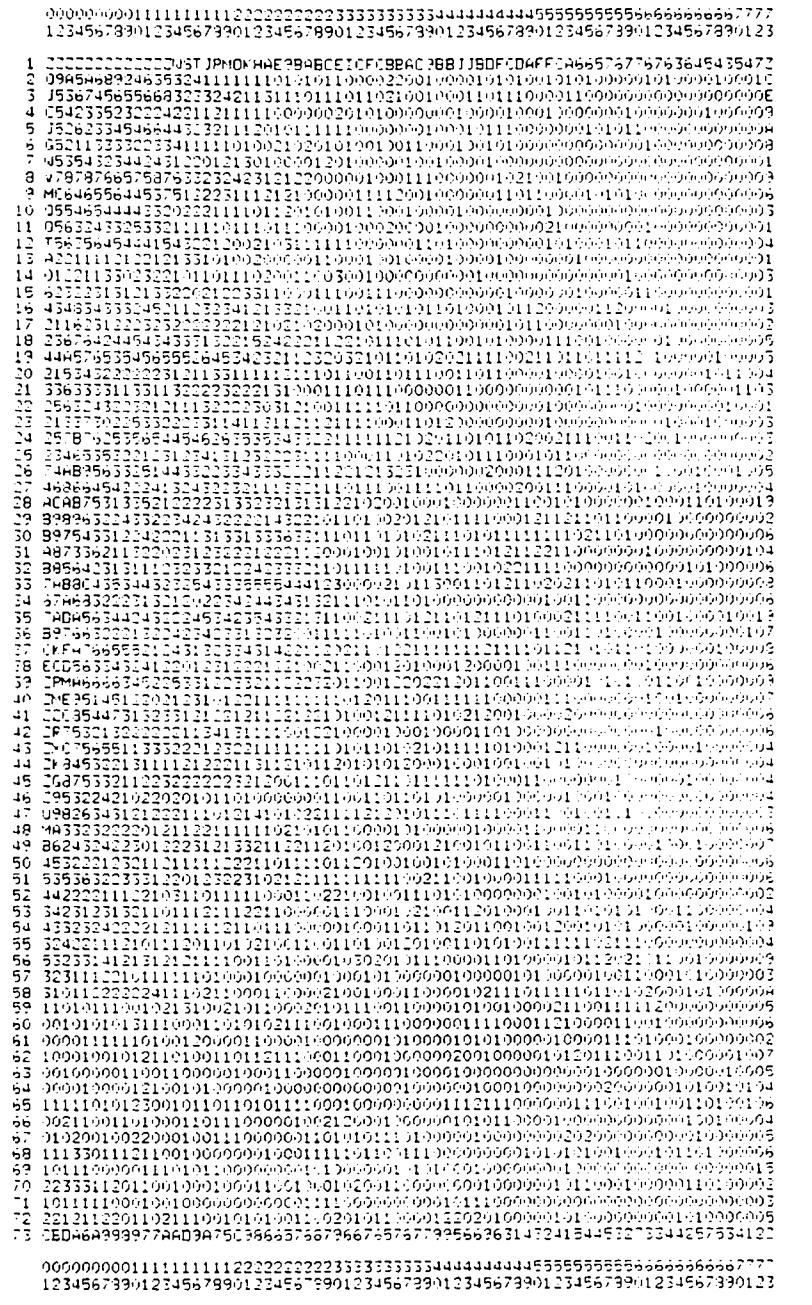
Fig. 2.4-1: An one character per "pixel" alpha-numeric print of H(Gpm,G). Each "pixel" value is represented by 0, 1, ..., 9, A, B, ..., Z, see the scale at the bottom of the figure. In order to show the "less populated" slot values in H(Gpm,G), the histogram peaks were clipped at 100.0 (highest value for the scale). (File: Inr009.fig)

Fig. 2.4-2: The labelled decision space Hl(Gpm,G), which is now a look-up table. The characters represent cluster labels (C=2, D=3, E=4, etc.). (File: Inr010.fig)

Fig. 2.4-3: An one character per "pixel" alpha-numeric print of H(Gpm,Gpa). (File: Inr013.fig)

Fig. 2.4-4: The labelled Hl(Gpm,Gpa) where the "don't care" or "take-me" label is indicated by @. (File: Inr014.fig)

Figures 2.4-5a & -5b: The ("raw") left Lh(i,j) and right Rh(i,j) images of class labels. One character per pixel prints of the raw results of hierarchical classification after absorbtion of the "don't care" pixels. (Files: Inr017 and Inr018.fig)



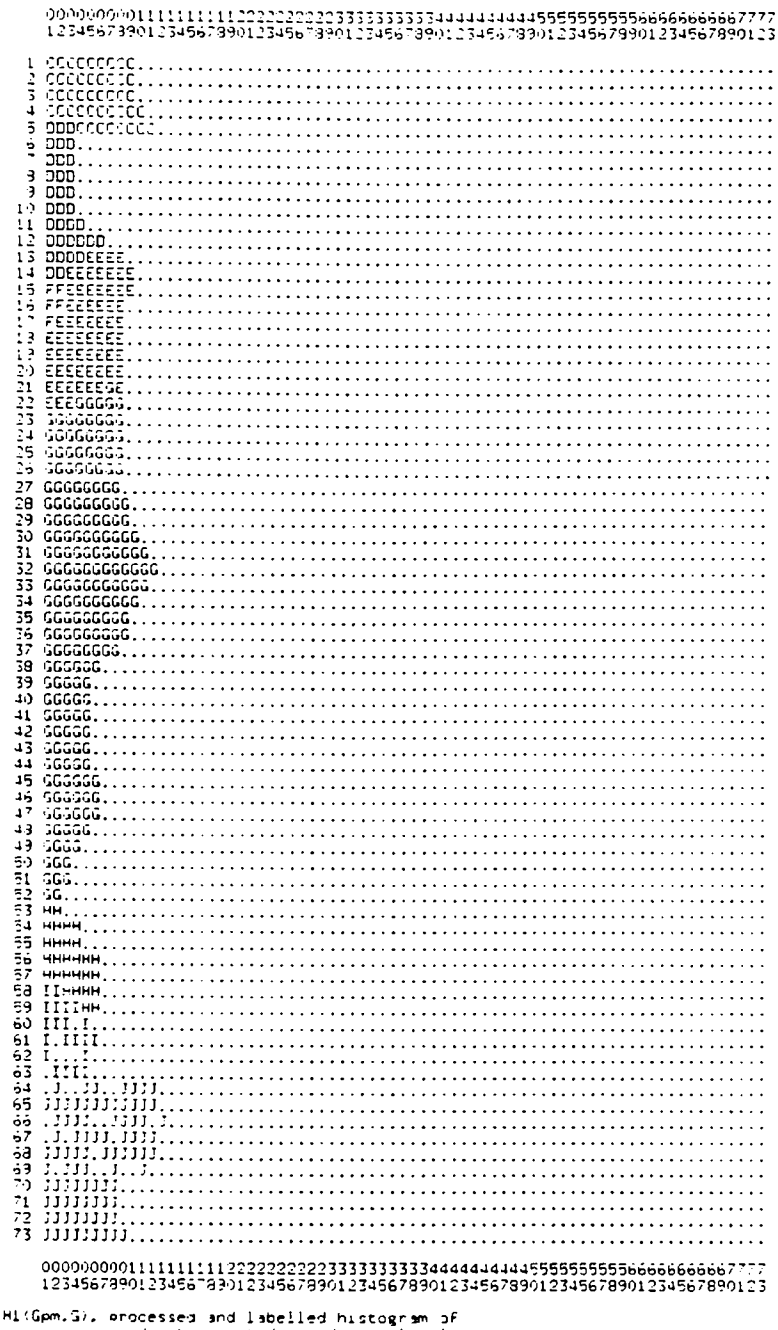
Hr(Gpm,G). raw histogram of gradient magnitudes and gray levels.

36 QUANTIZATION LEVELS. ALPHA-NUMERIC PRINT. PI>MAX.PI>MIN = 0.1000E+03 0.0000E+00

0.0000E+00	9>	0.2778E+01	1>	0.5556E+01	2>	0.8333E+01	3> 0.1111E+02
	4>	0.1389E+02	5>	0.1667E+02	6>	0.1944E+02	7> 0.2222E+02
	8>	0.2500E+02	9>	0.2778E+02	A>	0.3056E+02	B> 0.3333E+02
	C>	0.3611E+02	D>	0.3889E+02	E>	0.4167E+02	F> 0.4444E+02
	G>	0.4722E+02	H>	0.5000E+02	I>	0.5278E+02	J> 0.5556E+02
	K>	0.5833E+02	L>	0.6111E+02	M>	0.6389E+02	N> 0.6667E+02
	O>	0.6944E+02	P>	0.7222E+02	Q>	0.7500E+02	R> 0.7778E+02
	S>	0.8056E+02	T>	0.8333E+02	U>	0.8611E+02	V> 0.8889E+02
	W>	0.9167E+02	X>	0.9444E+02	Y>	0.9722E+02	Z> 0.1000E+03

File: Inc009.fig

Fig.: 2.4-1.



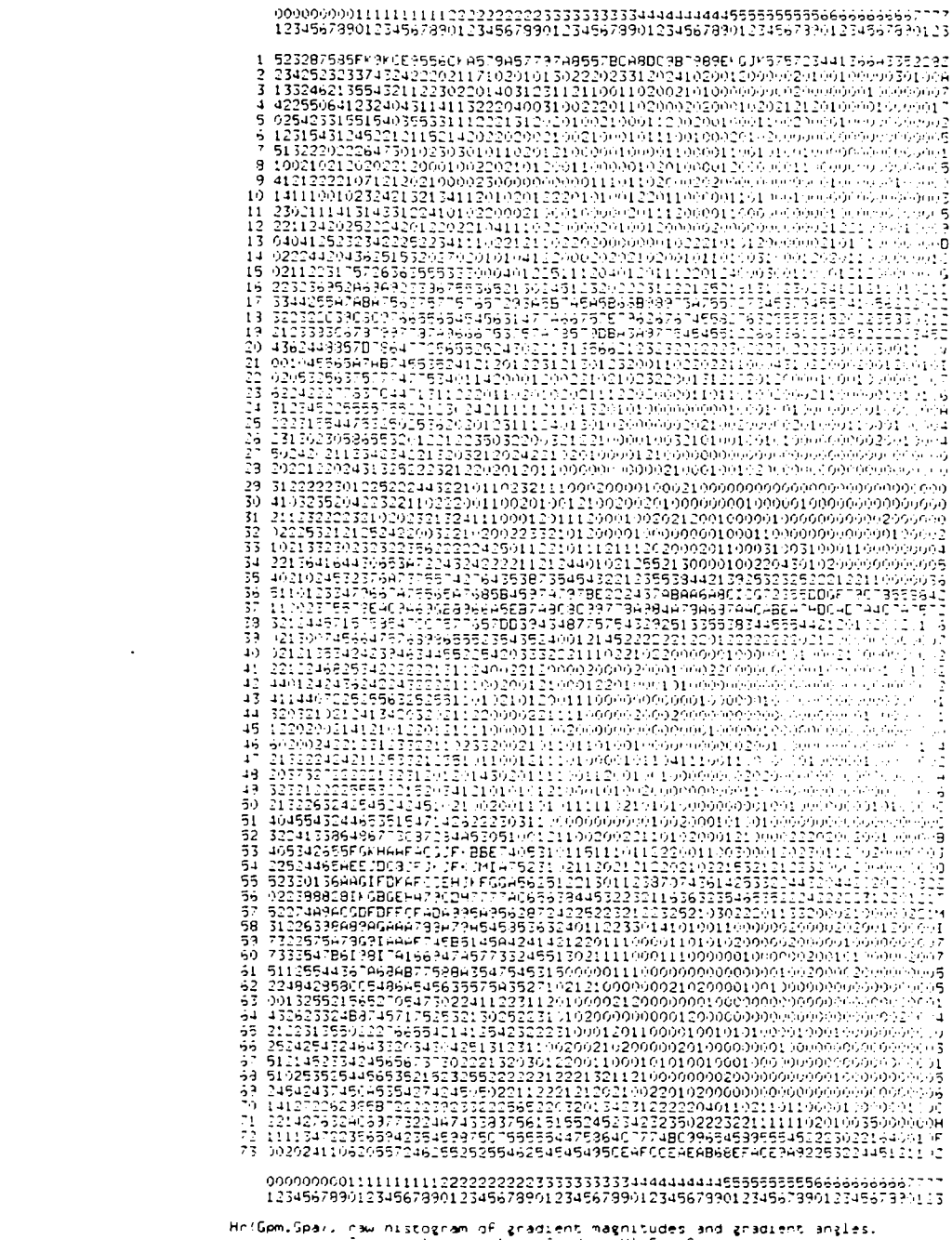
H1(Gpm.S), processed and labelled histogram of
gradient magnitudes and gray levels.

LABEL PRINTING

.= 0	*= 1	C= 2	D= 3	E= 4	F= 5	G= 6	H= 7		
I= 8	J= 9	K=10	L=11	M=12	N=13	O=14	P=15		
Q=16	R=17	S=18	T=19	U=20	V=21	W=22	X=23		
Y=24	Z=25	[=26	A=27	B=28	C=29	D=30	.ETC.		
.= 0	*= -1	c= -2	d= -3	e= -4	f= -5	g= -6	h= -7		
i= -8	j= -9	k=-10	l=-11	m=-12	n=-13	o=-14	p=-15		
q=-16	r=-17	s=-18	t=-19	u=-20	v=-21	w=-22	x=-23		
y=-24	z=-25	{=-26	a=-27	b=-28	c=-29	d=-30	.etc.		

File: Inr010.fig

Fig.: 2.4-2.



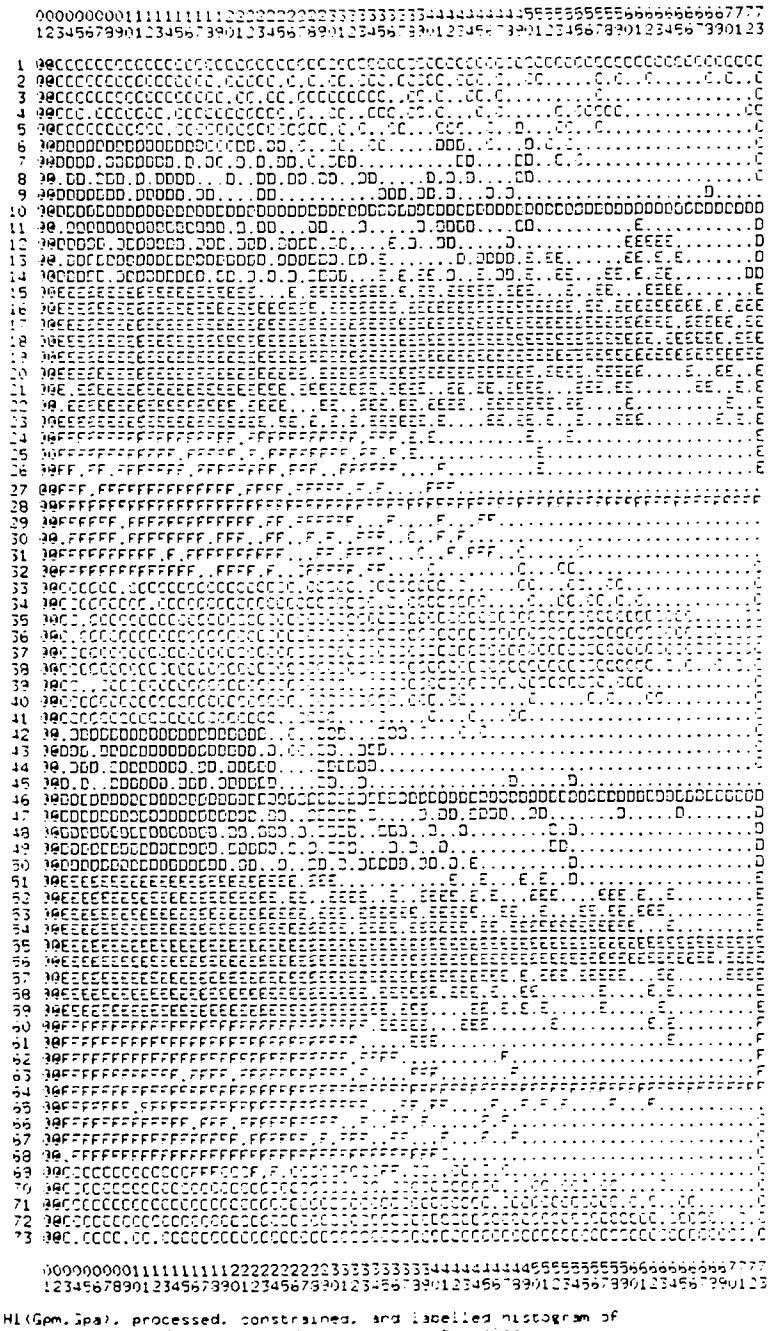
Hr(Gpm.Gpa), raw histogram of gradient magnitudes and gradient angles.
for pixels not classified by Hl Gpm.G).

36 QUANTIZATION LEVELS. ALPHA-NUMERIC PRINT. PIXMAX,PIXMIN = 0.5000E+02
0.0000E+00

0.0000E+00	<0>	0.1389E+01	<1>	0.2778E+01	<2>	0.4167E+01	<3>	0.5556E+01
	<4>	0.6944E+01	<5>	0.8333E+01	<6>	0.9722E+01	<7>	0.1111E+02
	<8>	0.1250E+02	<9>	0.1389E+02	<A>	0.1528E+02	<B>	0.1667E+02
	<C>	0.1806E+02	<D>	0.1944E+02	<E>	0.2083E+02	<F>	0.2222E+02
	<G>	0.2361E+02	<H>	0.2500E+02	<I>	0.2639E+02	<J>	0.2778E+02
	<K>	0.2917E+02	<L>	0.3056E+02	<M>	0.3194E+02	<N>	0.3333E+02
	<O>	0.3472E+02	<P>	0.3611E+02	<Q>	0.3750E+02	<R>	0.3889E+02
	<S>	0.4028E+02	<T>	0.4167E+02	<U>	0.4306E+02	<V>	0.4444E+02
	<W>	0.4583E+02	<X>	0.4722E+02	<Y>	0.4861E+02	<Z>	0.5000E+02

File: Inr013.fig

Fig.: 2.4-3.



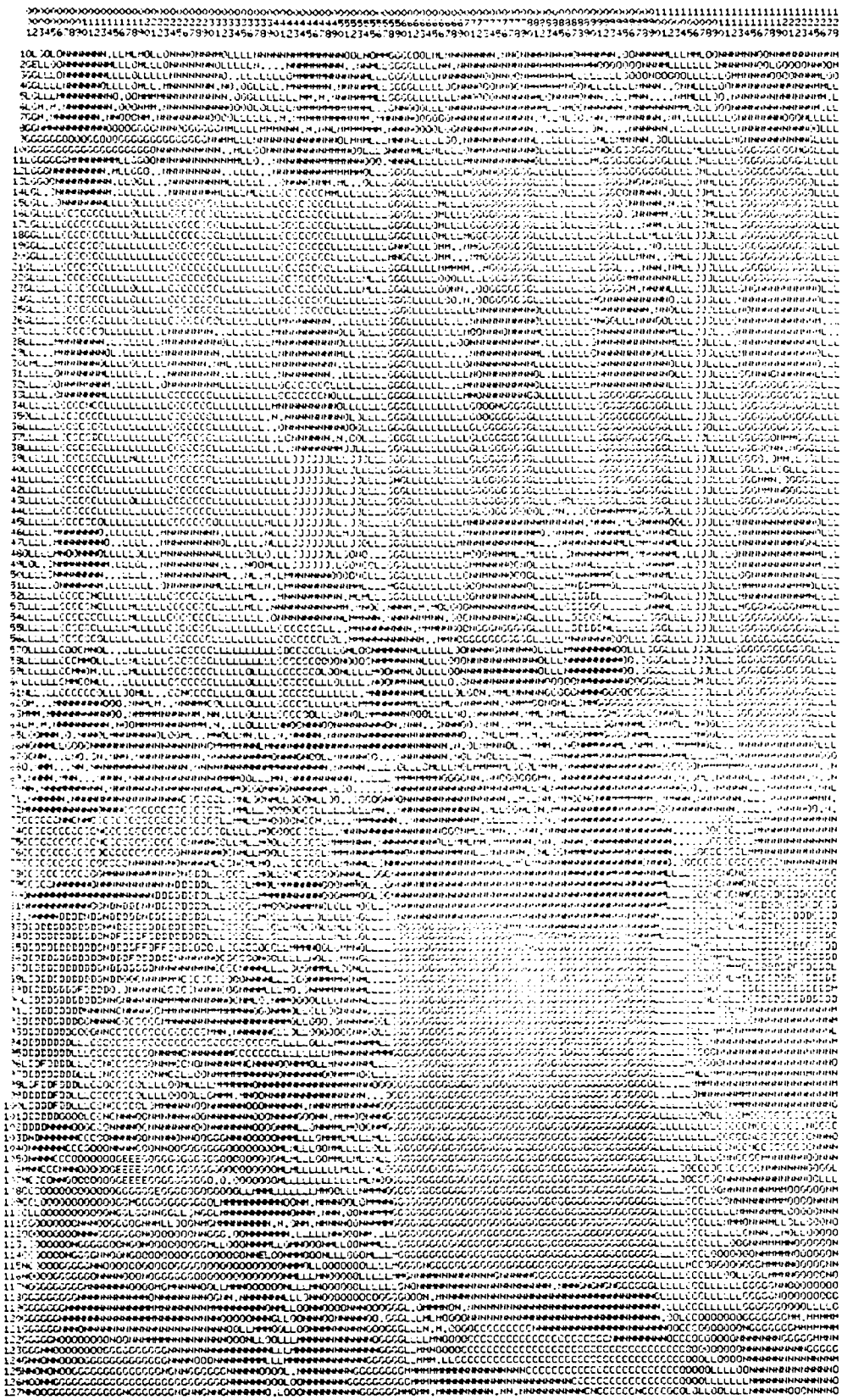
H1(Gpm.Ipa), processed, constrained, and labelled histogram of
gradient magnitudes and gradient angles.

LABEL PRINTING

.=	0	*=	1	C=	2	D=	3	E=	4	F=	5	G=	6	H=	7
I=	8	J=	9	K=	10	L=	11	M=	12	N=	13	O=	14	P=	15
Q=	16	R=	17	S=	18	T=	19	U=	20	V=	21	W=	22	X=	23
Y=	24	Z=	25	[=	26	\=	27	]=	28	^=	29	_=	30	.ETC.	
.=	0	A=	-1	b=	-2	c=	-3	d=	-4	e=	-5	f=	-6	g=	-7
h=	-8	i=	-9	j=	-10	k=	-11	l=	-12	m=	-13	n=	-14	o=	-15
p=	-16	q=	-17	r=	-18	s=	-19	t=	-20	u=	-21	v=	-22	w=	-23
x=	-24	y=	-25	z=	-26	{=	-27	|=	-28	}=	-29	~=	-30	.etc.	

File: Inr014.Fig

Fig.: 2.4-4.



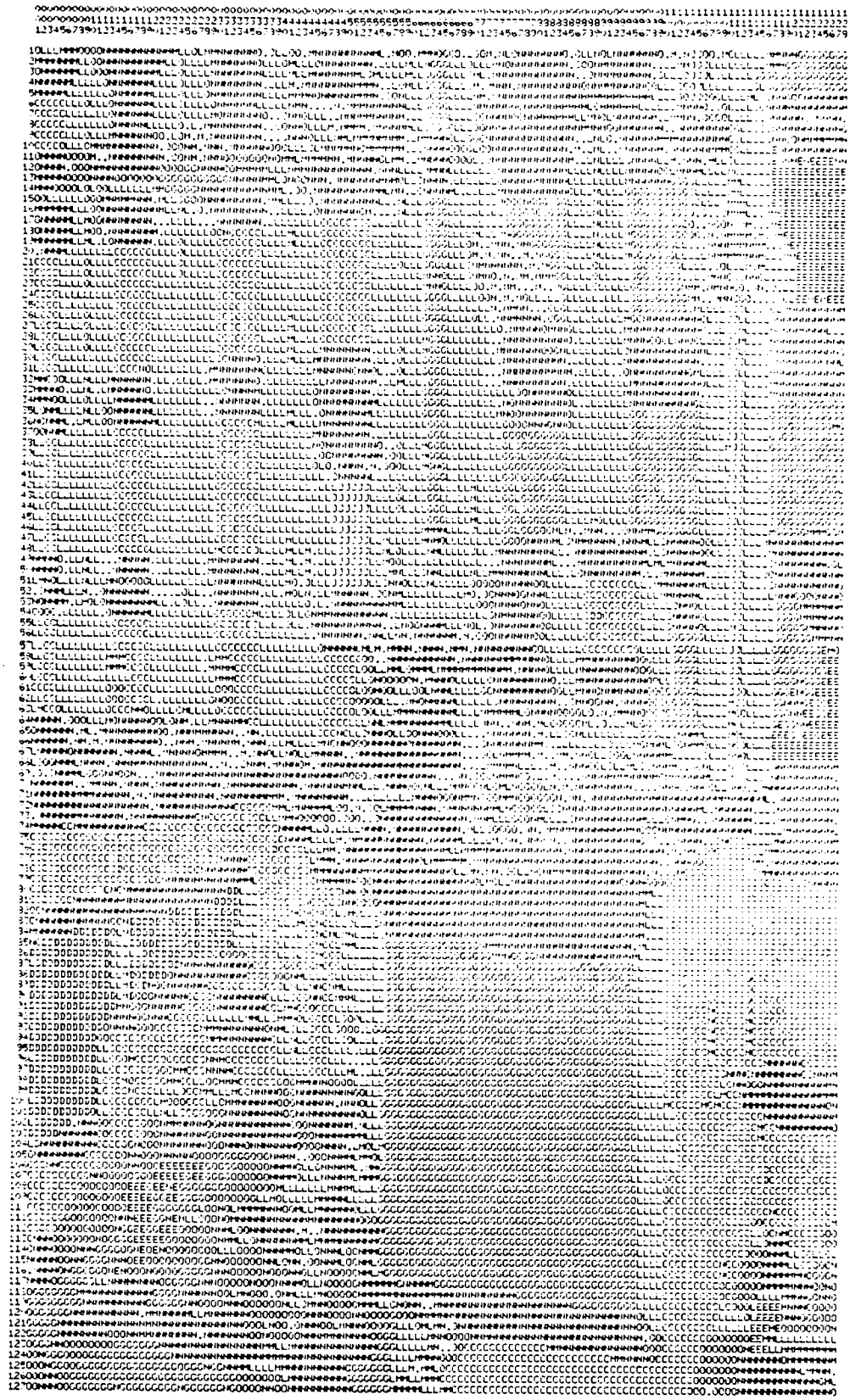
Ln(i,j), left image classified on Hl(Gpm,G) and Hl(Gpm,Gpa).

LABEL PRINTING

.= 0	== 1	C= 2	D= 3	E= 4	F= 5	G= 6	H= 7
I= 8	J= 9	K= 10	L= 11	M= 12	N= 13	O= 14	P= 15
Q= 16	R= 17	S= 18	T= 19	U= 20	V= 21	W= 22	X= 23
Y= 24	Z= 25	[= 26	A= 27	B= 28	C= 29	D= 30	.ETC.
.= 0	R= -1	c= -2	d= -3	e= -4	f= -5	g= -6	h= -7
i= -8	j= -9	k=-10	l=-11	m=-12	n=-13	o=-14	p=-15
q=-16	r=-17	s=-18	t=-19	u=-20	v=-21	w=-22	x=-23
y=-24	z=-25	<=-26	a=-27	b=-28	c=-29	d=-30	.etc.

File: Inr017.fig

Fig.: 2.4-5a.



Ph(i,j), right image classified on Hl(Gpm,G) and Hl(Gpm,Gpa).

LABEL PRINTING

A= 0	B= 1	C= 2	D= 3	E= 4	F= 5	G= 6	H= 7
I= 8	J= 9	K= 10	L= 11	M= 12	N= 13	O= 14	P= 15
Q= 16	R= 17	S= 18	T= 19	U= 20	V= 21	W= 22	X= 23
Y= 24	Z= 25	[= 26	A= 27	B= 28	C= 29	D= 30	.ETC.
.= 0	?= -1	c= -2	d= -3	e= -4	f= -5	g= -6	h= -7
i= -8	j= -9	k=-10	l=-11	m=-12	n=-13	o=-14	p=-15
q=-16	r=-17	s=-18	t=-19	u=-20	v=-21	w=-22	x=-23
y=-24	z=-25	[=-26	a=-27	b=-28	c=-29	d=-30	.etc.

File: Inn018.fig

Fig.: 2.4-5b.

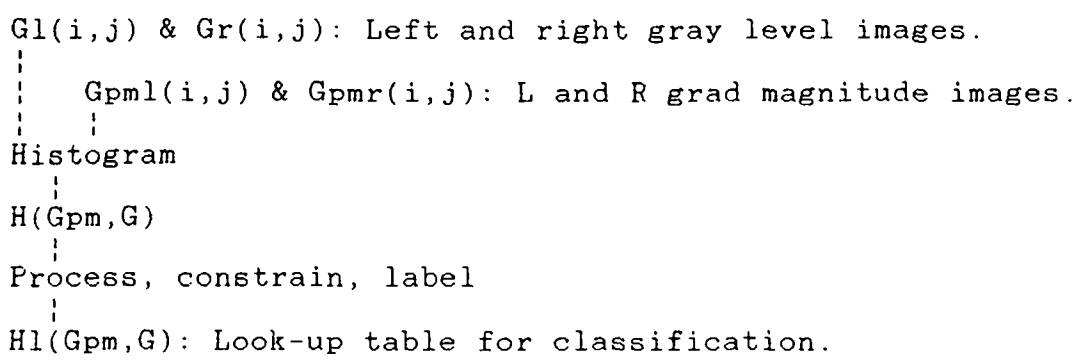


Figure 2.4-6a: Process steps 1 to 6, first level of hierarchy.

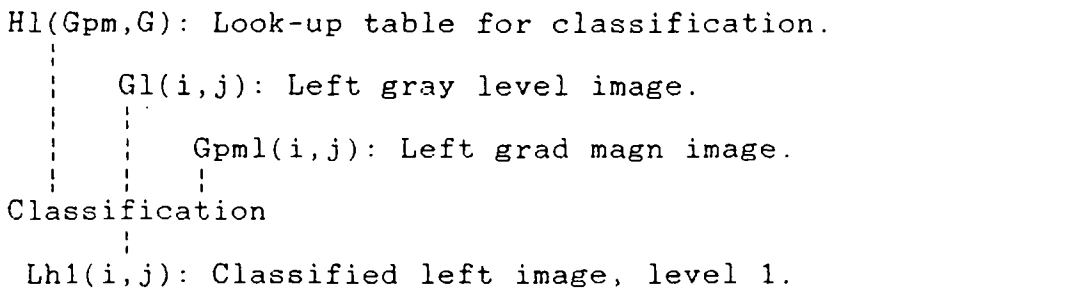


Figure 2.4-6b: Process step 7, L image, 1'st level of hierarchy.

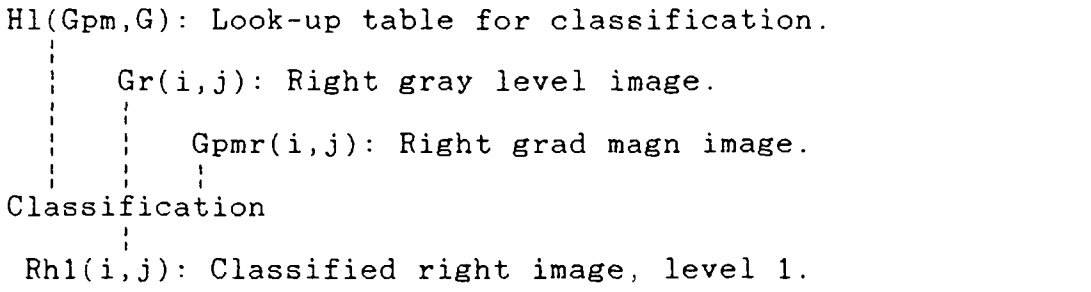


Figure 2.4-6c: Process step 7, R image, 1'st level of hierarchy.

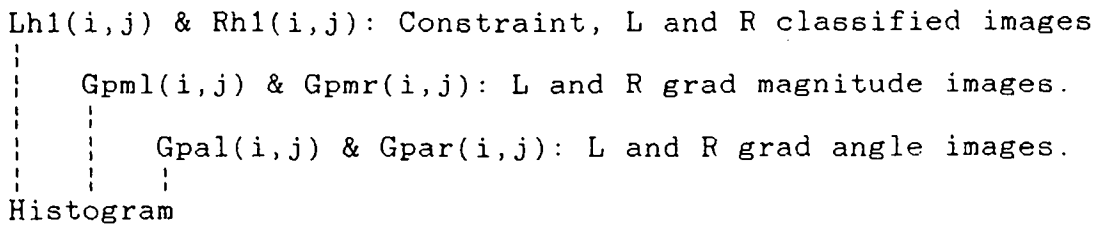


Figure 2.4-6d: Process steps 8 to 10, second level of hierarchy.

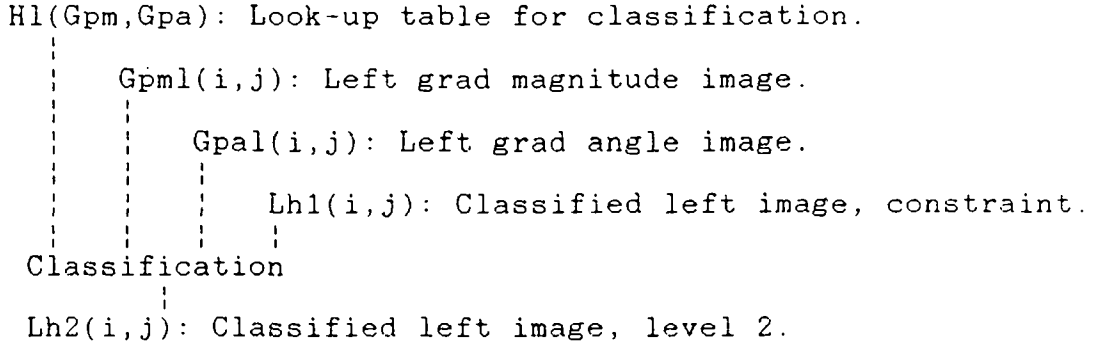


Figure 2.4-6e: Process step 11, L image, 2´nd level of hierarchy.

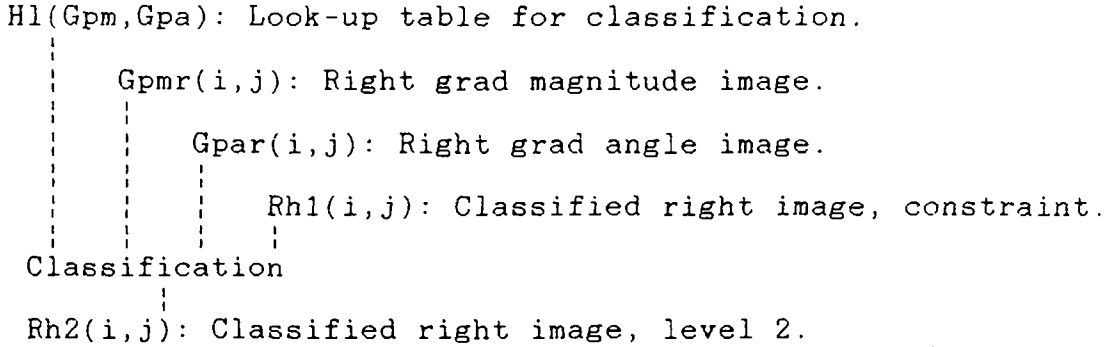


Figure 2.4-6f: Process step 11, R image, 2´nd level of hierarchy.

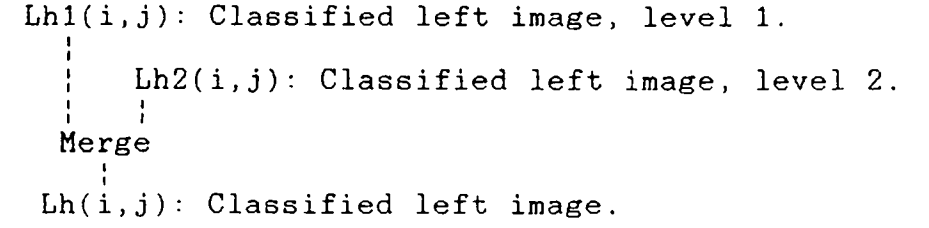


Figure 2.4-6g: Process step 12, merging left image classes.

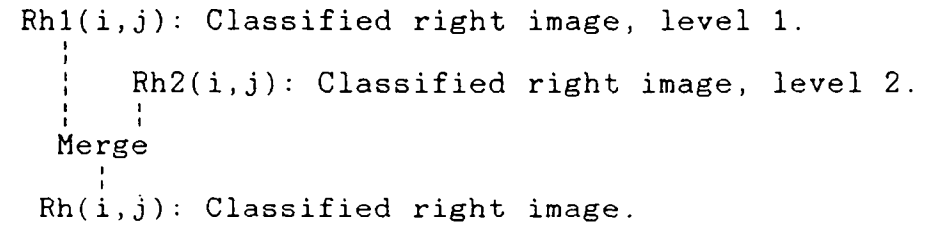


Figure 2.4-6h: Process step 12, merging right image classes.

Figure titles:

Figures 2.4-7a and -7b: The left Lh(i,j) and right Rh(i,j) images after post processing. The "take-me" labels were replace if possible, followed by one iteration for replacing 0-valued labels and one iteration of replacing "lonely" labels. (Files: Inr019.fig and Inr020.fig)

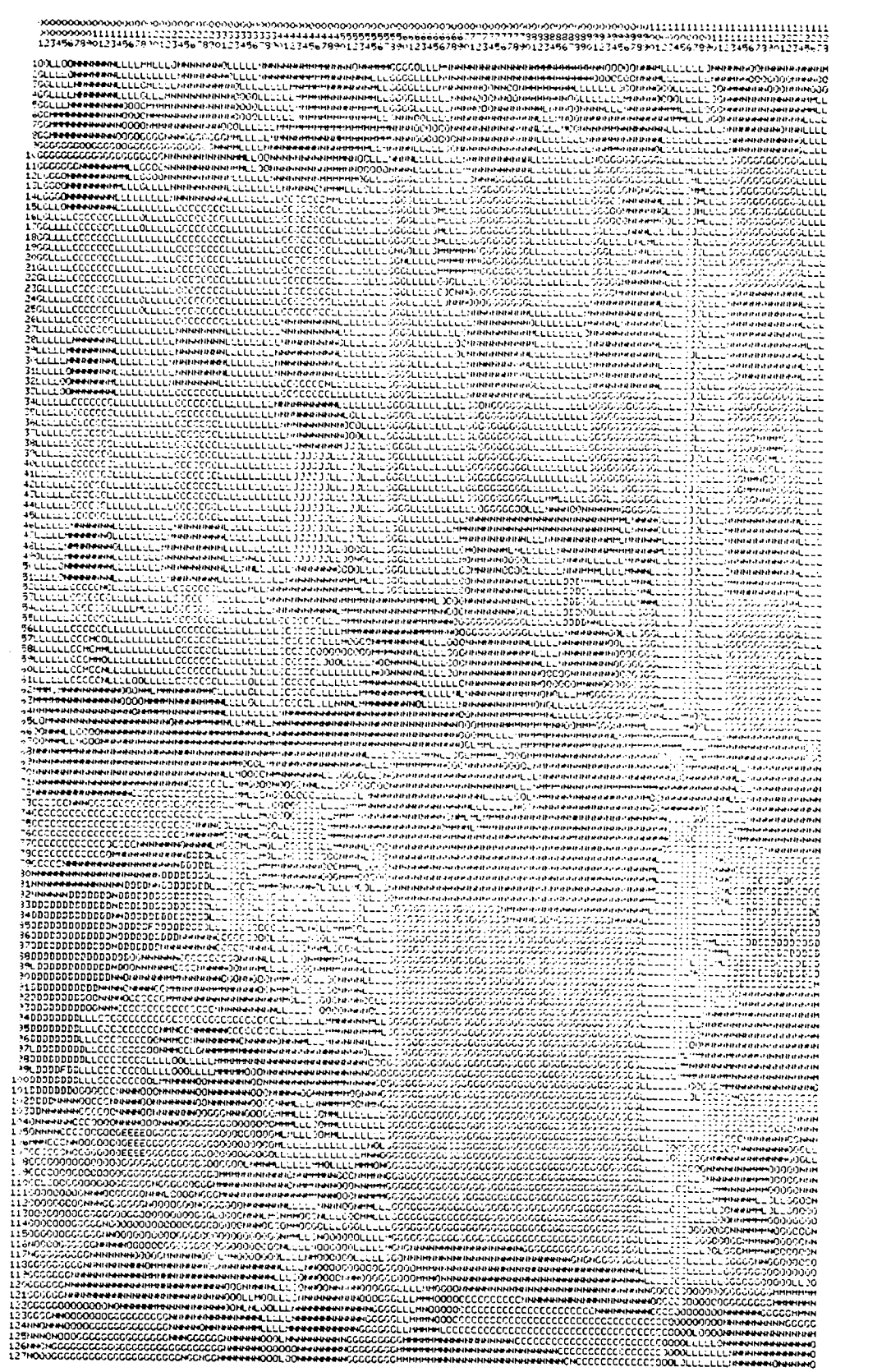
Figures 2.4-8a and -8b: The class average gray levels before post processing (-8a) and after post processing (-8b). (Files: Inr059.fig and Inr053.fig)

Figure 2.4-9: A stereo pair composed of average gray levels for each class.

Figures 2.4-10a and -10b: The adjacency matrix for the classes (-10b) and the new assigned colours (-10b). Six colours were needed. (Files: Inr054.fig and Inr055.fig)

Figure 2.4-11: A The stereo pair of class labelled images "painted" in six colours.

Figure 2.4-12: A magnified version of the stereo pair shown in Fig.2.4-11.



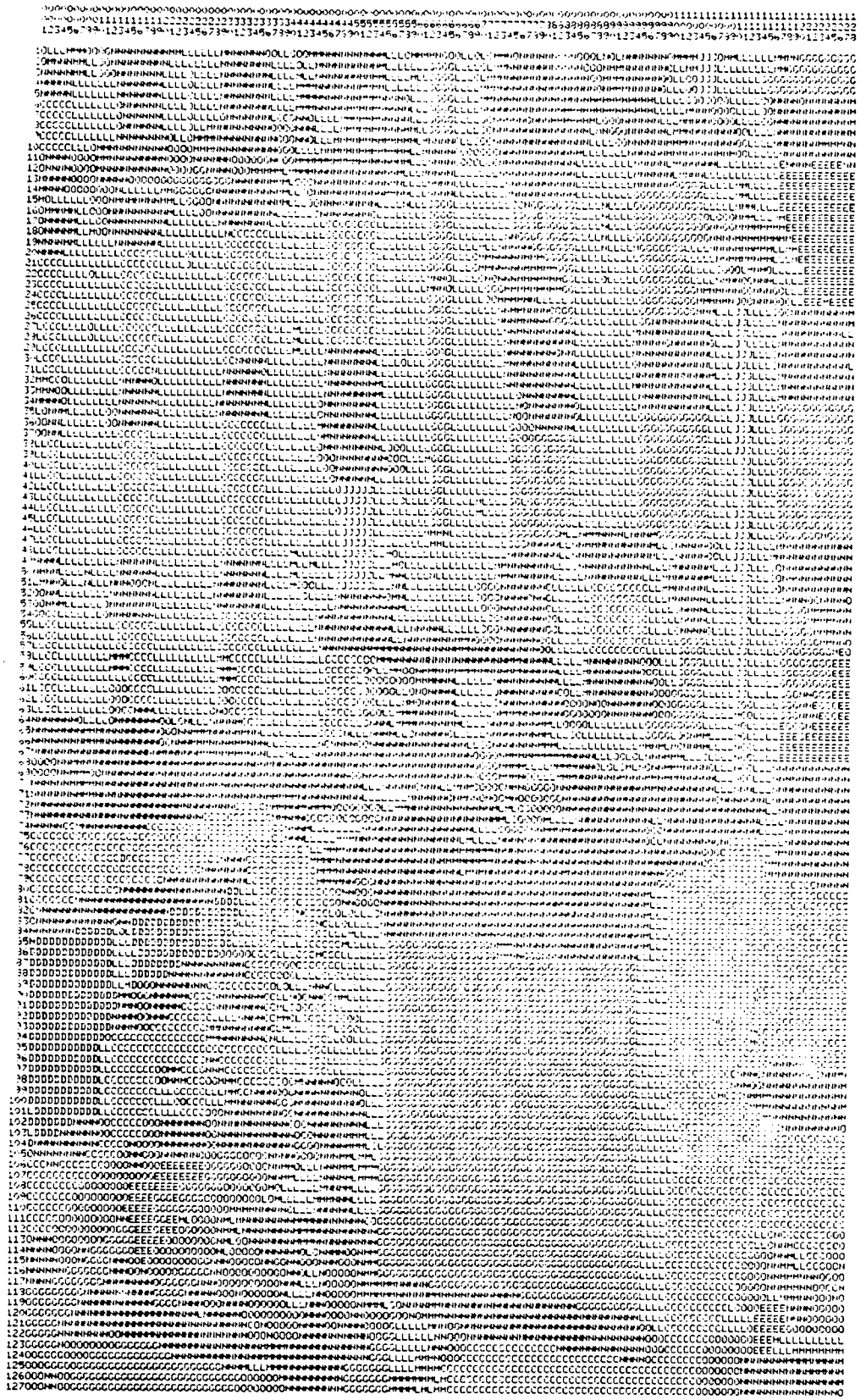
Ln(i,j). left, classified pixels, one cycle of "remove 0-s" and one cycle of "majority label", conn=4, majimal.

LABEL PRINTING

.= 0	#= 1	C= 2	D= 3	E= 4	F= 5	G= 6	H= 7		
I= 8	J= 9	K= 10	L= 11	M= 12	N= 13	O= 14	P= 15		
Q= 16	R= 17	S= 18	T= 19	U= 20	V= 21	W= 22	X= 23		
Y= 24	Z= 25	[= 26	\= 27	]= 28	^= 29	_= 30	.ETC.		
.= 0	#= -1	c= -2	d= -3	e= -4	f= -5	g= -6	h= -7		
i= -8	j= -9	k=-10	l=-11	m=-12	n=-13	o=-14	x=-15		
q=-16	r=-17	s=-18	t=-19	u=-20	v=-21	w=-22	x=-23		
y=-24	z=-25	{=-26	|=-27	}=-28	~=-29	=-30	.etc.		

File: Inm019.Fig

Fig.: 2.4-7a.



Rh(i,j). right. classified pixels. one cycle of "remove -s" and one cycle of "majority label". conn=4. majlim=1.

LABEL PRINTING

:=	;=	<=	==	>=	?=	@=	A=	B=	C=	D=	E=	F=	G=	H=	I=
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
J=	K=	L=	M=	N=	O=	P=	Q=	R=	S=	T=	U=	V=	W=	X=	Y=
16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	,ETC.
.=	/=	0=	1=	2=	3=	4=	5=	6=	7=	8=	9=	:=	;=	<=	==
-1	-2	-3	-4	-5	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15
>=	?=	@=	A=	B=	C=	D=	E=	F=	G=	H=	I=	J=	K=	L=	M=
-16	-17	-18	-19	-20	-21	-22	-23	-24	-25	-26	-27	-28	-29	-30	,etc.

File: Inm020.fig

Fig.: 2.4-7b.

Replace label region by averag gray from gray lev image

K	=	L	R	C	NNFL	NNFR	SSGL	SSGR
1	=	*	*	1	503	525	51.1	50.7
2	=	C	C	1	1445	2386	2.0	0.4
3	=	D	D	1	426	323	15.1	12.7
4	=	E	E	1	14	256	29.6	29.0
5	=	F	F	1	13	1	22.1	22.5
6	=	G	G	1	2823	2378	56.4	54.2
7	=	H	H	1	29	29	81.4	81.4
8	=	J	J	1	221	150	104.3	109.0
11	=	L	L	1	4081	3793	55.6	53.9
12	=	M	M	1	361	393	48.9	47.5
13	=	N	N	1	4170	4004	41.9	40.2
14	=	O	O	1	1570	1419	44.7	42.4

SC.GMAX = 0.20001E+01 0.12750E+03

Average gray levels for classes before post processing.
(Raw classification results)

Inr017.fig = Class labels, left.
Inr018.fig = Class labels, right.

K = Class label.
L = Class label, left image.
R = Class label, right image.
C = 1 if both labels in left and right images, else 0.
NNFL = Number of pixels per class, left image.
NNFR = Number of pixels per class, right image.
SSGL = Average gray level of class, left image.
SSGR = Average gray level of class, right image.

File: Inr053.fig

Fig.: 2.4-8a.

Replace label region by averag gray from gray lev image

K	=	L	R	C	NNFL	NNFR	SSGL	SSGR
1	=	*	*	1	6	1	29.6	21.5
2	=	C	C	1	1503	2485	2.4	[illegible]
3	=	D	D	1	457	319	15.1	12.7
4	=	E	E	1	11	267	29.6	29.3
5	=	F	F	1	2	0	21.5	[illegible]
6	=	G	G	1	2999	2425	56.3	54.1
7	=	H	H	1	33	30	81.5	81.5
8	=	J	J	1	230	156	103.7	107.6
11	=	L	L	1	4376	4061	54.7	53.4
12	=	M	M	1	523	344	51.9	52.1
13	=	N	N	1	4564	4414	43.1	40.9
14	=	O	O	1	1347	1254	44.6	43.1

SC.GMAX = 0.20001E+01 0.12750E+03

Average gray levels for classes after post processing.

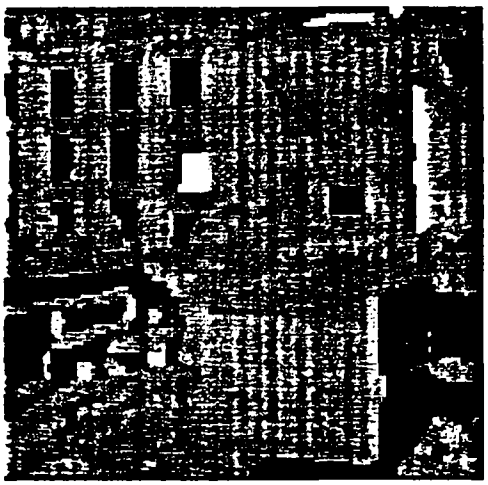
Inr019.fig = Class labels, left.
Inr020.fig = Class labels, right.

K = Class label.
L = Class label, left image.
R = Class label, right image.
C = 1 if both labels in left and right images, else 0.
NNFL = Number of pixels per class, left image.
NNFR = Number of pixels per class, right image.
SSGL = Average gray level of class, left image.
SSGR = Average gray level of class, right image.

File: Inr053.fig

Fig.: 2.4-8b.

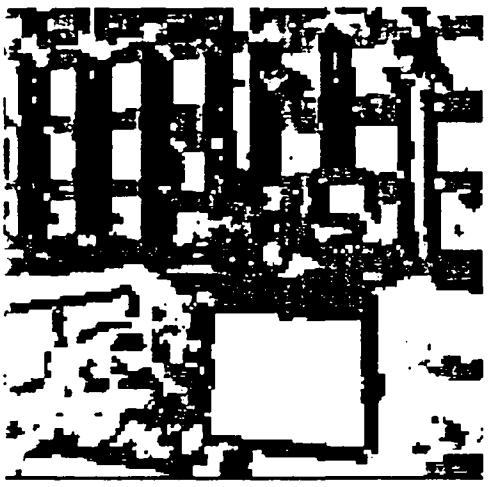
if8hlgry1.lgr



if8hlgry1.rgr

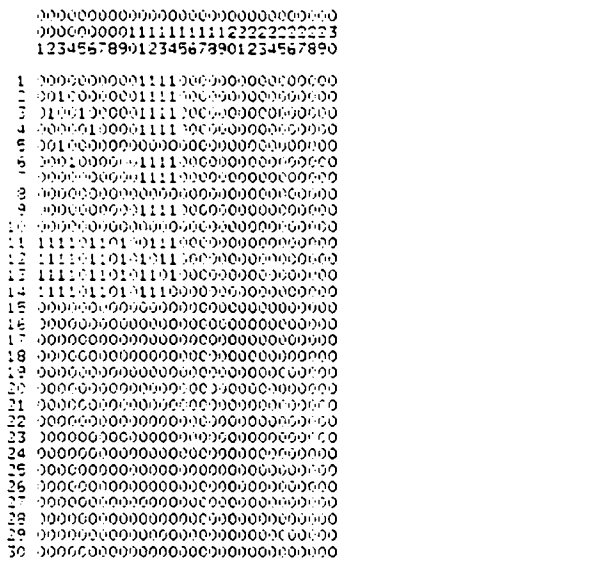
Fig.: 2.4-9.

if8hlgry2.lgr



if8hlgry2.rgr

Fig.: 2.4-11.

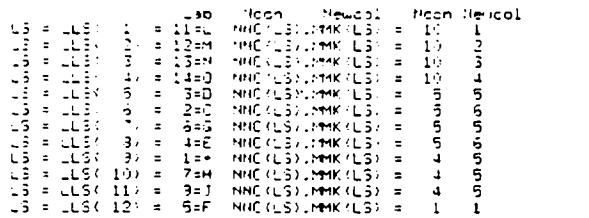


Adjacency matrix for class labelled left and right images.

Im#19.Fig = Class labels. left.
Im#20.Fig = Class labels. right.

File: Im#64.Fig

Fig.: 2.4-10a.



Colour assignments. Six colours were needed.

Lab .. = Class label in images.
Ncon . = Number of contacts with other labels.
Newcol = New colour assignment.

Im#19.Fig = Class labels. left.
Im#20).Fig = Class labels. right.

File: Im#65.Fig

Fig.: 2.4-10b.

if8hlgry2.lgr

if8hlgry2.rgr

Fig.: 2.4-12.

2.5 Analytic approximation and relaxation

The regions or segments produced in the classification stage were of two types, namely, segments with approximately uniform gray levels, and segments with "cylindrically distributed" gray levels. In general, if an analytic approximation is desired, the characteristics upon which the original segmentation was based have to correspond to the form of the analytic function chosen to approximate the gray levels of the segments. In principle, the flat segments may be approximated by first order and the curved segments by a second order polynomials. In the present case this distinction could be made but was not used and all the segments were approximated by a second order polynomial of the form:

$$Ga(x,y) = A + Bx + Cy + Dx^2 + Exy + Fy^2$$

where Ga(x,y) is the analytically computed illumination level of the segment, A, B, C, D, E, F are the coefficients to be computed, x = i, y = -j, and (i,j) are the pixel coordinates in the digitized image.

An analytic approximation of each segment (facet) serves the following purposes:

a) All the pixels belonging to a facet (segment) contribute to the approximation. Thus, the analytic approximation serves as an "umbrella" for representing (the gray levels of) each facet.

b) The "spreading effect" of the high gradient regions can be reduced with "relaxation".

c) Functions computed from the coefficients may be used to recognize the facets (according to their gray levels, gradient directions, etc.).

d) The image may be reconstructed from the analytic representations.

During these processing steps much additional information also becomes easily obtainable, for example, the size of each facet, the local thickness of each facet, numerous facet shape descriptions are possible, facet adjacencies (which facet is the neighbor of which other facet), and so on. The approximation and relaxation process is sketched in Fig. 2.5-1. The details are as follows:

For analytic approximation and "relaxation" the following processing steps are used:

1. Label the (4-connected) regions in the left Lh(i,j) and right Rh(i,j) classified images. This gives a unique "tag" or label for each connected region (facet). Clearly, size and

shape information is now immediately computable for each facet separately.

2. Remove facets which are too small to support a given analytic approximation. This process may be "finessed" by selecting approximations that are suitable for each facet or by removing only facets that cannot be approximated uniquely (for example, if all the pixels are on a straight line).

3. Relabel the remaining facets according to size. This is only a convenience feature to allow the facet labels to be used as addresses during computations and parameter tabulations. The (facet) labelled images are called Lf(i,j) and Rf(i,j) for the left L and right R images, respectively, and the "f" indicates a facet.

4. Compute the coefficients of the analytic approximation for each facet. In the present case the L1 norm was used in order not to overemphasize the large errors. The origin (x=0, y=0) of each facet may be placed at the image corner (i=0, j=0), at the centre of gravity of the facet, or where desired. The procedure consists of: Extract all pixels with a given label k, give these to the approximation program, receive the coefficients and tabulate them as A(k), B(k), etc. It should be noted that this approximation is not an "elastic" approximation (a "spline") of the gray levels. Here the approximation of a facet is independent of any other facets (as well as of its adjacent neighbors).

5. This step is optional but should be carried out to "evaluate by inspection" the accuracy of the approximation, i.e.:

i) Compute the gray level values for the left Gal(i,j) and right Gar(i,j) images from the analytic approximations and display the images. Here the "a" in Ga(i,j) indicates "analytic".

ii) Compute the error images E(i,j) = |G(i,j)-Ga(i,j)| and display the results.

However, it is advisable to carry out these "inspections" after some analytic relaxation of the images to remove minor errors (the regions removed in step 2 leave "holes" in Ga(i,j) which are highly disturbing to our vision).

6. The analytic relaxation procedure is based on comparing the actual gray levels G(i,j) (and their gradients Gx(i,j) and Gy(i,j), if desired) with the analytically computed gray levels Ga(i,j) (and their gradients, if desired). The information on "which pixel (i,j) belongs to which facet k" is available from the corresponding facet labelled image F(i,j) (F(i,j) = Lf(i,j) for the left and F(i,j) = Rf(i,j) for the right image). The parameters are available from the look-up tables A(k), B(k), C(k), D(k), E(k), F(k) for the left and right image, respectively.

In brief, the following "input" information is used for the left and right image, respectively:

G(i,j) = Gray level image.
Gx(i,j) = The x-component of the gradient of G(i,j).
Gy(i,j) = The y-component of the gradient of G(i,j).

F(i,j) = Facet labelled image.
A(k), B(k), C(k), D(k), E(k), F(k) = Parameter look-up tables.

$Gak(x,y) = A(k) + B(k)x + C(k)y + D(k)x^2 + E(k)xy + F(k)y^2.$
Gakx(x,y) = B(k) + 2D(k)x + E(k)y = x-gradient of Gak(x,y).
Gaky(x,y) = C(k) + E(k)x + 2F(k)y = y-gradient of Gak(x,y).

i) Self-relaxation: Clearly, whether or not the accuracy of the analytic approximation deviates "too much" from the actual values is easily detectable by computing the error E(i,j) at pixel (i,j). Thus, in brief:

For all (i,j)
G(i,j) is the original gray level
k=F(i,j) is the label for this pixel
$Gak(x,y) = A(k) + B(k)x + C(k)y + D(k)x^2 + E(k)xy + F(k)y^2.$
E(i,j) = |G(i,j) - Gak(i,j)| is the error
If (E(i,j) > Threshold) then set F(i,j) = 0

The label in F(i,j) is zeroed if the error E(i,j) exceeds some threshold and this pixel becomes "unassigned". The unassigned connected pixel regions are also labelled, and the parameters are recomputed for all the facets. In the present context this is one of the processes that can "bring out more details within a facet" after the left and right facets have been matched.

ii) Cross-relaxation: This form of "relaxation" occurs at the boundaries of the facets. A 3 by 3 neighborhood N with the centre pixel p is sufficient. The centre pixel p is on the boundary of one of the facets. Within the 3x3 neighborhood N3 (4-connected usually) there are pixels with other facet labels.

In brief, the gray level (and its derivatives, if desired) are computed for the centre pixel p from all the pixels (labels) in N that "participate" in the "competition" for assigning its label to the pixel p in F(i,j). The label that gives the best match with respect to G(i,j) (and its derivatives, if desired) is declared the "winner" and it puts its label at location p in F(i,j).

If there are some labels within the neighborhood N3 in F(i,j), two basic situations occur, namely, the label at pixel p = (i,j) is zero (0 = Label(p) = F(i,j)) and the label at p is non-zero (k = Label(p) = F(i,j)).

If Label(p) is 0 but there are non-zero labels in N3, then it is a question of "conquering unopposed territory" and the label that gives the lowest error term is declared the winner provided that the error is less than some threshold T. If T is large then the "best label wins" and after some iterations all the 0-valued (unassigned) pixels in F(i,j) become filled. This method was used to "fill the gaps" in F(i,j) which were created in step 2 since some of the facets were too small to allow a definite analytic approximation.

If Label(p) is not 0 and there are non-zero labels in N3, then it is a question of "conquering opposed territory". As before, the label that gives the lowest error term is declared the winner provided that the error is less than some threshold T. However, now the behaviour depends on how the error term is defined, on the value of T, and repeated iterations need not terminate with "the number of changed pixels equals zero".

"The number of changed pixels" decreases rapidly first but then "slows down" at small values and may even become slightly "oscillatory". The effect is easily explained by considering a "cylindrical" region which (for whatever reason) became approximated by three facets indicated by A, B, and C. A and C approximate the two sides of the "cylinder" while B approximates both sides of the "cylinder". A part of the contact region between facets A and C (indicated by C') may be best approximated by facet B and B starts to "invade" the contact between A and C, see Fig. 2.5-2.

The error term for a facet k at p=(i,j) is actually defined as a sum of gray level and gradient vector errors:

$$Ek(i,j) = |G(i,j) - Gak(i,j)| + W*|G'(i,j) - Gak'(i,j)|$$

where: $|G(i,j) - Gak(i,j)|$ = Abs value of gray level error.
$|G'(i,j) - Gak'(i,j)|$ = Gray level gradient error.
W = Weight to balance the two errors.

With the "best label in N3 wins" strategy, for a small W and large T, the gray levels are approximated as well as possible with the given set analytic coefficients and label distribution but the result may not be best for subsequent matching experiments. In Chapter 3 it is shown how the analytic relaxation process was used after the facets had been matched as well as possible.

However, in order to demonstrate the effects of relaxation, the labelled images in Figures 2.4-7a and -7b were "relaxed" resulting in images shown in Figures 2.5-3a and -3b. The label changes during the relaxation are shown in Fig. 2.5-4 (which is an edited extract of the "history file" during computing). The original parameter lists are given in Figures 2.5-5a and -5b. The parameters have been (in this case) computed with i=0 and j=0 as the origin of the coordinate system. In these lists "K" is the facet label and the "@" in "K=@" indicates the letter used in

printing the facet labelled images (for example, Fig. 2.5-3). The NNEL represents the number of pixels on the facet. NNHL=@ is intended to indicate the corresponding histogram (class) label, but is not used in the present list (set same as K). XX00 and YY00 give the centre of gravity of the facet (XX00=i, YY00=-j). Since the alphabetic code for label printing is modulu 26, the centre of gravity should be used to verify the label code in the image.

In order to show how the image "looks" to the computer the analytic parameters and the labels are used to create an analytic versions G_ana_l(i,j) and G_ana_r(i,j) of the left and right gray level images. The results are shown in Figure 2.5-6 before analytic relaxation and in Figure 2.5-7 after analytic relaxation. Black dots and "squarish" regions are created where the corresponding pixels have no label or the analytic approximation was "indefinite". Clearly, the relaxation process has "sharpened" the image, but it should be remembered that the initial emphasis was on creating regions that are easier to match rather than more pleasant to look at. Histograms of the gray level differences between the analytic and original images (|G_original(i,j)-G_analytic(i,j)|) are shown in Figure 2.5-8. Images of where the major errors occur are shown in "8-level alphanumeric gray" in Figure 2.5-9.

In order to estimate the effects of segmentation on "stereo fidelity" the analytically created gray level images are shown as stereo pairs in Figures 2.5-10a and -10b. The stereo effect is hardly visible in the raw approximation (Fig. 2.5-10a) but improves somewhat after relaxation (Fig. 2.5-10b). Our vision is greatly disturbed by the "noise points" (missing analytic values) and by the artificial gray level contours created in printing the images.

In order to create more "pleasant-looking" images, the analytically obtained gray level images were "corrected" by replacing all pixels in the analytic images with the corresponding gray level values from the original images if there was no analytic value or if the analytic value differed by more than 15 units. The gray level error of 15 was simply the first choice, being about 12% of the median gray. The pixels that were replaced are shown as "white" (.) in the labelled images (Figures 2.5-11a and -11b). Of course, one could compute new analytic approximations for these regions, if desired, and even correct the class label images Lh(i,j) and Rh(i,j). This may improve "matchability" in subsequent processing steps. The resultant "corrected" images are shown in Figure 2.5-12 on a "magnified" scale and as a stereo pair in Figure 2.5-13a. Except for the false gray level contours, most visible on the "desk end", the stereo effect is recaptured. For comparison, the original gray level image pair is shown in Figure 2.5-13b. However, even though the stereo effect has reappeared, this does not really mean very much since the facet boundaries are no longer visible.

Comments:

The analytic relaxation procedure was iterated on the same set of parameters as shown in Figure 2.5-4. In practice, during relaxation some regions may vanish, others become too small to allow proper analytic approximation, and some others may split into two or more separated connected regions (some of which are usually too small, etc.). Depending on the logic used in the relaxation, regions may also become "vacated" since none of the existing approximations apply. Hence, the regions should be relabelled, the small ones eliminated, and new labels computed according to the sizes of the remaining regions. Now the relaxation process may be reiterated. A block diagram of the process was shown in Figure 2.5-1.

In principle, specially if the analytic approximation uses several forms of analytic functions depending on the nature and size of the region (for example, 0-order for very small facets, 1-st order for "planar" regions, 2-nd order for "curved" regions) then the error can be reduced to very small values. However, in practice an additional form of instability, in addition to that illustrated in Figure 2.5-2 is likely to occur, where pixels on the borders of adjacent facets with very similar parameters will simply continue to "exchange labels". The remedy is to merge such regions first and then continue the iterations.

The analytic approximation and relaxation gives a set of coefficients that characterize the gray level distribution (or any other function computed over the image) for each facet. Since the view-point for a stereo pair of images must be different, the gray level distributions (and the parameters) cannot be exactly the same for the same facet in the left and right image. Consequently, the parameters cannot be compared directly, but several invariants may be computed from the parameters, such as the "curvatures" of the gray level distributions (2.4). Time has neither permitted a closer study of the variations between the parameters nor of the invariants.

A final warning is in order, i.e., if the segmentation (labelling, parameter computations, and relaxation) of the left and right image is done independently, there is no guarantee that the regions in the left and right stereo image are "matchable". In the present case the only constraints to ensure "matchability" were the common H(Gpm,G) and H(Gpm,Gpa) decision spaces. However, this is a rather weak constraint to "tie together" the left and right images.

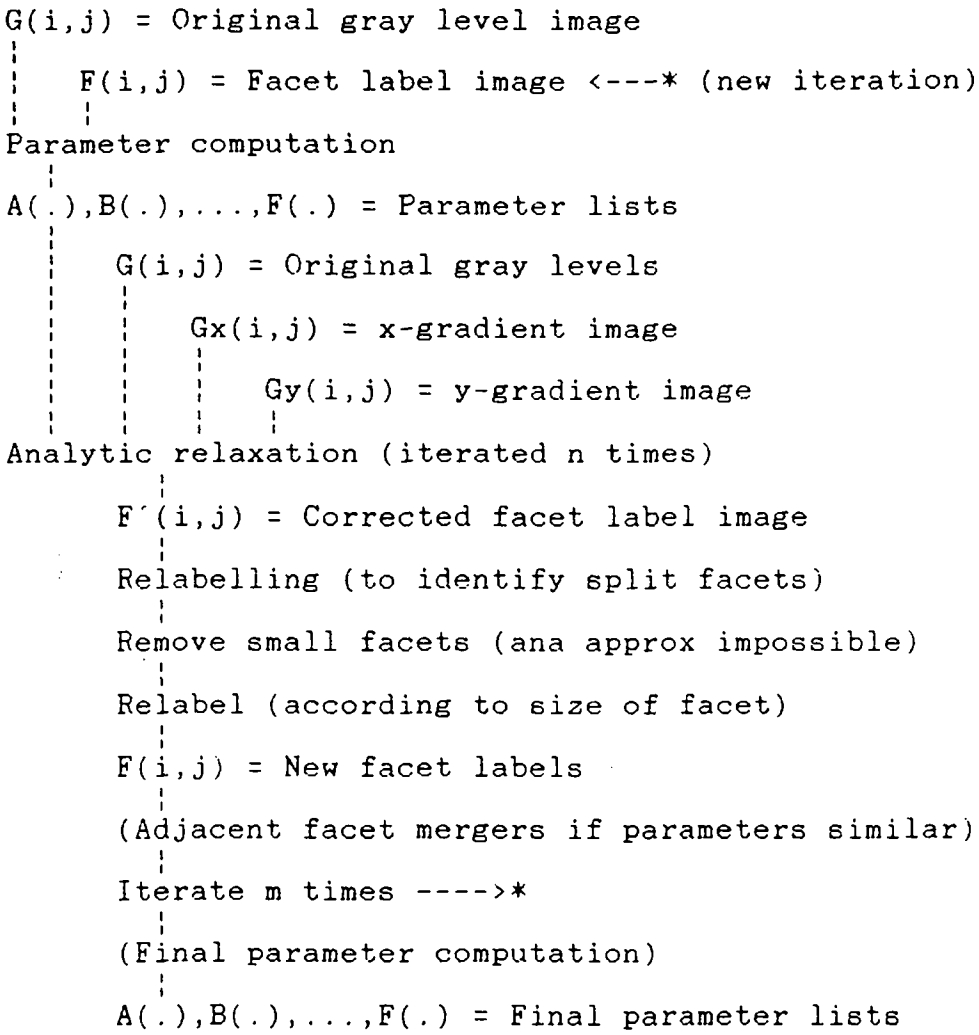


Figure 2.5-1: The analytic approximation and relaxation procedure.

```
A A A A A A A A B B B B B B B B   A A A A A A A A A A B B B B B B
A A A A A A A A A A B B B B B B B   A A A A A A A A A A B B B B B B
C C C C C C´C´C´C´B B B B B B B   C C C C C C B B B B B B B B B B
C C C C C C C C C C B B B B B B B   C C C C C C C C C C B B B B B B
C C C C C C C C C C B B B B B B B   C C C C C C C C C C B B B B B B
```

Before relaxation After relaxation

Figure 2.5-2: An example of instability in analytic relaxation.

Figure titles:

Figure 2.5-3a and -3b: The facet labelled images Lf(i,j) and Rf(i,j) after 10 cycles of iteration on the initial parameters. (Files: Inr045.fig and Inr046.fig)

Figure 2.5-4: An edited extract of the "history record" during analytic relaxation. (File: Inr033.fig)

Figures 2.5-5a and -5b: The initial analytic parameters used in the relaxation process. (Files: Inr029.fig and Inr030.fig)

Figure 2.5-6: The analytically recreated gray level images G_ana_l(i,j) and G_ana_r(i,j) before relaxation.

Figures 2.5-7: The analytically recreated gray level images G_ana_l(i,j) and G_ana_r(i,j) before relaxation.

Figure 2.5-8: Histograms of gray level differences between the analytic and original images (|G_original(i,j) - G_analytic(i,j)|). (File: Inr049.fig)

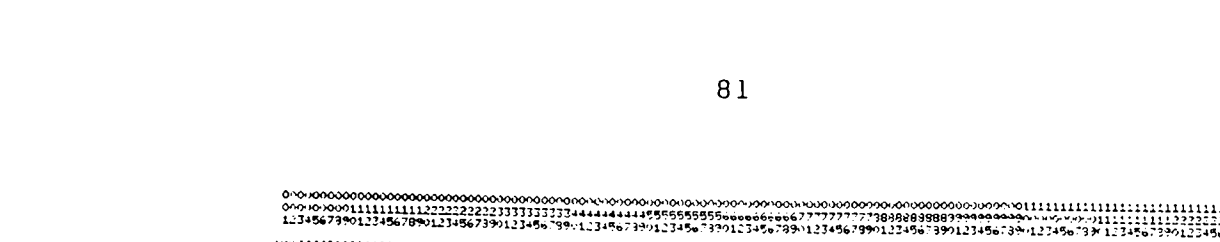
Figure 2.5-9: Images of the major errors between original and analytic gray level images, shown in "8-level alphanumeric gray". (Files: Inr039.fig and Inr040.fig)

Figures 2.5-10a and -10b: The analytically created gray level images are shown as stereo pairs. (a) The "raw" approximation after segmentation. (b) Segmentation after analytic relaxation.

Figures 2.5-11a and -11b: The replaced pixels shown as "white" (.) in the labelled images in order to "correct" the analytic approximations for "easier stereo viewing". The gray level error is 15 units or about 12% of the median gray. (Files: Inr051.fig and Inr052.fig)

Figure 2.5-12: The resultant "corrected" analytic images shown on a "magnified" scale.

Figures 2.5-13a and -13b: Stereo pairs. (a) The "corrected" analytic images. (b) The original gray level image pair shown for comparison.



Lf(i,j). left image after relaxation on initial parameters and ready for new parameter computation.

LABEL PRINTING

.=	)=	==	1=	2=	3=	D=	3	E=	4
F=	5	G=	6	H=	7	I=	8	J=	9
K=	10	L=	11	M=	12	N=	13	O=	14
P=	15	Q=	16	R=	17	S=	18	T=	19
U=	20	V=	21	W=	22	X=	23	Y=	24
Z=	25	[=	26	\=	27	]=	28	^=	29
_=	30	`=	-1	a=	-2	b=	-3	c=	-4
d=	-5	e=	-6	f=	-7	g=	-8	h=	-9
i=	-10	j=	-11	k=	-12	l=	-13	m=	-14
n=	-15	o=	-16	p=	-17	q=	-18	r=	-19
s=	-20	t=	-21	u=	-22	v=	-23	w=	-24
x=	-25	y=	-26	z=	-27	{=	-28	|=	-29
}=	-30	~=	.etc.						

File: Inv345.Fig

Fig.: 2.5-3a.

Ff(i,j). right image after relaxation on initial parameters and ready for new parameter computation.

LABEL PRINTING

a= 0	a= 1	b= 2	b= 3	c= 4	c= 5	d= 6	d= 7		
i= 8	j= 9	k= 10	l= 11	m= 12	m= 13	n= 14	n= 15		
o= 16	o= 17	p= 18	p= 19	q= 20	q= 21	r= 22	r= 23		
s= 24	s= 25	t= 26	t= 27	u= 28	u= 29	v= 30	.ETC.		
a= 0	a= -1	b= -2	b= -3	c= -4	c= -5	d= -6	d= -7		
i= -8	j= -9	k=-10	l=-11	m=-12	m=-13	n=-14	n=-15		
o=-16	o=-17	p=-18	p=-19	q=-20	q=-21	r=-22	r=-23		
s=-24	s=-25	t=-26	t=-27	u=-28	u=-29	v=-30	.etc.		

File: Inr046.fig

Fig.: 2.5-3b.

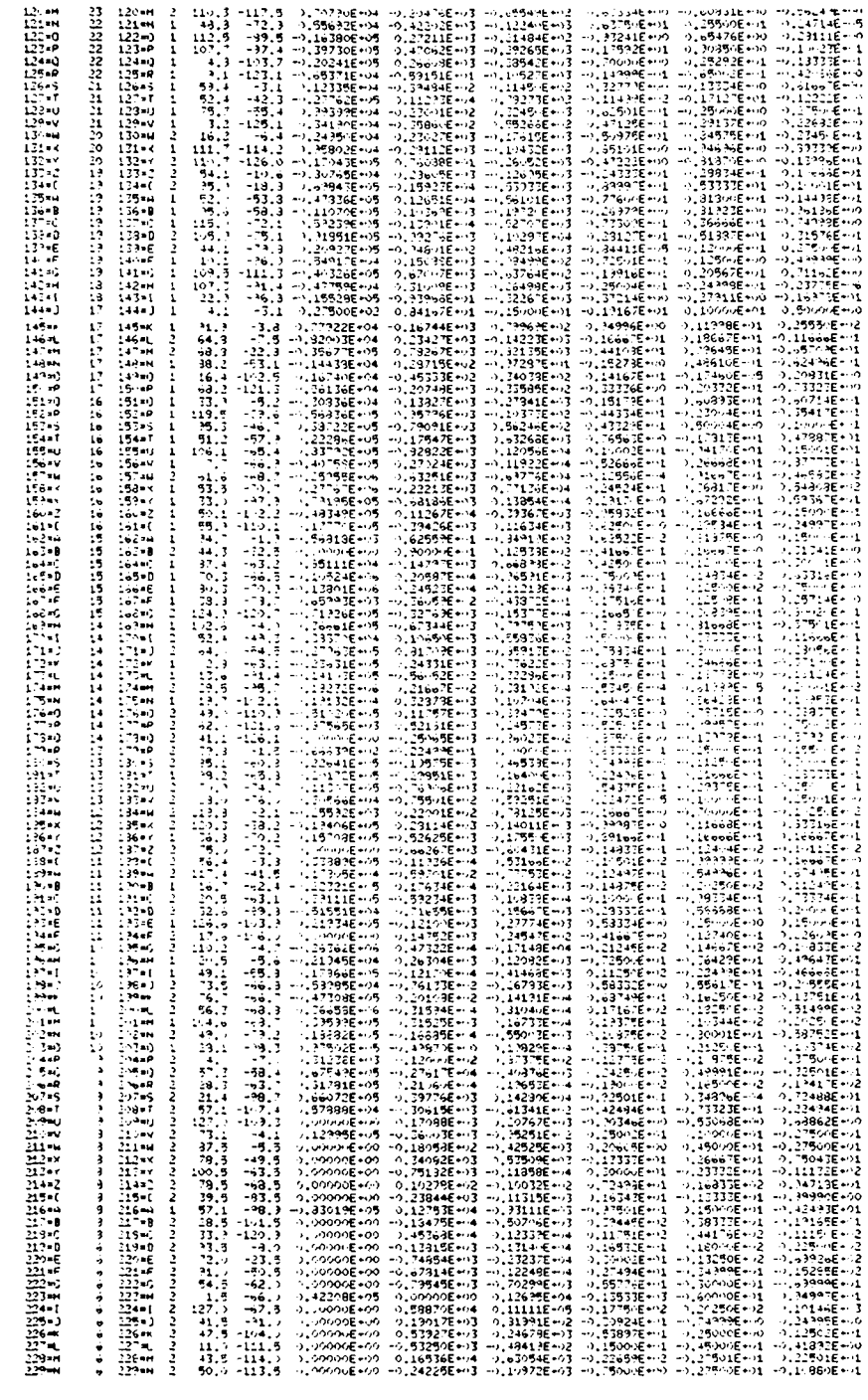
```
CROSS-CLEAN: Analytic relaxation of left image.

IOXFAP.IOXFAC.IOXG.IOXGX.IOXGY.IOXSCC =  36  33  11  37  39  41
MON327R: READ INITIAL SURFACE PARAMETER LISTS: IOFP.KDIM =  35  250
MON120: READ FILE: IVAX.IDIM.JDIM=   33  128  127
MON120: READ FILE: IVAX.IDIM.JDIM=   11  128  127
MON120: READ FILE: IVAX.IDIM.JDIM=   37  128  127
MON120: READ FILE: IVAX.IDIM.JDIM=   39  128  127
MON307A: ANALYTIC CROSS-CLEANING:
IDIM.JDIM.KDIM.NCYCLE.KONECT.IDENOR =    128   127   250    10     4     0
INDEF.IOPT.IPT.JPT.LPOUT.LPEX.NONE.P.LASTHRL =  1    -1     3     0     3     0     0  250
WEIGHTG.WEIGHTD =  0.10000E+01  0.50000E+00
EPLIM.ZMUL .... =  0.25000E-02  0.10000E+03

Iteration cycles on initial parameters:
CYCLE=   1 NUMBER OF CHANGED PIXELS=  0.36430E+04
CYCLE=   2 NUMBER OF CHANGED PIXELS=  0.31000E+03
CYCLE=   3 NUMBER OF CHANGED PIXELS=  0.22600E+03
CYCLE=   4 NUMBER OF CHANGED PIXELS=  0.12500E+03
CYCLE=   5 NUMBER OF CHANGED PIXELS=  0.86000E+02
CYCLE=   6 NUMBER OF CHANGED PIXELS=  0.44600E+02
CYCLE=   7 NUMBER OF CHANGED PIXELS=  0.24600E+02
AUTO-EXIT: TOT TOTNCI=  0.65360E-02
MON307A: TOTAL NUMBER OF PIXEL CHANGED =  0.43580E+04
MON121: WRITE FILE: IVAX.IDIM.JDIM=   41  128  127


CROSS-CLEAN: Analytic relaxation of right image.
IOXFAP.IOXFAC.IOXG.IOXGX.IOXGY.IOXSCC =  36  34  12  38  40  42
MON327R: READ INITIAL SURFACE PARAMETER LISTS: IOFP.KDIM =  36  251
MON120: READ FILE: IVAX.IDIM.JDIM=   34  128  127
MON120: READ FILE: IVAX.IDIM.JDIM=   12  128  127
MON120: READ FILE: IVAX.IDIM.JDIM=   38  128  127
MON120: READ FILE: IVAX.IDIM.JDIM=   40  128  127
MON307A: ANALYTIC CROSS-CLEANING:
IDIM.JDIM.KDIM.NCYCLE.KONECT.IDENOR =    128   127   250     1     4     0
INDEF.IOPT.IPT.JPT.LPOUT.LPEX.NONE.P.LASTHRL =  1    -1     3     0     3     0     0  250
WEIGHTG.WEIGHTD =  0.10000E+01  0.50000E+00
EPLIM.ZMUL .... =  0.25000E-02  0.10000E+03

Iteration cycles on initial parameters:
CYCLE=   1 NUMBER OF CHANGED PIXELS=  0.36040E+04
CYCLE=   2 NUMBER OF CHANGED PIXELS=  0.73800E+03
CYCLE=   3 NUMBER OF CHANGED PIXELS=  0.31400E+03
CYCLE=   4 NUMBER OF CHANGED PIXELS=  0.14000E+03
CYCLE=   5 NUMBER OF CHANGED PIXELS=  0.60000E+02
CYCLE=   6 NUMBER OF CHANGED PIXELS=  0.33000E+02
AUTO-EXIT: TOT TOTNCI=  0.34178E-02
MON307A: TOTAL NUMBER OF PIXEL CHANGED =  0.46430E+04
MON121: WRITE FILE: IVAX.IDIM.JDIM=   42  128  127

File: Inp033.Fig
```

Fig.: 2.5-4.

Left image facet parameters. IOXF+P.IOXFAC = 35 33
G(x,y) = A +B*x +C*y +D*x**2 +E*x*y +F*y**2 : x=1, y=-1
K = Facet label. NNEL = Number pixels. NNHL = Histo label
LH = 1 for valid paras. = 2 for indefinite paras. = 0 none

K=?	NNEL	NNHL=?	LH	XXOO	YYOO	SSAA	SSBB	SSCC	SSDD	SSEE	SSFF
2=C	1394	2=C	1	79.3	-100.7	0.41177E+02	0.88325E+00	0.33295E-01	-0.62086E-02	-0.19250E-02	-0.12937E-02
3=D	1063	3=D	1	31.4	-72.2	0.54612E+03	0.56291E-01	0.13350E+02	-0.16382E-02	0.68995E-01	0.17140E+00
4=E	568	4=E	1	36.1	-35.6	-0.58435E+04	0.32317E+03	-0.39472E+01	-0.45800E+01	0.30125E-01	[illegible]
5=F	527	5=F	1	107.0	-33.3	-0.25672E+05	0.48557E+03	0.34443E+01	-0.22883E+01	-0.10135E+00	-0.13224E-01
6=G	493	6=G	1	18.3	-37.2	-0.13874E+04	0.15402E+03	-0.26021E+01	-0.40931E+01	0.79277E-01	-0.12393E-01
7=H	363	7=H	1	65.4	-32.0	-0.11660E+05	0.35873E+03	-0.11380E+01	-0.27430E+01	0.21377E-02	-0.45711E-02
8=I	358	8=I	1	35.4	-25.2	-0.24239E+05	0.57239E+03	0.46223E+00	-0.33735E+01	-0.26425E-01	-0.16824E-01
9=J	314	9=J	1	11.5	-98.2	0.12357E+03	-0.25527E+00	0.27463E-01	-0.35863E-02	-0.72026E-02	0.15234E-01
10=K	294	10=K	1	99.2	-116.2	0.31756E+03	-0.64740E-01	0.35984E-01	0.12701E-01	-0.35029E-01	[illegible]
11=L	270	11=L	1	17.1	-67.4	-0.11354E+05	0.27637E+02	-0.35277E+03	-0.65939E-01	0.29122E+00	[illegible]
12=M	263	12=M	1	53.9	-28.3	-0.10677E+05	0.39511E+03	-0.62453E+01	-0.36312E+01	0.10205E+00	[illegible]
13=N	250	13=N	2	126.5	-35.2	-0.77218E+04	0.10807E+03	0.43067E+01	-0.36309E+00	-0.51110E-01	[illegible]
14=O	248	14=O	1	3.5	-39.3	0.32534E+02	0.84622E+01	-0.10861E+01	-0.28730E+01	0.67454E-02	-0.13562E-01
15=P	228	15=P	1	102.7	-99.5	0.16319E+05	-0.32049E+03	-0.42613E+01	0.14955E+01	[illegible]	[illegible]
16=Q	201	16=Q	1	24.9	-120.6	-0.50267E+04	0.30731E+01	-0.73851E+02	-0.48253E-02	0.62030E-01	[illegible]
17=R	137	17=R	1	17.5	-112.3	0.77957E+03	-0.10410E+02	0.12963E+02	0.42325E-01	-0.84605E-01	[illegible]
18=S	173	18=S	1	33.2	-113.8	-0.34563E+05	-0.45554E+02	-0.62383E+03	-0.68936E-02	-0.40224E+00	[illegible]
19=T	165	19=T	1	119.3	-95.7	-0.30492E+04	0.11427E+02	-0.15323E+03	-0.32616E-01	0.38720E-01	[illegible]
20=U	163	20=U	1	14.7	-74.7	0.25007E+01	0.22154E-05	0.14308E-05	-0.17388E-08	-0.37253E-07	[illegible]
21=V	158	21=V	1	112.9	-91.4	-0.37153E+03	0.67142E-01	-0.11504E+02	-0.19972E-01	0.25423E-01	-0.31647E-01
22=W	155	22=W	1	113.4	-16.1	-0.36855E+03	0.16774E+02	-0.56932E+01	-0.70862E-01	0.34015E-01	[illegible]
23=X	154	23=X	1	113.4	-59.1	-0.32809E+03	0.50621E+01	-0.32760E+01	-0.28748E-01	-0.27414E-01	[illegible]
24=Y	149	24=Y	1	18.3	-37.5	-0.16388E+03	0.23906E+00	-0.34540E+01	0.23607E-02	0.47358E-02	[illegible]
25=Z	145	25=Z	1	96.9	-25.0	-0.11442E+04	0.24040E+02	-0.17793E+01	-0.12304E+00	-0.17277E-02	[illegible]
26=(	143	26=(	2	106.3	-38.7	0.00000E+00	0.75268E-01	0.12393E+02	-0.63542E-01	-0.13342E+00	[illegible]
27=A	139	27=A	1	59.3	-31.1	-0.32632E+03	0.27464E+02	-0.26750E+01	-0.21389E+00	0.35711E-01	[illegible]
28=B	135	28=B	1	74.1	-6.0	0.27283E+03	-0.63871E+01	-0.26350E+02	0.55661E-01	0.28122E+00	[illegible]
29=C	135	29=C	1	47.3	-96.8	0.37627E+04	0.23254E+02	0.35727E+02	0.36537E-01	0.25663E+00	[illegible]
30=D	124	30=D	1	121.3	-35.3	0.48059E+03	-0.31677E+00	0.27437E+01	-0.43432E-02	-0.22649E-01	[illegible]
31=E	123	31=E	1	96.2	-38.1	-0.31313E+03	0.21547E+02	0.33114E+01	-0.10915E+00	0.74093E-02	[illegible]
32=F	119	32=F	1	113.3	-37.3	-0.12622E+03	-0.48744E+01	-0.30111E+01	0.12963E-01	-0.10545E-01	[illegible]
33=G	112	33=G	1	76.1	-17.7	-0.33359E+03	0.22627E+02	-0.13390E+00	-0.14096E+00	-0.33020E-01	[illegible]
34=H	109	34=H	1	3.2	-109.7	-0.33312E+04	0.63921E+02	-0.52453E+02	-0.35431E+00	0.43960E+00	[illegible]
35=I	109	35=I	1	113.1	-4.3	-0.20048E+04	0.76415E+02	0.57503E+01	-0.25837E+00	0.71401E-02	[illegible]
36=J	105	36=J	1	119.6	-27.0	0.76455E+03	-0.31771E+01	0.37557E+02	0.71683E-02	-0.70777E-01	[illegible]
37=K	104	37=K	1	10.3	-10.7	0.29413E+02	0.20514E+01	-0.11132E+01	-0.21290E-01	0.46834E-01	[illegible]
38=L	98	38=L	1	75.7	-39.0	-0.47762E+03	0.14633E+02	0.45526E+01	-0.10087E+00	-0.22562E-01	[illegible]
39=M	97	39=M	1	48.7	-6.3	-0.27274E+03	0.22239E+02	0.26275E+02	-0.34847E+00	-0.12633E-01	[illegible]
40=N	97	40=N	1	13.3	-125.5	0.29345E+04	-0.13993E+02	0.42217E+02	0.70230E-02	-0.15365E+00	[illegible]
41=O	96	41=O	1	44.3	-19.5	0.14751E+02	-0.11366E+01	-0.10616E+01	0.17793E-01	0.15779E-01	-0.12433E-01
42=P	96	42=P	1	17.3	-20.5	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	[illegible]
43=Q	91	43=Q	1	27.0	-39.0	0.21594E+02	-0.21631E+01	-0.30986E+00	0.50254E-01	0.20107E-01	[illegible]
44=R	90	44=R	1	75.5	-29.1	-0.17269E+03	0.16233E+02	0.27564E+02	-0.36551E-01	0.46174E-01	[illegible]
45=S	90	45=S	1	119.3	-49.4	0.27531E+04	-0.21631E+02	0.39289E+02	0.48634E-01	-0.20457E+00	[illegible]
46=T	88	46=T	1	13.2	-90.0	-0.42441E+04	0.15553E+02	-0.39015E+02	-0.24046E-01	0.17683E+00	[illegible]
47=U	87	47=U	1	49.1	-123.0	-0.64556E+04	-0.77600E+02	0.76752E+02	0.30848E+00	-0.28310E+00	[illegible]
48=V	84	48=V	1	10.0	-21.5	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	[illegible]
49=W	83	49=W	1	75.7	-60.3	0.10391E+05	0.42389E+02	0.33246E+03	0.68745E-01	-0.59142E+00	[illegible]
50=X	83	50=X	1	124.2	-123.3	0.36683E+04	-0.17566E+02	0.42943E+02	-0.31121E-01	0.11567E-01	[illegible]
51=Y	80	51=Y	1	9.9	-39.3	-0.37427E+02	0.33912E+01	0.45934E+00	-0.30985E+00	0.68687E-01	[illegible]
52=Z	80	52=Z	1	75.6	-49.5	-0.32770E+04	0.21636E+02	-0.10035E+03	-0.52231E-01	0.26173E+00	[illegible]
53=(	76	53=(	1	49.5	-105.4	-0.37540E+03	0.47532E+02	0.21503E+01	-0.17757E+00	0.29126E+00	[illegible]
54=A	71	54=A	1	94.4	-14.7	-0.73522E+03	0.17050E+02	0.24137E+01	-0.53677E-01	-0.51357E-01	[illegible]
55=B	71	55=B	1	27.0	-56.6	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	[illegible]
56=C	69	56=C	1	96.3	-6.1	-0.79924E+04	0.16314E+03	-0.30363E+02	-0.32302E+00	0.72967E+00	[illegible]
57=D	69	57=D	1	33.3	-52.3	0.12344E+05	-0.29091E+03	-0.16544E+02	0.16851E+01	0.10210E+00	[illegible]
58=E	69	58=E	1	120.0	-124.3	-0.14256E+05	-0.42979E+02	-0.26633E+03	-0.14238E-01	-0.36735E+00	[illegible]
59=F	67	59=F	1	27.0	-4.4	0.12849E+03	-0.62070E+01	0.47093E+02	0.65594E-01	-0.11649E-01	[illegible]
60=G	66	60=G	1	45.5	-44.0	-0.32139E+02	0.13603E+00	-0.73968E+01	-0.47639E-02	-0.35313E-02	[illegible]
61=H	64	61=H	1	56.3	-38.6	-0.14000E+05	0.46201E+03	-0.13125E+02	-0.40410E+01	-0.10822E+00	[illegible]
62=I	63	62=I	1	9.9	-4.7	0.37095E+02	0.36302E+00	0.23379E+02	-0.11230E+00	-0.29580E+00	[illegible]
63=J	62	63=J	1	45.4	-94.5	0.46362E+04	-0.63497E+02	0.72795E+02	0.21608E+00	-0.53435E+00	[illegible]
64=K	62	64=K	1	33.1	-90.7	-0.31277E+04	0.21720E+02	-0.17476E+03	-0.51502E-01	0.13767E+00	[illegible]
65=L	62	65=L	1	75.9	-104.7	-0.29024E+04	0.23777E+02	-0.34415E+02	-0.13959E+00	0.13907E+00	[illegible]
66=M	60	66=M	1	52.2	-116.0	0.12043E+05	-0.14655E+03	0.13770E+03	0.34227E+00	-0.22302E+00	[illegible]
67=N	59	67=N	1	52.5	-93.4	0.12899E+03	-0.59468E+01	-0.51452E-01	-0.10035E+00	-0.13422E+00	[illegible]
68=O	58	68=O	1	44.3	-28.6	-0.20741E+04	-0.23256E+01	-0.15032E+03	0.41664E-01	0.49995E-01	[illegible]
69=P	58	69=P	1	21.2	-113.9	0.13523E+05	-0.13674E+03	0.21641E+03	0.49944E+00	-0.10450E+01	[illegible]
70=Q	56	70=Q	1	38.3	-46.4	0.17329E+05	-0.86029E+02	0.56094E+03	0.44397E-01	-0.16407E+01	[illegible]
71=R	56	71=R	1	43.6	-58.3	0.32009E+03	-0.25048E+02	0.12197E+02	0.13651E+00	-0.12445E+00	[illegible]
72=S	56	72=S	1	116.7	-108.1	0.33573E+04	0.11124E+02	0.77133E+02	-0.11706E-01	0.72753E-01	[illegible]
73=T	55	73=T	1	35.1	-115.1	0.27524E+05	-0.23102E+03	0.40533E+03	0.70912E+00	-0.15632E+01	[illegible]
74=U	54	74=U	1	37.2	-60.4	-0.34153E+04	0.53896E+02	-0.13613E+02	-0.29067E+00	0.64481E-01	[illegible]
75=V	53	75=V	1	27.1	-11.3	0.44275E+03	-0.39929E+01	0.23857E+02	0.21427E+00	0.12500E+00	[illegible]
76=W	53	76=W	1	26.9	-48.6	-0.13483E+05	0.33609E+01	-0.55558E+03	-0.60009E-01	0.41313E-05	[illegible]
77=X	52	77=X	1	108.6	-123.1	0.30504E+03	-0.71312E+02	-0.47379E+02	0.25514E+00	-0.18301E+00	[illegible]
78=Y	51	78=Y	1	3.4	-55.8	0.13605E+03	0.10424E+02	0.79176E+01	-0.10709E+00	0.14255E+00	[illegible]
79=Z	51	79=Z	2	76.4	-126.2	0.00000E+00	0.50630E+02	0.39020E+02	0.23207E-01	0.47742E+00	[illegible]
80=(	50	80=(	1	45.0	-52.2	-0.27246E+05	0.11545E+03	-0.37703E+03	-0.39050E+00	0.44471E+00	-0.34727E-01
81=A	50	81=A	1	38.7	-110.7	0.23935E+05	0.25542E+02	0.42999E+03	-0.38219E+00	-0.35397E-01	0.13797E-01
82=B	49	82=B	2	43.2	-10.3	0.24003E+02	0.12000E+02	0.23667E+02	0.17381E-06	0.39999E+00	[illegible]
83=C	49	83=C	1	27.2	-29.5	-0.49651E+04	-0.37954E+01	-0.74240E+03	0.71424E-01	-0.39985E-05	[illegible]
84=D	49	84=D	1	95.0	-52.7	0.10401E+05	-0.22354E+03	0.25512E+01	0.10779E+01	-0.47043E+00	[illegible]
85=E	48	85=E	1	113.2	-119.5	-0.51498E+03	-0.23140E+02	-0.37507E+02	0.73013E-01	-0.36008E-01	[illegible]
86=F	47	86=F	1	45.3	-35.8	-0.31531E+04	0.12773E+02	-0.41240E+03	-0.25635E+00	-0.32914E+00	[illegible]
87=G	45	87=G	1	90.2	-58.1	0.57564E+04	-0.21099E+02	0.62368E+02	0.22299E+00	-0.38395E+00	[illegible]
88=H	43	88=H	1	3.3	-48.7	-0.12378E+05	0.49828E+01	-0.31058E+03	-0.45833E+00	-0.12000E+00	[illegible]
89=I	41	89=I	1	9.3	-30.5	-0.39727E+04	-0.44739E+01	-0.26322E+03	-0.11205E-04	-0.16000E+00	-0.47774E-01
90=J	41	90=J	2	35.4	-75.3	0.41480E+03	-0.25148E+02	0.36583E+00	0.14435E+01	0.33381E+00	[illegible]
91=K	41	91=K	1	64.5	-123.5	-0.17631E+05	0.43159E+03	-0.25048E+02	-0.11236E+01	0.27357E+01	[illegible]
92=L	38	92=L	1	9.7	-13.1	-0.77969E+02	-0.73534E+01	-0.23167E+02	-0.72631E-03	-0.22396E+00	[illegible]
93=M	37	93=M	1	31.0	-65.3	0.34831E+04	-0.15992E+03	-0.35361E+02	0.24472E+01	0.23997E+01	[illegible]
94=N	36	94=N	1	63.7	-60.2	-0.44070E+04	0.37809E+02	-0.41070E+02	-0.53293E+00	0.21633E+00	[illegible]
95=O	36	95=O	1	22.7	-38.2	-0.23644E+05	0.16667E+03	-0.49992E+03	-0.33282E+00	0.17752E+01	[illegible]
96=P	36	96=P	1	42.3	-119.9	-0.13739E+05	0.26014E+03	-0.10711E+03	-0.13999E+01	0.17142E+01	[illegible]
97=Q	34	97=Q	1	50.4	-60.9	-0.45342E+04	0.10736E+02	-0.13305E+03	0.73381E-01	-0.24295E+00	[illegible]
98=R	33	98=R	1	84.4	-2.4	0.27670E+03	-0.53570E+01	-0.23071E+02	0.35717E-01	-0.25000E+00	[illegible]
99=S	32	99=S	1	13.1	-2.5	-0.62320E+03	0.77391E+02	-0.73290E+02	-0.22252E+01	-0.15426E+01	[illegible]
100=T	32	100=T	1	112.0	-34.2	0.15192E+04	-0.53658E+02	-0.35774E+02	0.13457E+00	-0.21300E+00	[illegible]
101=U	32	101=U	1	46.3	-30.5	0.61205E+04	-0.26213E+03	-0.72229E+03	0.23036E+01	-0.13176E-04	[illegible]
102=V	31	102=V	2	49.9	-43.7	-0.23456E+05	0.33971E+03	-0.13570E+02	-0.33246E+01	0.48344E+00	[illegible]
103=W	31	103=W	1	44.3	-75.5	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	[illegible]
104=X	30	104=X	1	50.7	-2.3	-0.27611E+04	0.30233E+02	-0.27035E+03	-0.71666E+00	0.45000E+01	[illegible]
105=Y	30	105=Y	1	39.3	-48.5	0.18060E+05	-0.35437E+03	0.11306E+02	0.13763E+01	0.23651E+00	[illegible]
106=Z	30	106=Z	1	48.6	-37.4	-0.31230E+04	-0.14763E+03	-0.11035E+02	0.27500E+00	-0.13000E+01	[illegible]
107=(	29	107=(	1	63.3	-54.3	0.26607E+05	-0.67223E+03	0.10381E+03	0.43510E+01	-0.11570E+01	[illegible]
108=A	29	108=A	1	55.3	-30.7	0.74897E+03	-0.10475E+03	-0.46640E+02	0.11041E+01	0.13973E+00	[illegible]
109=B	29	109=B	1	44.5	-94.0	-0.30347E+03	-0.74737E+01	-0.10044E+02	0.13044E+00	0.43475E-01	[illegible]
110=C	28	110=C	1	24.2	-101.9	-0.21742E+05	-0.93981E+01	-0.41959E+03	0.31015E+00	-0.79593E-01	[illegible]
111=D	27	111=D	1	64.3	-3.7	-0.74448E+04	0.22175E+03	-0.68391E+02	-0.16334E+01	0.10439E+01	[illegible]
112=E	26	112=E	1	46.4	-122.0	0.26736E+05	-0.37937E+03	0.31749E+03	0.77310E+00	-0.23155E+01	[illegible]
113=F	26	113=F	1	118.5	-112.2	-0.30168E+04	0.11027E+03	-0.30324E+02	-0.24208E+00	0.52646E+00	[illegible]
114=G	25	114=G	1	56.8	-114.8	0.69581E+04	-0.11108E+03	0.67029E+02	0.16875E+01	0.65000E+00	0.44591E+00
115=H	24	115=H	1	68.4	-19.5	-0.31392E+04	0.17066E+03	0.24124E+03	-0.13749E+01	-0.38332E+01	[illegible]
116=I	24	116=I	1	41.0	-73.4	0.46726E+04	-0.42775E+02	0.21504E+02	0.66667E+00	-0.20555E+00	[illegible]
117=J	24	117=J	1	33.1	-101.2	-0.60487E+04	-0.23518E+03	-0.20815E+03	-0.70175E+00	-0.32544E+01	[illegible]
118=K	23	118=K	1	59.6	-9.6	0.16658E+04	-0.59352E+02	-0.29575E+02	0.50001E+00	-0.15000E+00	[illegible]
119=L	23	119=L	2	30.3	-81.6	-0.16246E+04	0.73077E+02	-0.14432E+02	-0.13988E+01	-0.32265E+01	[illegible]

Fig.: 2.5-5a, start.



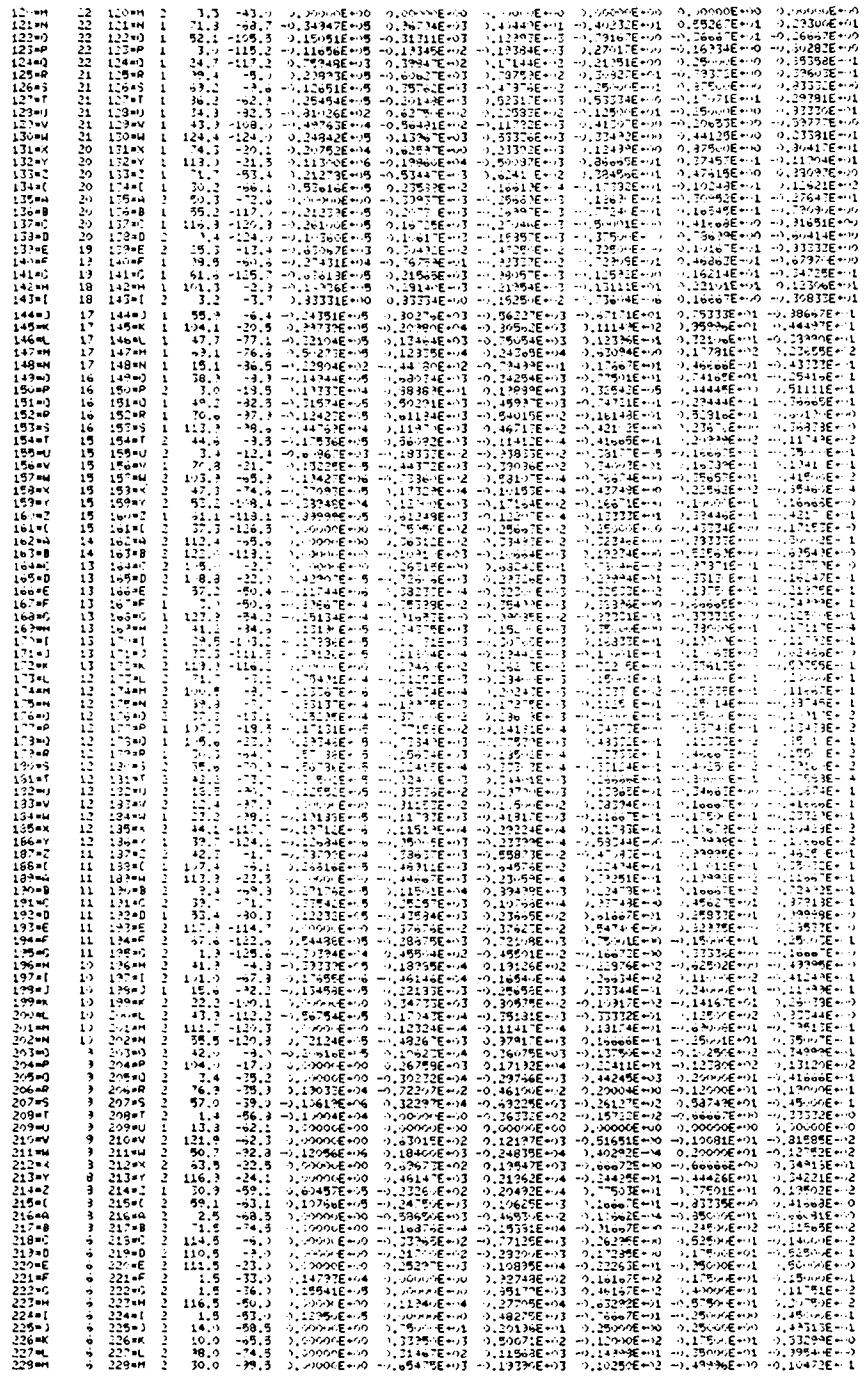
File: Inn29.fig

Fig.: 2.5-5a, end.

Right image facet parameters. IOXFAP,IOXFAC = 36 34
G(x,y) = A +Bx +Cy +Dx**2 +Exy +Fy**2 : x=i, y=j
K = Facet label. NNEL = Number pixels. NNHL = Histo label
LH = 1 for valid paras. = 2 for indefinite paras. = 0 none

K=J	NNEL	NNHL=J	LH	XXOO	YYOO	SSAA	SSBB	SSCC	SSDD	SSEE	SSFF
2=C	1394	2=C	1	63.3	-71.4	0.73218E+03	-0.26463E-01	0.13162E+02	-0.23329E-02	-0.23527E-01	0.10627E+00
3=D	1294	3=D	1	76.0	-102.0	-0.49530E+02	0.55538E+00	-0.21373E+01	-0.47319E-02	-0.13541E-02	-0.12761E-01
4=E	1150	4=E	1	109.0	-101.2	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00
5=F	744	5=F	1	63.6	-32.4	0.11147E+04	-0.36755E-02	-0.63513E+01	0.71601E+00	0.93379E-01	-0.41371E-02
6=G	584	6=G	1	24.0	-87.0	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00
7=H	493	7=H	1	41.3	-41.6	-0.25643E+04	0.47712E-03	0.74426E-01	-0.33234E+01	-0.14767E+00	-0.29724E-01
8=I	475	8=I	1	24.4	-40.0	-0.27206E+04	0.22597E-03	-0.14763E+01	-0.46029E+01	0.34444E-02	-0.28622E-01
9=J	409	9=J	1	38.5	-32.3	-0.25215E+05	0.56312E-03	-0.16586E-02	-0.31374E+01	0.14450E+00	-0.33130E-01
10=K	395	10=K	1	3.3	-40.4	-0.27481E+03	0.67676E-02	-0.14275E+01	-0.27922E+01	-0.13303E-01	-0.17779E-01
11=L	223	11=L	1	6.7	-92.3	0.15442E+03	-0.75277E+00	0.20177E+01	0.14033E-02	-0.10168E-01	0.15743E-01
12=M	208	12=M	1	114.9	-49.7	-0.47869E+05	0.37729E-03	-0.56639E-02	-0.40092E+01	-0.51416E+00	-0.20411E-01
13=N	197	13=N	1	48.2	-113.5	-0.34957E+04	0.43323E-02	-0.44732E+02	-0.20130E+00	0.21174E+00	-0.15816E+00
14=O	194	14=O	1	39.5	-100.6	0.61274E+04	-0.11355E-03	-0.35511E+01	0.34066E+00	0.27489E-01	-0.75654E-01
15=P	191	15=P	1	108.0	-48.0	-0.17495E+05	0.30821E-03	-0.44737E+00	-0.13504E+01	-0.12230E-01	-0.36276E-01
16=Q	174	16=Q	1	73.3	-7.4	0.22294E+03	-0.73770E+01	-0.72567E-02	0.56602E-01	0.11132E+00	-0.15395E-01
17=R	155	17=R	1	122.5	-40.0	-0.10795E+04	0.16946E-02	-0.60641E+01	-0.63529E-01	0.21976E-01	-0.13990E-01
18=S	146	18=S	1	30.7	-120.3	0.12580E+05	0.74902E-02	0.21731E-03	0.37546E-01	0.23667E+00	0.37144E+00
19=T	135	19=T	1	123.3	-17.1	-0.39107E+03	0.16637E-02	-0.34470E+01	-0.63753E-01	0.11723E-01	-0.37673E-01
20=U	129	20=U	1	100.1	-40.4	0.34436E+03	-0.24019E-01	0.34547E+01	0.14267E-01	-0.50709E-02	0.12402E+00
21=V	126	21=V	1	127.1	-73.1	-0.16396E+05	0.50177E-02	-0.28397E+02	-0.75142E+00	-0.30189E+00	-0.20617E-01
22=W	122	22=W	1	50.0	-92.1	-0.29064E+04	0.65512E-02	-0.21616E+02	-0.16329E+00	0.29707E+00	-0.12741E-01
23=X	113	23=X	1	98.6	-19.1	-0.37567E+03	0.13411E-02	-0.47221E-01	-0.34372E-01	-0.16026E-01	-0.21378E-01
24=Y	111	24=Y	1	111.2	-3.0	-0.20481E+05	0.77354E-03	0.49640E-02	-0.17416E+01	-0.43564E+00	0.44719E-01
25=Z	106	25=Z	1	19.0	-90.9	-0.10999E+05	0.32046E-02	-0.15744E-03	-0.31147E-01	0.45664E+00	-0.19040E-01
26=(	105	26=(	1	73.3	-50.3	-0.67680E+03	0.27399E-02	-0.27730E-02	-0.17307E+00	-0.15613E-01	0.42473E-01
27=A	102	27=A	1	50.0	-34.9	0.31067E+04	-0.41093E-02	-0.13605E-03	0.24739E+00	-0.22493E+00	0.16114E-01
28=B	102	28=B	1	29.2	-120.3	0.14572E+05	-0.43761E-02	0.27721E-03	-0.72443E-01	-0.40436E+00	0.37616E+00
29=C	99	29=C	1	15.3	-7.0	0.79315E+02	-0.13054E-01	0.22262E-02	0.23363E-01	-0.32743E+00	0.14154E-01
30=D	98	30=D	1	122.5	-29.1	0.13482E+04	-0.11579E-02	0.32256E-02	0.44444E-01	0.27774E-02	0.63211E-01
31=E	99	31=E	1	100.0	-30.1	0.52101E+03	0.79774E-01	0.38116E-02	-0.71431E-01	-0.20715E+00	0.62541E-01
32=F	97	32=F	1	43.5	-22.4	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00
33=G	96	33=G	1	73.4	-41.5	-0.13666E+04	0.47712E-02	-0.13731E-02	-0.22616E+00	0.33337E-01	-0.66667E-01
34=H	95	34=H	1	13.7	-125.5	-0.16503E+04	0.37495E-02	-0.27613E-02	0.75426E-03	0.25376E-02	-0.11171E-01
35=I	94	35=I	1	24.7	-34.4	0.37956E+03	-0.45572E-01	0.71574E-01	-0.10204E-01	-0.59564E-01	0.33375E-01
36=J	93	36=J	1	32.0	-6.5	0.33326E+03	-0.46229E-02	0.27754E-02	0.67330E+00	-0.76343E+00	0.13324E-01
37=K	93	37=K	1	12.7	-108.9	-0.63596E+04	0.71347E-02	-0.11473E-03	-0.17025E+00	0.75301E+00	-0.46957E-01
38=L	91	38=L	2	111.1	-42.3	0.00000E+00	0.75135E-01	-0.13377E-02	-0.63200E-01	-0.14641E+00	-0.25264E-01
39=M	91	39=M	1	33.0	-42.0	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00
40=N	85	40=N	1	33.0	-23.7	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00
41=O	79	41=O	1	29.5	-110.3	0.11597E+05	-0.83255E-02	0.19407E-03	0.27651E+00	-0.30193E+00	0.77371E-01
42=P	78	42=P	1	51.0	-3.3	-0.63323E+03	0.27366E-02	0.74413E+00	-0.73764E+00	-0.34035E+00	-0.21414E+01
43=Q	78	43=Q	1	7.4	-113.3	-0.22107E+04	0.31277E-01	0.73561E-02	-0.36521E-01	0.68477E-01	0.17079E+00
44=R	77	44=R	1	13.9	-113.4	-0.17947E+04	-0.24613E-02	0.27193E-02	0.53721E-02	-0.21371E+00	0.12786E+00
45=S	76	45=S	1	63.7	-73.3	-0.23075E+04	0.54493E-02	0.21607E+00	-0.72412E+00	-0.78889E-01	-0.22232E-01
46=T	76	46=T	1	37.0	-113.0	0.14339E+05	-0.74241E-02	0.12371E-03	-0.16613E+00	-0.36321E+00	0.77126E-01
47=U	75	47=U	1	121.7	-7.9	-0.26556E+03	0.14227E-02	0.12712E-03	-0.77034E-01	-0.25764E+00	0.37377E-01
48=V	75	48=V	1	127.4	-59.1	-0.15142E+04	0.13629E-02	0.37449E-02	-0.12684E+00	-0.12904E+00	0.13277E-01
49=W	71	49=W	1	16.3	-25.4	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00
50=X	69	50=X	1	121.6	-39.2	-0.14254E+05	-0.55555E-01	-0.13924E-03	-0.15810E-01	-0.12710E-01	-0.15477E-01
51=Y	68	51=Y	1	120.3	-69.4	-0.24413E+04	0.27827E-02	-0.77145E-01	-0.16050E+00	-0.16753E-01	-0.13337E+00
52=Z	67	52=Z	1	16.3	-42.4	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00
53=(	63	53=(	1	32.0	-54.0	-0.38017E+03	-0.60213E-01	-0.72970E-02	0.46402E-01	0.59660E-01	-0.15779E+00
54=A	62	54=A	1	24.0	-4.6	-0.17457E+04	0.14529E+03	-0.10914E-02	-0.29579E+01	0.30876E+00	-0.25629E+00
55=B	62	55=B	1	36.9	-102.5	-0.34873E+03	0.25777E-02	0.39430E-01	-0.53801E-01	0.20931E+00	0.12496E+00
56=C	61	56=C	1	50.3	-54.3	0.15933E+05	0.17912E-02	0.36764E-03	0.20824E-01	0.29645E+00	0.51066E-01
57=D	60	57=D	1	30.9	-18.5	-0.37263E+03	0.11235E-02	0.35417E-01	-0.69270E-01	-0.15510E+00	-0.37161E-01
58=E	58	58=E	1	49.6	-13.3	-0.16592E+03	-0.39626E-01	-0.37795E-02	0.23431E+00	0.14239E-01	-0.13731E-01
59=F	57	59=F	1	33.7	-58.2	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00
60=G	57	60=G	1	124.5	-63.9	-0.28023E+03	0.17557E-02	0.22733E-02	-0.10057E+00	-0.15798E+00	0.68067E-01
61=H	54	61=H	1	21.2	-109.6	-0.61747E+04	0.46996E-02	-0.10766E-03	-0.12727E+00	0.27793E+00	-0.47444E+00
62=I	52	62=I	1	50.9	-46.7	-0.36301E+03	0.17132E-02	-0.12391E-02	-0.15840E+00	0.30577E-01	-0.22296E+00
63=J	51	63=J	1	32.5	-60.3	0.18051E+04	-0.44413E-01	0.13557E-02	0.13322E-01	0.10833E-01	0.25785E+00
64=K	50	64=K	1	31.0	-48.1	0.16483E+05	0.70024E-01	0.67345E-03	-0.20000E+00	-0.53995E+00	0.61374E-01
65=L	50	65=L	1	113.4	-125.7	0.37799E+04	0.42144E-02	0.10009E-03	0.13926E-01	0.37621E+00	0.12161E-01
66=M	49	66=M	1	122.1	-2.7	-0.77651E+03	0.56817E-01	-0.75337E-02	-0.26789E-01	0.51111E+00	-0.22961E-01
67=N	49	67=N	1	3.3	-9.4	-0.27991E+03	0.60750E-02	-0.40590E-01	-0.36714E-01	0.35715E-01	-0.27571E+00
68=O	47	68=O	1	38.3	-63.9	0.64848E+04	-0.14777E-03	-0.17221E-01	0.35046E+00	0.69930E+00	0.61617E+00
69=P	46	69=P	1	113.1	-17.6	-0.79940E+04	0.14988E-03	-0.74446E-02	-0.72802E+00	-0.64752E+00	0.17986E+00
70=Q	46	70=Q	1	102.5	-60.9	-0.46137E+04	0.31774E-02	0.20396E-01	-0.42427E+00	0.76392E-01	0.36227E-01
71=R	45	71=R	1	22.6	-15.2	-0.14524E+04	0.13676E-02	-0.13342E-03	-0.37603E+00	-0.25468E+00	-0.67111E-01
72=S	44	72=S	1	47.7	-3.2	0.46573E+03	-0.17605E-02	0.36534E-02	0.14844E+00	-0.11375E-01	0.36135E-01
73=T	44	73=T	1	28.9	-98.3	-0.13923E+05	0.14111E-03	-0.41339E-03	-0.13165E+00	0.14359E-01	-0.12774E-01
74=U	44	74=U	1	125.7	-117.5	0.65748E+04	-0.46071E-02	0.71492E-02	0.13100E+00	-0.12753E-01	0.31756E+00
75=V	43	75=V	1	33.5	-112.3	0.34599E+05	-0.27752E-02	0.53642E-03	-0.17906E+00	-0.38582E+00	0.15437E-01
76=W	43	76=W	1	109.3	-124.4	-0.36579E+04	0.32616E-02	0.17533E-01	-0.40075E+00	-0.12730E-01	0.15609E-01
77=X	42	77=X	1	103.1	-50.3	-0.20092E+04	0.64346E-02	0.46363E-02	-0.39446E+00	0.34444E-01	0.53069E+00
78=Y	42	78=Y	1	39.7	-65.3	0.30419E+04	-0.67715E-02	0.15434E-03	0.25289E+00	-0.35995E+00	0.34276E+00
79=Z	41	79=Z	2	2.7	-25.9	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00
80=(	41	80=(	1	32.3	-72.9	-0.31646E+04	0.20417E-02	-0.17209E-03	-0.53935E-05	0.58333E+00	-0.23331E+01
81=A	40	81=A	1	17.0	-58.3	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00
82=B	40	82=B	1	48.7	-61.3	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00
83=C	39	83=C	1	35.3	-9.3	-0.15252E+04	0.46813E-02	0.16643E-02	-0.20233E+00	-0.57917E-05	0.79136E+01
84=D	39	84=D	1	15.9	-17.0	-0.64760E+03	0.39441E-01	-0.36332E-02	-0.17422E+00	0.26136E+00	-0.12750E-01
85=E	38	85=E	2	72.7	-51.1	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00
86=F	37	86=F	1	16.1	-51.6	-0.55123E+04	-0.15324E-02	-0.11453E-03	0.42776E-04	-0.29511E-05	-0.16932E-01
87=G	36	87=G	1	104.5	-9.1	-0.42635E+03	-0.63190E-01	-0.11345E-03	0.11740E+00	-0.20732E-01	-0.13321E-01
88=H	36	88=H	1	42.5	-89.6	-0.15213E+05	0.13494E-03	-0.15473E-03	-0.59374E+00	0.15120E-01	-0.11761E-01
89=I	35	89=I	1	69.4	-62.5	0.17583E+04	-0.13927E-02	0.55716E-02	-0.26841E+00	-0.34387E+00	-0.69297E-01
90=J	35	90=J	1	73.3	-67.6	0.10022E+05	-0.15547E-02	0.17721E-03	-0.13317E-01	-0.45312E-01	-0.30971E+00
91=K	35	91=K	1	74.1	-31.2	-0.17517E+05	-0.27612E-02	0.11379E-03	0.12556E+00	-0.13779E+00	-0.16747E-01
92=L	35	92=L	1	27.0	-108.3	-0.11720E+05	0.36122E-02	-0.11356E-03	-0.25001E+00	0.76253E+00	-0.20066E+00
93=M	33	93=M	1	77.6	-22.7	-0.12562E+05	0.15746E-03	-0.15346E-03	-0.13676E-01	-0.65136E+00	-0.64712E-01
94=N	33	94=N	1	75.1	-124.7	-0.39411E+04	0.84416E-02	-0.10713E-03	-0.31221E+00	0.40705E+00	-0.22961E+00
95=O	31	95=O	1	67.3	-9.1	0.36001E+04	-0.22756E-03	0.13906E-02	0.22433E-01	-0.33332E+00	-0.33334E+00
96=P	31	96=P	1	3.7	-13.3	-0.25963E+03	0.35630E-02	-0.16593E-02	-0.12319E-05	0.24022E-01	0.27174E+00
97=Q	31	97=Q	1	73.9	-71.7	0.50166E+05	-0.64257E+03	0.61046E-03	0.20413E-01	-0.34002E-01	0.22034E-01
98=R	31	98=R	1	50.0	-117.6	0.13271E+05	-0.12775E-03	0.27504E-03	0.45407E-01	-0.15994E-01	0.16032E-01
99=S	30	99=S	2	39.0	-5.2	-0.59986E+03	0.12037E-02	-0.15779E-02	-0.37037E-01	-0.62967E+00	-0.27407E-01
100=T	30	100=T	1	55.0	-60.3	0.12340E+05	-0.23366E-03	0.43351E-03	0.69762E+00	-0.26548E+01	0.22976E-01
101=U	30	101=U	1	6.9	-103.9	-0.30139E+05	0.12536E+03	-0.37097E-03	-0.34710E+00	0.13490E+01	-0.27033E+01
102=V	29	102=V	1	63.3	-4.9	0.63401E+03	-0.27945E-02	-0.31936E-02	0.21112E+00	0.46236E+00	-0.12222E+00
103=W	29	103=W	2	16.4	-34.1	-0.57600E+03	0.30915E-04	-0.34000E-02	0.74506E-05	0.14164E-05	-0.50000E+00
104=X	28	104=X	1	63.3	-13.0	0.17172E+04	-0.49400E-02	0.11300E-02	0.26666E+00	-0.16666E+00	0.66667E-01
105=Y	27	105=Y	1	21.7	-10.1	-0.10318E+04	0.44630E-02	-0.33516E-02	-0.62496E-01	0.33333E+01	0.45143E-01
106=Z	27	106=Z	1	46.4	-33.3	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00
107=(	27	107=(	1	24.3	-103.4	-0.16571E+05	-0.37748E-02	-0.33165E-03	0.31666E+00	-0.63998E+00	-0.16370E+01
108=A	27	108=A	1	114.4	-116.4	-0.75584E+04	0.47013E-02	-0.71472E-02	-0.35433E+00	-0.35342E+00	-0.44757E+00
109=B	26	109=B	1	42.3	-101.9	-0.32553E+03	-0.64810E-02	-0.29076E-02	-0.12547E+01	-0.13076E+01	-0.57045E+00
110=C	25	110=C	1	34.3	-2.4	0.29859E+04	-0.62334E-02	-0.61662E+02	0.33334E+00	-0.43233E-04	-0.30000E+01
111=D	25	111=D	1	3.0	-3.0	0.15741E+01	0.40741E+00	0.50000E+00	0.13510E-01	0.55556E-01	0.37037E-01
112=E	25	112=E	1	3.2	-58.3	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00
113=F	25	113=F	1	42.1	-35.0	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00
114=G	24	114=G	1	63.5	-11.5	0.11228E+04	-0.40740E-02	-0.40342E-02	0.34213E+00	0.23348E+00	-0.32105E+00
115=H	24	115=H	1	69.6	-122.5	0.66382E+05	-0.57958E-03	0.50047E-03	0.17500E+01	-0.29167E+01	0.25414E+01
116=I	23	116=I	2	73.4	-3.9	0.32361E+04	-0.43412E-03	0.39416E-02	0.57500E+01	-0.22917E+01	0.11250E+01
117=J	23	117=J	1	58.3	-2.3	-0.13532E+05	0.43277E-03	-0.24111E-02	-0.24583E+01	0.20833E+00	-0.39621E+00
118=K	22	118=K	1	6.3	-1.7	0.31714E+01	0.24121E-01	-0.72333E+01	-0.12917E+00	0.49048E+00	-0.30952E+00
119=L	22	119=L	2	1.3	-43.0	-0.62858E+01	0.00000E+00	-0.66666E+00	-0.17222E+01	0.33333E+00	0.31614E-07

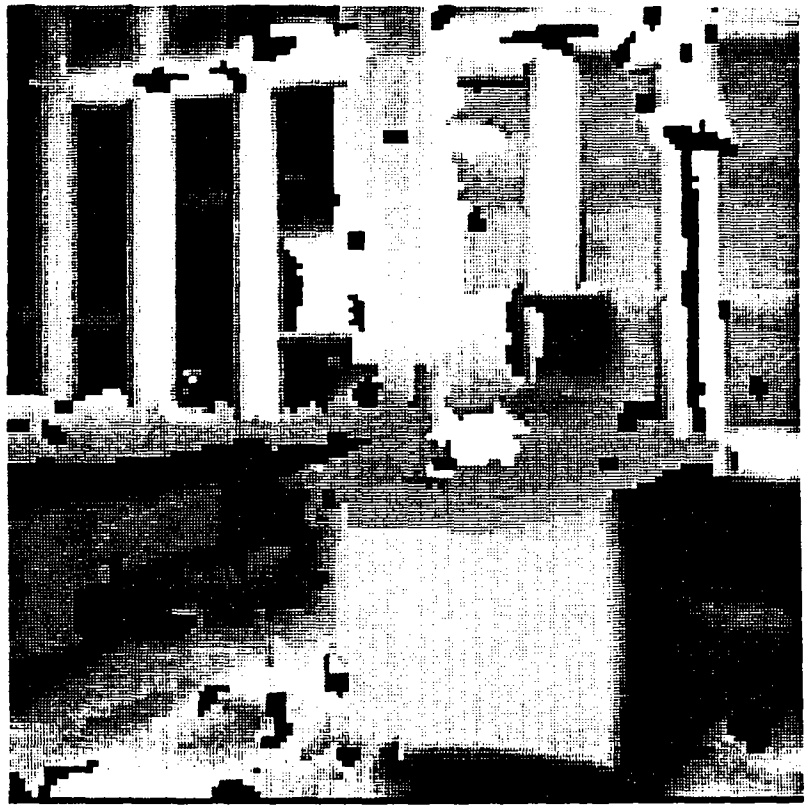
Fig.: 2.5-5b, start.



File: Inn030.Fig

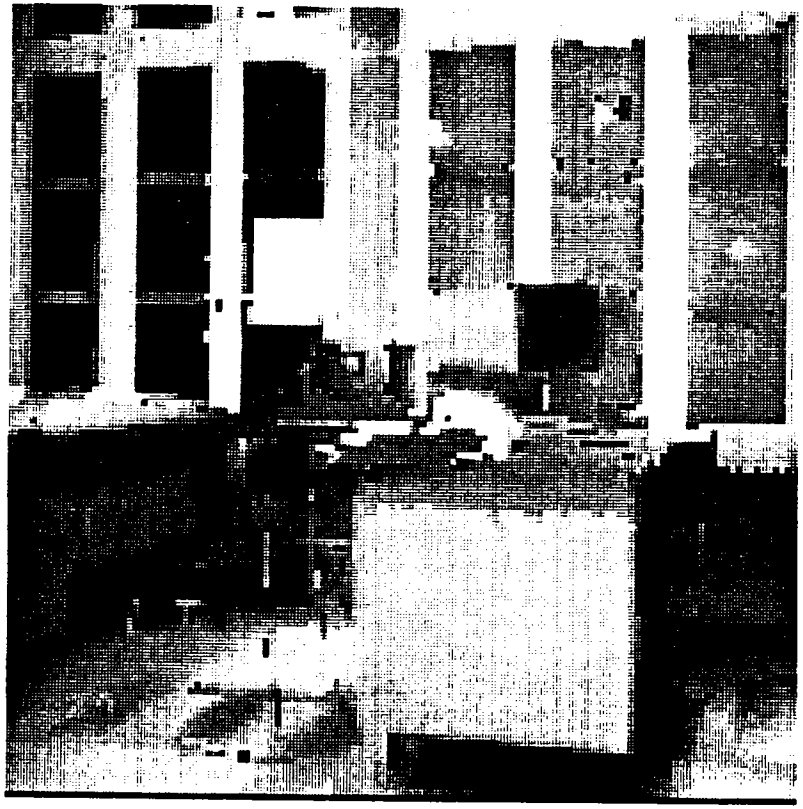
Fig.: 2.5-5b, end.

if8ganal.lgr

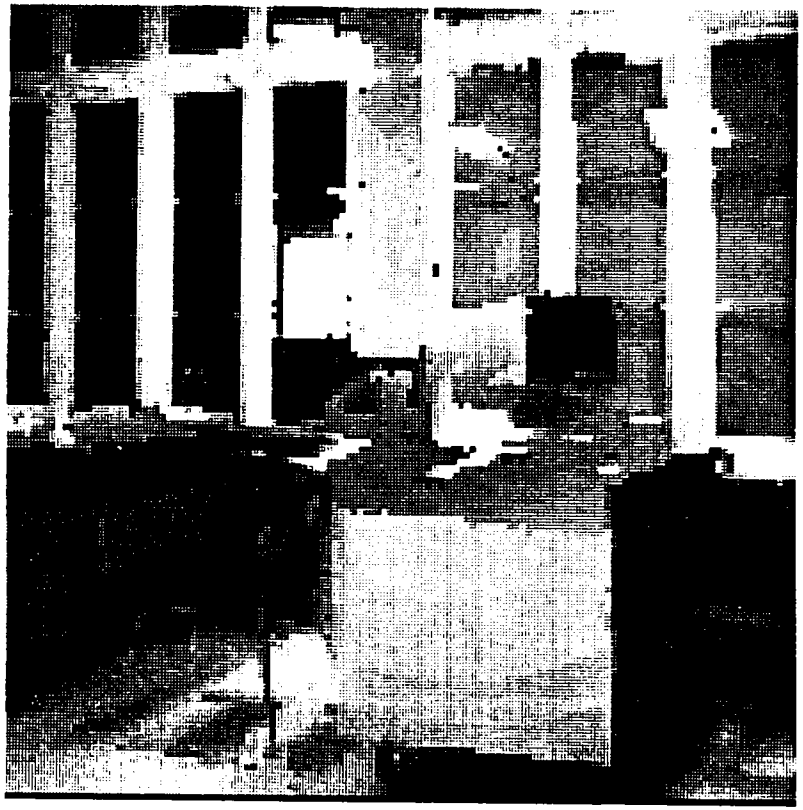


if8ganal.rgr

Fig.: 2.5-6.



if8gana2.lgr



if8gana2.rgr

Fig.: 2.5-7.

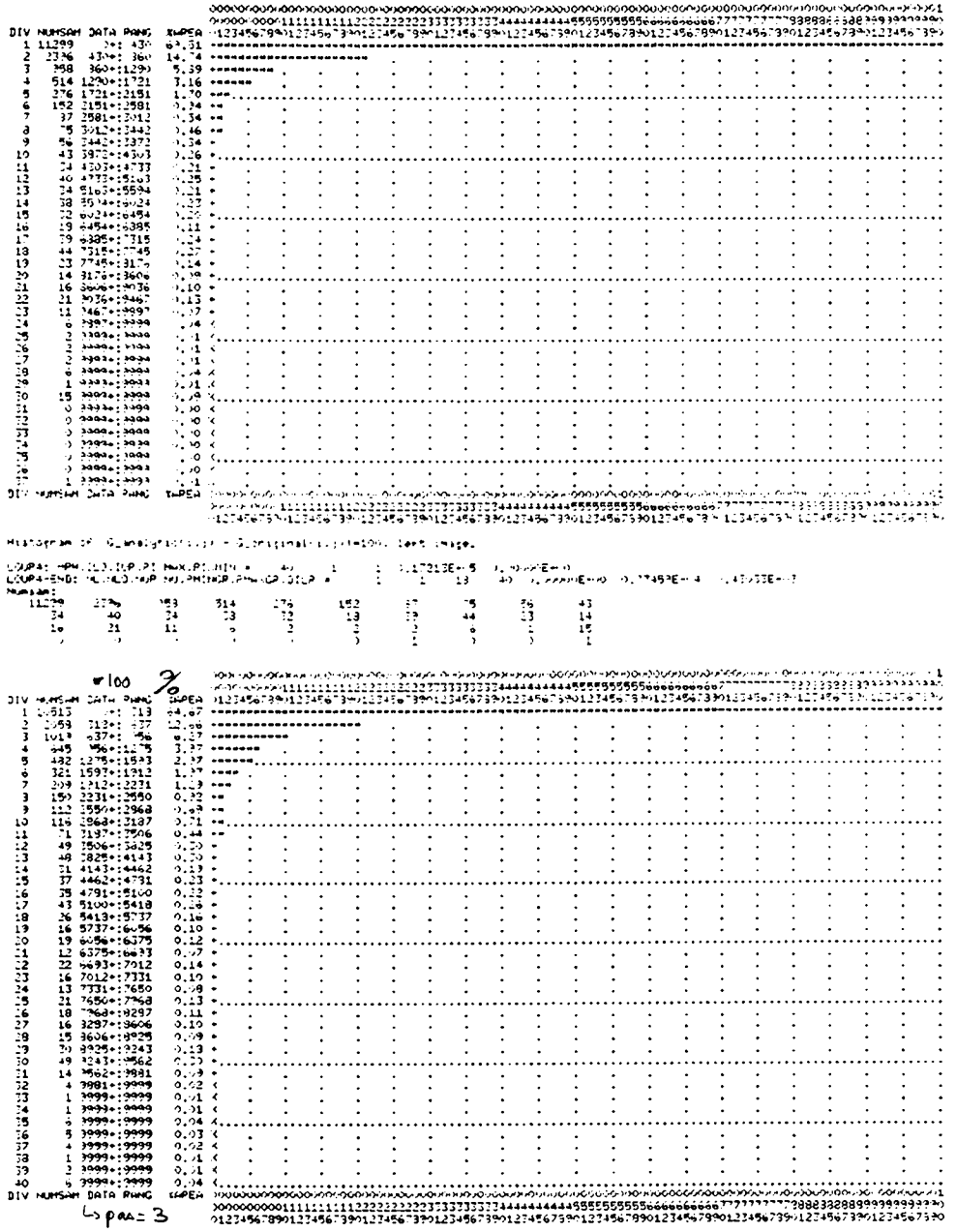


Fig.: 2.5-8.

$|g_{ana} - g_{sim}| * 100$

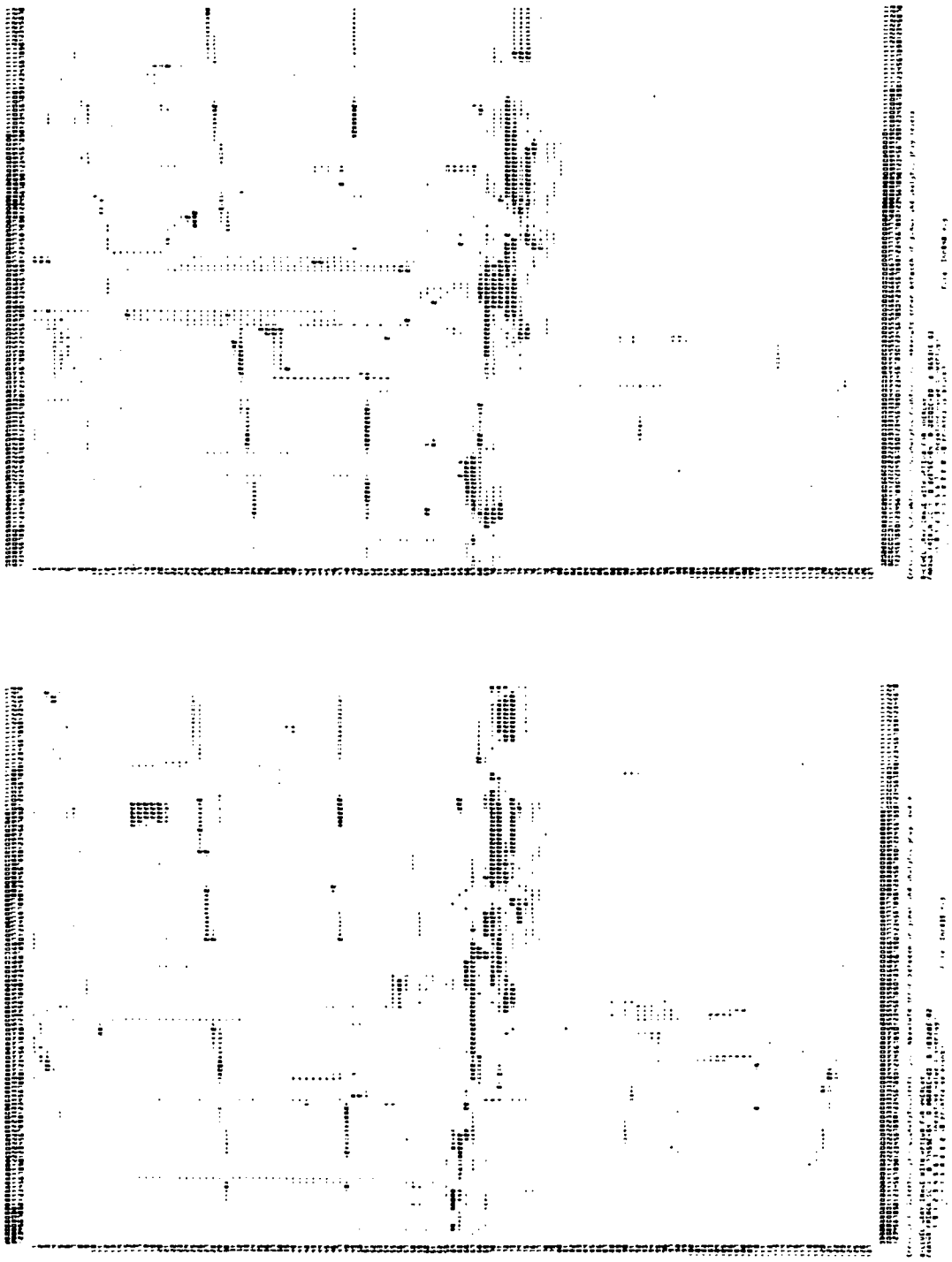


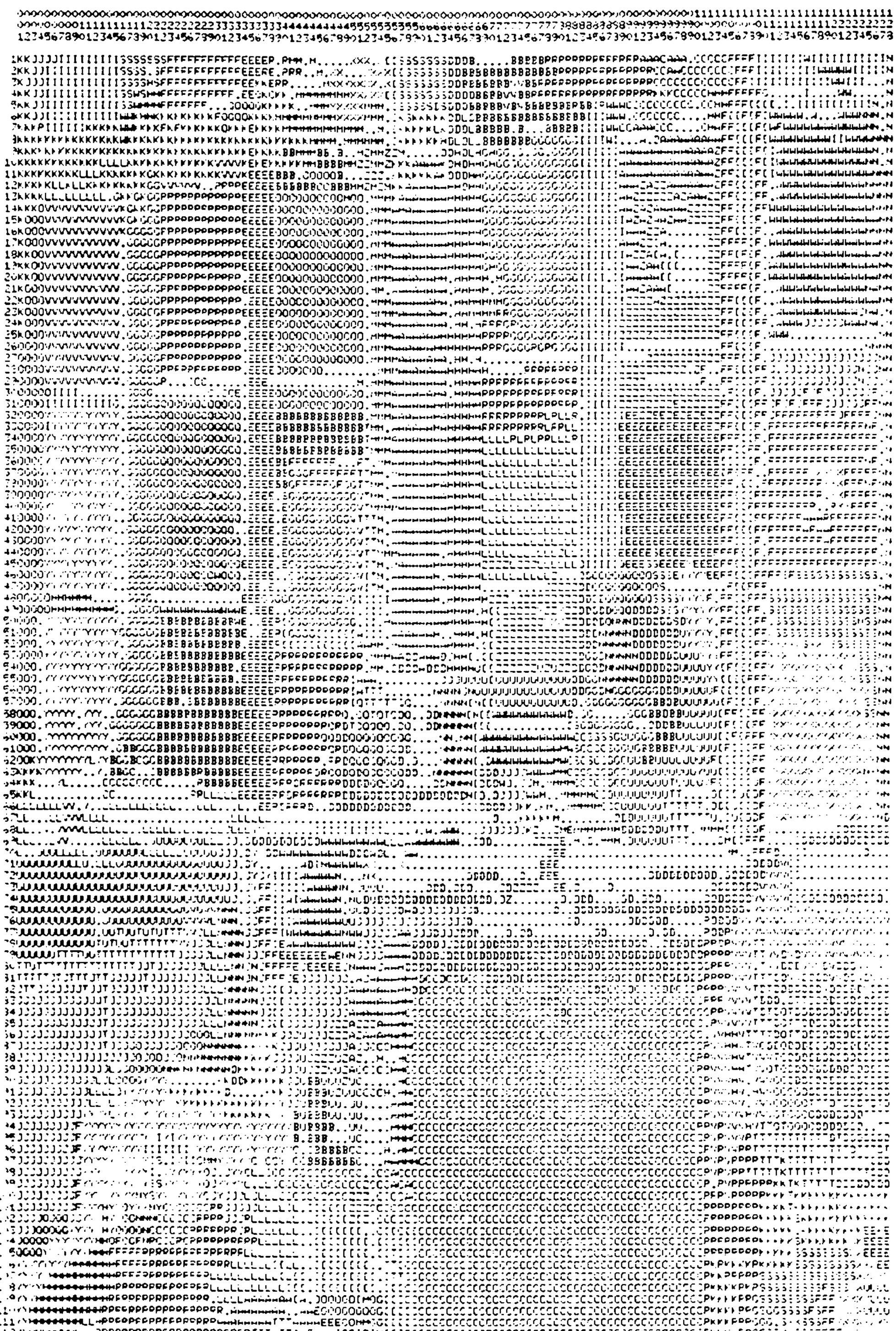
Fig.: 2.5-9.

if8ganal.lgr if8ganal.rgr

Fig.: 2.5-10a.

if8gana2.lgr if8gana2.rgr

Fig.: 2.5-10b.



LF(i,j). left. Corrected labels for error limit = 15.0. [IF(.G(...-Gana(...,1.Gt.Erlim) LF(i,j)=0]

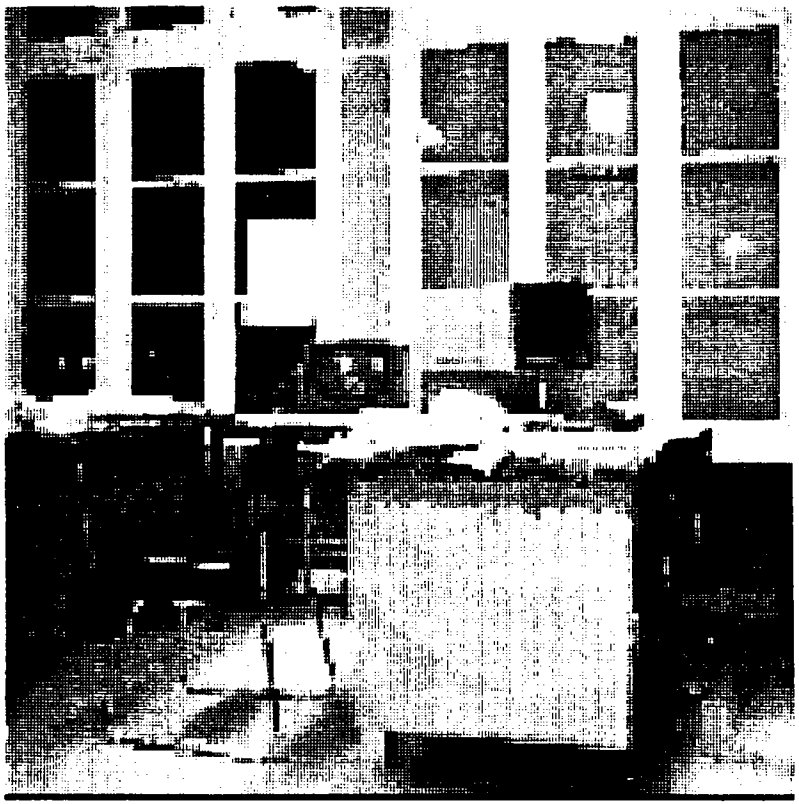
LABEL PRINTING

.= 0	*= 1	C= 2	D= 3	E= 4	F= 5	G= 6	H= 7
I= 8	J= 9	K= 10	L= 11	M= 12	N= 13	O= 14	P= 15
Q= 16	R= 17	S= 18.	T= 19	U= 20	V= 21	W= 22	X= 23
Y= 24	Z= 25	(= 26	A= 27	B= 28	C= 29	D= 30	.ETC.
.= 0	a= -1	c= -2	d= -3	e= -4	f= -5	g= -6	h= -7
i= -8	j= -9	k=-10	l=-11	m=-12	n=-13	o=-14	p=-15
q=-16	r=-17	s=-18	t=-19	u=-20	v=-21	w=-22	x=-23
y=-24	z=-25	(=-26	a=-27	b=-28	c=-29	d=-30	.etc.

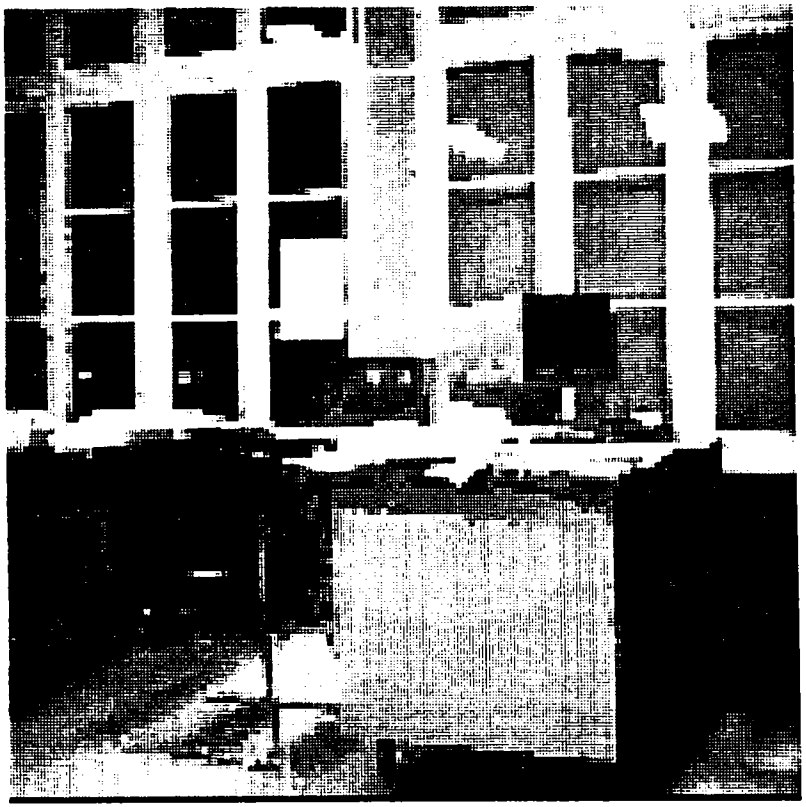
File: Inr51.Fig

Fig.: 2.5-11a.

Fig.: 2.5-11b.



if8gana3.lgr

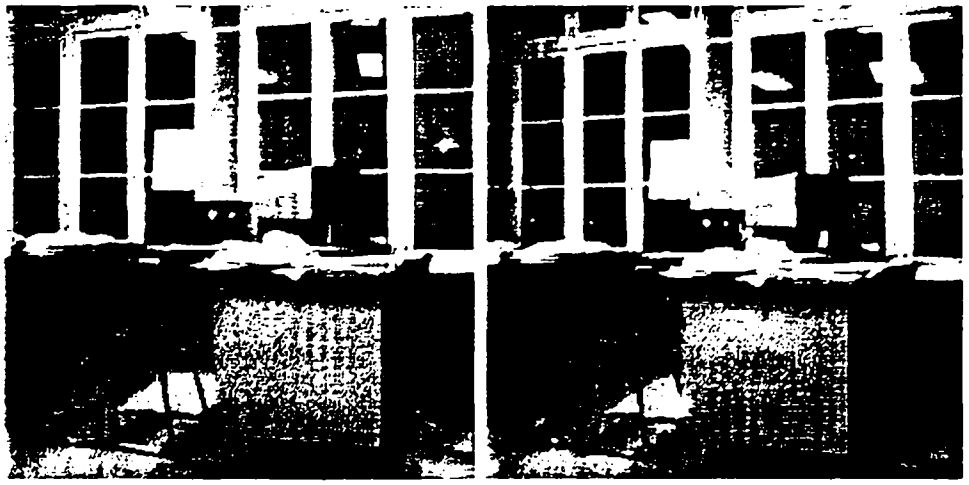


if8gana3.rgr

Fig.: 2.5-12.

if8gana3.lgr if8gana3.rgr

Fig.: 2.5-13a.



i001.log i001.rog

Fig.: 2.5-13b.

2.6 Super-regions versus scale-space

The initial segmentation described so far is dependent on the contents of the original image (image data), on the processes used to extract additional information (gray level gradients in the present case), on the size of the local operator (N), and on the decision spaces and the clustering procedures used (and to some extent on the "post processing" and on "analytic relaxation", if used). Of course, all the variables are kept constant when the left and right image are processed, but even then the segments produced are only "relatively similar" in the left and the right images of the stereo pair but the similarity is by no means "perfect". The differences occur since in some parts of the images the gray level structure is very different between the left and right image due to illumination changes, reflections, transparencies, or what is visible to one "eye" is not visible to the other due to the geometry of the scene. Thus, a certain amount of "miss-match" between the left and right segmented images is to be tolerated.

At least according to the currently popular opinions, if the size (N) of the local operators is increased then the results are expected to represent increasingly larger regions in the images (the "scale-space" effect). If N is "large enough" then the regions will be "large enough" and matching of the corresponding regions in the left and right images should become increasingly simpler. There is, of course, some truth in this argument, but we should remember that the scale-space approach only amounts to "smearing the information" in the images and is not likely to be the best method.

A more elegant method consists of not destroying the spatial resolution by increasingly larger local operators (N) but rather continuing the hierarchical approach based on the facets already found. There are numerous possibilities which, briefly summarized, are:

a) Continuing the clustering of non-adjacent similar facets. This actually has to be done if there are structures in the image composed of similar facets since such facets cannot be matched correctly between the left and right images without some additional distinguishing information. A similar problem was once studied in connection with "linear textures" (2.5).

b) Using adjacent facets to create "new recognition features" since two adjacent facets A and B in one image are also very likely to be adjacent in the other image.

In both cases "super-facets" can be created without destroying the spatial resolution. Time has not permitted experiments with these methods up to the present.

2.7 Conclusions

In this chapter some of the preliminary image processing and analysis steps for stereo vision have been experimented with in order to gain some practical experience with this problem. The image pair used is rather general and, probably, contains a fair example of the problems one may encounter with most gray level images. Exceptions, such as, images consisting only of textures, images of thin line structures, semitransparent scenes, etc., have not been investigated.

As briefly indicated, a very large number of pixel-level features are computable from the stereo pair of images, i.e., it is possible to extract much information from the images upon which to base segmentation. Of course, the same operations, in the same sequences, and with the same parameters, should be used when the two images are processed for pixel features. Hence, it is largely a question of willingness on our part to program the necessary procedures and to produce the computing power (special hardware) to make it practical.

If the original images differ, as is the case with the present pair of images, some preprocessing may be required to "equalize" the images (but this does not mean "histogram equalization"). However, if the information is lost due to saturation, then it is beyond recovery. In the present images this created some interesting problems in matching, see next Chapter. In regions where the pixel level features can be computed, clearly, for stereo vision to be possible, the images of these features must preserve "stereo fidelity", when we look at such a pair of images through a stereo viewer we have to be able to see "depth". Since our stereo vision ability is very powerful, if it exists at all, if an impression of depth cannot be seen then, clearly, there is "something seriously wrong" with the features and it is unlikely that further computing on such features has much meaning for stereo. The features used in the present study could all be "seen in stereo".

It should also be remembered that the segments in themselves do not need to "mean anything" in human terms, even though it is desirable if they do and they often do have a meaning if understood in terms of what the mathematical procedure is doing. In the stereo problem the only important criterion is "similarity" of the segments between the left and right images, and nothing else! Any level of merging is feasible if an hierarchical approach is used. For example, the experiments with the "automatic volume control" filter indicated that much detail can be "brought out" for continued matching on a finer scale. Visual inspection indicated that some of the "texture" brought out preserved texture integrity while, basically, these surfaces had no "intentional texture". In an active ("live") vision system such classification is carried out for each "glance" at the scene.

In the present case, the operations so far carried out and the sequences in which they have been carried out are summarized in the block diagram shown in Figure 2.7-1. The places where the stereo pairs have been created for visual verification of the results are shown in Figure 2.7-2. The abbreviations are explained in Figure 2.7-3.

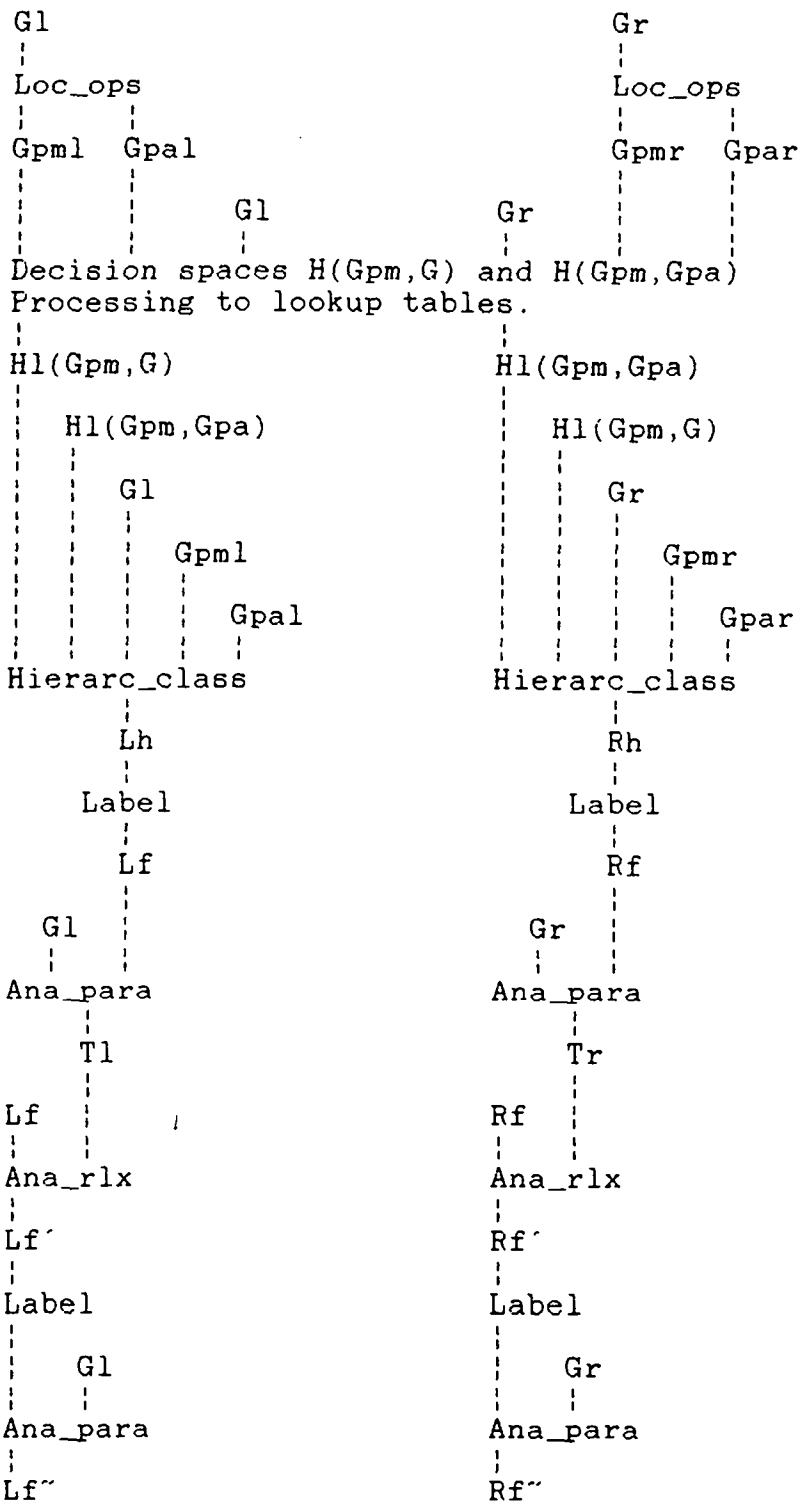


Figure 2.7-1 The basic analysis process. The abbreviations are listed in Fig. 2.7-3.

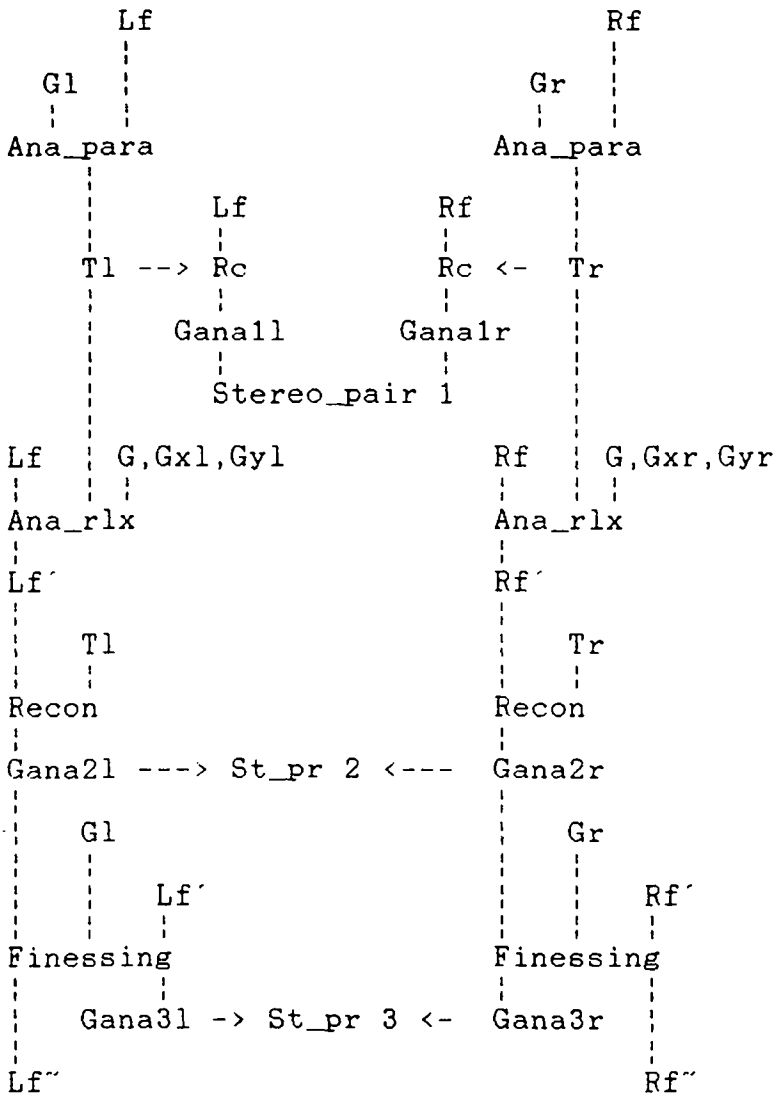


Figure 2.7-2: Stereo pair construction for visual verification. The abbreviations are explained in Fig. 2.7-3.

*l = Associated with left image.
*r = Associated with right image.
G = Gray level image.
Gl = Gray level image of stereo pair, left.
Gr = Gray level image of stereo pair, right.
Loc_ops = Local operators within neighborhood N.
Gpm = Gradient magnitude.
Gpml = Gradient magnitude, left.
Gpmr = Gradient magnitude, right.
Gpa = Gradient angle.
Gpal = Gradient angle, left.
Gpar = Gradient angle, right.
H(Gpm,G) = Decision space on Gpm and G for left and right.
H(Gpm,Gpa) = Decision space on Gpm and Gpa for left and right.
Hl(Gpm,G) = Lookup table for classifying Gpm and G.
Hl(Gpm,Gpa) = Lookup table for classifying Gpm and Gpa.
Hierarc_class = Hierarchical classification.
*h = Class labelled image.
Lh = Class labelled image, left.
Rh = Class labelled image, right.
Label = Region labelling, 4-connected.
*f = Facet (region) labelled image.
Lf = Facet (region) labelled image, left.
Rf = Facet (region) labelled image, right.
Gxl,Gyl = Gray level gradients, left.
Gxr,Gyr = Gray level gradients, right.
Ana_para = Analytic approximation parameters for facets.
Tl = Analytic parameter table, left.
Tr = Analytic parameter table, right.
Ana_rlx = Analytic relaxation.
Lf′ = Facet (region) labelled image after relaxation, left.
Rf′ = Facet (region) labelled image after relaxation, right.
Rc = Reconstruction of image from analytic parameters.
Recon = Reconstruction of image from analytic parameters.
Gana = Analytically reconstructed image.
Ganal1 = Analytically reconstructed image, 1′st, left.
Ganalr = Analytically reconstructed image, 1′st, right.
Stereo_pair = A stereo pair of images shown in text.
St_pr = A stereo pair of images shown in text.
Gana2l = Analytically reconstructed image, 2′nd, left.
Gana2r = Analytically reconstructed image, 2′nd, right.
Finessing = Replace large errors in Gana by G values.
Gana3l = Analytically reconstructed image, finessed, left.
Gana3r = Analytically reconstructed image, finessed, right.
Lf″ = Facet (region) labelled image, finessed, left.
Rf″ = Facet (region) labelled image, finessed, right.

Figure 2.7-3: Abbreviations used in Figures 2.7-1 and 2.7-2.

Chapter 3: FACET MATCHING

3.1 Introduction

Successful matching of the left (L) and right (R) image detail of a stereo pair is absolutely essential for "still-life" stereo-based depth vision. The match has to be carried out to pixel level, that is, exact (unique and exclusive) correspondence has to be established for all the pixels in the L and R images which represent regions in the scene visible to both "eyes". Facet matching cannot accomplish this alone, being only a step in an essentially recursive or iterative procedure (from larger facets to increasingly smaller until the facet becomes the size of the pixel and sub-pixel). A vision system that only produces a partial match will leave "unknown" regions in the scene.

For a "live" vision system the fact that a part of the scene is "unknown" or that a facet does not "match", is in itself information (a "centre of interest") to guide the "eyes" to such a location for "taking another look", until the whole 3D scene has been properly reconstructed from multiple "glances" at the scene. Even if the "scene" is only a stereo pair of images viewed through a set of stereo glasses, the "live" system can still "keep searching" until a complete match has been obtained. It is thus conjectured that a "live" vision system that can "move its eyes" (centre of fixation) to any desired place in the scene has a far easier problem to solve than the "still-life" stereo system.

In a "still-life" stereo system only two images are available and the "eyes" as such do not exist. We know from our own experience that such images can be seen "in stereo". However, we can still move our eyes during viewing. The author is not aware of any experiments to verify whether a stereo pair of "fixated images" can be fused by our visual system, but the answer is likely to be affirmative. This question has only been raised to alert the reader to the unlikely possibility that "still-life" stereo vision may not be a very realistic problem, except in photogrammetry.

However, be as it may, this chapter describes several experiments in attempting to match the facets in a left and right "still-life" stereo pair of gray level images. As was demonstrated in the previous chapter, there is no lack of information (individually, i.e., separately) computable from the two images upon which to base the matching experiments. As a brief summary, the information consists of:

a) Pixel level features belonging to each pixel (gray level, colour, and flicker). In the present case only the gray level is available.

b) Pixel level features obtained from local operators (over a neighborhood N). There is a profusion of such features.

c) Homogeneous regions obtained by variously grouping the pixel features from (a) and (b). There is a profusion of such groupings also. The basic criterion is that the facets (regions) in the images become "easier to match" than the original pixels (for example, they have to be "reasonably large" connected regions).

d) Super-regions of various types obtained by grouping the regions from (c) and/or by using operators that process the whole image as a unit.

Hierarchical matching strategies were suggested. The initial match may only be on the "super-region" level (d) but could also be on any level of detail if the features are unique. Presumably the super-regions are easiest to match since they practically always overlap in the two images (super-region intersection between L and R images is highly likely to be non-zero). This will assist the matching of the "homogeneous regions" (c) specially if they are subsets of super-regions (d). After this matching is accomplished, the local operator may be made more "sensitive" by decreasing the size of local region N (the "scale-space" approach) or the image is enhanced to bring out details within each homogeneous region, or the region boundaries are matched next, etc. In the end the regions should become "small enough" to allow unique pixel level feature matching.

Only the "similarity principle" is used in the matching, i.e., at whatever level the matching is to occur, the items matched are (rather) similar in the L and R images but must have unique characteristics to prevent confusion (combinatorial problems should not exist). The L and R similarity principle extends from pixel features to super-regions. Consequently, object recognition, structure recognition, scene models, etc., are unnecessary and "image understanding" is not needed to "see depth" in stereo images. By definition, the images to be matched must be unknown initially and remain so even after stereo matching.

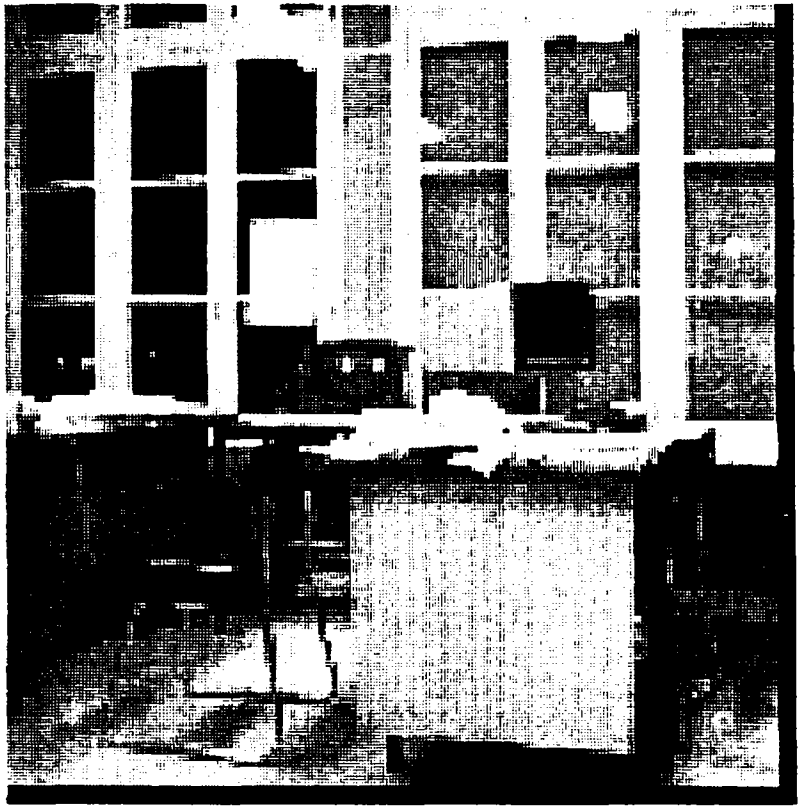
Since no technical information about the two "INRIA images" was available, the right image (R) was shifted slightly with respect to the left image (L) in an attempt to create and apparent "centre of fixation". The R image was shifted by 3 pixels to the left and 2 pixels up with respect to L, see Figure 3.1-1. The corresponding region in L was zeroed. This did not disturb stereo fidelity, as may be ascertained from Figures 3.1-2a and-2b. All feature images needed in the processing were shifted by the same amount (after the processing described in Chapter 2).

The experiments are far from complete both due to time and equipment limitations. About half a dozen matching experiments are described and even these experiments have been restricted to only the "homogeneous region" level (c). In retrospect, some of the experiments appear to have been rather poorly thought out before the experiment was started but, one learns by doing!

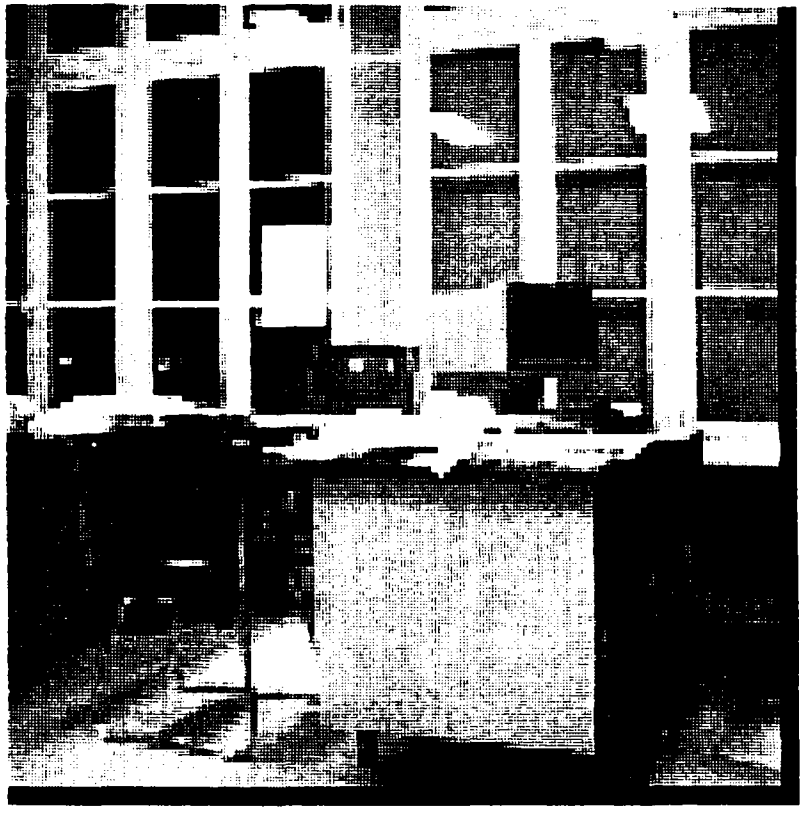
Figure titles:

Figure 3.1-1: The images after shifting in an attempt to create a "centre of fixation".

Figures 3.1-2a and -2b: Stereo pairs of the shifted gray level and the shifted gradient magnitude images.



if8gray.lsh

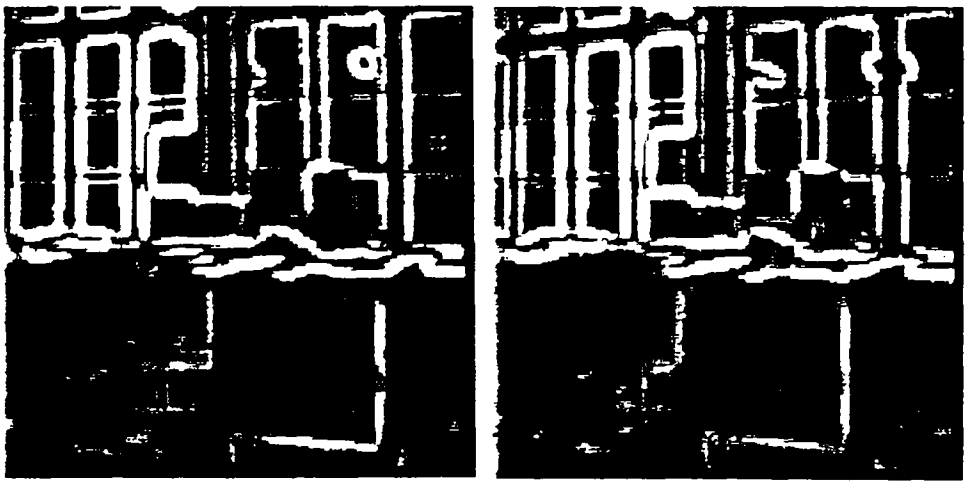


if8gray.rsh

Fig.: 3.1-1.

if8gray.lsh if8gray.rsh

Fig.: 3.1-2a.



if8gradm.lsh if8gradm.rsh

Fig.: 3.1-2b.

3.2 Matching methods

Several matching methods were tried. These are briefly described in the sub-sections to follow. Briefly:

1. The first method attempts to use "matched classification" in order to try to classify the pixels in the left and right image at the same time. In the present context the results form this case are used to illustrate the many difficulties one can encounter.

2. The second method is based on simple "and-ing" of class labels which are then post-processed. This method has been somewhat more "refined" to study how some of the problems can be partially "fixed". Except for a few results shown in the Abstract and in section 3.2.2, these results have been left out of the report to reduce its size.

3. The third approach is based on correlation. It is only a slight modification of the techniques used in binary image processing. (Discussed but not completed due to lack of time.)

4. The forth method is using "mutual recognition" of the facets in the left and right images. At present it is only a slightly modified "minimum distance" classifier.

5. The fifth approach attempted to use pixel feature matching to go directly to a depth image, but became constrained to a very restricted case to due equipment limitations. (Briefly discussed.)

In the first and second methods (1,2) the initial matched image is called L&R(i,j) and it is accompanied by the corresponding left L(i,j) and right R(i,j) images. The third method (3) gives a list of matching facet labels and also a list of "best shifts" between the facets in L and R. The forth method (4) only gives a list of matching facet labels in L and R but no shifts. The images of matched facets, Lm(i,j) and Rm(i,j), are constructed from these lists. In all cases, the L&Rm(i,j) and/or Lm(i,j) and Rm(i,j) images require considerable further processing before the results can be judged. The fifth method attempts to go directly from pixel features to 3D reconstruction.

3.2.1 Matched classification

The two hierarchical labelled decision spaces (look-up tables) used in "homogeneous" region detection (Hl(Gpm,G) followed by Hl(Gpm,Gpa), see Chapter 2) contained about a dozen separate region types but, of course, the actual number of region types is not important. As seen from the results in Chapter 2, independent classification of the L and R images produced "reasonably consistent" regions, except in some areas of the images. The "totally different" regions in L versus R were mainly due to missing gray levels and reflections of light sources.

It was argued that if a pixel p at coordinates (i,j) in the left image L belonged to a given class k, then there should be a pixel p´ in the right image R within "an allowable maximal displacement" di and dj form (i,j) which should belong to the same class k if the gray level (G) and gradient magnitude and angle values (Gpm and Gpa) were "perturbed" by increasing amounts after each search in di*dj. This should improve the similarity of the "reasonably consistent" regions but, of course, should not improve the "totally different" regions. Thus, it should be possible to construct a "matched and classified" image L&Rh(i,j) and "displacement images" di(i,j) and dj(i,j) containing di and dj, where

i(left image) +/- di = i(right image)
j(left image) +/- dj = j(right image).

A block diagram of the processing structure is shown in Figure 3.2.1-1a with explanation of the symbols in Figure 3.2.1-1b.

The "displacement images" di(i,j) and dj(i,j) could not be used since the features were essentially homogeneous within each facet and no pixel matching could take place. The number of features was too limited for unique identification and matching of pixels. The "matched and classified" image L&Rh(i,j) is shown in Fig. 3.2.1-2. The corresponding Lh(i,j) and Rh(i,j) images are in Figures 3.2.1-3a and -3b. In the present case the "raw" results are shown in order to better illustrate the reasons for many of the processes described in section 3.3, i.e., "post processing" was not used to "homogenize" the L&Rh(i,j), Lh(i,j), and Rh(i,j) triplet.

As in Chapter 2, attempts were made to display the results "in stereo" for visual verification. When each class was replaced by its average gray level (see Figures 3.2.1-4a and 3.2.1-5) the images became rather "bland" since the gray level gradients are missing (except at class boundaries). Nine "colours" were required to "paint" these regions in different "colours" (see Figures 3.2.1-4b, -4c and -6b). A stereo pair for the average gray levels is shown in Figure 3.2.1-6a and for the "painted" classes in Figure 3.2.1-6b. Mild stereo effects are visible, but the nonmatching "dots" in Figure 3.2.1-6b greatly disturb our vision.

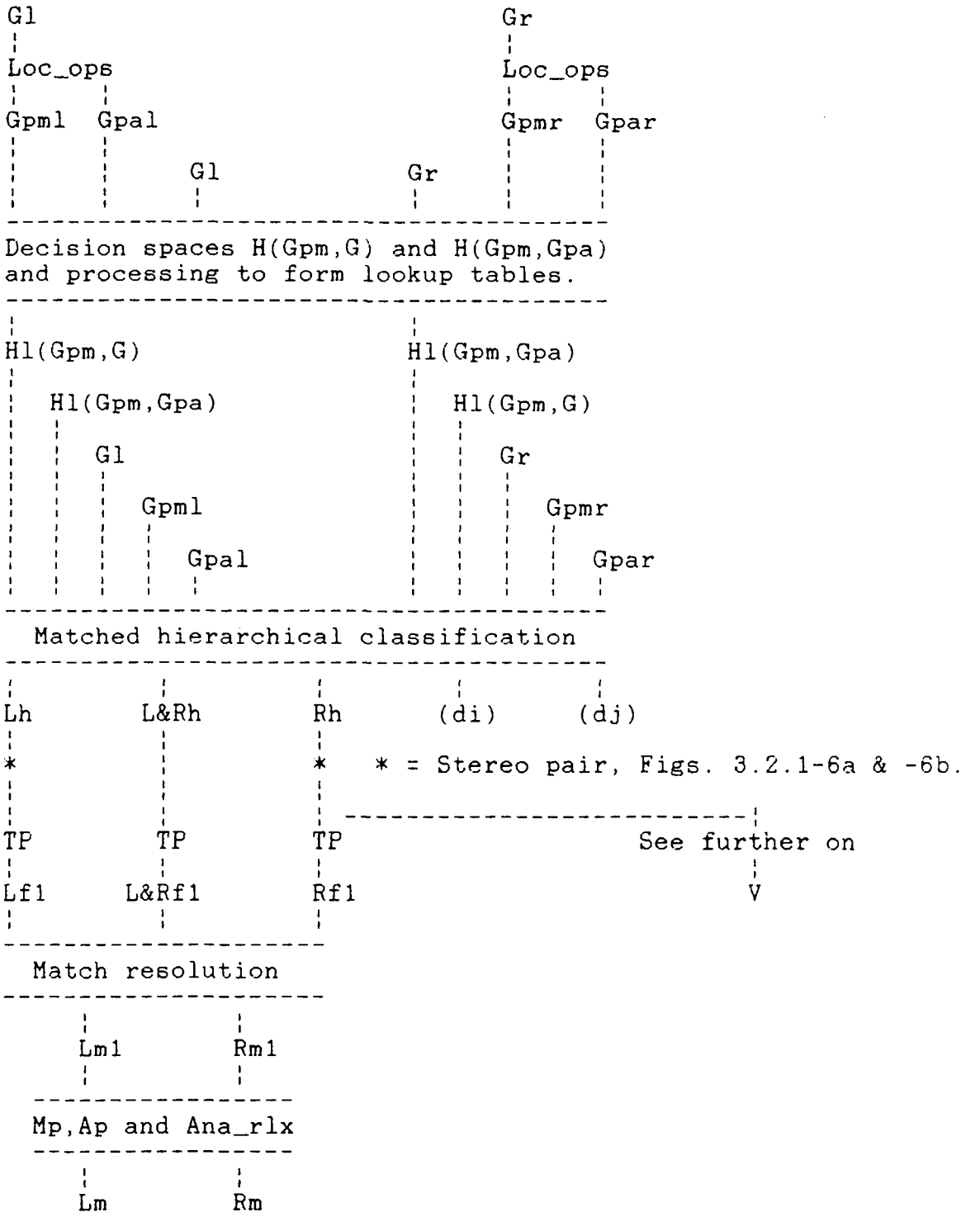


Figure 3.2.1-1a: Matched classification and match resolution. Process shown starting from stereo image pair Gl(i,j) and Gr(i,j). Explanation of symbols is in Fig. 3.2.1-1b.

*l = Associated with left image.
*r = Associated with right image.
G = Gray level image.
Gl = Gray level image of stereo pair, left.
Gr = Gray level image of stereo pair, right.
Loc_ops = Local operators within neighborhood N.
Gpm = Gradient magnitude.
Gpml = Gradient magnitude, left.
Gpmr = Gradient magnitude, right.
Gpa = Gradient angle.
Gpal = Gradient angle, left.
Gpar = Gradient angle, right.
H(Gpm,G)   = Decision space on Gpm and G for left and right.
H(Gpm,Gpa) = Decision space on Gpm and Gpa for left and right.
Hl(Gpm,G)   = Lookup table for classifying Gpm and G.
Hl(Gpm,Gpa) = Lookup table for classifying Gpm and Gpa.
*h = Class labelled image.
Lh = Class labelled image, left.
Rh = Class labelled image, right.
L&Rm = Matched class label image.
di = Displacement image for i-coordinate.
di = Displacement image for j-coordinate.
TP = Lh, L&Rh, and Rh triplet processing
*f = Facet (region) labelled image.
Lf1 = Facet (region) labelled image, left.
Rf1 = Facet (region) labelled image, right.
L&Rf1 = Facet (region) labelled image of L&Rm.
Lm1 = Matched image, left.
Rm1 = Matched image, right.
Mp,Ap and Ana_rlx = Subsequent processes.
Mp = Match processing.
Ap = Analytic approximation.
Ana_rlx = Analytic relaxation.
Lm = Matched image, left.
Rm = Matched image, right.

Figure 3.2.1-1b: Symbols for matched classification, see Fig. 3.2.1-1a.

(Relevant computer files: /user2/kasvand/ima2):
Lh = Inr065.fig L&Rh = Inr063.fig Rh = Inr066.fig
Lf1 = Inr069.fig L&Rf1 = Inr067.fig Rf1 = Inr070.fig
Match list = Inr071.fig
Lm1 = Inr073.fig Rm1(i,j) = Inr074.fig

The principal difficulties with matched classification are that if nothing is known about the scene then the "allowable maximal displacement" for di (along the direction of displacement of the cameras) can be the whole image width, and if the "perturbations" are allowed to be too large then any pixel in L will match any pixel in R and vice versa. However, when the "perturbations" and the "di and dj" tolerances where made small, then the method started to resemble direct "and-ing" of the Lh(i,j) and Rh(i,j) images. Chronologically, this was the second method tried, being a "watered down" version of direct "pixel feature matching", which was beyond the capabilities of the equipment.

Figure titles:

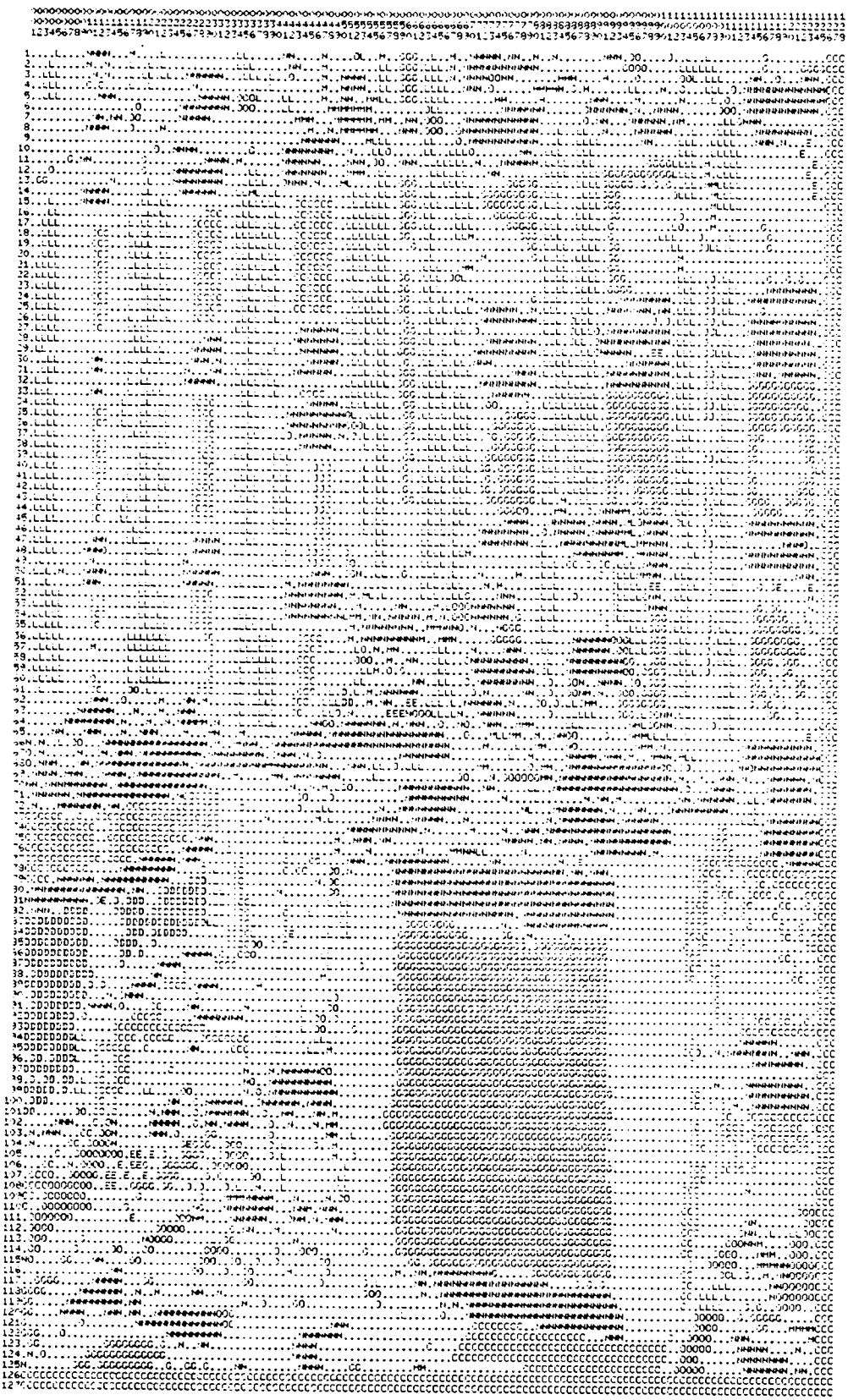
Figure 3.2.1-2: The "matched and classified" image L&Rh(i,j) showing the class labels from Hl(Gpm,G) and Hl(Gpm,Gpa). The labelling is continued from Hl(Gpm,G) to Hl(Gpm,Gpa). (File: Inr063.fig)

Figures 3.2.1-3a and -3b: The Lh(i,j) and Rh(i,j) class label images. No "post processing" has been used. (Files: Inr065.fig and Inr066.fig)

Figures 3.2.1-4a, -4b and -4c: Average gray levels per class, adjacency matrix for classes, and "colour" assignments for the classes. (Files: Inr095.fig, Inr096.fig, and Inr097.fig)

Figure 3.2.1-5: Each class was replaced by its average gray level.

Figures 3.2.1-6a and -6b: Stereo pairs of the Lh(i,j) and Rh(i,j) images. (a) Using average gray level for each class. (b) Classes "painted" in different "colours" where adjacent classes have different colours.



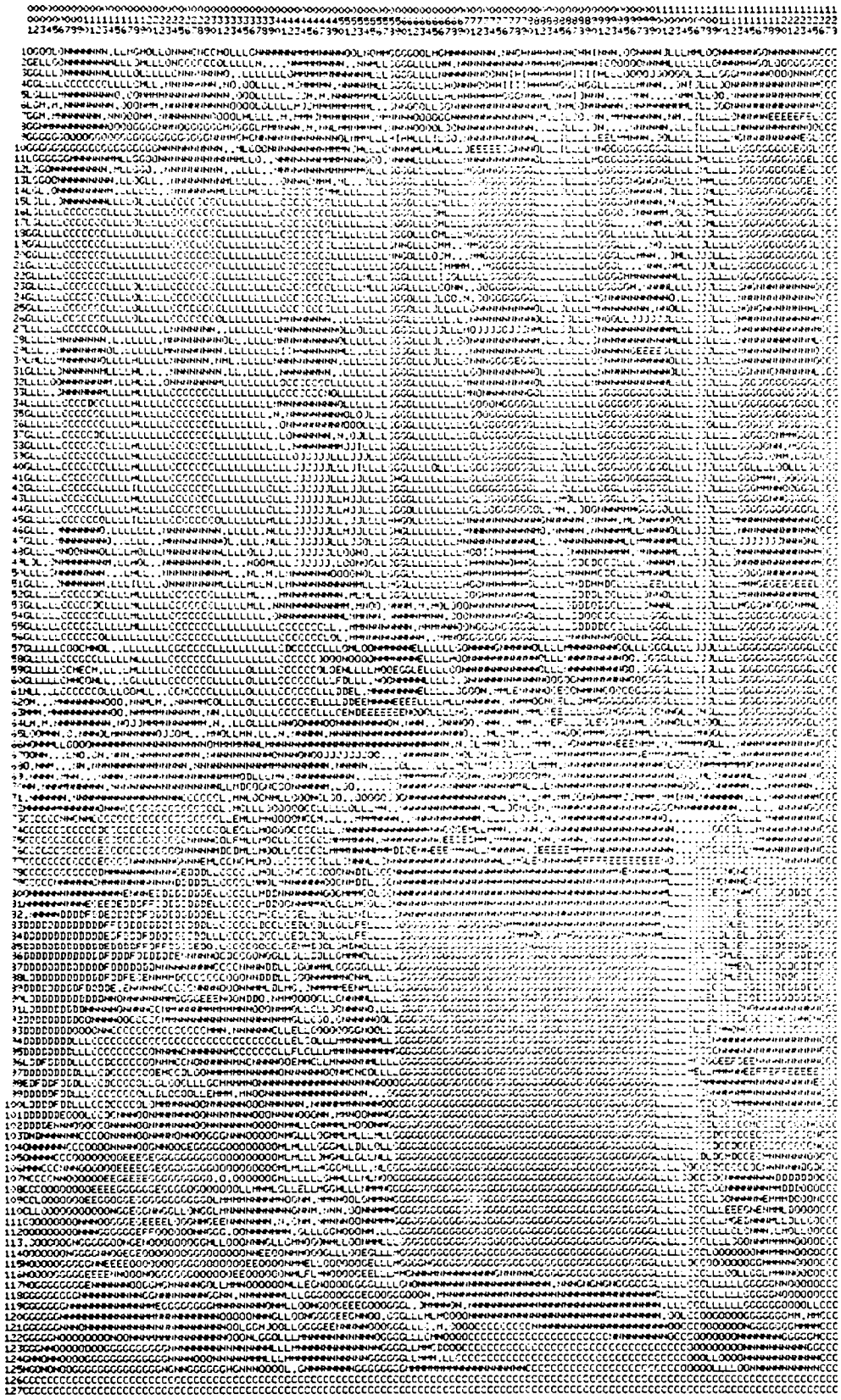
LiPh(i,j), class labels from matched classification. Iopt1,2,3 = 1,0,0.

LABEL PRINTING

A= 0	B= 1	C= 2	D= 3	E= 4	F= 5	G= 6	H= 7
I= 8	J= 9	K= 10	L= 11	M= 12	N= 13	O= 14	P= 15
Q= 16	R= 17	S= 18	T= 19	U= 20	V= 21	W= 22	X= 23
Y= 24	Z= 25	[= 26	A= 27	B= 28	C= 29	D= 30	.ETC.
i= -8	j= -9	k=-10	l=-11	m=-12	n=-13	o=-14	p=-15
q=-16	r=-17	s=-18	t=-19	u=-20	v=-21	w=-22	x=-23
y=-24	z=-25	=-26	a=-27	b=-28	c=-29	d=-30	.etc.

File: Inn063.fig

Fig.: 3.2.1-2.



Lh(i,j). left image using matched classification with Hl(Gpm,G) and Hl(Gpm,Gpa).

LABEL PRINTING

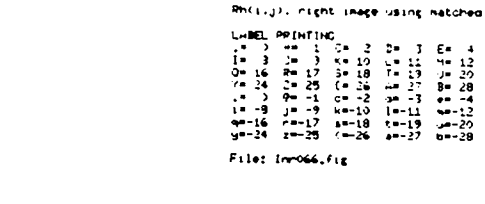
A= 0	B= 1	C= 2	D= 3	E= 4	F= 5	G= 6	H= 7
I= 8	J= 9	K= 10	L= 11	M= 12	N= 13	O= 14	P= 15
Q= 16	R= 17	S= 18	T= 19	U= 20	V= 21	W= 22	X= 23
Y= 24	Z= 25	[= 26	a= 27	B= 28	C= 29	D= 30	.ETC.
.= 0	?= -1	c= -2	d= -3	e= -4	f= -5	g= -6	h= -7
i= -8	j= -9	k=-10	l=-11	m=-12	n=-13	o=-14	p=-15
q=-16	r=-17	s=-18	t=-19	u=-20	v=-21	w=-22	x=-23
y=-24	z=-25	{=-26	a=-27	b=-28	c=-29	d=-30	.etc.

File: Inr065.fig

Some program parameters:
NSTEPDM,NSTEPFB,NSTEPSE,IOPT,LASTGMG,IOPT2,IOPT3,NJUMP0 =
2 3 4 -1 9 0 0 0
DGG,DGM,DGA = 0.50000E+03 0.20000E+03 0.20000E+01
FMULR = 0.10000E+03

Fig.: 3.2.1-3a.

Rh(i,j). right image using matched classification with Hi(Gpm,G) and Hi(Gpm,Gear).



File: Inn066.fig

Fig.: 3.2.1-3b.

```
 K = L R C   NNFL   NNFR    SSGL    SSGR

 1 = • • 1    485    504    50.5    50.8
 2 = C C 1   2129   2963     1.5     0.4
 3 = D D 1    509    343    15.1    13.5
 4 = E E 1    240    400    26.4    27.2
 5 = F F 1     41     12    22.3    23.7
 6 = G G 1   2946   2548    55.8    54.2
 7 = H H 1    113     37    80.5    81.2
 8 = I I 1     41     28    37.1    38.5
 9 = J J 1    315    277   105.0   107.3
10 =   K 0      0      5     0.0    96.4
11 = L L 1   3627   3472    55.4    54.0
12 = M M 1    310    738    49.6    47.8
13 = N N 1   3639   3603    42.1    40.6
14 = O O 1   1301   1206    44.6    41.3

 SC.GMAX =  0.20001E+01  0.12750E+03
```

Average gray levels for classes after
   matched classification.

Inr017.Fig = Class labels. left.
Inr018.Fig = Class labels. right.

K = Class label.
L = Class label. left image.
R = Class label. right image.
C = 1 if both labels in left and right images.
    else 0.
NNFL = Number of pixels per class. left image.
NNFR = Number of pixels per class. right image.
SSGL = Average gray level of class. left image.
SSGR = Average gray level of class. right image.

File: Inr095.Fig

Fig.: 3.2.1-4a.

```
    00000000000000000000
    00000000011111111112
    12345678901234567890

 1  01110111101111000000
 2  10111110001111000000
 3  11011100001111000000
 4  11101100001111000000
 5  01110000001111000000
 6  11110011101111000000
 7  11000101101111000000
 8  10000110111111000000
 9  10000111011111000000
10  00000001101010000000
11  11111111110111000000
12  11111111101011000000
13  11111111111101000000
14  11111111101110000000
15  00000000000000000000
16  00000000000000000000
17  00000000000000000000
18  00000000000000000000
19  00000000000000000000
20  00000000000000000000
```

Adjacency matrix for class labelled
   left and right images.

Inr017.Fig = Class labels. left.
Inr018.Fig = Class labels. right.

File: Inr096.Fig

Fig.: 3.2.1-4b.

```
                      Lab    Ncon     Newcol     Ncon Newcol
LS = LLS(  1)  = 11=L   NNC(LS).MNK(LS)  =  13    1
LS = LLS(  2)  = 13=N   NNC(LS).MNK(LS)  =  13    2
LS = LLS(  3)  = 12=M   NNC(LS).MNK(LS)  =  12    3
LS = LLS(  4)  = 14=O   NNC(LS).MNK(LS)  =  12    4
LS = LLS(  5)  =  6=G   NNC(LS).MNK(LS)  =  11    5
LS = LLS(  6)  =  1=•   NNC(LS).MNK(LS)  =  11    6
LS = LLS(  7)  =  2=C   NNC(LS).MNK(LS)  =  10    7
LS = LLS(  8)  =  3=D   NNC(LS).MNK(LS)  =   9    8
LS = LLS(  9)  =  4=E   NNC(LS).MNK(LS)  =   9    9
LS = LLS( 10)  =  7=H   NNC(LS).MNK(LS)  =   9    8
LS = LLS( 11)  =  8=I   NNC(LS).MNK(LS)  =   9    7
LS = LLS( 12)  =  9=J   NNC(LS).MNK(LS)  =   9    9
LS = LLS( 13)  =  5=F   NNC(LS).MNK(LS)  =   7    5
LS = LLS( 14)  = 10=K   NNC(LS).MNK(LS)  =   4    3
```

Colour assignments. Nine colours were needed.

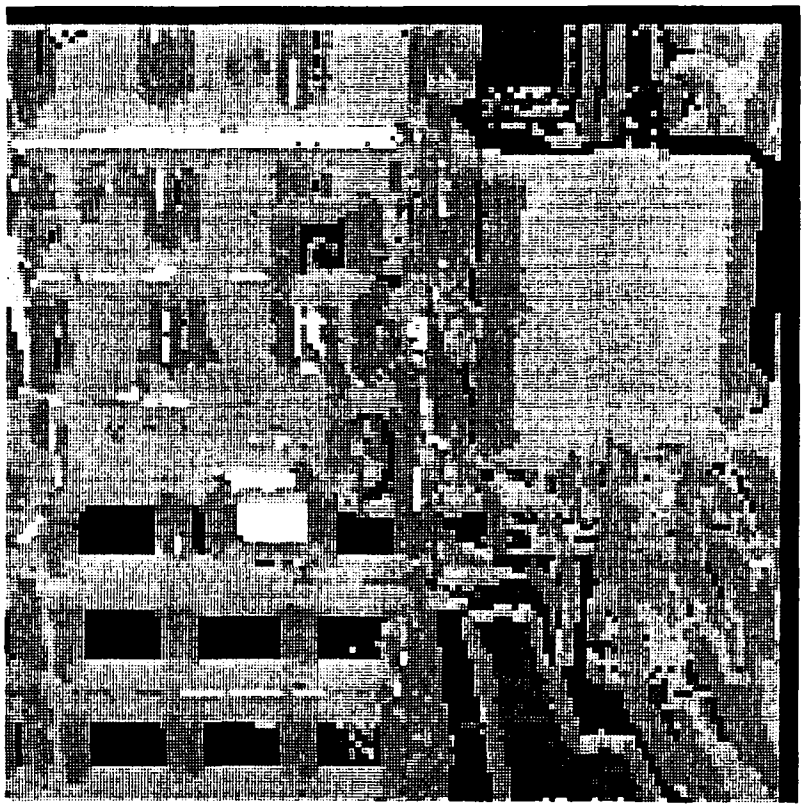
Lab .. = Class label in images.
Ncon . = Number of contacts with other labels.
Newcol = New colour assignment.

Inr017.Fig = Class labels. left.
Inr018.Fig = Class labels. right.

File: Inr037.Fig

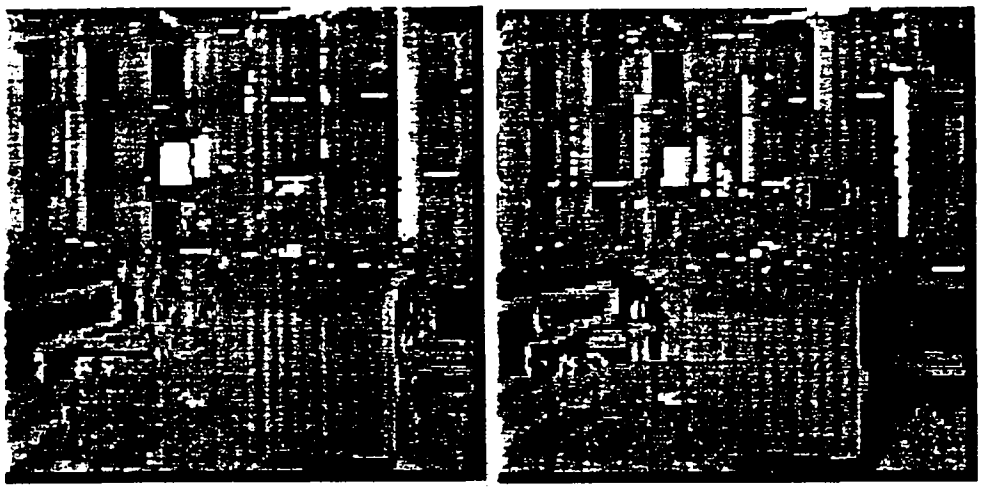
Fig.: 3.2.1-4c.

if8hlmc1.rgr



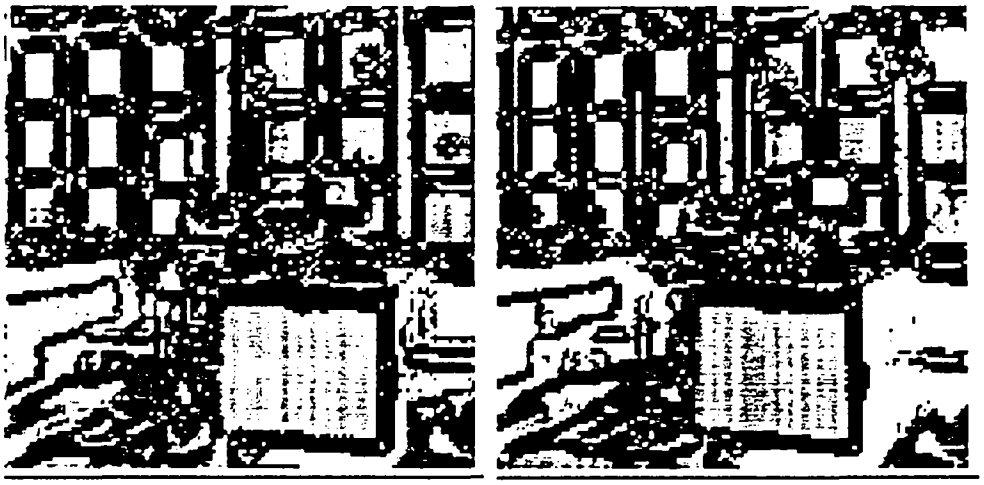
if8hlmc1.lgr

Fig.: 3.2.1-5.



if8h1mc1.1gr if8h1mc1.rgr

Fig.: 3.2.1-6a.



if8h1mc12.1gr if8h1mc12.rgr

Fig.: 3.2.1-6b.

3.2.2 Direct "and-ing"

The classified left Lh(i,j) and right Rh(i,j) images contain the class labels from Hl(Gpm,G) and Hl(Gpm,Gpa). Direct "and-ing" consists in comparing the labels in Lh(i,j) and Rh(i,j) pixel by pixel (for the same i and j in Lh(i,j) and Rh(i,j)) and if the labels are the same then this label is stored in L&Rh(i,j) else a zero is stored. Thus,

```
For all pixels (i,j)
If(Lh(i,j).Eq.Rh(i,j)) then L&Rh(i,j) = Lh(i,j)
                       else L&Rh(i,j) = 0
```

In block diagram form the process is shown in Figure 3.2.2-1a with explanation of symbols in Figure 3.2.2-1b. There is an obvious danger with this method, specially if the Lh(i,j) and Rh(i,j) images contain (horizontal) repetitive structures. As seen from Fig. 3.2.2-2 for an "one-dimensional" case, if the shift (disparity) between the L and R images is larger than the region size then the matching will be entirely incorrect. Clearly, "grouping" of repetitive structures is needed. Grouping is not a very difficult problem. Some discussion on an one-dimensional version of it may be found in Ref. 2.5.

The Lh(i,j) and Rh(i,j) images, shown in Chapter 2 as Figures 2.4-7a and -7b, are repeated as Figures 3.2.2-3a and -3b. The resultant L&Rh(i,j) image is shown in Fig. 3.2.2-3c. In this case the Lh and Rh images were subject to post-processing (as compared to "Matched classification") in order to illustrate the improvement (compare with Figs. 3.2.1-3a and -3b). As before, attempts are made to show the results in different ways for verification and evaluation. In Figure 3.2.2-4 the pixel classes are represented such that adjacent regions have different gray levels (seven "colours" were used). Figures 3.2.2-5a and -5b show the same "colours" printed as symbols. Figures 3.2.2-6a and -6b show the stereo pair for Figure 3.2.2-4 and for the original gray level images.

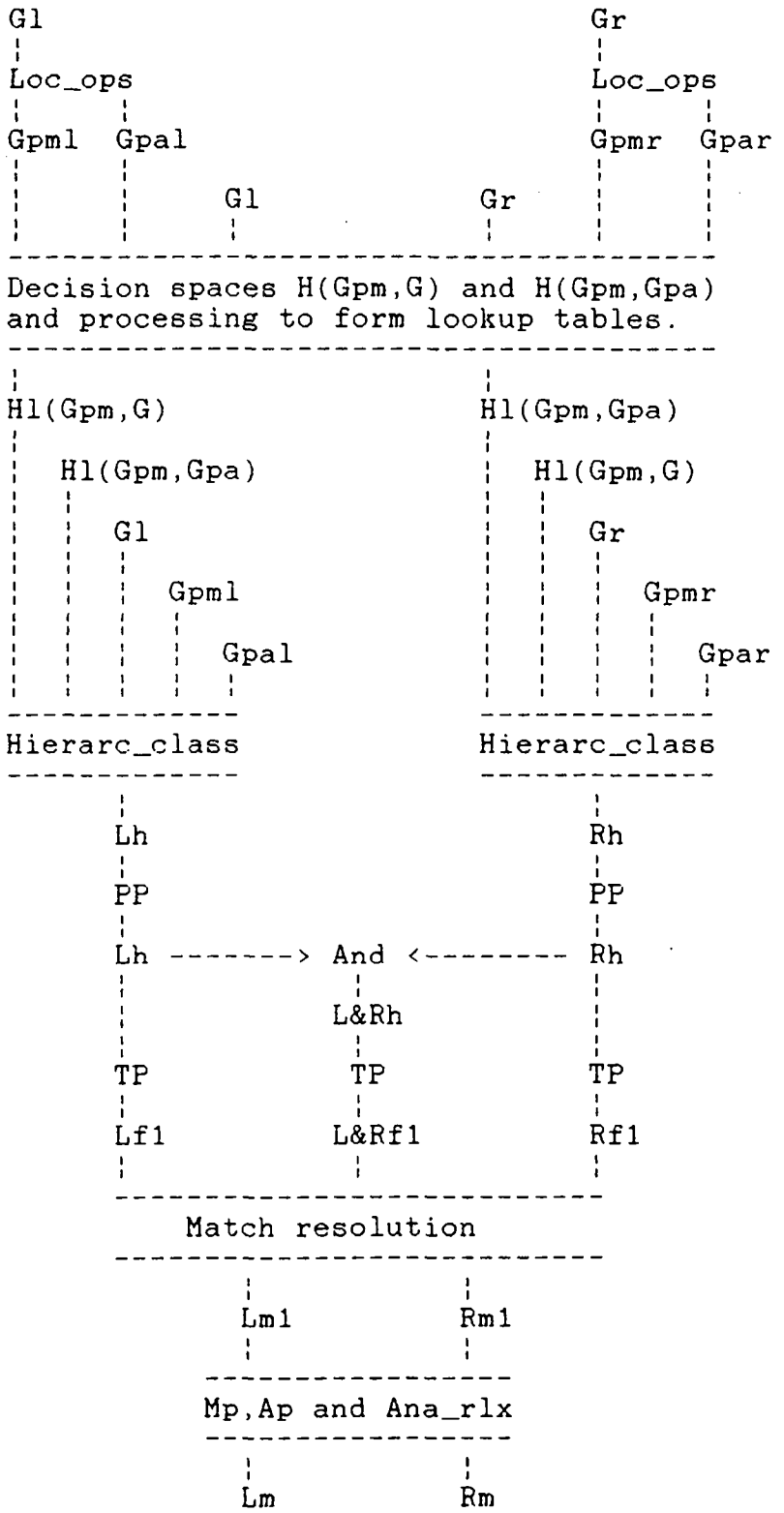


Figure 3.2.2-1a: "And-ing" of class labels and match resolution.
The symbols are in Fig. 3.2.2-1b.

*l = Associated with left image.
*r = Associated with right image.
G = Gray level image.
Gl = Gray level image of stereo pair, left.
Gr = Gray level image of stereo pair, right.
Loc_ops = Local operators within neighborhood N.
Gpm = Gradient magnitude.
Gpml = Gradient magnitude, left.
Gpmr = Gradient magnitude, right.
Gpa = Gradient angle.
Gpal = Gradient angle, left.
Gpar = Gradient angle, right.
H(Gpm,G)   = Decision space on Gpm and G for left and right.
H(Gpm,Gpa) = Decision space on Gpm and Gpa for left and right.
Hl(Gpm,G)   = Lookup table for classifying Gpm and G.
Hl(Gpm,Gpa) = Lookup table for classifying Gpm and Gpa.
Hierarc_class = Hierarchical classification.
*h = Class labelled image.
Lh = Class labelled image, left.
Rh = Class labelled image, right.
PP = Lh and Rh pair post-processing.
And = The "and-ing" process.
L&Rh = "And-ed" class labels.
TP = Lh, L&Rh, and Rh triplet processing.
*f = Facet (region) labelled image.
Lfl = Facet (region) labelled image, left.
Rfl = Facet (region) labelled image, right.
L&Rfl = Facet (region) labelled image of L&Rm.
Lml = Matched image, left.
Rml = Matched image, right.
Mp,Ap and Ana_rlx = Subsequent processes.
Mp = Match processing.
Ap = Analytic approximation.
Ana_rlx = Analytic relaxation.
Lm = Matched image, left.
Rm = Matched image, right.

Figure 3.2.2-1b: Explanation of symbols for matching based on "and-ing", see block diagram in Fig. 3.2.2-1a.

(Relevant computer files: /user2/kasvand/ima2):
L&Rh = Inr100.fig Lh = Inr101.fig Rh = Inr102.fig
Lfl = Inr103.fig Rfl = Inr104.fig
Lml = Inr107.fig Rml = Inr108.fig
Lm   = Inr115.fig Rm = Inr116.fig
Mp,Ap and ana_rlx: Inr109.fig Inr110.fig
Inr111.fig Inr112.fig Inr113.fig

```
.. AAAA BBBB AAAAA BBBBB AAAAAA BBBBBB ........ = Labels in Lh
.......... AAAA BBBB AAAAA BBBBB AAAAAA BBBBBB = Labels in Rh
............ AAA . BB .. AA ... B ............. = Labels in L&Rh
```

Figure 3.2.2-2: A case of incorrect results from "and-ing". The labels are A for one class and B for the other.

Figure titles:

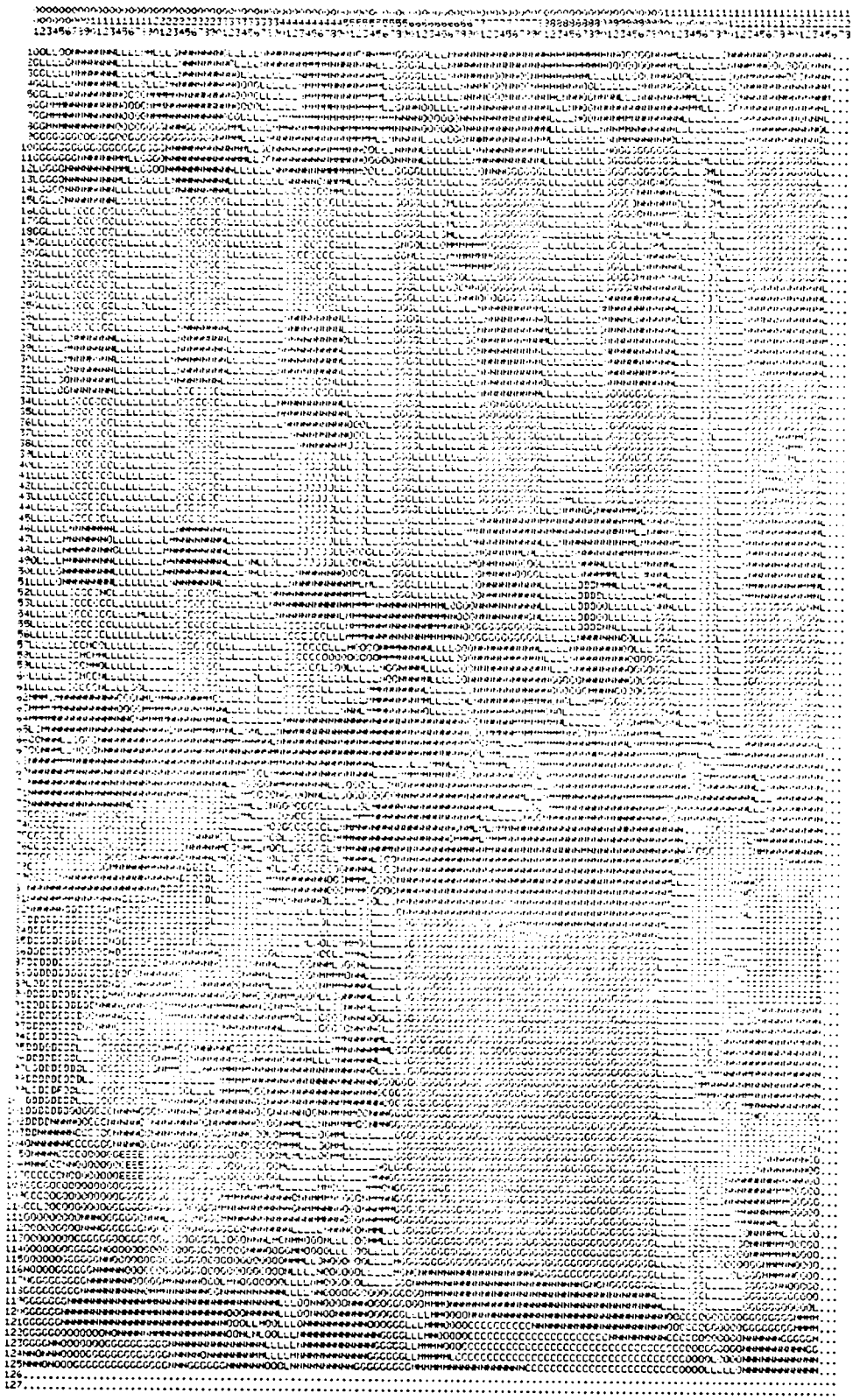
Figures 3.2.2-3a and -3b: The cluster (or class) labels from Hl(Gpm,G) and Hl(Gpm,Gpa) projected into the image space, after some processing. (Files: Inr101.fig, Inr102.fig)

Figure 3.2.2-3c: The "and-ed" L&Rh(i,j) image showing the class labels from Hl(Gpm,G) and Hl(Gpm,Gpa). (File: Inr100.fig)

Figure 3.2.2-4: The "painted" classes. The pixel classes are represented such that adjacent regions have different gray levels ("colours").

Figures 3.2.2-5a and -5b: The "painted" classes in Figure 3.2.2-4 shown with symbols. (Files: Inr193.fig, Inr194.fig)

Figures 3.2.2-6a and -6b: The stereo pairs for Figure 3.2.2-4 and for the original gray levels.



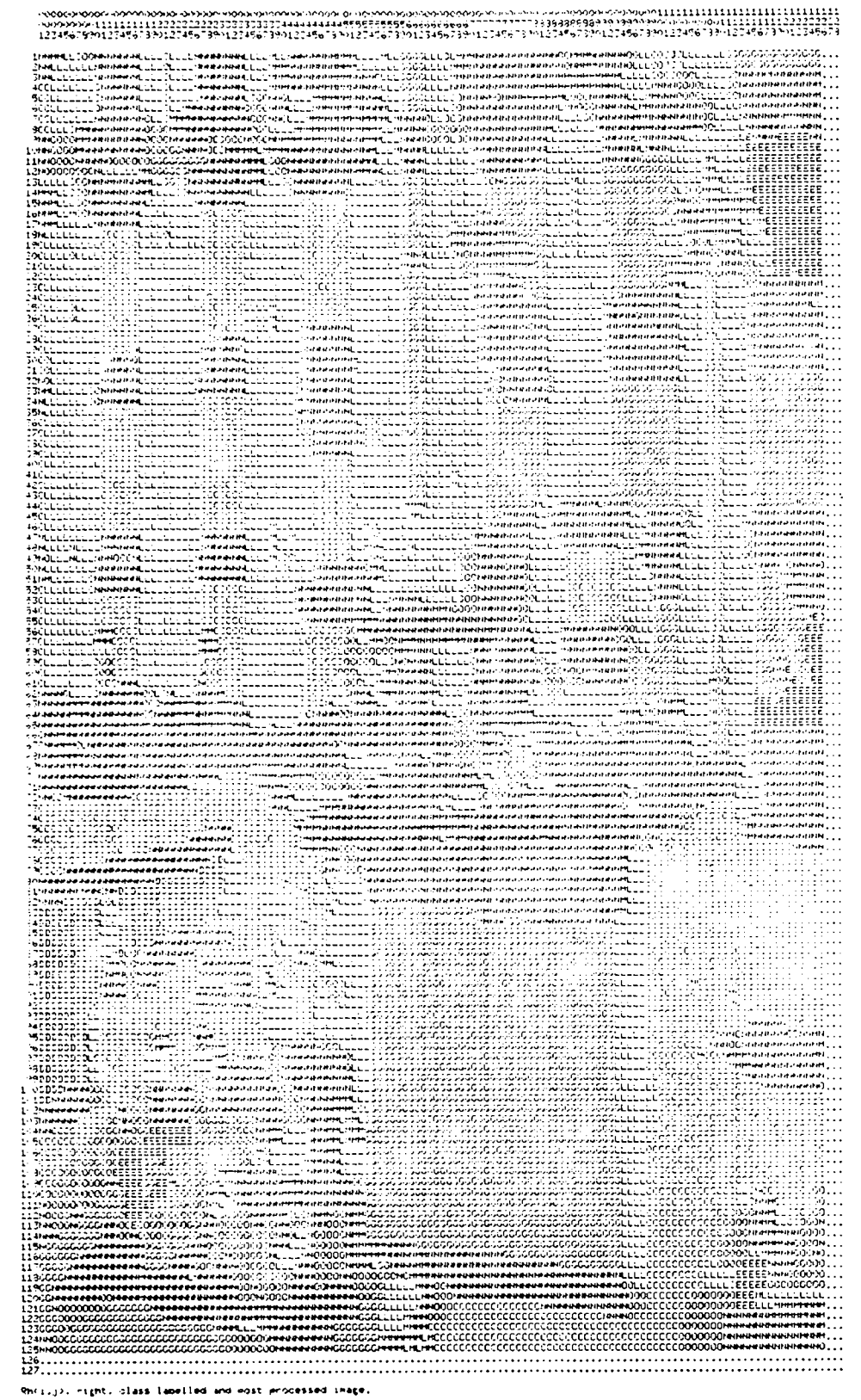
Lh(i,j), left, class labelled and post processed image.

LABEL PRINTING

.= 0	A= 1	C= 2	D= 3	E= 4	F= 5	G= 6	H= 7
I= 8	J= 9	K= 10	L= 11	M= 12	N= 13	O= 14	P= 15
Q= 16	R= 17	S= 18	T= 19	U= 20	V= 21	W= 22	X= 23
Y= 24	Z= 25	(= 26	A= 27	B= 28	C= 29	D= 30	.ETC.
.= 0	Q= -1	c= -2	d= -3	e= -4	f= -5	g= -6	h= -7
i= -8	j= -9	k=-10	l=-11	m=-12	n=-13	o=-14	p=-15
q=-16	r=-17	s=-18	t=-19	u=-20	v=-21	w=-22	x=-23
y=-24	z=-25	(=-26	a=-27	b=-28	c=-29	d=-30	.etc.

File: Inr101.fig

Fig.: 3.2.2-3a.



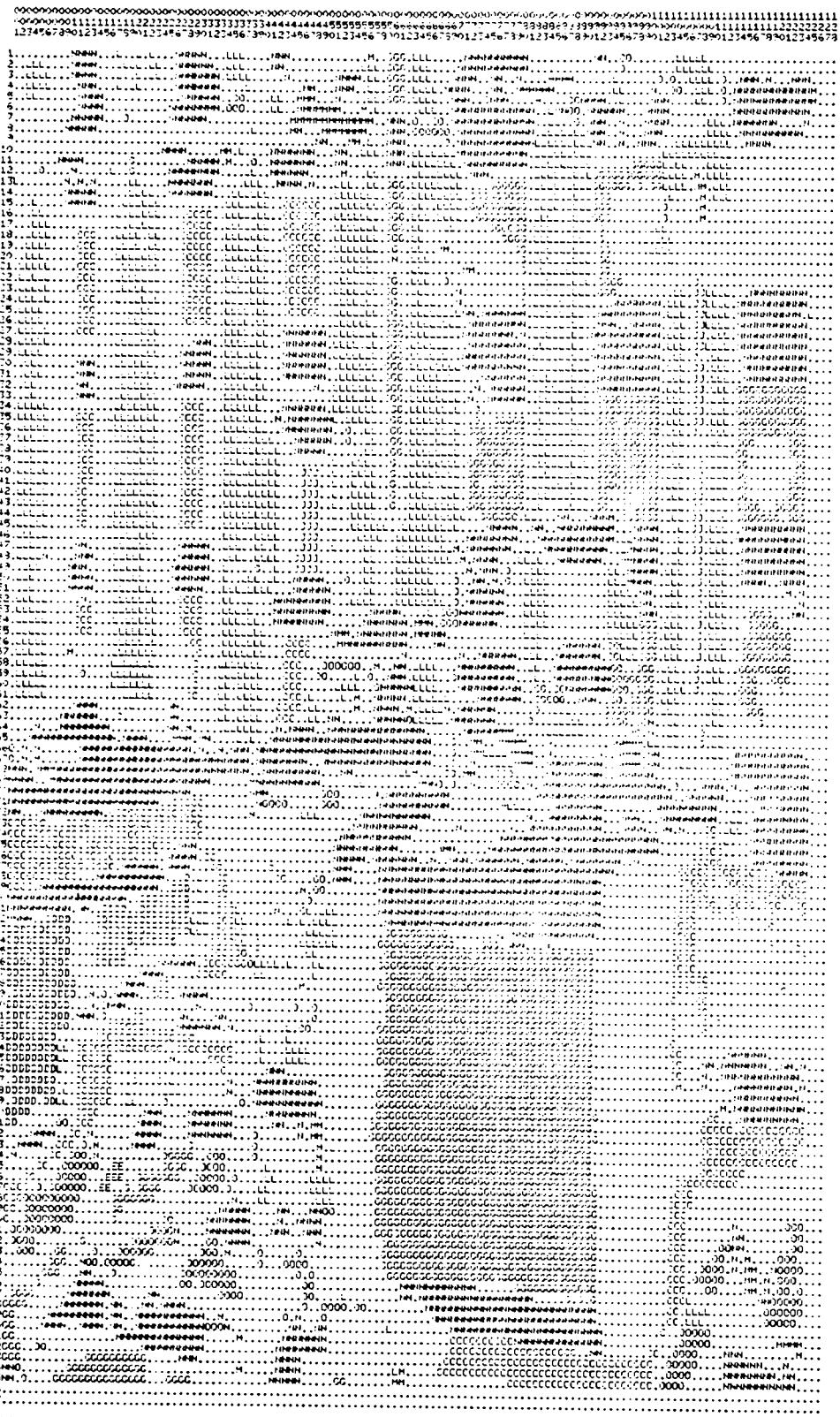
Ph(i,j). right, class labelled and most processed image.

LABEL PRINTING

.= 0	*= 1	C= 2	D= 3	E= 4	F= 5	G= 6	H= 7		
I= 8	J= 9	K=10	L=11	M=12	N=13	O=14	P=15		
Q=16	R=17	S=18	T=19	U=20	V=21	W=22	X=23		
Y=24	Z=25	[=26	\=27	]=28	^=29	_=30	.ETC.		
.= 0	a=-1	b=-2	c=-3	d=-4	e=-5	f=-6	g=-7	h=-8	
i=-9	j=-10	k=-11	l=-12	m=-13	n=-14	o=-15			
p=-16	q=-17	r=-18	s=-19	t=-20	u=-21	v=-22	w=-23		
x=-24	y=-25	z=-26	{=-27	|=-28	}=-29	~=-30	.etc.		

File: lnn192.fig

Fig.: 3.2.2-3b.



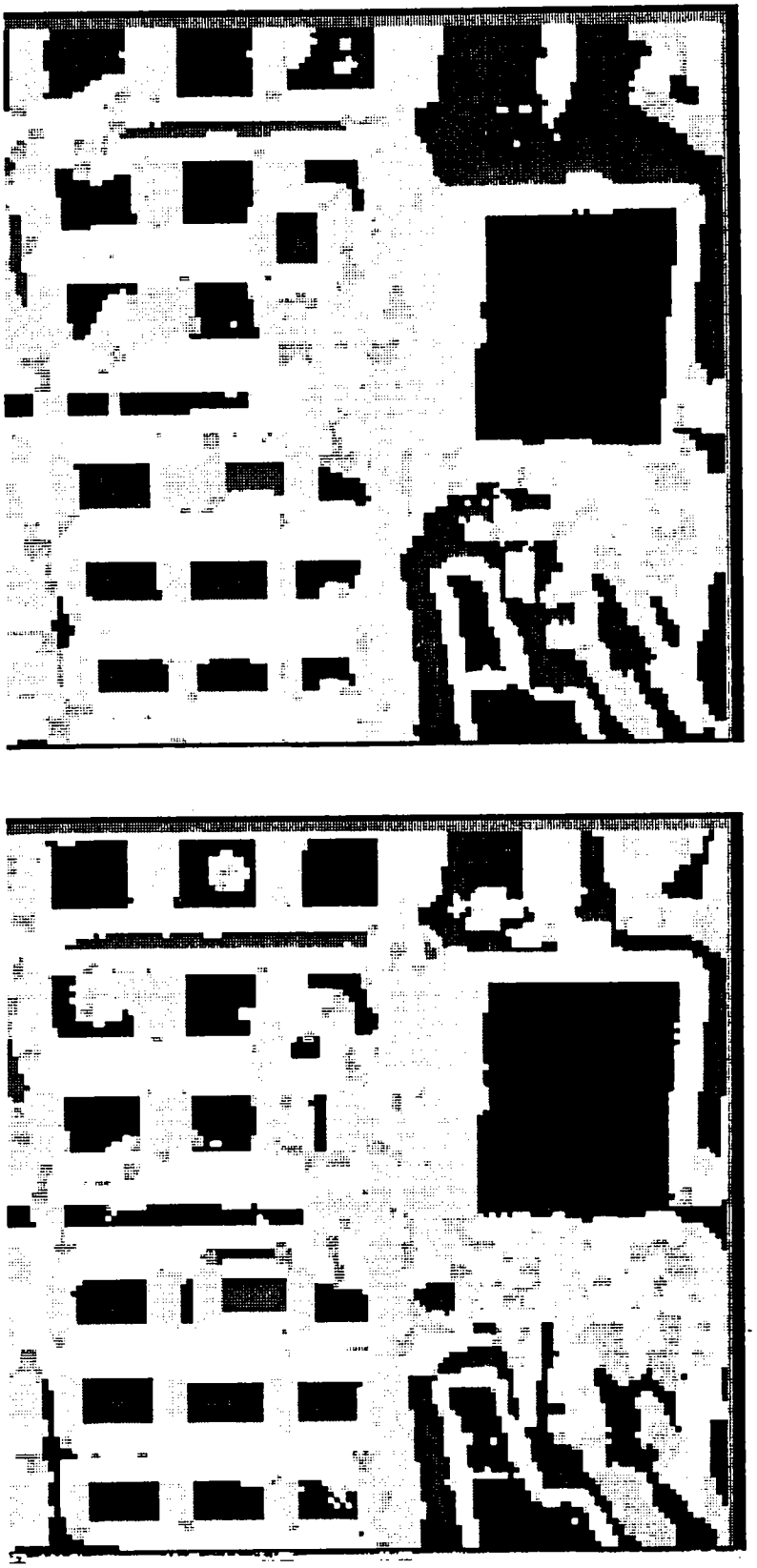
L&Rh(i,j), "and-or" Lh(i,j) and Rh(i,j).

LABEL PRINTING

.= 0	== 1	C= 2	D= 3	E= 4	F= 5	G= 6	H= 7				
I= 8	J= 9	K= 10	L= 11	M= 12	N= 13	O= 14	P= 15				
Q= 16	R= 17	S= 18	T= 19	U= 20	V= 21	W= 22	X= 23				
Y= 24	Z= 25	[= 26	\= 27	]= 28	^= 29	_= 30	.ETC.				
.= 0	a= -1	c= -2	d= -3	e= -4	f= -5	g= -6	h= -7				
i= -8	j= -9	k=-10	l=-11	m=-12	n=-13	o=-14	p=-15				
q=-16	r=-17	s=-18	t=-19	u=-20	v=-21	w=-22	x=-23				
y=-24	z=-25	{=-26	|=-27	}=-28	~=-29	=-30	.etc.				

File: Inn100.fig

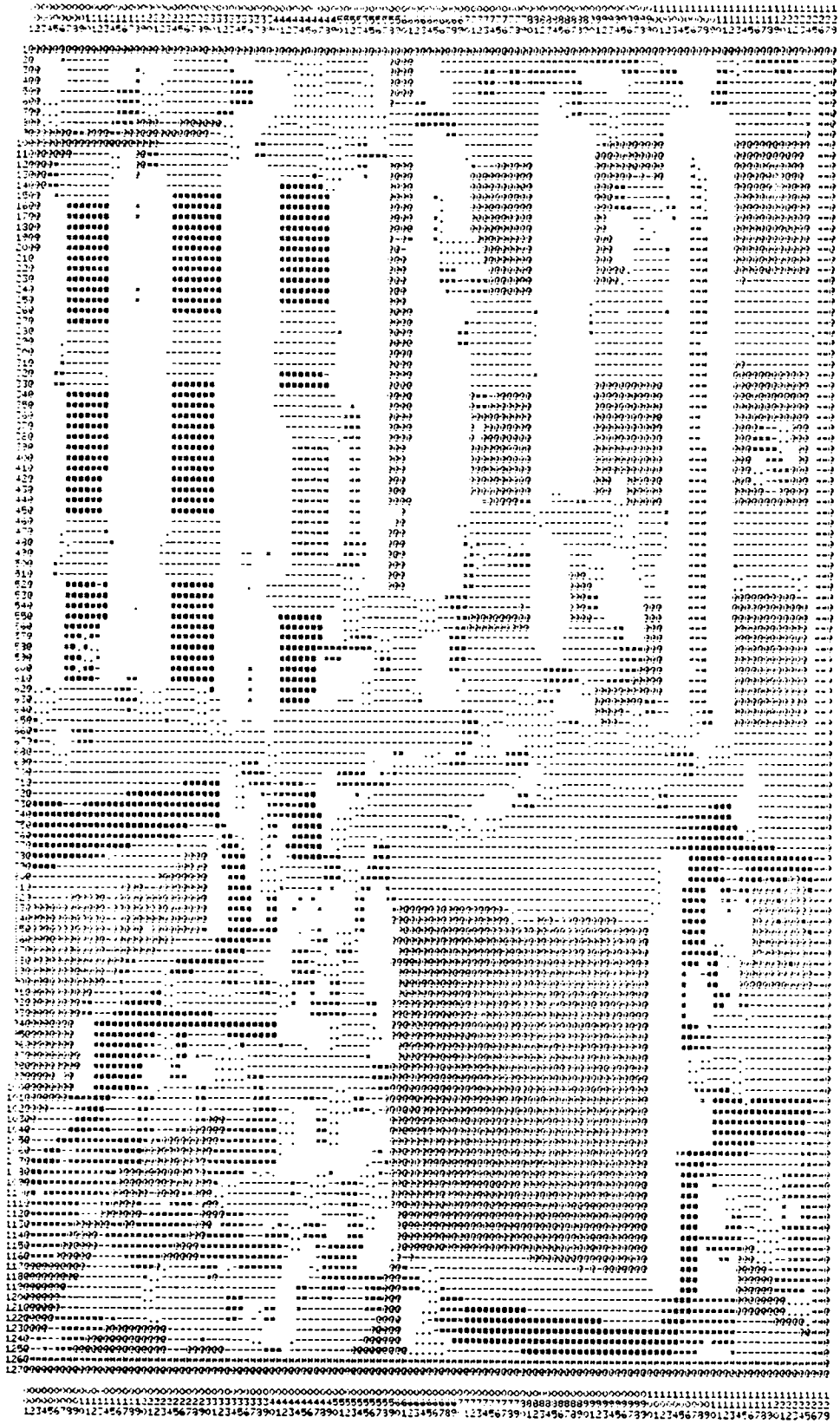
Fig.: 3.2.2-3c.



if8hkpp2.lgr

if8hkpp2.rgr

Fig.: 3.2.2-4.

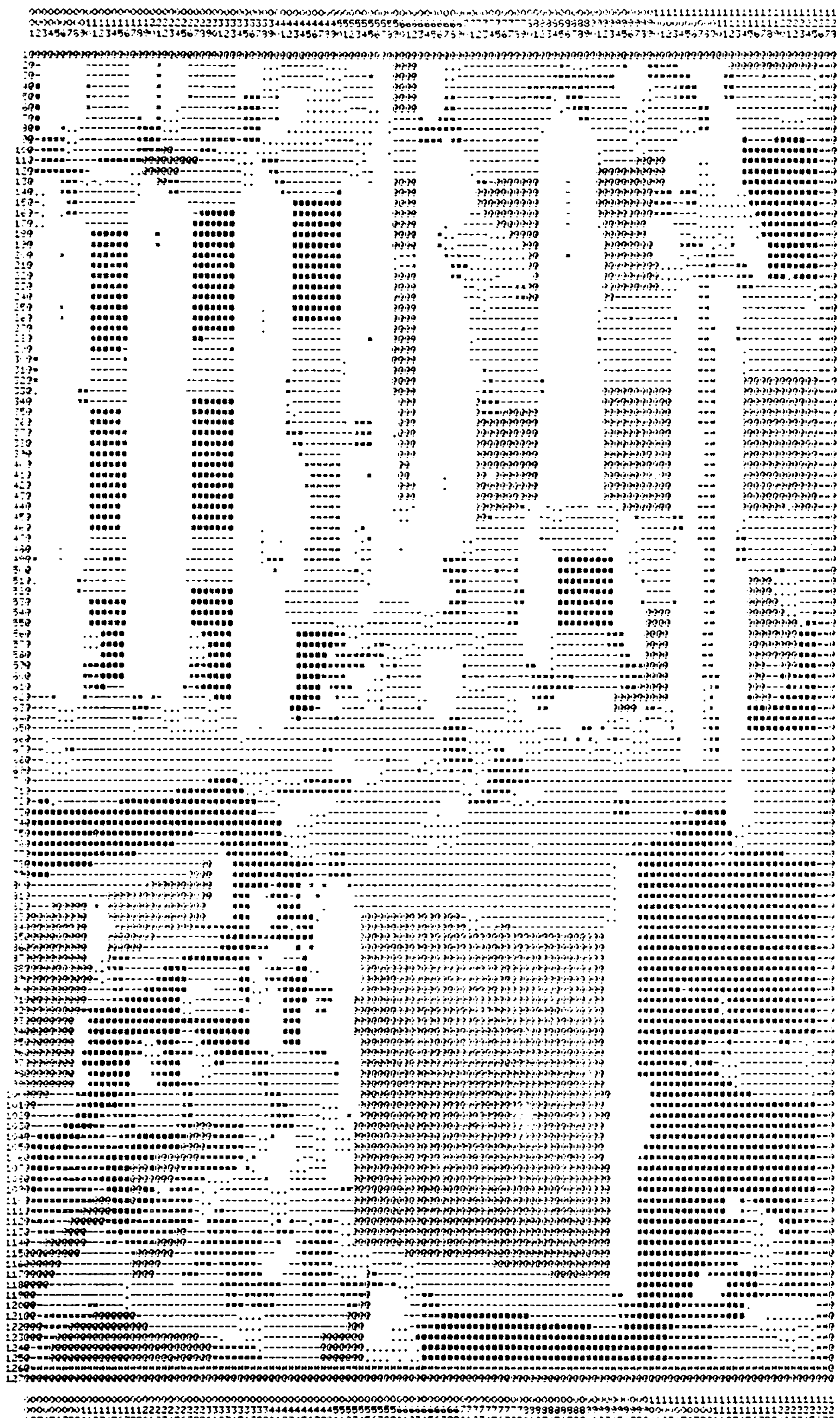


Lh(i,j) = Inr101.fig, left, class labels, most processed, "painted" in 7 colours.

FMAXGR.FMINGR.SC = 0.25500E+03 0.00000E+00 0.31373E-01
8-LEVEL GRAY IMAGE WITH OPTION FOR OVERLAY (inverse)
....-1 0 1 2 3 4 5 6 7.... (negative value = overlay)
.... . 0 * + + + - . (blank from 7 upwards)

File: Inr193.fig

Fig.: 3.2.2-5a.



Phi(i,j) = lnr102.fig, right, class labels, post processed, "painted" in 7 colours.

FMAXGR,FMINGR,SC = 0.25500E+03 0.00000E+00 0.31373E-01
3-LEVEL GRAY IMAGE WITH OPTION FOR OVERLAY (inverse)
....-1 0 1 2 3 4 5 6 7.... (negative value = overlay)
...., ? @ = = + - , (blank from 7 upwards)

File: lnr194.fig

Fig.: 3.2.2-5b.

if8hkpp3.lgr if8hkpp3.rgr

Fig.: 3.2.2-6a.

if8gray.lsh if8gray.rsh

Fig.: 3.2.2-6b.

3.2.3 Correlation matching (not completed)

Matching based on correlation has been used for a long time. In photogrammetry the correlation uses fixed size areas (regions) from the left and right images. The (possibly enhanced) gray levels are cross-correlated to find the "shift" for the best match. Alternately, Fourier methods are used for the same purpose. The present case differs in two aspects:

a) The regions to be correlated have been found by other means, i.e., they are not "arbitrarily" defined.

b) The size of the regions, their shape, and even combinations of regions may be used as additional "features" for matching.

In this form of matching each region (segment or facet) found in the L and R images is first given "individual identity" by labelling the connected groups of similar cluster labels. The cluster label may serve as a feature describing the facet or it may be ignored. Thus, either facets with similar features are correlated or all facets are correlated irrespective of their features. The shape of regions is "automatically included". The correlation method is, in principle, identical to the one used for matching binary images (where there is only one feature, i.e., 1 on a background of 0's).

In its simplest realization, the result is a hugh two-dimensional correspondence matrix C(Fm,Fn) giving the correlation between facets Fm and Fn, m = 1,2,3, ..., Nf and n = 1,2,3, ..., Nf, where Nf is the highest facet label number in the Lf(i,j) or Rf(i,j) image. Two additional matrices I(Fm,Fn) and J(Fm,Fn) give the values of the "best shift" in i- and j-coordinates indicating where the best correlation was found. The shift between the Lf(i,j) and Rf(i,j) images should correspond to physical reality (zero at fixation point and constrained to the vicinity of the epipolar line). In order to obtain unique correspondence the matrix C(Fm,Fn) is processed for the "best" correlation between a facet label pair (Fk,Fl) and the corresponding "best" shift (ik,jl) is saved, this pair is eliminated and the search is repeated for the "next best", etc. The labels in the original facet label images Lf(i,j) and Rf(i,j) are changed to new matching labels and the best shifts are used for continued processing. In a block diagram form the process is illustrated in Fig. 3.2.3-1. The programming has not been completed at the present moment.

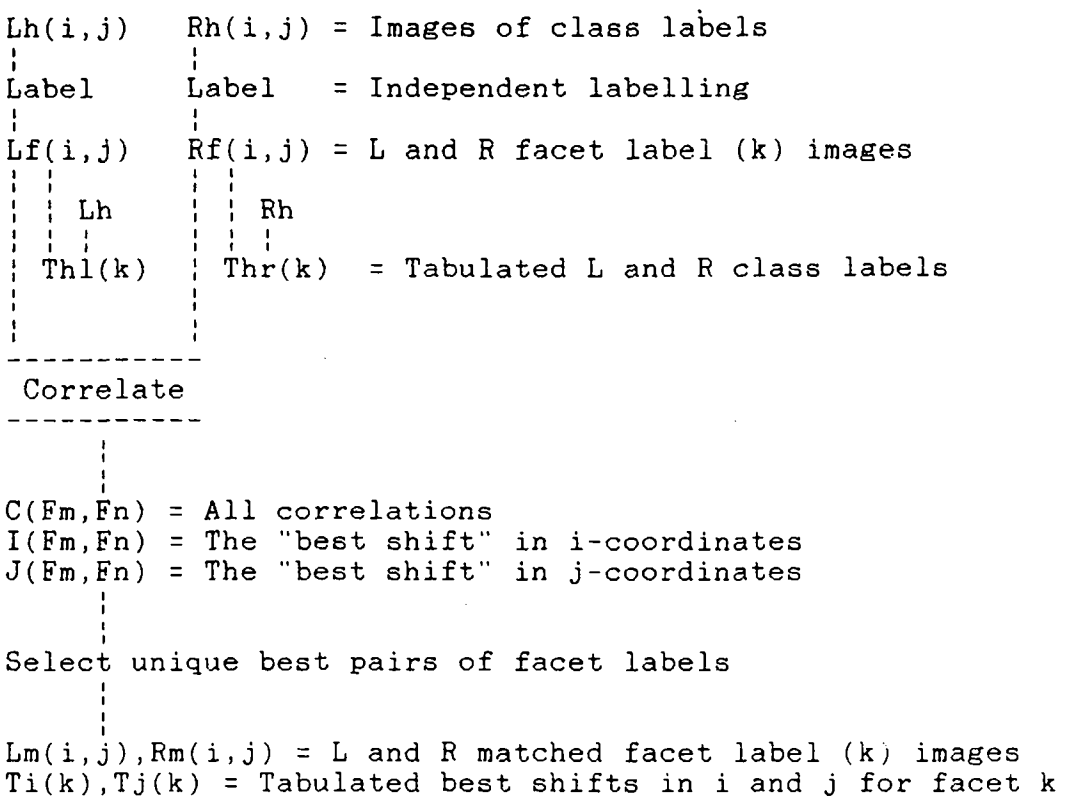


Figure 3.2.3-1: A block diagram of correlation matching.

3.2.4 Facet recognition matching

In matching based on some form of "mutual recognition" only the left versus right image "similarity principle" is used. The problem is to determine similarity between facets in the vicinity of the epipolar line given a set of facet descriptions.

In principle, the entire pattern recognition "paraphernalia" could be employed, and there are many methods available. Since there are many text on "pattern recognition", these will neither be described nor summarized. However, since only the "similarity principle" is needed and not "full recognition" of objects, these methods need to be "seen" from a slightly different point of view. For example, we may consider the left image features (near the epipolar line) as the "learning set" in terms of which the "objects" in the right image (near the epipolar line) are to be "recognized" (and/or vice versa). Thus, there is recognition in the "classical" terms (given a set of objects, acquire or learn the features for cluster or decision surface formulation, compare, etc.) but the objects "keep changing" during the process. However, the deviation from classical procedures is very slight and the needed modifications are obvious.

The number of possible descriptions for the facets is very large but it should be remembered that the segmentation process that produced the facets is prone to errors. Consequently, the descriptions should not be too "elaborate" at this stage. Furthermore, if the image contains repetitive (similar) facets then any "recognition" on the facet level is likely to produce ambiguous results (super-facets are needed). The presently used facet descriptions are:

a) Facet class label from Lh(i,j) and Rh(i,j).
b) Number of pixels per facet (area).
c) Average gray level.
d) Average x-gradient.
e) Average y-gradient.
f) A 4-dimensional shape feature.
g) Centre of gravity in image plane.

The present "shape feature" consists of the standard deviations of the projections of the facet pixels onto 0, 45, 90, and 135 degree lines (since a "central" standard deviation only requires the centre of gravity for each projection). Most likely the "best" features are those that require no definition of a coordinate system for the facet. Other shape features, such as invariant moments, P^2/A, projected shapes, and so on, are easily obtainable. However, as stated, this is only a "first try" to determine what can be achieved with very unsophisticated methods.

Since the above "features" are a mix of different measures (quantitative, ordinal, and qualitative) which are not mutually comparable (the area of a facet has no relation to its average gray level, the average gray level is at most marginally related

to the average gradient, etc.), the "decision function" was of the following form:

$$Dmn = Wf*Ef + Wa*Ea + Wg*Eg + Ws*Es + Wc*Ec$$

where:

Dmn = "Distance" between facets m and n.
Ef = 0 if facet class labels are the same and
= 1 if facet class labels differ for facets m and n.
Ea = Difference in area sizes between m and n, normalized.
Eg = Magnitude of gradient difference, normalized.
Es = Distance in 4-D "shape space", normalized.
Ec = Distance between centers of gravity.
Wf, Wa, Wg, Ws, Wc = weights.

Ea = |Am - An|/(Am + An)
Am = Area (pixel count) of facet m.
An = Area (pixel count) of facet n.
Eg = |G'm - G'n|/(|G'm| + |G'n|)
G'm = Gradient vector for facet m.
G'n = Gradient vector for facet n.
Es = Distance(m,n)/(Distance(m) + Distance(n)).
Distance(m,n) = Distance between m and n in sample space.
Distance(m) = Distance of sample m to origin.
Distance(n) = Distance of sample n to origin.
Ec = Distance((ic,jc)m,(ic,jc)n).

The match consisted in finding unique pairs of facets (m,n) constrained to the vicinity of the epipolar line for minimum Dmn. Thus, the "recognizer" is basically a minimum distance classifier. The weights allowed some "control" over the process.

The gradient directions on "cylindrical" facets are "opposed to each other" on the two sides of the "cylinder" and the directions on flat facets can be rather erratic if the gradient is small. Since the necessary modifications to compensate for these conditions were not included (at present) the weight Wg was set to zero (Wg=0) to eliminate the Eg term.

The shape descriptor Es consisted of the standard deviations of the projections of the facet pixels onto a line. Consequently, this descriptor depends upon the size of the facet and, furthermore, it was not very sensitive to the shape of the facet (as was found from experimental results). A crude normalization consisted of dividing the distance between the sample points m and n (Distance(m,n)) by the sum of their distances to the origin (Distance(m) + Distance(n)).

The distance Ec between the centers of gravity (ic,jc) was an attempt to force the facets (m and n) to be "not too far from each other" in their respective image planes. This has some validity for "far away" facets but not for near facets in the scene that are far from the centre of fixation.

As stated, the match consisted in finding unique pairs of facets (m,n) constrained to the vicinity of the epipolar line for minimum Dmn, where the weights allowed some "control" over the process. The method is crude, ad hoc, and only intended as a "first exploration" of this problem in a reasonably flexible setting. The unique difference between the present problem and the "classical" pattern recognition procedures is that a pairwise matching is required, i.e., an "object" from the left image is to be matched to exactly one "object" in the right image. This constraint partially exists in classical pattern recognition only if each object is required to be a separate class, but even then, the constraint that there must be only one object recognized per class (or none at all if the facet does not appear in one of the images) is not of the classical type. In addition there is a requirement (which in many cases is true) for an "ordering relation" between the recognized facets. However, these questions will not be considered any further.

The following two methods were used to try to resolve the uniqueness (one-to-one match) and the "mutual" recognition problems at the same time:

1. Find the largest facet "m" in the left (L) or the right (R) image (a facet which has not yet been marked as "done") and find the best matching facet "n" in the other image (R or L) which has not yet been "done". If the (minimal) distance Dmn between these facets was less than some threshold T, then declare this pair as "matched" and mark both as "done" to guarantee one-to-one matching. Iterate until all facets that can be matched are matched. The constraint to the epipolar line consisted of a highly rectangular box (Di*Dj). The centers of gravity of the two "candidates" (m and n) had to be within this box for the matching to take place. The problem with this method is that it tends to escalate errors if T is large. If an error is made at a particular facet size then all the smaller facets will have to "choose among the leftovers" for their "mate".

2. The second method was similar to the first (1) except that the "choice for a mate" was not made. Instead all the matches (Dmn) less than the threshold T were stored in a "similarity matrix" (actually a list). When no further matches could be found (or when the list was full, the number of entries stored is controlled by T), then the matches were chosen according to "best first" until no more matches were possible.

The results presented no "surprises" since the methods are classical. However, it is best to apply them as "additional help" to other methods or for verification of the results from other methods rather than to use an elementary version of recognition directly.

3.2.5 Pixel feature matching

As indicated in Chapter 2, in addition to super-regions and regions obtained from various forms of clustering to create homogeneous regions, there is a profusion of pixel features obtainable from the local operators and also some features which are strictly on the pixels level (gray level of pixel in this case). Some of these features are statistically independent but mostly the dependence is unknown. In many cases these features are not directly comparable since some may be quantitative, others ordinal, or qualitative. Preferably, the features should be "independent" but, whatever method is to be used, it should allow for various "incompatibilities".

Clearly, the more features there are the more likely it is that each pixel in the left image L and the corresponding pixel in the right image R only have one (and only one) unique combination of features, at least in the epipolar region. The problem of matching is now reduced to finding these unique combinations, whatever they may be. The hierarchical methods will help in defining the regions in the vicinity of the epipolar lines and thereby limit the possible pixels to be matched, but this form of matching will basically produce pixel level correspondences directly.

In order to carry out these experiments in an "interesting manner" (rather than via endless searching of lists) would have required in the order of M*M*N*F*P words of computer memory for an M*N image using F features and P processes to extract the left and right dependencies. The problem is rather similar to processing "stacks" of medical images from a CAT scanner, for example.

In order to see what such an approach might lead to and to accommodate the immediately available computational equipment, a very much scaled down experiment was made. The images L and R were 128*127 (M=128, N=127) with depth of 9 (M*M was restricted only to a displacement of 9 pixels), with only one feature (F=1) and one process (P=1). The feature was the cluster label used in earlier studies. (Chronologically, this was the first matching experiment as soon as the cluster processes in Chapter 2 were completed). In essence, the results of this experiment "pointed out" that the classified regions in the images could either be matched by direct "and-ing" or matched by using correlation. Much more could be said about this approach but these would only be speculations since the experiment over-taxed the computer resources and were discontinued.

3.2.6 Other methods

In the hierarchical method (structure) there are regions of different sizes (from super-regions down to a few pixels or even sub-pixels in final resampling) which are the results of various classifications (clusters) when these are represented in the image space. Preliminary matches can be obtained by "matched classification", "and-ing", correlation, or recognition, either singly, or better, in suitable combinations.

As a first approximation, these matched facets could be directly "projected" into the 3D space by estimating the rotation matrix (A) and translation vector (B), i.e., Fl = A*Fr + B, where Fl and Fr represent the pixels, centers of gravity, moments, etc., of the matching facets in the left and right images (3.1). Approximate shapes of the facets (Fl,Fr) are available from the image space, adjacencies are known in image space, etc. In the reconstructed 3D space the adjacencies are likely to remain, and the 3D space should not contain any "cracks" between the reconstructed facets. (The interiors of the facets are "empty" at present). Thus, numerous optimization methods ("energy" functions, "rubber templates", "snakes", and so on), could be formulated to "relax" all the facets at the same time. If, in addition, the facets can "acquire pixels most likely to belong to them" then the methods include segmentation correction based on both the image data and physical reality in the 3D space. A lack of time has at present prevented a "deeper look" at these possibilities.

3.2.7 Comments

The results of the matching experiments that were based on facets may be summarized as follows:

> Given large enough facets and a sufficient number of facets to "cover" the images then the matching problem is basically trivial. Exceptions occur for small facets, repetitive structures of similar facets, and in regions where the L and R images differ significantly.

Regions in the left (L) and right (R) images can differ markedly due to:

a) Reflections which superimpose more or less "transparent" facets onto existing facets. This creates a new type of image processing problem that has not been studied.

b) Loss of detail where the gray levels have become more or less "saturated" in one but not in the other image. If some significant information remains, these regions could be "equalized in gray" by some form of enhancement. The problem has not been studied in detail.

Matching failed when:

c) The facets were either too small and did not generate an "intersection" in the L&R(i,j) image or they were too different in shape to allow "mutual recognition" with sufficient reliability.

d) Due to classification errors where, for example, a small shift in the gray level in one image (L or R) places a well-matching facet into another category in the other image (R or L). This type of error is "fixed" in the next section.

e) Incorrect facet pairs are formed. These are basically impossible to detect (automatically) at this stage of processing. Only after 3D reconstruction may it be possible to find some inconsistencies, but even this is unlikely since we ourselves are subject to certain visual illusions which can be corrected by an "alternate interpretation" of the scene or by "having another look" usually from a different point of view.

f) Repetitive similar facets are present. If these facets are "displaced sufficiently" the "and-ing" process will create incorrect matches. If the facets are very similar then the recognition procedure cannot distinguish them. Clearly, such facets have to be first processed into "super-facets" by using similarity criteria before matching is attempted, i.e., the hierarchy in facet based matching needs several levels. This has not been done in the present case.

g) The same region in the 3D scene resulted in very different facets in the corresponding regions in the L and R images. For example, there is one large facet in one image but many smaller ones in the other image (see the lower right corner of the INRIA image pair). This is a rather "insidious" error and required considerable (and somewhat dubious) processing to correct.

As already implied, some of the matching errors appeared to be "correctable". Other needed refinements became also fairly obvious, which are partially investigated in the next section.

In summary, the major weaknesses of the various methods were as follows:

1. Matched classification on pixel features requires many more features than the few that were attempted. Consequently, the experiment is inconclusive.

2. Direct "and-ing" of class labels fails for small facets and can generate "catastrophic" errors if repetitive structures of (small) facets have not been grouped first.

3. Correlation matching produces ambiguous results on repetitive structures. At best these can be declared as "unmatchable" since the correlation is (nearly) the same. The experiments were not carried out.

4. Pattern recognition methods cannot distinguish between many similar facets in an image. At best these facets can be declared as "unmatchable" since they are very similar.

5. Direct pixel feature matching could not be carried out due to equipment limitations.

This leaves the question: Is there a "best combination" of methods, and if so, what is this combination?

i) The preliminary processing has to include detection of regular structures of similar facets. These structures become "super-facets". Clearly, the hierarchy needs several levels and not just the one described so far. One has to be able to "mask out" regions in the images during feature computations in order to "reprocess" selected regions. These aspects have been studied in connection with other problems but have not been included in the present case.

ii) More features and additional decision spaces should be studied.

iii) Use "and-ing" for large facets since the probability that they overlap in L and R is very high, but a few exceptions can be created if one of the cameras only "sees" the edge of a flat surface while the other sees much more of the same surface. The ultimate extreme is to place the cameras so that one camera sees one side and the other the other side of a thin flat sheet (in 3D space).

iv) Apply recognition procedures to the leftover facets after direct "and-ing" (iii).

v) Carry out various "matching correction" procedures to be described in the next section. Once the nature of a problem is known then something can be done about it, unless the weakness in a method is "fatal".

3.3 Modifying the various matches

The "raw" matching results from the previous section are of the following two types:

a) An "intersection" image L&Rh(i,j) giving the class labels which intersected in the two images. The intersection could simply be of the "and" type, or a more "refined" one obtained from "matched classification". The L&Rh(i,j) image is accompanied by the left and right class label images Lh(i,j) and Rh(i,j). Independent labelling of the connected regions followed by "match (intersection) resolution" produces a matched pair of images of the second type (b).

b) A left and right matched image pair Lm(i,j) and Rm(i,j) where the matching facets have the same label numbers. Non-matching facets are given negative and differing numbers in Lm(i,j) and Rm(i,j) in order not to loose facet information. The Lm(i,j) and Rm(i,j) image pair may or may not be accompanied by i- and j- coordinate "displacement" lists Ti(k) and Tj(k) for the matched facets k, k = 2,3,..., Kmax. The label k=1 is not used for "historic" reasons (old programs needed 1 to indicate "raw binary" data).

Several new problems were encountered, which will be considered in some detail in the following sub-sections. However, briefly stated:

1. Sets of connected regions in the L&Rh(i,j) "intersection" or "coincidence" image frequently "implicated" sets of facets in the Lh(i,j) and Rh(i,j) images. This required "match (intersection) resolution".

2. Depending on which process was used previously and the tolerance settings, two supposedly "matched" facets could be of markedly different shape.

3. Obvious errors occurred where correct matching was "missed" due to differences in original classification.

4. Questions arose about what to do with the unmatched, un-matchable, and undefined (unlabelled) facets.

5. Computation of suitably "balanced" analytic parameters for the facets in order to try "matched" analytic relaxation (in section 3.4).

Methods of checking that the results after processing are at least approximately correct were the same as before, i.e., visual inspection of stereo pairs of results and the "film-loop" in the Vicom image processing and display system.

3.3.1 The Lh, L&Rh, and Rh triplet

The "coincidence" or "intersection" image L&Rh(i,j) gives the class labels which intersected in the two individually classified Lh(i,j) and Rh(i,j) images. In the more "refined" or "matched classification", the L&Rh(i,j) image contained "matched" regions which could be considerably larger than those obtained in simple "and-ing". Both types of L&Rh(i,j) images were accompanied by the left and right class label images Lh(i,j) and Rh(i,j).

The regions in the "more refined" L&Rh(i,j) image were modified by requiring a match between L&Rh(i,j), Lh(i,j), and Rh(i,j) class labels or by requiring at least that one of the labels in Lh(i,j) or Rh(i,j) be nonzero. This operation reduced the L&Rh(i,j) image to the same form as that obtained from direct "and-ing". This simplified the subsequent processing.

The subsequent processing steps consisted of the following, see Figure 3.3.1-1:

1. Independently label the connected regions with similar class labels in the L&Rh(i,j), Lh(i,j), and Rh(i,j) images. Call these images L&Rf0(i,j), Lf0(i,j), and Rf0(i,j).

2. Compute the sizes and "thinnesses" of the labelled regions in L&Rf0(i,j), Lf0(i,j), and Rf0(i,j).

3. Reject all regions in L&Rf0(i,j), Lf0(i,j), and Rf0(i,j) that were either too small or too thin to allow subsequent analytic approximation.

4. Relabel according to area size the "surviving" facets in L&Rf0(i,j), Lf0(i,j), and Rf0(i,j). These images will now be referred to as L&Rf1(i,j), Lf1(i,j), and Rf1(i,j), where the "f" indicates facet labels and the "1" indicates an initial image.

5. Resolve the "set intersection" (match or intersection resolution) problem since sets of connected regions (facets) in the L&Rf1(i,j) image frequently indicated sets of facets in Lf1(i,j) and Rf1(i,j) images.

The problem (5) could be resolved by creating a hugh "coincidence" matrix, or by processing lists, which was used and which became an interesting programming exercise. The problem is resolved by finding unique label pairs in Lf1(i,j) and Rf1(i,j) which have the highest "pixel count" in L&Rf1(i,j), these are eliminated and the comparison problem is repeated until the pixel count is "too low". Additional requirements are that the matching pair in Lf1(i,j) and Rf1(i,j) have approximately the same areas (sizes) and that the intersection areas are "comparable" to the areas of the two facets. (Additional "pattern recognition" procedures have not yet been introduced). Thus:

```
If((Amn.Gt.T1).And.(Bmn.Gt.T2).And.(Cmn.Gt.T3)) Then
  assign a new label to the facet pair in Lf1(i,j) and Rf1(i,j).
```

where:

Anm = Intersection area from L&Rf1(i,j) for facets m and n in Lf1(i,j) and Rf1(i,j).

Bmn = Amn/((Am+An)/2) = Normalized intersection area size for facets m and n.

Cmn = |Am-An|/(Am+An) = Normalized area difference for facets m and n.

Am = Area of facet m.
An = Area of facet n.
T1 = Threshold
T2 = Threshold
T3 = Threshold

The new labels are tabulated in terms of the facet labels in Lf1(i,j) and Rf1(i,j) and a new matched set of labelled images called Lm1(i,j) and Rm1(i,j) is created. The labels that could not be matched are set negative and they differ in Lm1(i,j) and Rm1(i,j) (but the label number assignment is continued such that the absolute value of a label can be used in addressing).

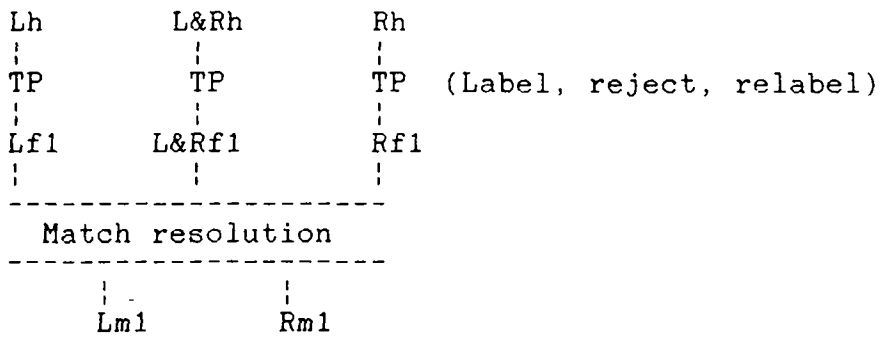
Three "heuristic" parameters, namely T1, T2, and T3, have been introduced for the following reasons:

T1 is to eliminate very small intersections in L&Rf1(i,j) which are basically unreliable.

T2 is a limit on the intersection area size for the two facets. Large regions should intersect more than small ones on a relative scale.

T3 is a "mutual recognition" limit. If the two facets differ widely in size, clearly they should not be considered as matching.

Prior experiments indicated that if the heuristic limits (T1, T2, T3) are absent then facets of rather different sizes could be associated despite the "best match first" condition in the pairing.



The process as shown previously.

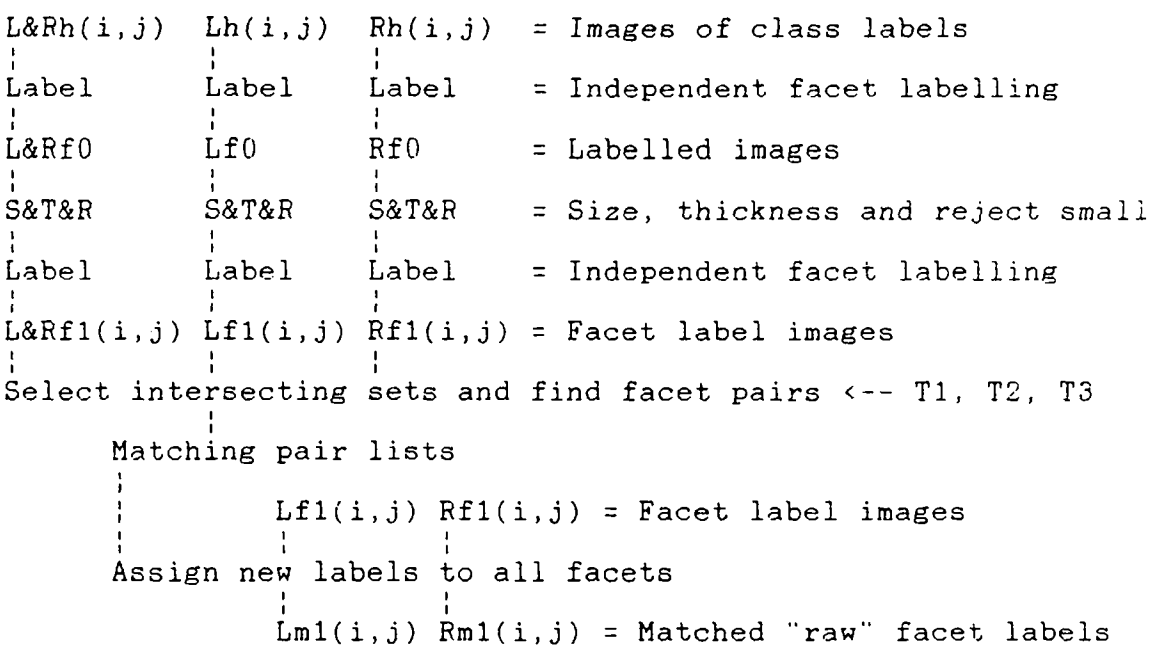


Figure 3.3.1-1: A block diagram of the processing of the L&Rh(i,-j), Lh(i,j), and Rh(i,j) images of class labels.

(Relevant computer files: /user2/kasvand/ima2):
L&Rh = Inr063.fig Lh = Inr065.fig Rh = Inr066.fig
Lf1 = Inr069.fig Rf1 = Inr070.fig
Matching pair lists = Inr071.fig
Lm1 = Inr073.fig Rm1 = Inr074.fig

A rather extreme output image triplet L&Rh(i,j), Lh(i,j), and Rh(i,j) was obtained from "Matched Classification" (Figures 3.2.1-2, 3.2.1-3a and -3b) and repeated as Figures 3.3.1-2a,-2b, and -2c. In this case the tolerances were intentionally made very liberal in order to illustrate what happens. After labelling and removal of small and/or very thin facets, the "surviving" facet labels Lf1(i,j) and Rf1(i,j) are shown in Figures 3.3.1-3a and -3b. The matching pair list is not shown since it contains about 220 rows of entries comprehensible only if the details of the programs are known. The resultant "raw" matched facet labels Lm1(i,j) and Rm1(i,j) are shown in Figures 3.3.1-4a and -4b. The matched facets are shown with capital letters while the unmatched ones use lower case letters.

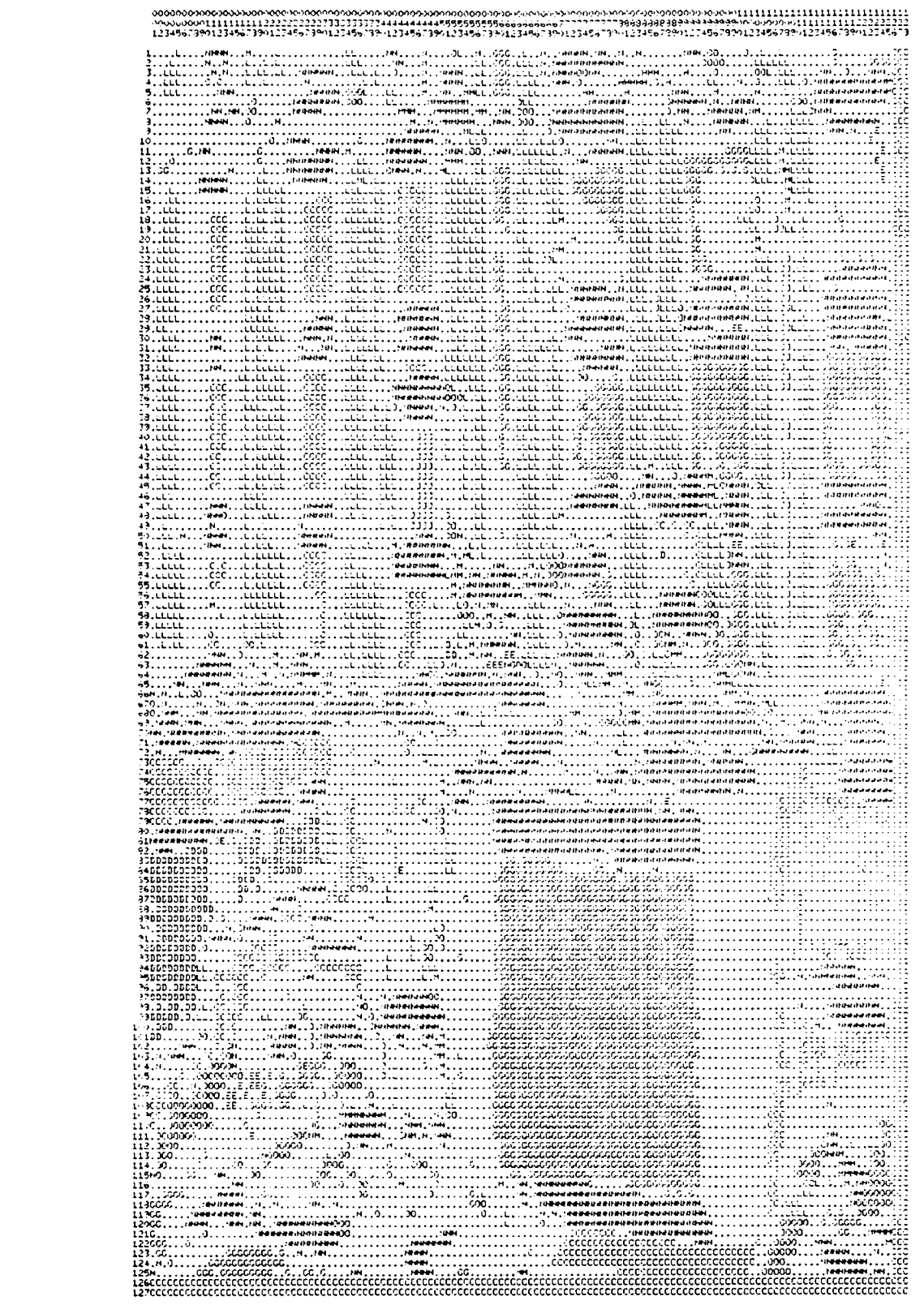
A closer examination of the results (Lm1(i,j) and Rm1(i,j) images) reveals that most of the matches appear to be correct, but there are also some rather comical matches which are not physically feasible in a 3D world. However, the match resolution program knows nothing about the 3D world. This matching should be followed by "recognition matching" in order to pair more facets since most of the unmatched facets are very similar but shifted "beyond the range" of these procedures.

Figure titles:

Figures 3.3.1-2a, -2b, and -2c: Output image triplet L&Rh(i,j), Lh(i,j), and Rh(i,j) obtained from "Matched Classification" (Figures 3.2.1-2, 3.2.1-3a and -3b).
(Files: Inr063.fig, Inr065.fig, Inr066.fig)

Figures 3.3.1-3a and -3b: The "surviving" facet labels Lf1(i,j) and Rf1(i,j) after removal of small and/or very thin facets.
(Files: Inr069.fig, Inr070.fig)

Figures 3.3.1-4a and -4b: The "raw" matched facet labels Lm1(i,j) and Rm1(i,j), same as Figs. 3.3.2-1a and -1b.
(Files: Inr073.fig, Inr074.fig, Matching list = Inr071.fig)



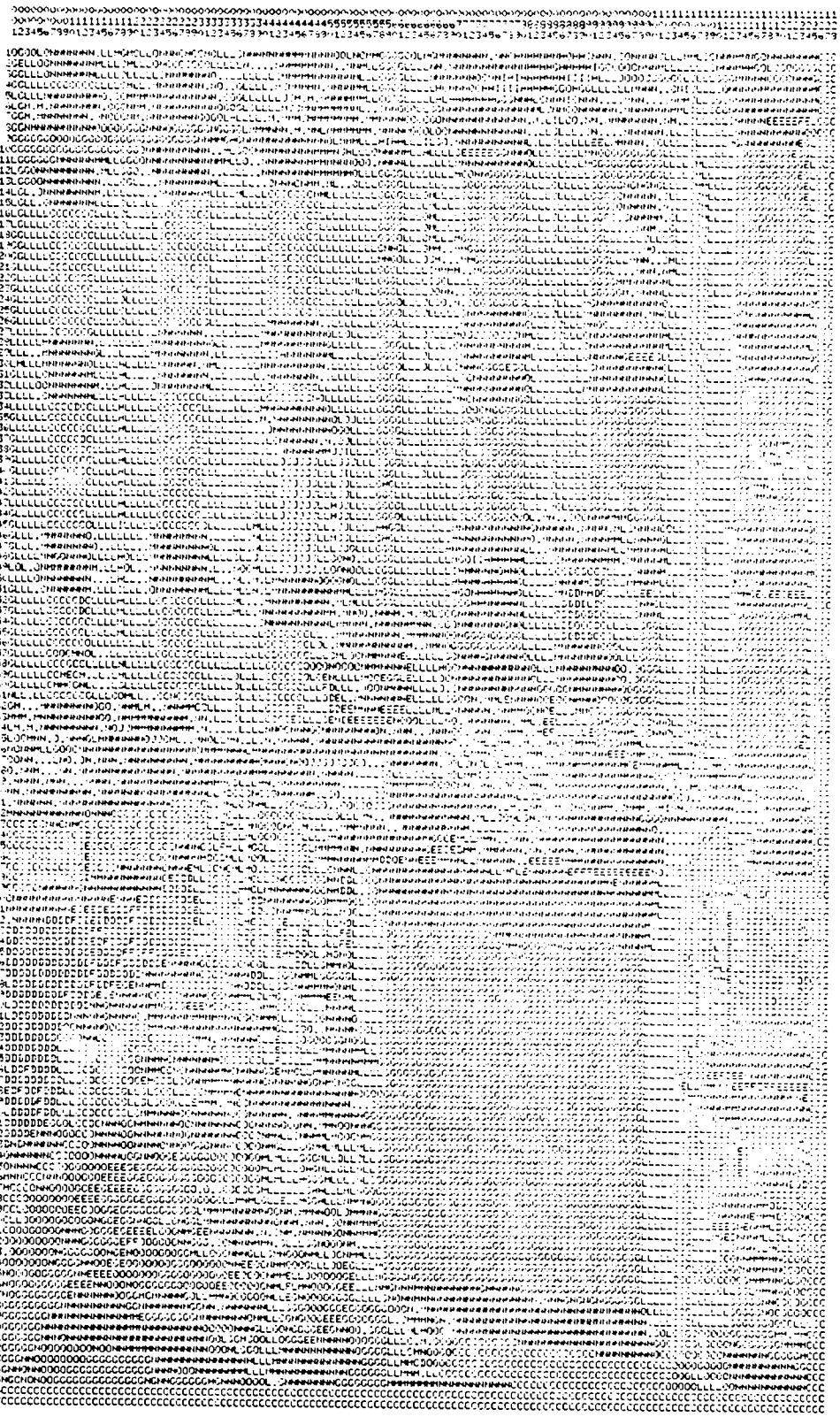
LERN(i,j). class labels from matched classification. font1,2,3 = 1,9,5.

LABEL PRINTING

.= 0	== 1	C= 2	D= 3	E= 4	F= 5	G= 6	H= 7		
I= 8	J= 9	K= 10	L= 11	M= 12	N= 13	O= 14	P= 15		
Q= 16	R= 17	S= 18	T= 19	U= 20	V= 21	W= 22	X= 23		
Y= 24	Z= 25	[= 26	\= 27	]= 28	^= 29	_= 30	`ETC.		
a= 0	b= -1	c= -2	d= -3	e= -4	f= -5	g= -6	h= -7		
i= -8	j= -9	k=-10	l=-11	m=-12	n=-13	o=-14	p=-15		
q=-16	r=-17	s=-18	t=-19	u=-20	v=-21	w=-22	x=-23		
y=-24	z=-25	{=-26	|=-27	}=-28	~=-29	=-30	.etc.		

File: lnr063.fig

Fig.: 3.3.1-2a.



Lh(i,j). left image using matched classification with Hl(Gpm,G) and Hl(Gpm,Gpa).

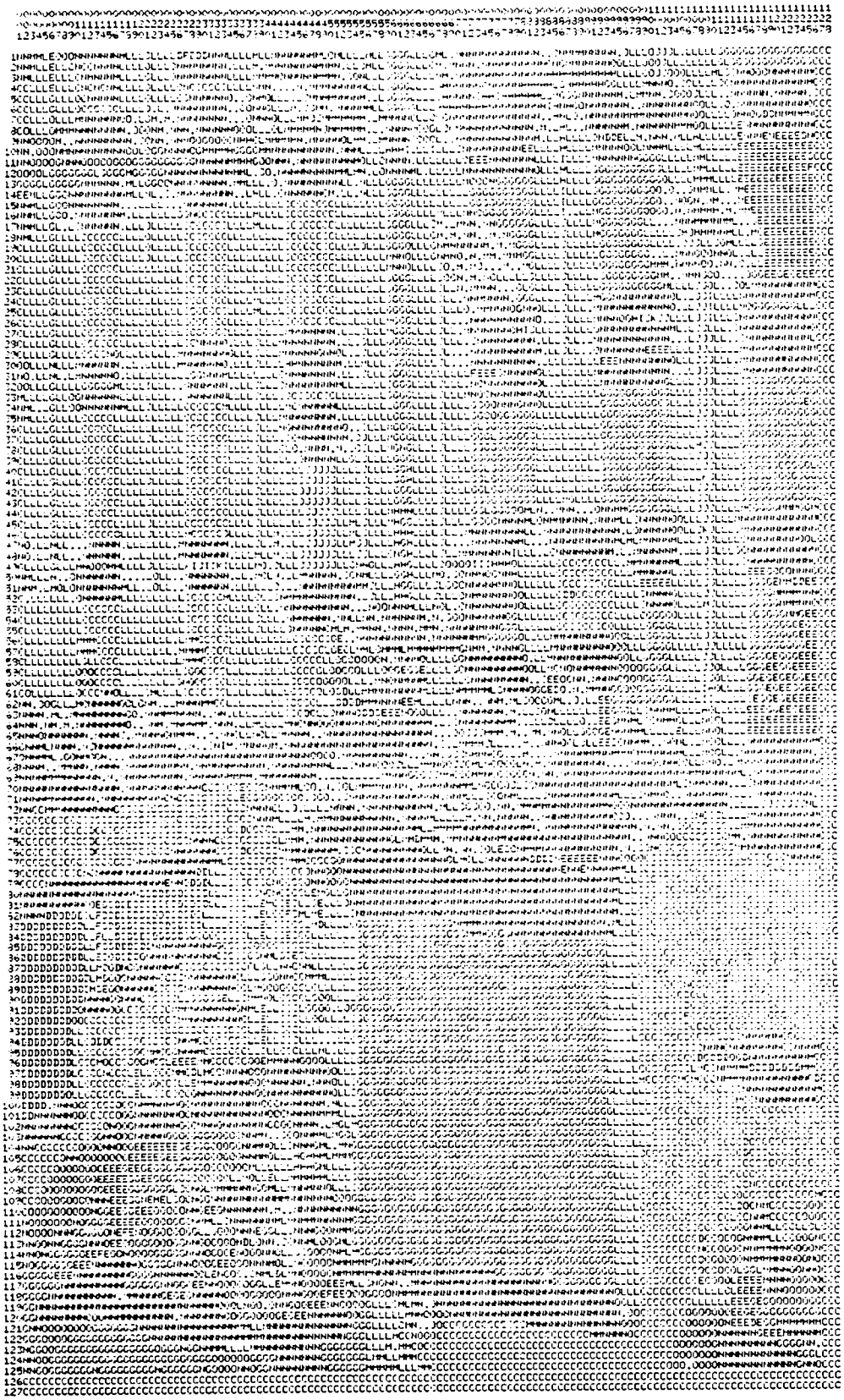
LABEL PRINTING

.= 0	A= 1	C= 2	D= 3	E= 4	F= 5	G= 6	H= 7
I= 8	J= 9	K= 10	L= 11	M= 12	N= 13	O= 14	P= 15
Q= 16	R= 17	S= 18	T= 19	U= 20	V= 21	W= 22	X= 23
Y= 24	Z= 25	[= 26	A= 27	B= 28	C= 29	D= 30	.ETC.
.= 0	@= -1	c= -2	d= -3	e= -4	f= -5	g= -6	h= -7
i= -8	j= -9	k=-10	l=-11	m=-12	n=-13	o=-14	p=-15
q=-16	r=-17	s=-18	t=-19	u=-20	v=-21	w=-22	x=-23
y=-24	z=-25	(=-26	a=-27	b=-28	c=-29	d=-30	.etc.

File: Inr065.fig

Some program parameters:
NSTEPDM.NSTEPFB.NSTEPSE.IOPT.LASTGMG.IOPT2.IOPT3.NJUMP0 =
2 3 4 -1 3 0))
DGG.DGM.DGA = 0.50000E+03 0.20000E+03 0.20000E+01
FMULR = 0.10000E+03

Fig.: 3.3.1-2b.



Rh(i,j), right image using matched classification with Hl(Gpm,G) and Hl(Gpm,Gpa).

LABEL PRINTING

.= 0	A= 1	C= 2	D= 3	E= 4	F= 5	G= 6	H= 7
I= 8	J= 9	K= 10	L= 11	M= 12	N= 13	O= 14	P= 15
Q= 16	R= 17	S= 18	T= 19	U= 20	V= 21	W= 22	X= 23
Y= 24	Z= 25	[= 26	A= 27	B= 28	C= 29	D= 30	.ETC.
.= 0	@= -1	c= -2	d= -3	e= -4	f= -5	g= -6	h= -7
i= -8	j= -9	k=-10	l=-11	m=-12	n=-13	o=-14	p=-15
q=-16	r=-17	s=-18	t=-19	u=-20	v=-21	w=-22	x=-23
y=-24	z=-25	<=-26	a=-27	b=-28	c=-29	d=-30	.etc.

File: Inr066.fig

Some program parameters:

NSTEPDM,NSTEPFB,NSTEPSE,IOPT,LASTGMG,IOPT2,IOPT3,NJUMPO =
2 3 4 -1 9 0 0 0
DGG,DGM,DGA = 0.50000E+03 0.20000E+03 0.20000E+01
FMULR = 0.10000E+03

Fig.: 3.3.1-2c.

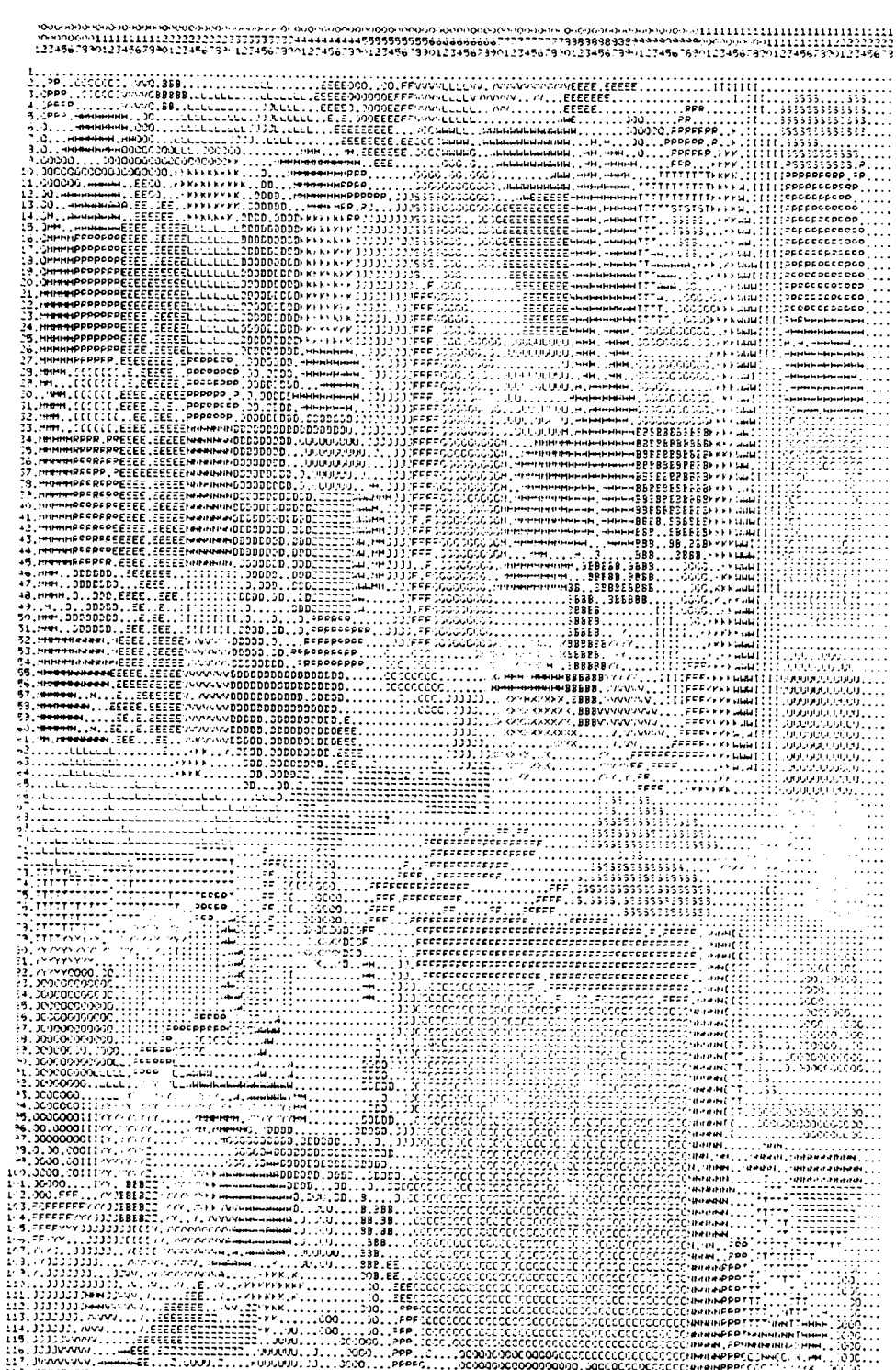
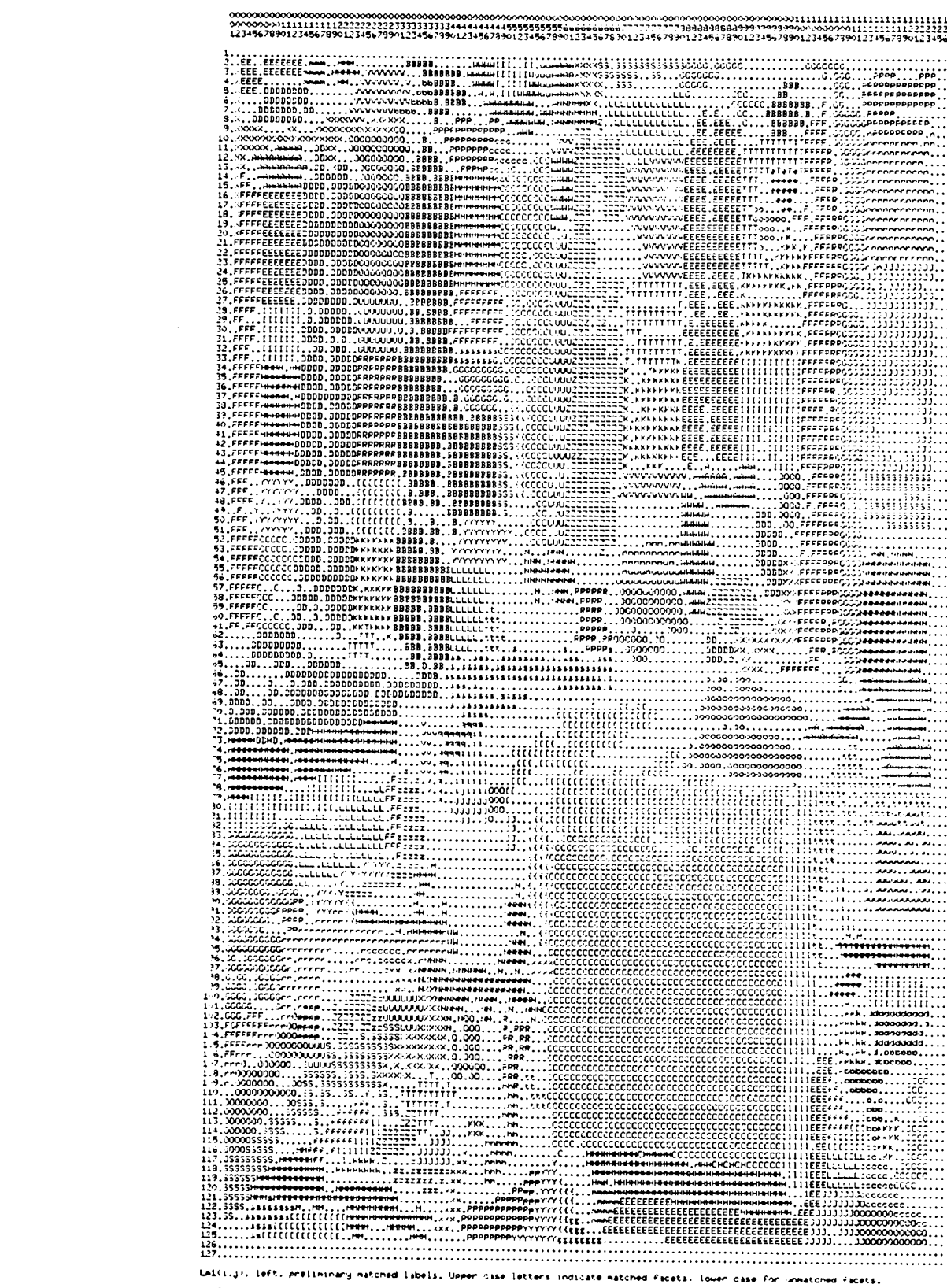


Fig.: 3.3.1-3a.

Fig.: 3.3.1-3b.



Lai(i,j), left, preliminary matched labels. Upper case letters indicate matched facets, lower case for unmatched facets.

LABEL PRINTING

.=	0	==	1	0=	2	D=	3	E=	4
I=	8	J=	9	K=	10	L=	11	M=	12
O=	16	P=	17	S=	18	T=	19	U=	20
V=	24	X=	25	[=	26	&=	27	$=	28
.=	0	9=	-1	c=	-2	d=	-3	e=	-4
i=	-8	j=	-9	k=	-10	l=	-11	m=	-12
q=	-16	r=	-17	s=	-18	t=	-19	u=	-20
y=	-24	z=	-25	{=	-26	&=	-27	$=	-28

F=	5	G=	6	H=	7				
N=	13	O=	14	P=	15				
V=	21	W=	22	X=	23				
D=	30	.ETC.							
f=	-5	g=	-6	h=	-7				
n=	-13	o=	-14	p=	-15				
v=	-21	w=	-22	x=	-23				
d=	-30	.etc.							

File: lw073.fig

Fig.: 3.3.1-4a.

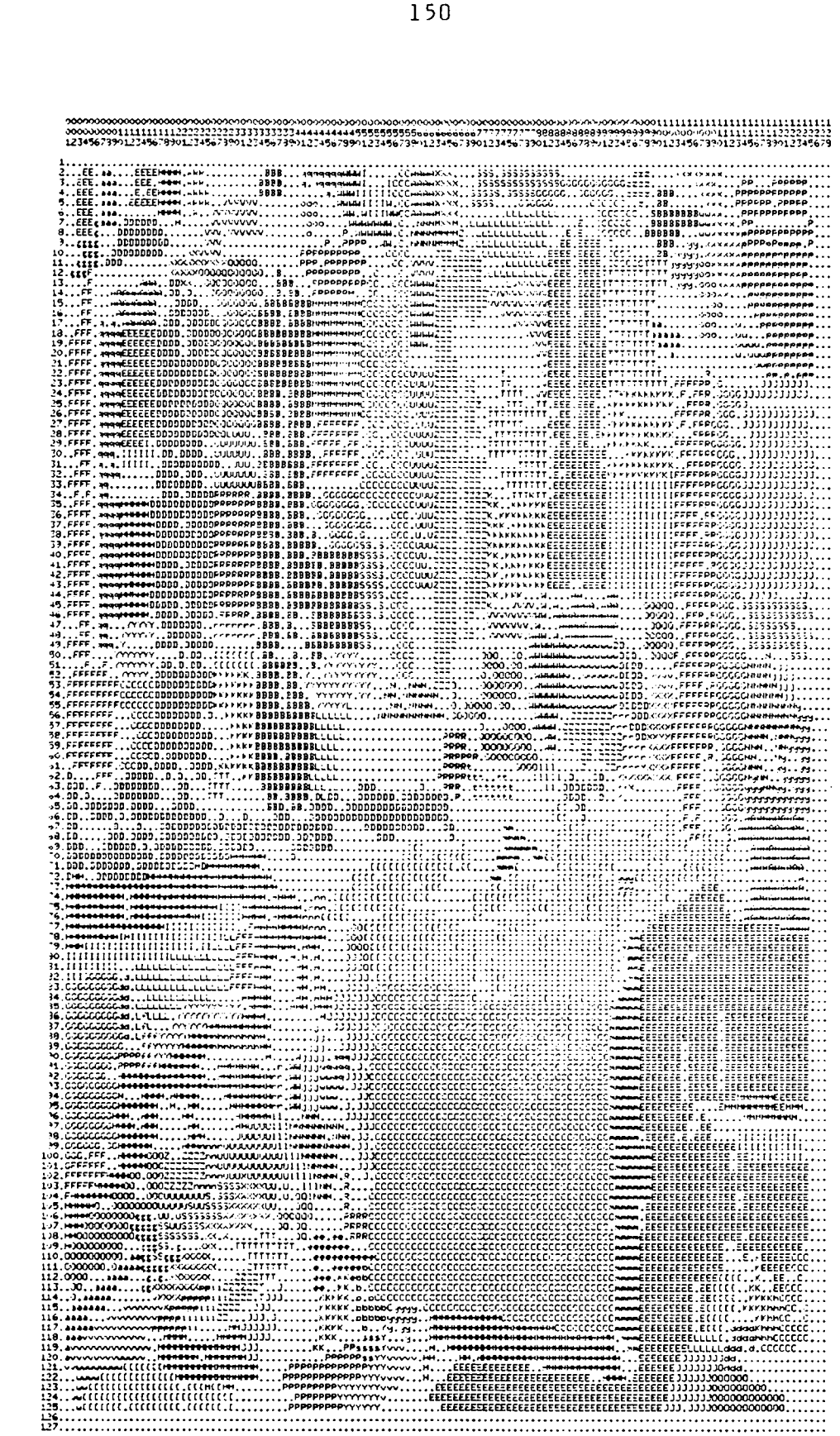


Fig.: 3.3.1-4b.

3.3.2 The raw Lm and Rm images

After the initial or "raw" matching was achieved (Lm1,Rm1) it was clear that there are numerous rather basic and possibly "philosophical" problems that have to be solved. The raw matched left and right images, i.e., Lm1(i,j) and Rm1(i,j), contained:

a) Mainly correctly matched facets but some of these facets could be rather different in size if the tolerances were lowered.

b) Unmatchable facets without one-to-one correspondence between the facets. For example, in one image a region could be split into two while in the other it remained a single region. The best matching "splinter" was matched leaving the other "splinter" as "unmatchable".

c) Facets that should have been matched but were not matched due to different classification. For example, the region in one image could be darker than in the other image and, if the gray level was close to a class boundary, the regions became classified as "different".

d) Unmatched facets that were "too far" from each other and did not generate an intersection in the L&Rh(i,j) image or the intersection was too small and was rejected.

e) Unmatched facets that could not be matched since there was no corresponding facet in the other image or its "partner" had already been "grabbed" by another facet (see (b)).

f) Matches which appeared to be errors since verification of the initial raw results is unreliable.

Most of these problems could be attributed to three causes, namely:

i) Independent segmentation of the images produces differing sizes of facets in certain regions of the images. The so-called "matched classification" and the use of common decision spaces did not improve the situation markedly. Thus, in "still-life" stereo a certain number of segmentation errors will have to be accepted and dealt with in subsequent processing.

ii) Facets which should have been "paired" could not be paired since their intersection in the L&Rh(i,j) image was zero or "too small", the initial classification was wrong, or the "partner" has already been "taken".

iii) Facets in one image did not exist in the other image or the scene was rather different in certain regions mainly due to reflections and partially due to the geometry of the scene.

Numerous attempts were made to try to deal with these problems. There are no particularly strong "theoretical" justifications to what has been attempted, except to say that "at the time they appeared to be 'reasonable' thing to do". A part of the image pair is shown in Figures 3.3.2-1a and -1b, which are portions of Figures 3.3.1-4a and -4b. Careful inspection reveals that most of errors can be found in these images. However, it should not be forgotten that the printing of the labels is in modulu 26 and it can happen that different but touching labels have the same letter code.

Figure titles:

Figures 3.3.2-1a and -1b: Magnified extracts of the "raw" matched facet label images Lm1(i,j) and Rm1(i,j), from Figures 3.3.1-4a and -4b. (Files: Inr073.fig, Inr074.fig)

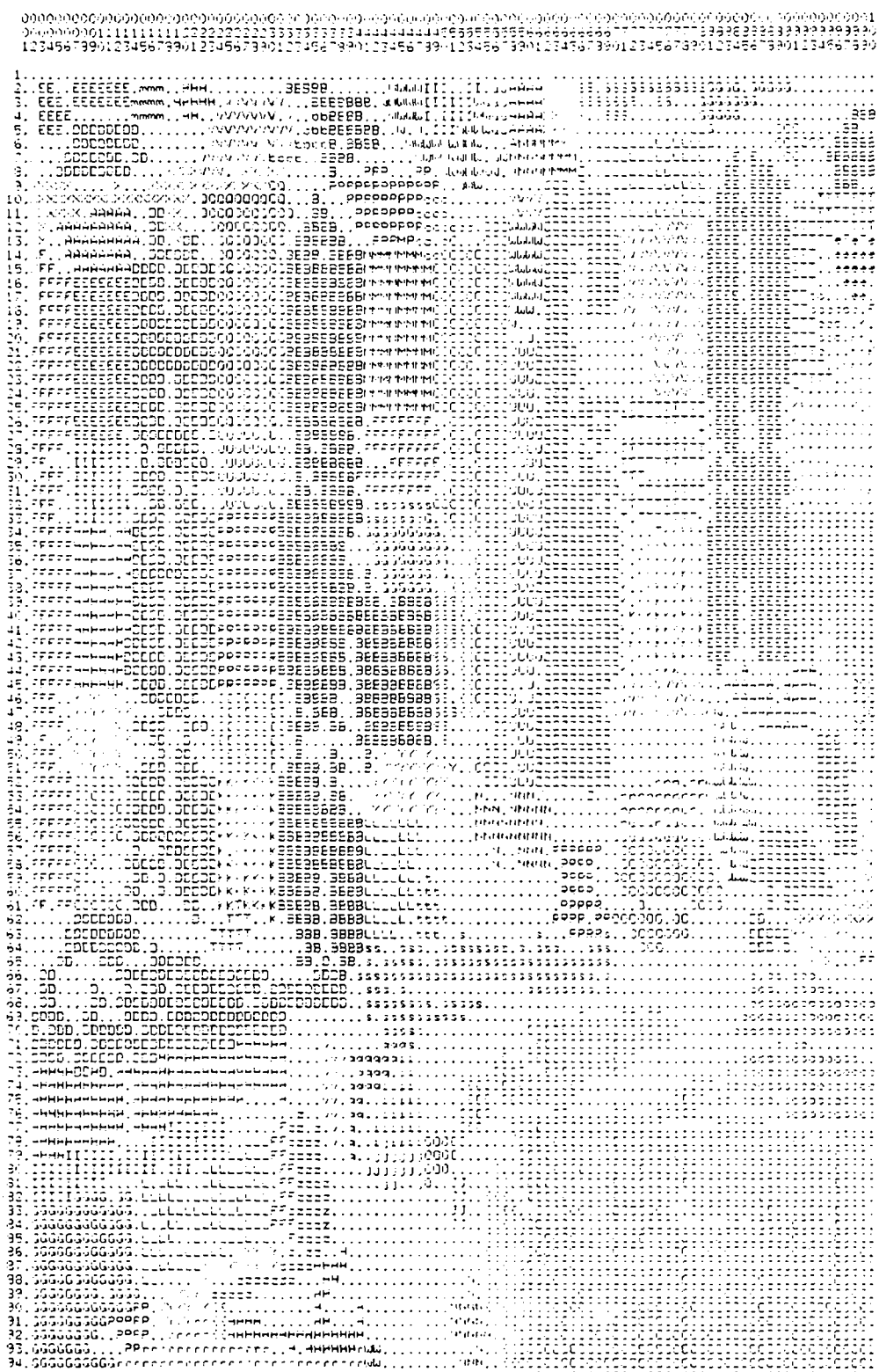


Fig.: 3.3.2-1a.

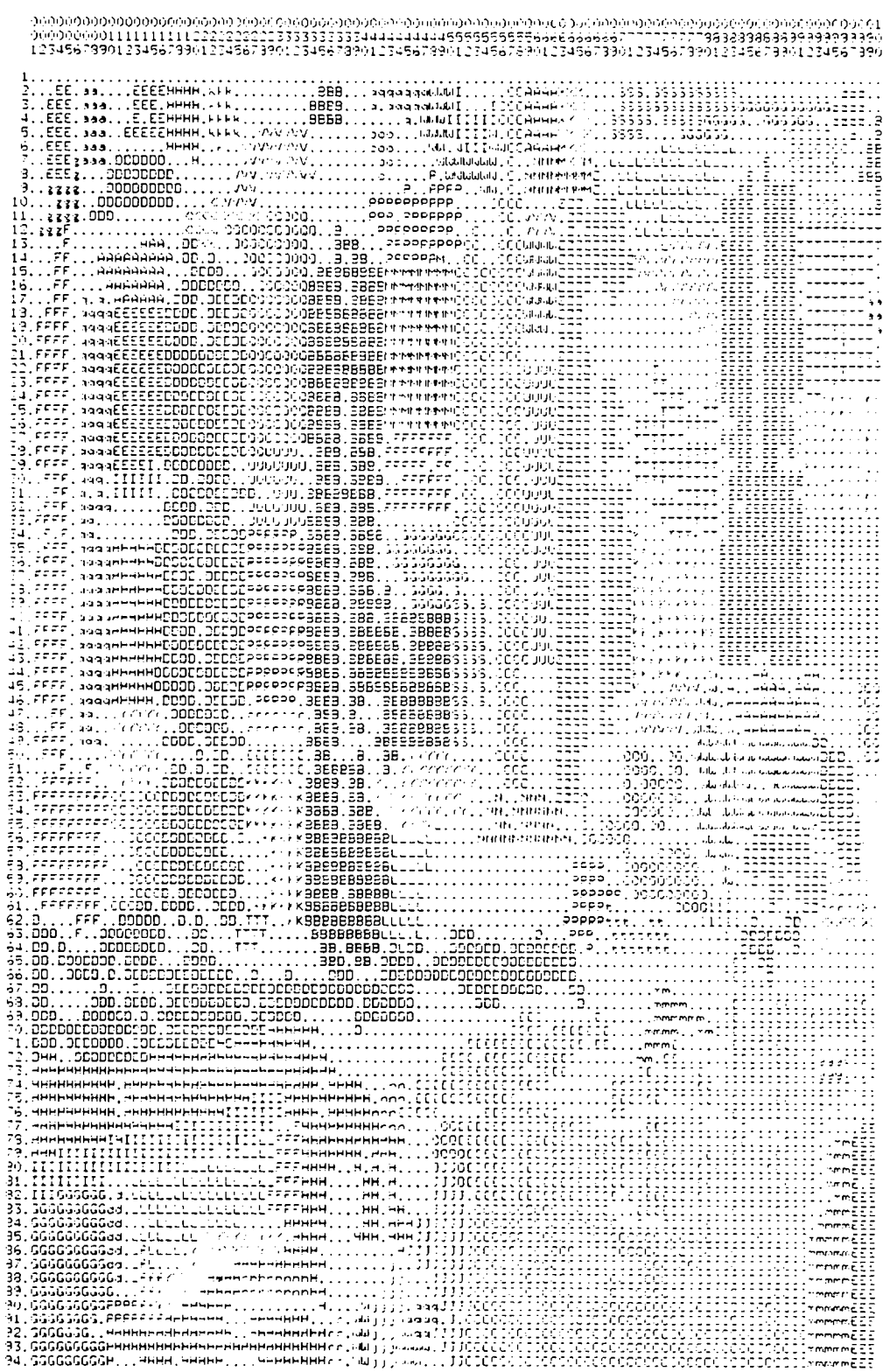


Fig.: 3.3.2-1b.

3.3.3 Misclassifications

Misclassifications were mainly caused by variations in gray levels between the two images (Gl(i,j) versus Gr(i,j)) when the gray levels were very close to a class boundary. Since the classes in this case were not "widely separated" this situation will always occur with a segmentation method based on classification (of the whole image in one process).

This problem is easily solved by constructing a coincidence matrix of the unpaired facets and by using the "best match first" principle for assigning the pair a new label. The process is identical to the initial pairing of facets but without requiring a match of class labels. A "finessing" consists of also computing a "distance" between the classes and including the distance in the "best match first" criterion. In the present case the pairing was simply terminated as in the case of initial matching (when the intersection pixel count was less than a threshold).

An edited extract of a part of the "history file" is shown in Figure 3.3.3-1. In this figure the coincidence matrix indicates that the match between labels "-121=n" and "-123=p" has 108 coincident pixels (with temporary addresses (1,1) in the coincidence matrix), located at (i,j)=(118,16) and (i,j)=(120,14), etc. The result after correcting some misclassifications is shown in Figures 3.3.3-2a and -2b.

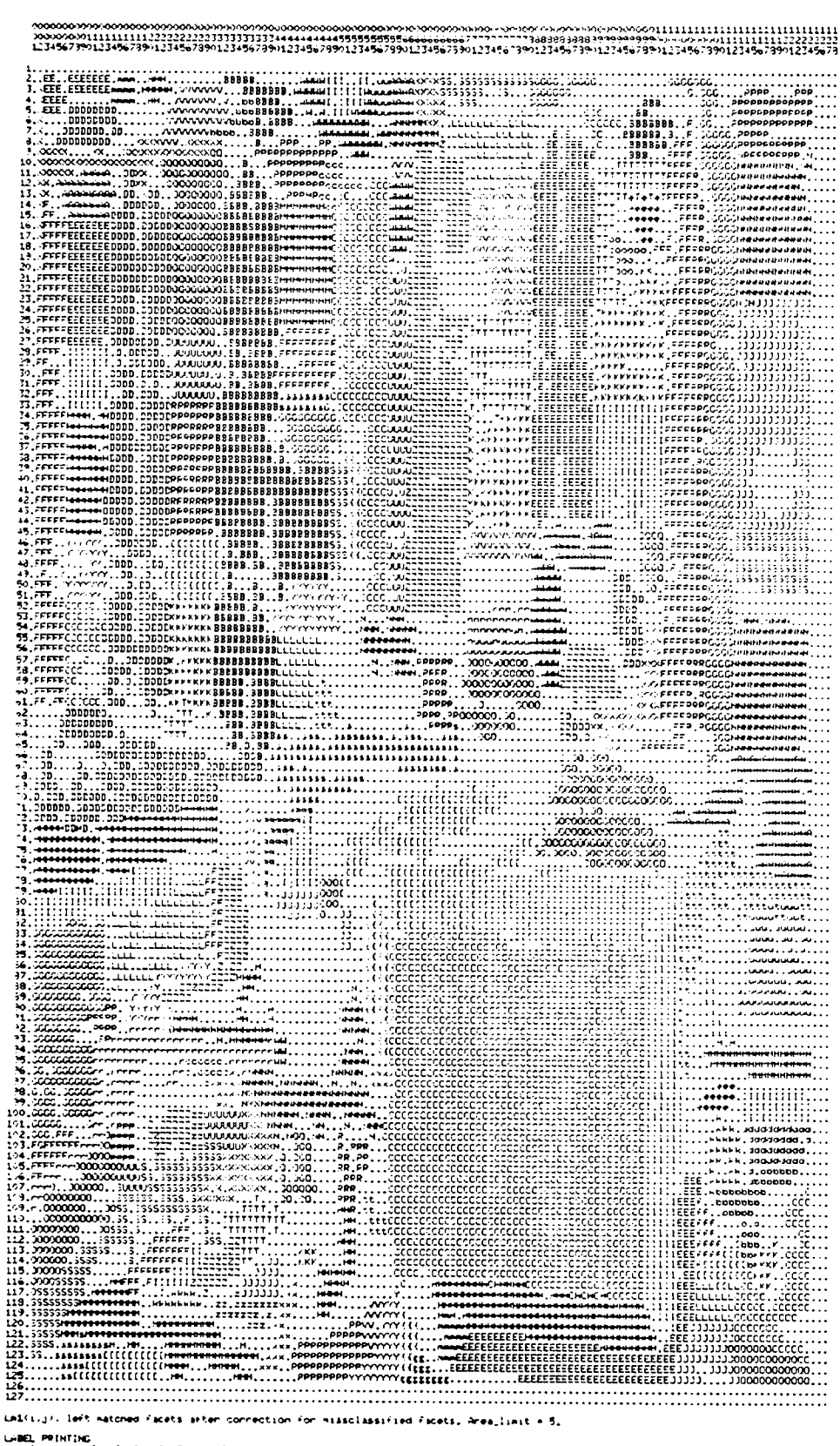
Figure titles:

Figure 3.3.3-1: Misclassification correction, edited extract from "history file". (File: Inr076.fig, extract from Inr075.fig)

Figures 3.3.3-2a and -2b: The result after correcting some misclassifications in the images shown in Figures 3.3.1-4a and -4b. (Files: Inr077.fig, Inr078.fig)

```
Corrections for miss-classifications.
(Find matching negative labels after 'match resolution'.)


Coincidence matrix:

 Ja\Ia
      1    2    3    4    5    6    7    8    9   10   11   12   13
  1 108    0    0    0    0    0    0    0    0    0    0    0    0
  2   0    0    0    0    0    2    0    0    0    0    0    0    0
  3   5    0    0    0    0    0    0    0    0    0    0    0    0
  4   0    0    0    0    0    0    2    0    0    0    0    0    0
  5   0    0    0    0    0   17    0    0    0    0    0    0    0
  6   0    0    0    0    0    0    0    0    0    0    0    6    0
  7   0    0    0    0    0    0    0    0    2    0    0    0    0
  8   0    0    0    0    0    0    8    0    0    0    0    0    0
  9   0    0    0    0    8    0    0    0    0    0    0    0    0
 10   0    0    0    7    0    0    0    0    0    0    0    0    0
 11   0    0    0    0    0    1    0    0    0    0   14    0    0
 12   0    0    0    0    0    0    0    0    0    0    0    0    1
 13  13    0    0    0    0    0    0    0    0    0    0    0    0
 14   0    0    0    0    0    0    0    0    0    1    0    0    0
 15   0    0    0    0    3    0    0    0    0    0    0    0    0
 16   0    0    0    0    0    0    0   10    0    0    0    0    0
 17   0    0    4    0    0    0    0    0    0    0    0    0    0
 18   0    6    0    0    0    0    0    0    0    0    0    0    0


NEW LABEL: MAX,XLMAX,XPMAX,LABL=0,LABR=0,LABNEW=0:
     108   1    1    -121=n -123=e  121=N
    - Labl=0 Labr=0 Nell Nelr  Llm  Lll  Llr  Il  Jl  Ir  Jr Ia Ja
  121 -121=n     )=.  154    )    0  126    0 118  16   0   0  1  0
  123     )=. -123=e    0  132    0    0  108   0   0 120  14  0  1

NEW LABEL: MAX,XLMAX,XPMAX,LABL=0,LABR=0,LABNEW=0:
      17   6   5   -140=F -150=e  140=F
    - Labl=0 Labr=0 Nell Nelr  Llm  Lll  Llr  Il  Jl  Ir  Jr Ia Ja
  140 -140=F     )=.   37    )    0   20    0  22 113   0   0  6  0
  150     )=. -150=e    0   22    0    0   17   0   0  21 115  0  6

NEW LABEL: MAX,XLMAX,XPMAX,LABL=0,LABR=0,LABNEW=0:
      14  11  11   -173=1 -173=1  173=I
    - Labl=0 Labr=0 Nell Nelr  Llm  Lll  Llr  Il  Jl  Ir  Jr Ia Ja
  173     )=. -173=1    0   16    0    0   15   0   0  26 115  0 11
  173 -173=1     )=.   16    0    0   14    0  26 115   1   0 11  0


 Edited extract of history file:

Labl = Left label.
Labr = Right label.
Nell = Number of elements, left.
Nelr = Number of elements, right.
Il,Jl = i,j coordinates of facet, left.
Ir,Jr = i,j coordinates of facet, right.
Ia,Ja = i,j coordinates of coincidence matrix.


File: Inn076.Fig
```

Fig.: 3.3.3-1.



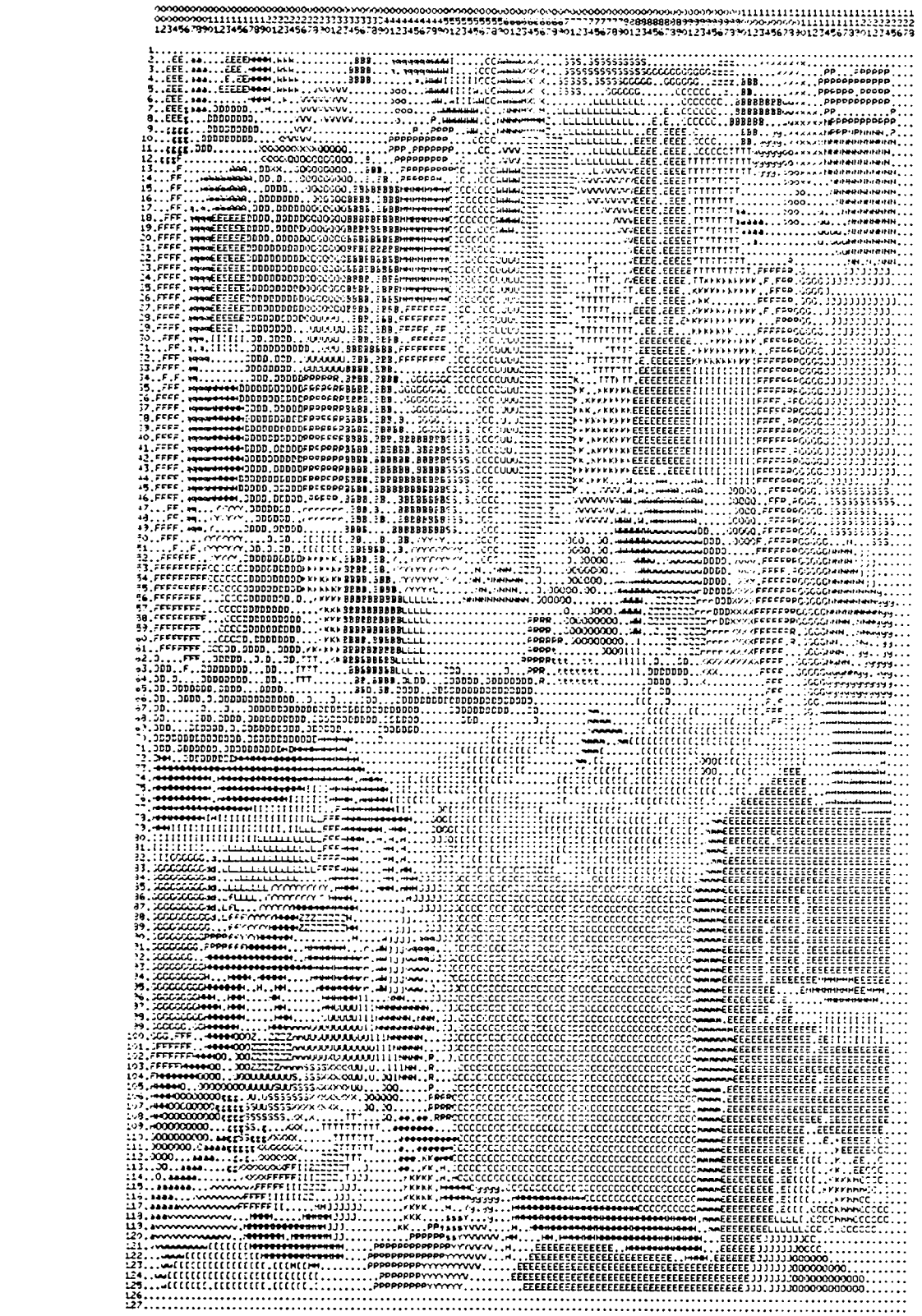
Lm1(i,j). left matched facets after correction for missclassified facets. Area_limit = 5.

LABEL PRINTING

A= 0	B= 1	C= 2	D= 3	E= 4	F= 5	G= 6	H= 7
I= 8	J= 9	K= 10	L= 11	M= 12	N= 13	O= 14	P= 15
Q= 16	R= 17	S= 18	T= 19	U= 20	V= 21	W= 22	X= 23
Y= 24	Z= 25	[= 26	A= 27	B= 28	C= 29	D= 30	.ETC.
a= 0	b= -1	c= -2	d= -3	e= -4	f= -5	g= -6	h= -7
i= -8	j= -9	k=-10	l=-11	m=-12	n=-13	o=-14	p=-15
q=-16	r=-17	s=-18	t=-19	u=-20	v=-21	w=-22	x=-23
y=-24	z=-25	<=-26	a=-27	b=-28	c=-29	d=-30	.etc.

File: Inr077.fig

Fig.: 3.3.3-2a.



Pm1(i,j), right matched facets after correction for misclassified facets. Area_limit = 5.

LABEL PRINTING

.=	)	==	1	C=	2	D=	3	E=	4
F=	5	G=	6	H=	7	I=	8	J=	9
K=	10	L=	11	M=	12	N=	13	O=	14
P=	15	Q=	16	R=	17	S=	18	T=	19
U=	20	V=	21	W=	22	X=	23	Y=	24
Z=	25	[=	26	\=	27	]=	28	^=	29
_=	30	.ETC.							
.=	)	Q=	-1	R=	-2	S=	-3	T=	-4
U=	-5	V=	-6	W=	-7	X=	-8	Y=	-9
Z=	-10	[=	-11	\=	-12	]=	-13	^=	-14
_=	-15	Q=	-16	R=	-17	S=	-18	T=	-19
U=	-20	V=	-21	W=	-22	X=	-23	Y=	-24
Z=	-25	[=	-26	\=	-27	]=	-28	^=	-29
_=	-30	.etc.							

File: Imro78.Fig

Fig.: 3.3.3-2b.

3.3.4 Cut

The matched facets that were markedly different in size could be "cut to size" since usually a facet in one image should not differ very much in size from the corresponding facet in the other image. Of course, exceptions exist which may cause new problems later. A block diagram of the "cut" process is shown in Figure 3.3.4-1.

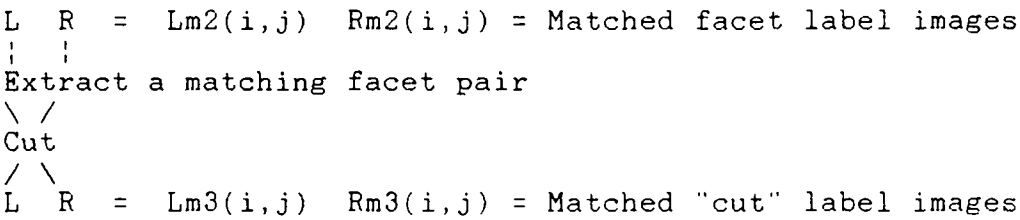


Figure 3.3.4-1: The "cutting" process.

The "cutting" process was based on defining a "maximal allowable" size difference rectangle (Di*Dj), where Di is the allowed size variation along the i-axis and Dj is the one along the j-axis. In the experiments Di = 2 and Dj = 7 pixels. The basic program logic for any matched facet F is as follows, where Fl is a facet label in L and Fr is the corresponding facet label in R (Fl=Fr). (The method assumes that the matching pair F of facets in L and R have been extracted and stored in a separate pair of images):

If, for any pixel at (i,j) in L with label Fl
the label Fr at ((i +/- Di),(j +/- Dj) in R is not equal to Fl
Then zero the pixel (i,j) in R
Else leave pixel unchanged. (Similar logic for R versus L)

Thus, for any pair of matched facets F in the left (L) and right (R) images, the pixels that were outside the "box" defined by Di and Dj are zeroed. These pixel regions are labelled with 0 and are called the "unknown" regions. Of course, the unmatched facets cannot be "cut" since they have no "partners".

Some results showing the effects of "cutting" are shown in Figures 3.3.4-2a and -2b. The most noticeable change occurred to the facet labelled "E" in Figure 3.3.1-4b, lower right corner, which has been "cut to size" in Figure 3.3.4-2b leaving a large "undefined" area. Experimental results tended to show that "cutting" was justified but the facets after "cutting" were still rather different in size since Di and Dj were quite "liberal".

Figure titles:

Figures 3.3.4-2a and -2b: Results showing the effects of "cutting". (Files: Inr081.fig, Inr082.fig)

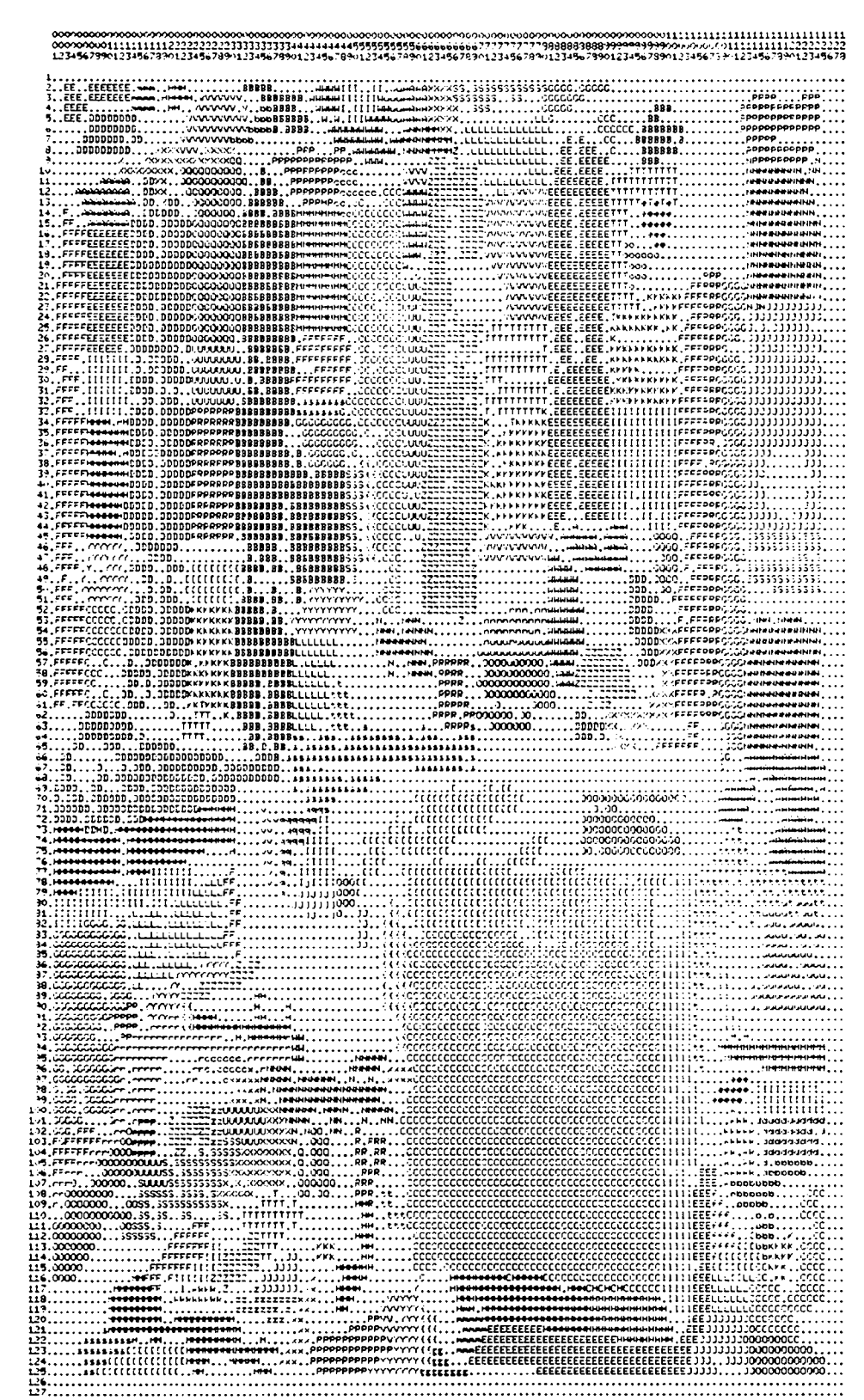
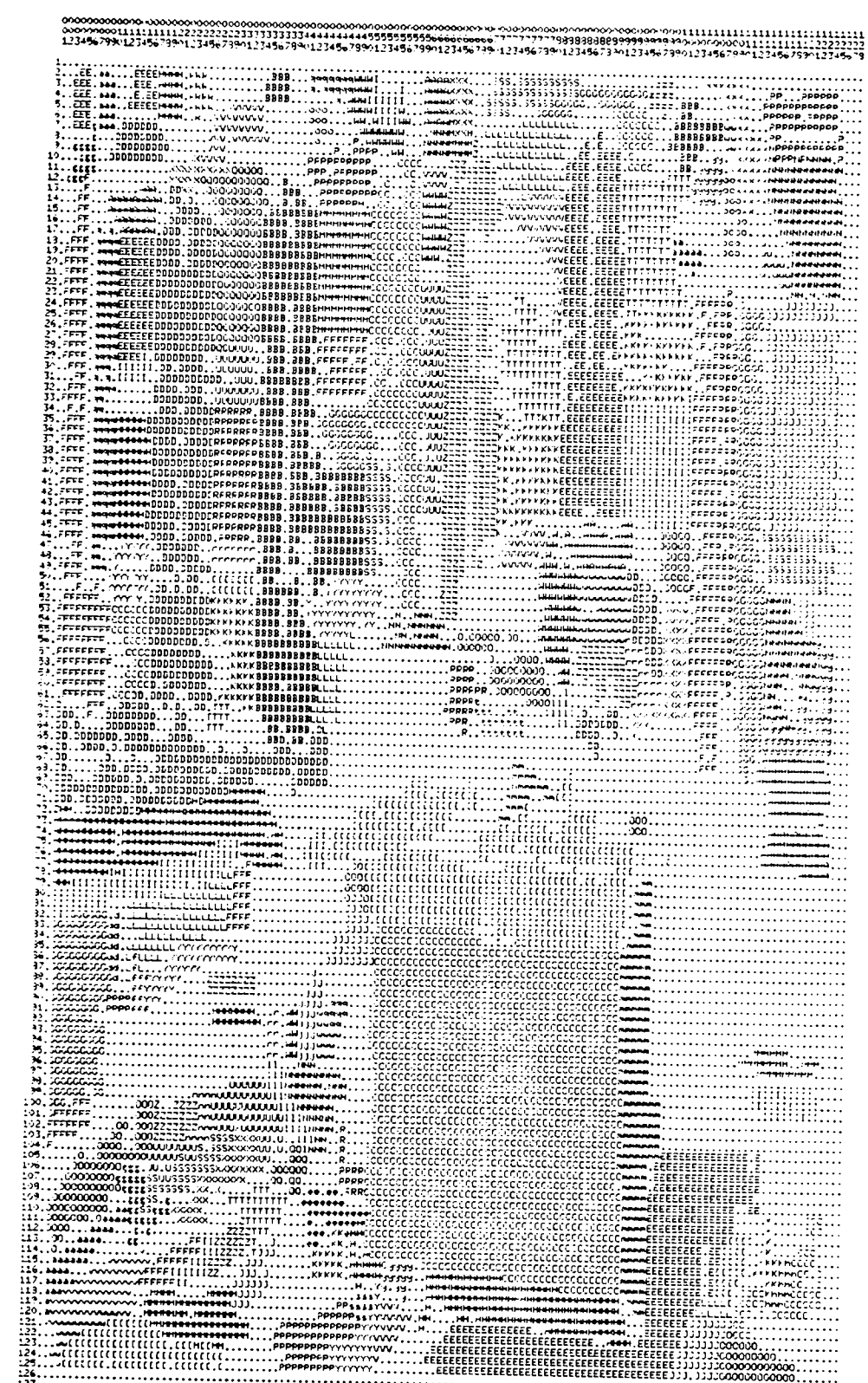


Fig.: 3.3.4-2a.



2m2(i,j). right, matched facets after "cutting" to reasonable disparity. Di = 7. Dj = 2.

LABEL PRINTING

:= 0	=+ 1	C= 2	D= 3	E= 4	F= 5	G= 6	H= 7
I= 8	J= 9	K= 10	L= 11	M= 12	N= 13	O= 14	P= 15
Q= 16	R= 17	S= 18	T= 19	U= 20	V= 21	W= 22	X= 23
Y= 24	Z= 25	[= 26	\= 27	]= 28	^= 29	_= 30	.ETC.
`= 0	a= -1	b= -2	c= -3	d= -4	e= -5	f= -6	g= -7
h= -8	i= -9	k= -10	l= -11	m= -12	n= -13	o= -14	p= -15
q= -16	r= -17	s= -18	t= -19	u= -20	v= -21	w= -22	x= -23
y= -24	z= -25	{= -26	|= -27	}= -28	~= -29	DEL= -30	.etc.

File: lnr082.fig

Fig.: 3.3.4-2b.

3.3.5 Transplant

The "transplanting" process was one attempt to try to fill the unknown regions in the L and R images. It was simply argued that, if a pixel (i,j) in R has no label (label=0) while there is a non-zero unmatched (negative) label in the same location (i,j) in L, then, why not simply "transplant" this label from L into R at location (i,j), and vice versa for R versus L. Thus, where .Lt. means "less than" and .Eq. means "equal to":

```
If((L(i,j).Lt.0).And.(R(i,j).Eq.0)) Then R(i,j)=L(i,j)
If((R(i,j).Lt.0).And.(L(i,j).Eq.0)) Then L(i,j)=R(i,j)
```

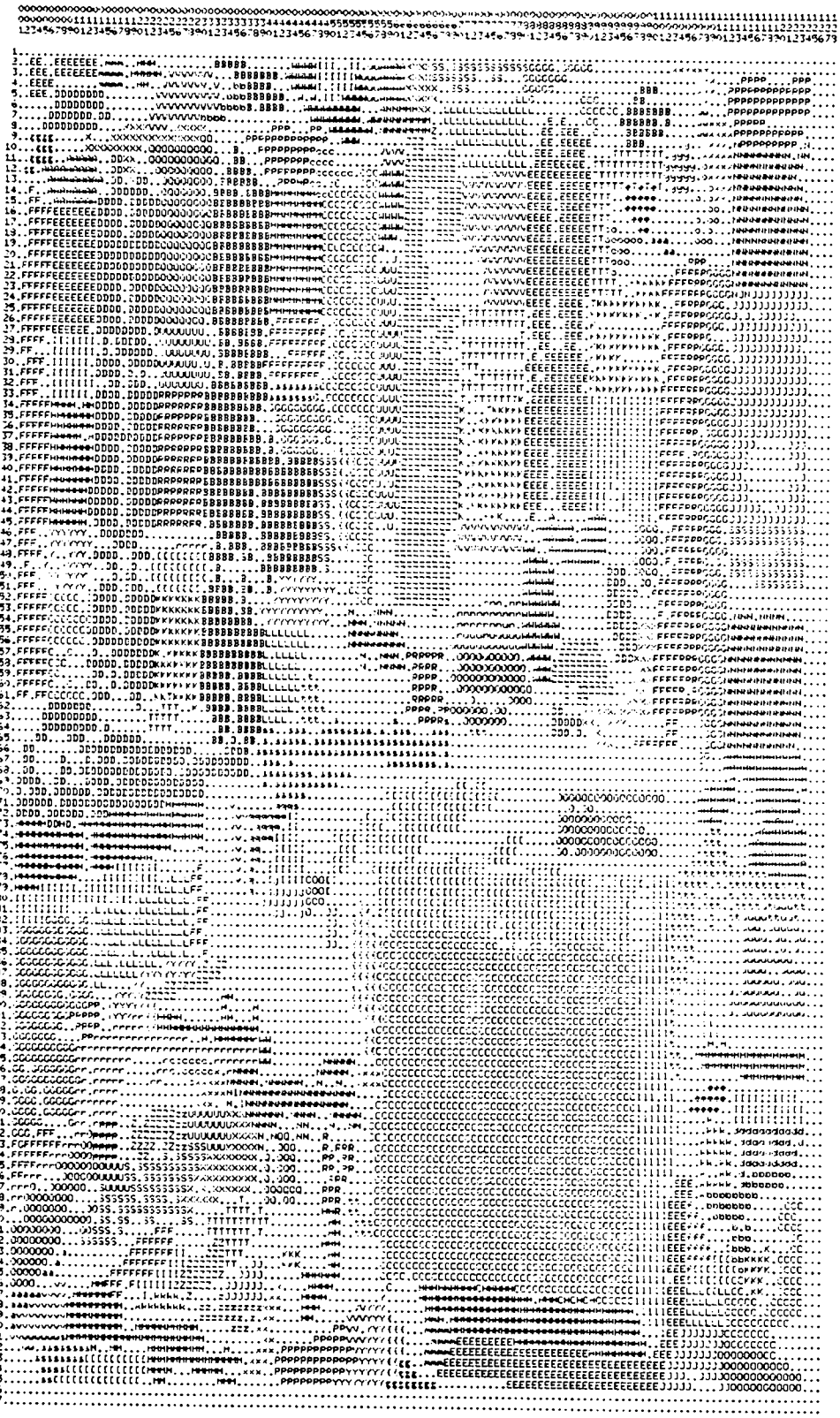
Of course, this is approximately valid only for "distant" facets in the scene but will generate drastic errors for "close" facets which are not near the centre of fixation. However, close facets should have large intersection areas and should not create this problem in the first place. A block diagram is shown in Figure 3.3.5-1 and some experimental results of "transplanting" are in Figures 3.3.5-2a and -2b. The most noticeable change again occurred in the lower right corner of the images where the original gray levels were rather different.

```
L  R  =  Lm3(i,j)  Rm3(i,j) = Matched "cut" label images
\ /
Transplant
/ \
L  R  =  Lm4(i,j)  Rm4(i,j) = Matched facet label images
```

Figure 3.3.5-1: The label "transplanting" process.

Figure titles:

Figures 3.3.5-2a and -2b: Some results of "transplanting" the labels. (Files: Inr083.fig, Inr084.fig)



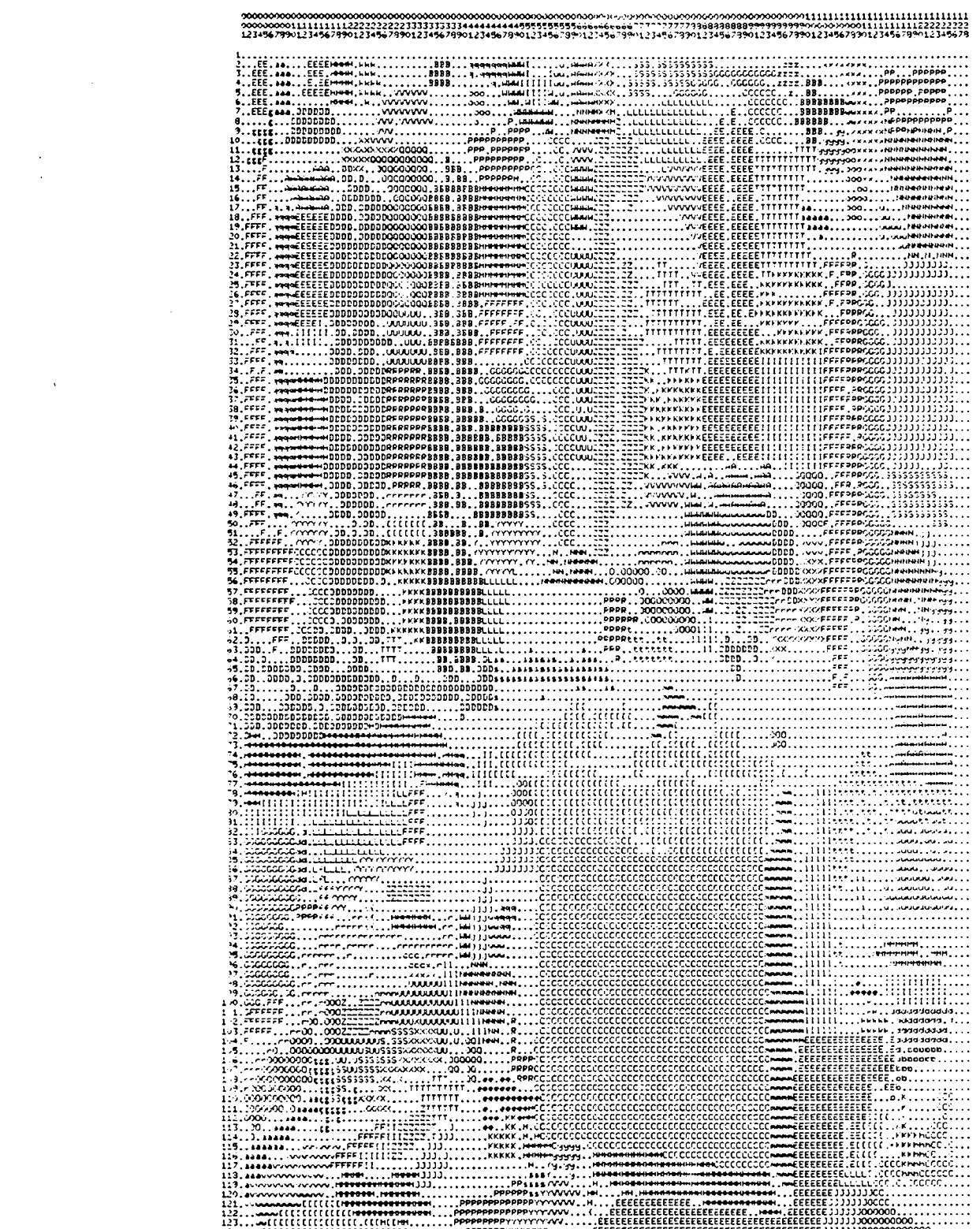
Lm(i,j), left, matched lables after "cut and transplant". Transplanted and unmatched lables are negative. Di=7 Dj=2.

LABEL PRINTING

.= 0	A= 1	C= 2	D= 3	E= 4	F= 5	G= 6	H= 7
I= 8	J= 9	K= 10	L= 11	M= 12	N= 13	O= 14	P= 15
Q= 16	R= 17	S= 18	T= 19	U= 20	V= 21	W= 22	X= 23
Y= 24	Z= 25	[= 26	A= 27	B= 28	C= 29	D= 30	.ETC.
.= 0	P= -1	c= -2	d= -3	e= -4	f= -5	g= -6	h= -7
i= -8	j= -9	k=-10	l=-11	m=-12	n=-13	o=-14	p=-15
q=-16	r=-17	s=-18	t=-19	u=-20	v=-21	w=-22	x=-23
y=-24	z=-25	{=-26	a=-27	b=-28	c=-29	d=-30	.etc.

File: Inr083.fig

Fig.: 3.3.5-2a.



Pm(i,j). right, matched lables after "cut and transplant". Transplanted and unmatched lables are negative. Di=7 Dj=2.

LABEL PRINTING

A=	0	a=	1	C=	2	D=	3	E=	4	F=	5
I=	8	J=	9	K=	10	L=	11	M=	12	N=	13
Q=	16	R=	17	S=	18	T=	19	U=	20	V=	21
Y=	24	Z=	25	[=	26	\=	27	]=	28	^=	29
_=	0	@=	-1	c=	-2	z=	-3	a=	-4	f=	-5
i=	-8	j=	-9	k=	-10	l=	-11	m=	-12	n=	-13
q=	-16	r=	-17	s=	-18	t=	-19	u=	-20	v=	-21
y=	-24	z=	-25	{=	-26	\=	-27	}=	-28	~=	-29

G=	6	H=	7								
P=	15										
W=	22	X=	23								
.ETC.											
b=	-2	c=	-3	d=	-4	e=	-5	f=	-6	g=	-7
o=	-14	p=	-15								
w=	-23	x=	-24								
d=	-30	.etc.									

Fig.: 3.3.5-2b.

File: Im~84.fig

3.3.6 Clip

The purpose of the so-called "clipping" process is to try to give an "even start" in the analytic relaxation process to each pair of matched facets. Two alternatives were considered:

A) Centre of gravity based clipping.

a) For any pair of matched facets in Lm1(i,j) and Rm1(i,j) compute their individual centers of gravity.

b) Shift one of the facets (of a matched pair) so that the centers of gravity in Lm1(i,j) and Rm1(i,j) overlap.

c) "And" the shifted facets. This reduces the facet pair to the same size in the new matched facet images Lm2(i,j) and Rm2(i,j).

B) Correlation based clipping (not completed).

a) Correlate the pair of matched facets in Lm1(i,j) and Rm1(i,j) to find the shift di and dj for the best match. (The di and dj lists are directly available form the correlation based matching.)

b) Shift one of the facets (of the matched pair) as given by di and dj for best overlap.

c) "And" the shifted facets (as in (A)). This reduces the facet pair to the same size in the new matched facet images Lm2(i-,j) and Rm2(i,j).

A block diagram of the process is in Figure 3.3.6-1.

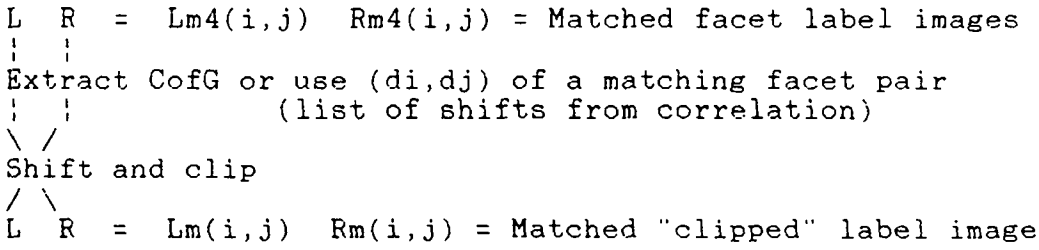


Figure 3.3.6-1: The label "clipping" process.

The "allowable displacements" between the two facets were limited by a "box" Di*Dj, as in previous cases. Some results are shown in Figures 3.3.6-2a and -2b. The matching facet pairs are reduced to the same sizes in the Lm(i,j) and Rm(i,j) images. This is thought to be "the best that can be done" for "balancing" the subsequent relaxation process, see section 3.4. Note that the clipping process (as it is at present) also created holes within facets. For example, if a region Fl in the left image has a hole but the corresponding region Fr in the right image does not, then the "and-ing" process will make a hole in Fr (a region of zero labels). If the shift between Fl and Fr is less than the "diameter" of the hole then some of the 0-valued regions will coincide, see further on under "gaps".

The matching pairs of facets are of the same size in the Lm(i,j) and Rm(i,j) images. Geometrically such images are impossible. However, when viewed stereoscopically, a rather strange scene is seen (but it may take some initial effort to "integrate" these images visually). Consequently, several versions of stereo pairs were printed.

Figures 3.3.6-3a and -3b show the matched facets (Fl,Fr) printed with averaged grays (Gray = (Gray(Fl)+Gray(Fr))/2). The unmatched facets were printed in average gray over both images. In Figures 3.3.6-4a and -4b the facets were "painted" such that adjacent facets have different gray levels. Five gray levels ("colours") were needed.

Some rather interesting stereo effects were visible in the original prints but the copying procedures may have reduced all these efforts to nought. Thus, in words, the shifted matched facets create a scene that generally appears correct but there are gaps between the facets. This creates an "exploded" view of the 3D scene. In addition, there are "freely hanging" facets which are "in front of" the room scene. Only the unmatched facets create visual disturbances. This creates interesting implications for the subsequent reconstruction:

1. The initially matched (and clipped) facets are likely to be immediately reconstructible since they "appear correct" in the 3D scene. However, since these facets are (mostly) "exploded and hanging in free space" the reconstruction will have gaps between the facets and, of course, the interiors of the facets are empty.

2. The facets that "hang in front of the 3D scene" do not integrate with the rest of the scene. They act to "obscure" parts of the background.

3. The unmatched facets create visual disturbances since they "want to integrate" with the scene.

The full implications of the above observations are not entirely comprehended at the moment. It is unlikely that any visual illusions are involved, but the author has seen "too much" of this particular stereo pair to be a reliable observer. Experiments on colleagues have indicated that they are mostly unfamiliar with stereo viewing. After some "training" they see some stereo effect to a certain degree in some images but "independent confirmation" has been hard to obtain with the existing image quality and equipment.

The obvious implications are:

a) The matched facets should be "expanded" by some form of "acclomerative pixel acquisition method" (acclomerative merging, snakes, etc.). In the present case this acclomeration is done via analytic approximation and relaxation, see section 3.4.

b) The facets that "hang in front of the scene" (case 2) should also "acquire a parallax" after the processing indicated in (a).

c) The unmatched facets should either be eliminated or not allowed to participate in the process (a). The unmatched facets are most likely to be errors but, of course, anything seen with one "eye" only will be unmatched.

Figure titles:

Figures 3.3.6-2a and -2b: The matched facets after clipping. (Files: Inr087.fig and Inr088.fig)

Figures 3.3.6-3a and -3b: The matched facets (Fl,Fr) printed with averaged grays (Gray = (Gray(Fl)+Gray(Fr))/2). The unmatched facets are printed in average gray over both images.

Figures 3.3.6-4a and -4b: The facets "painted" in five "colours" such that adjacent facets have different gray levels.

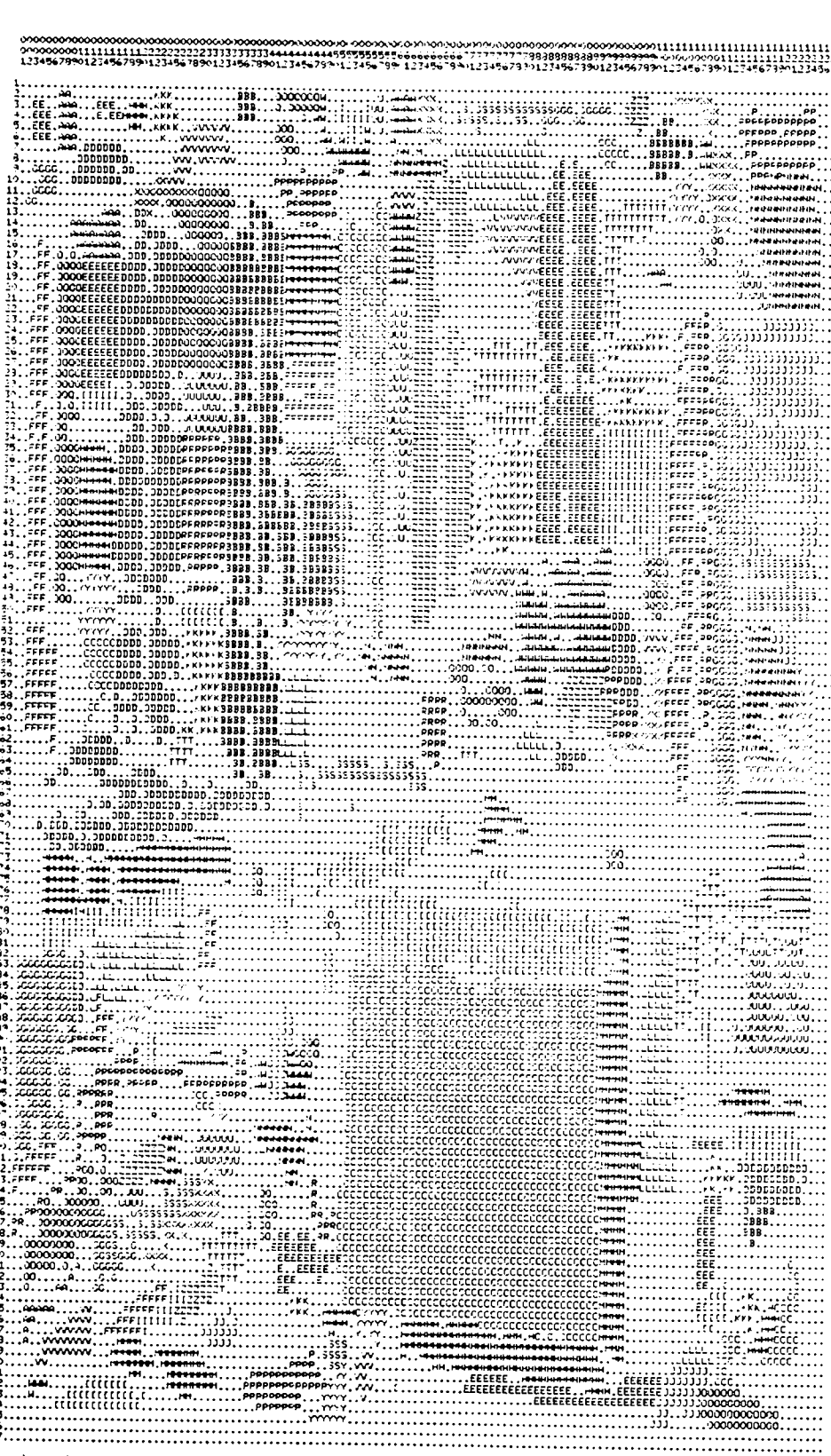
La(i,j), left, matched, cut, transplanted, and clipped facet labels, Di=7, Dj=2.

LABEL PRINTING

.= 0	a= 1	C= 2	D= 3	E= 4	F= 5	G= 6	H= 7		
I= 8	J= 9	K= 10	L= 11	M= 12	N= 13	O= 14	P= 15		
Q= 16	R= 17	S= 18	T= 19	U= 20	V= 21	W= 22	X= 23		
Y= 24	Z= 25	(= 26	A= 27	B= 28	C= 29	D= 30	,ETC.		
.= 0	0= -1	c= -2	z= -3	e= -4	f= -5	g= -6	h= -7		
i= -8	j= -9	k=-10	l=-11	m=-12	n=-13	o=-14	p=-15		
q=-16	r=-17	s=-18	t=-19	u=-20	v=-21	w=-22	x=-23		
y=-24	z=-25	(=-26	A=-27	B=-28	C=-29	D=-30	.etc.		

Fig.: 3.3.6-2a.

File: Inv087.fig



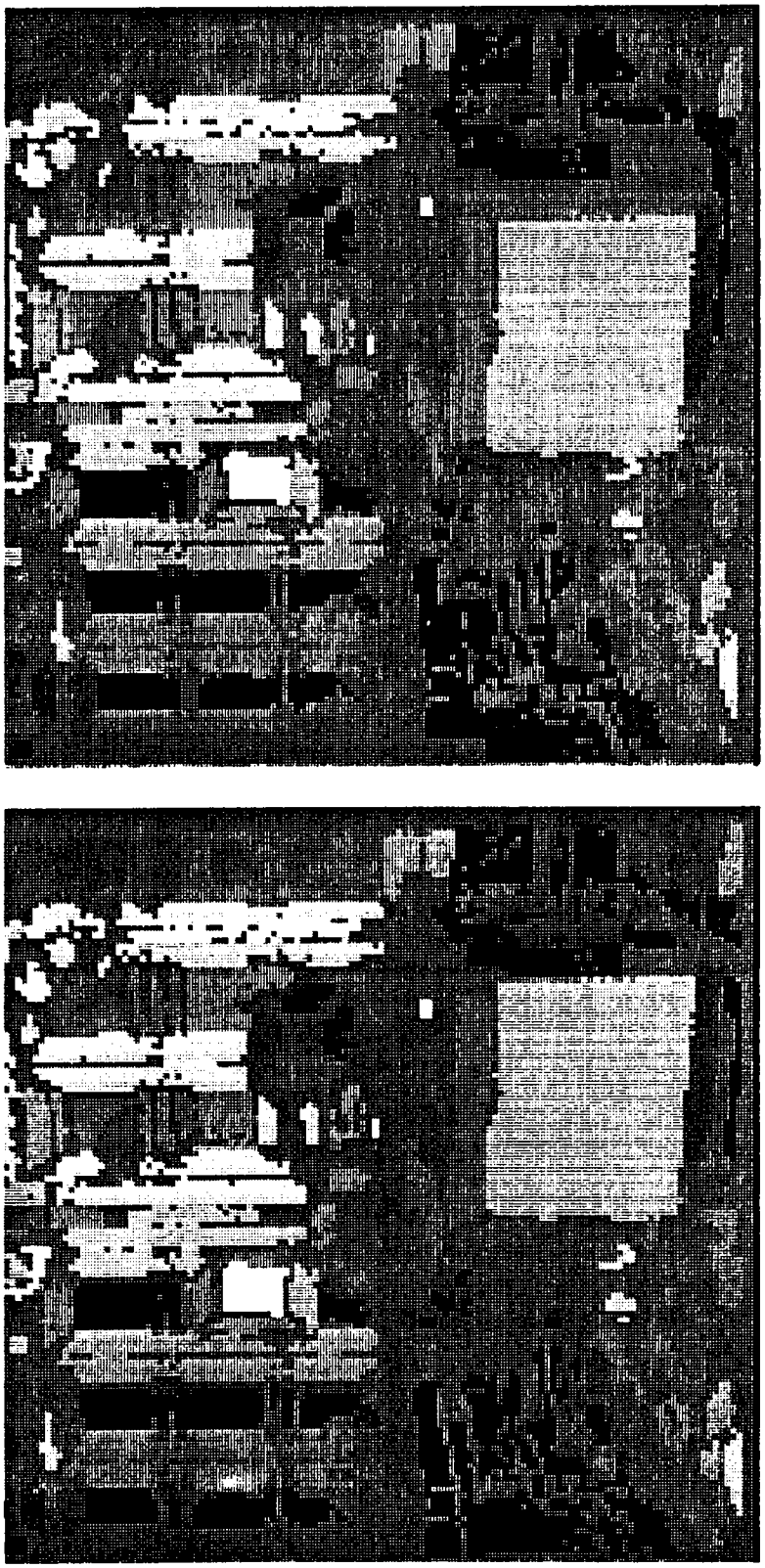
Rm(i,j), right, matched, cut, transplanted, and flipped facet labels, Di=7, Dj=2.

LABEL PRINTING

.= 0	== 1	>= 2	D= 3	E= 4	F= 5	G= 6	H= 7	
I= 8	J= 9	K= 10	L= 11	M= 12	N= 13	O= 14	P= 15	
Q= 16	R= 17	S= 18	T= 19	U= 20	V= 21	W= 22	X= 23	
Y= 24	Z= 25	[= 26	\= 27	]= 28	^= 29	_= 30	.ETC.	
,= 1	@= -1	a= -2	b= -3	c= -4	d= -5	e= -6	f= -7	
g= -8	h= -9	i=-10	j=-11	k=-12	l=-13	m=-14	n=-15	
o=-16	p=-17	q=-18	r=-19	s=-20	t=-21	u=-22	v=-23	
w=-24	x=-25	y=-26	z=-27	{=-28		=-29	}=-30	.etc.	

File: Inr088.fig

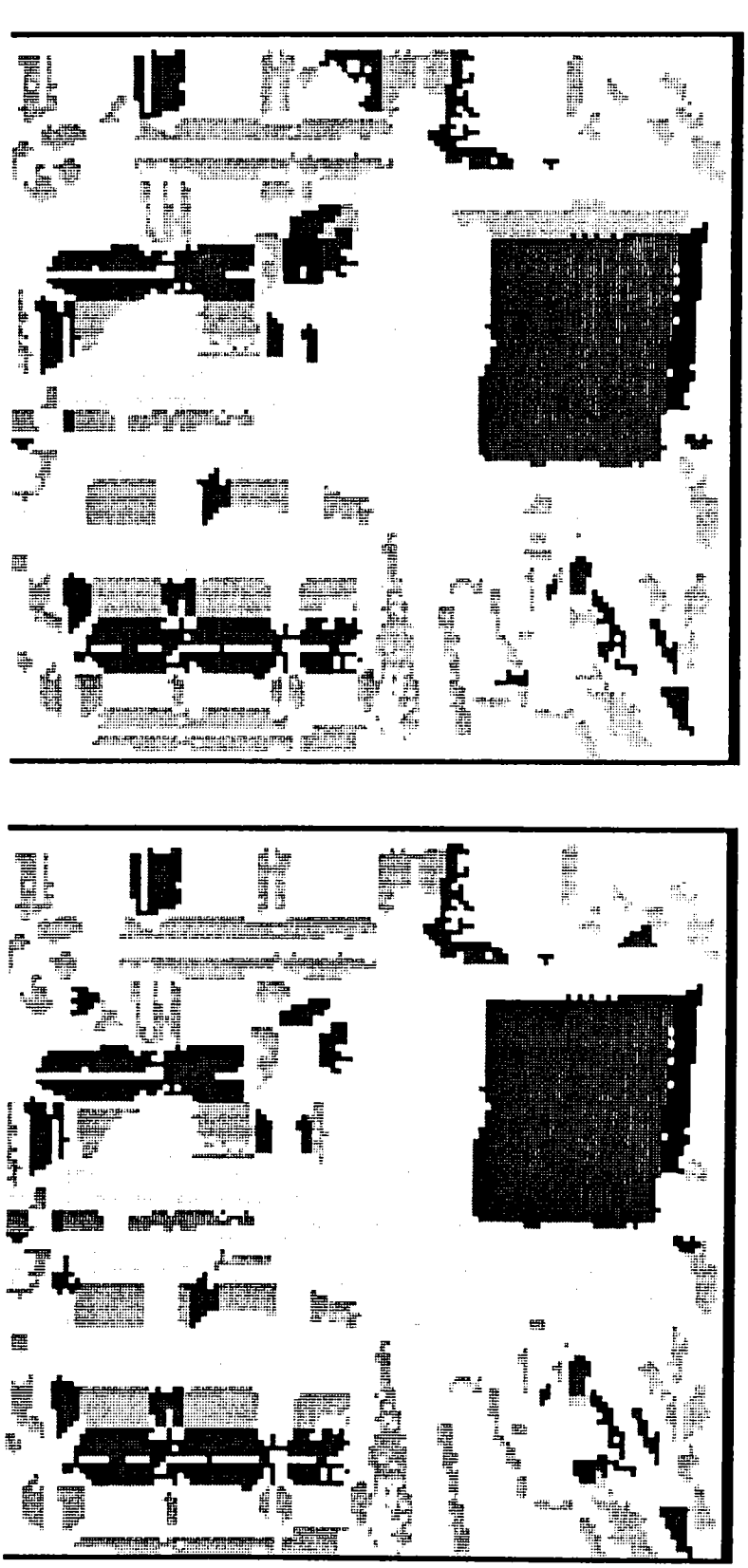
Fig.: 3.3.6-2b.



if8flgmc6.lgr

if8flgmc6.rgr

Fig.: 3.3.6-3a.

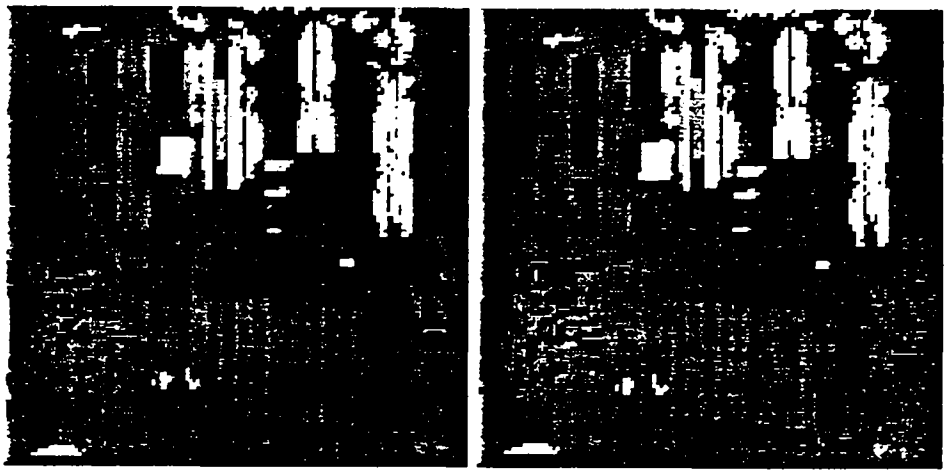


if8flgmc5.rgr

if8flgmc5.lgr

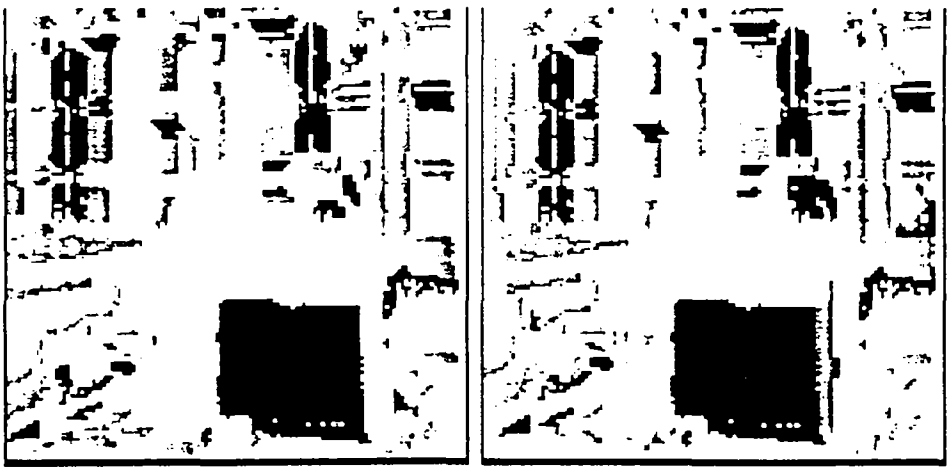
Fig.: 3.3.6-4a.

ly27.90



if8flgmc6.lgr if8flgmc6.rgr

Fig.: 3.3.6-3b.



if8flgmc5.lgr if8flgmc5.rgr

Fig.: 3.3.6-4b.

3.3.7 The gaps and the un-matchabels

There are three kinds of facets that are "left over" from the previous processes, i.e.:

1. The "facets" (regions) where the labels are zero. There is "nothing" at these pixels.

2. The "unmatched" facets where there is a facet in one image of the stereo pair but not in the other. Even after the "transplant" operation there remain some unmatched facets.

3. The "unmatched" facets that could not be matched even though a "partner" may exist in the other image. Reason for the inability to match are that the two facets are "far apart" or one of the facets has become "fragmented".

In case of the unlabelled regions (case 1), if there is "nothing" in both images at the same pixel (i,j) then the "common characteristic" (similarity) is "nothing" (which is information about similarity) and this can be used to create a region. The process is as follows:

i) Create regions of pixels indicating "nothing". i.e.,

```
If((L(i,j).Eq.0).And.(R(i,j).Eq.0)) Then B(i,j) = 1
Else B(i,j) = 0
Endif
```

where B(i,j) is a binary image, ".Eq." means "is equal to", and ".And." means logical "And". Since the previous processes may have reduced the sizes of the existing regions, this operation can create a binary image B(i,j) consisting of "blobs connected by strings". This brings the process directly into the domain of binary image processing.

ii) Erode B(i,j) to "disconnect" the "blobs".

iii) Label the "blobs" but continue the label numbers from highest label in Lm(i,j) and/or Rm(i,j).

iv) Put these regions "back" into the Lm(i,j) and Rm(i,j) images.

At least now the "larger" unlabelled regions have become matched and labelled. However, they may not perform too well in the subsequent analytic approximation and relaxation processes if the regions cannot be properly approximated by the simple functions used for the matched facets.

The unmatched facets (negatively labelled regions) that still remain after all these operations depend on the sequence in which the processes were carried out. In general several interpretations are possible:

a) These regions are only visible to one "eye".

b) The regions are caused by reflections (for example, distant bright objects) and are thus at different locations in the 3D scene than the region in which they appear in the image, or they are from transparent regions (a "hole" in the foreground scene). Frequently, the reflected region is greatly modified (broken up) by the reflecting surface.

3.3.8 Comments

The modifications that the matched facet labelled images Lm(i,j) and Rm(i,j) have been "subjected to" are summarized in Figure 3.3.8-1 and listed below. The purpose has been to illustrate some of the processing steps that could be used to "fix" the majority of errors. Hopefully, fewer errors have been introduced than have been fixed.

1. A certain number of misclassifications are removed by finding the "intersection" between unmatched facets.

2. Matched facets that are "too different in size" are "cut" to "reasonable size".

3. Since "cutting" leaves blank regions these are "filled in", if possible, by transplanting labels from the other image.

4. All the facets are then "clipped" to the same size in order to give different facets an "even chance" during the subsequent operations (analytic approximation and relaxation, see section 3.4).

5. The remaining "unknown" regions are then also given labels in order to allow them to "defend their territories" during subsequent processes.

6. The facet regions are next approximated analytically and the regions are corrected by analytic relaxation, see section 3.4.

The basic question is, how many of these processes are logically justifiable, but most of all, how do they perform in practice?

The correction of misclassifications may be looked upon as and extension or a correction to the label matching process and justified in the same terms as the initial matching. The cutting process can only be justified in probabilistic terms, i.e., it is more likely than not that the corresponding facets in the L and R images are approximately of the same size. Some exceptions were mentioned earlier. The transplantation process may also be justified in probabilistic terms since, if an unmatched region is

"large enough", it is very likely to contain facets that are similar to those visible in the other image, provided there are any. The subsequent clipping process is not logically justifiable except in terms of processes that follow. In the present case the subsequent process is analytic relaxation that is expected to "grow" the facets back to their proper sizes in L and R. Hence, the clipping process is only justified as "a method of giving an 'equal chance' to all the facets" in the subsequent analytic relaxation. The filling of the unknown regions in both images by "artificial facets" may also be viewed as a method of giving an "equal chance" to such regions in the analytic relaxation process. An alternative would be to reclassify the image for only these regions (see Chapter 2).

The most dubious processes are cutting and transplanting. Besides their ad hoc nature, the dilemma with the "cutting" and "transplanting" processes is:

i) Cutting is too "rough", it either leaves or removes a large area of a facet and leaves nothing in its place, depending on the Di and Dj. The logical "upper limit" for Di is the whole i-dimension of the image (in the vicinity of the epipolar line).

ii) Transplanting puts "something" into the vacated area but this "something" is only a part or a copy of the facets that were in the other image. Consequently, unless the facets are from "far away places", the "transplants" are in the wrong positions.

The "cutting" and "transplanting" processes were attempting to deal with unequal segmentation in the left (L) and right (R) images. An alternate approach may be envisaged, which is best explained in connection with a simple example. Consider a region in the image which in L has been split into two facets Fl1 and Fl2, while the region is only one facet Fr in R. Of course, the intersection of Fl1+Fl2 with Fr is not precise (+ indicates "union"). Both Fl1 and Fl2 "lay claims" on Fr. The amount of these "claims" may be the intersection areas Fl1*Fr and Fl2*Fr (* indicates "intersection"). At present the facet with majority wins. More subtle processes are possible but their implications have not been thought out in detail.

In the final analysis the important aspect is how these operations behave in practice on images for which such operations are appropriate. Experimental results tended to show that "cutting" and "transplanting" were largely justified but there were some facets that remained the same in L and R after "matched" analytic relaxation. The facets after "cutting" were still rather different in size since Di and Dj were rather "liberal". Consequently, the analytic parameters for the two "matching" facets were not really representing the situation in a satisfactory manner and the meaning of the matched analytic relaxation become dubious. Both of these problems were "alleviated" by including

the clipping process. The experiments so far are insufficient to either confirm or reject these as viable methods.

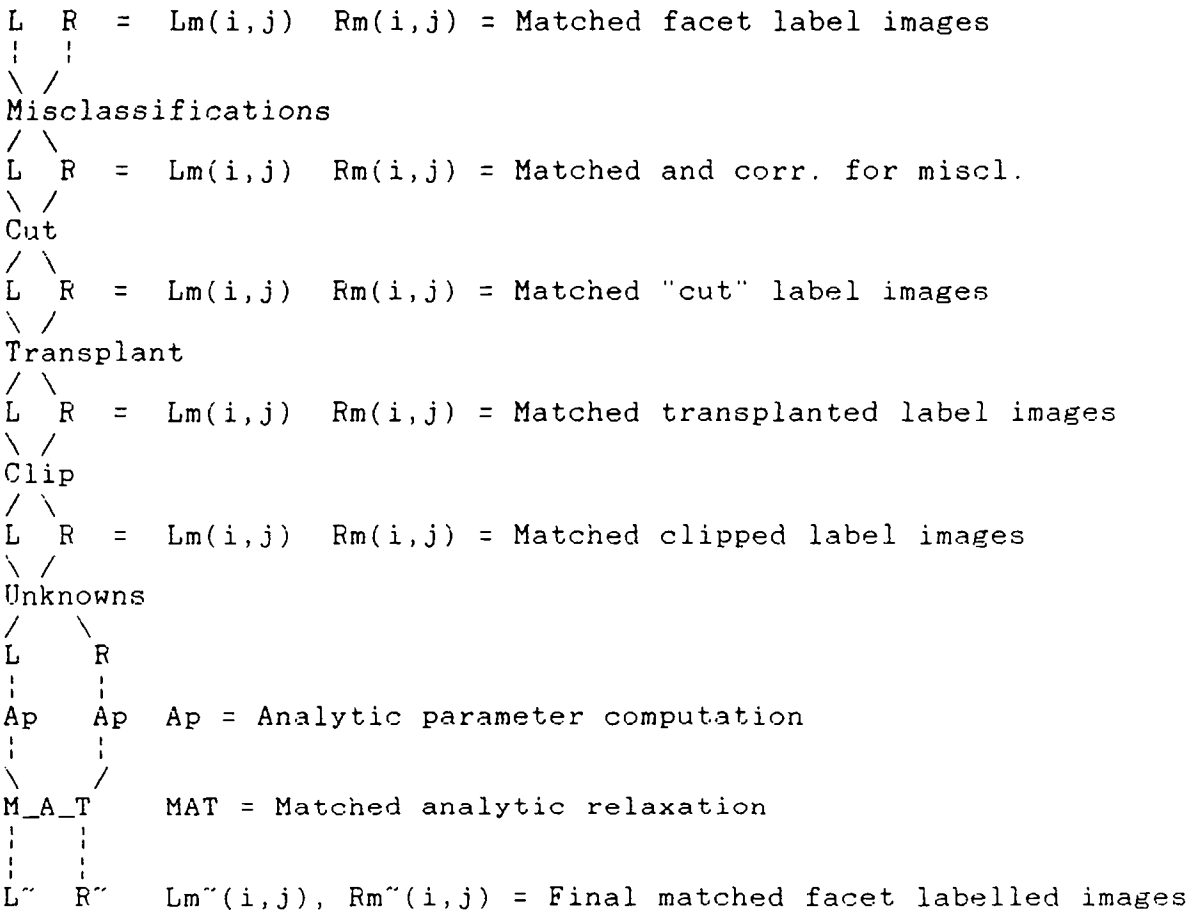


Figure 3.3.8-1: A sketch of the "fixing" processes.

3.4 Analytic methods

The analytic approximation of the gray levels of each facet is intended to serve two purposes which have been expressed here in ordinary words and concepts:

1. The approximation is an "umbrella" or "cover" for the facet which has been "defined" by all the presently known pixels that belong to the facet. The presently chosen second order polynomial approximation is a very crude model of light distribution on a facet.

2. This "umbrella" serves to "defend the interests" of the pixels belonging to the facet and also serves as the tool for acquiring new "like-minded" pixels during the analytic relaxation.

Hence, the approximation and subsequent analytic relaxation may be likened to a military government of a country in its role as the instrument of the country's coherence and during times of conflict the defender of the country's territorial integrity and the tool for its expansionist interests. The reader may translate this into mathematical symbols!

The process was intentionally expressed in "clear" language in order to explain some of the problems that are complicating the analytic approximation and relaxation strategy.

a) In order to have a balanced relaxation (a "fair fight" between the facets) the initial approximations for all the facets should be as "fair" as possible. In the absence of any better method, the supposedly paired facets were made of equal size, since it was noticed in prior experiments that unequal starting conditions were undesirable.

b) All regions in the images should have analytic expressions in order to properly participate in the analytic relaxation. If a facet was bordering an unlabelled region (an "undefended" region without an analytic approximation for it) then the labelled facets tended to expand into it during relaxation since the only constraint was the limit on the overall error term between the analytically computed and actual data.

c) The unmatched but labelled facets created a dilemma during relaxation.

If the unmatched facet was a "true" facet representing a region that was only visible in one image then, of course, this facet should "defend its territory" during relaxation. However, if the unmatched facet was false (a part of a matched region split off in one image but not in the other image), then the relaxation should cause the false facet to be absorbed by the region from which it was split off originally. In general, this does not happen since the analytic approximation of a partial

facet is usually rather different from that of a whole facet. Consequently, the "false" facet was able to "defend itself" and the desired correction did not take place. An alternative is to simply ignore such facets, i.e., the unmatched facets are treated as "unlabelled", which contradicts (b). Hence, considerable effort was spent in order to try to "match everything" before the analytic process was started, see previous section.

d) There is no such thing as "two falsely matched facets". Expressed in other terms, if a match has been found between two facets then, at the present stage of processing, there is no other information available to detect whether or not this match is correct. After the reconstruction stage it may be possible to detect some contradictions in the 3D scene, for example, "two visible facets obscure each other", but these aspects of the problem have not been studied.

e) If a region was "incorrectly" segmented (from human point of view) but the "incorrect" segments were matched correctly, then they are correct segments (despite human opinion). According to the similarity principle "recognition" or "understanding" of the facets is neither required nor needed.

3.4.1 The coupling

The greatest conceptual problem is created by the need to have some form of "coupling" or link between each of the left and right matched facets (Flk,Frk) during relaxation.

Let the k'th left facet be indicated by Flk (where "l" is for "left" image), let the k'th right facet be Frk ("r" indicates "right" image), and let F symbolize a facet. For a matched facet pair (Flk,Frk) the index k is the same in the labelled images Lm(i,j) and Rm(i,j), and the analytic parameters are obtained from tables addressed by k (the left image parameter tables are Tl(k,n) and the right image ones are Tr(k,n), for n = 1,2,... for the degree of the analytic approximation). Thus, the data management in the relaxation program is very simple. A block diagram of the process is shown in Figure 3.4.1-1. Some details of the relaxation process were given in Chapter 2.

The basic problem with the "coupling" between the left and right matched facets (Flk,Frk) during relaxation is that the procedure is analytic (and the author is not aware of any other form of mathematics which is both precise and general at the same time)! Even though the approximated values can be computed at any pixel (il,jl) in the left image L and at any pixel (ir,jr) in the right image R, for a "coupling" to exist between L and R the pixels (il,jl) and (ir,jr) have to be matched in the sense of stereo matching. Thus, ideally, the L and R images will have to be matched on the pixel level before the relaxation technique can

be correctly applied! [A similar problem in pattern recognition versus segmentation is well-known: An image has to be segmented correctly before the objects can be recognized but the objects have to be recognized before the image can be segmented correctly! In ordinary terms this is known as the "chicken and egg" puzzle, i.e., which came first, the chicken or the egg?]

Several strategies are possible in order to try to bypass this problem without resorting to "models". Models restrict the solution to a very small domain and create "combinatorial explosions" during matching.

1. Improve the segmentation independently in the two images. Attempts to simply try to "improve" the segmentation using the existing set of features may help but become futile "in the limit" since the images do differ. The alternative is to compute more features for each pixel and to improve both the classification and the matching techniques. In the limit this would resemble direct pixel feature matching and eliminates most of the intervening steps.

2. Try to solve the problem with approximations and iterations. As with any iteration technique, if the starting point is too far from the real solution then either no solution is obtained and the iterations do not converge, or some form of false convergence is obtained. Active or "live" vision is able to "look at a funny facet" and this process immediately produces a better match and corrections to the iterations since the centre of fixation is on the facet. In still-life vision the wrong solution may be just another way of seeing the 3d scene. In the absence of additional "knowledge" how is one to verify that this is indeed the wrong solution? Our own visual illusions tend to confirm this opinion.

The approximate matches from (2) should be very helpful in pixel feature matching (1) since the problem is now confined to the vicinity of each matched facet. As a final step pixel level matching is required anyway, whatever the intermediate methods in the hierarchy may have been.

3. Approximate reconstruction. If the 3D scene (depth map or image) is reconstructed from the presently matched information (facets, edges, pixels) then, of course, the results depend on the matches. In other words, the matching information can only be obtained from the left and right images and a partially reconstructed 3D scene is no better than the matches (in L and R) that were used. It is not clear at the moments whether a partially reconstructed 3D scene would be of any help in resolving the paradox without using additional information (more features, better analytic procedures, or even "extraneous" knowledge, i.e., models).

4. Apply constraints that are appropriate to the accuracy at which the matching is carried out, see next section.

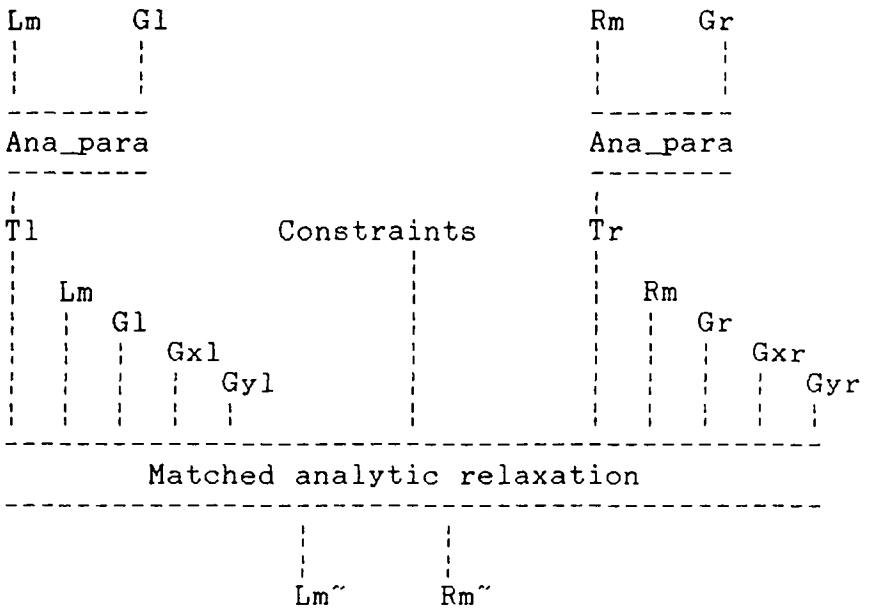


Figure 3.4.1-1: A sketch of the analytic process.

Lm, Rm = Left and right matched facet label images.
Gl, Gr = Left and right gray level images.
Ana_para = Analytic parameters for facets.
Tl, Tr = Tabulations of analytic parameters.
Gxl, Gyl = X and y components of gradient of Gl.
Gxr, Gyr = X and y components of gradient of Gr.
Lm˝, Lr˝ = Analytically corrected matched facet label images.

3.4.2 Constraints on relaxation

There is no lack of approximate constraints that could be placed on the relaxation process in order to "couple" the two images of the stereo pair. Given any two facets A and B in the L and R images, some of the constraints may be the following (which have been listed without any particular order of preference):

1. Edge within a "tolerance box" Di*Dj.
2. Presence of an edge for facet A.
3. Presence of an edge between facets A and B.
4. Adjacency of A to B.
5. Edge with a certain slope tolerance.
6. Edge with a certain gradient tolerance.
7. Edge with a certain curvature tolerance.
8. Local maximum of edge function at edge.
9. Combinations of above including local extrema.

a) The presence of an edge (border) for facet A in both L and R, while ignoring the other facet (B).

This simply requires the facet to be matched (A is both in the left 'L' and right 'R' images). The relaxation is allowed to progress independently in L and R provided the edge of A stays within a tolerance box Di*Dj centered on the epipolar line. The constraint on the "progress" of relaxation is obtained independently from the L and R images. Pixel-level matching between L and R is not required. The method is similar to the "independent relaxation" in Chapter 2, except that only the matched facets are allowed to "relax" provided they stay within Di*Dj.

b) Improvements on the first method (a) consist of requiring increasingly precise matches between the borders of facet A. For example:

i) Gray level and gray level gradient were introduced separately in Chapter 2. Here one may require a certain amount of similarity between these in both images at the same time.

ii) The local maximum of the "edge function" Ef (see Chapter 2) should be reached in both images, provided Ef > some tolerance. Ef(i,j) was defined as a function with "amplitude" rather than "on-off" for exactly this reason since it allows "hill climbing". A "smeared" on-off type edge function can serve in the same way but most likely with poorer results.

iii) The slope of the edge at the border of A in L and R should be approximately the same. For more or less straight edges this constraint does not require much pixel-level correspondence between L and R.

iv) The curvature of the edge at the border of A in L and R should be "similar". This places more stringent pixel-level

match requirements between L and R where the curvatures are high (corners).

It may be worth noticing that these methods of constraining the relaxation include a certain degree of "edge matching" between for a facet A in the left (L) and right (R) images. This information may be saved for subsequent use. Edge matching will not be discussed further due to lack of time.

c) Various more or less "artificial" constraints may be used, such as:

i) There has to be a contact between the matched facets A and B in both images.

ii) The displacement of the boundary where the relaxation takes place in L and R is to stay within a "reasonable tolerance limit" Di*Dj.

iii) Many combinations of methods (a), (b), and (c).

In the present experiments the constraints only consisted of the following:

1. Gray level and gray level gradient constraints applied individually as in independent relaxation.

2. The requirement of a contact between two matched facets (c-i) was optional but abandoned due to many unlabelled pixels.

3. The boundaries of a matched facet in L and R had to stay within a "box" (c-ii) around the epipolar line.

The reasons were twofold. Firstly, these were initial experiments before a "deeper comprehension" of what was involved and, secondly, the programs started to overwhelm the computational resources.

The facet labels from "matched classification", after various corrections, are shown in Figures 3.3.6-2a and -2b, together with several stereo pairs. As was described earlier, there were two kinds of facets, i.e.:

a) Those that were approximately in the "right position" when viewed stereoscopically (but there were "cracks" between them in the 3D scene).

b) The facets that were "hanging in free space" in front of the 3D scene.

The resulting facet labels after relaxation are shown in Figures 3.4.2-1a and -1b. In these figures the "acquired pixels" have been indicated by lower case letters to show the "conquests"

made by the various labelled regions. For example, the pixels conquered by facet E are indicated by e. Pixels that had no label and could not be "conquered" are indicated by dots (.). The iteration was run for ten cycles using the same set of "raw" or initial facet parameters. The number of pixels conquered in each iteration is shown in Figure 3.4.2-1c. As seen, the relaxation process is stable in a "global" sense but, as will be seen soon, the small details in the images do not obey "stereo fidelity" requirements.

If the gray level image is recreated (computed) from the analytic values, with the gray levels for the "missing" (0-valued) labels replaced by the gray values from the original image, then the result looks "reasonable", see Figures 3.4.2-2a and -2b. In this figure "gross errors" were also replaced, i.e., if |G_analytic - G_original| > 15 then the original gray level pixel was used instead of the analytic value. The "surviving labels" are shown in Figures 3.4.2-3a and -3b. The stereo pair is shown in Figure 3.4.2-4a together with a stereo pair of the original gray level image for comparison, see Figure 3.4.2-4b.

The replacement of "missing pixels" (zero-valued labels and where the approximation is "too poor") is a valid operation, being just another form of "correction". However, the facets to which these "missing pixels" should be assigned, or the new facets created by this operation, have not yet been determined. Unfortunately, these results are deceptive since the "stereo effect" is rather subtle.

To investigate further, the relaxed labels in Figure 3.4.2-1a and -1b were used for analytically recreating different gray level images. The gray levels of the 0-valued labels was replaced by average gray for all the 0-valued labels in the images. The results deteriorated dramatically, see Figures 3.4.2-5a and -5b. In order to see the errors the gray levels for the "missing labels" are taken from the original images and the pixels with the labels are shown as "average gray", see Figures 3.4.2-6a and -6b. The corresponding stereo pairs are shown in Figures 3.4.2-7a and -7b.

The same experiment was repeated with the "corrected" labels, see Figures 3.4.2-3a and -3b. Figures 3.4.2-8a and -8b show the analytic gray levels with average gray for the unlabelled pixels. Figures 3.4.2-9a and -9b show the errors with average grays for the labelled pixels. The corresponding stereo pairs are in Figures 3.4.2-10a and -10b. The results did not change significantly.

Next, the minimum number of "colours" was assigned to each facet in the two images such that adjacent facets in the two images have different "colours" (Figures 3.4.2-1a and -1b with conquered labels set positive). Six colours were needed. The "coloured" facets are shown in Figures 3.4.2-11a and -11b with the top left corner of the adjacency matrix in -11c. A print is

shown in Figures 3.4.2-12a and -12b, and the stereo pair in Figure 3.4.2-14a. The facets were also shown with average gray per facet, see Figures 3.4.2-13a and -13b, with the stereo pair in Figure 3.4.2-14b.

The profusion of displays in various combinations represents an attempt to "pin down" the critical aspects of a stereo image. The analytic results were or could be corrected and/or displayed in several ways:

1. Whenever the analytic value exists, i.e., the pixel has a label then, if the error "|G_ana-G_orig| > Limit" then eliminate this pixel (set the label to zero). This operation is equivalent to using another constraint on the output of analytic relaxation (a form of post-processing of analytic results).

2. A method that is similar to the first (1) but the labels are corrected only at the edges of the labelled regions. This has not yet been done.

3. A method that is similar to the first (1) or the second (2) but the analytic gray values for the pixels that are unlabelled or that "fail the test in 1" are replaced by the gray levels from the original image. The resultant gray level images were shown in Figures 3.4.2-2a and -2b, and as stereo pairs in Figures 3.4.2-4a. However, this method of presenting the results does not show the facets and prevent visual verification of facet shapes too see if the boundaries between the facets are approximately correct.

4. The adjacent facets are "painted" in different "colours". This was largely an attempt to bypass the weaknesses of the laser printer, but as seen, with little success.

All the results "point" in the same direction, i.e., the stereo appearance of these results is "disturbed" by the mismatching random "dots", by the differences in the details of the contours of the supposedly matching facets, and differences in adjacency relations between facets in the two images. Our stereo vision is greatly disturbed by relatively minor deviations between the facets and by inconsistent adjacency relations. The full implications remain to be investigated, especially for smooth contours without visible quantization effects, but the immediate conclusion is that pixel level correspondence is required (for highly spatially quantized contours and facets), and that adjacency relations cannot be violated (in a gross manner).

Figure titles:

Figures 3.4.2-1a, -1b, -1c: (1a) and (1b) show Lm˜(i,j) and Rm˜(i,j), the resulting facet labels after relaxation. The "acquired pixels" have been indicated by lower case letters to show the "conquests" made by the various labelled regions, i.e., the pixels conquered by facet E are indicated by e, etc. (1c) shows the number of pixels changed during relaxation (extract from history file). (Files: Inr093.fig, Inr094.fig, Inr092.fig).

Figures 3.4.2-2a, -2b: The analytically created gray level images after missing pixels and gross errors have been replaced.

Figures 3.4.2-3a, -3b: The "surviving" labels after analytic reconstruction with error limit of 15. (Files: Inr187.fig, Inr188.fig).

Figures 3.4.2-4a, -4b: (4a) A stereo pair of Figures 3.4.2-2a and -2b together with a stereo pair of the original gray level image for comparison (4b).

Figures 3.4.2-5a, -5b: The gray levels of the 0-valued labels replaced by average gray for all the 0-valued labels in the images.

Figures 3.4.2-6a, -6b: The errors when the gray levels for the "missing labels" are taken from the original images and the pixels with the labels are shown as "average gray". "Inverse" display compared to Fig. 3.4.2-5.

Figures 3.4.2-7a, -7b. The stereo pairs corresponding to Figs. 3.4.2-5 and -6.

Figures 3.4.2-8a, -8b: The analytic gray levels with average gray for the unlabelled pixels using the "corrected" labels in Figures 3.4.2-3a and -3b.

Figures 3.4.2-9a, -9b: The errors with average grays for the labelled pixels. The "inverse" to Fig. 3.4.2-8.

Figures 3.4.2-10a, -10b. The stereo pairs corresponding to Figs 3.4.2-8 and -9.

Figures 3.4.2-11a, -11b, -11c: (11a and 11b) "Coloured" facets from Figures 3.4.2-1a and -1b. The "colours" were assigned to each facet in the two images such that adjacent facets in the two images have different "colours". Six colours were needed. (11c) The top left corner of the adjacency matrix. (Files: Inr093b.fig, Inr094b.fig, Inr125a.fig)

Figures 3.4.2-12a, -12b: A print of the 'colored" facets shown in Figures 3.4.2-11a and -11b.

Figures 3.4.2-13a, -13b: The facets shown with average gray per facet.

Figures 3.4.2-14a, -14b: Stereo pairs, (14a) for Figure 3.4.2-12 and (14b) for Figure 3.4.2-13.

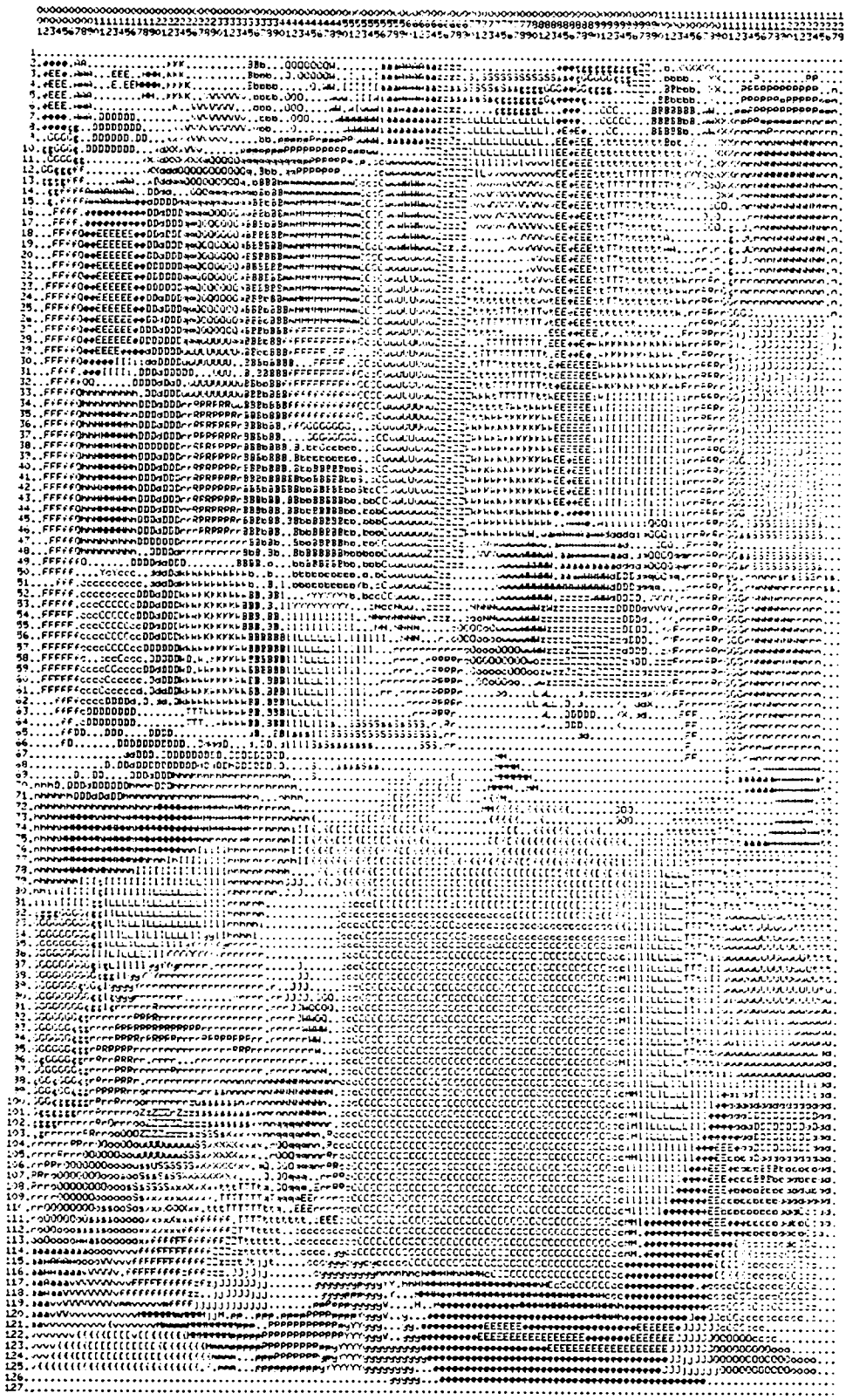
Left(i,j), left, facet labels after 10 iterations on initial parameters.

LABEL PRINTING

i=	0	j=	1	k=	2	l=	3	m=	4	n=	5	o=	6	p=	7
q=	8	r=	9	s=	10	t=	11	u=	12	v=	13	w=	14	x=	15
y=	16	z=	17	A=	18	B=	19	C=	20	D=	21	E=	22	F=	23
G=	24	H=	25	I=	26	J=	27	K=	28	L=	29	M=	30	,ETC.	
.=	0	@=	-1	a=	-2	b=	-3	c=	-4	d=	-5	e=	-6	f=	-7
g=	-8	h=	-9	i=	-10	j=	-11	k=	-12	l=	-13	m=	-14	n=	-15
o=	-16	p=	-17	q=	-18	r=	-19	s=	-20	t=	-21	u=	-22	v=	-23
w=	-24	x=	-25	y=	-26	z=	-27	A=	-28	B=	-29	C=	-30	.etc.	

File: lnr093.fig

Fig.: 3.4.2-1a.



RF(i,j), right, facet labels after 10 iterations on initial parameters.

LABEL PRINTING

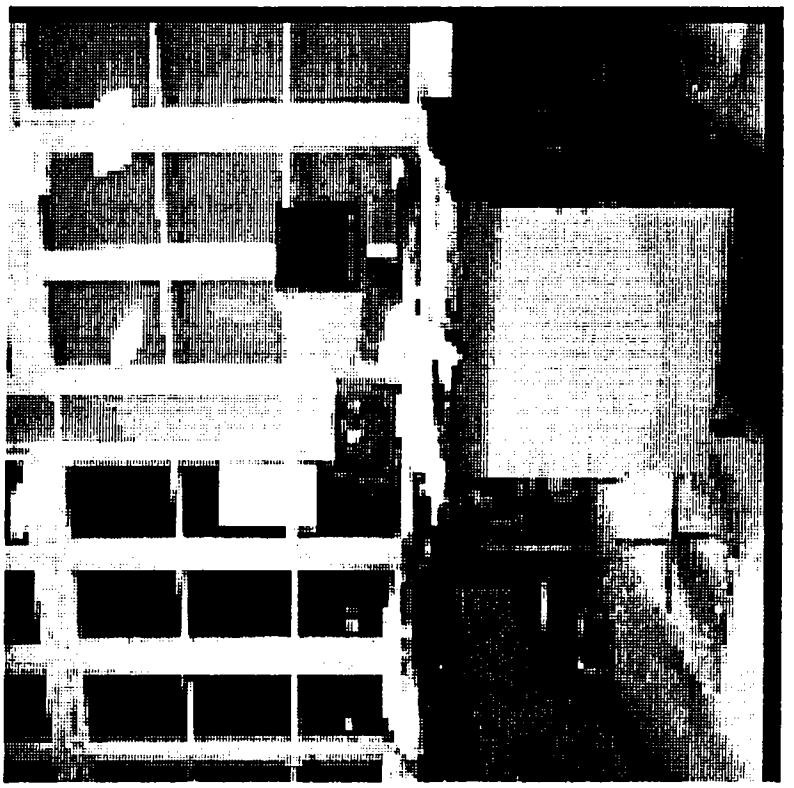
.= 0	== 1	C= 2	D= 3	E= 4	F= 5	G= 6	H= 7
I= 8	J= 9	K= 10	L= 11	M= 12	N= 13	O= 14	P= 15
Q= 16	R= 17	S= 18	T= 19	U= 20	V= 21	W= 22	X= 23
Y= 24	Z= 25	[= 26	A= 27	B= 28	C= 29	D= 30	.ETC.
.= 0	@= -1	c= -2	d= -3	e= -4	f= -5	g= -6	h= -7
i= -8	j= -9	k=-10	l=-11	m=-12	n=-13	o=-14	p=-15
q=-16	r=-17	s=-18	t=-19	u=-20	v=-21	w=-22	x=-23
y=-24	z=-25	[=-26	a=-27	b=-28	c=-29	d=-30	.etc.

File: Inr094.fig

Fig.: 3.4.2-1b.

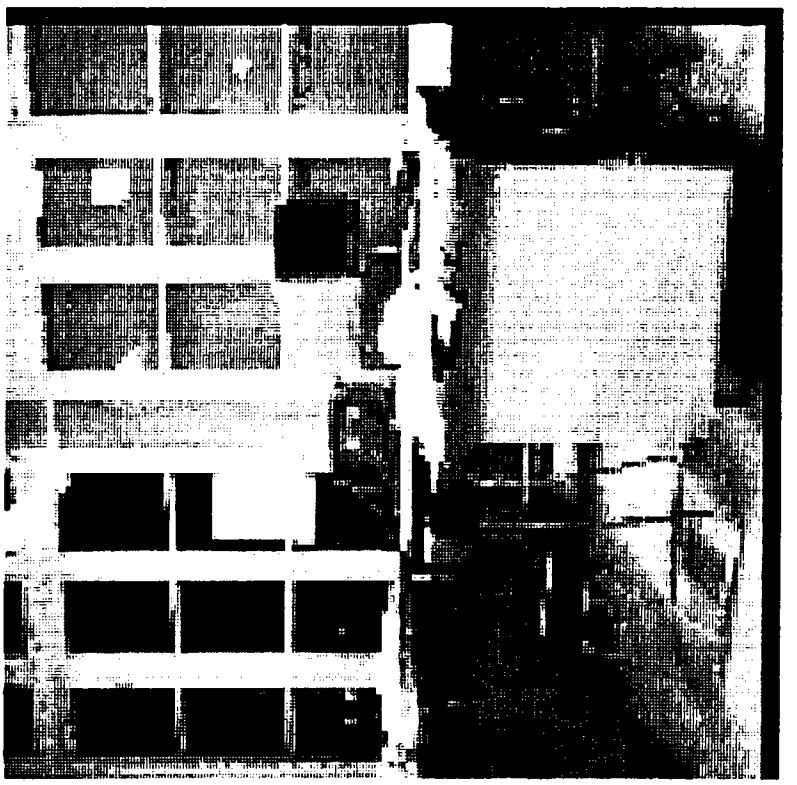
```
Analytic relax of matched facets.

MON397: ANA RELAX OF MATCHED PAIRS
IVAXL.IVAXR. IDIM.JDIM.KDIM =    55    56   129   127   500
JDIMW.KDIMW =   1   3
JSTART.JEND.JSTP.IFEDBK. =    1  127    1    0
IOPT.NCYCLE.KCHECT.NSTEPI.NSTEPJ.INDEF.KORDER = -1  10  4  7  2  0  2
ITEST0.ITEST1.ITEST2.ITEST3.LHBTEST =    0    0    0    0    0
IPT.JPT.FILMARK =    0    0    1
FMUL.WUEY =  0.10000E+03  0.10000E+01
WEIGHTG.WEIGHTD =  0.10000E+01  0.10000E+01
EPLIM.EPLIMO =  0.25000E+02  0.75000E+02

                                  NC      TOTL      TOTR       TOT
CYCLCHNG_PIX: NC.TOTL.TOTR.TOT =  1    2754.0    2715.0    3469.0
CYCLCHNG_PIX: NC.TOTL.TOTR.TOT =  2    1672.0    1694.0    3366.0
CYCLCHNG_PIX: NC.TOTL.TOTR.TOT =  3     921.0     942.0    1863.0
CYCLCHNG_PIX: NC.TOTL.TOTR.TOT =  4     602.0     558.0    1160.0
CYCLCHNG_PIX: NC.TOTL.TOTR.TOT =  5     415.0     357.0     772.0
CYCLCHNG_PIX: NC.TOTL.TOTR.TOT =  6     277.0     270.0     547.0
CYCLCHNG_PIX: NC.TOTL.TOTR.TOT =  7     203.0     213.0     416.0
CYCLCHNG_PIX: NC.TOTL.TOTR.TOT =  8     161.0     130.0     341.0
CYCLCHNG_PIX: NC.TOTL.TOTR.TOT =  9     120.0     140.0     260.0
CYCLCHNG_PIX: NC.TOTL.TOTR.TOT = 10      97.0     110.0     207.0


Extract of printout during analytic relaxation.
 Edited extract from history file: Inr091.fig).

NC   = Iteration count.
TOTL = Number of pixels changed per iteration in left image.
TOTR = Number of pixels changed per iteration in right image.
TOT  = Number of pixels changed per iteration in both images.

Weightg = Weight for error between analytic and actual gray values.
Weightd = Weight for error between analytic and actual gray value
          gradients.
Eplim   = Overall error limit. The formula is in Chapter 2.
Nstepi  = Di of the "box".
Nstepj  = Dj of the "box".

WEIGHTG.WEIGHTD.EPLIM =  0.10000E+01  0.10000E+01  0.25000E+02
NSTEPI.NSTEPJ =  7  2

File: Inr 92.fig
```

Fig.: 3.4.2-1c.



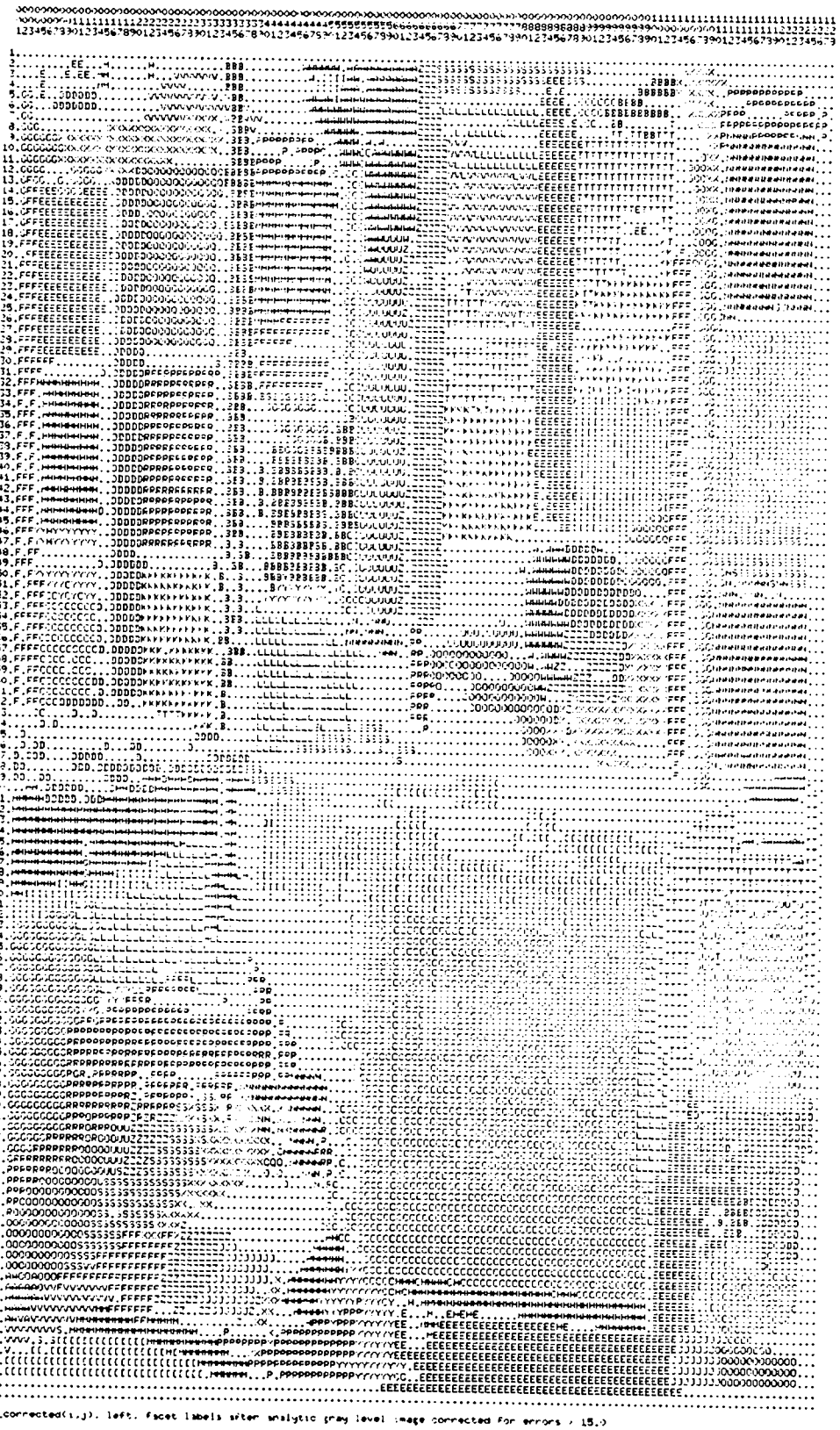
if8gacmc9.lgr

Fig.: 3.4.2-2a.



if8gacmc9.rgr

Fig.: 3.4.2-2b.



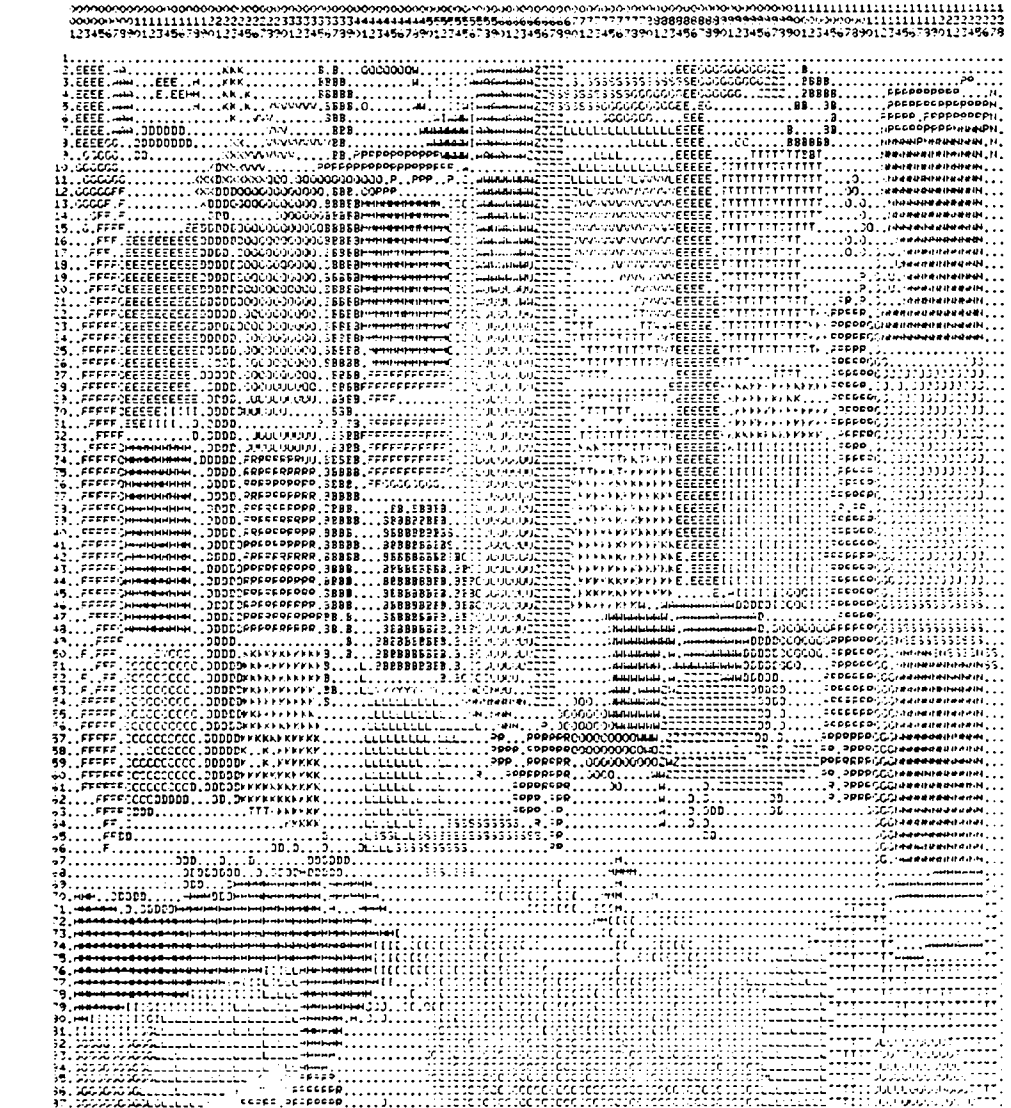
Lff_corrected(i,j). left. facet labels after analytic gray level image corrected for errors / 15.)

LABEL PRINTING

.=	0	==	1	C=	2	D=	3	E=	4	F=	5	G=	6	H=	7
I=	8	J=	9	K=	10	L=	11	M=	12	N=	13	O=	14	P=	15
Q=	16	R=	17	S=	18	T=	19	U=	20	V=	21	W=	22	X=	23
Y=	24	Z=	25	[=	26	\=	27	]=	28	^=	29	_=	30	.ETC.	
.=	0	`=	-1	a=	-2	b=	-3	c=	-4	d=	-5	e=	-6	f=	-7
g=	-8	h=	-9	i=	-10	j=	-11	k=	-12	l=	-13	m=	-14	n=	-15
o=	-16	p=	-17	q=	-18	r=	-19	s=	-20	t=	-21	u=	-22	v=	-23
w=	-24	x=	-25	y=	-26	z=	-27	{=	-28	|=	-29	}=	-30	.etc.	

File: lnr107.fig

Fig.: 3.4.2-3a.



Pf"_corrected(i,j), right. Facet labels after analytic gray level image corrected for errors - 15.0

LABEL PRINTING

.=	0	==	1	C=	2	D=	3	E=	4
I=	8	J=	9	K=	10	L=	11	M=	12
O=	16	P=	17	S=	18	T=	19	U=	20
Y=	24	Z=	25	[=	26	a=	27	b=	28
.=	0	9=	-1	c=	-2	d=	-3	e=	-4
i=	-8	j=	-9	k=	-10	l=	-11	m=	-12
q=	-16	r=	-17	s=	-18	t=	-19	u=	-20
y=	-24	z=	-25	{=	-26	a=	-27	b=	-28

F=	5	G=	6	H=	7				
Q=	14	R=	15						
W=	22	X=	23						
D=	30	.ETC.							
g=	-6	h=	-7						
o=	-14	p=	-15						
v=	-21	w=	-22	x=	-23				
c=	-29	d=	-30	.etc.					

File: Inv188.fig

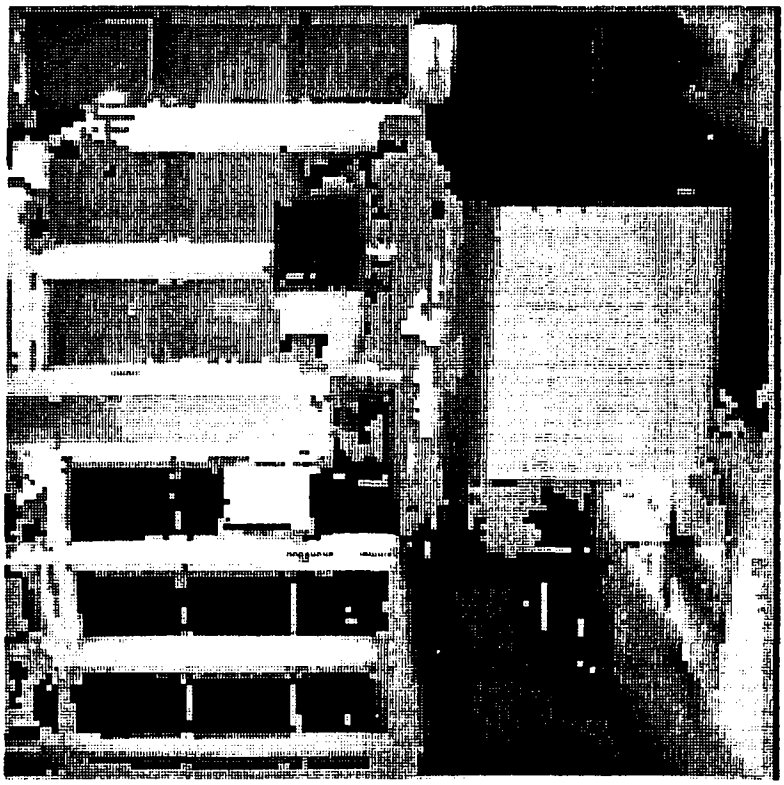
Fig.: 3.4.2-3b.

if8gacmc9.lgr if8gacmc9.rgr

Fig.: 3.4.2-4a.

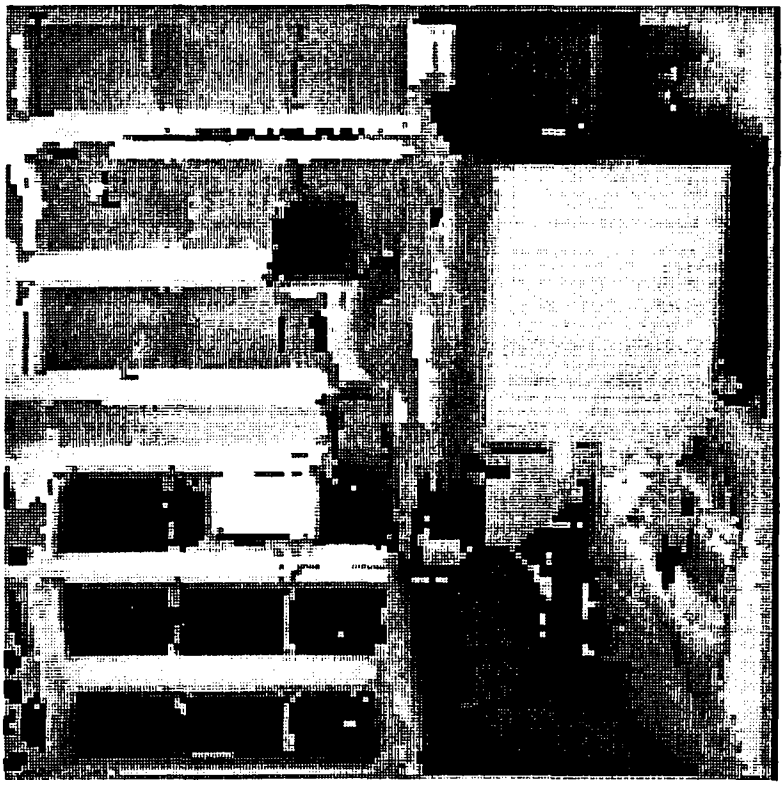
if8gray.lsh if8gray.rsh

Fig.: 3.4.2-4b.



if8ganmc.rgr

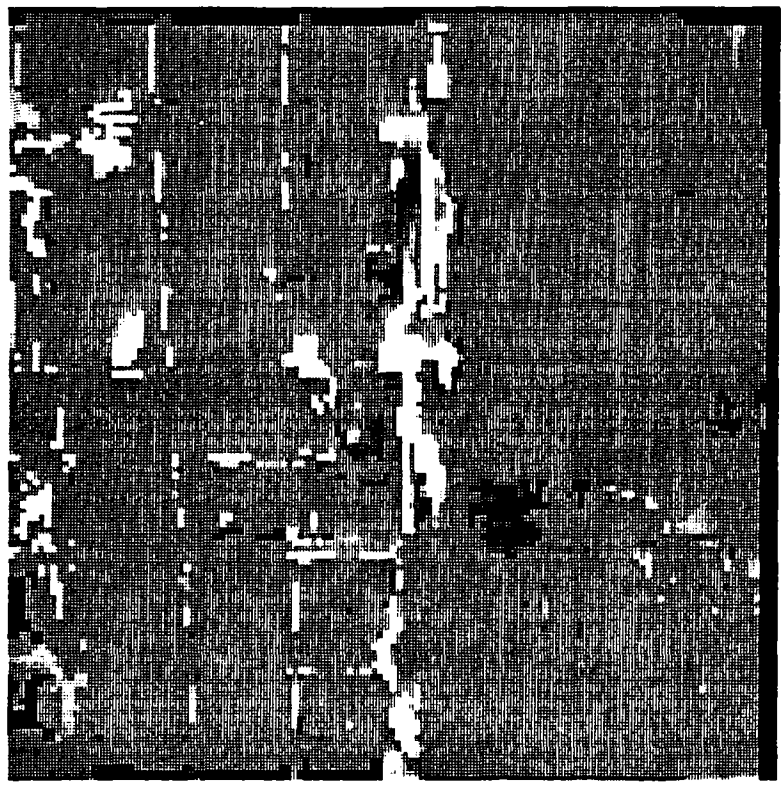
Fig.: 3.4.2-5b.



if8ganmc.lgr

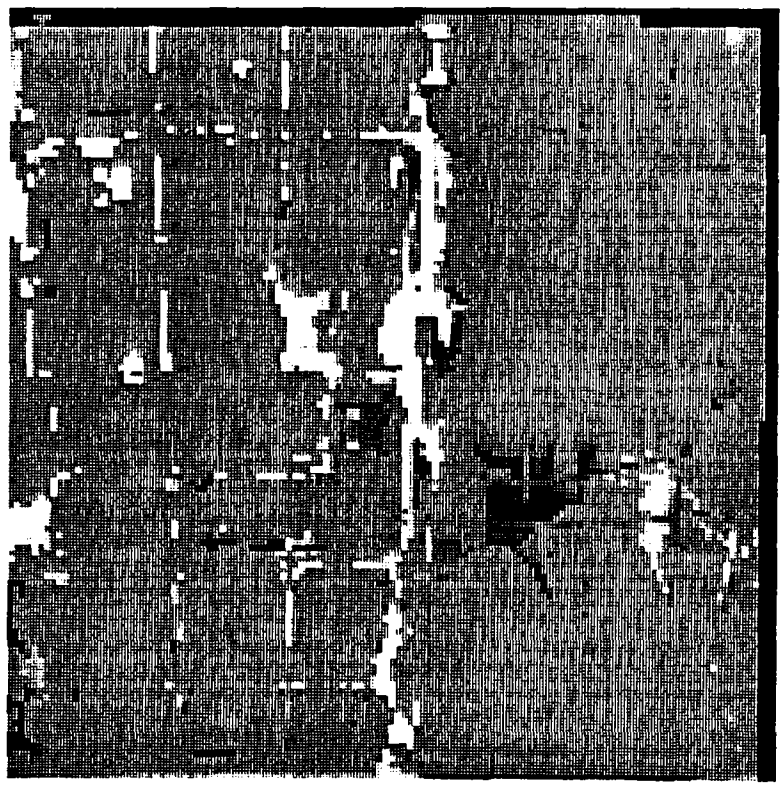
Fig.: 3.4.2-5a.

2 5,15,90



if8germc.lgr

Fig.: 3.4.2-6b.

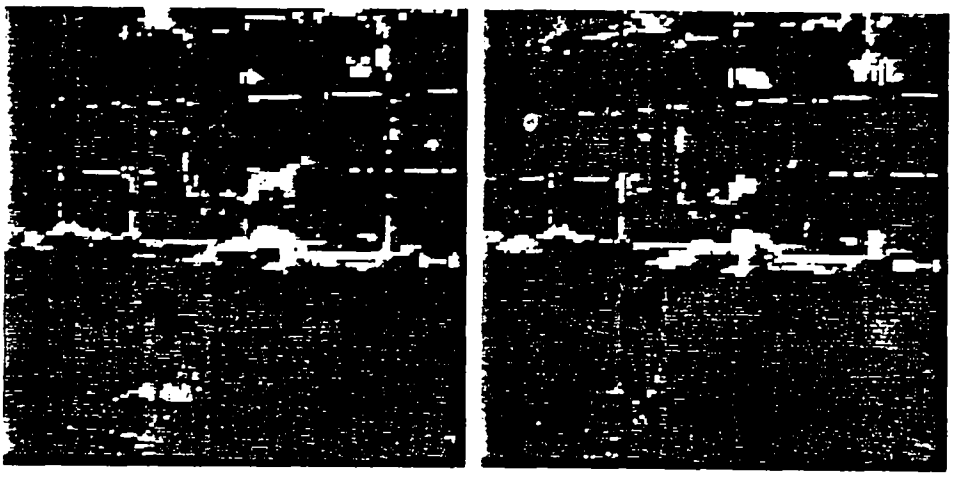


if8germc.lgr

Fig.: 3.4.2-6a.

if8ganmc.lgr if8ganmc.rgr

Fig.: 3.4.2-7a.



if8germc.lgr if8germc.rgr

Fig.: 3.4.2-7b.

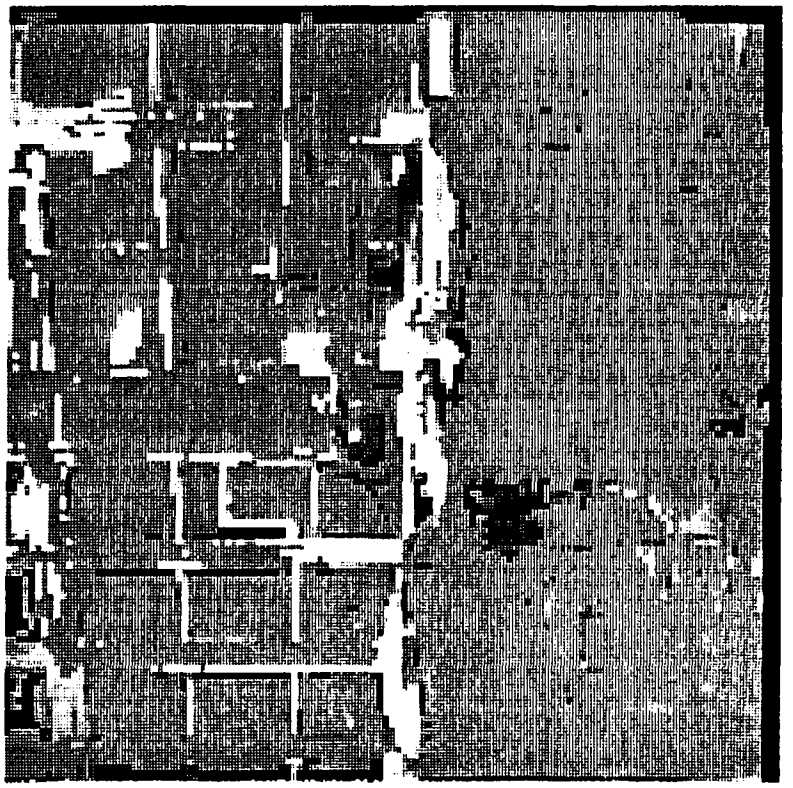
J. 7.15.90

if8flgac2.tgr

Fig.: 3.4.2-8a.

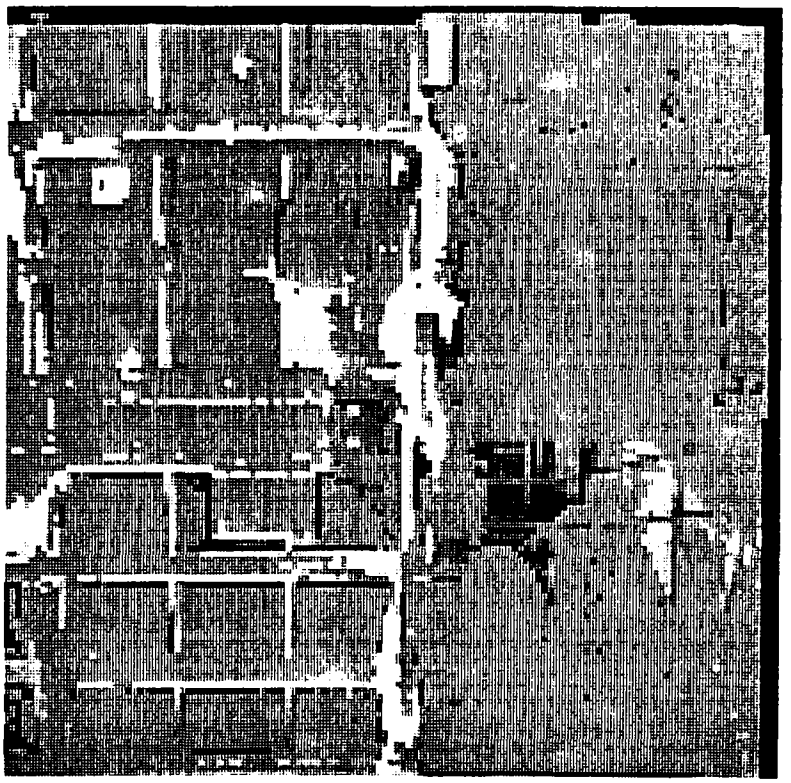
if8flgdc2.tgr

Fig.: 3.4.2-8b.



if8germc2.lgr

Fig.: 3.4.2-9b.

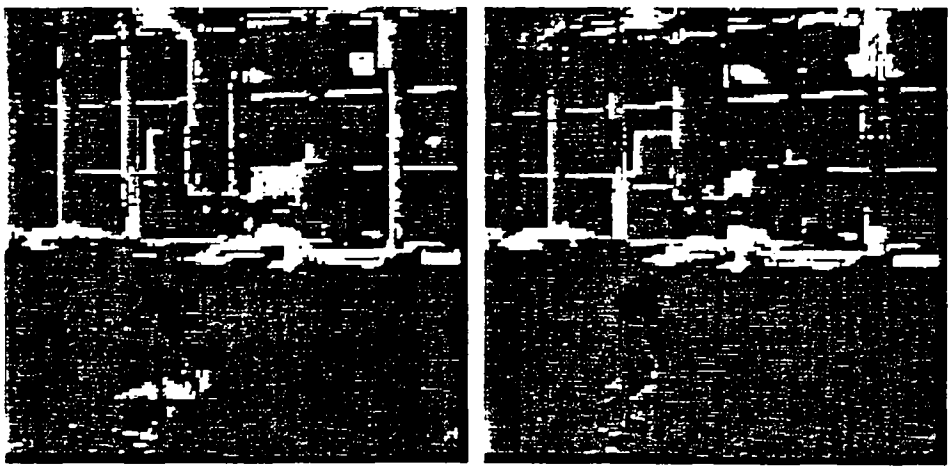


if8germc1.lgr

Fig.: 3.4.2-9a.

if8ganmc2.lgr if8ganmc2.rgr

Fig.: 3.4.2-10a.



if8germc2.lgr if8germc2.rgr

Fig.: 3.4.2-10b.

Fig.: 3.4.2-11a.

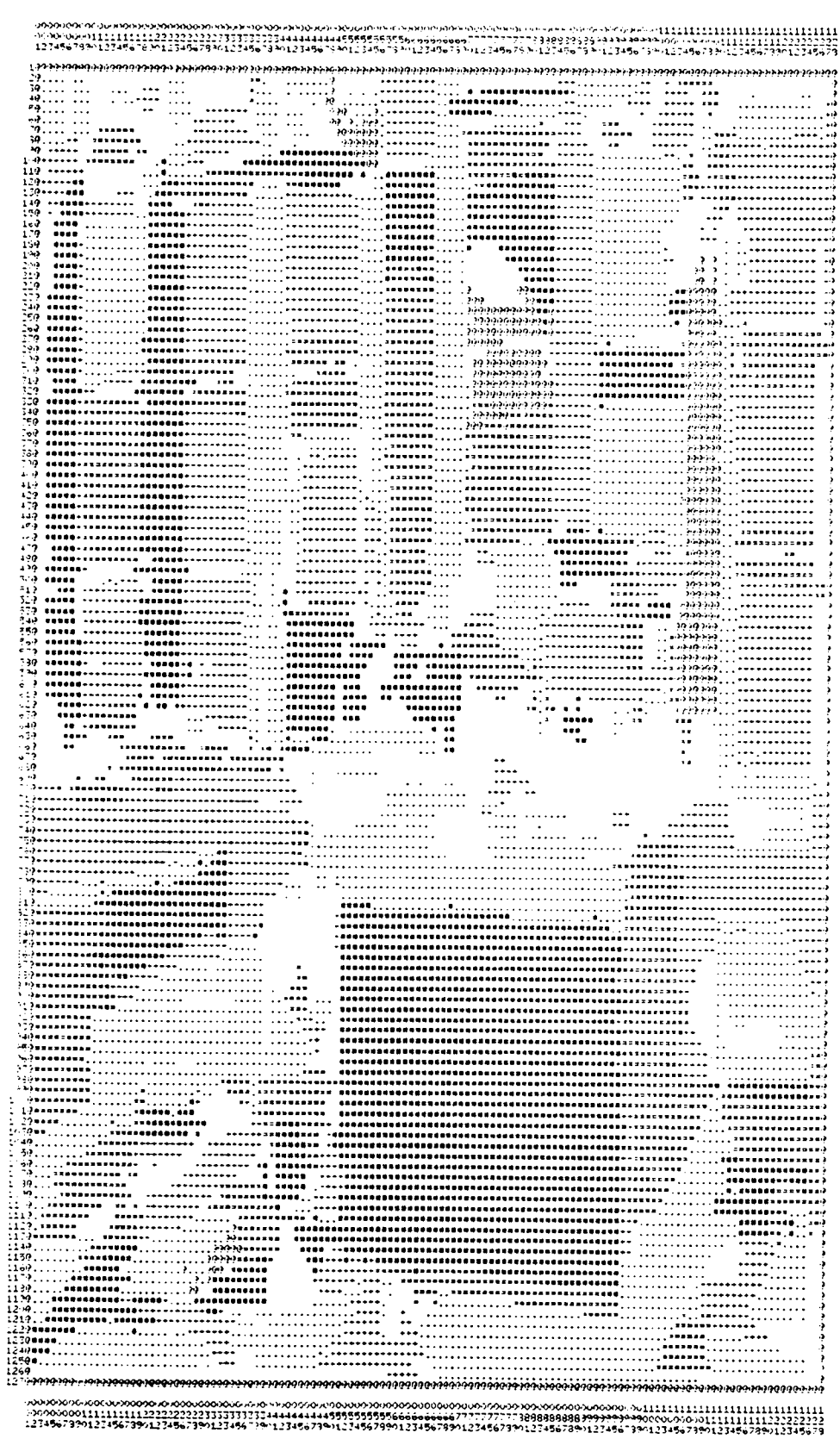


Fig.: 3.4.2-11b.

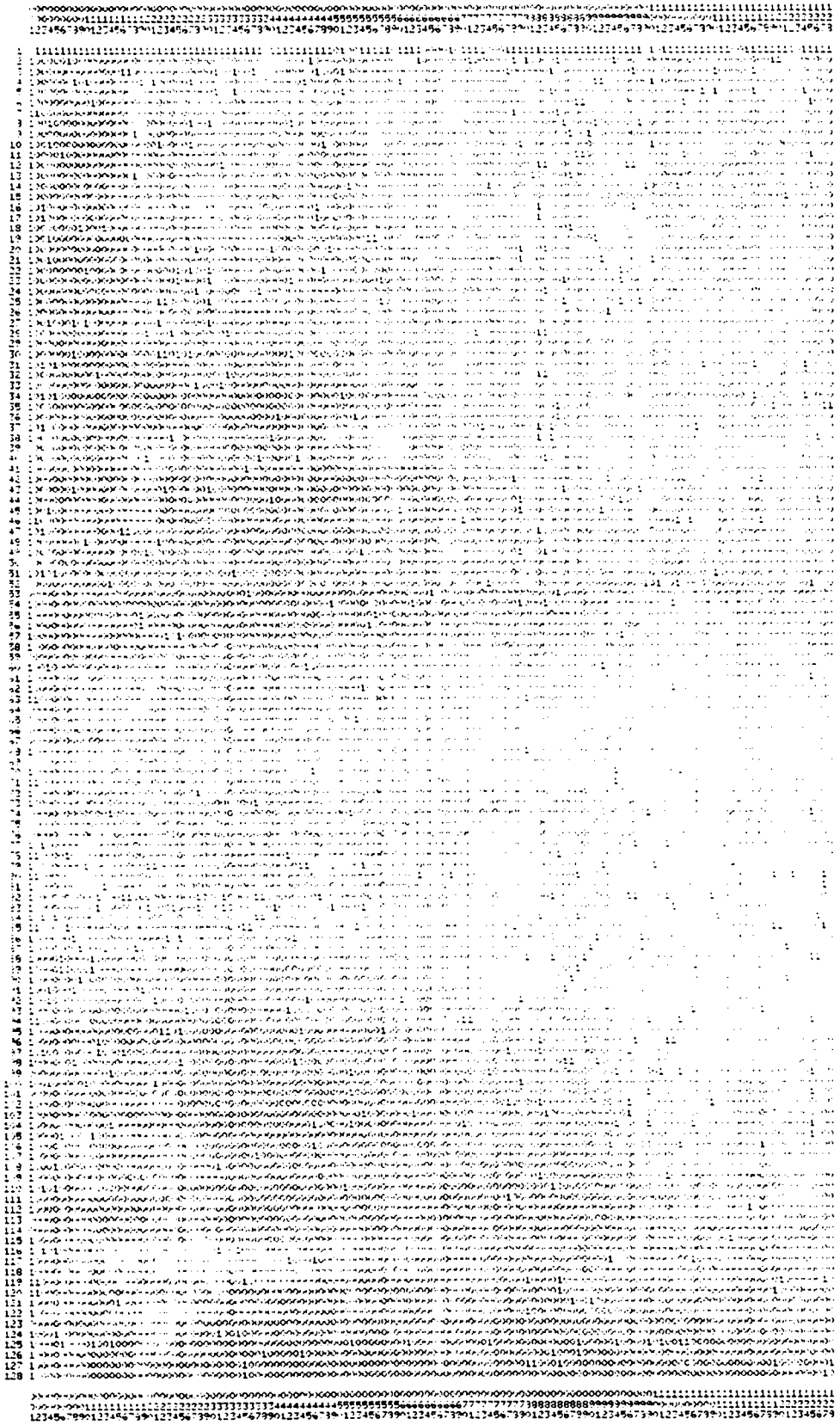
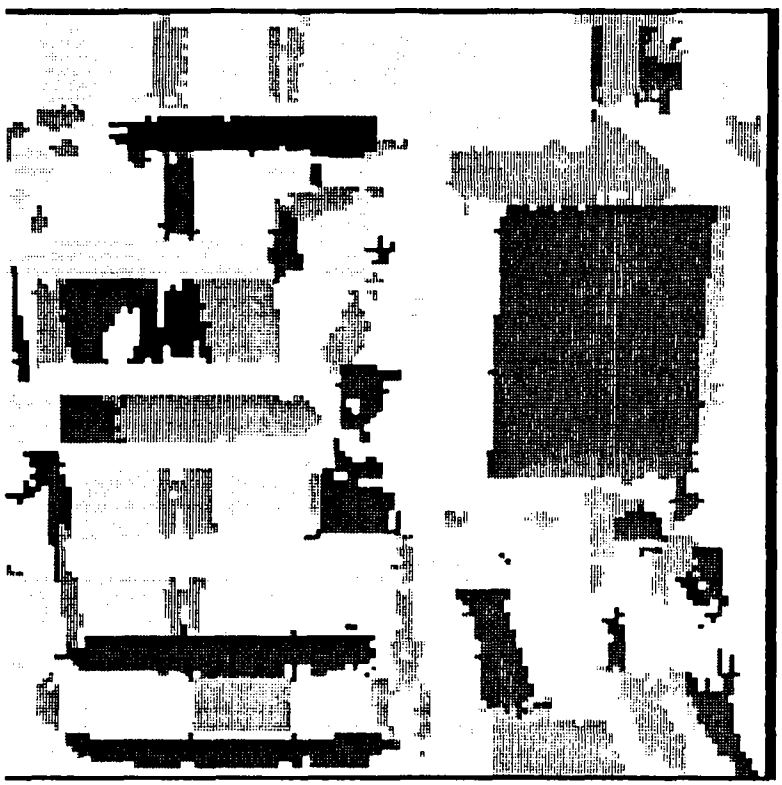
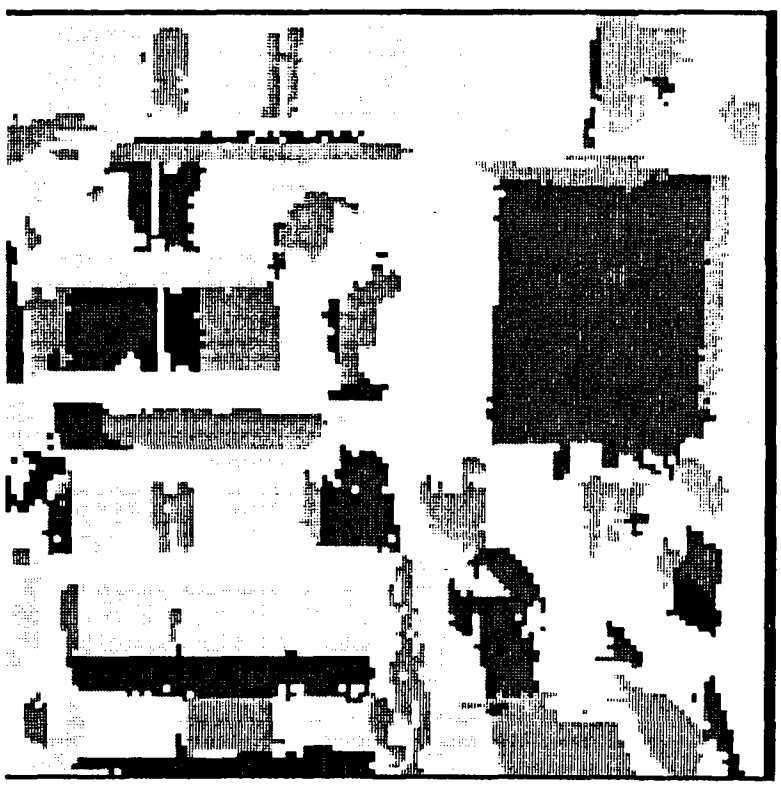


Fig.: 3.4.2-11c.



if8flkmc8.lgr

Fig.: 3.4.2-12a.



if8flkmc8.rgr

Fig.: 3.4.2-12b.

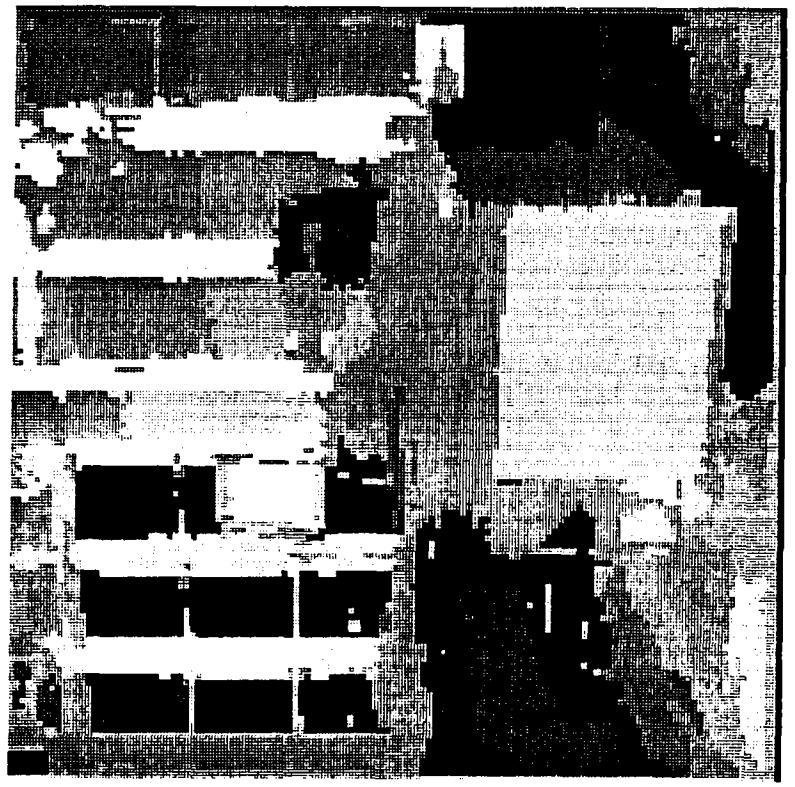


Fig.: 3.4.2-13b.

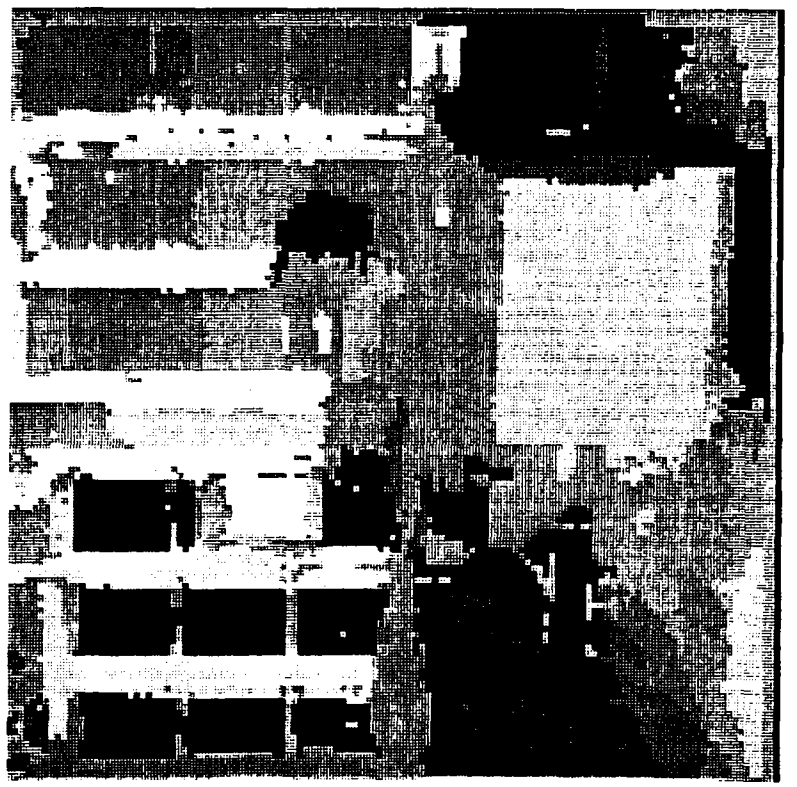
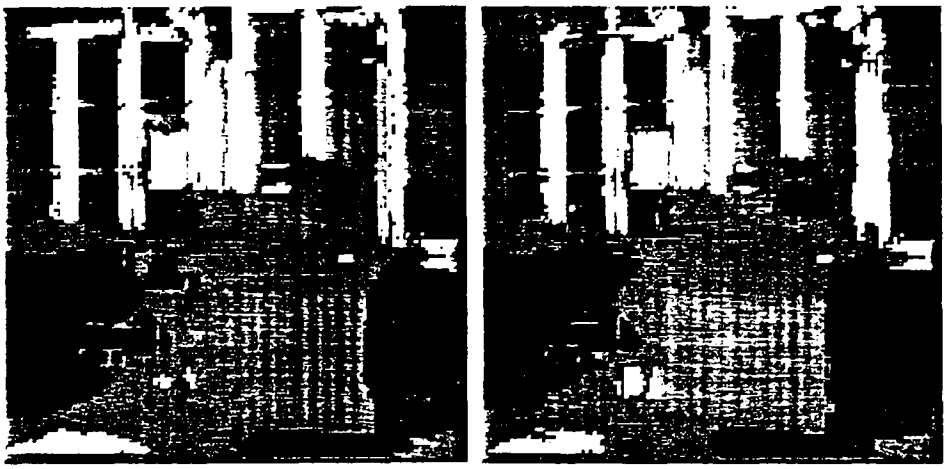


Fig.: 3.4.2-13a.

if8flkmc8.lgr if8flkmc8.rgr

Fig.: 3.4.2-14a.



if8flgmc~.lgr if8flgmc~.rgr

Fig.: 3.4.2-14b.

3.5 Comments

The processing steps following the matching are all well known, i.e., these consist of:

(a) Improving the edges since they are rather "ragged" and have violated the "similarity principle". The presently popular method is called "snakes". The better established methods are based on regularization.

(b) Segmentation of the edges and matching the edge pieces in the left and right images. If a piece-wise linear approximation is used then the results "interface" immediately to the edge based techniques. However, "edge segmentation" also has to obey the "similarity principle".

(c) A problem that requires "further thought" is how to guarantee valid adjacency relations and in what conditions are these relations "imperative".

(d) The interior of each facet is "just a smaller image". Consequently, the techniques in Chapter 2 apply again, but now on a smaller scale. However, there are also other methods, for example, the "rubber sheet" or "flexible template" approaches, as they used to be known.

However, instead of continuing, it is better to review and criticize the results since several weaknesses have become apparent.

REVIEW AND CRITIQUE

As described in the preceding chapters, the "similarity principle" is the "fundamental secret" for solving the 3D stereo problem since similarity requires no "higher level knowledge". For a considerable time the so-called "low level processing" of images has been considered either "well known requiring no further research" or "too low to be worthy of study". Immense efforts have been spent on "using high level knowledge" and on the application of "elegant mathematics" but without demonstratable results based on realistic images. Consequently, at least in the opinion of the present author, application of computers to image analysis, among other topics which are "natural and easy for ourselves", has been in a state of "Brownian motion without apparent drift" at least for the last decade or two.

In the present case all the "trends and fashions" were ignored and approximately half a year was spent on "thinking and experimenting". Brute computing power was available via the "INRIA net", ranging from "Suns to Connection machines", but a PC/AT was considered sufficient for preliminary experiments, especially since an image processing program library existed on the PC/AT. Considerable programming was done but most of the work consisted in modifying existing programs and assembling them into new running sequences. The experiments so far required 13 "machine loads" of programs (see "Inria*.for" in /scrt2/kasvand/ima3).

In brief, the processing steps may be summarized as follows, where the comments or observations have been included within the paragraph or inserted as "retrospective comments":

1. The computer is given two matrices (Gl and Gr) of integer numbers which are called the "left" (l or L) and the "right" (r or R) images of the stereo pair. No other information is neither available nor wanted since no "understanding" or other human concepts need to be evoked. The similarity principle is sufficient. Anything that is to be found is to be found from these two matrices of numbers.

2. These two matrices (Gl,Gr) are processed for gradients, which creates more matrices of numbers. Gl gives Gpml and Gpal, and Gr gives Gpmr and Gpar ("p" indicates gradient operator, "m" indicates magnitude, "a" indicates angle, and "l" and "r" are for "left" and "right" images). Notice that only two of the practically unlimited number of features have been used.

3. The pixels of these matrices (Gl,Gpml,Gpal and Gl,Gpml,Gpal) are clustered according to more or less classical pattern

recognition procedures. The regions in the images represented by the clusters are slightly "cleaned".

4. The connected regions of similar pixel classes are given "individual identity" (labelled) and the more "robust" regions are kept. The labelled regions are now called "facets".

5. The facets that produced an "intersection" (i.e., by "and-ing" the left and right images) are given similar labels in both images. Facets that could not be matched were also given new labels (negative), but these labels differed in the left and right images. This is a straight-forward application of well-known techniques and is but one of several methods for matching the left and right image regions.

Retrospective comments: The "new and unknown territory" begins at the end of step (3), in (4), and in (5). The "newness" is not in the techniques but in the modifications, improvements, and combinations required to make them "work better". The mistakes are, basically:

(a) The classical techniques are too "myopic", i.e., the 3x3 4-connected neighborhood is unsuitable (too small).

(b) More techniques should be used together to match the facets.

Now follows a long string of operations (steps 6 to 13) which were mainly studied for "curiosity" rather than necessity. In better quality images most of these operations are superfluous.

6. Some obvious misclassifications are corrected by not requiring class label coherence in the left and right images during matching.

7. The highly dissimilar but matched facets are "cut" to "reasonable" size by comparing them against each other. This step is justifiable in probabilistic terms but alternate methods should also be tried, for example, regions where the facets are very dissimilar should be reclassified.

8. Facet pairs which remained unmatched in "and-ing" even though they were very similar (since their intersection was zero or too small) are matched by recognition techniques. However, elementary pattern recognition methods can also produce many erroneous matches if many of the facets do not differ much from each other.

9. The facets are "clipped to the same size". This step was thought to be necessary to give a "fair start" to all the facets in the subsequent analytic relaxation procedure but otherwise has no strong logical justification. The use of

the centre of gravity of the "raw" facets, matching the centers of gravity, and "clipping" produces "good looking" facets but the adjacency relations between the facets are not properly preserved. Our stereo vision becomes "disturbed" by the "cracks" between the facets if the cracks do not match in the left and right images. Clearly, adjacency has to be preserved or recovered.

10. The facets that have remained unmatched are "transplanted" from one image into the other if the other image contains no labelled pixels at the same coordinate locations. This procedure, also, has only a probabilistic justification. The resultant facets, when viewed in stereo, "hang in free space in front of the scene".

11. The unmatched labels (negative) that still remain may be retained or eliminated by setting them to zero. Zeroing these labels has poor justification and should not be done but zeroing enlarges some of the "unknown" (zero-label) regions. This problem is at present unresolved.

12. The remaining unlabelled regions in the two images, if they "match", are extracted, "rounded" by binary processing, and labelled. This allows the "unknown facets" to "participate" in the analytic procedures but otherwise the procedure can only be justified in terms of probability.

13. The "unknown facets" found in the previous step (12) are put back into the left and right images. The left and right images are now rather well populated by "matched" facets.

The preliminary or "raw" matching of the facets in the left and right images is now complete. Most of the matches are correct and the matching method can be adjusted to the type of facet present, i.e., large facets can be "and-ed" for match, smaller facets require recognition techniques, including the adjacency relations to already matched facets, etc. In other words, this was a preliminary study to see what rather crude techniques can produce.

The next set of experiments are attempts to improve the shapes of the "raw" matched facets.

14. The gray level of each of the facets is approximated by an analytic function. A second order polynomial was used (since the programs were available from prior work).

15. The facets are now "relaxed" to try to assign as many pixels as possible to the various facets while preserving or recovering "stereo fidelity". The present relaxation program "ties together" the left and right images only by requiring that "Er < Erlim" in

Er = Wg*|G_org - G_ana| + Wd*|G´_org - G´_ana|, where

G_org = Original gray level value,
G_ana = Analytic gray level value,
G´_org = Original gray level gradient value,
G´_ana = Analytic gray level gradient value,
Wg, Wd = Weights,

provided that a pixel at (i,j) in the left image had an "homologue" which was within a box (i = +/-di, j = +/-dj; di,dj = 2,7) and the homologue was within the error tolerance (Erlim). This constraint was far too weak.

Retrospective comments: The list of mistakes made is rather long. Some of these "mistakes" were deliberate in order to see if it is possible to "get away with" some crude method. Briefly summarized:

a) The analytic approximations of the gray levels of the facets did not correspond to functions for surface brightness.

b) No analytic approximation was used for the edges of the facets and thus the shapes of the edges could not be used.

c) During relaxation decisions were made at pixel level without including sufficient information from the neighborhood around each pixel.

d) The relaxation procedure was not sufficiently constrained and, consequently, produced rather "noisy" results.

e) Most importantly of all, the relaxation procedure did not properly incorporate the similarity principle.

The present study ends at this stage since:

i) The processes in Chapter 3 require modifications in light of the experience acquired and the remaining time is insufficient to complete the changes for this report.

ii) The results can serve as starting points for numerous edge based techniques. The combinatorial explosion problem has been eliminated since only the edges of each of the matching facets have to be considered rather than all the edges in the image at the same time (1.2, 4.1).

iii) Each matched facet may be looked upon as a small image and processed in the same way. Thus, the process may be iterated for each facet in order to achieve the next level of correspondence.

References

(The author´ apologizes. Time did not allow a careful search and review of the available material. For additional refs., see 1.1, 1.2, 4.1, and books entitled "... pattern recognition ...".)

1.1) Manual of Photogrammetry, Fourth Edition, C. C. Slama, C. Theurer, and S. W. Henriksen, eds. 1980.

1.2) Faugeras, O., A few steps towards artificial 3D vision, INRIA, research report No. 790, February, 1988.

1.3) Julez, B., Foundations of Cyclopean Perception, University of Chicago Press, 1971.

1.4) Gibson, J.J. The Perception of the Visual World, Houghton Mifflin Co., 1950.

1.5) Horn, B.K.B., Obtaining shape from shading information. In P.H. Winston (Ed.), The Psychology of Computer Vision, McGraw Hill, 1975, pp. 115-155.

1.6) Grossberg, S., and Marshall, J.A., Stereo boundary fusion by cortical complex cells: A system of maps, filters, and feedback networks for multiplexing distributed data, Neural Networks, Vol. 2, 1989, pp. 29-51.

2.1) Peak viewer 2x, No.1994-2, made in Japan, approximate cost is $40.- Canadian.

2.2) Kass, M., Computing visual correspondence, in From Pixels to Predicates, A.P. Pentland, ed., Ablex Pub. Co., 1986, pp. 78-92.

2.3) Otsu, N., and Kasvand, T., A new family of nonlinear edge detectors for noisy images. Paper #435-02. SPIE.

2.4) Forsynth, D., Mundy, J.L., Zissermann, A., and Brown, C.M., Projective invariant representations using implicit algebraic curves, Computer Vision - ECCV90, pp. 427-436.

2.5) Kasvand, T., Extraction of lines of unspecified texture from unconstrained line drawings, Science of Form: Proc. 1st Internat. Symp. on Science of Form, Y. Kato, General editor, KTK Scientific Pub., Tokyo, 1986, pp. 421-430.

3.1) Peyret, M., and Gagalowicz, A., Reconstruction 3D basee sur une analyse en regions d´un couple d´images stereo, Inria, Aug. 1989.

4.1) Remion, Y., Stereovision par zones, outils et structures d´un systeme expert, these docteur ingenieur, l´Ecole National Superieure des Telecommunications, 8 Juin 1988.

Appendix: IMAGE FILES

All the results shown in this report, including results not discussed, are available as computer files in the "Bora" system in /user2/kasvand/ima2/ for images and /scrt2/kasvand/ima3/ for some of the programs and the text of this report.

1. Text of report. The text has been written using PC/AT Wordperfect 4.2 but stored in Bora without the control codes (#1 foreign file output mode). The control codes have to be inserted for other types of word processors.

   ich0p0n.prn = Beginning of report.
   ich1p0n.prn = Chapter 1.
   ich2p0n.prn = Chapter 2.
   ich3p0n.prn = Chapter 3.
   ich4p0n.prn = Review to end of report.

2. Images in one pixel per character alphabetic or alphanumeric form and "history files" of the runs for this report. Normal printing method is via X11 window using "small" print and "cat file_name". The images produce two pages of print which have to be "glued together" to produce the results shown.

   inr000.fig = Index to inr*.fig and to image files.
   inr*.fig = Image or history files, see index (inr000.fig).
   * = 001 to approximately 200.

3. Display files, 256x256 one byte. These may be displayed on the Vicom image display system or printed on the laser printer using ~/im_laser -T -s n (-inv -xcen -ycen) file_name. The size of the image is defined by "n", n =1, 2, ..., 6. Size 5 (n=5) is used for the "big" prints and size 3 (n=3) for the stereo pairs. Unfortunately, the gray level quality is very low and sinks with increasing n.

   if8*.l* = Left image of the pair.
   if8*.r* = Right image of the pair.
   The if8*. part is the same for the left and right image.
   The .l* and .r* portions may sometimes be inconsistent.
   ifort8*.l* = Some older file, left image of pair.
   ifort8*.r* = Some older file, right image of pair.
   n*.* = Normalized display file (Inrimage "norma" used).

4. Composite display files (512x512 by one byte) for the Vicom system. These were intended for showing integrated results but were discontinued due to lack of disk space.

   pvic*.l51 = Left image, composite of four results.
   pvic*.r51 = Right image, composite of four results.

Imprimé en France
par
l'Institut National de Recherche en Informatique et en Automatique.

ISSN 0249-6399